Technical Memorandum

FDOT Master University Agreement BDV29-977-55

A Synthesis on Data Mining Methods and Applications for Automated Fare Collection (AFC) Data

Final Report

Prepared For

Planning and Environmental Management Office

Florida Department of Transportation - District 6

Prepared By

Fabian Cevallos, Lilibeth Turino, and Xia Jin

Florida International University

10555 E Flagler St., EC 3603

Miami, FL 33174

Date

May 2021

DISCLAIMER

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation or the U.S. Department of Transportation.

Prepared in cooperation with the State of Florida Department of Transportation and the U.S. Department of Transportation.

METRIC CONVERSION CHART

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 L shall be shown in m^3				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or metric ton)	Mg (or t)
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F−32)/9 or (F−32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	pound force	4.45	newton	N
lbf/in^2	pound force per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or t)	mega grams (or metric ton)	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newton	0.225	pound force	lbf
kPa	kilopascals	0.145	pound force per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Synthesis on Data Mining Methods and Applications for Automated Fare Collection (AFC) Data		5. Report Date May 2021 6. Performing Organization Code
7. Author(s) Fabian Cevallos, Lilibeth Turino, and Xia Jin		8. Performing Organization Report No.
9. Performing Organization Name and Address Florida International University 10555 W. Flagler Street Miami, Florida 33174		10. Work Unit No. (TRAIS) 11. Contract or Grant No. BDV29-977-55
12. Sponsoring Agency Name and Address Florida Department of Transportation 605 Suwannee St., Tallahassee, FL 32301		13. Type of Report and Period Covered Final Report January 2020 - May 2021 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Automated Fare Collection (AFC) systems have been widely adopted by transit agencies around the world. AFC systems can provide a bias-free and low-cost approach to obtaining long-term continuous observations of transit passenger travel behavior. The availability of AFC data can help examine the behavior of passengers and the demand characteristics for public transportation services, which can be useful to support short-term planning and the development of long-term strategies.

This research project provides a broad review of existing knowledge and experience in utilizing AFC data, from both the practice and research perspectives. After an extensive review of the literature, it is clear that there is a lot of research material in this area. However, there is still a need for an efficient, comprehensive, and user-friendly system to fully take advantage of the AFC data.

An analytic and visualization tool that can mine the AFC data in combination with other datasets can be used to analyze passenger behavior and the demand characteristics of public transportation. Therefore, the research team presented a framework for the development of a data mining tool to analyze passenger behavior and the demand characteristics of public transportation. The proposed tool can be useful for general data analysis, but it can also be used for granular data analysis at a higher level of detail. The proposed tool can extend the utility of AFC systems and help transit agencies and planning organizations to have the necessary data and information to plan and deliver a more efficient and equitable transit service.

Future research is needed to better understand the spatial and temporal dynamics of transit demand. In addition, there is a lack of effective tools and sophisticated techniques to take full advantage of the AFC data. Therefore, the role of Transit ITS vendors and software developers need to be explored to improve transit service and promote the use of public transportation.

17. Key Words		18. Distribution Statement No restrictions	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 212	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

EXECUTIVE SUMMARY

Automated Fare Collection (AFC) systems, also referred to as Automatic Fare Collection or Electronic Fare Payment (EFP) systems, have been widely adopted by transit agencies around the world. AFC systems can provide a bias-free and low-cost approach to obtaining long-term continuous observations of transit passenger travel behavior. While providing time savings, convenience, and other advantages in revenue collection, they also produce massive, continuous, and anonymized digital records of fare transactions. The availability of AFC data and information can help examine the behavior of passengers and the demand characteristics for public transportation services, which can be useful to support short-term planning as well as the development of long-term strategies.

Many studies have taken advantage of AFC data and conducted research from several perspectives, including ridership statistics and performance indicators, to support operational analysis, travel pattern identification, and network analysis for service planning and behavior analysis to facilitate long-term planning for public transportation. The literature also noted some challenges in dealing with AFC data, including the lack of trip purpose information, lack of demographic information of the users of the transit system, and in bus systems the lack of alighting locations.

This research project aims to provide a broad review of existing knowledge and experience in utilizing AFC data, from both the practice and research perspectives. From the practice perspective, the project examined whether and how transit agencies have utilized AFC data to support planning and operational activities. From the research perspective, the project innovative data collection and analysis methods and applications were explored and summarized.

Summary of Literature Review

The research team realized the existence of a large number of research papers from the academic side that explore many areas that address the potential of transit smart card data such as passenger behavior analysis and market segmentation, system performance assessment, impact analysis of policy changes or service improvement, and data processing (such as trip purpose inference, etc.). Earlier projects were mostly focused on implementation and operational issues of smart card and AFC technologies, while newer studies have focused more on the uses of AFC data, and they can be wide-ranging. In general, transit agencies can benefit from the efficient use of AFC data to better understand the service needs of the transit users. For instance, AFC data can be used to assess the needs of special populations like older adults and people with disabilities by examining the preferred origins and destinations or the days and times of transit use. Furthermore, combining AFC data with other transit intelligent transportation systems (ITS) data can enhance the datasets available for decision making. To provide ideas on the potential uses of the AFC data, the following subjects were included: improving services for special populations, smart card applications, fare evasion, measuring activity similarity, inferring origin-destination demand, assessing transit loyalty, uses of electronic fare payment records, data for bus service and operations planning, travel behavior analysis, and mobility patterns of seniors, children and students, and adults.

In addition, the use of data mining and big data techniques have been widely recognized as a powerful instrument to analyze large-scale data, identify patterns, and derive meaningful information that can be used to support planning and predictive analysis. The use of big data can be grouped into six categories: Service/Performance, Travel Behavior, Travel Demand, Management, Resilience and Health/Safety, and Other Topics.

The selected work in this research study also included relevant North American and International experiences in the use of AFC data. The research identified innovative uses of the data and results from research studies in section 2.4. In this area, it is worthwhile mentioning the work of international researchers, as there is a vast amount of information that covers a variety of useful topics relevant to the use of AFC data. Although there is a wide variety of topics to cover, the areas captured in this research study include some of the most investigated areas, such as:

- Understanding fare increases
- Trip purpose inference
- Fare structure and social vulnerability
- Prioritizing bus schedule revisions
- Carbon emissions
- Jobs and housing relationships
- Commuting patterns
- Extracting boarding information
- Transport information services
- Fare collection interoperability
- Policy and planning
- Transit assignment modeling
- Analysis of transit service performance

Overall, the topics outlined in this research represent a good depiction of the available topics on the use of AFC data. The work in this research can help understand the challenges and opportunities researchers and practitioners face with the efficient collection, handling, and applications of the data from fare collection systems. Besides the identified topics for the potential use of AFC data, below are some observations from this study:

- AFC systems from transit agencies generate large amounts of data as part of their daily operations.
- The use of data can provide valuable insights to improve efficiencies in the delivery of public transportation services. However, the collected data are not being fully utilized in the planning and decision-making process.

- Combining AFC data with other ITS data presents many opportunities. Transit agencies can benefit from using enhanced datasets with information from the ITS systems. Therefore, there is a need for database systems that can take advantage of data from the all the existing ITS systems. To use effectively and efficiently the data from these systems, there is a need of sophisticated optimization models and decision support tools.
- The work in the advanced use of AFC data has been more in the academic research context. So, there is a need for technology transfer mechanisms for the research to be effectively transferred to practitioners for their use.
- There are many opportunities for using the data to improve the transportation services, but for this to materialize it requires long-term investment in infrastructure and talent. Further, there is a need of a roadmap that may require the support of county, state, or federal agencies. In addition, vendors and developers of these system may need good incentives to develop tools or systems that can be costly.

Case Studies

This section enhances the literature review by identifying three case studies. These are locations where transit agencies have utilized AFC data to develop approaches, methodologies, technologies, applications, tools, etc. to help improve their efficiencies. This was based on the information available through documentations, project reports, and research papers. The selected locations of the case studies are New York, Massachusetts, and Utah with their corresponding transit agencies: Metropolitan Transportation Authority (MTA), Massachusetts Bay Transportation Authority (MBTA), and the Utah Transit Authority (UTA).

These three case studies were selected for the diversity of relevant topics and the availability of the literature regarding the use of AFC data from the transit agencies at these locations. The topics vary from each agency, which offers a good overview on the research and studies that deal with the use of AFC data in different areas. They cover topics such as the general use of AFC data, the use of fare data in the decision making, and fare payment systems.

Discussion

The topics identified in the literature review and case studies cover different areas of public transportation. They provide many examples of how transit agencies can benefit from the efficient use of AFC data to better understand the service needs of the transit users. The following items highlight a summary of the key topics that were identified by the research team, as they relate to this project.

The Age of Big Data. Because of the increasing demand for high quality data, agencies should have the necessary resources to perform data analysis and create visualization for a better communication and comprehension with the stakeholders regarding the transit system condition, emphasizing the significance of the capital investments. The following are some of the tools that can be used to take advantage of the data and present the information to stakeholders:

- Business Intelligence and Analytics. Business intelligence (BI) is the term applied to the ability of an organization to collect, maintain, and organize data. The BI technologies can provide historical, current, and predictive information on business operations by transforming raw data into meaningful and useful information, which can be used to inform more effective strategic, tactical, and operational insights and decision-making.
- Data Visualization. Visualizations could provide a better understanding of the collected data by using pictures, images, or animations depending on the best way to communicate with an intended audience.

Using Integrated Electronic Data for Service Planning. With the collection of data from transit ITS systems, transit agencies should have the necessary information to improve the service in an efficient way. The information about riders boarding and alighting allows for estimations of passenger impacts on route re-designs. Replacing a manual collecting process with an electronic one enhances the reliability and accuracy of the information collected. It also provides a better picture of the condition of the transit system. Therefore, electronic data can be used to allocate the necessary resources to efficiently manage the system and allow innovative analysis to keep upgrading the system and the quality of the transportation services.

Visualizing Transportation Networks. Using different types of visualization can assist transit professionals with the improvement of the transit system. Visualizations can serve as a window to the data, assisting a particular audience with focusing on the right amount of information for their needs. Therefore, target audiences such as transportation consumers, transportation administrators, and civic interest groups can benefit from different transportation visualizations methods.

Automated Data to Improve Decision Making. Transit agencies can organize and use the existing AFC data to better understand the potential impacts of fare options on ridership and revenue. Better information about AFC data and the implications of fare changes enable transit agencies to support a more robust public debate, a higher degree of accountability, and ultimately wiser decision making.

Market Research and Demand Modeling. The implementation of automated fare collection (AFC) systems provides new opportunities for improving transportation services. Data can be mined to create inputs to operations planning and demand forecasting models. The information provided by these systems also allows the agencies to formulate questions about how current fare products are being purchased and used and how have those patterns changed over time to help them understand how these products are changing transit services. In addition, data can be used for demand modeling. The key considerations for selecting an appropriate model include the availability of historical data, the attributes that define the current and proposed fare products, and the availability of origin-destination-transfer information. Lastly, AFC data can help assess the performance of the system using market segmentation to better take into consideration the needs of different segments of the population.

Transit Equity. The collected data from the AFC systems open many possibilities for transit planners on developing different methods that can facilitate the prediction of people origins, destinations, and transfers when traveling. An origin-destination (OD) prediction can be used in the analysis and reporting of agencies' social goals, such as the provision of equitable service regardless of race, national origin, or ethnicity, which is federally required in the U.S. by Title VI of the Civil Rights Act of 1964. Title VI prevents agencies receiving federal funding from having a disparate impact with regard to race, ethnicity, or national origin. In complying with this law, transit agencies must report regularly on how their service is provided to populations with different demographics. Fare equity is a topic that will help planners and policymakers to design a better infrastructure that serve people needs and demands.

Transit Fares for Disadvantaged Populations. It is expected that by 2050 there will be a significant growth in the global elderly population. This projection will not only mean that the current lifespan will be extended, but it will bring as a consequence many changes, especially in the transportation services that need to be implemented to be able to cover requests and demands that are expected to greatly increase. For this reason, transit agencies must take actions to provide better transit and affordable options to the elderly population, in particular regarding disadvantaged segments of that population.

Recommendation for a Data Mining Tool

It is clear that there are a lot of research material in this area, where specific methodologies and tools have been developed mainly to support the planning and operations of transit agencies. However, there is still a need for an efficient, comprehensive, and user-friendly system to fully take advantage of AFC data. An analytic and visualization tool that can mine the AFC data in combination with other datasets such as General Transit Feed Specification (GTFS), census, and land use data can be used to help understand passenger behavior and the demand characteristics of public transportation. A framework for the development of a data mining tool was developed. The proposed tool can be useful for granular data analysis. It can be used to extend the utility of AFC data to better understand passenger travel behavior by fare type and identify particularities in the patterns of passenger travel. It can provide visual analytics of the transit system performance based on the AFC, GTFS, census, and land use data. The tool can support transit planning at the operational, tactical, and strategical levels. If developed, the web-based system will incorporate data integration, visualization, analysis, and query capabilities to demonstrate a variety of potential applications of AFC data in support of transit planning and operations activities. Sophisticated data mining and visualization techniques will be employed to help identify patterns, trends, and the general dynamics in passenger travel that can be used in the overall decision making process. This can help transit agencies have the necessary information to plan and deliver a more efficient and equitable service.

Future Research

Based on the lessons learned from this study, the research team suggests the following areas for future research:

- To better understand the spatial and temporal dynamics of transit demand, there is a need to capture and monitor the trends and impacts on the transit system. This is needed to address issues like the growing and aging of the population, the increasing congestion, and the rapidly evolving technologies and mobility services.
- Most transit agencies in the U.S. collect data from the electronic farebox and other transit ITS systems. However, there is a lack of effective tools or mechanisms that allow for the extraction and efficient use of data and information, so this is an area that needs particular attention.
- The development of sophisticated techniques is key for taking full advantage of the available data. More research studies are needed on how to effectively manage and utilize large amounts of data and how that can be used by transit agencies and decision-makers. The development of new techniques or the technology transfer of previous research can benefit not only transit agencies, but also the whole transportation sector.
- More work is needed to identify the role of Transit ITS vendors and software developers in the development of systems and tools for the transit industry. Perhaps, government agencies and decision makers need to look at the possibility of encouraging vendors and developers to create technological systems that combine ITS data with other datasets such demographic data, parcel data, census data, and other related datasets that can be used by transportation and transit agencies. This can help improve transit service efficiencies that can promote the use of public transportation.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

AC Alighting Controlled
ACS American Community Survey
ADA Americans with Disabilities Act
ADAMS Data Archiving and Mining System
ADCS Automated Data Collection Systems
AFC Automated Fare Collection
AI Artificial Intelligence
APC Automatic Passenger Counters
APTA American Public Transportation Association
APTS Advanced Public Transportation Systems
ATS-A Automated Train Supervision
AVL Automated Vehicle Location
AWS Amazon Web Services
BART Bay Area Rapid Transit
BC Boarding Controlled
BCT Broward County Transit
BFE Bus Fare Evasion
BI Business Intelligence
BMA Beijing Metropolitan Area
BP Back Propagation
BRT Bus Rapid Transit
BSI Bus Stop Inventory
BYOC Bring Your Own Chip
CAV Connected and Automated Vehicles
CBD Central Business District
CFMS Contactless Fare Media System
CNG Compressed Natural Gas
CRTP Comprehensive Regional Transit Plans
CSI College of Staten Island
CSV Comma-Separated Values
CTA Chicago Transit Authority
CTD Console Dispatcher
CTPP Census Transportation Planning Products
CTPS Central Transportation Planning Staff
CTSP Community Transportation Safety Plan
DOT Department of Transportation
DSS Decision Support System
DT Dwell Time
DTPM Public Transport Authority

EBS Electronic Booking System
EFP Electronic Fare Payment systems
EJ Environmental Justice
EM Expectation–Maximization
ETL Extract, Transform, and Load
EVC Eigen Vector Centrality
FDOT Florida Department of Transportation
FIU Florida International University
FTA Federal Transit Administration
FVM Fare Vending Machines
GA Genetic Algorithm
GIS Geographic Information Systems
GPS Global Positioning System
GTFS General Transit Feed Specification
HDB Housing and Development Board
HPC High Performance Computing
ICT Information and Communication Technology
IETT Istanbul Electric Tram and Funicular Company
ITS Intelligent Transportation Systems
IVHS Intelligent Vehicle Highway Systems
KOTI Korean Transport Institute
LBS Location-Based Services
LOS Level of Service
LPR License Plate Recognition
LRT Light Rail
MBTA Massachusetts Bay Transportation Authority
MTA Metropolitan Transportation Authority
NYCT New York City Transit
OD Origin – Destination
ODIP Origin – Destination Inference Problem
ODX Interchange Inference
OIG Office of the Inspector General
OLAP Online Analytical Processing
OLS Ordinary Least Squares
OP Operations Planning
PAYG Pay-As-You-Go
PCAC Permanent Citizens Advisory Committee
PPB Pre-Paid Benefits
RCC Rail Control Center
REDOs Regional Economic Development Organizations

RFID Radio-Frequency Identification
RTA Regional Transit Authorities
SDR System Data & Research
SFE Transit's Subway Fare Evasion
SIRE Service Intervention Recommendation Engine
SIVT Système d'information et de validation des titres
SLD Surface Line Dispatchers
SMA Seoul Metropolitan Area
SPSS Statistical Package for the Social Sciences
SSP Schedule-Based Shortest Path
STM Société de transport de Montréal
STO Société de transport de l'Outaouais
STP Space Time Prism
TAP Transportation Access Pass
TAZ Traffic Analysis Zone
TCL Transport en Commun Lyonnais
TCQSM Transit Capacity and Quality of Service Manual
TfL Transport for London
TMA Transportation Management Association
TPAP Trip Purpose Assignment Process
TSP Transit Signal Priority
TVM Ticket Vending Machines
UHTS Utah Household Travel Survey
UTA Utah Transit Authority
UTFS Universal Transit Farecard Standards
V/C Volume to Capacity
VKT Kilometers Traveled per Vehicle
WTP Willingness to Pay
YCIPTA Yuma County Intergovernmental Public Transportation Authority system

1 INTRODUCTION

With the growing and aging of the population, increasing congestion, and rapidly evolving technologies and mobility services, there is a pressing need to capture and monitor the trends and impacts on the transit system to have a better understanding of the spatial and temporal dynamics of transit demand. Public transportation is a key component of urban transportation solutions that can help mitigate congestion, reduce vehicle emissions, and promote sustainable growth. However, planners have limited and imprecise instruments to obtain transit data and information. This is usually acquired through household travel surveys or transit on-board surveys that are costly and time consuming. On the other hand, the massive and continuous data and information collected by the automated fare collection (AFC) systems present great opportunities as a bias-free and low-cost and more efficient approach compared to manual methods. The availability of AFC data and information can help examine the behavior of passengers and the demand characteristics for public transportation services, which can be useful to support short-term planning as well as the development of long-term strategies.

Automated fare collection (AFC) systems, also referred to as automatic fare collection or electronic fare payment (EFP) systems, have been widely adopted by transit agencies around the world. While providing time savings, convenience, and other advantages in revenue collection, they also produce massive, continuous, and anonymized digital records of fare transactions. AFC data present unprecedented opportunities for transit planners and researchers to capture and analyze passenger travel patterns as well as general ridership and revenue information. Fare media provide much more comprehensive, exhaustive, and spatially and temporally precise information than traditional household travel surveys or on-board surveys could ever provide.

AFC data can help discover the tendencies and dynamics in passenger travel and assess users' responses to changes in the transit system (e.g., fares, routes, and scheduling modifications) or reactions to external factors (e.g., social trends, technology, mobility service innovations, etc.). Despite the advantages of the data collected by AFC systems, it is not entirely clear how AFC data could be utilized to efficiently support operational analysis, inform transit service planning and scheduling, and facilitate long-term planning for public transportation. Many studies have taken advantage of AFC data and conducted research from several perspectives, including ridership statistics, performance indicators to support operational analysis, travel pattern identification, network analysis for service planning, and behavior analysis to facilitate long-term planning for public transportation (Pelletier et al., 2011).

This project aims to provide a comprehensive review of existing knowledge and experience in utilizing AFC data, from both practice and research perspectives. From the practice perspective, it will be interesting to know whether and how transit or planning agencies have utilized AFC data to support planning and operational activities. From the research perspective, innovative data collection and analysis methods and applications will be explored and summarized. The findings of this research can be used to assess if the development of a data analytics and mining tool is necessary to extend the utility of AFC data to examine system performance and demand characteristics of public transportation at a higher level of sophistication and detail. In this case, a data mining tool can provide additional insights on the network performance as well as the spatial and temporal dynamics of transit demand at a resolution and scale that are impossible to obtain via traditional data analysis methods.

2 LITERATURE REVIEW

In this section a comprehensive literature review of existing AFC systems is conducted. The focus of the literature review is to identify current practices and innovative methods and applications on the use of AFC data. Project reports and agency documentations were sought to understand the current state-of-the-practice in utilizing AFC data. Relevant literature regarding the AFC systems and uses of the data collected were also identified and summarized, in particular in areas that can enhance public transportation services.

2.1 Advanced Public Transportation Systems (APTS)

The Federal Transit Administration (FTA) developed the Advanced Public Transportation Systems (APTS) Program as part of the U.S. DOT Intelligent Vehicle Highway Systems (IVHS) (Casey and Collura, 1994). The goal was to promote research and development of innovative applications to enhance the ability of public transportation systems to satisfy customer needs and contribute towards the accomplishment of wide-ranging community goals and local objectives. These technologies included automated vehicle location systems, smart card systems, dynamic ridesharing systems, passenger information systems, high occupancy vehicle systems, and vehicle component monitoring systems.

As per Casey and Collura (1994), in the *Advanced Public Transportation Systems: Evaluation Guidelines* report, the original smart card systems (Automated Fare Collection Systems) envisioned using a contact or contactless plastic card, with a microchip and storage and processing capabilities, to facilitate the collection of fares, verification of travel, and the acquisition of information about passengers and vehicle usage. The system would increase the convenience of fare payments within and between modes and facilitate the implementation of a more equitable and efficient fare policy. For instance, the smart cart system could be used to facilitate the ability to provide discounted fares to special user groups (e.g., people with disabilities).

APTS systems are a group of technologies that help increase the efficiency and safety of public transportation systems and provides users better access to information. Since the implementation of APTS technologies, there has been a transformation on the way public transportation systems operate and the delivery of services offered by public transportation systems. In addition, APTS systems can provide decision makers with information needed to make intelligent decisions on operations and services provided that can help improve ridership and riders' confidence in the transit system. APTS technologies can be divided into five broad categories that describe the technologies in relevance to the different transit applications (Casey et al., 2000). The five APTS technology categories are presented in Table 1:

Table 1: APTS Technologies

Transit Application	APTS Technologies
Fleet Management Systems	• Automatic Vehicle Location Systems • Transit Operations Software • Communications Systems • Geographic Information Systems • Automatic Passenger Counters • Traffic Signal Priority Systems
Traveler Information Systems	• Pre-Trip Transit and Multimodal Traveler Information Systems • In-Terminal/Wayside Transit Information Systems • In-Vehicle Transit Information Systems
Electronic Payment Systems	• Smart Cards • Fare Distribution Systems • Clearinghouse
Transportation Demand Management	• Dynamic Ridesharing • Automated Service Coordination • Transportation Management Centers
The Transit Intelligent Vehicle Initiative	• Lane Change and Merge Collision Avoidance • Forward Collision Avoidance • Rear Impact Collision Mitigation • Tight Maneuvering/ Precision Docking

In the State of Florida, public transportation agencies are continuously seeking opportunities to increase ridership and improve operational efficiency and quality of service. Therefore, transit agencies looked into Advanced Public Transportation System (APTS) technologies as effective tools to address a full range of customer service and operational issues including the following items: providing better customer information, offering enhanced traveler information, improving on-time performance, prioritizing investments and allocation of resources, reducing the cost per passenger trip, improving system safety and security, and enhancing the amount and quality of data available for planning analysis (Cevallos, 2007).

Using Advanced Public Transportation Systems (APTS), also known as Transit Intelligent Transportation Systems (ITS), transit agencies can manage their transit fleet, monitor ridership demand, improve the collection of fares, announce upcoming stops, and provide real-time information to their customers. Therefore, APTS technologies can help transit agencies become more efficient by allowing them to better manage the transit system and make more intelligent decisions in the allocation of resources. In addition, archived data from these systems can be used to develop performance measures to monitor the performance of the entire system, examine particular routes, route segments, or transit stops. With such information, transit agencies can improve on-time performance, monitor ridership, and identify potential areas of improvement to deliver better service to the public.

The implementation of Intelligent Transportation Systems (ITS) applications has enhanced the transit services by improving their efficiency, safety, and reliability. Developments in ITS has mostly focused on the hardware part, upgrading the data collection system, and leaving the software behind which is related to planning and decision-making (Iliopoulou and Kepaptsoglou, 2019). Their *Combining ITS and optimization in public transportation planning: state of the art and future research paths* study focus on the review and optimization of public transport systems and services, using AVL, APCs, and AFC data.

Large quantities of data are collected by the ITS systems providing several opportunities for information analysis. ITS applications have allowed for the collection of data that can be useful for planning and operations. According to Iliopoulou and Kepaptsoglou (2019), in public transportation, different ITS applications have enabled the collection of data for assessing areas such as system performance, ridership, and demand patterns. For instance, Automated Vehicle Location (AVL) systems can be used to monitor schedule adherence and allow the development of more accurate schedules; Automatic Fare Collection (AFC) systems and Automatic Passenger Counters (APCs) allow for the collection of detailed ridership data; and the combination of data from these systems with GIS (Geographic Information Systems) can be very useful for transit planning. However, it appears that the potential uses of transit ITS data to improve strategic and operational planning have not been fully exploited by the research community.

Iliopoulou and Kepaptsoglou (2019) suggest four stages to treat planning problems efficiently for a better transportation system such as:

- Strategic stage. The design of the network and passenger assignment are typically examined as part of a long-term planning process. This process uses data from surveys leading to a limited analysis of travel pattern. However, with the AFC data analysis, the monitoring process of riders could be longer to provide a better understanding of travel patterns that includes temporal and spatial demand variations and dynamic patterns into existing models.
- Tactical stage. This refers to determining the operational characteristics of the transit services, namely frequencies and timetables. It benefits from using longitudinal ITS data, because collected data from APC, AFC, and AVL systems can obtain frequent mobility patterns and identify reliability issues. With AFC data, several studies have attempted to estimate origin – destination (OD) matrices in the context of timetable/frequency/level of service adjustments. Studies to obtain optimal timetables are often related to data availability and detail level.
- Operational stage. It refers to vehicle scheduling, driver rostering, maintenance planning, as well as parking and dispatching. Information collected by the AVL system can help to

develop optimal vehicle schedules by extracting periods of homogeneous running time and trip time probability distributions.

- Real-time stage. It deals with daily operations and refer to control strategies. AVL systems have been widely applied for real-time control of public transportation systems and particularly for alleviating bus bunching, large waiting times at stops, etc. Even tough in the presence of real-time information, computationally intensive transit planning models may be unsuitable to quickly generate optimal paths, research efforts have been directed towards efficient trip planning models, which explicitly incorporates real-time AVL data in order to accurately represent bus arrival times.

With the increasing of ITS, along with new studies opportunities, research gaps and limitations have also been found. Some of the discovered issues are as follows (Iliopoulou and Kepaptsoglou, 2019):

- Additional data processing required. The collected data by AVL and AFC systems need to be analyzed and processed in order to be useful for transit planners.
- Lack of integration among various data sources.
- Different degrees of fleet penetration. While AVL systems are typically installed on entire bus fleets, the same is not true for APC systems that limits the collected information.
- Current state of practice. Lacking the degree of realism required in practice problems.
- Increased computational requirements.
- Operators' data-sharing policies. Certain operators have adopted data-sharing that stimulates ITS related research. However, this is not the typical case, as there may be some limitations due to privacy concerns and operators' restrictions.

In addition, Iliopoulou and Kepaptsoglou (2019) identified several factors that limit the research such as:

- Data quality considerations. Developing accurate models depends on the quality of the data. For instance, older and less advanced AVL systems produce reports which contain vehicle trajectory data that lacks stop-level information. Therefore, matching algorithms are needed to couple raw location data to route maps and schedules. To reduce this problem, a series of data manipulation procedures have been implemented, but ultimately it depends on public transport operator policies, through maintaining quality control, and post-processing procedures.
- Supplementary data requirements. AFC and APC data have limitations for some analyses, as critical elements required for decoding traveler choices are missing. For this reason, additional manual surveys are recommended to obtain accurate information.
- Computational effort. Computational costs associated with processing ITS data and optimization models across all planning stages are increasing inevitably due to high technology required for such procedures.

Transit planners and decisions makers can use operations research methods to transform data into valuable information that can be used not only to study the public transport performance, but also predict future conditions and generate solutions to planning problems (Iliopoulou and Kepaptsoglou, 2019). Nevertheless, the usage of the data for developing suitable planning and operational tools is limited and the research has not extensively investigated the potential of combining ITS data for developing optimization models and decision support tools. Therefore, there are many opportunities for using ITS data for developing models and methodologies in public transport planning and operations in the areas of route planning, scheduling, and resource allocation in real time.

Despite the limitations and gaps, the large datasets may be used by transit planners and operators to enhance the transit services. There exists some work on the development of data-driven platforms for public transportation planning which integrates mining methods, regression models, and visualization techniques to assist in performance monitoring and predict and evaluate potential impacts of different transit strategies that can provide a more comprehensive understanding of network dynamics overall.

2.2 Automated Fare Collection (AFC) Systems

Automatic Fare Collection (AFC) systems, or Electronic Fare Payment (EFP) systems, use electronic communication, data processing, and data storage techniques to automate the manual fare collection process. The use of this technology offers multiple benefits. The most significant advantage is that AFC systems make fare payment more convenient for travelers and revenue collection less costly for transit providers. AFC systems help reduce the cost of labor-intensive cash handling and the risk of theft, improve reliability and maintainability of fareboxes, and permit sophisticated fare pricing based on time of day or distance traveled. AFC systems can also help with the automation of financial and accounting processes (Cevallos, 2007).

Electronic fare payment systems provide an automated means of collecting and processing fares for public transportation services. According to Hwang, et al. (2006), the description of electronic fare payments can be organized into three major categories:

1. Fare Systems
2. Fare Media
3. Clearinghouse/Regional Service Center

These categories were structured using the work undertaken by the Universal Transit Farecard Standards (UTFS) Program. These three elements need to work together in a transit electronic fare payment system. In addition, the Clearinghouse/Regional Service Center highlights the importance of integration when deploying a regional fare system that involves multiple agencies. In regional fare systems, integration can occur at a number of levels to help facilitate seamless fare payment, including an integrated approach to fare policies, fare media or card reading standards, fare equipment, and revenue reconciliation (Hwang et al., 2006).

Automatic Fare Collection systems combine fare media, such as magnetic stripe cards, smart cards, and cash with electronic communications systems, data processing computers, and data storage systems. They can provide a range of potential benefits to transit agencies and their customers, including simplifying fare payment, increasing customer convenience, reducing boarding times, improving revenue security, increasing information about ridership and fare payment trends, and potentially eliminating the acceptance of cash, coins, and tokens.

Electronic transactions also facilitate data gathering and analysis because of the additional fare and ridership information obtained electronically (e.g., the time, location, and type of each transaction). These systems store large amounts of data in computerized files that include all the transactions that take place during the fare collection process. The technology used in automated fare collection and passenger recording can help collect large amounts of data regularly. This is important, as the collection, processing, and safekeeping of passenger fare revenues and ridership data are vital to a public transit agency. However, without proper database storage, tools, and processes to take advantage of the data, planners and transit staff cannot begin to analyze transit usage and trends or identify travel characteristics and patterns.

2.2.1 The Fare Structure

Transit agencies around the country have implemented AFC technologies to improve efficiencies. The AFC implementation includes the setup of the fare structure. The transit fare structure is a key element in the collection and use of data from AFC systems. While smart cards and stored-value cards are already popular, other cash fares and fare media like monthly, weekly, and daily passes are all implemented.

To illustrate frequently used fares by transit agencies, Table 2 shows the Broward County Transit (BCT) adult fares for ages 19-64[*]. This is part of the whole BCT fare structure. In fare collection systems, a fare structure can be created with different fare types to address different segments of the population. For example, college students may use the student pass and older adults and people with disabilities may use reduced fares. This can help transit agencies organize the data collected from the AFC systems, so that they can provide information to transit planners regarding to travel patterns and preferences of the various population groups included in the fare structure. Further, with appropriate data from AFC and other transit ITS systems, transit planners can have the necessary information they need for the optimum allocation of resources.

Table 2. BCT Fare Structure: Adult Fares (Ages 19-64) *

One-way Cash Fare	Basic fare for one-way travel.
3 Day Bus Pass	Unlimited rides for 3 consecutive days. Starts the first day the card is swiped on the bus.
7 Day Bus Pass	Unlimited rides for 7 consecutive days. Starts the first day the card is swiped on the bus.
10 Ride Bus Pass	Expires after the 10th ride is taken.
All Day Pass	Available for purchase on the bus. Unlimited rides all day on BCT fixed routes.
31-Day-Adult	Unlimited rides for 31 consecutive days. Starts on the first day the card is used. Expires after the 31st day.
Premium Express Fares	Premium Express One-way Cash Fare Premium Express 10 Ride Bus Pass

* BCT also offers discounted fares for the following groups: Senior Fares (65 and older), Youth Fares (18 years or younger), College Bus Pass, and Veterans.

2.2.2 Implementing AFC Technology in Small Transit Agencies

Implementation of a smart cards system for small or rural transit agencies could provide several benefits to their riders because the automation of fare collection not only will replace the outdated cash handling manipulation on the buses, but also will improve the agency services and efficiency. One of the main reasons of why these transit agencies do not have the AFC systems implemented is because they cannot afford the expensive technology of smart card solutions which is typically offered by the major suppliers of fare collection system (Allen et al., 2016).

As stated in Implementation of Smart Card Automatic Fare Collection (AFC) Technology in Small Transit Agencies for Standards Development report (Allen, W. E et al., 2016), deployment of the American Public Transportation Association (APTA) Contactless Fare Media System (CFMS) Standard eliminates the proprietary solution, meaning the cost of implementation is lowered, giving the opportunity to small and rural agencies of affording fare payment systems. Under a contract from the National Academy of Sciences, Transportation Research Board, and IDEA Program, Acumen Building Enterprise, Inc., a consulting firm with expertise on transit systems engineering, implemented the APTA CFMS on a rural agency system.

The project team chose the existing proprietary Yuma County Intergovernmental Public Transportation Authority (YCIPTA) system because it had a contactless smart card system installed and in operation using a proprietary smart card data structure. YCIPTA provides fixed-route service throughout southwestern Yuma County and operates 18 buses on 11 routes, Monday through Saturday.

The APTA Standard specifies the interface between a smart card and a smart card reader, as well as the interface between the local equipment and the back-office fare collection database system. The goal of the APTA CFMS is to promote interoperability of fare media among transit agencies. Acumen successfully implemented the APTA CFMS on the YCIPTA bus system. They followed four major procedures as shown on Figure 1 (Allen et al., 2016):

- Passenger and Bus Center shows the passenger transactions using smart cards.
- YCIPTA AcuFare Management Center that manages the AcuFare Readers on the buses.
- Acumen Host Processing Center that maintains the smart card transaction data.
- Payment Gateway that handles the processing of credit or debit card transactions and uses PayPal to facilitate the payments.

The Transit IDEA 79 Project was a success demonstrating the potential of small agencies in operating a smart card system, and consequently the evaluation of other possible small transit agencies for fare collection system implementation. Acumen also improved their services by adding a connection to PayPal for credit card purchases, meaning a passenger could connect to the YCIPTA website and buy fare products like cash fares, rides, daily passes, monthly passes etc. using a PayPal account. Acumen also manufactured hardware components into the General Services Administration (GSA) pricing schedule to provide small agencies with the opportunity to purchase devices at the lowest market price that can be used into a CFMS-compliant smart card system (Allen et al., 2016).

The lessons learned include challenges and opportunities that can be applied to improve the implementation efficiency and similar undertakings. The study grouped the lessons learned into implementation, procedural, and demographic lessons. They covered many subjects dealing with the implementation of AFC technology in small transit agencies, including cost considerations, operating issues, equipment capabilities, standards' requirements, income issues, passenger perceptions, and special needs persons. It is also important to mention that small agencies have particular challenges. As mentioned in the study, small or rural transit agencies has limited or no technical support for the implementation of a high technology project like the AFC system. This presents a challenge to support and effectively use the AFC system (Allen et al., 2016).

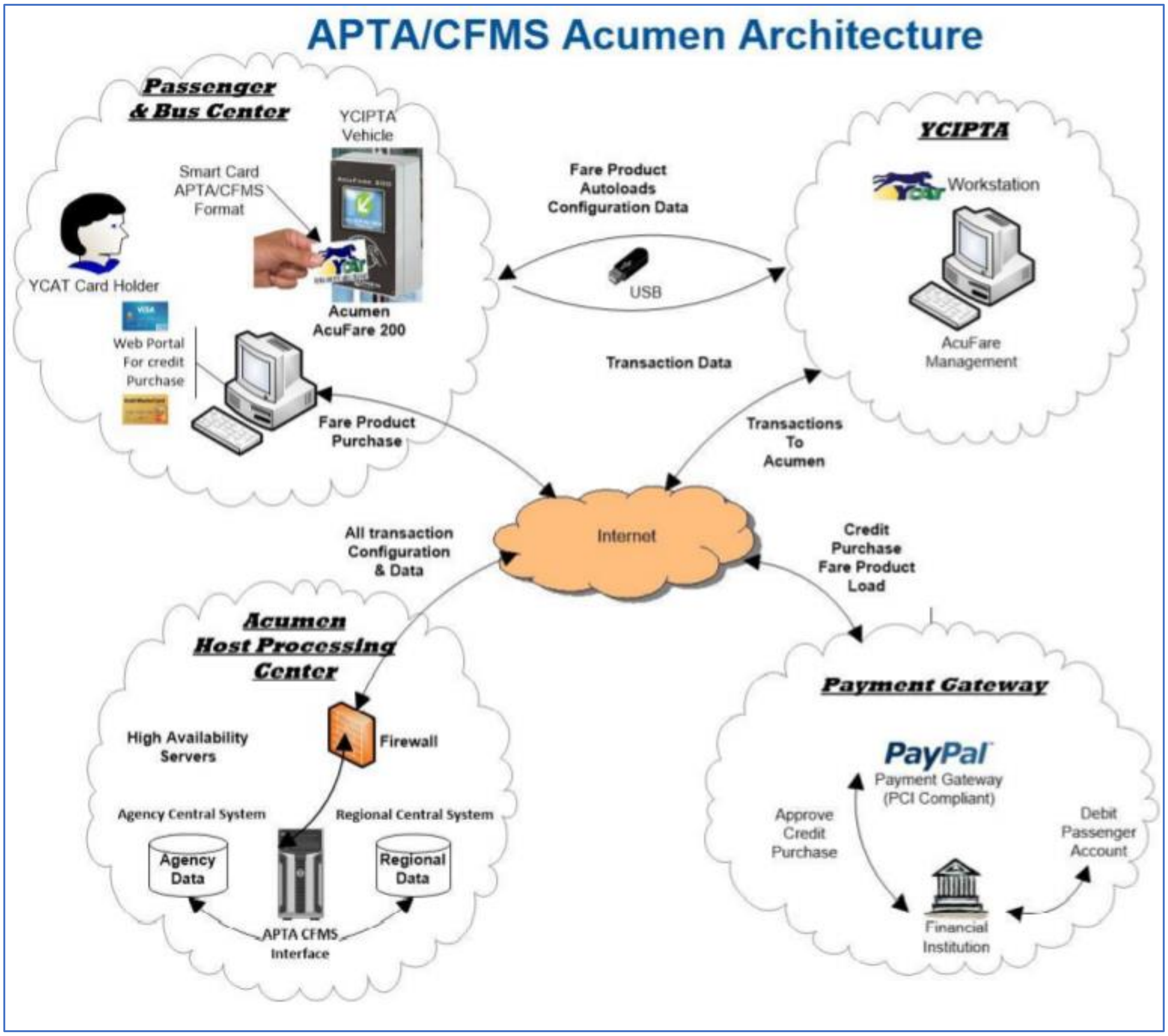


Figure 1. APTA CFMS System Architecture for YCIPTA

2.2.3 Using Data from AFC Systems

Most transit agencies in the U.S. collect data from electronic farebox systems that usually contain critical information that can be used to determine the travel characteristics and behavior of the riders using the transit system. Knowing the travel patterns and preferences of the riders can provide useful information for planning and improving the delivery of the services provided.

The information from AFC systems in combination with other ITS systems can be used to prioritize the investment on transit infrastructure like shelters, provide services at particular time of day or at particular locations, comply with the Americans with Disabilities Act (ADA), and help improve the overall mobility for all segments of the population, including older adults and people with disabilities (Cevallos et al., 2010b).

To fully understand the processes associated with obtaining and utilizing data from an AFC system, it is essential that transit agency staff understand how data flow throughout the system. AFC data is usually generated at the farebox onboard the agency vehicles. The data are loaded and stored into a database, which can be queried to retrieve information. A query is essentially a precise request for information retrieval within a database or information system. For example, a user may want to find the total number of passengers that rode the transit system on a given day. Upon properly requesting the data, the query will return the requested results.

Retrieving data from a farebox system is an extremely important. Not all farebox systems work identically, but for the most part the usage, storage, and retrieval of information are similar. After the data has been loaded into a database or a data warehouse, data can be queried and retrieved. For example, data from the Automatic Fare Collection systems can be used to answer the following questions (Cevallos et al., 2010a):

- *What bus routes have the highest older adults and people with disabilities ridership?* Determining which routes are being used by older adults and people with disabilities can provide key information that can help agencies determine where and how to prioritize route services and stop infrastructure.
- *At what times do most older adults and riders with disabilities prefer to travel?* The output data could be used to create charts or graphs by time of day detailing the number or percentage of older adults and riders with disabilities. Having this information can be used to assess or improve the services provided to these segments of the population.
- *Identifying travel trends of older adults and riders with disabilities.* This information can be used to then study travel trends of special populations. With these data, transit planners can have the additional information necessary to effectively plan for services to the older adults and people with disabilities population.
- *Examining Origin / Destination Data.* Smart card data can be used to determine travel trends for particular riders, based on trips recorded for each individual card. This can be combined with other operational data included in the AFC systems to study travel trends and preferred origin and destination locations.
- *Combining AFC Data with other Transit ITS Data*. AFC systems generate large amounts of data on a daily basis. However, there are other transit ITS/APTS technologies that also generate large amounts of data. All the data collected by these ITS systems can be combined and used by transit planners to help with the decision making and allocate resources properly.

To illustrate how the smart card system works and how the data flow, Figure 2 (Agard et al., 2006) present a diagram of the Smart Card Information System. When the user boards the buses, the smart card's fare gets validated. The validation process is done following these steps: the bus system usually contains operational information such as the planned routes and runs for the day. If the system has a Global Positioning System (GPS) reader, the bus identifies the location where the boarding is made. The system validates the route and the run at this location. The card number, date, time, validation status, and location are stored at each boarding. The location information can also be associated to a bus stop number. In addition, the central server can be fed with service operation information and smart card point-of-sale data. The collected data is downloaded to the AFC server that it is shown in Figure 2 as the SIVT server (Système d'information et de validation des titres). After the data are saved to the AFC server, an analysis could be performed. For instance, classify the people in different groups based on their age as well as their boarding times and routes. This assessment could help transit planners understand the customer habits and patterns when traveling using the public transportation system; thus, making it easier for transit agencies to make the necessary improvements.

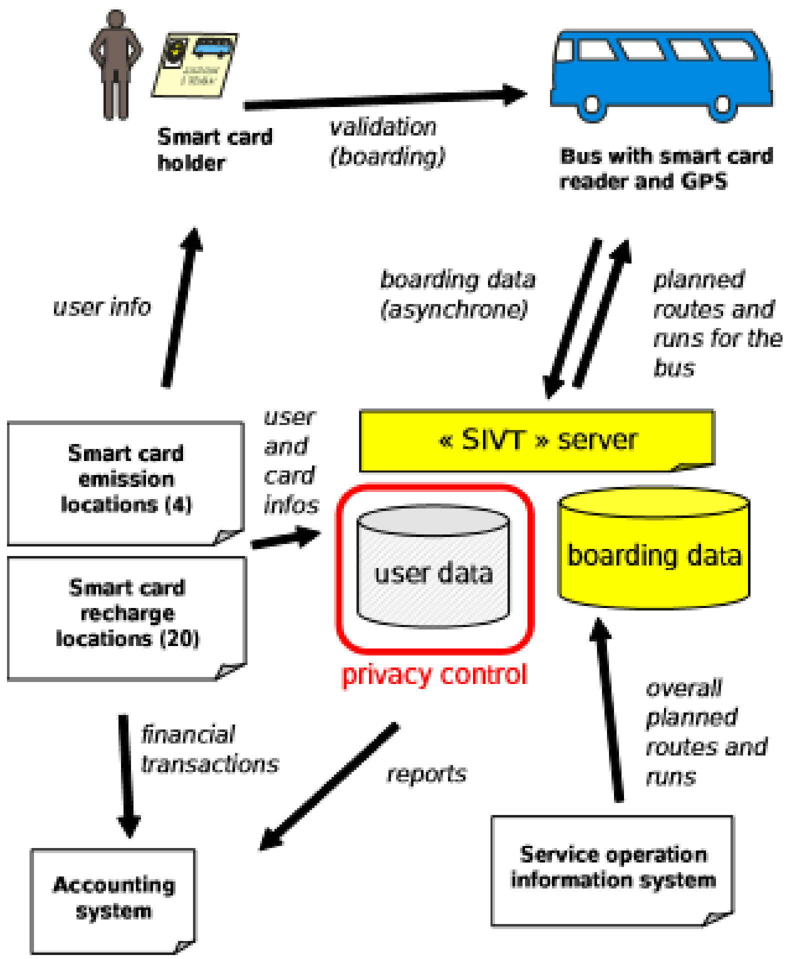


Figure 2. Smart Card Information System

The databases of AFC systems usually include operational information such as route number, run number, trip number, direction, etc. along with the data collected from the fare events and transactions such as timestamp, revenue by fare type, transaction ID, transaction sequence number, smart card ID, etc. Traditionally, farebox records have been used to calculate transit agency performance measures. Obtaining measures such as ridership, revenue, fare recovery ratio, passengers per hour, and passengers per revenue mile are all common practices. However, as the farebox data provides detailed information, it also allows to discover a wealth of information about the riders. This can be used by transit agencies to improve the delivery of their services (Cevallos, 2007).

In the Smart card data use in public transit: A literature review research paper, Pelletier et al. (2011) examined different aspects on the use of data from smart cards in the public transit context. Section 4 of that paper summarizes information from studies that have been conducted on the use of transit smart card data. This was grouped in three categories, based on the management level and type of analysis: strategic, tactical, and operational. Table 3 depicts a review of studies on smart card systems for strategic transit planning. As it can be seen in this table, smart card data can be used in many areas such as understanding the user habits and behaviors, identifying market segments and demand forecasting, and assessing the loyalty of transit users. This can help transit agencies provide better service based on readily available data from AFC systems.

Table 3. Uses of Smart Card Data for Transit Planning

Author(s)	Data	Analysis/use	Benefits
Agard et al. (2006)	Boarding date, time, and location. Card type	Define typical user type and measure their trip habits. Analyze use variability according to the day, the week, or the season	Better understand user behavior
Bagchi and White (2005)	Time, space, and structure. Personal and travel data	Turnover analysis. Marketing	Analyze the consistency of users' travel behavior over time. Produce targeted marketing campaigns to retain users in specific groups
Blythe (2004)	Route load profiles	Manage the demand through the network	Make public transit more attractive
Chu and Chapleau (2008) and Chu et al. (2009)	Boarding date, time, and location. Estimated alighting point. Card type	Runtime estimation. Itinerary reconstruction. Spatio-temporal portrait of the network. Concept of Driver Assisted Bus Interview (DABI)	Make adjustments to network geometry and schedules. Obtain richer information than that from a travel survey. Adapt the network to user needs
Utsunomiya et al. (2006)	Personal information. User address, boarding point, frequency of use. Trip information, demand elasticity	Development of a mailing list for service change announcements. Fare policy analysis. Marketing analysis: identification of market segments with low penetration, conduct targeted surveys. Analysis of a demographic profile of riders by route or station	Allow users to follow an alternative itinerary. Improve user trust in the service. Fare adjustment according to user needs. Demand forecasting
Park and Kim (2008)	Historical data	Future trend estimation. Creation of a future demand matrix	Service adjustment (long-term). Network extension and adaptation
Trépanier et al. (2004)	Boarding date, time, and location	Transportation Object-Oriented Modeling. Planning of the public transit network	Anticipate network extensions
Trépanier et al. (2009a,b)	Boarding date, time and location. Estimated alighting point. Card type	Comparison of smart card data with household survey data (bus use, temporal and spatial distribution of trip)	Improve the accuracy of data from both sides. Complete the survey with smart card data
Trépanier and Morency (2010)	Boarding transactions by card. Starting and ending dates	Calculate the lifespan if the card and the ratio of use of smart card users	Model the loyalty of users from smart card data. Would help to focus on loyalty and retention improvements

Pelletier et al. (2011) also identified many potential challenges that transit operators will have to address in the future, as presented below. It is important to note that, although this research was published in 2011, some of the identified challenges remain relevant.

- Technological improvements: Smart card systems need to evolve in the coming years to make them more robust and reliable. Standards will have to be improved and partnerships developed between operators and commercial partners.

- Data validation: Smart card systems are complex and generate large amounts of data. Therefore, database validation logic and rules will need to be developed to clean and validate the data generated by transit smart card payment systems. Getting accurate geographical location for boarding activities and attaining additional operational information from other technology systems remain a challenge for transit operators.
- Economic benefits: The full benefits of using smart card data are difficult to quantify. There is a need to better understand its benefits and assess the gains of using fare collection data in areas such as fraud reduction, staff reorganization, fare management, and many other benefits.
- Journey information: Although fare collection systems data can provide detailed information on each trip, the trips must be linked to retrieve individual journeys. In addition, algorithms need to be developed to estimate the alightings, when the system only collects boardings.
- Modeling methods: To be able to handle massive amount of disaggregate data, classical models cannot be used to deal with such detailed level of resolution. Therefore, new modeling methods are needed. Linking fare collection data with socio-demographic data should be a priority.
- New analysis approaches: Data for fare collection systems can be used to generate travel behavior data and information on the regular basis. Transit agencies can benefit with the development of new analysis approaches that can be useful for longitudinal studies of ridership data or time series modeling.

It is expected that, in the coming years, smart card fare collection systems will become the most widely used method of payment in transit systems. Millions of smart cards will provide detailed and continuous source of data that have the potential for using this data for strategic, tactical, and operational purposes (Pelletier et al., 2011). In addition, with a large amount of data, planners and researchers can have a good source of data to gain a better understanding of the behavior of transit users that can help improve the public transportation system.

The following section presents samples on the uses of AFC data. As the literature has a vast amount of information in this area, the research team attempted to capture the most relevant topics that refer to the use of AFC data. Additional information can be found in the Relevant Experiences section that presents North American and International experiences and the Case Studies section that discusses specific topics from selected transit agencies.

2.2.3.1 Transit ITS

The use of AFC data can be enhanced when combining with other transit ITS/APTS systems like Automatic Vehicle Location (AVL), Automatic Passenger Counter (APC), Geographic Information System (GIS), and Stop Inventories. They can provide the needed information for better decision making. With pertinent data from these ITS/APTS systems readily available, transit planners have more information to help them decide how and where to allocate resources for improving the transit system (Cevallos, 2007).

The information obtained from Transit ITS systems can assist planners in their decision to improve transit service such as the assignment of low-floor vehicles or the allocation of infrastructure such as the placement of bus shelters at specific locations. For instance, combining APC and AFC data can help transit agencies assess the activity of older adults and persons with disabilities at the system, route, and stop level, so improvements can be made accordingly.

To demonstrate the use of data from the different transit technologies, the APTS Data Archiving and Mining System (ADAMS) was developed (Cevallos, 2007). The data flow diagram for ADAMS is presented in Figure 3. The ADAMS system included data from the following systems: AVL, APC, AFC, Bus Stop Inventory (BSI), and GIS. After some quality assurance and data manipulation tasks, the datasets from these systems were loaded to a SQL Server database. Then, the ADAMS system was setup to allow the processing and querying of data using data mining and analysis tasks. The system provided outputs for different data mining models and reports. Internet users could make use of these data models or reports through a web application and reporting tools as the front end of the ADAMS system (Cevallos and Wang, 2008).

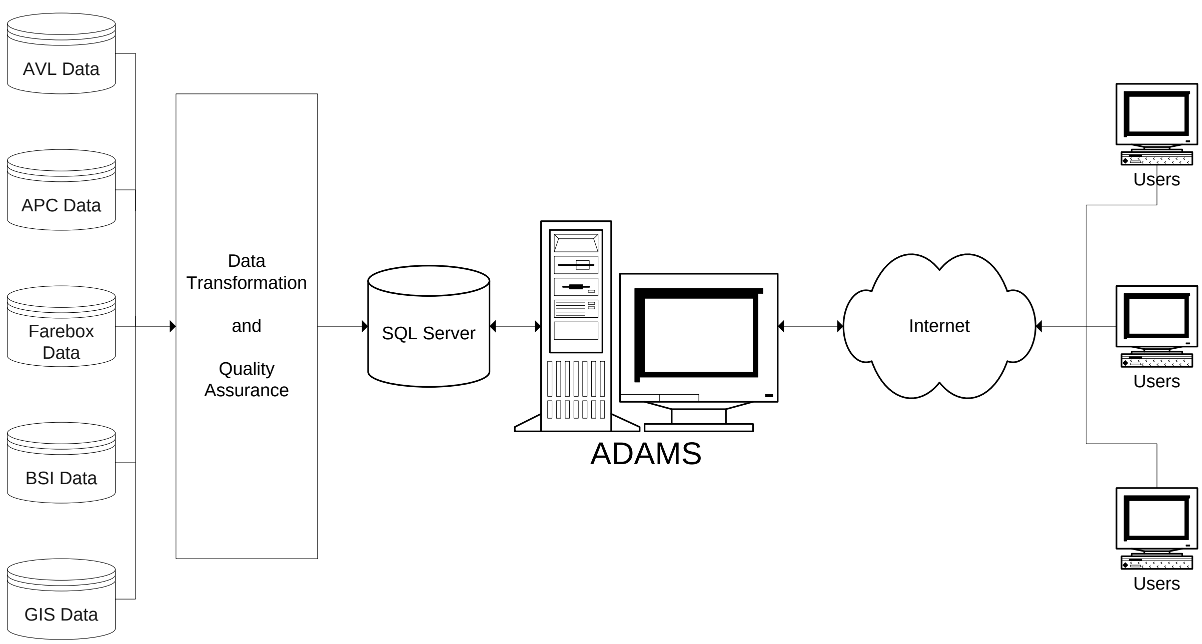


Figure 3. ADAMS Data Flow Diagram

2.2.3.2 Improving Services for Special Populations

Traditionally, farebox records have been used to calculate transit agency performance measures such as passengers per hour and passengers per revenue mile. However, few individuals realize that the same farebox data can be analyzed to learn detailed information about specific populations.

Although it is not a common practice yet, fareboxes offer a wealth of information about elderly and disabled riders.

Using farebox data to examine the travel behavior and trends of specific populations like the elderly or people with disabilities can be accomplished the same way that more general AFC analyses are performed. However, in order to isolate the data records associated with special populations, additional procedures must be instituted. In most cases, specific queries must be made to only select and certain records. The following items are examples on how the AFC data that can be useful to transit planners by providing better services to older adults and people with disabilities (Cevallos et al., 2010b):

- **What bus routes have the most elderly and disabled riders?**
  Determining which routes have highest older adults and people with disabilities is valuable knowledge that can help agencies determine where and how to prioritize route services and stop infrastructure. Although transit agencies may already have an idea of what routes have the most elderly riders, AFC data can easily sort through the millions of associated records and confirm or reject previous assumptions.

- **What is the monthly ridership for some of these routes?**
  Monitoring monthly ridership is important because many areas of the country have large populations of seasonal residents. These individuals, often retired, may live in two or three different residences during a typical year. In Broward County and South Florida in general, additional older seasonal residents and visitors create added demand on bus services during the winter months. If large variations in older adult ridership exist, then transit agencies may need to formulate a plan for providing additional services during busy times of the year and shifting those services elsewhere when few seasonal riders are utilizing services.

- **What bus routes have the largest percentage of elderly and disabled riders?**
  Determining routes with a high percentage of older adults and riders with disabilities can also be quite useful. When allocating scarce resources, this variable might be the deciding factor. Routes with high percentages of these riders ideally should be served by low floor or kneeling buses and assigned to the friendliest and safest bus drivers. Also, popular stops should have the necessary infrastructures installed for safe waiting and boarding. If the data query reveals routes with too few riders, a minimum threshold may be set to ensure that only transit routes with at least a certain number or percentage of older adults and riders with disabilities are selected.

- **When (at what times of day) do older adults and people with disabilities ride the most?**
  Sometimes it is important to determine what are the dates and times that most older adults and people with disabilities would take the bus. Thus, transit agencies can help allocate more buses with ADA equipment in that time period.

- **What bus routes and times are the busiest?**
  Although bus routes with large numbers of passengers may signify healthy revenues, they can also lead to some problems. Crowded conditions might be just an annoyance for most riders, but they can pose serious problems for older adults and people with disabilities. A rider in a wheelchair or with a walker is likely to have significant problems maneuvering within the bus if the vehicle is so crowded that other transit riders must stand in the aisles. Boarding and exiting the vehicle in a timely manner may become very difficult, and at times impossible.

  Likewise, an older rider may have trouble safely finding a place to sit or a pole to hold. This type of situation may increase the likelihood of falls and injuries. Older adults and people with disabilities may be safeguarded against such crowds if transit agencies monitor ridership and consistently supply appropriate transit service to meet demand.

  By examining the busiest segments of the day, agencies can determine where additional resources should be allocated. The routes with extremely high numbers of riders may be considered for increased service frequency as a means to alleviate onboard crowding. Alternatively, larger buses may be assigned to serve the route, thereby providing additional seating and more space for riders.

2.2.3.3 Smart Card Applications

Transportation agencies are using the data collected from Automated Fare Collection systems for a variety of purposes. The datasets have great potential to study not only the travel patterns but also passengers' behaviors, service demand, and transit performance. Before the implementation of smart cards systems, the data collection was conducted from surveys in form of diaries, through phones, person interviews, and GPS loggers. Little information was collected, and operation costs were high. On the other hand, the data collected through AFC is unbiased, more accurate, and provide additional attributes. Information obtained from these systems could help transit planners to improve the public service, make policy changes, advertising, and target marketing (Faroqi et al, 2018b).

As per *Applications of transit smart cards beyond a fare collection tool: a literature review*, Faroqi, H et al. (2018b) state that the AFC system became more streamlined after the growth and positioning of Information and Communication Technology (ICT) systems. The AVL system grew simultaneously with the GPS system directly affecting the AFC system implementation, because it could dynamically locate the position of the transit vehicles with a minimum range of errors. The Faroqi et al. (2018b) paper reviews the most recent developments in the applications of AFC data in diverse domains:

- AFC data: It is considered the database of the fare transactions, where each transaction stores information of the passengers such as: location and time of the boarding or alighting stops, route, direction of the vehicle, fare types, transfers, and sociodemographic characteristics.

- OD estimation: The collected data from AFC include the boarding and alighting transactions per passenger which it is used to determine the OD matrix. The transaction must be linked by a common activity to be classified as a valid transaction for the OD estimation. The problem encountered is the validation time because the time passed between alighting and boarding could depend on the trip purpose.
- Mining travel patterns: Travel patterns reflect regularity, frequency, and relations between trips. Algorithms were developed to analyze the information and categorize the patterns and relations stored in an AFC dataset. The algorithms were classified in supervised, which are recognized as classification or labelling algorithms and need training samples to learn, and unsupervised such as clustering that do not need training. The clustering algorithm has been implemented to extract travel patterns behavior from AFC data focused on detecting spatial travel patterns or determining popular stops.
- Trip purpose detection: Knowing the purpose of each trip it is important when enhancing the public transportation system. This important attribute could be deduced from the time of starting or ending of the activity, duration of the activity, land use in the proximity, and the sociodemographic characteristic of the individual.
- Other applications:
    - Modelling the passengers path based on their trip behaviors.
    - Evaluating the public transit performance.
    - Recording effects before and after a transit policy was changed.
- Limitations: The AFC system has some limitations that can be used as new research opportunities such as: short life expectancy of smart cards, lack of appropriate datasets for validating estimated missing attributes, privacy issues, and fare evasion

Many are the applications that can be created based on the collected data through AFC system to improve the transportation system nowadays. The most important goal is to cover the needs and demand of the passengers by providing efficient and smoother routes.

2.2.3.4 Fare Evasion

Automated fare collection data can also be used by public transit operators to detect the different levels of fare evasion. In the *Can we estimate accurately fare evasion without a survey?* paper, Egu and Bonnel (2020) present the results from a data comparison approach in Lyon, France using fare collection data, fare inspection data, and counting data. The research analyzes the differences on estimating the fare evasion level by the traditional methods and by using the automated data. It was observed that using fare inspection logs, the rate of fare evasion was irregularity low when comparing to automated counts and farebox transactions. This may indicate a level of irregularity.

Measuring the level of fare evasion with automated fare collection data is relatively new. Transport en Commun Lyonnais (TCL) is the commercial name of the public transport network of Lyon, France. The TCL network was used to evaluate the different methods of determine the different levels of fare evasion. The current fare transaction system was implemented in 2002 giving access to customers to a variety of fares. With the automated system, riders are required to validate their tickets and this information is stored in a system where it can be evaluated. There are three different providers of data (Egu and Bonnel, 2020):

- The first provider of data is the fare collection system where every transaction is recorded and stored.
- The second provider of data are the different counting systems where the system stored the number of boardings per day, but the reliability of the data is based on the technology used.
- The third provider of data is the fare inspection system. This is equipment with magnetic and radio-frequency identification (RFID) readers that can verify electronically smart card and paper ticket validity and issue penalties.

The on-site surveys are the other form used to determine fare evasion levels. The information is recorded manually by surveyors in a face-to-face interview with the riders for further analysis.

These methods have their advantages and disadvantages. The data obtained from manual collection could be subjected to bias and its accuracy depends on how well the clients provide the information needed. Sending inspectors to the field is expensive, but on the other hand, the information gathered could provide a better idea of the real situation regarding the fare evasion.

The fare collection system generates data automatically at an incredibly low cost and with the difference being that no human interaction is required. The data is continuously collected, and calculation analysis is programmed providing different levels of the fare evasion information. The disadvantage of the AFC system is that passengers who do not validate their tickets are considered evaders and does not provide precise information on the type of evasion that can be derived from this indicator (Egu and Bonnel, 2020).

2.2.3.5 Measuring Activity Similarity

The implementation of Automatic Fare Collection systems is increasing in transport systems allowing agencies to conduct research for better understanding of the network performance. Smart cards store different kinds of information that can be used to develop better transport models. The paper *A model for measuring activity similarity between public transit passengers using smart card data* (Faroqi et al., 2018a) studies the activity similarity among passengers based on the following characteristics: location and time of boarding and alighting of passengers and the kind of land use at the time of the activity.

The Faroqi et al. (2018a) proposed model consists of two parallel steps: one for the spatiotemporal aspects and the other for the activity type. Figure 4 shows the methodology of the process. The first step uses the concept of Space Time Prism (STP) to measure the spatiotemporal similarity of two activities in a three-dimensional continuous space. The second step uses a probabilistic decision tree to measure the activity type similarity. The STP can model the spatiotemporal aspects of similar activities between passengers using time and location where a probabilistic decision tree approach is used to model the activity type based on the available land use, start time, and duration of the activity.

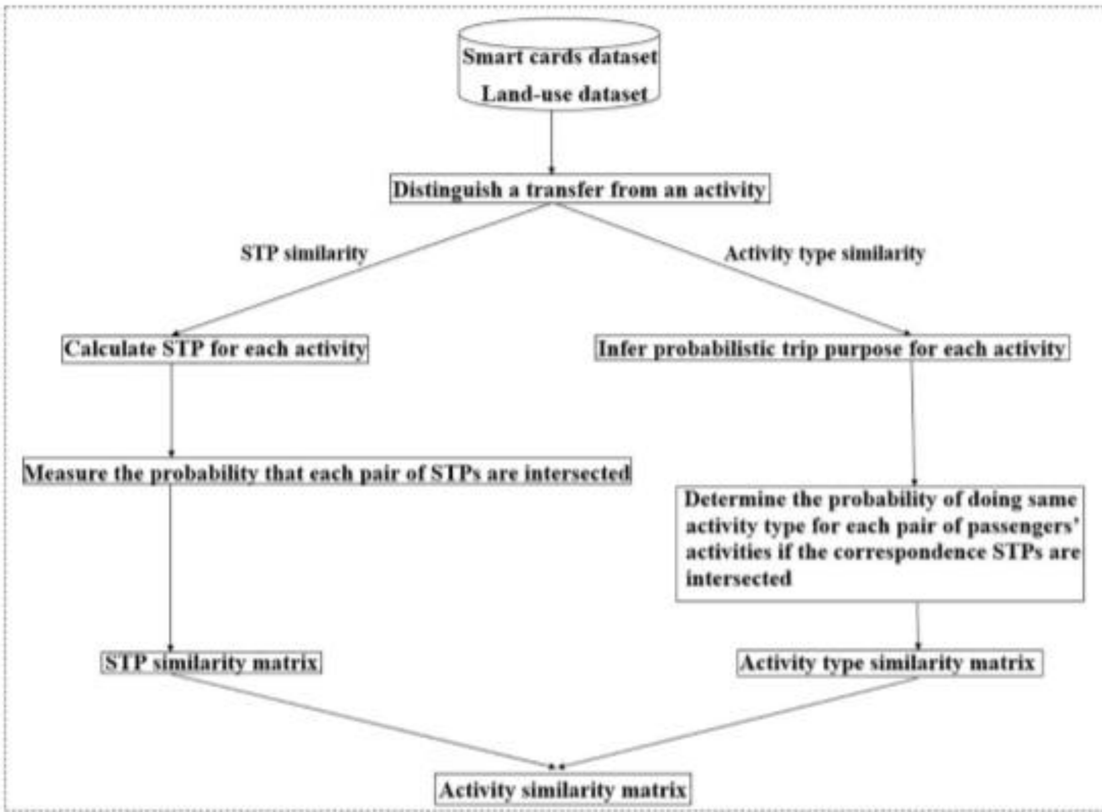


Figure 4. Methodology of the Two Parallel Procedures

Two points in a three-dimensional space defines a STP to identify the probability or span of time and locations of a moving object between those points. Constraints are used to quantify movement possibilities such as places, walking speed, and the maximum walking distance. The two main points are assumed to be the alighting and boarding transactions of a passenger, and the constraints can determine all places and times that a passenger can go after alighting from the public transit vehicle and before boarding it again. The activity similarity measures the probability of passengers doing the same activity. Further, the validation analysis shows that the proposed model can provide an accurate measure for activity similarity of transit passengers when the alighting stops, before the activity, and the boarding stops after the activity are known.

The model was implemented on a four-day smart card dataset from the public transport network in Brisbane, Queensland (Australia). For each day, 20,000 passengers were chosen randomly, and a land use database was also created. This model poses significant advantages because it can be used to develop applications for the public transit networks. First, the use of location, time, and land use could help business owners in making better decisions to cover the needs of their clients. For example, if two groups of passengers perform different activities at a similar time and location, business owners can adapt their businesses to the people's preferences in that area for successful results. Second, the model can show similar activities between passengers, allowing researchers to create applications targeting those groups of passengers to promote products or offers discounts programs (Faroqi et al., 2018a).

The proposed model focuses on passengers' activity after alighting and before the next boarding transaction, using activity similarity and relying on inference methods. This can lead to the development of potential applications in business planning and group-based applications in the public transit network that can help discover groups of passengers who perform their activities together.

2.2.3.6 Inferring Origin-Destination Demand

Obtaining accurate data has become easier with the use of the automated fare collection (AFC) system. This is because, besides its main use of fare collection, the system also stores valuable travel information such as the point of origin and final destination of trips, which can be useful for transit operators and planners in developing new methods to improve the transportation system. The data collected from AFC systems are recorded as transactions and not trips. One transaction contains information as boarding station and time, alighting station and time, price paid, etc., at the point where the user's card is swiped (Wu et al., 2020). For multimodal trips where the card is swiped many times, the data needs to be analyzed to identify the origins and destinations.

As per *Inferring Origin-Destination Demand and User Preferences in a Multi-modal Travel Environment Using Automated Fare Collection Data* (Wu et al., 2020), the knowledge of origin-destination (OD) and route choice logic are essential in building a better transport system that addresses the needs of the customers. Transit ODs are a key information for policy makers, while the evaluation of route preferences is important for planning to assess the efficiencies of the transit system. This paper focuses on identifying true ODs along with preferences in making multimodal route choice decisions with one method addressing both problems.

One of the challenges encountered using AFC data is deducing the OD of each traveler because many factors have to be taken into considerations such as the assumption that the travelers make decisions to maximize their travel quality and convenience. This assumes that the correct OD was chosen among a finite set of eligible ODs (Wu et al., 2020). Regarding traveler preferences, it was found that the possibility of choosing similar routes could be influenced by the trip's crowdedness, meaning that those routes are crowded because they have high demands.

The OD Inference Problem (ODIP) models the route selection by travelers based on their ODs, under variable travel environment conditions. It enables the inference of the ODs based on the observations of the trips which are taken as part of those routes and recorded in the AFC system data. To infer true ODs from AFC data, the AFC data needs to be tailored for ODIP validation.

For this purpose, Wu et al. (2020) made use of a "typical trip", as depicted in Figure 5, where the three different levels of the travel sequence are shown: Walking from Home, Traveling on Bus or Metro, and then Walking to the Workplace. The OD inference task assumed that some levels of the trip where unknown and needed to be inferred. Further, a set of the eligible ODs (w_i) for traveler i are assumed to be the first boarding and last alighting stations.

The known information includes all the route attributes along this OD w_i. The ODIP problem is modeled as quasi station-level OD inference problem. The bus stations and unknowns were inferred from the observation of the use of inbound-outbound metro stations.

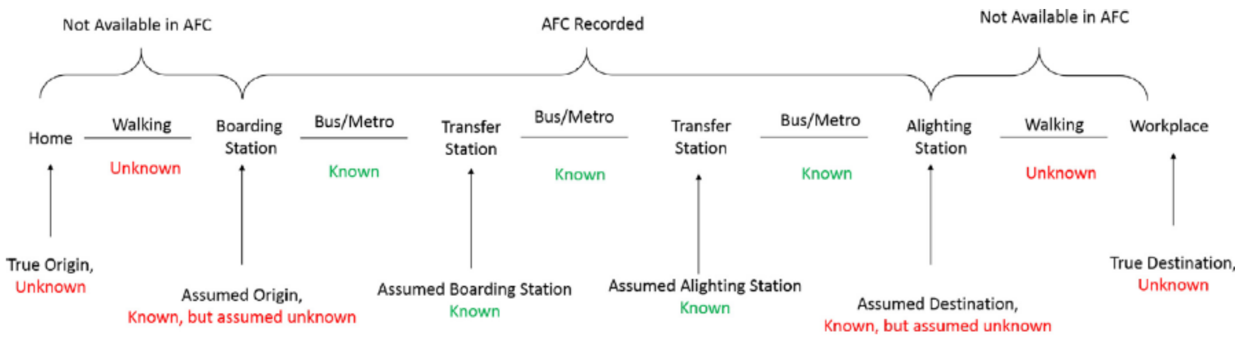


Figure 5. Typical Trip Travel Sequence

A solution to this problem returns the most likely OD for each traveler. Data collected from the AFC system, of the metro and bus routes within Seoul Metropolitan area for 12 weeks, was tested for the ODIP model. The Expectation–Maximization (EM) algorithm was also used in combination with OD analysis having a successful approach of inferring 89% of true ODs from the total user base (Wu et al., 2020).

Future research will focus on improving the efficiency of extracting eligible ODs from AFC systems and other sources to make the data-driven method more appealing. Identifying true ODs using the AFC systems could be helpful when developing new transit routes, because it will offer attractive solutions to the travelers when their preferences are being taken into consideration.

2.2.3.7 Assessing Transit Loyalty

The collected data through Automated Fare Collection (AFC) systems provide valuable information than can be used for different purposes. This study focusses on measuring the user's loyalty to public service based on the data analysis from smart cards system. In order to evaluate large amounts of data, transit agencies need to have access to continuous data on the behaviors of transit users. Determining loyalty or retention rate could be useful when developing marketing strategies, fare packages, or to increase attractiveness of the system (Trépanier et al., 2012).

Previous studies have shown that the customer level of satisfaction is related to transit loyalty. Better service frequency, lower waiting and travel times, and vehicle comfort and cleanliness are usually identified as key factors for retaining customers. This paper presents the application of a discrete time hazard model using data collected for 5 years, from January 1, 2004 to December 31, 2008 of a medium-size transit authority Société de transport de l'Outaouais (STO) in Canada. The STO operates 200 buses and provide services to 240,000 people where more than 80% of the users have smart cards (Trépanier et al., 2012).

The interest of this paper was to investigate the factors that can influence the usage of smart cards by analyzing the information collected per transactions such as the starting and ending month of individual cards and frequency of monthly usage of the card. In the case that the personal information was limited, then the following attributes could be used: average zonal socio-demographic, land use, and the transportation system. Using these attributes, a model can be developed of transit smart card usage as a function of the available attributes (Trépanier et al., 2012).

A Hazard model is then the appropriate one to determine the duration of a user's smart card. As stated in *Are transit users loyal? Revelations from a hazard model based on smart card data* (Trépanier et al., 2012), the concept of the hazard model relates to the fact that the probability to continue the use of a smart card by a user, at a particular time, is conditional to the probability of not cancelling the card before the end of the time period. For this case, the study only used adults' smart cards transactions, because they use public transportation more frequently than students as shown in Figure 6. Information was collected from more than 68,100 smart cards showing a mean average of 16.26 months. The average proportion of active months over life duration is 87.9% and reaches 100% for 6 out of 10 cards (i.e., 60% of the time). This means that some 40% of the users will interrupt their use of the system and skip one or more monthly fares during their overall transit lifespan.

From this study, it was deduced that different scenarios can affect, either positive or negative, the users' loyalty to public transportation (Trépanier et al., 2012):

- As far as the hazard rate, users from different regions present different characteristics on the usage of smart cards. For instance, suburb users have a lower positive impact on the hazard rate than other regions, as it seems that people who live far from the city are more loyal to public transit than people who live in the downtown areas. More populated areas have a negative impact on the hazard rate; however, it is shown than the usage of smart cards is longer in higher-density urban areas.
- Modal share of bike and walk is another indicator of transit loyalty because the higher the modal share of bike and walk, the higher the hazard rate and shorter the smart card usage. But, when referring to transit modal share, the opposite occurs. Higher transit modal share has a low effect on hazard rate and promotes the usage of smart cards.

- In the case of regional population structure, it is shown that that a younger population induces longer card usage and consequently a lower hazard rate. It is also deduced that a higher man-woman ratio increased the hazard rate and reduces card usage.
- Unemployment and low-income families have a higher effect on hazard rate and influence on the decreasing of the duration of smart card usage due to the termination and inconsistency of income, respectively.

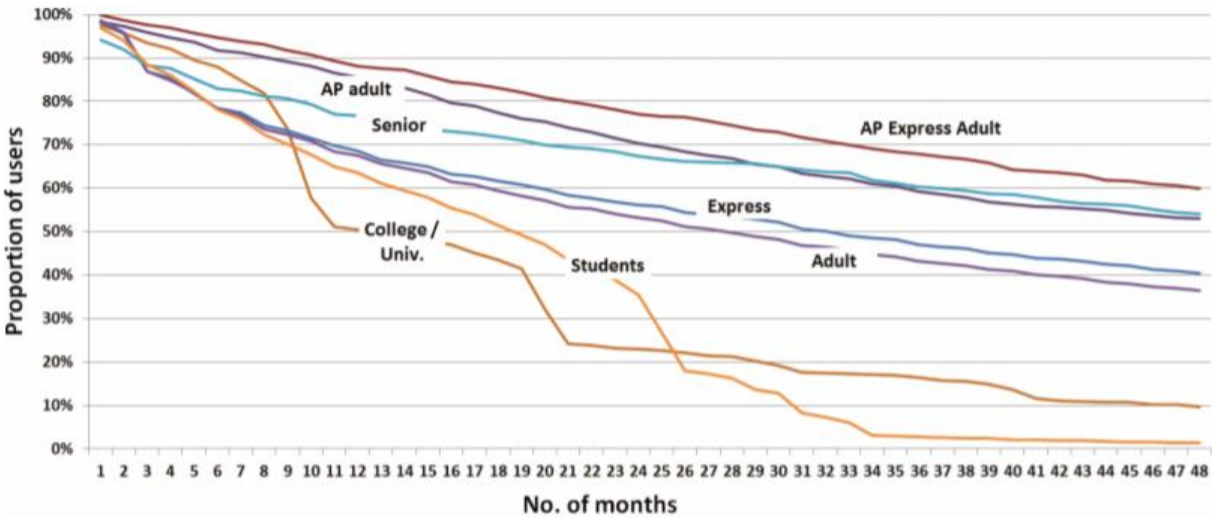


Figure 6. Retention Rates of Smart Cards Users

This study provides useful knowledge of the collected data of STO for 5 years. The effect on the hazard rate by different situations is crucial on determining the riders smart card usage. However, for further investigations, more attributes could be included to enhance the estimation of the hazard model. Defining user's loyalty level is important for the transportation system because it can be used by transit agencies to develop a more focused marketing strategy to increase the public ridership and properly establish fare rates.

2.2.3.8 Uses of Electronic Fare Payment Records

Demand of a better transportation service is growing every day. Planning and design of new routes to meet people's needs and demand has become a challenge for operators and transportation planners. The challenge it is not only adding more buses, but a study needs to be completed to develop the proper services to satisfy customers demands and comply with transit agencies requirements.

For this reason, using the data collected by Automated Fare Collections (AFC) systems, transit agencies could perform various investigations to determine and implement the needed services. Nowadays, various technologies are used to collect data because of their capacity of storing large amount of information. The automatic passenger counter (APC) systems are capable of recording arrival and departure time at each stop and can be analyzed off-line. There are also some automatic vehicle location (AVL) systems which, even though they record information in real time, they cannot be used for off-line analysis. However, these systems can be integrated providing new opportunities for improving the quantity, variety, and quality of data collected and archived (Sahin and Altun, 2007).

As stated in the Potential Uses of Electronic Fare Payment Records for Public Transit Agencies paper (Sahin and Altun, 2007), the study focuses on showing the vast information that it is collected through the AFC systems and that can be used to solve some of the problems that transportation systems are actually facing. The presented method is not for replacement of existing data-gathering technologies. However, it illustrates that in order of enhancing the current transit services, it is necessary to gather the necessary information to perform various analysis.

For the purpose of the paper (Sahin and Altun, 2007), Istanbul, Turkey was chosen as a case study. At that time, the city had 500 fixed bus routes and 7,834 bus stops throughout Istanbul. The dominant public transit agency was the Istanbul Electric Tram and Funicular Company (IETT) providing a service with 2,478 buses. The agency initiated an electronic fare payment system (Akbil) in 1995 using an iButton which is similar to a smart card. Each button has a unique and unalterable address (or user ID) that is laser-etched onto its chip inside the can. Each transaction in IETT stores the following information: user ID, fare category (exact/discounted/pass), payment amount, credit left, agency code, route number, bus ID, date, and time of day. The data was collected every 30 min from October 7 to 13, 2002 (Monday to Sunday) between 5:00 a.m. and 12:00 a.m. and a graph was plotted as shown in Figure 7, indicating ridership trends.

Transaction records of a bus at different stops are valuable information, because with the help of geographic information system (GIS), any particular route and distance between consecutives stops can be determined. Some of the problems that could be encountered such as a passing a bus stop without any boardings or determining the arrival at the last stop can be solved with the information stored in the GIS system (Sahin and Altun, 2007).

Public transit agencies are conducting various analysis to enhance the current transportation systems to meet the needs and demands of the passengers. Data from Smart cards provides useful information to transit planners that can be used for analysis of ridership and operations information. Many studies have taken place using the collected data from the AFC system by utilizing the user ID data of each transaction record and each individual's travel pattern also could be analyzed (linked trips). These linked trips can also help to determine transfer facility locations. There is a vast amount of information that can be used for several purposes to offer a better and improved transit service to the people (Sahin and Altun, 2007).

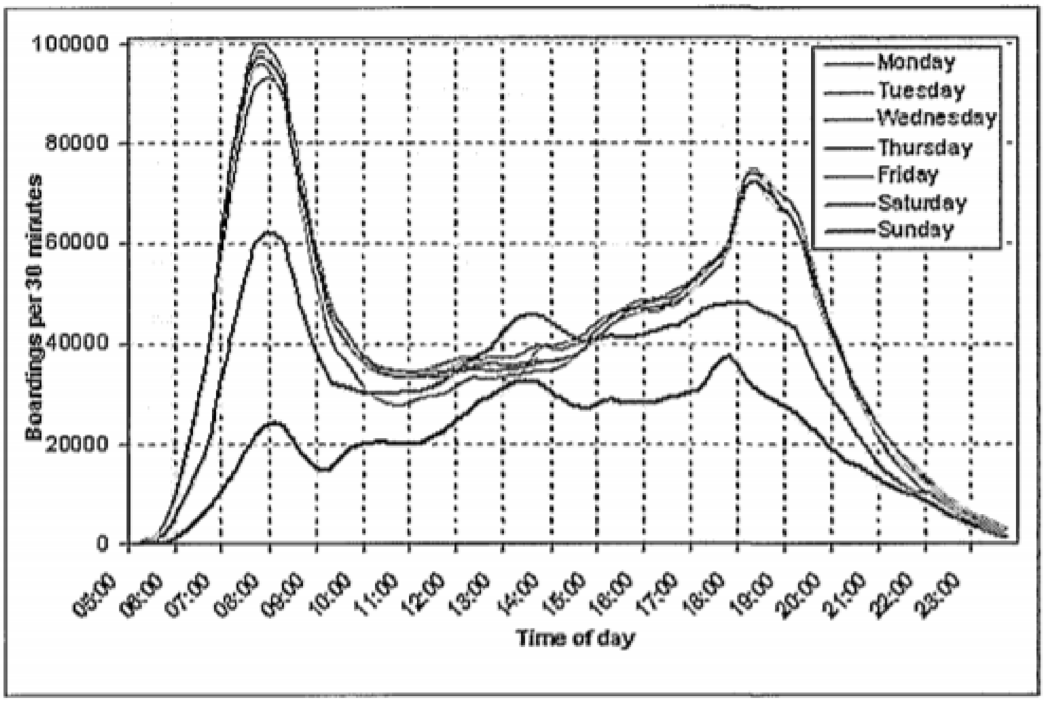


Figure 7. Ridership Trend

<u>2.2.3.9 Data for Bus Service and Operations Planning</u>

Technology has significantly evolved in the past decades and transportation systems have evolved along with it. Different systems have been created such as Automatic Vehicle Location (AVL), Automatic Passenger Counters (APC), and Automated Fare Collection (AFC) to automatize bus operations and providing a vast amount of data that can be used for transit planners to make better and accurate decisions related to transit service (Shireman, 2011). This study can be useful for transit agencies that want to expand their public transit service, but they have limited resources of buses and operators.

This thesis focuses on the importance of collecting data for an improved bus service planning process (Shireman, 2011). A portion of the MBTA network serving the Boston suburbs of Somerville and Medford is used as a case study. This consists of 18 bus routes in Somerville and Medford, Massachusetts, using several types of automated data collection systems (ADCS)-based performance indicators. It evaluates several service scenarios using GIRO's NetPlan software package, which is a sketch service planning and timetabling tool linked to its HASTUS scheduling system with a major goal of determining the number of required buses.

The results of the ridership and running time analyses were used as inputs into the bus service scenario planning process. First, the existing bus schedule was slightly changed allowing the service to adjust to any deficiencies encountered through an interlining process. Then, performing an analysis of the AVL data, current running times were adjusted to enhance reliability. In addition, based on demand, the frequency of two service change scenarios were modified, routes that serve the same route segments were synchronized, and selected changes in routing were incorporated (Shireman, 2011).

As per Shireman, (2011), there are many elements that affect the scheduled and observed running times:

1- All schedules must be set, so that a high percentage of trips can start the next trip on-time. However, if too much running time is allocated, then bus operators may reach timepoints early or may have to drive at a slower speed.
2- Congestion directly affects the trip time especially during peak hours in the peak direction. This has an impact on the scheduled cycle times.
3- Dwell time at bus stops can significantly affect the running times. When buses are crowded, the times for boarding and alighting are longer.

Many transit agencies set their running times using similar times from the mean observed running times. The half-cycle time (running time plus recovery time) is generally set at between the 85th - 95th percentile of the actual running times for a time period. Setting the half-cycle time at the 85th percentile means that there is a 15 percent probability the bus will arrive so late that it will be unable to start the next trip on schedule, because tardiness from one trip propagates to the next one. A bus at 95th percentile round-trip running time should be able to start the next trip on time if the boarding and alighting in terminals takes only a few minutes. Compared to half-cycle running time analysis, two to three minutes of buffer time (round-trip) have been added to the observed running times to provide some additional reliability when proposing new cycle and running times. Figure 8 shows the round-trip running time distributions by time of day for Route 83 using three months of data (Shireman, 2011).

As stated in the *Using Automatically Collected Data for Bus Service and Operations Planning* thesis, NetPlan can perform thousands of trips shifting iterations minimizing the number of vehicles required for a given schedule and consider interlining, as an opportunity to reduce the excess layover time in the network. Interlining can provide a solution to the problem of having limited buses by combining two or more routes that serve the same terminal.

From this thesis (Shireman, 2011), it was deduced that the collected data from the automated data collection systems (ADCS) play an important role in the service planning process, because it provides transit agencies the necessary information that they need to develop more efficient bus schedules using accurate running times. Further, ridership data from APCs or AFCs should be used to adjust service frequencies on bus routes with frequency changes being prioritized, so that the maximum wait time and scheduled delay savings can be achieved.

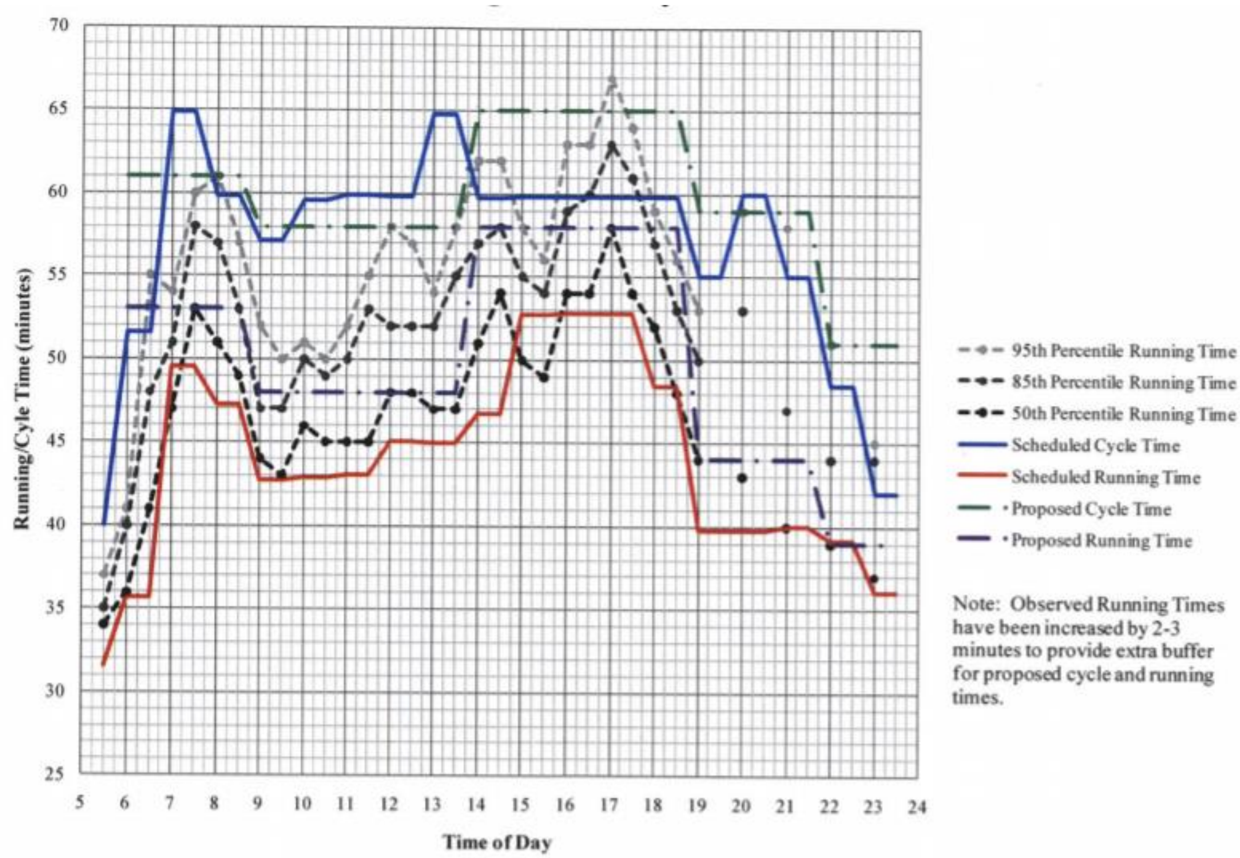


Figure 8. Running Time Analysis for Route 83

Nevertheless, it is recommended to determine the number of vehicles available in the network before making any decision, because changing schedules could affect not only the number of vehicles but also the crew requirements. This study suggests using clock-face headways, because these are easier for customers to remember, understand, and use. A recommendation is to consider the NetPlan tool to implement any modification needed like exploring the interlining option to enhance the service provided to the transit passengers (Shireman, 2011).

There are many research opportunities using the collected data of automatized systems particularly when the Somerville-Medford Green Line Extension Project is completed (Shireman, 2011):

- Analyze the bus service changes required for after the Green Line Extension opens.
  A new analysis will be needed when the Green Line Extension to College Avenue and Union Square is opened to determine the number of buses that will be needed to cover the demand.
- Create stop-level OD matrices. AFC and AVL data can be linked dataset to estimate the origin-destination matrices. The creation of OD matrices will provide detailed information of the route segment and stop level that can be used when setting the frequencies and revised routing.

- Automatically select running time periods. In this thesis, the running time periods were determined mainly by visual inspection of the AVL running time distribution plots. In addition, the HASTUS-ATP module includes algorithms that suggest the most appropriate time periods for a route once the available AVL data is analyzed. However, it is important to study the transitions between time periods on a more frequent basis to implement methods to adjust running and cycle times.
- Set the running time between timing points based on AVL data. The running time analysis focused on setting the end-to-end running times and terminal recovery times. However, little attention was given to the scheduled running times between timepoints.
  More research will be needed to study the schedule setting of running times between timepoints and running time between stops to maximize reliable passenger information and synchronization of routes along common route segments.
- Add more routes to the automated timetabling scenarios. The Somerville and Medford routes were studied separately from most of the other MBTA routes. For future research, the routes that were excluded in this study could be added to investigate possible interlining at terminals near the study boundary. Many people may benefit if interlining is implemented, not only because of the route extension but also it reduces the need of transferring from one bus to another. For this reason, for terminal nodes where there are many bus-to-bus transfers, synchronization factors can be applied so that some of these transfers can be accommodated by interlining.

The implementation of ADCS is an effective way of processing buses transactions and a valuable and inexpensive method of collecting information that can be used to enhance the transportation system. The collected data is useful for analyzing the demand and running times for bus routes, and the outputs of these analyses may be used as inputs into service planning scenarios. Transit service agencies should base their decisions utilizing this collected data because of its accuracy and the possibilities it offers of improving the public service (Shireman, 2011).

2.2.3.10 Travel Behavior Analysis

The transportation system has been affected in a positive way with growing technology. Various software has been developed to enhance the transit services such as the smart cards which not only facilitate the services offered to the people, but also generate a large amount of data that can be used by transit agencies for different purposes to keep enhancing the transportation system. As stated in the Travel Behavior Analysis using the Smart Card Data (Ali et al., 2015), this paper attempts to utilize the smart card data as input for large-scale activity based public transport simulation for analyzing travel behavior pertaining to transit users, using MATSim, an open source agent-based transport simulation package. MATSim is a dynamic simulation toolkit that is implemented in Java modular structure and can be modified to various conditions to handle different scenarios. At first, it was used as traffic simulation, but then a new model was develop giving reliable results and its functionality was extended to support detailed public transport simulation.

Transit ridership during Summer occupancy levels, calculate vehicle speed and in-vehicle travel time etc. A transportation service Korean Transport Institute (KOTI) in Seoul, South Korea was chosen as a case study since the service is the 2nd largest public transport modal share in the world with approximately 63% public transport ridership after Hong Kong where modal share is 90% for mass transit. The public transport system contains 19 urban railway lines and more than 400 bus routes in greater Seoul Metropolitan Area (SMA) (Ali et al., 2015).

Seoul Metropolitan Government introduced a new system of fare collection called "T-Money" where users must validate their boarding and alighting when using the network services except for transfers within the subway system. Each transaction records individual's information like card ID, boarding time, boarding station ID, alight time, alight station ID, transport method (subway, regional bus, circular bus, etc.), bus route ID, passenger type (adult, youth, children), total fare and total distance travelled. Data used were provided by the Korean Transport Institute (KOTI) for five weekdays from June 11 to June 15, 2012 with approximately 28 million transactions made on average working day. The methodology used to study the dataset was completed with simple coding and formulation, and spatial analysis was carried out using ArcGIS. The methodology steps are (Ali et al., 2015):

1) Assigning each transit stop a zone code based on its spatial location: Seoul is divided into 25 administrative districts called "gu", and those districts are further divided into sub-districts called "dong" with a total of 467 stops All the transit stops are first located on a map with district shapefiles as a base layer for assigning the zone codes to the stops.
2) Join the zone attributes with commercial statistical software SPSS: For each trip segment, xy coordinates and zone code attributes for each boarding and alighting are added into the database by rotating transit stations.
3) Differentiating trips and trip segments: A trip is defined as a movement from point of origin to the point of destination where activity is being performed. A trip may be composed of several trip segments and/or different modes; therefore, it is mandatory to distinguish between trips and its segments.
4) To study the travel behavior patterns, it is important to have the information of the trip starting and ending point and have in-depth knowledge about spatio-temporal distribution of trips. Several observations were made through this study such as the impacts due to the cease of educational activities.
5) Under bad weather conditions, ridership trends will be inclined to the use of metro as a preference choice of transportation.

OD matrices can be generating using different attributes like card type, trip start and end time, trip duration, trip regularity over a large period and transport mode. The places where the trips are generated could be defined by using socio-demographic data along with facility shapefiles for activity-based microsimulations to be performed in MATSim. Usually, the start and end trip of a passenger is located at their homes. It was also noticed that 91% of the total trips started within 1 km of the preceding trip's alighting location meaning there was a major probability that riders did not use any other transportation mode between trips such as taxi, walk, bicycle, or car sharing (Ali, et al., 2015).

Smart card systems provide practical information about a person trip but lack of a more personal perspective. Using the collected data, a model was created to simulate public transit activity-based travel demand for public transportation at a micro level providing accurate results, but further studies would be needed to identify the short trips made by the customers within the transfer window. For this reason, more advanced data mining techniques are required to define the travel behavior patterns and characteristics from the fare collection system dataset. However, this study (Ali et al., 2015) examines some of the users' attributes like waiting time and transfer volumes at different metro stations, that can be used by transit planners and operators to enhance the transportation system and maintain adequate level of service (LOS).

2.2.3.11 Mobility Patterns of Seniors, Children/Students, and Adults

In urban areas, commutes follow interesting travel patterns that are often stored in regional transportation databases. These patterns can vary based on the day of the week, the time of the day, and commuter type. Studying the pattern mobility of different individuals can help to make better decisions regarding the transportation and land use to contribute to a better understanding of urban dynamics. Going to work, school, or travel to other places are daily activities that can be tracked through the smart cards systems to allow further analysis. These activities over time could be developed into travel patterns that can be grouped by different ages for a better understanding. As stated in *Understanding Spatio-Temporal Mobility Patterns for Seniors, Child/Student and Adult Using Smart Card Data*, Huang and Tan (2014) propose a method to detect the underlying spatiotemporal mobility patterns by analyzing the use of smart cards among seniors, children/students, and adult commuters in Singapore.

The goal is to explore the relationship between smart cards users and the primary urban geography because it provides valuable insights to transit planners about the functionality of the urban dynamic. The main contributions of this study include (Huang and Tan, 2014):

1) Detecting human flow patterns using smart card records as the data source.
2) Using quantitative measures such as graph properties to obtain an overview of travel demand.

The fare collection system in Singapore was chosen as a case study. The analyzed data used over 36 million individual trips for one complete week from March 19, 2012 (Monday) until March 25, 2012 (Sunday). The passengers were divided into three different categories according to age distribution: Adult, Child/Student, and Senior citizens. The information recorded by the smart card is presented in Table 4 (Huang and Tan, 2014).

Table 4. Smart Card Collected Data

Field	Description
Journey ID	The unique number for a journey
Card ID	The coded number for a stored value card
Passenger Type	Derived based on the card type the commuter hold – Adult, Senior Citizen and Child (including students)
Travel Mode	Refer to transport mode of the ride – Bus or Metro.
Boarding_STOP_STN	Boarding bus stop for a bus ride, or station of entering rail system for a metro ride.
Alighting_STOP_STN	Alighting bus stop for a bus ride, or station of leaving rail system for a metro ride.
Ride start date	The date of a ride started.
Ride start time	The time of a ride started.
Ride distance	The ride distance in km.
Ride time	The time interval (minutes) between the boarding and the alighting of a ride.
FarePaid	Fare paid for the ride
Transfer_number	The transfer sequence number of a journey

The bus and metro stops were the main study places because they provide boardings and alightings volume (density) information and duration of relevant stops. The study was divided into 24 time-intervals for weekdays and weekends and temporal patterns are shown in Figures 9 and 10 (Huang and Tan, 2014).

From those figures, it can be deduced that in Figure 9, there are two peaks in the morning and evening for children/students and adults, whereas the seniors' pattern suggests they are not restricted to office hours. Figure 10 shows similar patterns starting with seniors and then followed by adults and children/students. The duration people stay at these locations, to conduct certain activities such as working, studying, social visits, etc. is calculated using the time interval between the alighting time at one specific station for a trip and the boarding time of the same station for the following trip.

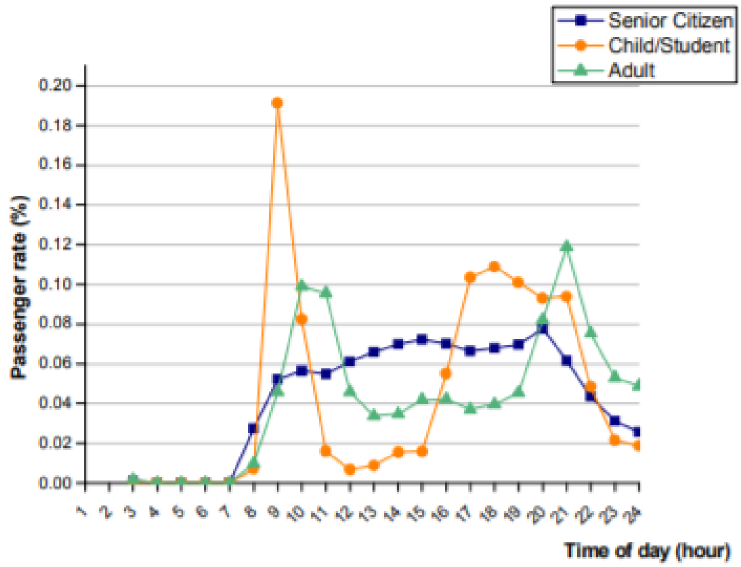


Figure 9. Weekdays Density

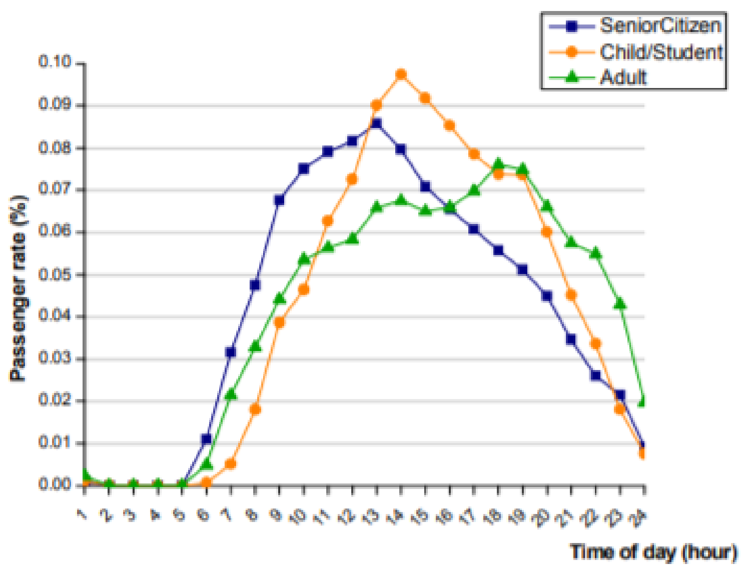


Figure 10. Weekends Density

The research method was based on a synthesis of network science and spatial analysis was developed to identify the mobility pattern variations between the three groups: Adult, Child/Student, and Senior citizens. Specifically, descriptive statistics, network analysis, and spatial analysis methods were presented. Descriptive variables were proposed as density and duration to detect temporal features of people and a directed weighted graph was defined to analyze the global network properties of every pair of the transportation link in the city during an average workday for all three categories.

In addition, to gain a better understanding of the spatial distribution of interested locations, an attractiveness index was selected as the attribute field to be projected into geographical space to determine hot spots, which reflect spatial patterns of individuals, and the eigen vector centrality (EVC) was selected as the attractiveness measure. EVC is known as influence measures, therefore, a node that has high eigenvector score is one that is adjacent to nodes that are themselves high scorers. An analysis was conducted for the three groups that presented some differences as listed below (Huang and Tan, 2014):

- Senior citizens preferred the Central Region and Northeast Region. This could be explained because the Housing and Development Board (HDB) which is a self-contained housing state and has a town center acting as a focal point for the shopping and entertainment needs of the residents located in that area.
- Children/students prefer the West Region, which contains several educational institutions.
- Adults prefers the Central Area where has most of the commercial areas.

This study explored methods to understand human mobility patterns with quantitative measurements of spatio-temporal features by analyzing the behavior patterns of senior citizens, adults, and children/students using smart card data. It was shown, their travel patterns follow a certain degree of temporal and spatial universality but also displays unique patterns within their own specialties on both bus and metro links. Some variables were proposed such as density and duration to detect temporal features and network centrality indices especially eigenvector centrality to compare spatial variations for the three categories. One of the limitations encountered was that the collected data was not enough because only one week of data was used. It is recommended to use a larger dataset to reflect a real situation of human flow over a period of time. For future studies, multiple data sources could de integrated to obtain a higher level of understanding of urban dynamics (Huang and Tan, 2014).

2.3 Data Mining and Big Data

2.3.1 Data Mining

Data mining involves dealing with large amounts of data and converting the data into useful information. It is often used to automatically find hidden patterns not easily found through conventional methods. As per *Mining Public Transport User Behaviour from Smart Card Data* (Agard et al., 2006), data mining functions are categorized as follows:

- Classification involves assigning labels to new data based on the knowledge extracted from historical data.
- Estimation deals with filling in missing values in the fields of a new record as a function of fields in other records.
- Segmentation (or Clustering) divides a population into smaller sub-populations with similar behavior according to a predefined metric. It maximizes homogeneity within a group and maximizes heterogeneity between the groups.

- Description and Visualization are used to explain the relationships among the data.

Data mining can be used to solve hundreds of tedious problems in a few minutes, compared to the lengthy amounts of time needed for human workers. Data mining is also part of Business Intelligence (BI), Online Analytical Processing (OLAP), Enterprise Reporting, and Extract, Transform and Load (ETL), and other related systems. Based on the specifics of each problem, one or more of the following data mining tasks may be utilized (Cevallos and Wang, 2010a):

Classification—Classification is a commonly used data mining task. Solving problems for customer attributes and ad targeting typically involve classification. During classification data are assigned into categories based on a predictable attribute. Each datum contains a set of attributes, one of which is the class or predictable attribute. Classification then searches for a class attribute as a function of input attributes. Common algorithms used for classification include decision trees, neural network, and Naïve Bayes.

Clustering—Clustering is also known as segmentation. The clustering task is commonly used to identify natural groupings of cases based on attributes. This can be an unsupervised data mining task with no single attribute used to guide the training process, as all input attributes are treated equally. Most clustering algorithms build the model through a number of iterations and stop when the model converges.

Association—Association is another commonly used data mining task. For example, association is used by businesses seeking to analyze which particular products are likely to be sold together. By identifying common sets of purchased items, businesses can develop strategies for marketing and advertising. Besides identifying frequent items, most association type algorithms also find rules. An association rule has the form A, B => C with a probability, where A, B, C are all frequent item sets. The probability, in data mining, is also referred to as confidence. In general, the probability is a threshold value that a user can specify before training an association model.

Regression—Regression is another popular data mining technique that can be used for modeling, forecasting, and trend analysis. For smaller datasets, regression can actually be performed on several computer programs, including Microsoft Excel. However, for huge databases with millions of records, data mining is a more appropriate solution. Regression is applicable for any number of variables and it can be performed for both bivariate and multivariate data sets.

Forecasting—Forecasting is another important data mining task. Forecasting is frequently used with a time series dataset, for example, the arrival times of planes at a major airport. The time series data typically contains observations that are chronologically dependent. Time variables for an airplane example might include arrival, cleaning, boarding, gate clearance, taxi time, runway wait, and take off information. Forecasting techniques are then applied to predict expected time trends between each stage of flight operations.

To maximize the benefits of using transit ITS/APTS data, data mining and business intelligence techniques are absolutely necessary. Therefore, it is advisable that transit agencies maintain these ITS systems and address any issues associated with the hardware, communication infrastructure, software applications, and available resources, so that they are able to take full advantage of the collected data from the transit technologies.

2.3.2 Big Data

Welch and Widita (2019) present a review of big data and its applications in public transportation. They state that the collection of big data has become more viable in the past decade, as alternative to traditional resource intensive data collection methods. Further, the availability of big data opens new opportunities for sophisticated analysis to support decision making. The use of big data to solve transportation problems provides new insights that were not be able to be done with traditional datasets.

Based on the literature review, previous studies indicate that big data has been mainly used to understand transit users' travel behavior and to transit service quality (Welch and Widita, 2019). Nevertheless, while the collection of big data continues to grow, there is still a need for improved analysis methods to take advantage of the data. To show the interest in big data, Figure 11 is a chart that depicts the number of Google searches with the word big data. The figure shows, on a proportional scale from o to 100, that starting in 2011, the United States showed an increased interest in big data with a peak at the end of 2014. However, the rest of the world appears to have started a little later in 2012, but continue to show high interest in the topic, as shown in the Goggle Trends Chart.

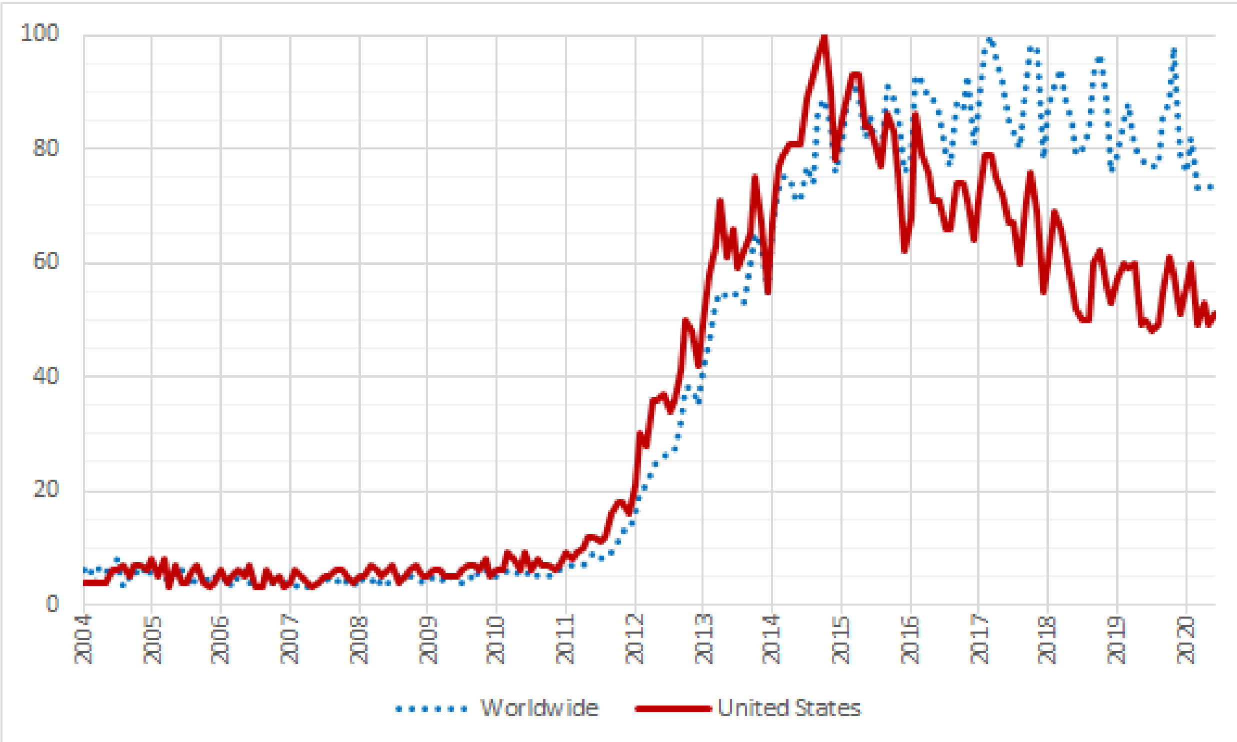


Figure 11. Big Data Google Search

In *Big data in public transportation: a review of sources and methods*, Welch and Widita (2019) attempt to fill the gap in transportation analyses research by highlighting the typical big data sources and studies and the challenges and opportunities presenting some suggestion for addressing the big data applications in public transportation. They identified GPS, smart card data, automated passenger counter (APC), automated fare collection (AFC), automated vehicle location (AVL), sensor data, mobile phone data, web data, and social media data as the source of big data in previous studies. Further, they grouped this study into six categories: Service/Performance, Travel Behavior, Travel Demand, Management, Resilience, Health/Safety, and Other Topics. Below is a description of the Service/Performance, Travel Behavior, and Travel Demand categories:

- Service/Performance: There are many studies that have dealt with the service and performance category, as big data can be used to help agencies evaluate their services and identify potential improvements. The uses of big data are extensive and can go from building an optimal route network with stop locations and visualizing the performance of their bus fleets to using AVL and AFC data to allow the determination of passengers' origins and destinations. In general, big data can be used by transit providers to conduct performance evaluation to help optimize the overall transit service.
- Transit Users' Behavior. Using big data is crucial for understanding passengers' behavior that can help by transit agencies in the decision-making process. Big data can be used to understand short-term and long-term trip patterns, extreme travel behaviors, and factors

that impact passengers' behaviors. With this information, transit agencies can make intelligent decisions in planning and scheduling the transit system. In addition, a combination of smart card with other datasets like GTFS enables the investigation of passenger trajectories and trip chaining. Lastly, the combination of smart card with survey data can help address the missing socioeconomic attributes in smart card data.

- Travel Demand. This is a category that is closely related to travel behavior. However, many studies in this area have focused mainly on the calculation of origin-destination matrices using big data approaches. Further, using AFC and AVL helps understand the trip itinerary and flow of transit passengers. Tools can be developed to improve the public transportation planning process. This can provide information such as passenger flow by day of week, hour of day, inbound and outbound trips, and the usage of passes or cards. Travel demand models have been used to predict passenger flows and to enable what-if scenarios. This can help with more sophisticated analysis such as the analysis of ridership changes due to variations in the level of service or modeling the transportation mode split for estimating the shares of car, public transportation, and walking.

The following list presents the data sources used in these studies:

- Automated data. This is obtained from automated sources such as automated passenger counters (APCs), automated fare collection (AFC), and automated vehicle location (AVL) systems.
- Global Positioning System (GPS). This is the data collected using satellites and ground receivers to identify geographic location.
- Mobile phone data. This covers the data collected from the mobile phones of transit users, including cellular data, call records, and location information.
- Sensor. This is the data collected from devices placed at specific locations to gather information such as Bluetooth sensors.
- Smart card data. This refers to data stored in smart cards. It includes identification, authorization, and payment information. As smart cards data is useful for transit data collection and facilitates the data analysis, they are becoming increasingly common as replacements for traditional magnetic stripe cards.
- Other systems. This includes data sources from other systems like radio-frequency identification (RFID), license plate recognition (LPR), web data extracted from websites, and social media websites. It is important to mention the low utilization of social media data, which is a disadvantage for big data and transit research. The user-created social media content provides good opportunities to gain valuable information from transit and non-transit users. However, the collection and use of social media can be complex with the increasing privacy concerns.

The use of big data for public transit is increasing as more agencies and researchers see potential opportunities. However, the utility of using big data to this point has been limited. Of the 81 papers reviewed in this study (Welch and Widita, 2019), only eight (less than 10%) used advanced methods like machine learning techniques. Nevertheless, using big data for understanding travel behavior and calculating travel demand are important elements for providing effective transportation services. Without having a good understanding about the passengers, agencies will not be able to provide optimal services. Using new methods for calculating travel demand and travel behavior that generate better information will allow public transit providers to make intelligent decisions on how optimize their services.

As the amount of data available from public transit continues to grow, the appropriate use of big data is essential to take advantage of the available information. However, further research is needed to expand knowledge base to allow transit agencies use the data for more informed decisions. The development of new techniques can benefit not only public transit, but also can be beneficial for research in transportation, city planning, and other related areas. With the availability of big data, there are many opportunities for research. So, more research studies are needed on how to effectively manage and utilize large amounts of data and how that can be used by transit agencies and decision-makers.

In the *Smart Card Data Mining of Public Transport Destination: A Literature Review* paper, Li et al. (2018) focuses on describing methods to deduce the public traffic destination estimation using the collected data through AFC systems. These systems are being implemented in major transit systems throughout the world. Collected information from smart cards, not only make charging and management more convenient, but also allow researchers analyze the data than can be used in different fields such as: transit riders' travel patterns, behavior analysis, performance assessment of bus transport reform, and planning of the public transportation system (Li et al., 2018).

Most of the smart cards system records only boarding information but there are some systems that are recording boarding and alighting locations like the AFC system in South East Queensland, Australia and the AFC system in Seoul, South Korea. There are three main destination estimation models to infer trip destinations such as (Li et al., 2018):

- Trip Chaining Model: This model is applied to infer the origin and destination trip tables and it is based on three basic hypotheses. Firstly, there is no other mode of transportation (e.g., car, motorcycle, bicycle, etc.) between two consecutive trips. Secondly, travelers will not walk a long distance when transferring. Thirdly, travelers will end their last trip of the day at the station where they begin their first trip of the day.
- Probability Model: This model is applied to estimate passengers' OD matrix. The model calculates the alighting probability considering travel distance and passenger boarding numbers. The model is improved by adding elements of station transfer capacity and land use levels around the station.
- Deep Learning Model: This model is used to solve traffic problems, such as travel mode choice predication, short-term traffic flow prediction and destination forecast of bus passengers with the help of Artificial Intelligence (AI) applications. In China, a modified back propagation (BP) artificial neural network was used to deduce the bus OD matrix, but

the dataset needs to be of a considerable size to achieve accurate results. In Korea, a more recent deep-learning model was developed to estimate the alighting number using the collected data (target boarding time, number of transfers, network travel times, generalized travel times) through AFC and land use information (residential floor area, commercial floor area, cultural floor area). Table 5 summarizes the advantages and disadvantages of each model (Li et al., 2018).

Table 5. Model's Advantages and Disadvantages

Destination Estimation Model	Advantage	Disadvantage
Trip Chaining Model	• Only need smart card data • The algorithm is relatively simple • Can infer the alight station of each passenger	❖ Difficult to validate the model with sufficient data
Probability Model	• The considerations are more comprehensive	❖ Only infer the total on-off passenger number without individual alight information
Deep learning model	• The considerations are very comprehensive • Can infer the alight station of each passenger • The model can be validated by numerous travel data	❖ Need abundant data ❖ Only appropriate for entry-exit system ❖ The algorithm is more complex

Different factors can intervene in the quality and performance of the models when estimating the destination of alighting stations using smart cards. These factors are divided into two main categories: most comprehensive factors (e.g., boarding locations and time, alighting locations and time) and transport network variable (e.g., bus stop and line density), but also including the land use variable. The Trip Chaining Model considers boarding locations/time and walking time/distance variables but does not include land use.

The Probability Model uses boarding number and network travel distance variables. Taken into consideration the variables used by each model, it can be concluded that the Deep Learning Model is the most comprehensive of the three, but it is impossible to point out if these are the factors affecting the results of these models, because there are other factors such as sample size or data cleaning that could be altering the results as well (Li et al., 2018).

Validation of the algorithms were completed with an entry-only system and the sample size is relatively small. Whereas using entry-exit systems, the validation of alighting data size was larger. Many problems can be encountered during data collecting like the ones caused by software, erroneous data, faulty hardware, or the users. Three types of problems were identified with the smart card data (Li et al., 2018):

- Missing data (containing entries or alighting data, whole transaction, next boarding information, or direction of travel)
- Illogical values
- Duplicate transactions.

Even though this paper does not explore any modification of the three models to infer the destination estimation process, it provides a detailed analysis that can be used for researchers to improve them.

2.3.2.1 Leveraging Big Data

The American Public Transportation Association (APTA) conducted a study on big data and the public transit industry. This was an APTA initiative that involved discussions with the public transit agencies and the private sector. As part of this effort, a survey was conducted that revealed that 94 percent of agencies are using big data techniques and methods to improve efficiencies. *APTA's Leveraging Big Data in the Public Transportation Industry* (Dickens and Hughes-Cromwick, 2019) provides valuable insights on how to use of big data to enhance public transportation services.

The term big data usually refers to large and continuous datasets that need to be structured before analyzing and used to inform decisions and solve problems. Many organizations in different industries have used large quantities of data to reduce costs, increase efficiencies, and improve decision making. Transit agencies are using the large amount of data collected by the different systems to improve their service and efficiency. These systems include automatic vehicle location, automatic passenger counting, and automatic fare collection systems. There are other sources of data that are not automatically created, but they are entered by agency employees. Breaking down the collected data is another important step to complete the necessary analysis of the information. In addition, new technologies like high performance computing (HPC) or Amazon Web Services (AWS) can help with processing big data. Having a cloud storage permits the sharing of data across an entire organization, allowing convenient access for different departments (Dickens and Hughes-Cromwick, 2019).

Big data can have many functions. Transit agencies are using it to improve and optimize operations and maintenance by allowing a comprehensive analysis of the equipment stating their current conditions. Once the analysis is completed, the AWS system can be used to create a predictive model of the equipment condition and performance to help the agency to foresee bus breakdowns. Predictive modeling is being used in other capacities to improve efficiencies and reduce costs. For example, predictive modeling can be used to monitor, manage, and react to potential operator absenteeism, allowing the agencies to provide backup in order to avoid service interruptions (Dickens and Hughes-Cromwick, 2019).

Transportation agencies are starting to use big data for safety and security purposes. Big data tools can be used to analyze the network environment and determine possible abnormalities allowing the agencies to take the necessary measures against any hacking threat. But also, the use of big data can play a part in improving the safety of public transit operations for passengers and people in the outside environment. Since collected data includes information about collision avoidance sensors and on-board cameras that monitor the streets and intersections used by the transit vehicles, a combination of these with the vehicle tracking information can identify streets and intersections where it is most likely to occur transit accidents (Dickens and Hughes-Cromwick, 2019).

Data collected through mobile apps can help the planners to make better decisions when designing transit service since it could collect information of trips and transfers and can even track trips among multiple regional transit agencies. Customers who use the smart cards system could even receive alerts of potential delays of their usual trips offering other alternatives (Dickens and Hughes-Cromwick, 2019).

There is also potential use of big data for external purposes such as advertising, advocacy, and public affairs. Aggregate transit data combined with other data sources can provide new insights on the value of transit networks. For example, Columbus, Ohio, made a connection between public transit and infant mortality in its Smart City grant bid, reducing infant mortality by 40 percent by 2020. The city laid out a transportation plan to improve bus frequency, add transit amenities, and develop a specialized transportation pilot to connect expectant mothers to healthcare. Other transit agencies are following the Columbus model and are researching for additional connections to improve their transportation services by linking them to jobs, health care, and schools (Dickens and Hughes-Cromwick, 2019).

However, public transit agencies have found several obstacles when analyzing the collected data, as reported by Dickens and Hughes-Cromwick (2019):

- The collection systems must be upgraded in order to obtain decent data.
- As agencies continue to expand their technology systems, there is a needed process for standardizing and developing commonality with the data. There must be a common reference point between the datasets for them to merge and derive meaningful information from them. This is because, some of the systems and applications may be from different vendors that can use different technology, making it difficult the access to the data.
- There are some datasets that transit agencies would like to access but may not be available. For instance, ride-hailing data is only rarely accumulated by transit agencies preventing them from taking advantage of additional information on customer travel patterns. Further, external datasets may not be well organized to be merged with other transit datasets.
- Privacy concern is also on the mind of public transit agencies when advancing their data applications. It has the potential to complicate some of the external data purposes listed in the prior section, including targeted advertising and precise route planning. Finding ways to anonymize individual data is crucial for transit agencies and private organizations take advantage of the data.

- Promoting a culture of data analysis should be a long-term commitment of transit agencies. However, this requires the support at the leadership level and requires adequate funding for technology systems and data analysis tools. It also requires adequate training and the hiring of qualified employees who have the skills to manage the technology systems and analyze the data.

What is evident from this study is that collecting and presenting quality transit data is becoming more important, particularly from a funding perspective. In some grant applications, the Federal Transit Administration requires that applicants identify the data that would be collected to inform decisions before, during, and after project implementation of the proposed project. Learning from other agencies' experiences is important because it can help inform transit agencies about new platforms and technologies. APTA encourages transit agencies the use the big data because of its many benefits. Big data cannot only help the transit agencies to make better decisions, but also can help improve their service efficiency (Dickens and Hughes-Cromwick, 2019).

2.4 Relevant Experiences

This section presents selected case studies from national and international experiences on the use of AFC data. The key findings from these experiences will be presented in the lessons learned section of this report.

2.4.1 North American Experiences

2.4.1.1 Boston, Massachusetts

Understanding Fare Increases

Automated Fare Collection (AFC) systems are being implemented in the majority of transit systems and provide useful data that can be analyzed for a variety of purposes. Transit agencies, due to budget issues, usually increase fares or reduce services as mechanisms to raise revenues or to lower costs. These changes can have a significant impact on the transit system. Pincus, K. S., 2014 on the *Analysis of the 2012 Massachusetts Bay Transportation Authority* thesis presents a study on the 2012 fare increase by the Massachusetts Bay Transportation Authority (MBTA). Data from the MBTA AFC system is used to understand of the impacts of the fare increase.

Elasticity of demand or fare elastic criteria is important for the estimation of ridership and revenue because it can measure the reaction to different fare changes. It measures the response of ridership to changes in fare levels or service quality. Fare elasticity with respect to demand is negative because fare and demand are inversely related. In the thesis, it is mentioned that fare increases will generate revenue when demand elasticity with respect to fare is between 0 and -1 and that revenue will decrease if the elasticity is less than -1. They also mentioned that elasticities tend to be greater for larger fare increases. Lastly, the commonly used Simpson-Curtin rule of thumb is included: the elasticity of demand with respect to fare is approximately -0.3. The Simpson-Curtin elasticity rule can be used to estimate overall changes in ridership due to fare changes.

When dealing with fare elasticities, different methodologies can be used. Below are three approaches to estimate fare elasticities:

- Preference surveys that can be used to evaluate expected responses to proposed changes, but the response may be different than actual reactions after a change is made.
- Shrinkage analysis measures ridership before and after fare increase by calculating the ratio of the percent change in demand to the percent change in fare.
- Econometric studies can incorporate other factors of the transit system and user population when estimating ridership demand based on historical data.

The objective of this study was to have better insight of the fare increase impacts on people's behavior regarding transportation habits by using the MBTA data. The system experienced a fare increase of 23% based on fare type and trip category on July 1, 2012. The periods chosen to complete the analysis of the fare increase were the period of April – June 2012, which occurred before the fare increase, and the period of April – June 2013, which served as a year-over-year comparison to account for seasonal effects. The September – November period was also analyzed in 2012 and 2013 to provide a baseline year-over-year comparison in absence of a fare increase. Each fare transaction was characterized by fare type category based on the type of card used to pay a fare. Consequently, the number of trips made with Regular CharlieTickets (stored value) and Regular CharlieCards (stored value) decreased year-over-year following the fare increase, while the number of trips made with Regular Monthly Passes remained relatively constant (Pincus, 2014).

When analyzing the data, it is important to take into consideration that one card or ticket does not necessarily belong to one person only and it can have multiple functions depending on the situation. For this reason, the number of active cards is not an accurate value, but it gives an idea of people's trip patterns. Four classifications were defined to analyze the individual card level usage for Regular Monthly Pass and Regular CharlieCard (stored value) customers:

1. Average weekday unlinked trips per week measures the frequency of trips made during the work week.
2. Average unlinked trips per week was calculated using the same methodology as the previous metric except it also includes usage from weekend days.
3. Average weekdays with usage per week measure the intensity of usage during the work week.
4. Average days with usage per week measure the intensity of usage over all nonholidays, including weekends.

Analysis of the average weekday trips per week showed that before the fare increase, the Regular Monthly Passes transactions were very similar for each time period year-over-year, while the Regular CharlieCard (stored value) transactions were a little higher. Two longitudinal panels were created to study the changes related to cards usage before and after the fare increase (Pincus, 2014):

- The Fare Increase Panel consisted of cards active in April – June 2012 and April – June 2013 to observe changes over the time period of the fare increase.
- The Baseline Panel of cards active in both September – November 2012 and September – November 2013 was created for comparison of individual year-over-year usage changes without a fare increase.

Cards' usage in the corresponding three-month period in 2012 suggested that the fare increase did not make a significant impact for the customers in the Fare Increase Panel, as it did with users from The Baseline Panel. This evaluation shows that the changes related to the active cards outside the panels are difficult to monitor with the AFC data. The panels were also used to study the heterogeneity within user groups with respect to the fare change. Other factors were also taken into consideration such as internal factors of individual frequency of use in 2012 and participation in the corporate program for monthly passes, as well as the external factors of weather, employment, and gas prices (Pincus, 2014).

Through this study it was concluded that the fare change did not made a big impact on regular users, it showed the trips lowered by 1 or 2 travel transactions. Also, the relationship between ridership and the external factors were also evaluated giving as a result that those factors did not have substantial consequences for the studied period, but it could be significant for a longer one (Pincus, 2014).

A regression model was developed to try to separate the effects of external factors from the internals ones on the dependent variable of change in average weekly trips from 2012 to 2013. Low R-squared values for both models indicated the heterogeneity of this variation of the dependent variable could be due to other factors like income, home location, and car ownership over time besides the ones included in the models. Measuring the elasticity of the total demand using the collected data, it ranged from -0.25 to -0.11 with the exception of Senior/T.A.P. CharlieCards, which had an elasticity of -0.85, and Student CharlieCard (stored value), which had a positive elasticity. Table 6 present the calculated fare elasticities for the different fare categories (Pincus, 2014).

Collected data through the AFC system could help to improve some of the problems the transit system is facing nowadays. Further studies need to be completed to have a better comprehension of the connection between a rider and a smart card. Analyzing the AFC could provide valuable insights of the impact of the fare changes and its consequences and could suggest possible fares for distinct transportation methods. Pincus, 2014 suggests that future research is needed to have a better understanding of the relationship between a user and a smart card or ticket in the system, as this would enhance the analysis of the impacts of the fare increase. The use of surveys or other tools could assist with the data analysis and understand the usage of the different fare types and shifts between the different fare categories in response to the fare increase.

Table 6. Fare Elasticity of Total Trips per Week by Fare Type

User Type	Medium Type	Total Unlinked Trips per Week				Elasticity
		2012	2013	Change	Percent Change	
Regular	CharlieTicket (SV)	323,887	310,057	-13,830	-4%	-0.15
Regular	CharlieCard (SV)	1,387,664	1,321,019	-66,645	-5%	-0.25
Senior	CharlieCard (SV)	114,191	104,546	-9,645	-8%	-0.11
Senior/ T.A.P.	CharlieCard (SV)	4,177	1,296	-2,881	-69%	-0.85
Student	CharlieCard (SV)	63,762	83,855	20,093	32%	1.39
T.A.P.	CharlieCard (SV)	118,230	98,438	-19,792	-17%	-0.21
Regular	1 Day	15,541	16,279	738	5%	0.21
Regular	7 Day	705,203	706,070	867	0%	0.01
Regular	Monthly Pass	2,012,312	1,983,855	-28,457	-1%	-0.08
Senior	Monthly Pass	96,748	106,734	9,986	10%	0.26
Student	Monthly Pass	230,336	237,033	6,697	3%	0.12
T.A.P.	Monthly Pass	118,909	133,369	14,459	12%	0.30
Regular	Cash	126,547	100,994	-25,553	-20%	-0.69
Senior	Cash	170,725	141,059	-29,666	-17%	-0.21
Student	Cash	161,569	149,606	-11,963	-7%	-0.27
T.A.P.	Cash	8,009	6,102	-1,907	-24%	-0.28

2.4.1.2 Minneapolis/St. Paul, USA

Trip purpose Inference

Using smart card data from Metro Transit in the Minneapolis/St. Paul metropolitan area, a research study demonstrates the process of trip purpose inference. Lee and Hickman (2014) make use of massive amounts of smart card transaction data to derive information about the behavior of transit passengers. In their research, they show how the automated fare collection data can be used to infer trip purpose and to reveal travel patterns.

Automated fare collection (AFC) systems generate a huge amount of data that contains spatial and temporal information on the characteristics of travel patterns. With such data, many questions can be answered:

- Can AFC data be utilized to assist transit service planning?
- Is it possible to replace costly origin-destination surveys with AFC data?
- Can the weakness of AFC data, regarding the lack of information on passengers' trip purpose and/or socio-economic characteristics, be addressed?

As opposed to the plentiful research on the use of AFC data for various applications, Lee and Hickman (2014) in *Trip purpose inference using automated fare collection data* stated that there is limited research on the effects of these applications. Therefore, to contribute to this line of research, the goal of this study is to explore and implement a potential method of deriving trip purpose from AFC data. The paper uses farecard transaction data for deriving useful information about transit passenger behavior such as trip purpose or activity. Under the assumption that detailed land uses within transit catchment areas or adjacent to origins and destinations affects travel decisions, trip purpose can be derived in combination with user information and time/space consistency. Further, to infer trip purpose, the assumption is that every transaction is made in a sequential trip chain. The following are the key assumptions used in this study:

- The destination of a trip is also the origin of the following trip.
- The last trip of the day ends at the origin of the first trip of the day.
- Transit users do not walk a long distance to board at a different stop from the one they previously alighted.
- Transit users do not use any other modes within their given sequence of daily transit trips.

This study focused mainly on mandatory activities/purposes like work and school-related trips, and other. In this study, a series of rules were examined for the trip purpose assignment. To build the rules for the Trip Purpose Assignment Process (TPAP), a cluster analysis was conducted using the Statistical Package for the Social Sciences (SPSS). Figure 12 presents a rule-based decision trees for trip purpose assignment process.

A decision tree classification algorithm was used because it makes it easy to construct rules for making predictions about individual cases and it also provides validation tools for exploratory classification analysis.

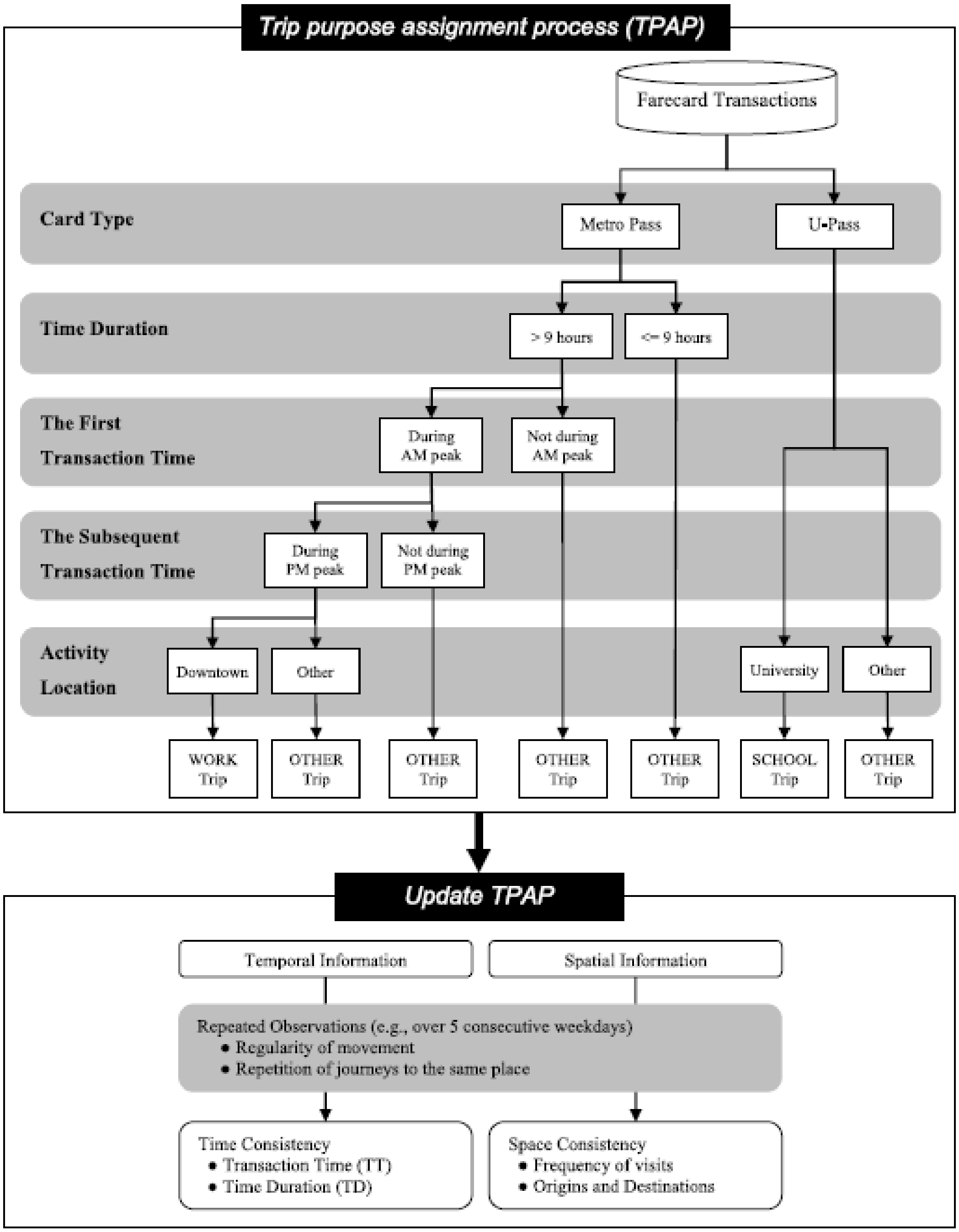


Figure 12. Decision Tree for Trip Purpose Assignment

Understanding transit trip patterns and behavior of transit users opens new opportunities in public transit planning and research. This study provided preliminary findings on the use of the farecard transaction data for deriving useful information about transit passenger behavior. The utilization of large datasets and various other data sources, including AFC, GTFS, and parcel-level land use data, supports more quantitative and direct measures to infer individuals' trip purpose. It was found that it is possible to successfully infer passengers' trip purpose and activity with limited resources using the AFC data. It is more focused on useful features of the farecard transaction data, such as user characteristics and spatial and temporal information, which contribute a great deal to the development of heuristic rules for inferring trip purpose. Even though the transaction data do not contain important information about the passengers and their trip purposes, inferences can be made through the trip purpose assignment process. Instead of looking at travel patterns at any point in time or at a particular location, it is necessary to consider repeated observations in temporal and spatial dimension for better understanding of travel patterns.

The results demonstrated the benefit of using an accurate learning algorithm. After building the decision tree, the proposed model can be used to predict other transaction data. As stated in this research, "making sense of transit users' trip purpose from the farecard transaction data is still unexplored territory". Although this research is enough to open a new promising line of investigation, more classification processes, interpretation, and validation efforts against on-board survey data still needs to be done. Other learning algorithms with diverse settings and further evaluation of the learning effectiveness are being explored, Lee and Hickman (2014).

2.4.1.3 Montreal, Canada

Fare Structure and Social Vulnerability

As stated by Verbich, and El-Geneidy (2017), the research on social equity, as it relates to transportation, commonly deals with the way residents in a region have access to desirable destinations. However, there is limited information on the relationship between transit fare structures and equity. In the *Public transit fare structure and social vulnerability in Montreal, Canada* paper, Verbich, and El-Geneidy (2017) examined transit fare purchases in Montreal, Canada. In their research, they found out that fare vendors in low-income neighborhoods and/or neighborhoods with a high proportion of unemployed residents are likely to sell more weekly passes than vendors in high-income neighborhoods and low rates of unemployment. As individuals residing in marginalized neighborhoods are likely to spend more money on transit fares over the course of a month, the findings of this research raise concerns regarding the financial burden that the existing fare structure in the city of Montreal imposes on low-income individuals. The methodology and findings from this research study can help address concerns with providing an equitable transit system.

The focus of transit research has focused on scheduling and operations or travel behavior and mode share. However, only a few studies have explored fare structures and purchasing, in particular in the context of social equity (Verbich, and El-Geneidy, 2017). Therefore, they investigated the relationship between the purchases of different transit fares and income and unemployment, which can be indicators of social equity. Their research studied purchases of transit fares using the OPUS card that is the smart card of the transit agency in Montreal, the Société de transport de Montréal (STM).

They modeled the number of total monthly fares purchased, total weekly fares purchased, and the number of riders who purchased three or more weekly passes. The hypothesis was that that some low-income individuals and unemployed residents may not be able to purchase monthly passes because of the high upfront cost and purchased weekly passes instead. It was expected to find concentrations of weekly fare transactions mainly in low-income neighborhoods. A total of 1,010,720 fare purchases were included in this study. These fare transactions, generated by 602,609 unique OPUS cards, that took place at both at transit stations (70 locations/Metro or transit stations) and points of sales (363 vendors). Figure 13 shows the data flow process for preparing the data for this project.

In the demographic and spatial variables section, it is mentioned that to link the neighborhood characteristics to each vendor, 500-m buffers around the centroid of every vendor's postal code were created using the Montreal street network. These buffers were intersected with census tracts to assign median-income households around each vendor. The same procedure was used for metro stations but using a 1-km network buffer. This assumes a larger draw of transit stations, instead of using the assumed walking distances to transit in Montreal. A similar procedure was used to calculate the population within the buffer. The unemployed residents were also considered that they would be prone to purchase weekly passes due to the travel uncertainty of the inability to afford a monthly pass. Finally, variables related to the built environment (metro stations, hub stations, bus stops, commuter stations, and other vendors) were determined and the distance to the downtown center point (CBD) was calculated using network analyst in a GIS.

Considering the non-normal and skewed nature of the dependent variables, transit fare purchases, and the fact that the dependent variables are count data, a negative binomial regression modeling was used. In addition, to account for the variability in transit fare sales between points of sale located within the same census tract, multilevel regression modeling was used when appropriate. Three different models were developed to determine how different variables contribute to the purchasing of different fare types. Further, other variables that are considered as proxies or indicators of social vulnerability were also tested. This included immigration status, educational achievement, and worker skill level. The preliminary results revealed that greater social vulnerability predicted greater fare purchases of all types, with a larger number of weekly fares rather than monthly fares. It was noted that all the models were tested using different model specifications and modeling techniques and key variables like income and unemployment were found to be stable and statistically significant.

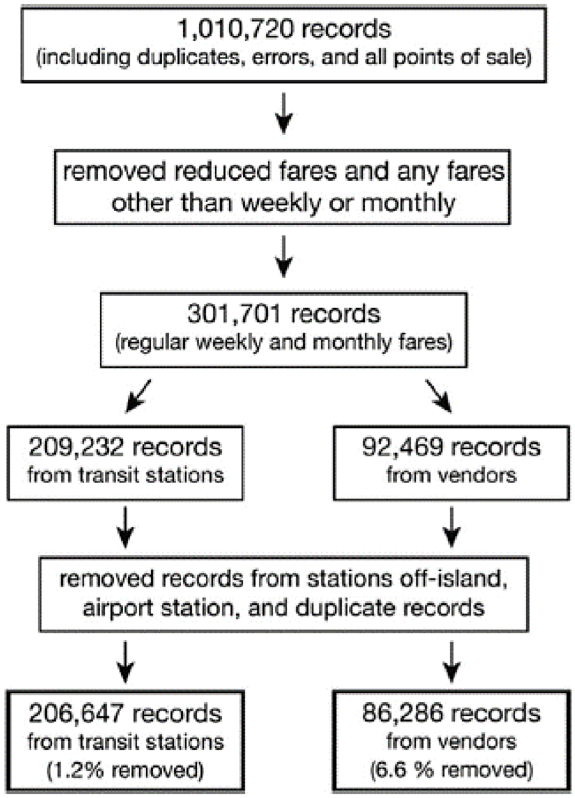


Figure 13. Data Preparation Flow Chart

The findings of this research highlight the necessity of conducting fieldwork to uncover the purchasing habits of vulnerable riders. In addition to low-income individuals, there are other vulnerable populations that are transit-dependent, such as single-parents, immigrants, and individuals with low educational attainment. This research indicates that these groups may also purchase multiple weekly fares as a substitute for monthly fares. Besides the time frame limitation of using only a single month of purchase data, there was also the use of neighborhood-level variables as proxies for demographics of fare purchasers. STM uses smart cards, but do not collect personal information or link that information to the smart card usage. Therefore, in future studies it is advised to collect individual and household demographic information to help stratify income groups and determine the impacts of fare purchases. Future work could also attempt to model single fare transactions and its potential relation with social vulnerability (Verbich, and El-Geneidy, 2017).

2.4.1.4 New York, USA

Prioritizing Bus Schedule Revisions

The New York City Transit (NYCT) operates the majority of New York City bus routes. The network is a crucial feeder for the subway system and function as a primary moving method in areas where the services are abysmal. Due to new developments, the public transit has been affected by the impact of the ridership patterns and bus speeds on many routes. However, with the implementation of Automated Fare Collection (AFC) systems, the service has been adjusted to the recent changes (Coleman et al., 2018).

Planning and scheduling routes were mainly based on manually collected data by traffic surveyors, a time consuming and costly process with inaccurate results. With an integration of AFC and automatic vehicle location (AVL) systems, transit agencies have been able to develop a model that determines passenger boarding and alighting locations at individual stops and trip levels (Coleman et al., 2018).

NYCT operates 235 bus routes which constitute over 650 distinct schedules. The implementation of smart cards has offered new insights to the schedule making and revision process making it feasible by analyzing the most recent data and allowing it to perform changes where needed. Along with the new development of NYCT, a combination of AVL data, AFC data from the MetroCard system, and general transit feed specification (GTFS) schedule data, provides detailed, stop-level boarding and alighting information for every trip on a given day (Coleman et al., 2018).

Even though the manually collected data was upgraded to AFC systems, it is still used to validate information from a specific route where the data could have large errors or higher fare evasion rates. This study allows bus planners and schedulers analyze the collected data to have a better understanding of passenger's behavior and demand to perform the necessary routes changes and headways, and also helps to create a list of recommended routes for revision based on their limitations (Coleman et al., 2018).

Coleman et al., 2018 explored three case studies that included route B32, M20 and S93 to demonstrate how the new methodology reduces the gap between changes in a route's performance and consequential schedule update. Route B32 was analyzed because it experienced various changes over the past five years. The route serves Williamsburg and Greenpoint in Brooklyn, and Long Island City in Queens. This route was introduced in September 2013 with the intent of meeting the increasing demand in the area showing that ridership on this route has increased 10% between years 2015 and 2016. Contrary to route B32, route M20 which runs in Manhattan from 66th Street down to South Ferry at the southern tip of Manhattan, experimented a decrease of 40% in ridership in 2016 from 2011. This was the result of a service reduction during peak hours negatively impacting passenger volumes. Route S93 runs between Staten Island and Brooklyn across the Verrazano-Narrows Bridge with limited stops. One of the stops is at the College of Staten Island (CSI), a campus of 13,775 students which helped to double ridership between 2012 and 2016. Due to the increased ridership, the schedule for this route suffered some changes in order to cover the growing demand.

Recommendations were also provided related to the constraints of the buses to find a balance between adding more routes and making reductions wherever it is necessary (Coleman et al).

The methodology evaluates schedules on two metrics (Coleman et al., 2018):

- the ratio of passenger volumes to capacity (V/C).
- the difference between actual and scheduled running times.

As stated i*n A Data-Driven Approach to Prioritizing Bus Schedule Revisions at New York City Transit*, both the capacity and running time measures may fail for two possible reasons: 1) having insufficient capacity or scheduled running time, or 2) having surplus capacity or running time. Identifying routes with too much capacity or running time is crucial because the extra busses can be redistributed to the areas where routes are insufficient for providing adequate service.

NYCT has taken an important step to make the schedule revision process faster and more responsive to changing conditions (Coleman, M. et al., 2018). By using the collected data from AFC systems, schedules could be adjusted to meet people needs and demands in an effective way by expanding the services to areas where they lack public transportation.

2.4.2 International Experiences

As there are many research papers on the use of AFC and other transit ITS data that are based on international experiences, this section presents a selection of this research. The goal is to learn from these experiences to identify key issues, methodologies, or findings that deal with interesting concepts that can be helpful in achieving the objective of this project.

2.4.2.1 Beijing, China

Carbon Emissions

The data collected from Automated Fare Collection system will help to provide new insights on how to alleviate urban traffic and consequently lessen gas emissions because traffic congestion makes not only people's lives difficult but also increases carbon emissions to the environment which is not healthy for human beings. To moderate this problem, Beijing, China has implemented a restriction policy to decrease the number of cars on the streets, for example: from Monday to Friday, automobiles with license end numbers of 1 or 6, 2 or 7, 3 or 8, 4 or 9, and 5 or 0, sequentially, would cease traveling on public road spaces (Zhang et al. 2017).

Since Beijing applied this kind of restriction, it was used as case study evaluating 10 consecutive weeks of smart cards data provided by the Beijing Transport Committee. Analyzing this data could help to develop a method to calculate the CO_2 emissions to help the transit agencies take proper measures in addressing this issue to reduce the greenhouse effect and improving air quality. But in order to have an accurate algorithm, the AFC data was combined with a geographic information system (GIS) (Zhang et al. 2017).

As result of the restriction policy, private cars users should have turned to public transportation and therefore by using the collected data from AFC system, an estimation of the carbon emissions before and after the policy implementation could be projected. Three problems must be taken into consideration when evaluation the carbon emission reduction to the environment (Zhang et al, 2017):

1. How to get the number of passengers turning to public transport because of the restriction policy based on continuous historical AFC data.
2. How to collect information about paths that passengers use in entering and leaving the public transport system, namely, sections B and D in Figure 14.
3. How to collect the complete trip path information for car users without the restriction policy in place, namely, section A in Figure 14, as well as the public transport trip paths to which these drivers turn with the implementation of restriction policies, namely, section C in Figure 14.

Two approaches were estimated (Zhang et al, 2017):

- Top-down approach which estimates CO_2 emissions using aggregate data for total energy consumed or the size of the vehicle fleet and average kilometers traveled per vehicle (VKT)
- Bottom-up approach which estimates emissions from fewer aggregate travel attributes, including trip frequency, mode choice, and vehicle kilometers traveled for each trip.

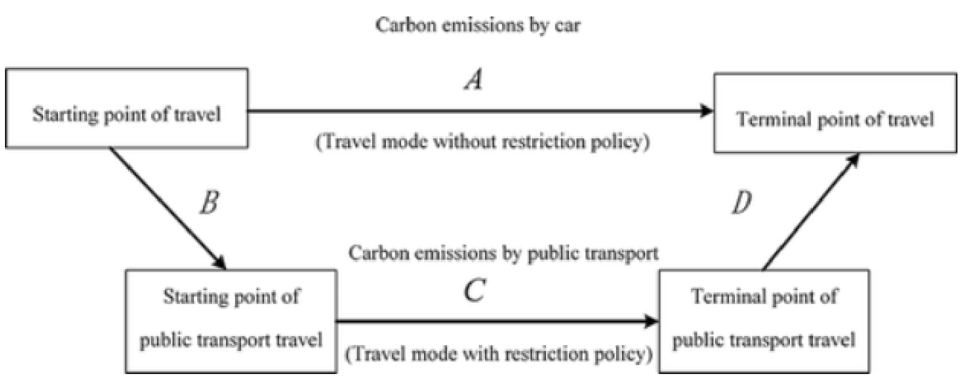


Figure 14. Travel Mode Shift and Emissions Reduction

As stated in *Evaluation of carbon emission reductions promoted by private driving restrictions based on automatic fare collection data in Beijing, China*, mining long-term historical data from public transport cards according to trip regulations could help to estimate the number of drivers turning to public transportation due to the restriction policy. From the collected data through the AFC system, the starting and end point of the trips could be obtained, and carbon emission levels could be calculated. But calculating the carbon emissions requires the distance of passenger’s trips and even though travel distances could not be obtained from AFC system, the routes could be calculated using TransCAD which is a geographic information system (GIS) designed to store, display, manage, and analyze transportation data.

It was appreciated that 85% of the passengers either chose to walk or ride using the metro system which displays the result that subway station relatively meets the needs of their clients. The study also shows, riders chose Metro because it is more convenient and punctual, thus it can be concluded that improving the public transport system will attract more customers and help to reduce the CO_2 emissions (Zhang et al. 2017).

The restriction policy implemented in Beijing forecasted favorable results because many people turned from private cars to public transit resulting in carbon emission reductions and an alleviation of traffic jams (Zhang et al. 2017).

Jobs Housing Relationships

Long and Thill (2015) in *Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing* presents a perspective for spatiotemporally analyzing dynamic urban Systems using Location Based Services (LBS). They argue that research has investigated urban dynamics using LBS, but less attention has been paid to the analysis of urban structure in particular when dealing with commuting patterns using smart card data that is are widely available in most large cities in China. The research paper combines bus smart card data with a household travel survey and parcel-level land use to identify job-housing locations as well as commuting trip routes in Beijing. The results are aggregated into the bus stop and traffic analysis zone (TAZ) scales.

The paper identifies job–housing location and commuting patterns in Beijing using smart card data that stores daily trip information of bus passengers. It proposes and implements a method for deriving commuting patterns from smart card data to provide information to city planners and transit managers about patterns of transit usage across space and time as well as about mobility patterns in the city region.

The widespread use of location-based services and positioning technologies have enabled the creation of large-scale and high-quality space-time datasets. This has created new opportunities to better describe and understand urban environments, with the analysis of the relationships between housing and jobs. Additionally, the geo-tagged smart card systems generate data necessary to analyze urban spatial structures. This paper presents a showcase for the use of smart card data for addressing the job–housing interaction and facilitates the commuting pattern analysis. The data used in this project included bus routes, bus stops, Beijing TAZs, Geographic Information Systems (GIS) layers of bus routes and stops, and smart card data. Land-use type for each land parcel in the Beijing Metropolitan Area (BMA) is also introduced to identify home and job locations. The land-use pattern is used to calculate the probability of each bus stop servicing a home or job location. Lastly, travel behavior surveys that were conducted in 1986, 2000 and 2005 in Beijing were also included to set rules for identifying job–housing location related to commuting trips.

The data processing and analytical approach included data pre-processing, identification of home and job location, and commuting trips based on the identified home and job locations. This yielded the following results: job–housing locations identification, aggregation of the bus stop and TAZ scales and comparison with observed data, the identification of commuting trips and comparison with the 2005 behavior survey, and the visualization of commuting trips for the whole region and for selected zones. Figure 15 depicts the commuting patterns in Beijing, as trip links at the TAZ level while the arrows represent the commuting direction from the home location to the job location.

The paper makes three main contributions: 1) investigate urban dynamics based on readily available LBS data using rules generated from conventional travel behavior surveys and GIS layers, 2) the processing of smart card data that tracked individual cardholder trips to analyze commuting patterns in Beijing, a decision tree framework for identifying job–housing location of smart card cardholders, and the retrieval of explicit spatial commuting patterns for Beijing based on more accurate information than conventional questionnaires or travel behavior surveys. Among the limitations of this research is the fact that the study was limited to the bus mode. The validation of smart card data could be strengthened by analyzing socioeconomic attributes of bus travelers in the travel surveys and surveying local bus passengers with smart cards. In addition, the spatiotemporal smart card information generated by fixed-fare bus routes is incomplete as the anonymity of the smart card data prevents the inclusion of socio-demographic information, making it hard to conduct behavioral study at the cardholder level. The authors mentioned that all these limitations may be addressed in future studies.

As per Long and Thill (2015), this study demonstrates the effectiveness of using smart card data for the analysis of job–housing relationships, including the evaluation of spatiotemporal dynamics of bus commuting system, identifying job–housing locations and commuting trips, and analyzing spatial and temporal commuting patterns. It also demonstrates the feasibility of analysis of urban environments using smart card data as alternative to conventional travel behavior surveys. In addition, the paper tests novel methods for obtaining useful information from large geo-tag datasets using rules retrieved from conventional questionnaires or surveys.

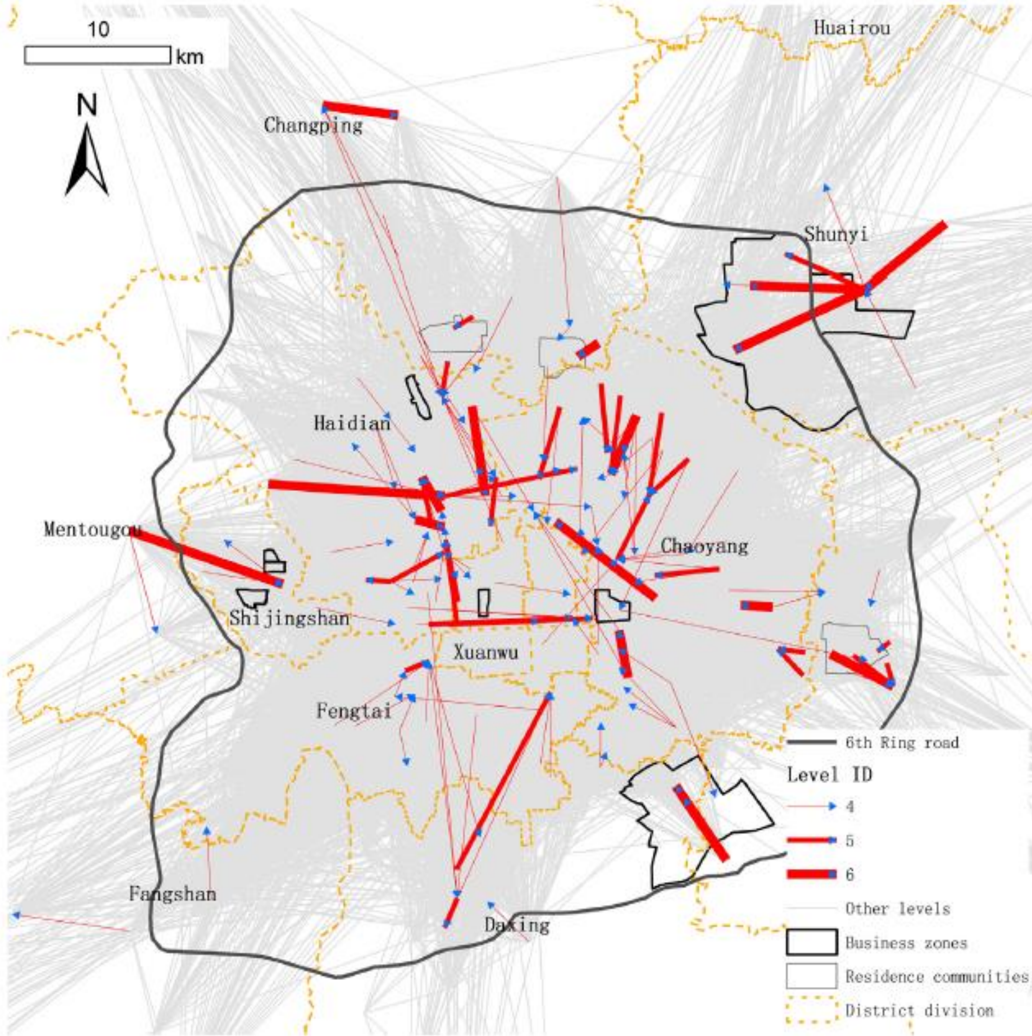


Figure 15. Visualization of Trip Links

Commuting Patterns

Moving from one place to another is an activity many people perform every day for diverse reasons. Nowadays, commuting can be influenced by the urban traffic conditions and for that reason, transportation planners and operators have the responsibility of enhancing the transit system to offer better mobility options. Prioritizing the public transit system is considered an effective strategy because it can significantly reduce car dependency, mitigate traffic congestion, and alleviate air pollution (Ma et al., 2016).

Studying the commuting patterns could help to understand the relationships between transit commuters' residences and workplaces, and it could provide valuable insights to design an efficient transportation network. This study focused on using the smart cards system to collect passenger information to identify transit commuters from noncommuters based on their travel patterns (Ma et al., 2016). As defined in the *Understanding commuting patterns using transit smart card data* paper, a commuter is a regular transit rider who performs periodically recurring trips between home and other nonresidential locations.

Beijing's transportation system was analyzed as case study. Since January 2015, all buses and subway lines have adopted distance-based fare strategies in which both passenger tap-in and tap-out data (e.g., route ID, transaction times, and boarding and alighting stops) for an individual transit rider are recorded (Ma et al., 2016). For the purpose of this study, the first and last trip of a passenger are assumed the most important ones for his/her commuting behavior.

For the home-to-work trip, the departure times is defined as T_h and for the work to home trip, it is defined as T_w. The number of occurrences of T_h and T_w are represented as N_{Th} and N_{Tw}, respectively. The similarity of departure time N_{time} can be calculated using the equation below (Ma et al., 2016):

$$N_{time} = N_{T_h} + N_{T_w}$$

The value of N_{time} represents the regularity of departure times for each transit commuter. A series of stop origin IDs home-to-work and vice versa were recorded. The number of occurrences of each stop were counted and the stop ID with larger number of concurrences was considered the most frequent stop or residence S_h. The same procedure was used for the stop IDs work-to-home and it was defined as S_w. The number of occurrences of S_h and S_w are represented by N_{Sh} and N_{Sw}, respectively. The similarity of stop N_{stop} can be calculated using the equation below (2) (Ma et al., 2016):

$$N_{stop} = N_{S_h} + N_{S_w}$$

Other routes were also identified from work-to-home and vice versa defined as $N_{Rh'}$ and $N_{Rw'}$ respectively. The similarity of route N_{route} can be calculated using the equation below (Ma et al., 2016):

$$N_{route} = N_{R_h} + N_{R_w} + N'_{R_h} + N'_{R_w}$$

Where N_{route} not only represents the spatial regularity of these frequently visited routes, but it also includes the alternatives routes from home-to-work and work-to-home for each passenger, thus improving the accuracy of transit commuter identification. The TOPSIS method measure the distance among targets of each transit rider, if the value is closed to one, the client is considered to be a commuter (Ma et al., 2016).

Figures 16 and 17 show the commuter and noncommuters patterns in Beijing. It can be concluded that most commuters take public transit around the peak hour times either in the morning or afternoon while noncommuters take the bus at any time during the day. The study shows that commuters travel more days than noncommuters with an average distance of 10.99 km while noncommuters travel less than 10 days per month with an average distance below 5 km (Ma et al., 2016).

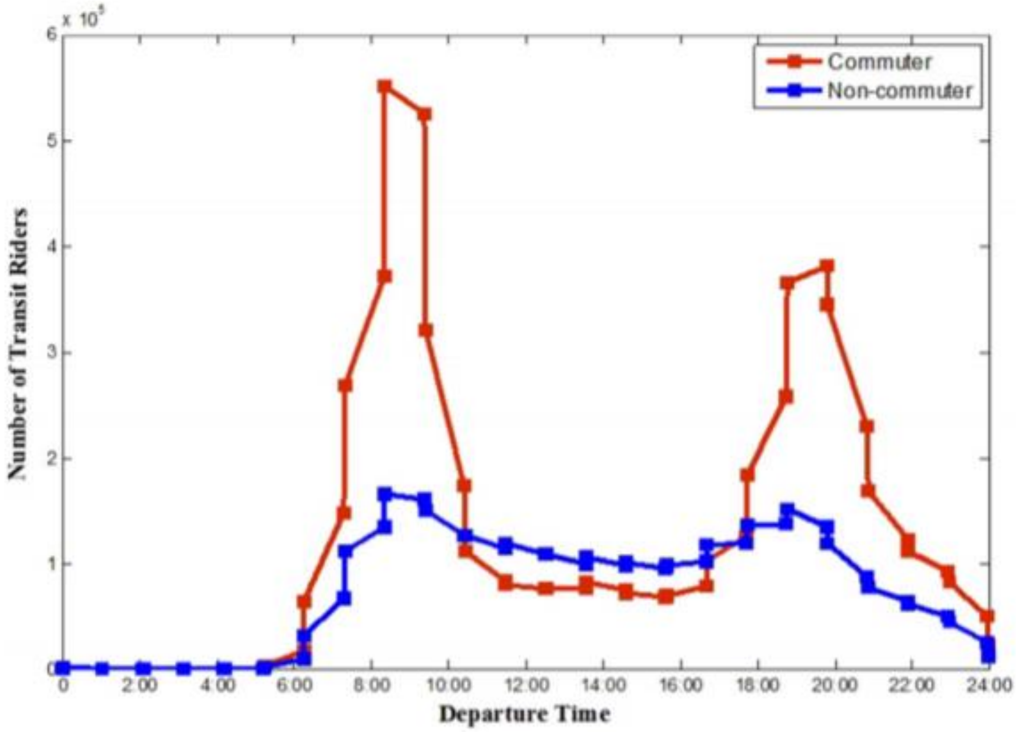


Figure 16. Numbers of Riders vs. Departure Time

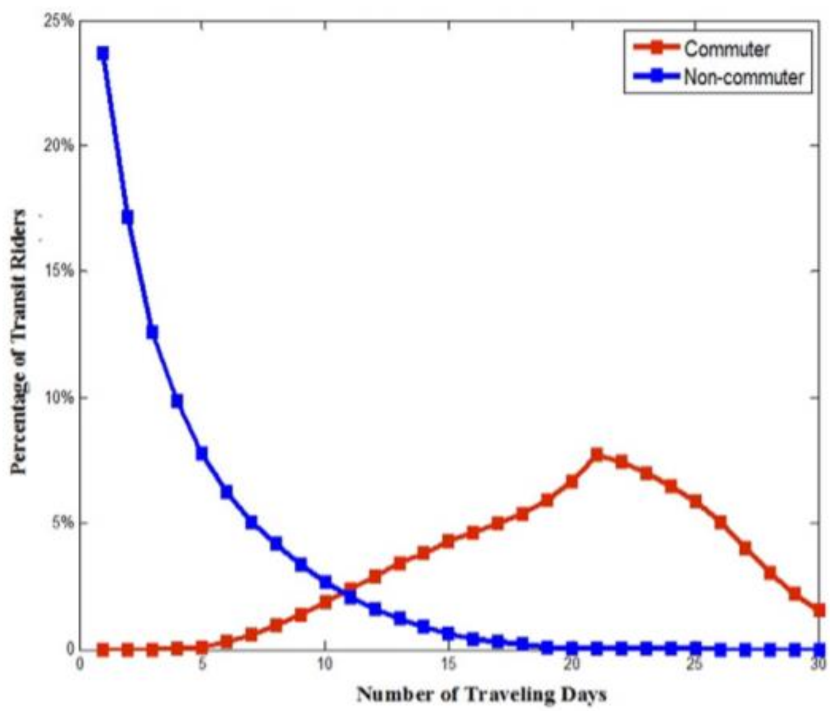


Figure 17. Percentage of Transit Riders vs. Traveling Days

This paper provides useful information for policymakers to improve the existing transportation system into a more effective one to cover the needs of the transit riders. The insights of this study could to not only design a better transportation system but also to reduce car dependency and traffic congestion.

2.4.2.2 Guangzhou, China

Extracting Boarding Information

From the data collected through the Automated Fare Collection (AFC) systems, useful information can be obtained and used for many applications with the objective of improving the transportation system. Smart card data is replacing the method of collecting information through surveys where the operational costs were high, and the information was not accurate. Many functions can be developed with the data from AFC such as the estimation of Origin-Destination (OD) matrices, which are valuable for any transportation study (Chen and Fan, 2018).

As stated in the *Extracting bus transit boarding stop information using smart card transaction data*, the purpose of this research is to develop a systematic approach to illustrate how the boarding information can be mined from AFC system without GPS and Automated Vehicle Location (AVL) system support.

It is important to know that the OD estimation is based on the boarding and alighting information per passenger and its accuracy is determined by knowing the location of the vehicles. Creating precise OD matrices can help decision makers plan, design, operate and manage an efficient public transportation system. However, this study will contribute to develop a method that can be applied to most smart card systems generating the same results without the help of GPS or AVL systems (Chen and Fan, 2018).

To conduct this study, the transit smart card database from Guangzhou, China was used for OD estimation and trip purposes. Data on 4 bus routes in 5 days were used for analyses which included 100,000 transactions. A transaction will collect the following information (Chen and Fan, 2018):

- Route: This field indicates which route the transit record belongs to.
- Card ID: The card ID is a 16-digit number which uniquely identifies a smart card. This field is critical as it allows selection of transit trips made by a passenger.
- Card type: The different categories of the card can help detect transit users' activity characteristics in an effective manner.
- Bus ID: This field indicates on which bus the transaction occurred, and kelp to identify the direction of the transit.
- Transaction time: The Transaction time field contains accurate timestamp information at the second level. This field is necessary to sort the transit trips (transactions) of a passenger in a sequential order.

A series of procedures need to be completed to clean the data. It is also important to take into consideration that consecutive swipes of the same card could generate some errors. For this reason, records with more than three swipes are deleted from the datasets (Chen and Fan, 2018).

Concerning route information, it can be calculated based on the operating times and length between stations, the number of intersections, and the total distance of the route. For the direction attribute, many studies utilized a 30 minute time gap as the time threshold for direction identification. But there are scenarios where the gap is different than 30 minutes, and for those cases it is necessary to labeling the direction using the following criteria (Chen and Fan, 2018):

- The direction will be changed when the time gap between the current transaction and last transaction is more than 30 minutes.
- The time difference between current record and first record of last transaction sequence is close to the operation time on schedule.
- The records with the transaction time before 6:00 a.m. on any day will be classified as belonging to the trip series of the previous day.

Based on the transaction records, it will be convenient and more accurate to categorize several transaction records together as boarding clusters to record all the different situations that could happen at each bus stop. Once the boarding clusters are identified, the boarding stop information can be extracted based on the difference in timestamps between adjacent boarding clusters assuming the first boarding cluster as the first stop of the route (Chen and Fan, 2018).

As a result of this study, there were a total of 100,000 smart card transactions with 98,632 of them being error free. Passengers average use was also calculated to have a better understanding of the passenger's behavior while using the route, and it was possible to estimate the frequency of boarding activities at each stop (Chen and Fan, 2018).

Developing a method to estimate ODs using only the information of boarding recorded by AFC system was a challenge because without the help of GPS and AVL, determining the vehicle's locations could have been difficult. However, the algorithm developed was tested, and the errors were minimized by following the processes in order of the identification of boarding direction, boarding clusters, and boarding stops (Chen and Fan, 2018).

2.4.2.3 London, England

Transport Information Services

The use of smart cards in transportation could help urban planners, designers, and policy makers to develop better and more personalized routes. As per *Individuals among Commuters: Building Personalized Transport Information Services from Fare Collection Systems,* Lathia et al. (2012) explain how individuals that use public transportation can show differences between individual travelers and provide more accurate information relating to their travels. A personalized public transit could address more efficiently people's preferences and needs to increase public transportation usage and reduce the pollution to the environment. The Transport for London (TfL) public transport system was used as a case study to analyze a fare collection dataset of travelers. Three different perspectives were taken into consideration (Lathia et al., 2012):

- The bird's-eye, system-level view
- Comparing individuals to each other
- Comparing individuals' transit times to the published travel times from two selected stations.

Lathia et al. (2012) focuses on using the information from the AFC databases to implement travel patterns with improved estimated times. The method used to predict client's trips collects information in a transparent way, allowing travelers to know how their data are being used. AFC databases have been studied to assess service quality and demand. However, the information collected through the fare system could be used for a variety of other improvements, for example:

- Blending it with the network structure (e.g., quantifying the number of route choices between an origin and destination)
- Examining cross-user similarity to predict ''cold-start'' trips (i.e., trips that a user has never taken in the past)
- Investigating inter-trip similarity, for example, estimate a trip from A to C based on trip segments (A, B), (B, C).

The AFC system used is called Oyster smart cards which besides eliminating the paper-based tickets, this system records the sequences of individual travelers' trips allowing the transport operator to build a fine-grained record of all passenger's movements within their network. Each record details the day, anonymized user id, the origin and destination stations, entry time and exit time (measured as accurately as the minute of entry/exit), as well as departure and arrival zone. It was noted the collected data did not show the actual route the client took, only the entry and exit point which, given the complexity of London Underground network, could be more than one possible route between the origin and destination points. Before analyzing the data, some values were eliminated to avoid inconsistencies such as: entries with the end time earlier than the start time, and equal origin and destination with duplicated station ids. Three main hypotheses were analyzed (Lathia et al., 2012):

(i) there exist fundamental differences between individuals' usage of public transport systems.
(ii) these differences can be discovered by analyzing AFC datasets.
(iii) these individual differences can be used to build personalized travel services.

An analysis of the collected data was completed showing the difference in behaviors of riders during weekdays and weekends in Figures 18 and 19. During weekdays, the peak hours are around 8:00 and 18:00 hours, however this pattern is not the same during the weekends. The study found that the distribution of trip times and cumulative distribution of standard deviations, meaning 86% of trips, have observations with a standard deviation of less than 10 min and around 32% have a standard deviation of less than 5 min (Lathia et al., 2012). This indicates that besides the complexity of London Underground system, the trip times are favorably steady.

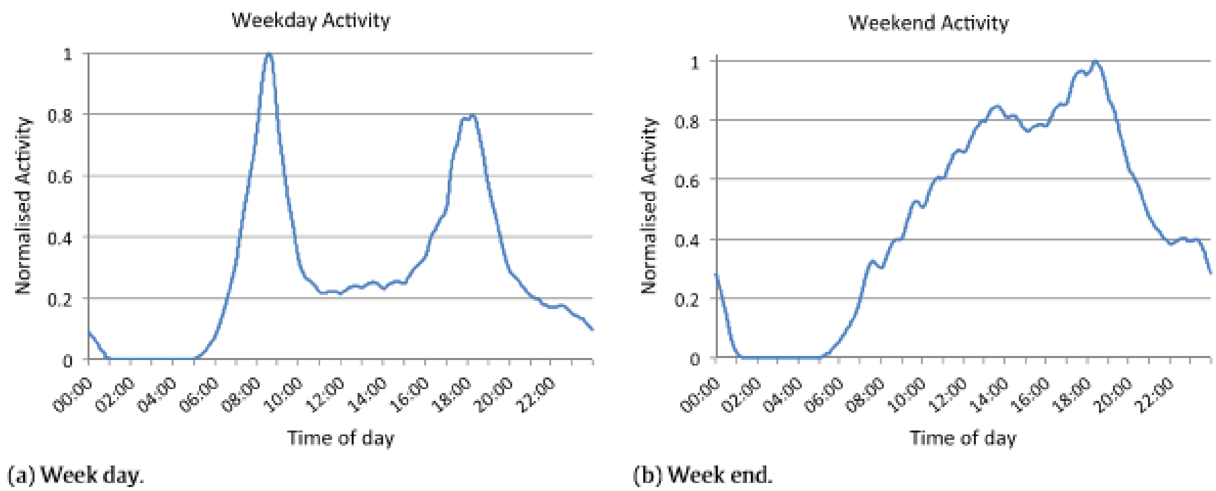


Figure 18. Travelers Activity in London Based on AFC system

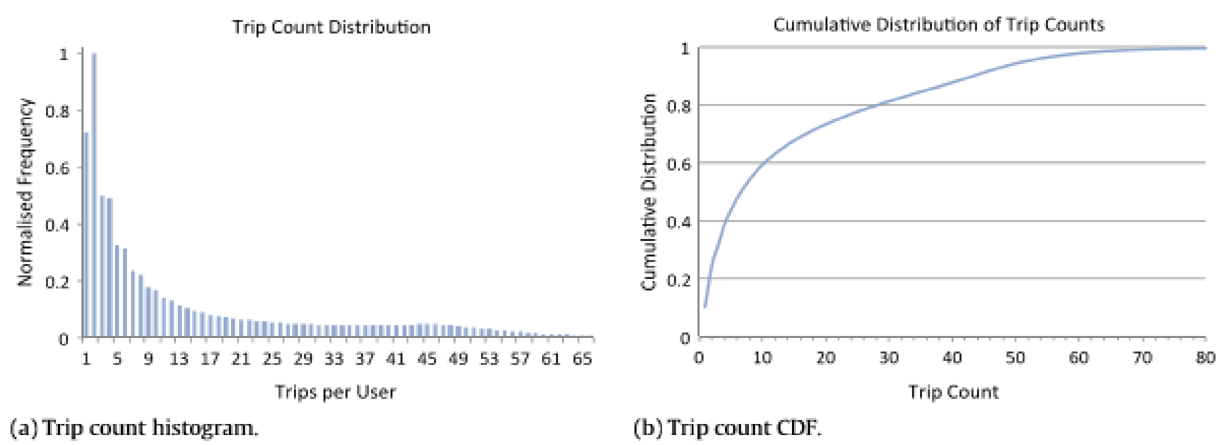


Figure 19. Trip Distribution and Cumulative Distribution Counts

An extensive analysis was completed for the collected data and drawing the following conclusions (Lathia et al., 2012):

- More than one traveler type exists.
- The crowd's travel time significantly differs from that of the individual travelers. The AFC system view fails not only to describe the travel patterns of most travelers, but also to capture the time it takes for individual travelers to complete their journeys.

- Published travel times fail to capture substantial differences in travelers' transit times. Individual travel time cannot be computed using the data collected from AFC and neither can be done by using published travel time estimates.

For one of the hypotheses, several methods are proposed in order to develop personalized travel services (Lathia et al., 2012):

- Baselines: it details the distance and travel times between the stations of the London Underground system
- Personalized approaches: it uses algorithms to provide personalized trips with better estimated times. First, the data is incredibly sparse and subject to each users' habits; thus, they do not hold a complete view of how travel time over all (*o*,*d*) pairs varies, for example, by time of day. Second, the data is subject to noisy entries, such as when users encounter delays or other events whose occurrence significantly changes their normal travel time.

It was concluded that data collected through the Oyster smart cards system is an immediate available and accurate source to develop travel times predictions algorithms providing a more personalized transit service. Besides the use of fare collection data to analyze travelers' patterns, it could also aid to offer the riders places and offers near their locations based on their trip preferences. There are many issues to be addressed using the data from the fare systems, predicting better trips times is only one of them. A series of processes for understanding, measuring, and engaging with urban residents, via this data, exists and is waiting to be harvested (Lathia et al., 2012).

2.4.2.4 Poland

Fare Collection Interoperability

Polish cities have incorporated smart cards to their transportation system offering convenience and easier transactions to either the customers or the operators. Automated fare collection system has been implemented in cities where the integration of fares and tickets across operators and modes are absent. In other words, a passenger cannot take a public transport smart card from one city and use it to ride on the commuter rail system. This document highlights the variety of possibilities of implementing an interoperable standardized system for Poland that builds on the benefits of smart media for passengers and operators alike (The World Bank, 2016).

Technology plays an important role when implementing a ticketing system, for example mobile apps offer a variety of advantages as to the use of bank cards with minimal risks of fraud and operational costs. It also facilitates the interoperability process by sharing a single ticket where the data elements, ticket specifications and smart card keys are shared between all parties and settlement through a common third party, or a Smart Ticketing Waller which is used to hold all the individual tickets needed for the journey in a manner that appears seamless to the end user.

The goal is to make the payment of transport fares simpler allowing the passengers to choose the payment method through their devices at their own convenience (The World Bank, 2016).

As stated in *Public Transport Automatic Fare Collection Interoperability: Assessing Options for Poland*, one of the negative impacts of an integrated public transport ticketing is the revenues. Since Government of Poland cannot restrict transit agencies to purchase from a single supplier, it is recommended to create standards that all suppliers must meet. Developing such standards could be approached in two different ways:

- Spend significant time and money developing them from scratch.
- Take a working system specification and convert that into a national standard with suitable accreditation and certification processes in place.

The Poland current situation regarding to tickets is the selling of single paper tickets which are non-transferable, surcharges for suburban zones are typical, and transfers have to be among 20, 30 or 60 minutes or not transfer at all. The selling of long and medium-term tickets (mostly 30 to 90 days and 24 hours respectively) contributes to the ticket sales in most cities. Whereas the selling of the different types of tickets are the predominant way of ticketing in Poland, the ticket sales over the internet are below 1 percent. However, smart cards are being introduced to the transit system by different entities by mainly selling mostly the long-term tickets. A survey was conducted that confirmed the high popularity of smart card ticketing in Poland (The World Bank, 2016).

In order to develop an integrated fare system, the following conditions should be taken into consideration (The World Bank):

- Acceptance of contactless bank cards on buses and at fare gates
- Two or more public transport operators arrange to accept each other's closed stored-value payment products.
- Acceptance of multiple-payment-enabled devices.

Implementing an integrated fare system in Poland will contribute not only to an increase of the usage of public transit but also will provide comfortability to passengers when paying for tickets at their own convenience using the online services.

2.4.2.5 Santiago, Chile

Policy and Planning

In the *Using smart card and GPS data for policy and planning: The case of Transantiago* paper, Gschwender et al. (2016) present an interesting approach for processing smart card and GPS data based on an innovative agreement between the university and the public transport authority in Santiago, Chile.

In 2007, the introduction of Transantiago, the public transport system in Santiago de Chile, brought new technologies that allowed the collection of large amounts of data. This also included the availability of obtaining vital information that the agency can use from processing this data. To assist with this effort, an agreement was signed between the University (Universidad de Chile) and the Public Transport Authority (DTPM). The agreement set forth the terms and conditions for sharing data and methods that were beneficial for both parties. From the University perspective, processing the data was a methodological challenge that allow researchers to explore potential solutions. From the Public Transport Authority perspective, this opened the possibility of obtaining detailed information at a relatively low cost for planning and monitoring.

Overall, the agreement allowed the development of methodologies and software tools that helped the transit agency obtain valuable information from the technology systems (i.e., GPS, smart card, and GIS information). The following items show a list of some of the methods developed to obtain valuable information from the available data:

- The processes required to perform more sophisticated analyses were to allocate bus GPS points to bus routes and to match the smart card transactions and positions databases, based on the bus plates or station codes and times.
- The data collection costs of expensive field measurements were reduced and allowed the public transport authority to conduct more sophisticated analysis.
- Obtained valuable information such as trips origin-destination matrices, speed profiles of buses, and service quality indicators.
- The number of passengers boarding each bus route at every bus stop could be identified for different time intervals.
- Developed a technique to estimate the alightings at a bus stop level or Metro station for each trip-stage.
- Used data and information to identify locations where particular attention to the infrastructure or the operation of the system is required.
- The information is being used as the key input to assess the impact on the creation, removal or modification of specific bus routes or the prioritization of new infrastructure like transfer stations, bus stops, and to define their size and characteristics.
- This is even being used to implement focused information campaigns to determine where and how much printed information to distribute to users.
- The use of information from large amounts of data is changing the way planning is conducted. The data provides a level of detail that allows the examination of the data in both time and space and permits analysis of not only average values, but it can also capture the data variability, which is a key aspect in assessing quality of service in public transport.
- Improved the long-term monitoring of key quality variables of the public transport system.

As an example of the information developed by this research, Figure 20 depicts the travel time in minutes from each zone of Santiago to the Centra Business District (CBD) during the morning peak; the red dots are Metro stations.

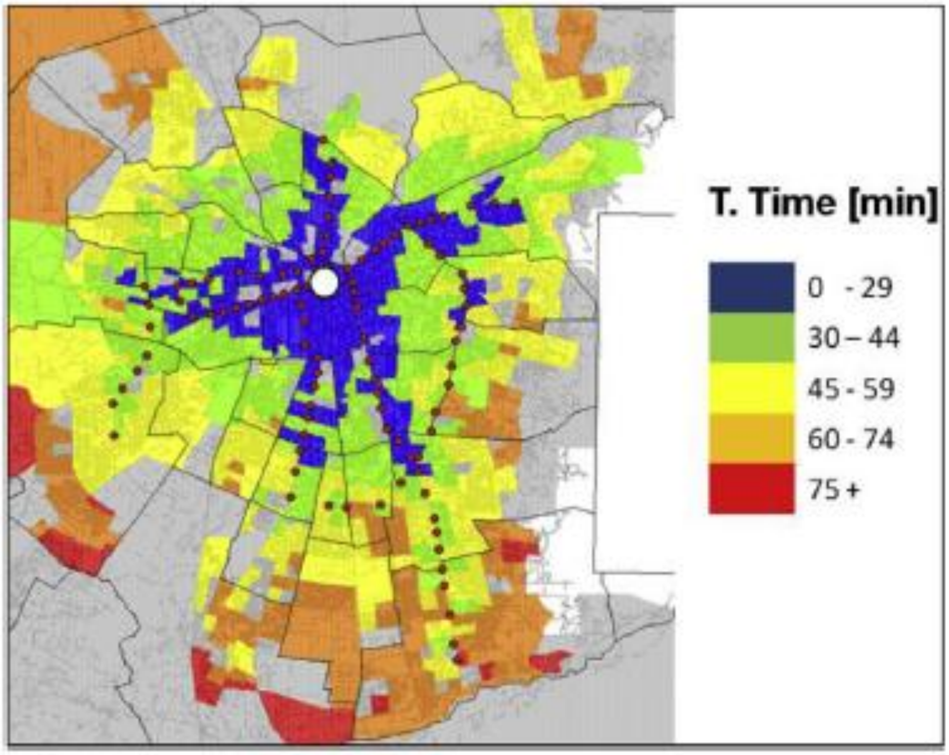


Figure 20. Morning Peak Travel Time to the Santiago CBD

The authors of the Transantiago paper (Gschwender et al., 2016) mentioned that there are several directions in which the methodologies can be enhanced to improve the information obtained from the data. This requires additional data that will have to be collected manually such as surveys or passenger counts. For instance, a methodology on fare evasion can be useful, as this is a relevant issue in Santiago especially with the bus system. Another area of further work is the development of visualization tools to help navigate the information generated from large amounts of disaggregated data analyses of large amounts of data. Without appropriate tools, the analyses can be difficult, time consuming, and requires specific knowledge.

2.4.2.6 Seoul, South Korea

Transit Assignment Modeling

As stated in *Data-driven stochastic transit assignment modeling using an automatic fare collection system* (Cheon et al., 2019), an Automatic Fare Collection (AFC) system was implemented in South Korea's Transit System Reformation project in 2004. This system, also known as smart card systems, was utilized to develop a stochastic approach using the information from every person regarding travel preferences such as trip patterns, transport mode, travel time and transfers recorded from the card readers installed in buses or subway stations. Figure 21 below shows the smart card process collection in a subway and on a bus.

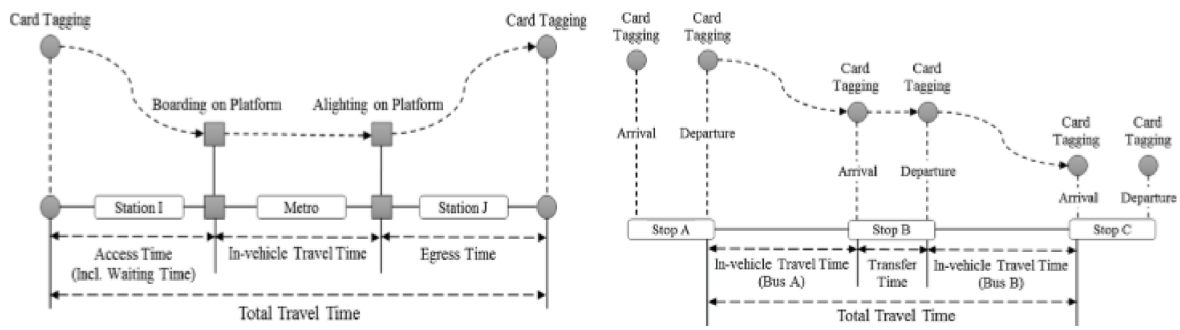


Figure 21. Data flow of the Automatic Fare Collection System

The implementation of AFC systems allows the collection of reliable data, which can be used to analyze travel behavior due to their accuracy. Another advantage is the diverse information that can be collected as shown in Table 7. The only disadvantage it presents is that it does not record the travel behaviors including various transfers movement and complicated trips across modes (Cheon et al., 2019).

Table 7. Data Field Collected from Seoul's AFC System

	Field	Description
1	D/**/Encrypted ID	Encrypted personal information
2	Card number for the day	Card number assigned for the first trip of the day
3	Card classification	Categorized by the card company
4	Boarding time	Boarding time (year/month/day/hour/minute/second)
5	Transaction ID	Distinguished ID for transfer
6	Type of mode	Bus (local/main/feeder/metropolitan/circle bus) or subway
7	Number of transfers	Number of transfers made (from 0 to 4)
8	ID of bus route	Number assigned to every bus route
9	ID of bus company	Number assigned to every bus company
10	ID of bus vehicle	Number assigned to every operated bus
11	Type of passenger	Adult, student, child, or elderly
12	Time of operation	Time at which bus travel begins
13	Boarding bus stop ID	Identification details of boarding bus stop
14	Alighting time	Card tag time of alighting
15	Alighting bus stop ID	Identification details of alighting bus stop
16	Number of passengers	Number of passengers
17	Boarding fare	Fare paid when boarding
18	Alighting fare	Fare paid when alighting

Using collected data from an AFC system can be used to construct not only single-mode transportation networks, but also incorporate multimodal transit networks since it is rare that a person takes only one mode of transportation. The suggested method to address this issue includes six steps (Cheon et al., 2019):

(1) Data extraction and processing for unnecessary information and trips.
(2) Extracting origins and destinations using station/stop-based nodes and the transferable boundary.

(3) Constructing OD pairs.
(4) Reflecting internal transfers as ascertained in the subway survey.
(5) Identifying OD pairs with more than two alternative paths for analysis.
(6) Constructing a rejected alternative set by averaging the alternative paths from the OD pairs.

This research proposes a stochastic approach for multimodal networks using real time smart card data from AFC system. The case study confirmed the reliability of the proposed data-driven approach. In sum, 34,852 bus stops and 539 subway stations, as well as 3,614,875 trips from smart card data, were used in the analysis. The final model was verified by comparing actual and assigned trips. The proposed method exhibits high accuracy (83.93%) and a high R-square value (0.981), which strengthens the superiority of the proposed stochastic transit assignment model (Cheon et al., 2019).

However, there are some limitations that need to be taking in consideration from this study such as this method is useful to analyze travel behaviors not individual travel behaviors in transit systems. Another limitation is the transfer variable should be defined because it could be influenced by the location of transfer points, transfer times, transfer service types, and transfer facilities (Cheon et al., 2019). The implementation of AFC system is increasing considerably, and the collected data should be used to implement effective transit networks to improve the current transportation options.

Analysis of Transit Service Performance

Automated Fare Collection systems are being implemented in transportation system worldwide to enhance to service offered to the people. Accessing the large amount of data collected through smart cards, not only helps to analyze the transit service but also create room for improvement. As stated in *Analysis of Public Transit Service Performance using Transit Smart Card Data in Seoul* (Eom et al., 2015), the main purpose of this study is to evaluate transit service performances by using the collected data of smart card systems considering three types of service performance: punctuality, crowdedness of vehicle, and operational speed.

The collected data through the smart cards system of the capital of South Korea, Seoul, was used as a case study. The AFC system was introduced in 2004 and it was reported than more than 90% of public transit passenger used the smart cards system for fare payment in 2010. For the purpose of this study, two transit lines were selected: one for commuter rail (Incheon metro No. 1) and another for a city bus (Bus No. 7024). The time chosen was during morning peak period from 7:00 AM to 9:00 AM with an average interval of 4 min for a metro train and 10 min for a bus. Four errors were encountered while examining the collected data (Eom et al., 2015):

1) Any one of the data fields was missing.
2) The tag for boarding and alighting is the same at the same stop or station.
3) The recorded route is not operational.
4) The recorded tag time is inappropriate.

Whether to determine the metro punctuality or not, it is important to understand the mechanism of the tag process of transit users because metro was not equipped with a card reader. Smart card data only provides the departure and arrival time for the passengers; therefore, a method needs to be developed to define when the riders take an on-time train at each station. To following logic was written based on two assumptions: 1) the metro arrives at the station either on time or is delayed (based on the schedule), and 2) passengers must use a train that arrived before the arrival tag time (Eom et al., 2015):

If $(T_i < T_i^x | T_j^x < T_j \leq T_j^{x+1})$, then use the vehicle x (on time).

If $(T_i^x \leq T_i < T_i^{x+1} | T_j^x < T_j \leq T_j^{x+1})$, then use the vehicle x (delayed).

T_i = Tag time at the station i

T_i^x = Scheduled departure time of vehicle x at the station i

T_i^{x+1} = Scheduled departure time of the next vehicle $x+1$ at the station i

T_j = Tag time at the station j

T_j^x = Scheduled arrival time of the vehicle x at the station j

T_j^{x+1} = Scheduled arrival time of the next vehicle $x+1$ at the station j

It is important to mention that the transit card data refers only to individual departure and arrival tag times; however, it does not denote the actual vehicle arrival and departure. This includes access and egress times from the gate or the platform at each station. In the calculation of vehicle punctuality, the alighting tag is done after the metro departure and before next metro arrival to avoid potential inconsistencies. It is also assumed that the in-vehicle time between stations is the same as the one in the schedule. Nevertheless, the limitation of this logic is that if a passenger behaves abnormally like running in-and-out of the gate or waiting for someone inside the gate, while accessing or egressing, the logic cannot take this into consideration in the calculations. In addition, if the metro travels faster than expected, the in-vehicle time will be changed and the logic may not evaluate the punctuality measure properly (Eom et al., 2015).

Figure 22 shows the distribution of the passengers for boarding and alighting during morning hours. It can be shown that the highest point was at station BP and then it decreased until the metro reached the last station. From this analysis it was deduced that the delay occurred if passengers take more time than expected to either board or descend from the metro and this will affect the rest of the stations (Eom et al).

Figure 22. Passengers Loads by Metro Stations

The Level of Services (LOS) of passenger occupancy (crowdedness) of Incheon metro 1 was calculated based on the passenger boarding and alighting numbers. Since there are currently no criteria for service evaluation on transfer times, transfer convenience, mobility, and equity between subway and bus, this research classified LOS in 6 levels from A to F, with LOS being the highest number of clustering. Based on the collected data, it was shown that even though high passenger load caused delays and that it could be reflected in the remaining stations, the LOS of passenger occupancy was A and improved as the morning peak hours ended (Eom et al., 2015).

To determine the bus operational speed, it is crucial to select a reasonable tag time for boarding and alighting at each stop. The tag process for buses occurs when passengers tag their card on detectors during boarding and alighting. For this reason, it is important to establish properly passengers' tags values at the bus stops. In case where improper time data was collected, two conditions were considered such as (Eom et al., 2015):

- At least two records for boarding and alighting because with only one record bus arrival and departure times cannot be specified simultaneously.
- The consecutive tag time should be within 5 seconds.

For metro occupancy analysis in buses, OD pairs were obtained from the transit smart card data to calculate passenger occupancy between stops (Eom et al., 2015).

The collected data from the Automated Fare Collection systems are allowing transit agencies and transport planners to have a better understanding of passenger's travel and behavior and such knowledge can be used to enhance the transportation services offered to the people. This study provides valuable information regarding the evaluation of service quality from passenger perspectives. Future studies need to be completed for system management considering other criteria like the operation of an express train or changes to train start positions to alleviate delays at congested stations (Eom et al., 2015).

3 CASE STUDIES

In the previous task, a comprehensive literature review of existing AFC systems was conducted. The literature review helped identify current practices and innovative methods and applications on the use of AFC data. This task enhances the Literature Review by identifying three case studies. These are locations where transit agencies have utilized AFC data to develop approaches, methodologies, technologies, applications, tools, etc. to help improve their efficiencies. This was based on the available information available through documentations, project reports, and research papers. The case studies focus on the use of AFC data, the analysis, and the methods used to retrieve, analyze, and visualize the data for transit planning applications. The selected locations of the case studies are New York, Massachusetts, and Utah with their corresponding transit agencies: Metropolitan Transportation Authority (MTA), Massachusetts Bay Transportation Authority (MBTA), and the Utah Transit Authority (UTA). Below is the general information of each of the agencies:

MTA

The MTA (Metropolitan Transportation Authority) is North America's largest transportation network, serving a population of 15.3 million people in the 5,000-square-mile area from New York City through Long Island, southeastern New York State, and Connecticut. The MTA network comprises the nation's largest bus fleet and more subway and commuter rail cars than all other U.S. transit systems combined. The MTA's operating agencies are MTA New York City Transit, MTA Bus, Long Island Rail Road, Metro-North Railroad, and MTA Bridges and Tunnels. The MTA transit network operates 24 hours per day.

The MTA subway had a daily ridership of approximately 5.5 million and an annual ridership in 2019 of roughly 1.698 billion. It includes 472 subway stations and more than 6,600 subway cars, which collectively traveled about 365 million miles in 2019. At the end of 2019, the MTA Bus and New York City Transit bus system had 327 routes: 234 local, 20 Select Bus Service, and 73 express routes. The bus fleet had a total of 5,927 vehicles, all 100% accessible to riders with disabilities. The bus daily ridership was approximately 2.2 million and an annual ridership of 678 million in 2019.

MBTA

The Massachusetts Bay Transportation Authority, often referred to as the MBTA or simply "The T", is the public operator of most bus, subway, commuter rail and ferry systems in the greater Boston, Massachusetts, area. It is the largest transit provider in New England, and the fifth largest in the country. The MBTA district is made up of 175 communities with a total population of 4.7 million. MBTA fixed route service serves the Greater Boston area. The subway is the largest part of Boston's public transit system with 4 main subway lines: the Green, Blue, Orange, and Red lines serving throughout Allston, Braintree, Brighton, Brookline, Cambridge, Dorchester, East Boston, Jamaica Plain, Malden, Mattapan, Medford, Milton, Mission Hill, Newton, Quincy, Revere, Roxbury, Somerville, and South Boston. The hours of operation vary, but most bus and train service start around 5:00 AM and ends around 1:00 AM.

In Boston, 55% of all work trips and 42% of all trips into downtown are made by transit. It roughly provided a weekday average of 1.3 million rides of all modes and 90% of those rides were on bus, heavy rail, and light rail. The bus service has a weekday average ridership of 153,000 as of July 2020. The MBTA operates 171 bus routes and 4 rapid transit routes in the Greater Boston area, with connections to the subway and commuter rail. As per July 2020 data, the subway has a ridership of 140,000 trips each weekday. It has 128 stops making easier the connections to or from other subway lines, MBTA buses, Commuter Rail, Amtrak, and regional bus services.

UTA

Utah Transit Authority (UTA) provides an integrated system of innovative, accessible, and efficient public transportation services that contribute to increased access to opportunities and a healthy environment for all people of the Wasatch region. It operates bus, bus rapid transit, light rail, commuter rail, vanpool, streetcar, and Paratransit services in Box Elder, Davis, Salt Lake, Tooele, Utah, and Weber Counties. In 2018, UTA ridership reached more than 44 million trips with a daily ridership of 152,826 trips. 100% of UTA's fixed route bus and TRAX light rail service is wheelchair accessible with lift-equipped or low floor buses and trains.

Riders can choose from more than 120 bus routes across UTA’s 1,400-mile service area. The UTA’s bus fleet had a total of 520 buses in 2018. They also provide a paratransit service counting with 80 paratransit vehicles for qualifying riders. As of 2018, the bus service had a ridership of more than 74,000 trips. Bus schedule varies from 4:00 AM to 1:00 AM seven days a week. TRAX, UTA’s light rail system has a total of 114 vehicles with 42.5 miles of line and 50 stations offering convenient connections to community destinations like shopping centers, schools and universities, FrontRunner stations, bus hubs, and Park & Ride lots throughout the Salt Lake valley. TRAX runs seven days per week, with 15-minute headways during peak times. In 2018, the FrontRunner commuter service had an average weekday boardings of 18,431 and the TRAX light rail 57,103.

3.1 New York MTA

3.1.1 Data and AFC Systems

3.1.1.1 The Data Transit Riders Want

In 2018, the TransitCenter published an interesting report about the data transit riders want. They mentioned that producing high quality data must be a priority for transit agencies. So, they can offer better information that could be used for either the agencies or the riders in order to make better decisions related to the transportation system. It allows to transit planners and operators to develop a functional system that can benefit both the agencies and the customers. However, data programs at most public agencies lag behind the industry best practices and are inconsistent with existing data specifications. With many agencies releasing low-quality data, it makes travel more difficult for transit riders. Therefore, there is a need of champions at the senior level as well as better allocation of resources to keep up with the rapid technological change of the rest of the transportation industry. The imminent change is inevitable, which brings new opportunities for alignment between the public and private sectors with the goal of finding solutions to provide a better transportation system that covers the needs of the today's riders (TransitCenter, 2018).

The TransitCenter and RMI offer some data recommendations to public transportation agencies and transit application developers (TransitCenter, 2018), as follows:

I. Data management and policy: Producing high quality, publicly available data must be a priority for transit agencies that seek to improve their service for riders.
II. Data quality: Comprehensive and widely available data is only valuable if it is accurate and timely.
III. Data specifications: Agencies and application developers will need to continue working together to expand on General Transit Feed Specification (GTFS) to be able to bring riders and agencies themselves more information, such as in-station routing and temporary changes to transit service.

High quality data is essential for agencies who wants to attract and keep riders because customers are the ones who use the transit applications for various purposes, so it is important to provide the most updated information to keep them coming back. High quality GTFS data make it easy for riders to find their bus stops and to know when the bus or train should come. Inaccurate locations could result in losing customers and give a bad image of the transit system. Real-time information provides the exact location of a bus and the time it takes to get to the stop, offering the customer a valuable insight of the bus whereabouts. This particular condition has been shown to increase ridership by approximately two percent. Managing the real-time data could also benefit the customers in reducing the bus waiting time because transit agencies could build real-time dispatch tools to actively manage headways and keep frequent transit service more evenly spaced (TransitCenter, 2018).

The accuracy of the data could also improve the information provided by the transit applications to inform the passengers about real time operations decisions saving them of an undesirable experience while waiting for a bus which will be late due to any problem. Figure 23 provides a simplified version of the processes that generate GTFS schedule and real-time data, as well as a sampling of rider-facing outputs that depend on those data (TransitCenter, 2018)

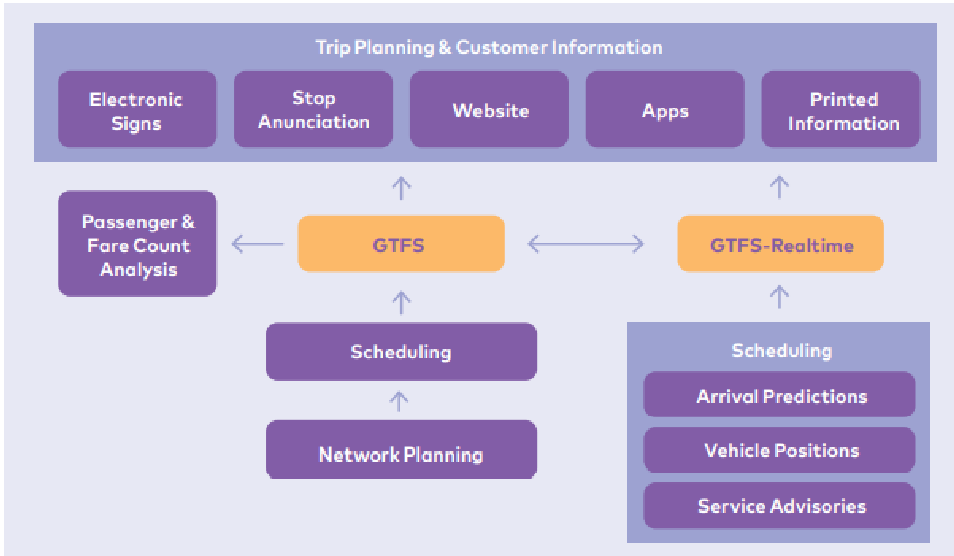


Figure 23. Transit Data Pipelines and Rides-Facing Outputs

Public agencies are the ones responsible to verify the data quality obtained from the customer's transactions. As stated in *The Data Transit Riders Want* report, there is no "state of good repair" mindset around improving data infrastructure. Safety, smooth operations, capital projects, and successful maintenance get the urgent attention, while data infrastructure is not receiving the required care.

Transit agencies are not taking advantage of the collected data for developing good data transit practices. Much of the end-user behavior and usage metrics for third-party applications is held by private firms, which is not necessarily a bad thing since their design application developers have an extremely good experience in that field. However, the quality of the final product mostly relies on the quality of the data provided which is collected by transit agencies (TransitCenter, 2018).

The interest in providing a good quality data is increasing in the private sector which it also reflects among transit agencies and application developers. If the data provided by transit agencies is accurate and complete, from application developers' perspective, then it will save time and money when those application developers start to editing and enhancing the collected data to provide to the riders the best possible information regardless of which app or service is ultimately delivering it. But this process could be obstructed by the lack of clear guidelines on how to improve the data (TransitCenter, 2018).

To address this situation, RMI convened a GTFS working group consisting of 19 transit data stakeholders and based on their input, they develop a set of practices (figure 24) for the agencies to follow.

- Datasets should be published at a public, permanent URL
- GTFS data is published in iterations so that a single file at a stable location always contains the latest official description of service
- Use consistent id fields for stop_id, route_id, and agency_id
- One GTFS dataset should contain current and upcoming service valid for at least the next 7 days
- Remove old services from the feed
- If a service modification will go into effect in 7 days or fewer, express this service change through a GTFS-realtime feed GTFS data should be configured to correctly report the file modification date

Source: Rocky Mountain Institute's GTFS Working Groups' GTFS Best Practices Document

Figure 24. Best Practices for High Quality Transit Data Feeds

Besides these practices, there are other needs that are not included on the GTFS best practices guidelines, which are important to incorporate in the applications developer's analysis to enhance the quality of the final results, such as vehicle information; fare data integration; station amenities; and temporary service changes like due to planned maintenance, inclement weather, or major events (TransitCenter, 2018).

3.1.1.1.1 Data Management and Policy

There are many opportunities that can be created when public and private sectors work together to ensure a high quality of the collected data (TransitCenter, 2018):

- Agencies can create powerful internal incentives to maintain and improve data quality by committing to use the same data feeds that the agency publishes for use by application developers and, by extension, the general public.
- Application developers, who are often at the front lines of transit rider communication and by promoting their own transit trip-planning services they are also, in effect, marketing on behalf of transit agencies.
- Agencies can advance data specifications through work with on-call contractors or through their day-to-day work. MTA New York City Transit is currently working to implement subway station navigation as part of its data offerings, which has required an experimental

extension of the GTFS specification as well as improvements to OpenTripPlanner that could be used by other agencies around the world.

- Agencies can support each other by sharing emerging challenges and new knowledge publicly.
- Agencies can benefit from making it as easy as possible for application developers to get in touch with dedicated agency staff.
- Private application developers can be allies to agency staff in building the case for better data, by providing statistics or other analysis that supports the need for improved data quality.
- Application developers can provide a valuable perspective of the best practices and emerging transit needs since they work with a diverse list of public agencies.

3.1.1.1.2 Data Quality

Data quality is crucial in developing a functional tool that best serve the customer's needs. Providing the riders with outdated information could lead to experience a significant decrease in ridership. Specific opportunities to improve transit data quality include (TransitCenter, 2018):

1. Agencies should upgrade hardware to improve real-time source data, e.g., AVL polling rates and GPS accuracy.
2. Agencies and application developers should develop their software to improve customer-facing real-time information, e.g., data latency and arrival time prediction reliability.
3. Application developers should share their validation tools and techniques with others in the industry in a replicable and/or easily implementable way.
4. Agencies should use validation tools that check for adherence to best practices (not just baseline specification compliance) and actively solicit application developer input to improve transit data quality and accuracy on an ongoing basis.

Improving the data quality will not only increase the accuracy of the real-time operational tools, but it will also improve the information provided to customers who count on accurate data to plan their daily transit trips.

3.1.1.1.3 Data Specifications

Improving the data quality is important to enhance the today's transit data best practices, but improvement to the underlying data specifications can expand what is possible. These improvements could include the capability to describe stop and station amenities, represent planned and unplanned service changes, integrate fare schedules and payment options, and provide schedule and/or real-time information for on-demand transit services (TransitCenter, 2018).

Describing the station amenities benefit especially older adults and disabled people. Currently, transit agencies are working on different aspects of these problems to show the real situation of the bus stations. For instance, NYC Transit offers a real-time feed describing elevator status, but it is not yet integrated with GTFS (TransitCenter. 2018).

Another important service the transit providers and application developers are working is to provide accurate information about service changes and detours. Schedules could be updated with the recent changes which are going to be better that printed announcements, the website, or social media. Application developers can assist in this effort. For instance, Transit App applies MTA New York City Transit (NYC Transit) service changes using a partially automated review of published changes, with human oversight (TransitCenter, 2018).

The GTFS fare model is limited because describes only single-trip zone and route-based fares with transfers, but not regional transit fares, pass products, and other pricing schedules such as time-of day and distance-based fares. Meaning that every fare application needs a specific programming code and the third-party trip planner software often have a fare information deficiency. This situation prevents the customer from having a smooth transaction when purchasing a transit ticket because of the lack of a standardized interface for transit trip planners and third-party applications with a ticketing and fare payment systems. There are at least two possible avenues that can be used for users and application interoperability (TransitCenter, 2018):

1. Least complicated but least integrated: Add ticketing app field to GTFS to enable deep links from trip planning applications to a mobile ticketing app.
2. More complicated with full integration: Develop an open payments API or SDK to third-party mobile applications to sell transit tickets without requiring users to leave the app.

It is helpful for the riders to provide the transit vehicle information. For example, showing a specific color of the bus give an idea to the riders if the bus is full or not. Also, it is nice to show some of the features the bus might have like onboard WiFi, bicycle-loading amenities (and bicycle occupancy), electrical outlets, or air conditioning (TransitCenter, 2018).

Transit agencies are looking for new ways to improve their services so they can offer a high-quality trip to the riders. Since a passenger plans a trip using the public transportation, the responsibility relies on the agencies to make it as the best experience as possible to maintain the existing riders and attract new ones. Access to timely and accurate data before and during a trip has been shown to increase ridership and improve the satisfaction of existing riders, that is way it is becoming essential on improving the transportation services. Transit data can also enhance the efficiency of the system by having monitored the service allowing the internal management to rapidly provide the service where it is needed. There are many advantages in using the data collected by public transit, but to do this transit agencies must internally prioritize data as an essential infrastructure and to continue improving the data quality. Transit agencies and developers must seek opportunities to work together for the benefit transit riders (TransitCenter, 2018).

3.1.1.2 The Age of Big Data

The Permanent Citizens Advisory Committee to the MTA (PCAC) oversaw writing a report to highlight the Metropolitan Transportation Authority (MTA) performance by using visualization techniques to provide a better understanding of the MTA's performance. The research projected 30 years of the capital investment in the MTA system and how it impacted the operational performance (Shannon & Bellisio, 2013).

The MTA produces over a thousand pages of text and statistical information, which could take a long time to interpret. For this reason, a data visualization tool could speed up the process to make the information clearer and more accessible to the stakeholders. The PCAC not only looked at the ridership impact, but also developed three indicators to capture the importance of the daily ridership impact. The indicators are (Shannon & Bellisio, 2013):

a) Commute Speeds (track investments and expansion): How improved track conditions and an increase in express train service have changed commute speeds.
b) Service Frequency (rolling stock expansion): How an increase in service frequency allows less waiting time between trains.
c) Major Delay Frequency and Recovery Time (signal and communication modernization): How the recovery speed from a major delay has changed in thirty years.

Even though the required data were available for the research, the PCAC wanted to understand how the collected information is managed to create those visualizations. They learned that the creation of those visualizations was often impeded due to the lack of good data architecture, which could lead to various problems. Because of the increasing ridership demand, the MTA should have the necessary resources to perform data analysis and create visualizations to improve communications with stakeholders and aid comprehension of the transit system's condition, emphasizing the meaning of its capital investment (Shannon & Bellisio, 2013).

Tools that help demonstrate the value of MTA performance data include the following (Shannon & Bellisio, 2013):

- Asset Management: The implementation of an asset management program will improve the interdepartmental and interagency communication, will aid in change management as a business philosophy, facilitate a whole-life cost analysis of assets, and increase transparency.
- Business intelligence and Analytics: Business intelligence (BI) is the term applied to the ability of an organization to collect, maintain, and organize data. The BI technologies can provide historical, current, and predictive information on business operations by transforming raw data into meaningful and useful information, which can be used to inform more effective strategic, tactical, and operational insights and decision making.

- Data Visualization: Visualizations could provide a better understanding of the collected data by using pictures, images or animations depending on the best way to communicate with an intended audience.
- Transparency: An agency must share information to the public about the decision-making process to create a transparency system. Transparency is not simply opening all data, but rather it is the delivery of information along with the tools required to process that information.

Improving the data analysis process will also result in a better way to visualize the information. Some of the developed visualization application could be for internal or external use through desktop intranet software or mobile mediums (Shannon & Bellisio, 2013):

- Desktop Software

  Many applications can transform an Excel spreadsheet into an interactive dashboard to aid the analysis of data, which can then be shared electronically. Adding this software to the MTA system, the interdepartmental communications could be enhanced, and hundreds of information pages could be more controllable. Figure 25 depicts the results of an interactive visualization system that allows the use of Excel data to present business analytics in a dashboard-like format.

- Mobile Applications

  A mobile software was developed called Roambi Analytics which is an online mobile publisher that allows a user to create interactive visualizations for an iPad or iPhone from an Excel sheet.

- Interactive Reports

  New interactive publishing platforms allow users to create interactive mobile reports. These reports combine a written narrative with the interactivity of a mobile application.

Figure 25. Tableau Desktop Visualization Program

3.1.1.3 MTA Stakeholder Benefits

Internal Stakeholders include managers and staff of all the MTA agencies as well as board members of all the MTA agencies and contractors and vendors hired by the MTA. Particularly, asset management could benefit the manager job significantly. It could provide access to recent and past trends of operational, station, or maintenance data whether the manager has access or not. It could help with the decision making when related to the maintenance of stations. Asset management can also help the communication among departments become more successful in achieving their goals. And finally, it could also forecast economic development scenarios before and after a station rehabilitation or service investments to the Real Estate department, as well as visualize the city's reduced carbon footprint due to the MTA system and the value of its carbon credits energy initiative (Shannon & Bellisio, 2013).

External stakeholders include everyone the MTA benefits including federal, state, and city agencies, elected officials, transit advocates, the media, and academic researchers. One reason for the adoption of a consistent data management plan throughout the MTA is the need to tell many elements of the MTA story as shown in Figure 26 (Shannon & Bellisio, 2013).

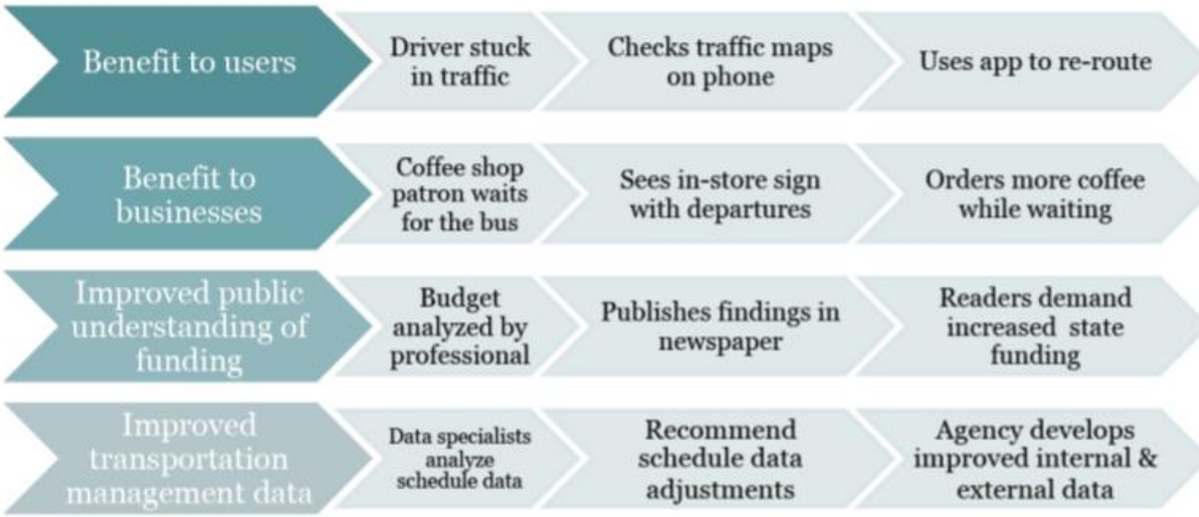


Figure 26. Typical Transportation Data Benefits

As stated in *The MTA in the Age of Big Data* report, starting with ISTEA in 1991, all recent Federal Transportation Bills have mandated public participation in the making decision process. Such participation has been useful with the help of interactive data visualizations where people can have a better knowledge of the changes effect (Shannon & Bellisio, 2013).

An analysis of 30 years of trend lines information could be ideal to have a better comprehension about how wisely capital dollars have been spent and what past investments have accomplished. With this concern, a deeply evaluation was conducted to understand what was obstructing the accessibility within the agencies. Some of the findings were (Shannon & Bellisio, 2013):

- IT Management Actions (2011-2012)
  - A Snapshot of the Complexity of the MTA IT Departments
  - The Open Data Task Force examine what data needs to be publicly available and identify the steps required to do so.
  - The MTA's CIO Council that will be addressing the need to create common platforms, standard software programs, and improve the long-term planning efforts.
  - New York City Transit Asset Management Program

- Corporate Culture: Obstacles to Open, Visual, Interactive Performance Data
  - Current Corporate Culture: although there are staff eager with the installation of new technologies, there are others who are resistance of new changes.
  - Data Silos vs. Relational Databases: Valuable data contained in silos are not included on the MTA website's Open Data portal, preventing outside stakeholders

from analyzing the MTA's fiscal, capital investment, and operational performance, which reduces the MTA's credibility.

    - Lack of Adequate Staffing to Fill Data Requests
    - Lack of Internal Data Visualization Capabilities

- IT Budget and Funding
    - Budget Reductions
    - MTA Funding

- Looking to the Future: Asset Management Planning
    - New York City Transit: Senior management at NYCT has taken the lead in creating a plan to incorporate asset management practices at its Agency. The plan includes a focus on corporate culture, optimizing business practices, and enhancing information technologies. It also recognizes the importance of educating members throughout the agencies of the benefits of an asset management program to ensure its success.

Significant changes were identified that could interfere with the improvement of data accessibility and increase the appreciation of the value of MTA data and the stories it can tell. One of the biggest issues encountered was that the MTA's information knowledge base, comprised of the Authority's business practices, corporate culture, information technology capabilities, and institutional knowledge, is underfunded, understaffed, and under supported. This puts the MTA in a disadvantageous position to implement new technologies (Shannon & Bellisio, 2013).

However, New York City Transit leadership was already working in a management plan to address many of the identified issues and PCAC offered their support for the required efforts of changing existing practices, systems, and culture. Besides the funding limitation of the MTA, there is of vital importance to invest on the information technology software to become more efficient and have more updated business methods (Shannon & Bellisio, 2013).

The recommendations made by the PCAC to the MTA system are (Shannon & Bellisio, 2013):

- Realize the Strategic Value of the MTA's Data
- Improve Performance Reporting
- Increase Open Data Efforts
- Demonstrate the benefits of MTA investments through data visualization

Three projects were recommended for showing the benefits showing information through the visualization tool. The first one is an interactive mobile and desktop application that displays trend lines and financial data with performance indicators by month and year. The second one is an interactive mobile MTA Board reports with that can access live web links including 10 years of trend lines and generating interactive charts. The third one is an app that combines the electronic board books with existing MTA data that act as a portal where user can have access to information and performance analysis.

However, MTA have several concerns about the new changes which were brought to the attention of the PCAC. Some of the questions are listed below (Shannon & Bellisio, 2013):

1. Has anyone done a cost benefit analysis for releasing data?
   In 2011 the Transportation Research Board conducted a study to learn about the use and deployment of real-time transit information founding that the costs of providing real-time information were not well understood and have not been thoroughly studied.

2. Who is in the Forefront of Data Visualization?
   The Federal Highway Administration (FHWA) has long used visualizations like heat maps and roadway models to illustrate traffic patterns and congestion. The PCAC considers of great importance the creation of an Operations Dashboard at the MTA. The one used by Tri-Met where performance of the bus and rail system can be shown allowing managers to take the proper actions.

3. Does asset management really work?
   It is proven that an implementation of an asset management program, the transportation system could experience major improvements. For example, London Underground has realized maintenance efficiencies in the areas of fleet and depots, tracks, signals and operations, and stations through the adoption of an asset management program.

4. How do you prevent misuse of data?
   One way to protect operational data is by creating a dashboard. If open data is simultaneously interpreted and displayed on the transit agency’s website, that agency will be protected against groups who want to “lie with statistics” because the true statistics are clearly presented.

5. How can this save the MTA money?
   Opening specific data and creating some visualizations will save time now and therefore money in the future. Having all the data in one place will facilitate the creation of new projects and programs and eliminate redundancies in data collection.

To better understand the different types of visualization that can be used, it is recommended that their definitions are known. Static visualization can be defined as a visualization that presents information in the form of pictures such as charts, graphs, and drawings. Dynamic visualization can be used to illustrate the change over time or steps in a process and can also be presented as animated charts, graphs, and drawings. Lastly, Interactive visualizations allows users to interact with the data, so they can choose a set of variables to generate maps or charts. Figure 27 shows the different visualization types (Shannon & Bellisio, 2013).

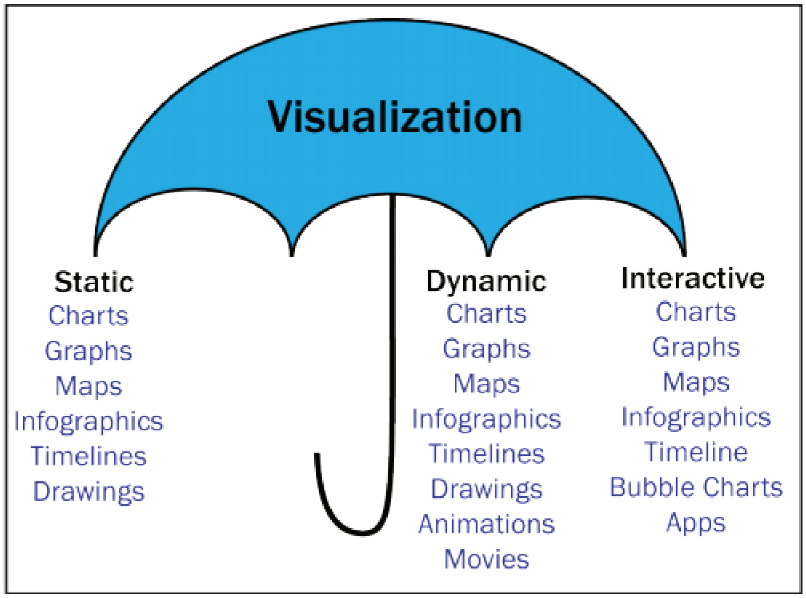


Figure 27. Visualization Definitions

They type of visualization can vary based on the needs and the data source. For instance, a basic visualization can display data in a format that it is easier to understand, rather than using tables, like the ones created by Excel. In addition to the static, dynamic, or interactive visualization, an Infographic can be another visualization type. This frequently developed by graphic designers and can provide a full picture of the information. This can also be interactive, but most of the time infographics are static. In general, interactive maps are an excellent way to display location-based information. In New York, most people are familiarized with this type of visualization that allows them to interact with the map and look for any specific information. Figures 28, 29 and 30 show examples of the different styles of visualization used by MTA (Shannon & Bellisio, 2013).

The Ten Busiest Subway Stations 2011

Station and Subway Lines	Borough	Annual Ridership
1. Times Sq-42 St N Q R S 1 2 3 7 / 42 St A C E	Manhattan	60,604,822
2. Grand Central-42 St S 4 5 6 7	Manhattan	42,795,505
3. 34 St-Herald Sq B D F M N Q R	Manhattan	37,731,386
4. 14 St-Union Sq L N Q R 4 5 6	Manhattan	34,927,178
5. 34 St-Penn Station 1 2 3	Manhattan	26,758,623
6. 34 St-Penn Station A C E	Manhattan	24,751,771
7. 59 St-Columbus Circle A B C D 1	Manhattan	21,300,892
8. Lexington Av N Q R / 59 St 4 5 6	Manhattan	20,377,141
9. 86 St 4 5 6	Manhattan	19,425,347
10. Flushing-Main St 7	Queens	18,967,751

Figure 28. Busiest Subway Stations as a List

Figure 29. Busiest Subway Stations as an Infographic

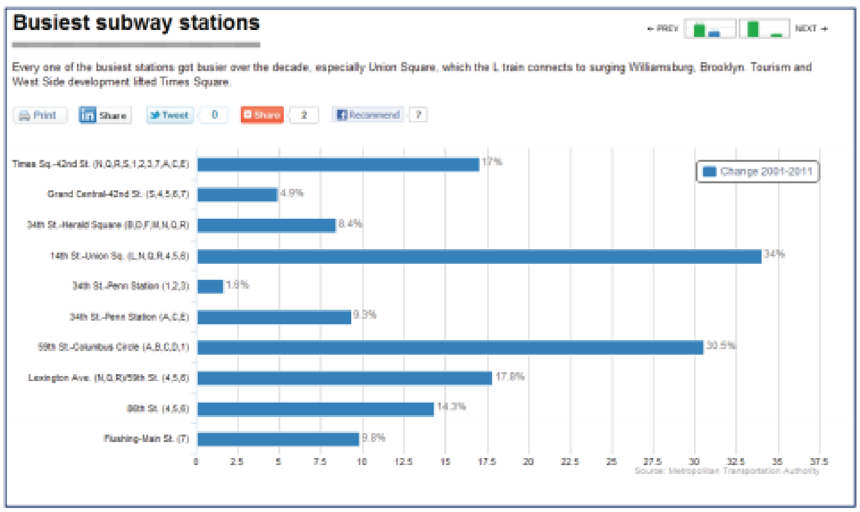


Figure 30. Busiest Subway Station as a Chart

3.1.2 Transit Applications

3.1.2.1 Using Integrated Electronic Data Service Planning

The MTA (Metropolitan Transportation Authority) provides transit service to the 5 areas of New York City. At the agency, surveyors have the responsibility of collecting data for the process of scheduling, headway determination, and route planning. However, with the implementation of AFC data with AVL data, the process could be completed at 100 % with more accurate information yielding trip level loads and passenger origin-destination (OD) data at both the trip and neighborhood levels on single-mode and inter-modal trips (Hanft et al., 2016).

With the AFC and AVL systems, the decision making would be based on quantitative evidence. One of the benefits of the AVL, besides the accuracy of their collected data, also help to improve the schedule making process because it provides a mechanism to analyze bus running times associated with stop and route modifications, temporal traffic variability, roadway construction, bus lane implementations, traffic signal timing, special events, and shuttle operations. This associated with the AFC data allowed to estimate the boarding and alighting locations of all passengers.

In the *Transforming Bus Service Planning Using Integrated Electronic Data Sources at NYC Transit* report, two case studies and the resulting service planning potential from having access to fully integrated big data sources. The focus is on two case studies that demonstrate how these data sources were used to analyze the following topics: route performance, running time, service changes, boarding and alighting locations, average passenger trip length, passenger transfers, passenger type analysis, OD patterns, and ridership impacts. The first case-study describes a neighborhood bus service analysis in the Co-op City section of the Bronx and the second details the redesign of a bus route in Manhattan.

<u>3.1.2.1.1 Co-op City Neighborhood Analysis</u>

Co-op City is a neighborhood of high-rise apartments in eastern Bronx with a population of about 35,000 people and consisting of 5 residential sections, 3 shopping centers, and additional commercial facilities throughout. It has a mix of students, workers, families and retirees, and a large number of older adults live in the residential towers but account for just under 20% of the population. Due to a budget limit in 2010, MTA had to made changes related with the transit service including re-routing several routes and discontinuing one, among others. But in 2013, as petition of community leaders, a bus service planning study was completed to evaluate the service in the area taken into consideration passengers boarding or alighting within Co-op City boundaries. The study objectives were organized into several broad categories (Hanft et al., 2016):

- Analyze whether existing service was provided as scheduled and assess the quality of service being provided.
- Determine if bus routing within Co-op City provided sufficient intra-neighborhood travel options.
- Study bus boardings and alightings to identify key stops and areas that were under or overserved.
- Analyze OD patterns of residents and visitors to determine if service was designed optimally to meet their needs and whether a high number of passengers were forced to make several transfers to reach their destinations.

It was estimated that 92% of all passenger trips started within Co-op City had as final destinations outside the Co-op City. There were tracked more than 10,000 bus–subway and subway–bus transfers coming from Co-op City during work hours highlighting the benefits of the use of electronic data and the matching of OD pairs to determine the areas where service is more needed and establish ridership service patterns for further analysis (Hanft et al., 2016).

Tracking journeys identified that only 8% of Co-op City residents' journeys began and ended within Co-op City, another 55% were bound for Bronx destinations outside of Co-op City, 35% were bound for other boroughs, and the remainder were bound for neighboring counties. This analysis was the result of OD ridership patterns where it was observed if service provided matched the monitored journeys. It was concluded that many residents were taking long trips with transfers mainly using the subway service. The new double transfer trips created in 2010 impacted fewer than 50 riders which instead of using numerous transfers, with only one the problem could have been solved or having increased the walking distance (Hanft et al., 2016).

The conclusion generated by this study and the positive reception of the Co-op City allowed the study to be extended to other areas of NYC with even larger geographic and ridership scope such as the Northeast Queens (Hanft et al., 2016).

<u>3.1.2.1.2 Route Redesign in Manhattan</u>

Along with the success of the Co-op City study, a single route in Manhattan, the M5, was analyzed at a more micro level of detail. The M5 bus route runs from the George Washington Bridge Terminal (Broadway and W 178 St) in Upper Manhattan to South Ferry in Lower Manhattan. Average weekday ridership exceeds 13,000 riders and is one of the longest bus routes in Manhattan (12 miles). The M5 operates limited-stop service on weekdays and local service overnights and weekends. The service suffered the same service cuts as the Co-op City due to budget limitations in 2010 resulting in discontinuing the M6 route and therefore the M5 expanded to the southern end of Manhattan to cover the gap (Hanft, et al., 2016).

Due to the route extension, significant delays concerns, gaps in the service, and the number of buses being short turned at various points along the route have been reported from the community and elected officials. These complaints about the M5 route were confirmed by using the AVL data coinciding that the route was one of the lowest-performing buses in the network (Hanft et al., 2016).

The service was chosen to be studied for route improvement. Some of the common adjustment made by surface line dispatchers (SLDs) are short turns, dark-to movements (beginning service a later stop than the schedule origin), or skipping stops through part of the bus trip. An electronic booking system (EBS) which manually record service changes such as short turns and other unplanned events was also installed to enhance the service. However, when analyzing the recorded data by EBS and the suggested service changes from AVL data, it was discovered that AVL provide more reliable sources of service interventions (Hanft et al., 2016).

Since the data collected by AFC and AVL were the most reliable, this information was the one used to guide the planning process of new route design for M5. The matched OD pairs showed that 75% of riders made no more than two trips a couple of neighborhoods beyond the boarding location. AFC data included the farecard purchased (e.g., pay per ride, 7-day unlimited, 30-day unlimited, etc.) along with details about discounted fares (e.g., senior, disabled) allowing the analysis to also focus on how to develop a specialized service for older adults and disabled people (Hanft et al., 2016).

Using the collected data by AFC, rider's behavior was analyzed indicating that older adult or disabled riders may prefer or require a bus trip to avoid stairs and longer access distance to underground subways. The data also indicated that approximately one-quarter of the ridership of the M5 transfer to subway which service is about every 10 blocks or closer in almost a parallel direction to the M5 which could have been easily covered by walking instead of using a bus (Hanft et al., 2016).

A route split was proposed to address the issues related with the long route of the bus. The negative side is that people will need to transfer to complete their journeys. Therefore, to provide the best solution to his problem, three constraints were defined, including northern boundary for southern split of route, southern boundary for northern split of route, and allowable size of overlap. While attempting to minimize passenger impacts, operational and cost considerations also were noted, and a proposed split point for the M5 was identified nearby layover space in midtown Manhattan within the first-pass optimal split bounds (Hanft et al., 2016).

These case studies showed that the use of electronic data could help to improve the service in an efficient way. The information about riders boarding and alighting allowed for estimations of passenger impacts on route redesigns. Replacing the manual collecting process to the electronic one, enhance the reliability of information collected. It also provides better scenarios of the ridership situation to allocate the necessary resources to manage the problem and allows for further analysis to keep upgrading the system and the quality of the transportation services.

3.1.2.2 Using Automated Data for Operations Control

The New York City Transit's (NYCT) Department of Subways implemented Automated Fare Collection (AFC) systems to improve their operational functions and service management. They focused on two separate tools which combine the data collected through AFC of real-time vehicle movements providing a better understanding of the train services to console dispatchers and superintendents. Through this data analysis, models to measure the crowded level of platforms could be forecasted to propose real-time recommendations to avoid service malfunctions (Reddy & Levine, 2016).

As stated in *Using Automated Data for Operations Control at NYCT* report, the Automated Train Supervision (ATS-A) system was rolled out at NYCT's Rail Control Center (RCC) in 2010. Even though the system shows a real-time track occupancy information and train identification, due to the lack of contextual data, it is difficult to know exactly what kind of action to take to improve the service when spacing between trains becomes uneven.

NYCT’s System Data & Research (SDR) group was working with the dispatchers to develop a tool designed to collect more contextual information to help with their decision-making process because it was found that RCC institutional knowledge did not match up with the real demographic data since major population changes have been occurred for the last decades. A pilot was developed to be put it in practice to determine how best to incorporate these tools into their workflow and to measure before and after improvements in service delivery and operational performance metrics (Reddy & Levine, 2016).

The following items provide information of the two tools developed Reddy & Levine, (2016):

- Platform Crowding Prediction Tool. The tool combines three different data sources to predict the crowding levels on a platform in terms of quantity and severity for up to 15 minutes in the future.
- Service Intervention Recommendation Engine (SIRE). This application takes as input multiple data sources including real-time train movement data from GTFS data, system layout, historical ridership, real-time crowding predictions, and ADA stations and produces a set of recommended service intervention actions that should be taken by a Console Dispatcher (CTD). The real-time inputs and the output recommendations are updated every 30 seconds.

Both tools are important for the train service in New York City. The first tool predicted high level of crowding based on upcoming gaps in service helping to the dispatchers with decisions of holding a train or skipped several stops to relieve any crowding situation. The second tool provides real-time recommendation on service control actions based on current service conditions making a positive impact on the subway service offered to the people. SIRE can recommend two types of service actions (holds and skips) that can help even spacing during regularly scheduled service (Reddy & Levine, 2016).

3.1.2.3 Using AFC Data to Estimate Transit Trips

Automatic Fare Collection (AFC) systems are being implemented in various transportations system. This new technology is helping not only to speed up the card transactions when passengers are boarding the vehicles, but also the information collected in their system is of a great value to operators and transit planners because it could help them during the decision-making process about a particular service. In addition, data can be mined to create inputs to operations planning and demand forecasting models.

Previous research studies have pointed that AFC data could be used to estimate the origin-destinations patterns. This paper focus on applying that method to all transit modes in New York City including subway, local and express buses, ferry, and tramway. The goals included improving the estimation process for subway trips, generating origin-destination patterns by traffic analysis zone, creating load profiles for bus and subway routes, and extending the analysis beyond the AM peak period (Barry et al., 2009).

The MTA's MetroCard system is an entry-only AFC system meaning the system will only record the boarding information of the passengers, there will be not record when the passenger exits a station or a bus. For this reason, this study approach is to define the trip leg by identifying the boarding and alighting locations. The research was based on a sample dataset of MetroCard transactions provided for a two-week study period in late April 2004 with almost 95 million records. It is known that NYCT never stops working and runs 24 hours a day, so 3:00 am was chosen as the start and end of a day to minimize the overlapping trips (Barry et al., 2009).

For the subway system, an algorithm was developed called the schedule-based shortest path (SSP) to make an accurate representation of the transit system by estimating the paths through the system, the location of transfers between routes and arrival times so that trips can be linked. Since the collected information by AFC contains records of every transaction, considerable efforts were required to clean up the data to provide a readable dataset that can be used to determine route patterns that can be matched with the schedules for either the subway or the bus services (Barry et al., 2009).

To estimate the boarding locations, subway and tramway used the turnstile fare collection information. This provides an identification of the station, but not the route(s) that has been boarded when there is more than one in the same station (Barry et al., 2009).

For the bus, the boarding location is estimated during the boarding transaction time which is a challenging process in New York City because their buses do not have an automated vehicle location (AVL) technology installed. As stated in *Using Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City*, the most difficult portion of this project was trying to figure out a method to clean up the bus trip tables and match them against the schedule because of inconsistent numbers corresponding to one trip, in order to be used to locate the bus boardings. After numerous failed attempts to clean up the data to be able to match the trips with the schedules, a successful strategy was finally developed based on the following observations:

1. The most important goal of matching the bus trip to the schedule is to determine the route pattern and the location of and relative times between stops.
2. Bus positions can be localized using transfers observed in the data.

To estimate the alighting locations, three assumptions were made (Barry et al., 2009):

1. Most riders start their next trip at or near the destination of their previous trip.
2. Most riders end their last trip of the day at or near the start of their first trip of the day.
3. The pattern of single-fare card users is similar to that of multiple-fare card users at a given boarding location.

A chaining procedure is applied to determine the likely alighting locations for riders that need to use two or more MetroCard transactions on a particular day. Each transaction defines a segment of the unlinked trip along the whole passenger's journey. The location of the next transaction in the MetroCard is used to determine the nearest bus stop or subway station for the alighting location of the current segment. The assumption is that many riders start their next movement near the end of their prior movement (Barry et al., 2009).

Since the chaining procedure will not determine all transactions' alighting, two expansion procedures that use sampling to assign alighting locations are applied (Barry et al., 2009):

- For subway transactions, an alighting stop is assigned by uniformly sampling based on the observed distribution from riders boarding at the same station with assigned alighting stations.
- For bus passengers, a similar approach is used based on a distribution of alighting stops for all passengers boarding at the same stop during the day for that route pattern.

Linking two or more movements to one rider into a single trip when they occur within a short period of time, help to determine the alighting time (Barry et al., 2009).

- For subway, trips are determined using the SSP algorithm, which utilizes the complete subway schedule and geographic representations of all route patterns, to predict the route traveled through the subway system and the time of arrival.
- For bus, alighting times are determined using the estimated arrival time for the bus at the alighting stop.

Various methods were developed to validate the location and linkage procedure while it was being developed (Barry et al., 2009):

- For subway transactions, the entrance and exit counts were tabulated by subway station complex for the full day and by four-hour periods, matching the periods used at each station to match the polling of the registers; these periods vary in their starting hour.
- For bus transactions, ride check data is an alternate source of information. Counts of boardings, alightings, and overall load are provided for each stop along each bus trip.

For this project, a software was created that processed more the 7 million of MetroCard fare transactions and created geo-located linked passenger trips that could then be analyzed using GIS-based query software to create reports, maps, OD matrices, load profiles, and new datasets. The powerful query software was created that allows almost any conceivable query to be answered. The software works in two steps: trip/leg selection and output creation. Queries can be made on either the linked trips table or the unlinked legs table (Barry et al., 2009).

Looking for a basic trip information like origin and destination in a New York City is a challenging task. Therefore, the Citywide Transit Travel Database with its query tool in TransCAD is very convenient to riders because the database provides a better understanding of the routes travel patterns by providing the origin and destination information including a key distinction between linked and unlinked trips. This tool is not only resourceful to passengers but for the transit planners as well since the software itself produce reports, maps, extracts and matrices as needed to support various planning and operational needs (Barry et al., 2009).

The collected data through AFC offers valued information that can be used to keep improving the transportation system. The developed method for an entry-only AFC system data is considered one of the first to actually handle this type of service, but the results could have been better with some easier algorithms as follows (Barry et al., 2009):

1. Improve the accuracy of recorded MetroCard transactions to a minute, a second or better. A bus can travel a significant distance in six minutes.
2. Preserve the orderings of the MetroCard bus boardings.
3. Improve the route system to the point where it matches the schedules exactly and has accurate geographic locations for all the stops.
4. Improve the bus trip logging system to allow an easier recovery of trip records. This could include having the drivers enter sign codes more consistently. Even better would be to have a GPS-based system to provide bus locations.

3.1.2.4 Visualizing Transportation Networks

The New York City Metropolitan Transportation Authority (MTA) could enhance their services using visualizations of the network to have a better comprehension of the system functionality. The NYC MTA was chosen as a case study because of its data availability and the size and complexity of the system which led to more possibilities for making choices on how to focus efforts. Therefore, as stated in the *Visualizing Transportation Networks: MTA Case Study* report, this study conducts and analysis of the types of standard visualizations that might contribute with the improvement of the transit system on a regular basis.

Visualizations can serve as a window to data, assisting a particular audience with focusing on the right amount of information for their needs. Therefore, three target audiences were selected to study different transportation visualizations methods (Danifo et al., 2014):

- Transportation consumers are the individuals trying to get from point A to point B in the way that will best suit their needs.
- Transportation administrators are the people who make transportation decisions, either government or private companies, who are trying to understand how transportation is used in aggregate and predict how they can impact the overall network for greater service or cost efficiency.
- Civic interest group are the citizens with some transit interests.

To avoid a contradiction for any overlaps among the defined audiences and to guarantee the best selection of visualizations, the following classification was specified to focus on the dominant use per visualization (Danifo et al., 2014):

1. Transit Use, in this case by MTA customers who are not concerned on how the system works but rather how to use it to move from one place to another at specific time.
2. Transit Administration including MTA employees who can make better decisions about where to invest time and resources if they have visualization information available.
3. Civic Interest. This is a very broad group of people who have tangential interests in transportation; they might eventually vote on issues related to public transportation or invest in private transportation companies.

Several technologies were integrated to create appropriate visualizations for the MTA data. First raw MTA system data was obtained from a variety of resources. MTA subway transit data was obtained from the MTA's developer pack providing CSV files containing operation and performance data from all MTA operated transit lines. Google Transit Feed data was obtained in the form of a database, which included stops and trip times for all MTA subway and bus lines. Additional MTA JSON files for paths, stops, and transfers in the MTA system were obtained from the Mapperly project (Danifo et al., 2014).

Several computer-based technologies were used to format, manipulate, and visualize the MTA transit data. The data was analyzed as a network using Python and NumPy. OpenRefine and Python Pandas were used to facilitate the formatting and merger of the various data sets into a standardized JSON file format. NetworkX was employed to create an object representation of the MTA network, with stops as nodes and transit lines as edges (Danifo, P. et al., 2014). To have a better interpretation of the results, the visualizations created (Figures 31 to 35) were evaluated against the defined criteria (Danifo et al., 2014):

Figure 31 shows the geographic view of the transit lines. This first visualization gives a user a quick overview of how MTA lines cover the geography of New York. This figure could be used by any of the audiences we described, however would be more useful if augmented with additional layers of information. Figure 32 shows the mean wait time by line. The second visualization builds from the first, but this time layers in detail about wait times. A transit user could use this visualization to decide which lines to avoid using. Transportation administrators could use this visualization to quickly see patterns among line wait times. This visualization could be particularly useful for monitoring any changes over time (either expected or unexpected).

Figure 31. Geographical View of Lines

Figure 33 shows the directed graph. This visualization provides information that is conducive to comparing against classified operations that a user might need to execute to analyze data.

Figure 32 Mean Wait Time by Line

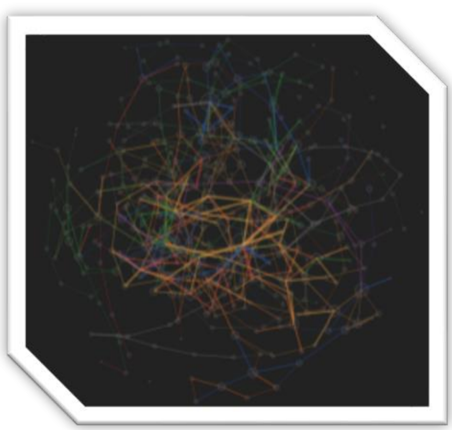


Figure 33 Directed Graph

Figures 34 and 35: MTA Bridges and Stations by Network Degree. The visualization for MTA stations and bridges by network degree is evaluated using the same criteria:

- Locate and Identify: the ample size of the bubbles, the contrasting color scheme, and the quality of the map serve to enhance the location and identification operations.
- Distinguish and Categorize: these operations do not apply to this visualization.
- Cluster and Distribution: clustering occurs naturally in this visualization due to the geographical nature and the rendering of the bubbles and distribution is based on this same geographical property.
- Rank: the size of the nodes serves to rank the stations by network degree, where the largest bubbles represent the station locations with the highest network degree.
- Compare/compare within and between relations: comparisons among the nodes are achieved by size differences.
- Associate and Correlate: associations and correlations are implied in these visualizations, existing only through geographical interpretation.

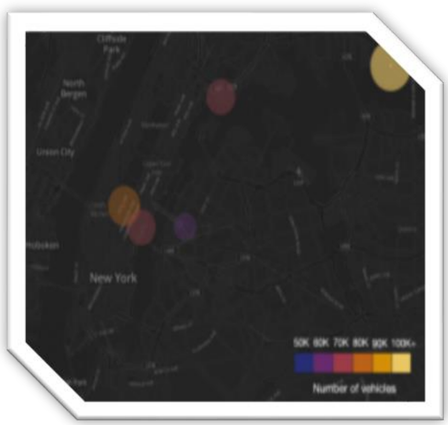


Figure 34 Bridges by Network Degree

The visualizations have multiple features like colored lines, circular nodes, nodes of varying sizes, an animated layout, meaningful edges (to signify the endpoints of subway lines), and hover-over text descriptions, making them more readable. Some of these features could be of great value for customers and transit employees. For example, the node and centrality properties highlight the most highly trafficked stations which would be of great value for customers identifying the most useful stations and for system administrators when need to direct resources and maintenance. In addition, the location feature will help riders determine which lines are more reliable in terms of waiting time (Danifo et al., 2014).

Figure 35 MTA Stations

The development of different visualizations should be taken into consideration by the MTA to enhance the transportation system in New York City because they can be helpful not only to riders when making decisions about the transit mode that better suits their needs, but also to transit administrators who can monitor on a regular basis the system functionality to prevent unnecessary adversities (Danifo et al., 2014).

3.1.2.5 Fare Evasion

The Office of the Metropolitan Transportation Authority Inspector General (OIG) reviewed the methods used by y MTA New York City Transit (NYC Transit or Transit) to evaluate the fare evasion situation on both subways and buses. Even though measures had been taken to control to reduce fare evasion, without tangible information of evasion levels and trends, will be difficult to address the problem (MTA, 2019).

The OIG review the Transit's Subway Fare Evasion (SFE) and Bus Fare Evasion (BFE) surveys from mid-2017 through the end of 2018, focusing on sampling methodologies like statistical assumptions, sample sizes, sample selection, confidence intervals, and margin of error formulas. These surveys could improve not only the management decision making, but also a simplification of the collected data by Traffic Checkers could be beneficial to determine the fare evasion levels (MTA, 2019).

As stated in *NYC Transit Fare Evasion Surveys: OIG Concerns*, Transit's Operations Planning Division (OP) conducts quarterly fare evasion surveys that produce statistical estimates of Bus Fare Evasion and Subway Fare Evasion. A methodology was created by OP to measure fare evasion by doing a random sample off are control areas stratified by 20 levels of activity (average ridership across fare control area and hour of day), two median income levels, and three parts of the week (weekday, Saturday and Sunday).

The Subway Fare Evasion (SFE) survey began around 2008, after MTA Audit Services found that fare evasion was a lot higher than estimated by Station Agents (2.1 percent vs. 0.32 percent). Therefore, OP conducted a pilot study where auditors measured fare evasion at randomly selected fare control areas (i.e., contiguous groups of turnstiles/gates within a station). After analyzing the results of the pilot study, OP created a methodology for measuring fare evasion. This was a simple random sample of fare control areas stratified by 20 levels of activity (average ridership across fare control area and hour of day), two median income levels, and three parts of the week (weekday, Saturday, and Sunday). Nevertheless, due to a number of situations, OIG had several questions and concerns to the SFE survey (MTA, 2019):

Concerns About Survey Design:

1. Some differences in margins errors using the OP simple random sampling method have brought attention and adjustments of the formula will be required to fit the requirements of a clustered sampling method bringing as consequence more review and validation of the sampling method and margin of error formula.
2. The lack of surveys completion by OP is rising questions as: What is the impact of this shortfall on the accuracy of the estimates, and thus their representativeness of the entire population of riders? Also, do the missed surveys bias the results (for example by disproportionately occurring in certain areas or times of day)?

Concerns About Data Collection Practices:

1. Certain fare control areas have multiple banks of turnstiles and/or visual obstructions that make it difficult to see all gates/turnstiles from any given vantage point.
2. The instruction to Checkers to watch a single gate in a fare control area with multiple gates would typically lead to undercounting illegal gate entries.
3. OP has not created a manual with instructions for SFE Traffic Checkers.
4. SFE Traffic Checkers are required to wear MTA uniforms. (BFE checkers are not.) Has OP analyzed the impact of the presence of uniformed personnel on evasion behavior?
5. Currently, SFE Traffic Checkers have limited availability, are shared with other surveys, and often cannot be substituted for in their absence.

The Bus Fare Evasion (BFE) survey began around 2008. This was part of an effort from OP to report ridership and daily bus passenger-miles to the Federal Transit Administration (FTA). At that time, NYC Transit was seeking FTA’s approval for a way of estimating bus passenger-miles using AFC data. Because the AFC system only captures data from paying customers, OP designed a survey to estimate the proportion of riders that do not interact with the farebox (e.g., fare evaders, children riding free with an adult, and uniformed officials using a flash-pass). Due to the results of this survey, OP was able to estimate the fare evasion rate for the whole system.

Traffic Checkers ride on the buses recording each boarding under one of 13 different categories of legal (multiple types of flash passes, badges, and passengers in uniform) or illegal (front door evaders, rear door boardings, and multiple types of incomplete payments) entries (MTA, 2019).

OP collected double the number of surveys needed to produce statistically valid estimate of fare evasion. This could be the result of the changes in the method for determining the sample size raising some question and concerns about the BFE survey (MTA, 2019):

Concerns About Survey Design:

1. The sample selection is not fully random because, for practical reasons, each sampled trip is paired with a return trip on the same route. Also, the results from one quarter, include observations from two or three different random samples. OP should assess what implications these factors have on the margin of error.
2. OP told us that it completed 321 samples in the 4th Quarter 2018 when it did not complete its full sample, is there statistical bias (day of the week, time, geography, etc.) in which surveys are omitted?
3. The margin of error calculation has the same value for all routes. This is questionable because some routes are known for having higher levels of fare evasion than others. Is the new margin of error formula now compatible with the longstanding approach to sampling and data collection, which OP has not revised?
4. In order to determine how many bus trips must be surveyed to obtain the target number of boardings, OP uses estimates of the average number of passengers per bus trip (load factor) in each area. However, OP does not calculate separate load factors for weekday/Saturday/Sunday. Has OP considered calculating load factors with AFC data covering all bus trips in a given stratum (adjusted for average fare evasion)?
5. Express bus service is not covered by any of the fare evasion surveys.

Concerns About Data Collection Practices:

1. Traffic Checker availability issues make it difficult to cover all sampled trips: OP has only six dedicated BFE Checkers, whose shifts can only include only one weekend day. OP expects around 20 percent Checker unavailability.
2. Can one Checker accurately monitor all-door evasion, especially on more crowded buses, and given the detailed categorization required?

An OP manager raised the concern about how new changes to the methods used in the surveys could alter the results and not be comparable with the results from past years. Other OP staff said that these changes could also affect the error margins calculations. However, OIG considers that doing the necessary methodology changes could considerably affect the point estimates of fare evasion, but it would still be beneficial having more accurate results. As a result of these observations, MTA New York City Transit was encouraged to conduct a series of revisions of the Subway and Bus surveys to make the necessary adjustment of the design methodologies and data collection used to evaluate fare evasion levels to provide reliable estimates and trends (MTA. 2019).

3.2 Massachusetts MBTA

3.2.1 Data and AFC Systems

3.2.1.1 MBTA Tariff

With the objective of set the terms of fare rate and payment options across the MBTA system, a Tariff was developed named Tariff and Statement of Fare and Transfer Rules. As stated in *MBTA Tariff and Statement of Fare and Transfer Rules* report (MBTA, 2018), the MBTA fares are defined by multiple variables like the mode of transportation a passenger is traveling, fare media used to pay and discount fares for eligible people. MBTA also offer a variety passes for unlimited times in a period and discount fares por eligible people based on their age, student status or disability. In addition, it provides a paratransit service called The RIDE offering a door-to-door service por eligible people that are not able to take the fixed transit. The Tariff lists multiple fare products offered to people.

<u>3.2.1.1.1 MBTA Fare Media</u>

There are several ways to pay for a ride on any MBTA mode of transportation. Some of this fare media grant fare discounts validating them and determining which privileges could be transferred (MBTA, 2018)

- CharlieCard

The CharlieCard is a durable, contactless smart card which uses a chip embedded in the card to store value information and communicate with MBTA faregates and fareboxes. Customers can add money to the card or purchase passes and load them onto it. The CharlieCard is accepted on MBTA subway and light rail, Local Bus, and Express Bus. The card provides discounts to passengers when they used CharlieTicket or cash.

- CharlieTicket

CharlieTickets are paper tickets produced by fare vending machines (FVMs) and other MBTA devices. Tickets hold fare information on a magnetic stripe which can be fed into a ticket reader. FVMs also print information on the ticket regarding the value or product stored on the ticket so in some cases a CharlieTicket can be used as a flash pass for visual inspection. When a CharlieTicket is used to store value, no single-ride discount applies.

- Paper Tickets

Paper tickets are used for single-ride ferry and boat tickets, private carrier bus service, as well as some special event Commuter Rail tickets. They do not contain a magnetic stripe and are only verified by visual inspection by conductors or operators. They do not allow transfers to any other service.

- Cash

Cash is accepted directly by fareboxes on buses and Green and Mattapan Line vehicles at ungated stops. No transfers are available when using cash. At gated stations, cash may only be used at a Fare Vending Machines (FVM) to add stored value or a pass to either a CharlieTicket or a CharlieCard.

- Commuter Checks and benefit cards.

Holders of employer or organization-provided transit benefit cards or checks may apply those funds toward the value of a purchase.

- mTicket

Mobile tickets are available for the Commuter Rail and Boat systems only. Customers can purchase mobile tickets electronically using the MBTA mTicket app. Single-ride tickets, TenRide tickets and monthly passes are available on mTicket. Paid tickets are stored on the app and must be activated by the customer before boarding the vehicle. Paid tickets are stored on the app and must be activated by the customer before boarding the vehicle. mTickets remain valid for one and a half times the length of the scheduled trip and are not currently accepted on MBTA bus or subway service.

<u>3.2.1.1.2 MBTA Fare Vending and Validation</u>

- Fare Vending Machines

Fare Vending Machines (FVMs) are located throughout the MBTA system. They issue CharlieTickets, but not CharlieCards. However, a customer with a CharlieCard can re-load it at an FVM.

- On-board Fareboxes

Fareboxes are located on-board buses and Green and Mattapan Line vehicles. Fareboxes are commonly used to deduct stored value or validate passes on a CharlieCard or CharlieTicket. Fareboxes accept cash, either as direct payment for a single ride or to add stored value to a CharlieCard.

- Platform Validators, Portable and Handheld Validators

Validators deduct stored value or validate passes on both Charlie Tickets and CharlieCards and then issue a receipt that allows the customer to board any door of a bus or trolley. Portable validators operate similarly to platform validators by deducting stored value or validating passes on both Charlie Tickets and CharlieCards, but they do not issue a receipt. Handheld Validators are operated by MBTA personnel assigned to assist with fare payment. They may either inspect the last transaction made with a CharlieCard to ensure that a customer has paid a valid fare, or they may deduct stored value or validate passes.

- Mobile

Mobile tickets are currently accepted only for Commuter Rail and Boat rides. Customers must download the MBTA mTicket app and follow the instructions. Major credit and debit cards are accepted through the mTicket app. Tickets purchased in the mobile app are subject to the same expiration rules.

3.2.1.1.3 Reduced and Free Fare Eligibility

The MBTA offers fare discounts to eligible customers like (MBTA, 2018):

- Seniors

Customers, age 65 and over, are eligible to receive a Senior CharlieCard. Senior single ride fares are approximately 50% of the full adult fare. With a Senior CharlieCard a passenger can purchase a reduced fare 10 ride ticket for Commuter Rail or Ferry.

- Persons with Disabilities

Persons with disabilities and Medicare cardholders may apply for a Transportation Access Pass (TAP). This card entitles the holder to the same discounts as the Senior CharlieCard. TAP CharlieCards are valid for a period of either 1 year or 5 years (depending on judgement of the disability stemming from the TAP Application).

- Middle and High School Students

Student CharlieCards are available for students in grades 5 – 12 only. They can hold up to $30 in stored valued to use for reduced single ride fares. The regular Student CharlieCard allows the student to pay the reduced fare rate for single rides on all modes or to purchase the student monthly pass, which is valid on Rapid Transit, Local Bus, and Express Bus. The Student CharlieCards are also valid for a period of one year.

- Youth

The Youth Pass is a reduced fare product for young people who are not eligible for the Student Pass and who are low-income and live-in participating cities and towns in the Greater Boston area. Youth are eligible if they live in a participating city and meet the MBTA's eligibility criteria. Youth Pass participants are not eligible for reduced fare Commuter Rail tickets or passes.

- Free Passengers

Free fares are given to certain types of passengers on MBTA services, not including The RIDE.

- Children 11 and under
- Blind
- Employee
- Military, Police and Fire Fighters
- Commonwealth Officials

The RIDE is the MBTA's paratransit service. It provides a door-to-door service and a shared-ride transportation to people with disabilities that impede to have access to fixed route transit. Persons with The RIDE CharlieCard may ride the MBTA system free of charge and the Commuter Rail up to Zone 5 (MBTA, 2018).

3.2.1.1.4 Third Parties

- Regional Interoperability (RTAs)

The MBTA and Massachusetts's Regional Transit Authorities (RTAs) provide fixed route and paratransit service in communities across the state. The partnership between MBTA and RTAs allows customers to use the CharlieCard on RTAs vehicles.

- Private Carrier Services

Private carrier buses have the same cost ride as MBTA Local bus routes (CharlieCard fare) but are not equipped with the MBTA's fareboxes. The service payment on these vehicles is either cash or with an MBTA pass that may be visually inspected.

- Third-Party Fare Media (Bring Your Own Chip)

The MBTA allows organizations and entrepreneurs to contract with the MBTA's approved manufacturer to embed MBTA CharlieCard-compatible chips in some ID cards and some other personal devices. This is referred to as "Bring Your Own Chip" (BYOC) in this document.

1. Contracting for encoding: BYOC Creators must contract directly with MBTA's smart card encoding provider (Encoder) for services to have Chips encoded for use in BYOC products.
2. Costs: The MBTA will neither charge nor reimburse BYOC Creators for participation in this program. MBTA's Encoder will charge BYOC Creators directly for charges.
3. Manufacturing requirements: A "MiFare Classic 1k 4-byte non-unique identifier NXP chip" (Chip) must be used. The Chip serial number and sequence number must be permanently etched or stamped onto the BYOC Product in a human-readable format. BYOC Creators must source and purchase these chips.
4. Fare validity, replacement, and expiration: This Chip is accepted by all MBTA electronic fare equipment and it will be treated as a CharlieCard. The same two methods available for the replacement of CharlieCards– at the CharlieCard Store, or by mail –apply to BYOC Products. BYOC Products without etched serial numbers are not eligible for replacement and may not be eligible to hold reduced fare products.

5. Data ownership and rights: All data contained on the Chip is the intellectual property of the MBTA except as defined in the MBTA Privacy Policy and it could be disabled by the MBTA for actual or suspected misuse.
6. Product customer service: BYOC Creators must clearly communicate customer service policies instructions and expected lifetime to customers. They are responsible for design, development, testing, and quality assurance of the BYOC products. The MBTA shall not be liable to BYOC Creators or users of BYOC products except as specified by express written agreement.

3.2.1.2 Automated Data to Improve Decision Making

Fare changes will impact transit ridership in a short and long term. A creation of a fare policy will not only help the transit agency to offer a good service, but also will help the cities to function and grow efficiently and equitably. Further, Incremental changes in fare policy are expected to have substantial long-term impacts on transit ridership and revenue. As stated in *Transit Fare Policy: Use of Automated Data to Improve Incremental Decision Making*, a procedural framework was proposed to organize analysis of incremental fare changes, linking exploration of current pricing strategies to estimation of behavioral parameters and modeling of fare change scenarios.

Two transit agencies were chosen for such study: the Massachusetts Bay Transportation Authority (MBTA) and the Chicago Transit Authority (CTA) which increased their fare prices in recent years as an answer of their budget shortfalls. Both agencies are experiencing a decline in bus ridership and a growth in competing travel alternatives like ride hailing (Stuntz, 2018).

The main focus of this thesis is related with the use of automated fare collection (AFC) data to help with the decision-making process about changes to fare structure and levels. This broad scope can be broken down into four components (Stuntz, 2018):

- Fare structure and levels: It describes how the interrelated drivers, functions, and technologies play a role in transit fare policy.
- Incremental changes: It refers to changes that modify fare levels and adjust products and policies. These changes are generally less dramatic, but they are the most common fare policy decisions made by existing transit agencies.
- Decision making: This research studies ways that transit agencies might organize and use existing AFC data to understand the potential impacts of fare change options more quickly, cheaply, and accurately on ridership and revenue. The hope is that better information about the implications of fare policy changes, both internally at transit agencies and externally for the public, will be one ingredient for more robust public debate, a higher degree of accountability, and ultimately wiser decision making.
- AFC data: This thesis focuses on the use of disaggregate AFC data, which is already available at most major transit agencies. AFC data provides cheaper and more accurate

ridership information than surveys and counts, and it is already used widely for agency reporting and analysis of aggregate ridership and revenue.

The MBTA serves 175 of the many cities and towns that make up the Boston metropolitan area. It provides five different transportation services in the Greater Boston metropolitan area: Subway, Bus, Commuter Rail, Ferry, and Paratransit. It roughly provided a weekday average of 1.3 million rides of all modes and 90% of those rides were on bus, heavy rail, and light rail. The MBTA commuter rail carries between one third and one half of all commuters traveling into Boston during the peak period. MBTA ridership has been relatively flat over the last 15 years, with a slightly decline of commuter rail ridership over that time but it has been stable for the last few years (Stuntz, 2018).

Under the MBTA's fare structure, fares are differentiated in four important ways: by service, tariff, user type, and medium. The most significant differences are by service. Different fares or prices are charged for seven different services: Local Bus, Inner Express Bus, Outer Express Bus, Rapid Transit (including bus rapid transit, light rail, and heavy rail), Commuter Rail, Ferry, and paratransit (The RIDE). Pricing for bus and rapid transit is flat, meaning that it does not vary by trip distance, zone, or origin-destination pair. Pricing for commuter rail is zone-based, increasing for longer trips according to pre-defined zones. The MBTA also offers multiple tariffs like the following (Stuntz, 2018):

- pay-per-use (also called single-ride, pay-as-you-go, or stored value), available on all services.
- multi-ride tickets, available for round trips or 10 rides on commuter rail
- monthly passes, available separately for each service and each commuter rail zone, and typically allowing unlimited travel on all less expensive services (for example any Commuter Rail pass covers travel on closer Commuter Rail zones, Rapid Transit, and Local Bus)
- 7-day and 1-day passes, available only for Rapid Transit (also covering Local Bus).

The MBTA collects fare payment through a variety of different sale channels, but mainly from the fare vending machines (FVMs) located in rail stations and major bus terminals, and the Corporate Program. It has an integrated fare collection system called the Charlie system which consists of FVMs, fare validation and fare media that interact with these devices (Stuntz, 2018).

The CTA serves Chicago and 35 surrounding suburbs. It provides bus and heavy rail transportation in the Chicago metropolitan area. The CTA operates alongside two other transit agencies – Metra providing regional commuter rail services, and Pace providing suburban bus and paratransit services. From 2002 to 2012, bus ridership was flat at around 300 million unlinked trips per year while rail ridership grew about 28%. Beginning in 2012, growth in rail ridership began to slow and bus ridership began to decline; by 2017, bus had declined about 21% from 2012 levels and rail had also begun to decline. This situation has brought questions about what factors are causing this situation (Stuntz, 2018).

The CTA offers several tariffs (or payment structures) such as pay-per-use and four different rolling period passes valid for 1, 3, 7, or 30 days. The passes are valid on CTA rail and bus; 30-day passes are also valid on Pace buses, and 7-day passes are valid on Pace for a small premium. The CTA also offers a discounted monthly pass to Metra (commuter rail) monthly pass-holders, valid on the CTA only during weekday peak periods (Stuntz, 2018).

The CTA use different sale channels for fare products and collects payments. As shown in Table 8, the major revenue is from ticket vending machines (TVMs), but also the online sales account for about a 1/3 of revenue recorded in the Ventra system (Stuntz, 2018).

Table 8. CTA Ventra System Fare Revenue by Sale Channel, 2017

Sale Channel	**Revenue in Ventra ($millions)**
Ticket Vending Machine (TVM)	$217.3
Mobile Ventra	$73.1
Retailers	$69.4
Threshold Autoload	$55.8
Pre-Paid Benefits (PPB)	$54.0
Patron Website	$36.0
Distributor Order	$16.9
Other	$8.3
Mobile Metra	*-$5.7*
Other Non-CTA Spending	*~-$37.3*
Total	$487.8

The data collected through AFC system provide useful information that can help with estimation and prediction parameters. There are four guiding question that can help the agencies to clarify how they relates because there are better opportunities to use existing estimates when behavioral parameters used for prediction take standards forms (Stuntz, 2018):

1. What are the fare policies scenarios in this example?
   The range of possible modeling scenarios is limited by focusing on two simple decisions – whether to offer a pass and where to price it. One scenario could be when an agency currently offers a single tariff option (flat, pay-per-use fares that are differentiated by transit mode) but it starts to explore self-selection of frequent transit users as a new strategy.

2. What is the theory of behavior?
   The theory of behavior used for modeling has to sufficiently describe customer decision making under both current and potential future fare structures – in this case, both pay-per-use and a new weekly pass. There are three key behaviors or customer decisions that impacts fare changes:

   – *Which fare product to choose?* Or, from the agency's perspective, how many customers will switch from pay-per-use to the new pass?
   – *What additional rides to take with a pass?* For existing customers that do switch to the new pass, how will their ridership change?
   – *Whether and how much to ride?* For potential transit customers that currently use other modes, how many would purchase the new pass and start using transit? (If pay-per-use fares were changing, this would also include consideration of current customers' decisions about whether to continue using transit and how frequently to ride.)

   Figure 36 shows the relation among the three key behaviors mentioned above.

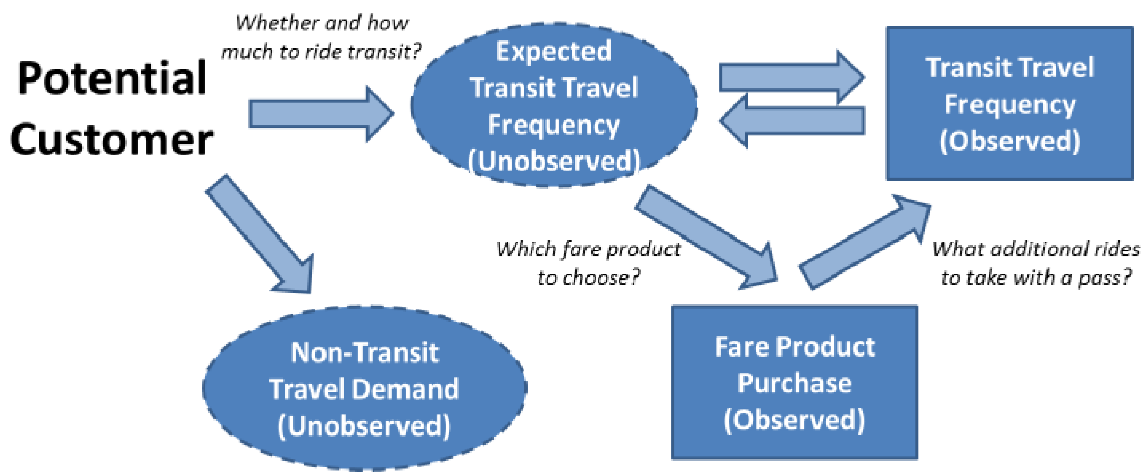


Figure 36 Relation among the Three Key Behaviors that Impact Fare Changes

3. How can these three behaviors be represented in a set of calculations to predict the impacts of a new pass?
   There are three parameters that can form the basis of a prediction procedure which can be combined sequentially in calculations that first allocate customers to different fare products (pass or pay-per-use) using a logit fare product choice formula and then scale ridership and revenue using induced ride factors and elasticities.

- *Fare product choice logit utility parameters*, specifically parameters representing sensitivity of product choices to a) expected weekly cost of transit travel, and b) a customer's inherent preference for one product over another.
- *Induced ride factors*, multiplicative factors describing the relative increase in ridership when a customer switches from riding transit with pay-per-use fares to riding with a zero-marginal-cost pass.
- *Elasticities*, additional multiplicative factors that scale ridership on a fare product up or down linearly with the percent change in price for that fare product (capturing both the switching between transit and other modes as well as changes in ridership *frequency* for continuing transit users).

4. How can these parameters be estimated?
Fare product choice models and induced ride factors can be estimated or approximated from cross-sectional AFC data at agencies offering both passes and pay-per-use fares. For situations when information is not sufficient, then in order to estimate these parameters the agencies would need to use data form prior studies.

3.2.1.2.1 Numerical Results and Implications

Analyzing the ridership shape and impacts revenue from introducing a pass at different prices, it could have completed with a single model. Figure 37 illustrates the customers weekly trip frequency at each level of transit ridership adapted from Ventra data for major account-based fare products at the CTA (Stuntz, 2018).

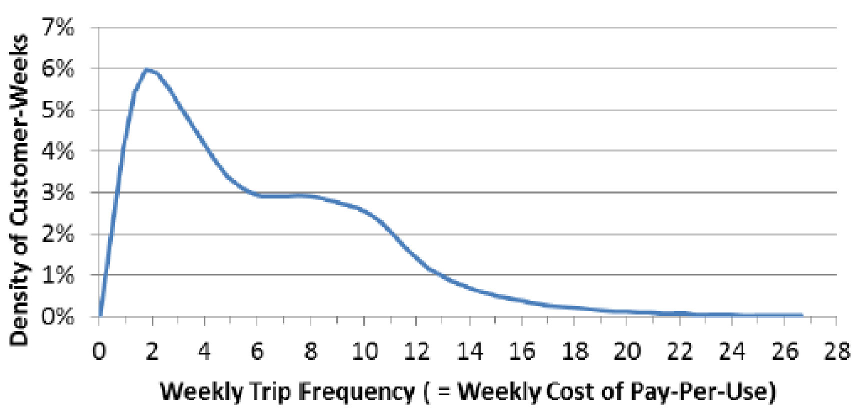


Figure 37. Assumed Weekly Trip Frequency Distribution for Pass Pricing Model Example

As stated in *Transit Fare Policy: Use of Automated Data to Improve Incremental Decision Making* (Stuntz, A. W., 2018), the ridership distribution follows the expected bimodal pattern. Many active customers take a small number of trips, and other transit users follow a distribution centered on commuter travel frequencies. For convenience, the pay-per-use fare is assumed to be $1 per trip and is never changed in this example analysis; weekly ridership is the same as the weekly cost of using pay-per-use. For this projection, single-use tickets and cash data was excluded to avoid any bias of these kind of products because they would look more favorable than others.

There are four behavioral parameters in the simple model. Two describe the choice between fare products using a logit formula, where the systematic utilities are (Stuntz, 2018):

$$V_{PPU} = \beta * (\text{weekly use value})$$
$$V_{Pass} = \alpha + \beta * (\text{weekly use value})$$

The parameter β describes sensitivity to weekly costs, and α describes inherent preference for the Ventra passes (V_{Pass}) relative to pay-per-use (V_{PPU}). An elasticity parameter, ε, captures the marginal impact of changes in transit costs on *mode* choice (switching between transit and driving). Finally, an induced ride factor, f, describes the additional rides taken when a customer uses a pass instead of pay-per-use. When a pass is introduced at a certain price P, some existing customers will shift from pay-per-use to the pass, and is calculated for customers with a given level of transit use using the logit choice formula (Stuntz, 2018):

$$P(pass) = \frac{e^{V_{Pass}}}{e^{V_{PPU}} + e^{V_{Pass}}}.$$

As expected, as fare prices increase, ridership grows meaning customers could change their current fare products to cheaper ones, consequently, revenues increase as well. However, eventually the incremental revenue loss (from existing and new high-frequency pass purchasers) will be greater than the incremental revenue gains (from pay-per-use customers who buy passes and customers who switch from other modes). There are a few basic lessons for pass pricing in this one example (Stuntz, 2018):

- **Use of passes can potentially increase revenue and ridership.** This is clearly not a theoretical inevitability, and it may have resulted from agencies prioritizing ridership gains and setting low initial pass multiples.
- **There is a revenue-maximizing pass price that should never be exceeded.** The revenue-maximizing weekly pass multiple under the initial parameter assumptions is 13.6 with a resulting revenue increase of 7.5% and ridership increase of 3.3%. Ridership increases monotonically as pass prices decrease, so it would never make sense to price a pass any *higher* than this revenue-maximizing point.
- **Considering ridership, pass prices should be set below the revenue-maximizing price.** Revenue is relatively flat around revenue-maximizing price, so there is an exceptional ridership return from foregoing a modest amount of revenue and pricing a pass *lower* than this point. Accepting 2.5% lower revenue (a 5% overall revenue gain) at a pass multiple of 10.3 nearly doubles the ridership increase to +6.3%, and a revenue-neutral pass product at a weekly pass multiple of 8.3 would increase ridership by 9.4%.

Transit agencies can enhance their analytical capabilities related to change in fare structure and fare levels by following three steps process (Stuntz, 2018):

1) Identifying current pricing strategies to organize their efforts. The process should also be accessible to any agency that maintains an automated, transaction-based fare collection system.
2) A simple modeling example focused on a widely relevant fare policy question.
3) The third step identify two important limitations. It is focused on incremental changes to existing fare products and does not necessarily apply to fundamental changes in fare structure (such as totally new products or fare integration with new services).

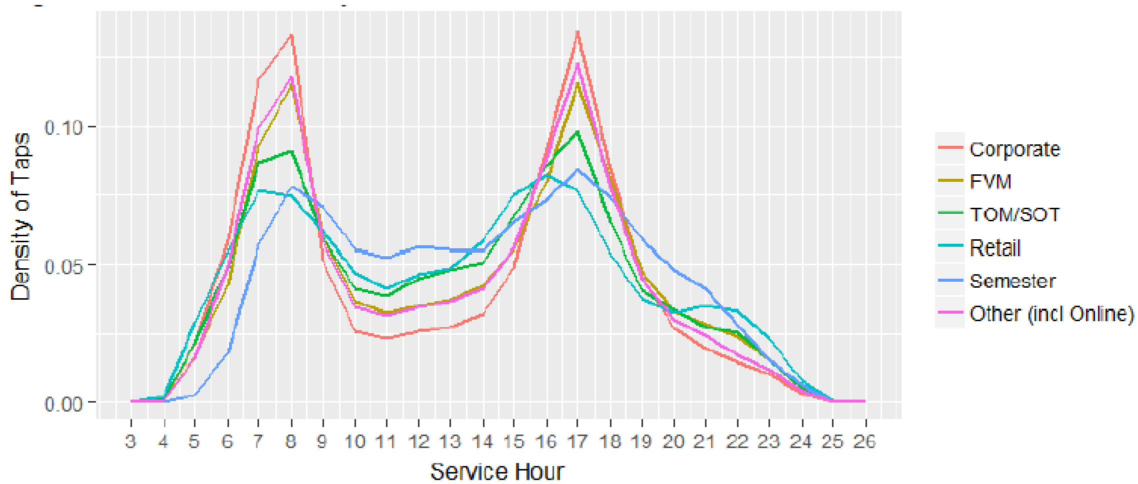
3.2.1.2.2 Using AFC Data to Describe the Role of Pass Sale Channels at the CTA and MBTA

AFC data does not provide demographic information of customers, but the collected information does allow to create a pattern of the customers travel behavior which it is useful for numerous analysis. Significant variation was observed in the pass types sold through each sale channel in both the MBTA and CTA. The largest two sale channels are the Corporate Program and fare vending machines (FVMs), which account for 51% and 30% of the MBTA's full-fare pass revenue (respectively) (Stuntz, 2018):

- Ticket offices and the mTicket app have much lower total sales but represent a sizeable share of commuter rail and commuter boat pass sales.
- Online, Retail, and Semester Pass sale channels represent small shares of pass revenue.
- Rapid Transit and bus passes are distributed similarly across the different sale channels; however, a lower share of Local Bus passes is sold through the Corporate Program, and a larger share is sold in the Retail Network.

With the study of the tickets sales, it can be concluded that sale channels are associated with fare products choices. For example, a customer in the Corporate Program or at a ticket window is more likely to use commuter rail or commuter boat. FVM customers are more likely to purchase 7-day passes, and Retail Network customers are more likely to purchase Local Bus or 7-day passes (Stuntz, 2018).

AFC also allows to explore when fare products are used differently across sale channels. Figure 38 shows that passes sold by Corporate Program are the most used, then passes sold by the FVM and Other/Online sale channels are the second most used followed by the Semester and Retail Network passes are least used and have higher evening and late night ridership (Stuntz, 2018).

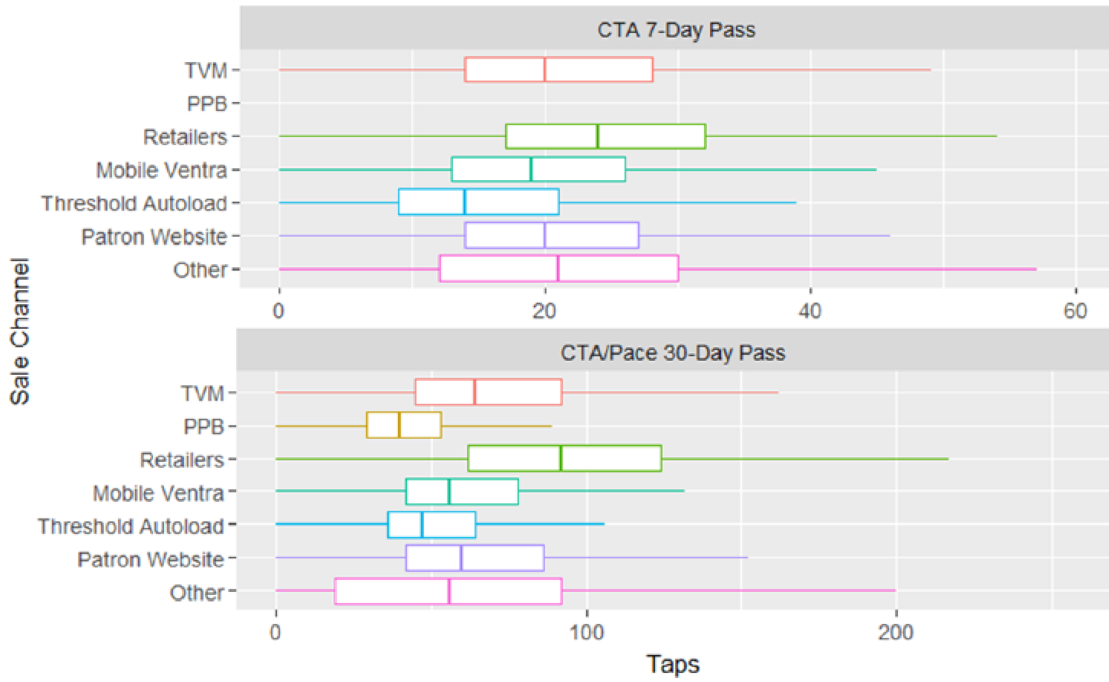


Sources: MBTA AFC, MBTA pass sale program administrative data
Notes: All days of the week are combined

Figure 38. MBTA Monthly LinkPass Time of Use, October 2016

The CTA has only two major pass types that account for over 90% of full-fare pass revenue - a 30-Day Pass valid on CTA rail, CTA bus, and Pace bus (the suburban bus operator), and a 7-day pass valid on the CTA rail and bus. These two passes have different primary sale channels; about 70% of 30-day passes are sold through the Pre-Paid Benefits Program and online channels (mobile app, auto load, and web site), while about 70% of 7-day passes are sold at Ticket Vending Machines and in retail stores. This indicates that the CTA's sale channels reach different groups of customers (Stuntz, 2018).

Figure 39 shows the different pattern of passes sales through different channels. Passes sold pre-tax through payroll deductions (Pre-Paid Benefits/PPB) and passes sold online (via mobile app, auto load, and web site) are used less frequently and more at peak times than passes sold at fare/ticket vending machines (FVM/TVM) and on the retail network. This behavior highlights an interest distinction between seven-day passes and 30-day passes having the seven-day passes much higher tap frequency than the 30-day passes on a weekly basis (Stuntz, 2018).



Source: CTA Ventra
Notes: Horizontal axes on the two graphs aligned such that 30-day pass taps equal 30/7 times 7-day pass taps (for comparability across pass types).

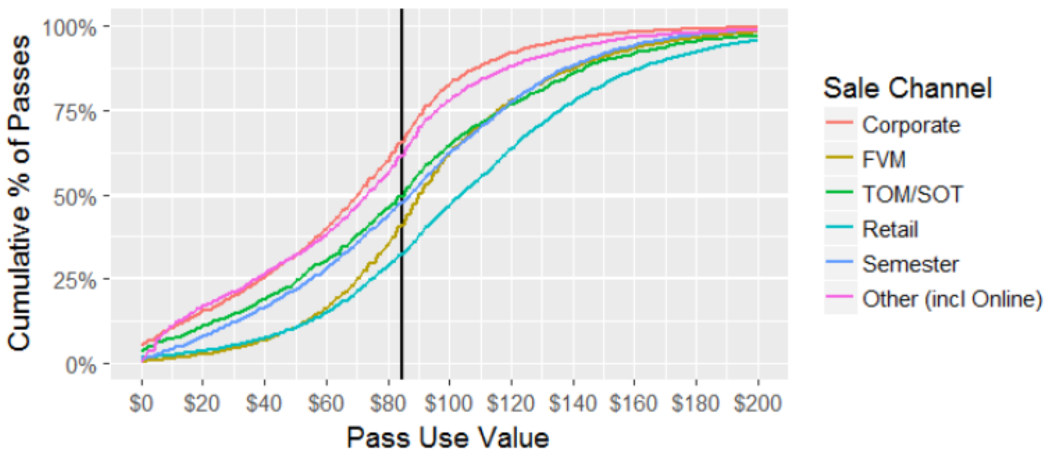
Figure 39. CTA Pass Frequency of Use (Taps), October 2017

3.2.1.2.3 Economic Efficiency and Behavioral Targeting

From the information above, it can be concluded that the collected data through AFC allow the development of patterns based on different customers groups behavior. It was observed that the frequency use of passes is connected to willingness to pay (WTP) for passes, resulting in two variations which would be important for a targeted pricing strategy (Stuntz, 2018).

- First, the opportunity cost of a pass could vary. If one group of customers has higher use value, the opportunity cost of an unlimited-use pass will be lower, and they will be willing to pay more for the pass (on average). Alternatively, if the effective price of a pass is lower for one group of customers (such as customers who can purchase a pass pre-tax through payroll deduction), the opportunity cost of purchasing a pass will also be lower for that group.
- Second, sensitivity to opportunity cost could vary across groups of customers; at any given expected use value and pass price, willingness to pay or the probability of selecting a pass over pay-per-use could vary. For example, between otherwise similar groups of customers, a group that places greater value on the convenience of a pass will be willing to pay more on average than a group that places less value in that convenience.

However, this evaluation does not distinguish between these two variations because it analyzes the pass use across sale channels. Figure 40 shows the use of value distributions for the Monthly LinkPass at the MBTA with pass prices overlaid as black vertical lines. As with earlier plots of tap frequency, passes sold through pre-tax payroll deduction via employers and through online sale channels have considerably lower use value distributions than passes sold in retail stores. Passes sold at TVMs fall in the middle of the use value distributions at the MBTA (Stuntz, 2018).



Sources: MBTA AFC, MBTA pass sale program administrative data
Notes: Includes passes that are sold but never used (use value = $0)

Figure 40. Use Value of MBTA Monthly LinkPasses Sold in October 2016

These differences in use value of the passes across the sale channels between MBTA and CTA could be the results of different aspects such as (Stuntz, 2018):

- Different underlying demand for different groups of customers.
- Different effective prices in different sale channels.
- Differences in price salience.
- Different preferences over upfront costs and convenience for different groups of customers.

3.2.1.2.4 Estimating Behavioral Parameters of Fare Product Purchase and Use

The MBTA and CTA offer unlimited use passes and pay-per-use fares, developing some customers behaviors that need to be identified in order to predict the potential impacts of future incremental changes in fare structure and fare levels such as (Stuntz, 2018):

- Induced ridership. When a transit agency offers both unlimited-use pass products and pay-per-use fares, the zero-marginal-cost nature of the pass products will affect a customer's ridership on a pass relative to the same customer's ridership using pay-per-use.
- Fare product choice. Customers choose which fare products to buy depending on many categories, but the prices of such products are often the main category that helps them decide.

- Elasticity. Having identified their preferred fare product, customers decide when, how, and how often to ride on transit. Ridership decisions are frequently based on service availability and quality, but they are also affected by the marginal financial cost of riding.

Information collected through the AFC varies from long to narrow data. Long data captures ridership information for every card and ticket in a transit system over long periods of time, and narrow data records a limited set of information about fare products and transit rides and typically does not include demographic information about customers (such as home address), data on customer activity outside of the transit system, or data on external factors such as weather or gas prices. Therefore, the information collected can be organized in different format for empirical modeling of customer behaviors related to fare policy such as (Stuntz, 2018):

- Cross-sectional data. Captures variation in attributes and choices across customers, such as transit ridership and resulting weekly cost under a selected fare product, without requiring any special attention to churn between fare products or churn in cards (since cards are not traced over time). There are two primary disadvantages. First, it does not provide any variation within customers; without a good instrumental variable (which has not been identified), it offers no way to separate the impact of transit ridership level on fare product choice from the simultaneous impact of fare product choice on ridership. Second, it cannot be used to evaluate the impacts of changes in fares over time.
- Aggregated time series data. AFC data can alternatively be aggregated into time series data. This is still computationally cheap and still nets out the impact of churn in fare products and cards, but it also captures exogenous variation in fare levels and other fare policies. The challenges of aggregated time series data in a single-agency context are sample size and limited variation.
- Panel data. Preserves a large sample size and maximizes variation in the dataset – both across customers and over time (including individual choices and system-wide fare policy changes). The combination of realistic individual-level choice settings and exogenous variation over time presents opportunities to parse out the relationship between fare product choice and ridership levels. As challenges, the exogenous variation that could be used to identify panel data models it is limited by the infrequent fare policy changes. Also, the panel data is computationally intensive relative to cross-sectional data and aggregated time series data, requiring processing of individual-level information over extended periods of time.

Depending on the format, AZFC data contains variations of in fares, fare product choices, and ridership that are needed to understand about fare-related customer behavior. At any one point in time, one of the variations could be the different fare product prices resulting in customers different purchases. Over time, the AFC data also captures variations within any transit card or account for example customers fare products purchase decision day to day, week to week, or month to month.

These type of over time variations presents two challenges referred as churn. The first is churn between fare products and the second d is churn in cards or accounts (Stuntz, 2018).

3.2.1.2.5 Past CTA Models: Survey Data vs. AFC Data

There are two important differences between the two models (Stuntz, 2018):

- First, previous CTA choice and elasticity models were based on customer surveys allowing the study of some variables such as income and car ownership which are not observed in AFC. However, the survey model could lead to potential errors in scaling factors besides of being and expensive model. AFC, on the other hand, developed logit model estimates using very simple specifications and the resulting logit model formula is applied directly to segments of AFC data. AFC also have some disadvantages such as the limited to incremental fare change scenarios, for example dramatic changes in fare structure or introduction of new fare products could still require customer surveys. Nevertheless, the AFC model could be improved by exploring alternative specifications and introducing AFC-based segmentation.

- Second, the Cambridge Systematics model calculated one system-wide average fare for the baseline and one systemwide average fare for the scenario based on product-level prices; the percent change between these average fare levels was applied to elasticities for four broad segments of CTA ridership (combinations of bus/rail and weekday/weekend), and the resulting total system-wide elasticity impact was re-allocated to fare products based on predicted scenario market shares.

The purpose of the choice and elasticity spreadsheet model is to inform the actual design and selection of fare changes. The primary incentive for a fare change in 2018 was closing a projected operating budget deficit. Following the major reduction in pass market share after the last fare change in 2013, CTA staff were also interested in focusing fare increases on pay-per-use fares to make pass products relatively more attractive (Stuntz, 2018).

Figure 41 shows predicted changes in ridership and revenue by fare product from the Baseline (in the center of each graph) under different pay-per-use fare levels, leaving pass prices unchanged. The black lines show the increasing pay-per-use fares is predicted to decrease total ridership and increase total revenue. The right graph of changes in revenue shows the counterintuitive prediction that pay-per-use revenue could actually decrease if pay-per-use fares are increased, even as net revenue increases. This is due to customers switching from pay-per-use to passes, which is a function of the relative prices of passes and pay-per-use in the Baseline (Stuntz, 2018).

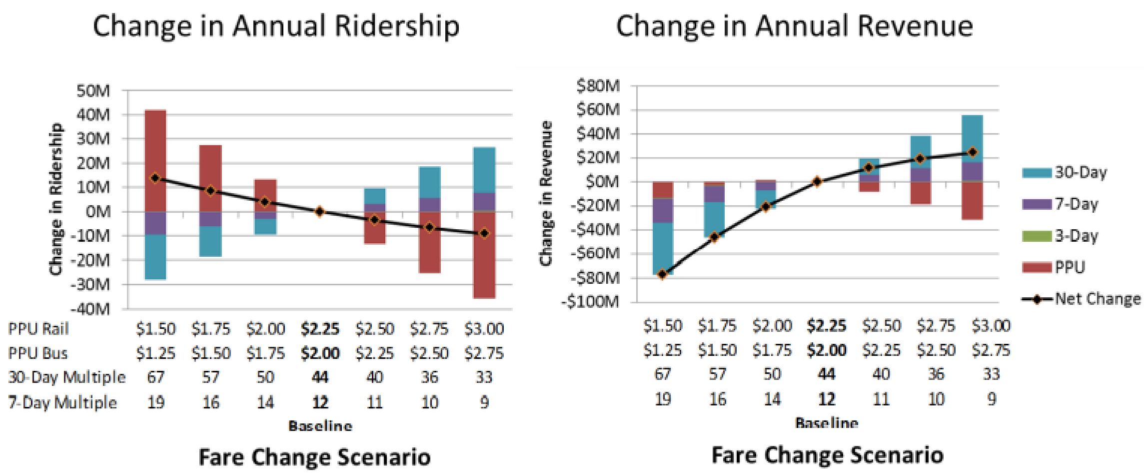


Figure 41. Predicted Change in Ridership and Revenue under Alternative Pay-Per-Use

The procedural framework can be applied to all transit agencies, and the empirical applications are relevant to agencies that collect AFC data and offer multiple payment structures. Many conclusions and recommendations can be deducted from the study such as (Stuntz, 2018):

1. Readily available AFC data provides many opportunities to learn about fare policies and customer behaviors and to improve analysis of incremental fare changes. However, the AFC-based tools and techniques illustrated in this thesis may take too much time to develop during a short political window of opportunity to make changes in fare policy.
2. As seen in the January 2013 fare change at the CTA and the July 2016 fare change at the MBTA, altering the relative prices of pass products and pay-per-use fares (or pass multiples) can result in significant switching between fare products, which in turn affects net changes in sales and ridership.
3. Transit agencies should consider promoting pass products and favoring a high pass market share. A simple model of pass pricing shows that passes should always be priced somewhat below revenue maximization to capture inexpensive gains in ridership.
4. There are several ways a transit agency might increase its pass market share. The most straightforward option is typically to increase pay-per-use fares more than pass prices, improving pass multiples. Ridership losses from increasing pay-per-use fares are mitigated by some customers shifting to passes, and losses can be further reduced by concentrating fare increases on lower-elasticity trip types like rail and peak-period trips. Another option is to make passes more attractive without changing prices, which would increase revenue and ridership by shifting lower-use customers into passes.

Transit fare policy analysis is an area that can be continuously improved. Below are some of the many topics that can be implemented (Stuntz, 2018):

- The pre-tax sale of transit passes through employee payroll deduction was found to generate additional revenue and ridership at the MBTA and CTA. However, this tax benefit is only available at a subset of employers who participate in direct pass sale programs or use third-party employee benefits administrators. Extending tax-free transit purchases to all transit commuters would correct this inequity and would increase agency revenue and ridership. Policies to achieve this could be explored and evaluated.

- Pay-per-use elasticities at the MBTA were estimated for a limited set of trip types. Additional trip features could be developed (trip distance, trips ending downtown, etc.) and used to further differentiate elasticity estimates.
- Fare product choice model estimation with CTA Ventra data used a very simple specification and did not make full use of the panel nature of the Ventra data. Logit choice model specification and estimation could be improved, and machine learning techniques for fare product choice prediction could be explored.
- One limitation of the CTA product choice and elasticity scenario model presented in this thesis is the inability to capture rapid changes in ride hailing as a competing transportation alternative. Cross-elasticities with respect to ride hailing prices and service could be estimated (if ride hailing data was available), or the model could be extended to explicitly model mode choices between transit, ride hailing, and other modes.
- Analyses in this thesis were disaggregated in various ways, but never spatialized. Exploratory AFC analysis, fare-related parameter estimation, and fare change scenario modeling could all be segmented or disaggregated by geography.

3.2.1.3 AFC Data for Market Research and Demand Modeling

The implementation of Automated Fare Collection (AFC) systems provides new opportunities for research to improve the transportation services. The information provided by these systems allow the agencies to formulate questions about the fare services like: How are current fare products being purchased and used, and how have those patterns changed over time? to help them understand how these products are changing the transit services. This study evaluates the potential applications of AFC at the Massachusetts Bay Transportation Authority (MBTA) to analyze the fare policies (Stuntz et al., 2016).

The study of the MBTA fare products illustrates how AFC data can be used to develop intuition and insights about current and potential fare products. First, attributes of the current fare product were identified. Then the role of those fare products in terms of revenue generation and use of patterns were summarized using the information provided by AFC to develop an initial market segmentation. And finally, using partial origin, destination, and interchange inference (ODX) inference information to further segment fare product purchases and use by location and trip type, the summaries are completed to be used to develop fare policy proposals (Stuntz et al., 2016).

Few transit agencies take advantage of the collected data through the AFC to create models to project ridership performance. As stated in *Fare Policy Analysis Using Automated Fare Collection Data: Market Research and Demand Modeling at the MBTA*, key considerations for selecting an appropriate model include the availability of historical data, the attributes that define the current and proposed fare products, and the availability of origin-destination-transfer information.

After the proposals being developed, the study of MBTA continued to evaluate the performance using AFC-based market segmentation. The new model allowed the calculation of expected ridership and revenue impacts, and also the metrics for alternative fare policy objectives such as equity and operational efficiency were estimated (Stuntz et al., 2016).

3.2.1.4 New Fare Collection System

The new AFC 2.0 system will enhance the transportation system due to its many advantages such as (Sikorski, 2019):

- Improve the customer service: it will easier the method of payment by offering convenient options to customers and will create more places to purchases and reload.
- Ensure equal access: it will improve accessibility for seniors, disabled people, low-income families, and other disadvantages groups.
- Upgrade assets: It will update the hardware and the software providing better network communications and exchange data while keeping the assets in a good state of repair.
- Improve revenue control: It reduce the system wide cost of fare collection, provide fully reconciled auditable and accurate revenue deposits and reports, control fare evasion and prevent fraud.
- Focus on core operations: Improve ridership and revenue data and reduce the time of boarding and payment in the vehicles.
- Support the future MBTA: Integrate other agencies, modes, carriers and services. Provide configuration and operational flexibility allow for fare policy innovation.

The AFC 2.0 will upgrade the transit system at MBTA improving many characteristics of the service covering people needs and demands. It provides many advantages like customers could pay before boarding using their card at stations validators installed on the different stations or bust stops, or they could even travel without card using smartphone or contactless credit card. When on board, handheld devices will be used to check the validity of fare media. Multiple card readers will be installed in all doors of the vehicle so customers could board at any door where they could tap their cards facilitating the boarding process. It has been projected the bus speed will improve 10% with this new system. After the transition to the new system is completed, fare boxes will be removed from the vehicles (Sikorski, 2019).

The new system has the following components as illustrated on Figure 42 (Sikorski, 2019):

Figure 42. Components of the New System

To make the transition easier to the new fare media, all the gates will be equipped with the tap targets so people could be getting used to the new system. After the media transaction is completed, a development for a new construction design will be needed to fully replace the old hardware for the new one and wider the gate aisles because tap target will be installed on entry and exit sides to support the option for future implementation of tap out (Sikorski, 2019).

Another advantage of the new system is the facility of manage your account everywhere using a smartphone or a computer where people can (Sikorski, 2019):

- Check balance.
- Purchase value or passes.
- Set auto-recharge.
- View account history
  – Travel taps and charges
  – Payments and purchases
  – Inspections
- Request a new or replacement fare card.
- Register a fare card for loss protection.

- Set personal preferences.
  - Language
  - Accessibility
  - Alerts

As stated in *AFC 2.0 Next generation MBTA fare system* presentation, The MBTA is using two simultaneous contracts (Figure 43) for the implementation of the AFC 2.0 system.

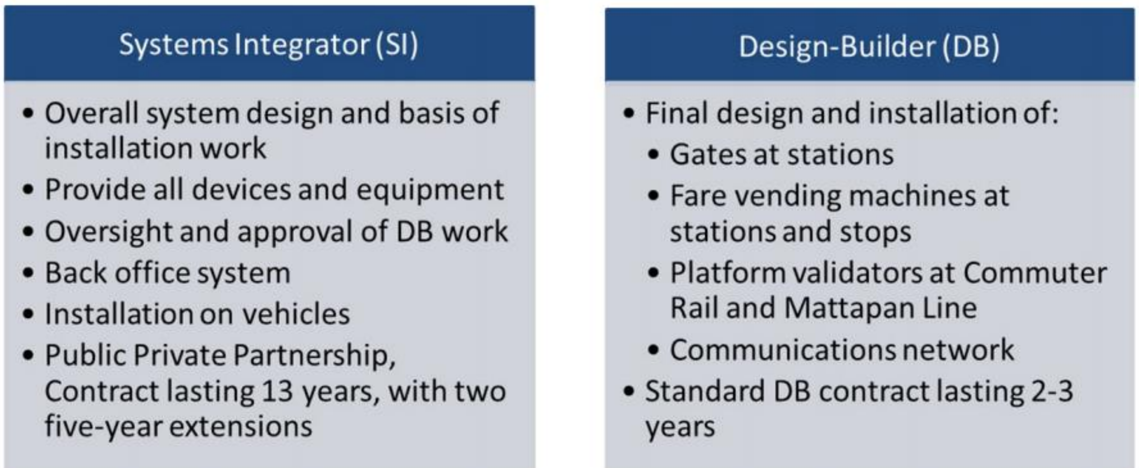


Figure 43. Implementation Contracts for AFC 2.0

3.2.2 Transit Applications

3.2.2.1 Planning Transit Networks

There are many factors, either long or short term, that limit the bus planning process. A new method has been developed that offer the agencies the possibilities of enhancing their planning services, which uses farecard and vehicle location data to provide a previously unavailable level of geographically precise disaggregate data on passengers' linked trips. This thesis studies how automatically collected data can benefit the bus service planning process. The method is based on works that had used AVL and AFC data to infer passengers' origin and destination locations (OD) and to link trip stages into full journeys by identifying transfers. This algorithm, called origin, destination, and interchange inference (ODX), produces a richly detailed data set of passengers' journeys that can be utilized to evaluate the rider's travel patterns and how they can be altered by changes to the bus network (Vanderwaart, 2016).

The new method proposed five phases: identification of target locations, analysis of those locations, development of proposed service changes, evaluation of those proposals, and post-implementation review. As a case study, the Massachusetts Bay Transportation Authority (MBTA) was chosen. Some location's services were analyzed to evaluate or propose changes like the addition of a new route, additional frequency of an existing route, and extensions of an existing route as well as an analysis of the ODX data to improve the current services (Vanderwaart, 2016).

As stated in *Planning Transit Networks with Origin, Destination, and Interchange Inference* (Vanderwaart, 2016), the bus system is the most flexible and adaptable mode available to transit planners, compared to other modes of transportation, because its infrastructure is minimal. Meaning that in a growing city, the bus system is the most convenient and easy to change to cover the needs and demands of the people.

The last year the MBTA added a new bus route was 1994, and it did not change due to a series of winter storms in February 2015 that hampered the transit services. For the first time, in 20 years since the new route addition, the MBTA evaluated its bus network with the possibility of changing its planning methods. However, the MBTA updated their technology by implementing different programs to increase their serviceability (Vanderwaart, 2016):

- An automated vehicle location system (AVL) records the position of each bus along every route in the system over the course of the day.
- Plastic, contactless CharlieCards and magnetic-stripe paper CharlieTickets replaced the old subway tokens in 2006.
- Automated fare collection (AFC) system records which farebox or subway faregate a passenger uses to pay for a ride and at what time.
- Automated passenger counters (APCs) on a portion of the bus fleet record how many people board or alight at each stop and therefore how many people are on a bus at any given time.

This research develops a framework for using ODX data for medium- to long-range bus planning, focusing on areas with constrained resources that need a new service plan to improve and increase accessibility to transit services. This framework will not only serve Boston but any city in the United States that has ODX available. This method is designed to provide an analytical approach to transit planners and help them in the decision-making process. It is important to highlight that the method should be customized to any city based on their characteristics and needs of their current transit operations (Vanderwaart, 2016).

In order to develop these new applications, it is important to understand the data sources and processing algorithms used to create the data set (Vanderwaart, 2016):

- ➢ Automated vehicle location data (AVL) records the location of buses and trains over time. Bus AVL typically uses GPS technology, sometimes supplemented by odometer readings, while trains and light rail vehicles may use GPS or data from track signals, circuits, or sensors. For buses in Boston, there are three types of AVL data: timepoint data, which records the time at which each vehicle passes a handful of set timepoints on a route; “heartbeat" data, which records GPS coordinates for each vehicle every 60 seconds; and what is generally referred to as "announcements" data which is a system that provides announcement for the visually impaired saying the name of the stop each time the bus doors are opened and the name of the next one.

- The second major source of automatically collected data is the automated fare collection system (AFC). In Boston it is known as the CharlieCard system, which it also handles paper CharlieTickets. All transactions at ticket vending machines, station faregates, and bus and light rail fareboxes are recorded, with the most common being paying for a trip and adding value to a card or ticket. Cash transactions at bus fareboxes are also recorded. The most relevant parts of an AFC record are typically the time of the transaction, the ticket or card number, the farebox or faregate number, and the fare type.
- Automated passenger counters (APCs) provide a third source of automatically collected data. Sensors on both front and rear doors detect passengers entering and exiting at each stop, resulting in a record of the number of boardings, the number of alightings, and the total number of passengers on the bus (load) after each stop.

The use of collected data through the automated systems became accessible in recent years. Traditionally, the evaluation of the transit system was mostly completed using customer surveys which take a tremendous amount of time to process and a bigger budget. The bus networks have evolved over the years and instead of designing it from the beginning, the routes were added as needed. In older cities such as Boston, transit networks were initially developed by the private sector and routes were introduced or frequency was increased when revenue exceeded operating costs (Vanderwaart, 2016).

As population increase, it is crucial to identify all possible improvements that can lead to positive changes in the coming years. The Boston region's street network is complex, with many very narrow streets, one-way streets, and complicated intersections. Developing a proper optimization problem would require a deep knowledge of the city geometric design. Current optimization methods assume a fixed level of demand which needs to be changed and have the imminent population growth in consideration (Vanderwaart, 2016).

At MBTA, real-time data is used for operations and customer information, while historical data is mostly used for performance reporting. The archived AVL data is used as an input for the Hastus scheduling system to make schedule adjustments based on actual running times. However, there is little use of archived data for long-range network planning. The MBTA Service Planning department receives information from the AVL, AFC, and APC systems. However, they have little capacity for more sophisticated and complex data analysis that combines multiple data sources.

Suggestion of origins and destinations using ADCS has been a subject of research since AVL and AFC systems became widespread. Early inference methods often had limitations such as only covering a single mode, inferring destination locations but not times, relying on schedules or fare transaction times rather than on AVL data, or assigning boardings to zones rather than individual stops. The MBTA's inference system is based on both entry and exit transactions, and inter-modal transfers.

ODX method provides valuable information that can be used for different applications varying from identifying changes in travel patterns to uncovering reliability problems to determining the effect of service changes on existing riders. The latent demand or demand by riders who are not currently using transit can also be measured. The MBTA current service is more focused on how the service meet the service standards rather than how well it meets the region transportation needs being an insufficient approach to address the transit problems. For this reason, a new method is proposed to improve the existing ones and it is composed of five phases (Vanderwaart, 2016):

1. Identifying where changes are most likely to be needed.
   The first phase of the process is identifying a set of target locations where the transit services are really needed. The MBTA system has been static in recent years, so locations that have experienced significant changes during that time are likely to be less well-served than those that have not changed much at all. It is important to consider a variety of location types for several reasons. One is that the area could become politically controversial. The locations should be spread throughout the city. Having a variety of locations is more likely to lead to equitable outcomes and address more issues of concern.

    - Economic Growth Centers
      Employment data from the U.S. Census, the Census Transportation Planning Products Program (CTPP), or state sources can show where the number of jobs has increased significantly over a recent 5- or 10-year period. These data indicated the employment rates which can be used for the MBTA to adapt their transit services to cover people's needs.
    - Residential Areas with Below-Average Transit Outcomes
      A study confirmed that African American commuters have longer commutes than white commuters in the Boston area. While these studies focused on the travel differences of two racial groups, the methods can be used to examine transit equity more broadly among racial and ethnic groups as well as across income levels.
    - Areas with Changing Travel Patterns.
      While ODX data is currently only available for a few months, over time it will become available for longer periods. This will enable longitudinal analysis of travel patterns. Where travel patterns are changing, the transit network may need to change as well in order to appropriately serve the new demand. Each trip and journey can then be assigned a starting and ending zone and the travel in each zone can be examined in the aggregate. Zones where the change in boardings, alightings, or transfers is out of sync with systemwide ridership patterns may indicate important target locations. Changes in the proportion of travel occurring at different times of day may also indicate a location with changing transit needs. ODX makes possible to determine what role a particular location plays in individuals' travel and to identify some limited demographic information.
    - Areas with Significant Transit Crowding

With the boarding and alighting information from the ODX data set, load profiles can be developed showing the level of crowding at different points along a route. This information is useful to reduce the crowding level and improve the speed and reliability of the service in the area.

- Areas with Significant Traffic Congestion
  Traffic congestion areas are good locations to be examine for two different reasons. First, congested areas are more likely to experience reliability problems due to transit vehicles that are stuck in traffic. The second reason is that they may present an opportunity to encourage mode shift away from cars and towards transit by providing better transit service.
- Areas with Poor Accessibility
  Improving the transportation services of low accessibility areas will lead to an economic development. Connecting people to places where they usually travel for different reasons make a big impact in the society. A common metric to have in consideration is the number of jobs that can be reached within a certain time, such as 30 minutes.

2. Analyzing existing travel patterns to and from these locations to determine the challenges at each. The goal is to identify problems and challenges during a person trip. Some of the statistics that can be easily calculated in this manner through the ODX method include:

   - The number of journeys from each zone to the location over a specific time period.
   - The median or average travel time to or from the location during a particular time period (such as the morning peak). A map showing the area that can be reached within 30, 45, or 60 minutes of a location would provide similar information.
   - The average speed of journeys to the location, either the straight-line speed or the speed along the journey's path.
   - The average time spent waiting during transfers on the way to the location.
   - The transit mode split among the journeys to the location (bus, rail, and light rail)
   - The variability of travel time to the location.

   As well as other criteria could de added such as population figures, information about other modes, or additional information about the existing transit network:

   - The number of trips per capita to the location
   - The mode split among car, walk, bicycle, and transit for travel to the location.
   - The travel time by car or bicycle to the location, and how this compares to typical travel times by transit.
   - The area that can currently be reached from the location with no transfers (sometimes called single-seat rides), determined based on the map of the transit network.

3. Developing proposals for service changes to address these challenges.
   This phase propones solutions to the identified challenges from phase 2. Problems with the geometry of the route, the solution will require a change in the geometry design like adding, altering, splitting, or extending a single route, or combining multiple routes. Problems of system performance require different solutions that include increased frequency on a route, operational strategies, bus priority measures, and other improvements along the route. Another problem could be the journeys that are effective in determined hours of the day and ineffective at other times due to a limited span of service or very infrequent service during certain times of the day. Changes to the span service or the frequency are the common solution to this problem, but when the resources are limited, another way need to be found to improve the service between two locations.

   It is important to know that only solving the bus transit problems will not enhance the transportation service in general. If rail service is not efficient, then the bus service could be affected. Other problems may be difficult to fully address due to topological and geometric factors, such as a limited number of streets crossing a river or railroad, a convoluted street network, a large park located between the origin and destination, or streets that are too narrow to accommodate turning buses.

4. Evaluating these proposals and determining which should be implemented.
   Costs will generally be straightforward to determine. The main cost of most bus service proposals will be vehicle hours of service at each period of the day, which can be found using expected new running times and headways for any route that has changed or been added. Evaluating the benefits of each proposal is more complicated. Each proposal may have impacts on initial waiting time, in-vehicle time, waiting time at transfers, the number of transfers that riders make, crowding, reliability, accessibility, and ridership.

   There are two methods that can be implemented to estimate the effect of service changes on demand. The first is a simple elasticity model. The second is a direct demand model, which is developed by performing a regression analysis on a number of variables affecting transit demand, such as population and distance from the central business district. After evaluating the different proposals, then a decision could be made about which ones give the best solution to the problems. The final group of proposals selected for implementation should maximize benefits while minimizing costs, address the most significant of the challenges identified by this process, spread benefits to different areas and constituencies within the region, and be feasible with available resources.

5. Implementing the selected proposals and performing post-hoc evaluation of the service changes. The components of implementation will depend on the type and scale of the

change, but may include funding, operations planning, issuing new timetables, and public information campaigns to announce the change. This post-hoc evaluation may include a variety of elements. At a basic level, the analysis should examine how travel times, speeds, and number of transfers have changed in geographic areas affected by the new or altered route. The implementation of the selected proposals could also raise some questions about whether a service change was a positive one and whether it is having the intended effect on the network.

This approach main goal is to provide transit service in areas where it is really needed based on the existing mobility patterns. This is a crucial change of thinking from the point of view of the transit provider to the customers perspective. Programs like AVL and APC are used to analyze reliability and running times and for load information respectively allowing the transit planners to take advantage of the collected information such as: population and employment trends, demographic and socioeconomic data, economic development patterns, and traffic information, among others, to have a better idea the real problem, obtain reliable results, and have a better decision-making process (Vanderwaart, 2016).

This study created a framework (Figure 44) for bus planning using ODX data to increase and improve the services in determined areas. For this research, the model was created based on the information provided by the MBTA, but it could be modified to any transit agency. Several studies are being performed due to projected population increase and the concern that transportation system is not equipped to cover the needs and demands of the passengers. Even though the MBTA service planners and Central Transportation Planning Staff (CTPS) rely on AFC, AVL, and APC data in their analyses, the framework developed in this research using ODX can enhance the service planning process by focusing in the needed areas of the transportation service using practical tools and methods. Service planning can never be completely automated, and will always rely on detailed local knowledge, the judgment to balance competing goals, and an understanding of the political aspects of planning transit service. This framework is a toolkit to transit planners to optimize the transportation system rather that reconfigure the entire bus network radically (Vanderwaart, 2016).

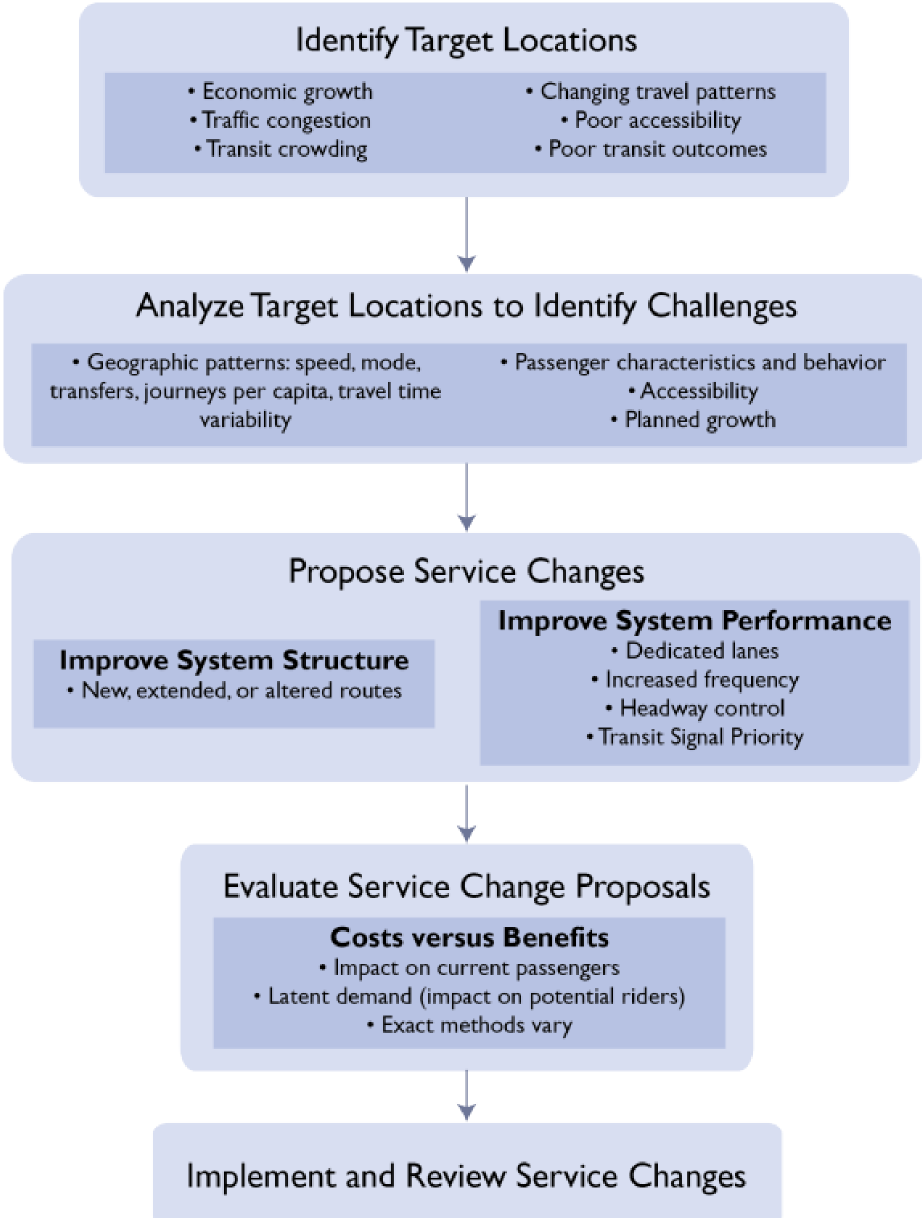


Figure 44. Framework for Bus Planning with ODX

3.2.2.2 Transit Equity

The collected data through Automated Fare Collection systems are opening various possibilities for transit planners on developing different methods that can facilitate the prediction of people origins, destinations, and transfers during the transit trips. An origin-destination (OD) prediction can be used in the analysis and reporting of agencies' social goals, such as the provision of equitable service regardless of race, national origin, or ethnicity, which is federally required in the USA by Title VI of the Civil Rights Act of 1964. Title VI prevents agencies receiving federal funding from having a disparate impact with regards to race, ethnicity, or national origin. In complying with this law, large transit agencies must report regularly on how their service is provided to populations with different demographics. As stated in *Analyzing Transit Equity Using Automatically Collected Data* report, the objective of this study is to demonstrate a preliminary methodology to link automatically inferred OD information from regular transit users to the demographic data of public transit commuters from the US Census's American Community Survey, and to examine variation in passenger-centric metrics such as journey time and speed.

Massachusetts Bay Transportation Authority (MBTA) was chosen as a case study to analyze their OD projections by using their passenger's information obtained though the automatic data collection systems (ADCS). MBTA is the transit agency responsible for the operation of bus, light-rail, and heavy-rail transit in the Boston metropolitan area and oversees the operation of contracted commuter rail and paratransit service in the urban core. The transit service is provided by 191 bus routes including four bus rapid transit (BRT) routes, three subway lines, and a light rail line with four branches that operates as a subway in the downtown core. The MBTA network includes 7,691 bus stops and 127 light-rail (LRT), heavy-rail, and BRT stations (Dumas, 2016).

The use of this automated fare collection data allows not only to determine the connection of riders with transportation network, but also to study their behavior patterns in order to infer the origins and destinations (OD) of their trips. This type of data also provides a more accurate and reliable information than the traditional surveys (Dumas, 2016).

This thesis is based in the synthesis of passenger-centric public transit information and primarily updates the work of Gordon about inferring origins, destinations, and full journeys in London. The motivation for this study is explained below (Dumas, 2016):

- State of the Art: Using ADCS to Infer Travel Behavior: The state of the art has moved in several directions, from improved methodologies, to inferring activities from OD, or using alternative massive passively collected data sets to infer travel behavior.

- Activity Inference: Moving beyond inferring the origin and destination of a trip, researchers have developed methodologies to elucidate trip purpose from the user and trip characteristics as well as land use characteristics.

- Alternative Data Sources: Researchers have developed methodologies to use more global data sources to infer travel behavior.

  - ADCS Applications in Boston: The MBTA's fare data have been used to analyze the effects of different fare policies. It determined the impacts of fare increase on 2012. It estimated the revenue made by the MBTA's Corporate Pass program. And the recorder transactions history can be used by customers to remember their trips. Researchers have also used ADCS to improve transit operations by identifying reasons for delays on the subway and then piloted and evaluated solutions to these. As well as designing real-time control strategies on high-frequency bus routes using real-time AVL data.

- Motivation: Using ADCS to Analyze and Improve Transit Outcomes: Researchers are exploring how ADCS can inform social policy. In contexts with robust censuses, these applications are intended to supplement censuses by providing intermediate diagnostics in between censuses. This can allow for rapid feedback on policy changes without the need for intermediate surveys of target populations or to wait for census results. An analysis of the data from 2011 of the Boston Metropolitan Area found a significant commute time penalty for Black commuters versus White ones across all modes which was most pronounced on the bus (on average an extra 70 hours per year)

This thesis focuses on using the collected fare data to study and infer origin and destinations of the customers trips to provide a finer grained equity analysis to satisfy the Federal Title VI. To accomplish this goal, the following objectives needs to be met (Dumas, 2016).

- Infer boarding and alighting locations and times for bus journey stages in Boston.
- Infer alighting locations and times for rail journey stages in Boston.
- Infer interchanges between journey stages of any AFC-enabled mode.
- Prepare OD inference process to be run over months.
- Develop methodologies for analyzing spatial variation of transit effectiveness.
- Link users to demographic census data
- Determine if differences exist in home-based journey characteristics across demographics and space.
- Propose and evaluate a set of solutions to differences in transit.

The thesis is divided into two parts: the work required to process and infer ODs in the Boston context and the subsequent use of this OD to analyze spatial variation in transit effectiveness in Boston. By creating ODs, the origin (location and time) and destination (location and time) of passenger's trips could be projected. The use of boarding and alighting times well as the coordinates of the origin and destinations are important when a journey is predicted. By applying heuristics to the time, a user spends between stages, as well as the spatial characteristics of these stages, one can link stages together into journeys if no trip-generating activity can be inferred to have occurred between stages (Dumas, 2016).

The different among open, closed and hybrid AFC payment systems is that open AFC system collects the least amount of information about user behavior: collecting a fee and recording a timestamp only when users enter the system. The closed payment system with distance-based and/or time-based fares calculate the customer's fare on each exit transaction recording the destination location and time. The hybrid payment system includes a combination of open and closed modes, typically an open bus system and a closed rail system (Dumas, 2016).

The fare information collected by the automated systems help to determine the patterns or sequence of stops in a route and directions of a vehicle. This process is performed to filter the amount of stops to that the AVL system GPS records may be matched to infer the boarding or alighting stops. Then the patterns are assigned to customers to limits the set of stops events at which the customer can board or alight. This assume that the customers will not stay on the vehicle after the trip is ended (Dumas, 2016).

The MBTA must synthesize the stop arrival and departure times from raw AVL data and from the set of scheduled stops and their coordinates. From the identified pattern, for both the vehicle and the transaction, a possible OD projection could be made by inferring the origin when matching the vehicle location to the user based on the transaction time. Also, by examining the user's transaction history, it is possible to determine the next trip origin and with OD-inference algorithms find the information necessary to infer destinations. This method assumes the customers do not use any other mode of transportation during the projected trip (Dumas, 2016).

For rail systems that have gated entry, the times of origin and locations are recorded by the AFC transaction. The pattern used in the destination inference include the network of stations that can be accessed from behind the entry gate. The arrival time can then be inferred from observed stop times at the inferred destination or from scheduled travel times (Dumas, 2016).

To determine the origin and destination of a route, a high accuracy data is required for a better analysis. Any error of a few seconds can cause the process to choose a different origin than the one used by the user and can cause a discrepancy between transaction and trip-start times at terminals resulting in having passengers in the wrong vehicles (Dumas, 2016).

The MBTA's AFC system collects fares on bus, LRT, and subway. Commuter rail fares are currently validated by conductors and are not recorded automatically, however passes exist that can be used on both commuter rail and the rapid transit network. The AFC system records detailed transaction information for cash, magnetic-stripe paper tickets (Charlie Tickets), and Radio Frequency Identification (RFID)-equipped smart cards (Charlie Cards). The figure below shows the entity relational diagram for this preprocessing with the output table on the right (Dumas, 2016).

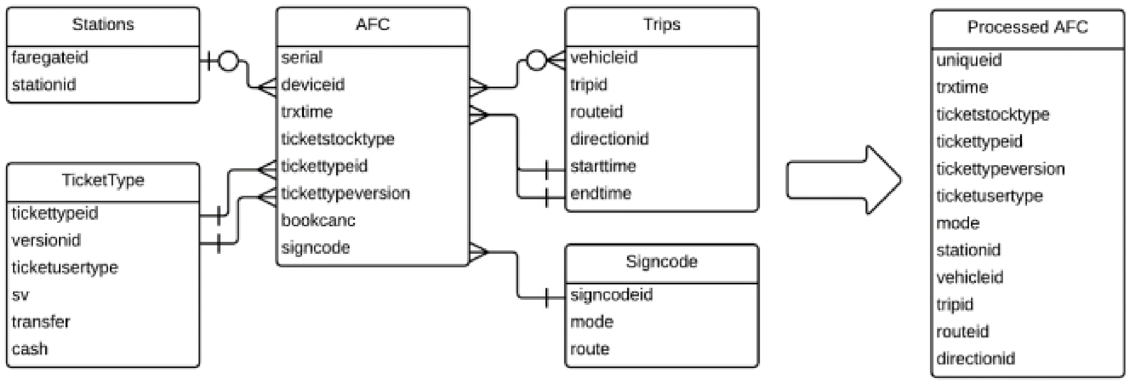


Figure 45. AFC Preprocessing Relational Diagram

The goals of the preprocessing are to (Dumas, 2016):

- Create unique identifiers for fare payers.
- Assign origins for LRT and other stations.
- Assign a mode to every transaction.
- Assign a vehicle trip to bus transactions.

Since the identifier is not unique to every customer due to the mix of the three different payments (cash, ticket and card), a compound identifier is developed to assign a unique identifier to the users for the OD inference algorithm. The surface portion of the Green Line LRT did not have accurate stop-level AVL data, so origins were inferred at the branch-level on the surface portion of the Green Line. This was done by matching the signcode to a signcode lookup table and using the LRT branch as the origin (Dumas, 2016).

There are three different modes a transaction can have for the purpose of origin inference. The transaction can (Dumas, 2016):

1. be made at a station gate and require its destination to be inferred within the network of stations that can be accessed without making another transaction (all subway, and subterranean LRT & BRT)
2. be made on a vehicle and require its origin to be inferred while requiring its destination to be inferred within the network of stations that can be accessed without making another transaction (surface LRT and select surface BRT)
3. be made on a vehicle and require its origin to be inferred and require its destination to be inferred along a route (bus and surface BRT).

These modes are assigned in the AFC preprocessing using the following mutually exclusive conditional statements, respectively (Dumas, 2016):

1. If the transaction's farebox ID is matched to a station.

2. If the signcode is a surface LRT or the transaction is assigned to a trip on one of the BRT modes that enters the Silver Line Tunnel (Silver Line Shuttle, Silver Line 1 and Silver Line 2).
3. If the transaction's farebox ID is matched to a bus and that transaction occurred within a trip.

A transaction is assigned to a bus strip to limit the set of stops during the search of potential origin and destinations. This is completed by matching the transaction to a trip performed by that bus based on the transaction time and the trip's start and end time (Dumas, 2016).

The slowly running of a vehicle farebox clock could provide inaccurate inference of origin and destination. It was found that clock was recalibrated when vehicles were in the garage and the farebox communicates with a central server during refueling or cash extraction. The analysis also found that clocks ramble by roughly 7 seconds, which it is a significant error for many purposes but for infer origin and destinations, it could lead to making error when inferring origin and destination. To address the issue of clock drift, the timestamps of AFC records are corrected by interpolating the temporal error of each farebox between clock calibrations. Clock correction is performed using the following methodology (Dumas, 2016):

1. The temporal error between the two systems' observation of this event is determined to be the farebox clock drift since the previous clock calibration.
2. An automated linear regression analysis is performed for each farebox with clock drift as the dependent variable and the independent variable the amount of time since the previous calibration.
3. For each regression, the slope of the regression line, the rate of drift per day, will be used to estimate each farebox's drift for each transaction using that farebox's rate of drift and the time since the previous calibration.
4. If the regression for a given farebox has too small a sample or too low a coefficient of determination (r^2), the median rate of drift from valid regressions is used.
5. Finally, the time of each fare transaction is corrected using the equation below, by adding the product of the time since the device's most recent calibration and the estimated drift per day.

$$Corrected\ Time\ =\ Uncorrected\ Time\ +(Time\ Since\ Last\ Probing\ *\ Drift\ Rate)$$

In a search for publicly available Title VI service monitoring reports, Dumas, (2016) looked at 12 Metros in the USA with subways. They only found reports for the MBTA and the Bay Area Rapid Transit (BART). This confirms that the FTA triennial Title VI reports are not widely distributed by agencies. Nevertheless, it is worth noting that despite of not making the triennial Title VI reports publicly available to the FTA, New York City Transit (now known as MTA) has published their used methodologies in peer-reviewed documents.

Note that because of the 2008 recession, NYCT faced a significant budget shortfall in 2009 having them to think in other ways to analyze equity service. They developed a method to estimate average fares experienced by individual farecard holders and assigned minority/non-minority and low-income/high-income status to individual farecard holders based on the status of the stations or bus routes that they first swiped in at. For subways and bus route services, the routes were individually analyzed based on the travel time metric (Dumas, 2016).

For the MBTA minority classification, census data is used. "Census tracts are designated as protected if they exceed the average proportion of minority residents of 26.2% for the entire MBTA service area". 175 municipalities were served by bus, rapid transit, boat, or commuter rail within Massachusetts. Prior to 2010, the analysis was performed for 2 different zones: the urban fixed-route service area (65 municipalities) and the commuter rail service area (175 municipalities) (Dumas, 2016).

The following metrics were reported by Dumas, (2016) on the service provided to minority areas:

Bus

- Vehicle Load: Disparate impact is based on the percentage of routes that pass the standard, which is based on the peak passenger load relative to the number of seats.
- Schedule Adherence: For bus, 75% of timepoints on route must pass the on-time criteria which is either schedule-based for service with a headway greater than 10 minutes or headway-based for high-frequency service.

Rail

- All heavy rail lines have over 40% of boardings occurring in minority tracts so no comparison is possible.

Light-rail

- Vehicle load: All lines pass the vehicle load standard during weekdays.
- Schedule adherence: None of the lines pass the schedule adherence on Weekdays except for the Mattapan High-Speed Line. This results in no disparate impact since a greater proportion of minority light-rail lines passes the standard (1 of 4) than non-minority (0 of 1).

As a result of Boston analyses, it is shown that the travel times are different for the different ethnicity groups. The problem is to treat the minority population as a homogeneous group rather than investigating different groups individually. For bus travel, when riders have disparate travel times is the result of not incorporating ridership into findings of disparate impacts (Dumas, 2016).

To improve the Title VI & environmental justice (EJ) Reporting using ADCS, it requires the following analyses (Dumas, 2016):

1. Surveying demographics and ridership patterns. Since American Community Survey (ACS) is collecting information about race, color, national origin, language spoken at home, household income and travel patterns except English proficiency, the combination of matching inferred OD and originating neighborhood demographics could satisfy the spirit of the requirements.
2. Analyzing fare changes and major service changes. For all service changes, travel time should be the metric used for analysis and disparate impact examining distributions of changes in travel times due to proposed modifications. For span reductions, analysis should first examine the potential for customers to use alternative service. If alternate service is infeasible, then should be determined whether the proportion of affected riders is disproportionate with respect to the population.
3. Service monitoring. The use of inferred origins, or APC if inferred OD is not available, and ACS data should be used to determine the number of riders with protected status affected by vehicle loads and schedule adherence that fail performance standards to determine disparate impacts. Vehicle load is a good measure of passenger discomfort from failing service, while also providing some information about where, when, and whether passengers are left behind. Schedule adherence should be replaced with a passenger-centered metric such as excess journey time to better measure poor service.

3.2.2.3 Limitations of Inferred OD

There are three categories of population for which OD information, by its nature, provides limited insight for equity analyses (Dumas, 2016):

1. Individuals who can switch to modes such as auto to avoid burdensome transit trips.
2. Individuals who have no available transit .
3. Individuals who access transit by non-walking modes such as driving or biking.

The first two categories raise questions about which population must service be equitable. The MBTA guidelines for service provision are different for areas where population density area greater than 5,000 persons per square mile. In this case, it is assumed that users walk to areas where there is bus, light rail, or heavy rail service, at most ¼ mile. This does not apply to areas with densities with less than 5,000 persons per square mile, since it is assumed that users in those areas drive (Dumas, 2016).

The use of inferred OD assumes an equivalent of access time. Note that due to housing availability in the inner city, or low stop density in the suburbs, some users may have to travel greater distances to access transit. Inferred OD only measures the trips that users really make and there are limits to this information (Dumas, 2016).

3.2.2.3.1 OD Processing for Analysis of Spatial Variability of Transit Effectiveness

A link between a customer and a farecard will be created to obtain demographic information like the racial/ethnic proportions of transit commuters to infer OD near the area (Dumas, 2016).

Figure 46 shows start time for bus, streetcar, and heavy rail commuters according to the American Community Survey 5-year estimates for 2013. The start time discrepancy could be due to the required time a customer needs to access transit services from their homes. The difference between the number of fare cards and the estimated number of commuters may be explained by several factors such as the use of multiple fare cards, the presence of regular transit users in the AFC sample who are not considered workers by the ACS, and commuter rail users who transfer to rapid transit (Dumas, 2016).

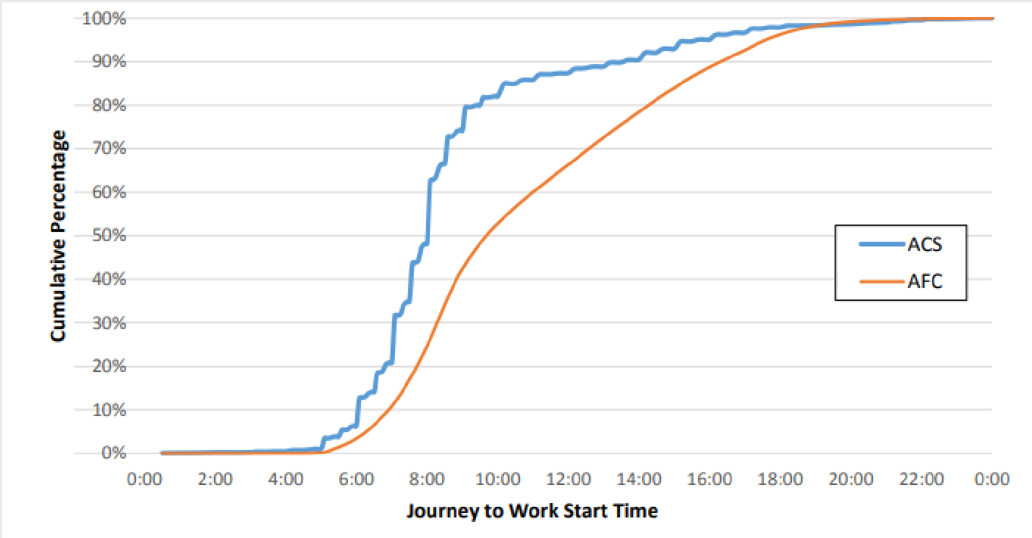


Figure 46. Distribution of Users for the Journey to Work

The 95th percentile of ACS commuters was used to exclude late start times, which corresponds to a journey to work start time of 15:10. The problem could be the nocturnal workers because they could multiply the spans service days. According to the data from the ACS, this means that as many as 4,000 users (2%) would be night workers. So, they would appear in the AFC data as having early transactions, since their first transaction of the service day would be the return journey from work around 3:00 a.m. Nevertheless, the proportion of AFC users with early median starts is smaller than the ones found in the ACS data. Further, there is little evidence that night workers are a large group as compared with regular transit users (Dumas, 2016).

3.2.2.3.2 Analysis of Spatial Variation in Transit Travel Times Using OD Data

The purpose of Title VI and EJ analysis is to identify disparate outcomes, not to prove discriminatory intent. Analyzing passengers experience using OD data it is better than supply-side metrics because it can highlight the deficiencies and inequities of the transit system (Dumas, 2016).

For this study, a comparison between Black and White commuters was chosen to evaluate the travel time differences between them. From the AFC data, regular users were selected and the linkage between them and their farecard was completed using the public transit demographic proportions from the American Community Survey at the census tract level and their deduced home locations.

The number of public transit commuters by race, income, citizenship status, occupation, and travel time are identified by the U.S. Census Bureau's 2013 five-year estimates of the Means of Transportation to Work by Selected Characteristics using the 2010 census tracts. Figure 47 shows the shows the proportion of different races of public transit riders identified by the ACS for census tracts which include categories where there is not AFC inferred OD information such as commuter rail and boat (Dumas, 2016).

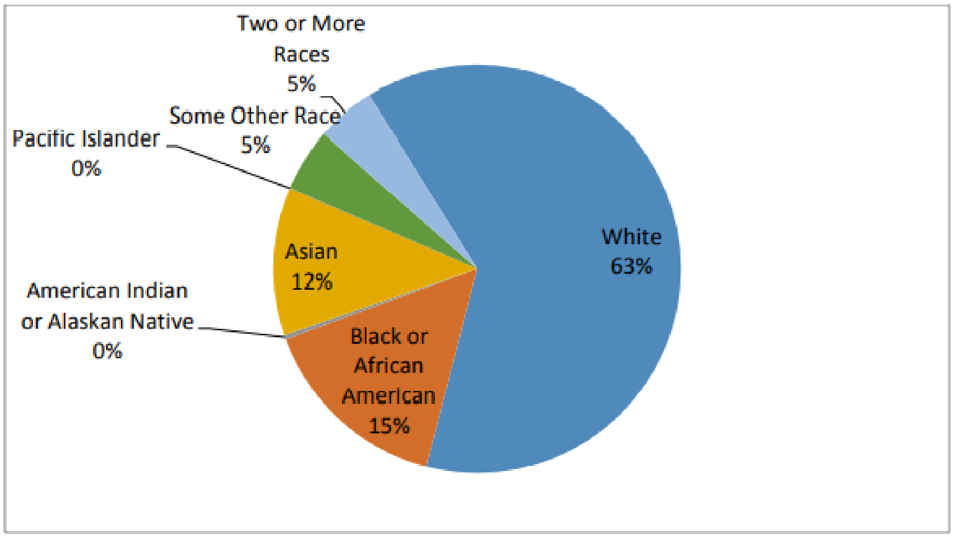


Figure 47. Proportion of Public Transit Riders by Race

AFC farecards are linked to census tract by estimating the probability of the passengers residing in that area based on the intersection of a user's home area with the tract geometry. Of the 4.23 million first weekday journeys that regular commuters took in April 2014, 3.38 million (79.8%) started from that user's inferred home and 1.96 million (46.3%, for a destination inference rate of 58.0%) of those have an inferred destination and arrival time (Dumas, 2016).

The collected information through ADCS could be used to create numerous methods to improve the transportation system. This thesis developed a method to infer origin and destination of a passenger trips using AFC data. A linkage between user's information and their farecard was developed given an insight of the area demographics of their neighborhoods at the census tract level by inferring users' home locations from the first trips they make on weekdays. The predicted ODX information was based on journey outcomes such as speed and journey time that were developed and compared, giving a passenger-centric perspective on the spatial analysis of transit effectiveness (Dumas, 2016).

The MBTA adopted the bus ODX-inference algorithm developed and validated for Transport for London for their bus and rail network. Using the data from April 2014, customers who were likely to walk from their homes to the near transit access were selected to be able to infer their home locations through the collected fare information. The demographics were also observed making a comparison between areas with high concentrations of White non-Hispanic public transit commuters (White Alone) and those with high concentrations of Black or African American public transit commuters. A difference increase was observed as the concentration of Black commuters increases beyond the area average proportion that the Federal Transit Administration (FTA) would recommend in its guidance. This difference is the result of the need of Black or African American commuters' areas to transfer to complete journeys of equivalent distance (Dumas, 2016).

After inferring the waiting times, it was possible to calculate the average travel times from all trips. These resulted in an overall average travel times of 32.6 minutes for commuters from Black tracts and 29.5 minutes for commuters from White tracts, resulting in a difference of 3.1 minutes. It was observed that this gap is greater than the gap of the individual modes because of the different mode splits and the relative travel times for each mode (Dumas, 2016).

There are some factors contributing with this difference (Dumas, 2016):

- First, since the ACS data is self-reported, differences in averages will be amplified due to respondents tending to report their travel times in 10 to 15 minutes increments given that commuters from Black tracts tend to transfer more.
- Second, the average rail journey times are around 15 minutes shorter than the self-reported averages, and the bus travel times are 20 minutes shorter generally. This is because the AFC observations do not include access and egress times as part of journey time, whereas the ACS does. This implies that walking to access transit can be substantial, and that Black commuters on average may reside further from rail or bus lines.
- Third, it is possible that the ACS data are simply inaccurately over-reported or weighted and that this leads to the large differences in travel times.

The speed differences occurred mainly during the early morning, before peak hour, when commuters from White tracts had much faster service on the Red Line than commuters from Black tracts on the same line. However, an accurate analysis cannot be performed due to the lack of OD information for boardings of the Green Lines leading to an exaggeration increase of rail service to White areas (Dumas, 2016).

An analysis of speeds by distance shows the greatest difference in bus speeds for trips travelling between 5 and 11 km, which represent 13% of trips from Black tracts. The bulk of bus trips, 69% of trips from Black tracts, are in the shorter range from 500m to 4000m with insignificant difference in travel speeds. For trips involving bus and rail, speeds were slower due to more transfers being required and a greater proportion of trips transferred to the slower Orange Line (Dumas, 2016).

3.2.2.3.3 Recommendations

Numerous recommendations were proposed based on the thesis results that could improve the operational speed for Black commuters. Figure 48 illustrates some of these potential solutions obtaining a 3.1 minute decrease in average journeys (Dumas, 2016).

	Potential Solution	Potential Impact on Travel Time Difference (% of Current Gap)	Comments
Operations	**Improve Bus Departure Reliability at Dudley and Forest Hills**	**2%**	**Doesn't include wait time reductions for users starting journeys at those stations.**
	Through-routing most Heavily Used Bus Route Pairs	**0.3%-1.0% per route**	
Fare	**Reduce Commuter Rail Fares at Hyde Park**	**Further analysis required**	
Capital Improvements	**Rapid Transit Frequencies on the Fairmount Line**	**25-35%**	**Use of a network model to examine all disaggregate benefits to users who could use the line required.**
	Increase Heavy Rail Frequencies on Orange Line and Ashmont Branch	**Further analysis required**	
	Reconfigure Bus Network	**Further analysis required**	**Examine bus speed improvements from BRT Examine benefits from reorganizing bus network to reduce transfers**

Figure 48. Potential Solutions

Recommendations were also given on the different transit modes such as bus, heavy and commuter rail which could enhance their travel times along with the service and functionality (Dumas, 2016).

Bus:

- Improving the reliability of departures at terminals in Dudley Square and Forest Hills, the transfer times would be enhanced resulting in a decrease in the difference in average travel times by 2%. This improving could also lead to a decrease in-vehicle travel times on routes departing those terminals and waiting times for users starting their bus journeys on those routes.
- Another solution could be the merging of either of the Silver Line Washington St. branches with routes 28 or 23 which could provide one seat per customer. This would decrease their travel times by an average of 4 minutes, and result in a 0.5% decrease in the overall difference in travel times.
- Extending the SL4 through Washington Street in Downtown Boston to the Blue Line ought to be further investigated to provide users' better access to the Blue Line and slightly shorter walking distances to the Orange Line.

- Routes 32 and 39 could also be merged to increase speed to access the Longwood Medical Area. This would decrease travel times using those routes by an average of 4 minutes, and reduce the overall difference in travel times by 0.7%

Heavy Rail:

- Investing in Communication-Based Train Control and Automatic Train Operation at the same time new vehicles are introduced to Red and Orange Lines, could improve reliability, capacity, and peak frequency for rail and bus services. Early morning frequencies on the Orange Line and the Ashmont branch could be increased in order to reduce waiting times.

Commuter Rail:

- Reducing the fare on the Main Line commuter rail line at Hyde Park from Zone 1 to Zone 1a would allow users currently riding the 32 to the Orange Line to access commuter rail stations such as Ruggles and Back Bay, and locations near South Station to have a faster trip.
- The introduction of Diesel Multiple Unit (DMU) service running at Orange Line frequencies on the Fairmount Line, which runs in the middle of the Black or African American area of Roxbury, Dorchester, and Mattapan would improve frequency and running times and could reduce the travel time gap between 25 and 35%.
- The service within a ½-mile (800 m) radius of station on the Fairmount was evaluated with and improvement of 5-minute headways and 16% faster running times than current schedule. From this evaluation it was observed that half of the 5% of Black trips meeting that criteria would switch to this improved Fairmount service, benefiting from an average decrease in travel time of 13.5 minutes. This switch would lead to an overall decrease in the average travel time difference between White and Black commuters of 10%
- Because the walk-only trips represent only a small portion of the trips which could benefit from the Fairmount Line, their benefits were extrapolated in aggregate for other trips which could benefit from the Fairmount. This led to the upper bound of estimated potential travel time difference reduction of 35%.

3.2.2.4 Prioritizing Equitable Growth

As stated in *Prioritizing Equitable Growth Through Fare Policy* (Haney et al., 2019), The Governor's Commission on the Future of Transportation stressed that Massachusetts has a unique opportunity to convert its 400-mile commuter rail network into a multidirectional regional rail system that supports more geographically balanced economic development across the Eastern half of the state. Transit-oriented development (TOD) in Gateway Cities is key to positioning commuter rail to fuel growth in this manner. However, a new commuter rail fare policy is needed to achieve equitable results for low- and moderate-income households. Guarantying the equity is a top priority due to the disconnections among affordable transportation options, affordable housing options, and economic opportunity.

The improvements in rail service in major cities throughout the United States is leading to reinvestment in urban neighborhoods bringing as consequence that low-income households that depend on public transportations are often displaced from the neighborhood due to new line services being adopted. Shifting from commuter rail to regional rail will take years, but it is a necessary change that can make incremental progress with immediate benefits (Haney et al., 2019).

Various fare equity concepts were explored analyzing the MBTA's current commuter rail fare policies with an equity lens. MassINC has published numerous reports over the years documenting the steady rise in income inequality that accompanied our economy's dramatic shift to knowledge industries. Fare equity is a topic that comes into debate every time a public charge is implemented. The basic equity framework for evaluating such tradeoffs has two dimensions (Haney et al., 2019):

- Horizontal such that all groups pay the same price for the level of service they use.
- Vertical such that those with greater means contribute proportionate to their ability to pay.

Arriving at the optimal mix of horizontal and vertical equity is partially a question of values and partially a question of efficiency and effectiveness. It is particularly important to have an agreement with these tradeoffs because transportation networks provide access to fundamentals such as employment, education, and healthcare, and are thus essential to equality of opportunity (Haney et al., 2019).

Fare equity is a topic that will help planners and policymakers to design a better infrastructure that serve people needs and demands. Unfortunately, this paradigm for equitable TOD has rarely been realized. As a result, many Gateway City residents and community-based organizations serving low- and moderate-income families are skeptical that future transit improvements and associated development will benefit disadvantaged members of the community (Haney et al., 2019).

Transit agencies are creating discounts programs to help the disadvantaged population, but this activity has been limited to bus and subway systems. The MBTA could become the first to make commuter rail travel affordable if the new standards are adapted to all systems. The following recommendations were offered to the MBTA to achieve that goal such as (Haney et al., 2019):

1. Experiment with means-tested fares.
2. Lower fares for reverse commuters.
3. Reduce fares for off-peak travel.
4. Develop a standard definition of equity and apply it consistently to all planning and policy studies.

3.2.2.5 Transit Fares for Disadvantaged Populations

It is expected by 2050 a significant growth in the globally elderly population. This projection will not only mean that the current lifespan will be extended, but it will bring as a consequence a lot of changes, especially in the transportation services need to be implemented to be able to cover the requests and demands that are expected a greatly increase. For this reason, the Massachusetts Bay

Transit Authority (MBTA) is taking actions to provide better transit and affordable options to the elderly population.

The MBTA conducted a case study to gain knowledge on how the aging population impact the transportations services. It is projected that by 2030, the Massachusetts population will be aged 65 and up and even though the transit services are considered one of the oldest and most comprehensive public transportation systems in the U.S, it is important to take the necessary measures to be prepared for the future. As stated in *Transit Fare Discount Processing Improvements for Disadvantaged Populations* report the highest concentration of senior transportation service providers (including medical and nonmedical transportation) is in Boston, leading this analysis to sample the older adults from that area. The collected information shows that the adult population is becoming increasingly racially, ethnically, economically, and linguistically diverse, posing future challenges to the MBTA to develop a better service to this portion of the population. The MBTA is also designing a new fare payment system (Automated Fare Collection 2.0 or AFC 2.0, implementation delayed until further notice) to centralize the fare collection and expands payment options and uses group designations to assign fare privileges to specific populations, including older adults.

The study was completed with a panel of 23 adults, all over the age of 85, who live in the greater metro-Boston area, to obtain some understanding about their experience in accessing the MBTA benefits. Three-quarters (75%) of participants report they use driving as their primary mode of transportation, while over 60% report they drive every day or almost every day. They consider driving more convenient than other modes of transportation because it is associated with comfort, reliability, efficiency, enjoyment, and accessibility (Patskanick & D'Ambrosio, 2019).

As years pass, driving percentage among older adults decrease. Many factors could lead them to take this action such as medical conditions, anxiety, stress, increase dependency on their relatives etc. Also living in a crowded city is not particularly easy for them because they must deal with parking difficulties, road safety, and may have trouble seeing in the dark. All these factors could lead them to stop driving and that is when public transportation became essential because it could offer a solution to their transportation situation (Patskanick & D'Ambrosio, 2019).

Participant in the survey stated they frequently use the paratransit service "The RIDE" and they are pleased with it rather than other public and private ride-sharing transportations because the drivers know their public and provide a proper service. However, some said that to travel using the RIDE they have to plan with plenty of time because the routes are large, and it could take longer than expected to arrive on time (Patskanick & D'Ambrosio, 2019).

The MBTA provide benefits for older adults on their transit services, however many of them stated they had had a hard time accessing them. One of the difficulties is that in order to start the application, they have to go to the Downtown Crossing CharlieCard store in Boston and provide a valid government-issued license or ID for proof of age. The problem is that many of them do not have the necessary means to get to the office or lack of a valid ID affecting their probabilities of receiving the benefits (Patskanick & D'Ambrosio, 2019).

Transportation is part of our daily activities. Even when we grew older, moving from one place to the other is still important in our lives. For this reason, it is fundamental to condition the transit system to meet the demands and needs of older adults as they are projected to increase in the coming years. Most of the participants in the survey are satisfied with the transportation services available to them, but they did not have the knowledge of the benefits through MBTA. The panel selected for this study is relatively well-resourced in terms of education, wealth and access to technology and information. However, there are gaps related to information they did not know. As changes in the transit system are being implemented, it is important to develop a communication method that reach the older population regarding their access to technology (Patskanick & D'Ambrosio, 2019).

One limitation this study had is the selected area for the study which it was geographically exclusive to the metro-Boston area, bringing as consequences that the recommendations do not apply to communities outside the region. For future research, a broader range is suggested to include a larger sample of older adults with different characteristics and situations to have a better understand of how the information related to the benefits of public transportation are reached to them or not. Another recommendation is to investigate with senior service providers and/or agencies that host Senior CharlieCard application to improve the fare application benefits for seniors (Patskanick & D'Ambrosio, 2019).

Even though the panel participants highlighted that driving is more convenient, many of them use the RIDE services, which could be the first step in changing the perspective of seniors about public transportation. Therefore, MBTA is offering transportation benefits and options designed for older adults to help change that perception by increasing senior ridership and providing a more reliable and efficient mode of travel (Patskanick & D'Ambrosio, 2019).

3.2.2.6 A Vision for the Future

Trying to improve the transit system is becoming essential for the people who relies their daily activities on public transportation. For this reason, the Governor's Commission on the Future of Transportation provided valuable recommendations regarding the transportation system stating as his first recommendation to "Prioritize investment in public transit as the foundation of a robust, reliable, clean and efficient transportation system" (Governor's Commission, 2019).

Fifteen Regional Transit Authorities (RTAs) provide public transportation service to millions of Bay Staters in Massachusetts, outside of the Greater Boston region. This service is crucial for the community because help people to move from one place to another serving as a connection to works, schools, hospitals, grocery stores, social activities, and family support networks, contributing to the growing of the area and the economy (Governor's Commission, 2019).

As stated in *A Vision for the Future of Massachusetts' Regional Transit Authorities* report (Governor's Commission, 2019), a Task Force on Regional Transit Authority Performance and Funding was created in the fall of 2018 by Outside Section 72 of the FY 2019 Massachusetts State Budget which purpose was to investigate the challenges and opportunities facing transit service providers. Most RTAs were facing budget problems despite the increase in funding for Fiscal Year 2019, having as a consequence that many agencies have limitations when trying to enhance their transportation services. Improving the public transit should be a priority for the Commonwealth, meaning that other methods should be implemented to emphasize such importance. For this reason, five categories were defined to help with the projection of the RTAs: Investment & Performance, Accountability, Service Decisions, Quality of Service, and Environmental Sustainability.

- Investment & Performance:
  - The legislature should fund the RTAs in fiscal year 2020 with a base of $90.5 million in state contract assistance.
  - Provide communities with the tools they need to increase local contributions to RTA funding, including through regional ballot initiatives.
  - Establish a Human Services Transportation working group to explore ways to better collaborate, improve service and save money through the brokerage system.

- Accountability:
  - Maintain local control of day-to-day operations and management of the RTAs, while standardizing performance metrics for level and quality of service and increasing regional collaboration to present a statewide vision for public transportation in the Commonwealth.

    MassDOT and RTAs should work in cooperation for the planning and decision-making process for a better management. Anyways, the RTAs make annually reports of performance measurements like Ridership, Customer Service and Satisfaction, Asset Management, and Financial Performance to the state because most of their funding comes from it. With this cooperation, MassDOT and the RTAs could define which data could be labeled as estimated since it varies from year to year creating some categories to portrait it in a yearly comparison for a better comprehension.
  - There should be a reinvigorated RTA Council that fosters greater collaboration, promotes best practices, and provides a statewide vision for RTAs.

- Service Decisions:
  - RTAs will continue to succeed by understanding their markets and by aiming to have their service networks meet the current and future mobility needs of their region as well as support connectivity to other regions where possible.

    Public transit system plays an important role in the growing society because can support the economic development, job creation, and reduce pollution. It is part of the daily lives of many people and its demand is projected to increase in future years. Taken into consideration these principles, the RTAs should develop or update a Comprehensive Regional Transit Plans (CRTP) to identify community needs for service and restructuring and ensure consistency with state and regional goals. Then, and evaluation of the sustainability of existing service will be performed and an analysis of the benefits and cost will be completed before providing any new service. RTAs and MassDOT will discuss the need for additional resources to meet the unmet need if available financial resources do not allow for current service or needed expansions. And lastly, to offer better transportation services to the community, RTAs will work with local partners, including TMAs, municipalities, regional economic development organizations (REDOs), Chambers of Commerce, employment centers, Workforce Investment Boards, and major business, healthcare and educational institutions.

- RTAs should identify routes in their service areas where there is a demonstrated community need for seven day a week and evening and night service.
- In communities that sit on the border of two RTAs, RTAs should work together to increase access to cross RTA services to better unify the Commonwealth's public transportation and increase access to more robust services.

- Quality of Service:
  - RTAs should determine which routes are prone to bus crowding and address the issues that cause bus crowding.
  - RTAs should ensure fixed routes, or on demand services where appropriate, maximize multimodal connectivity. To this same end, new infrastructure which addresses first-last mile problems, especially sidewalks, bike lanes, racks, bikeshares, and/or lockers, should be prioritized.
  - RTAs should formally include the public in decision-making on matters related to new projects, fare changes and service planning.
  - MassDOT and the RTAs should carefully utilize farebox recovery ratio as a performance metric, considering the ratio in context with other factors and balancing the need to maintain the affordability of service.

    Farebox recovery ratio is a performance metric that should be used carefully to avoid creating big problems for example: withholding transit services in areas populated by seniors. It should also avoid imminent fare policies that increase access, like providing discounted fares to low-income riders, or students. Another potentially useful metric may be to measure an RTA's total own-source revenue recovery—a measure that would include revenue generated from parking, advertising, special grants and other sources.
  - RTAs should modernize and standardize fare collection by partnering with the MBTA and adopting the new AFC 2.0 system on a statewide basis, while still maintaining an accessible system for cash customers as appropriate.

MassDOT and the RTAs should cooperate to incorporate technological advances into transit service. Microtransit, Automated Passenger Counters (APCs), Automated Fare Collection (AFC), and Automatic Vehicle Locator (AVL) technologies will not only speed up the process of payments avoiding unnecessary delays, but also provide valuable information which could be used to enhance the transportation system. Updating the service with these new technologies could also increase the quality of customer service and attract new customers due to the facility it provides to riders of using the same payment card in different methods of transportation. With Automated Fare Collection 2.0, or AFC 2.0, riders will be able to tap and board at any door with a fare card, smartphone, or contactless credit card. They will be able to reload using cash or credit card at vending machines at all stations and some bus stops or go online to manage one's account 24 hours a day. For these reasons, the RTAs should partner with the MBTA to implement the AFC system to enhance the service provide to the population. However, they should maintain a system for on-board cash payment for riders that are not tech savvy or do not have other option.

- Environmental Sustainability:
  - In order to reduce greenhouse gas emissions from the transportation sector by at least 40 percent by 2040, the RTAs and MassDOT should determine the mode shift that will be required to meet that goal, as well as work with local partners to create a long-term environmental sustainability plan.
  - In keeping with the state's environmental goals related to transportation, all public transit bus purchases should be zero-emissions by 2035.

Various recommendations were made to the RTAs to enhance their transportation system. This is an important topic that needs to be address because public transportation demands are projected to increase in future years, for that reason it is critical to implement the necessary measures nowadays to create the proper conditions for future changes. Besides the recommendations mentioned above, there are other topics that could be taken into consideration such as: the use of nontraditional performance metrics the use of centralized, standardized and customer-friendly technologies for RTA operations, the use of technology that could make paratransit more useful and responsive, and how mobility can be enhanced in communities not served by the RTAs (Governor's Commission, 2019).

3.3 Utah UTA

3.3.1 Data and AFC Systems

3.3.1.1 UTA Fare Payment Systems

Transit agencies around the world are studying and executing new methods to increase their ridership and improve their services. One of the most innovative is the contactless smart card payment systems which allow customers to pay their fares in a most efficient way. These systems typically issue agency-branded smart cards that are either used exclusively by riders in a closed payment system or, in some cases, used in retail locations established as extensions of the closed system updating its value on each card use. With the implementation of contactless credits and debit payment card program, the transit agencies are changing their roles of payment media issuer to acting more like a merchant in an open system increasing the transactions spread in compliance with the transit payments requirements. The main reasons of this research are (Smart Card Alliance, 2011):

- To inform the transit industry of the opportunities and benefits of accepting contactless open bank cards for fare payment
- To inform the bank card industry of unique requirements for transit fare collection

To fulfill the necessary functions by the Automated Fare Collection (AFC) systems, the following requirement must be met (Smart Card Alliance, 2011):

- Transaction Speeds
- Flexible Transit Fare Policy Support and Pricing
- Data Integrity and Customer Service
- Multiple Payment Options
- Data Security and User Privacy
- Transit Benefits Program Support

The use of contactless credit, debit, and prepaid cards offer many benefits either for the customers and the transit agencies by speeding the payments transactions on the vehicles and opening opportunities for agencies partnerships, co-promotion, and new revenue streams. Some of these benefits are (Smart Card Alliance, 2011):

- Account-Based Architecture to enable acceptance of contactless credit, debit, and prepaid cards.
- Interagency Interoperability
- Flexible Implementation of Government and Pretax Program Benefits on Prepaid Cards
- Flexible Use of ID and Access Media for Fare Payment
- Merchant Role for Transit Agency
- Well-Defined and Globally-Accepted Security Standards
- Reduced Payment Media Issuance
- Lower Fare Collection Costs
- Improved Customer Service
- Speed of Deployment
- Additional Revenue Opportunities

As a case study, the Utah Transit Authority (UTA) was chosen because it has fully implemented and launched a new electronic fare collection system that accepts contactless bank cards for payment of transit fares at the point of entry showing the benefits and challenges the new system have faced and how they have been addressed (Smart Card Alliance, 2011).

The Utah Transit Authority (UTA) is the regional transit provider for the primary urbanized areas of Utah. Its service area primarily consists of a 100 by 20-mile corridor bounded by the west face of the Wasatch Mountains and the Great Salt Lake and Utah Lake. A population of 1.8 million is served. UTA operates 520 buses out of four garages, 80 paratransit vehicles, four light rail lines over 35 miles, and a 44-mile commuter rail line (Smart Card Alliance, 2011).

There were many concerns when they first started the investigation of the electronic fare collection system such as: the convenience and ease of use for its customers; the efficiency and effectiveness of revenue collection; and the data that could be collected to understand transit system use and guide service planning. When the payment industry launched the new electronic fare collection system under the brands of American Express ExpressPay, MasterCard PayPass and Visa payWave, the variety of companies involved were attractive to UTA for the following factors (Smart Card Alliance, 2011):

- Issuance of payment media by other organizations
- Integration with the payments mainstream: payment at the fare box, gate or platform as a merchant POS transaction.
- Automatic interagency interoperability
- Customer service with issuers
- Security standards
- Flexible architecture for product development
- Robustness of the open payment's ecosystem
- Commoditization of devices
- Potential for a pathway to eliminate cash.
- Speed of deployment
- Cost
- Co-promotion by issuers

UTA's pilot was performed on 41 ski service buses with the objectives mentioned below. The completion of the pilot was a success where UTA acquire a perception of how to manage the new system when implemented in all vehicles (Smart Card Alliance, 2011).

1) Solve an immediate problem – accounting for the use of resort customer season and employee passes issued by and paid for by ski resorts.
2) Learn whether transit fares could be collected using the new contactless open credit and debit cards being issued under the open payment brands.

On January 1, 2009, UTA launched the new system. It included an infrastructure of readers at all doors of 520 fixed route buses and 170 validators installed on the 35 TRAX light rail and

FrontRunner commuter rail platforms; wireless communications gateways on buses with both 3G connectivity and WiFi connections at garages and optical fiber to all platforms; and Internet links from each device to a hosted back office. The initial fare products were contactless bank card acceptance for single adult fares, including honor of transfer rules and use of third party paid passes (ECO Pass for employers, Ed Pass for colleges and universities, and Ski Pass for five ski resorts within the UTA service area). The service architecture has the following characteristics (Smart Card Alliance, 2011):

- Tap-on/tap-off: Tapping at the entrance and exit of each bus or rail trip leg allows collection of linked origin/destination data that is invaluable for service planning.
- Account based architecture: Fares are calculated, and transactions are processed in the back office easing implementation of fare changes and creation and launch of new fare products.
- Open payments acceptance: The system and card acceptance devices are certified, and process contactless cards issued under the American Express ExpressPay, Discover Zip, MasterCard PayPass and Visa payWave brands.
- Near-real-time and real-time authentication: The price of each trip is calculated as the tap-on/tap-off actions are received and then submitted for full authorization and settlement.
- Inspection: Inspection devices for proof-of-payment rail services are using NFC and 3G smart phones to interrogate payment cards and determine through the back office that they had been previously presented and accepted by the platform validators.

The information collected by the system is planned to be used to predict origination-destination trips to improve the service and facilitate a better planning. The system is intended to gather information about the performed trips, not about the person who is traveling. The privacy of the riders has been a priority from the initial planning of the electronic fare collection system.

With the growing population, transit services are projected to be affected because of people needs and demands. For this reason, numerous methods have been analyzed and put into practice to speed the fare payment transactions, increase the payment choices and guaranteeing payment security. Even though the implementation of the automated fare collection system had been a success, there are some challenges that needs to be addressed such as: the means to attract rider markets either unfamiliar with or not needing the purchase of these specialized fare instruments; and the burden of owning, operating, and maintaining proprietary card-based systems (Smart Card Alliance, 2011).

The banking and payment industries understand consumer desires for faster transactions. Studying the origins of customer's preferences is a priority to later implement a system that works for both, the transit agency and the passenger. As stated in *Transit and Contactless Open Payments: An Emerging Approach for Fare Collection* (Smart Card Alliance, 2011), this white paper examines the confluence of two industries, transit and financial payments, moving toward the mutually compatible goal of market expansion through customer convenience, transaction speed, and data security. The document outlines the mechanics of the bank card payment process, including payment aggregation and advanced processing techniques, to address transit needs dealing with authentication, authorization, and approval in real time or near real time.

Two new technologies are being introduced (Smart Card Alliance, 2011):

- NFC, a short distance wireless communications technology, may completely alter the payment landscape by allowing purchases from mobile phone users and enabling location-based advertising and communication.
- EMV (Europay, MasterCard, and Visa) is an open standard specification for smart card payments and acceptance devices designed to improve the security of bank card transactions.

Many transit agencies are struggling with multiple standards and the challenges of interoperability, which could be achieved if the new electronic fare collection system is adopted. Smart Card Alliance, (2011) highlights the benefits of the open payment architecture, hoping to foster a new understanding in the transit community about the importance of a fare collection system change to better respond to the needs of customers.

3.3.1.2 Contactless Open Payment System

The transit service in Salt Lake City, Utah, is using an automated fare collection system where passengers need to tap their cards when using public transportation, allowing the collection of valuable information that can help the transit authority to incorporate more products in the future. The fare payments systems have been implemented decades ago, but because these systems are proprietary, it is difficult for transit systems to reduce their costs, develop innovative services, or generate new revenue streams. Also, fare payment systems are expensive to maintain. First Data Government and Transit Task Force, (2010) goal is to study the current trends in payment systems and the benefits they hold for transit authorities.

Various pilot programs have been implemented using new fare collection methods, showing the opportunity for cost saving, new revenue streams, improved customer convenience, easier integration with other transit modes, and innovative partnerships. For this reason, many agencies are searching for new ways of enhancing their fare systems by upgrading their electronic payments based on contactless bankcards that support Visa payWave, MasterCard Worldwide's PayPass, or other open bankcard contactless applications (First Data Government and Transit Task Force, 2010).

Most transit authorities have used a closed loop system which in term of a payment system restricts what goods or service can be purchased and from which merchants. This system is developed for transportation industry only and the transit authorities must act as its own bank for issuing the stored-value payment cards as well as serve as the financial clearinghouse for transaction settlement. Using this kind of system is expensive for the transit agencies as it cannot achieve the economy level that other financial services companies can. Some disadvantages are (First Data Government and Transit Task Force, 2010):

- Cost – Ongoing support and maintenance costs are high for systems whose components are unique or obsolete.
- Vendor lock-in – Proprietary hardware and software lock the transit authority into using one vendor for the long term.
- Resource allocation – Enormous resources must be devoted to the process of fare collection, which comprises customer service, information technology, maintenance and fare handling.
- Inconvenience for customers – Riders must purchase and carry an intermediate form of payment, such as a ticket, ride pass or token, instead of paying for a ride with a payment medium existing in their wallet, such as a standard debit, credit or prepaid card.
- Lost revenue opportunities – Transit authorities that operate a proprietary fare payment system miss out on revenue opportunities, such as revenue sharing from cobranded bankcard interchange or usage fees.

Along with the technology development, new transit fare collection systems are also emerging to benefit both the passengers and the agencies because of the many advantages they offer. Some of the new methods are (First Data Government and Transit Task Force, 2010):

- Electronic payments: Consumers have embraced electronic payment because they are quite comfortable with using their debit, credit and prepaid cards to pay for low-dollar-value transactions.
- Prepaid cards: The use of open loop prepaid cards is on a steep growth path. These cards are started to be used widely by employers to provide pay to employees; by government and nonprofit agencies to provide payments or assistance to constituents; by unbanked individuals who do not have a regular debit or credit account; by tax preparation agencies to issue tax refunds; and even by parents to give a spending allowance to teenagers. And since they are commonly used, people want to use them in more payments situations like pay for a ride.
- Open loop payment systems: An open loop prepaid card is typically branded by one of the major card brands such as Visa, MasterCard or American Express. The card can be used anywhere that payments via the brand are accepted and payments are processed over the regular card payment network. The advantage this method has for transit authorities that accept open bankcard payments is that they do not have to support their own expensive infrastructure in order to process and settle fare payments for themselves.
- Contactless payment systems: The growth of contactless payment systems is steadily trending upward. For this reason, several transit agencies are implementing this kind of system encouraging their customers to become comfortable with the concept of using

contactless devices based on radio frequency identification (RFID) and ultimately near-field communication (NFC) for initiating secure payments.

- ISO standards: There are open standards issued by the International Standards Organization (ISO) that dictate how cards and other devices interact with hardware such as readers.

The "Open bankcard standards" for payments are set by the world's most common financial networks, including Visa and MasterCard. Customers would be able to use electronic payment methods like credit, debit and prepaid that get processed through an existing financial network. If transit agencies adopt this new payment method, they are considered merchant rather than proprietaries allowing them to procure a merchant acquirer and a prepaid program competitively. The open bankcard payment standards offer many benefits listed below. Therefore, transit authorities are studying the possibility of changing their closed loop system for an improvement service opportunity:

- Fare cards would no longer be issued by the transit authority. The fare products would be tied to bankcards and maintained by a systems integrator, fare systems provider, or transit authority.
- The card readers and other equipment deployed at turnstiles, faregates and fareboxes would adhere to ISO standards. Numerous vendors could supply these devices to the transit authority, which would make them easier and less expensive to purchase, replace, service, and upgrade relative to the proprietary equipment commonly used today.
- The payment networks, with their well-defined processes for settling payments, would replace the proprietary systems the agencies use now for transaction settlement.
- The payment card industry has a security standard that reduces the risk that merchants will lose money to fraudulent transactions.

Contactless fare payments are projected to be the future of small-ticket transactions. Contactless cards and other devices such as cell phones use embedded or attached RFID or NFC chips to exchange data with a reader device that is just a few inches away. The technology is mature, secure, and standardized meaning no proprietary equipment is required to read the data on the cards or phones. Transit authorities are interested in implementing this type of payment and are already testing the technology. Contactless payments will enhance the transportations services by restructuring the fare payment, reducing customer service costs, increasing data, and decreasing the bus dwell time.

Changing to a new system generally comes with challenges. Therefore, it is convenient to implement the necessary methods to help people during the transition period and maintain the transaction speed such as (First Data Government and Transit Task Force, 2010):

- Aggregation (credit and debit). The AFC software can have a policy whereby there is a set amount pre-authorization (pre-auth) value on credit or debit card.

- Shadow balances (prepaid). The processor of a prepaid card (such as First Data or another processor) knows what the value on the card is at any given time.
- Negative files (credit, debit, prepaid). Payment card regulations change regularly and must be reviewed with the merchant processor to ensure adherence to the latest rules and regulations.
- Near-real-time authorization (credit, debit, prepaid). A rider taps his card at the faregate, and it opens immediately to allow him through. At the same time, the transaction is being sent to the host server for authorization.
- Direct connect to host (prepaid). Similar to the process for a shadow account, a fare collection system provider can connect directly to the prepaid card provider to get the card balance.

The Utah Transit Authority (UTA) was a pioneer in implementing a contactless open payment fare collection system. In October 2009, the agency was awarded the American Public Transportation Association Innovation Award for the development and launch of its new electronic fare collection (EFC) system. The launch of the EFC system made UTA the first transit agency in the United States to roll out to its entire fleet a transit payment system based on the open payment network (First Data Government and Transit Task Force, 2010).

This system allows the passenger to pay for their rides using contactless smart cards over electronic readers when boarding. Several credit and debit cards are accepted like Visa payWave, MasterCard PayPass, Discover Network Zip, and American Express ExpressPay. UTA also developed two prepaid smart cards: Eco Pass for workers and Ed Pass for students. The EFC system also accepts standard open payment contactless cards, but the riders need to register them first to be able to use them when boarding any vehicle. Passengers are asked to tap on/tap off their cards when boarding and alighting. By doing this, UTA can collect valuable information that can help transit planners to understand the different patterns of the riders to design or modify the routes when it is needed to improve the transit services (First Data Government and Transit Task Force, 2010).

New methods are being created to improve the efficiency and reliability of the transportation systems. Transit agencies need to consider whether to keep their closed loop system and not being able to cover the needs of their customers due to its several restrictions or change to open bankcard payments standards with the opportunity of increasing their revenue sources and improve customer service. The most important objective is to provide an excellence service and the new payment innovations will help to achieve it by adding convenience, cutting expenses, and enhancing the entire transportation experience (First Data Government and Transit Task Force, 2010).

3.3.1.3 UTA Performance Audit

Utah Transit Authority (UTA) is large multi-modal transit agency considered one of the largest transit systems in the country with a service area in six counties that includes almost 80 percent of the state's population. A deep analysis was completed to the performed audits in 2008 and 2012 to the UTA with the objective of identifying the problems where the agency needs to be focused to improve their service efficiency and functionality. Since UTA provide several services to the population like buses, light rail, commuter rail, vanpool and paratransit services to approved or eligible individuals, its mission and goal was defined more broadly. The UTA board and management are reviewing the recommendations provided by the report for further implementation and it designed a 2020 strategic plan centering the customer needs as its main objective based on valuable input from employees, municipalities, businesses, and customers from 2013. The goals of the plan are shown in Figure 49.

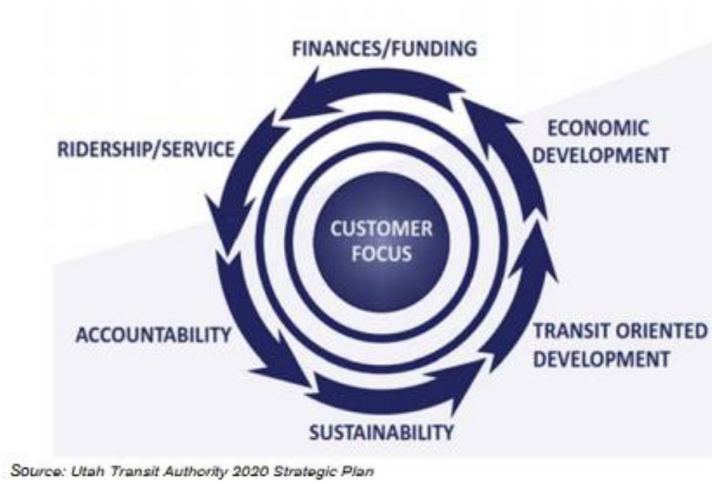


Figure 49. UTA's 2020 Strategic Plan

To improve the problems a system could have; it is important to identify them first and understand their nature. The performance audit reflects the current situation of the UTA fare collection system showing the current findings but also the previous audit recommendations on figure 50 to conduct a profound analysis and suggest possible solutions (Legislative Auditor General, 2014).

Due to the issues encountered, several questions were raised about the farebox policy implemented to the UTA. One of the questions is about the equity of the discounted programs due to audit reports from 2008 and 2012 reported widely different subsidy levels among passenger types and transit modes. UTA hired a consultant to complete an analysis of the fare's policies and how they might be improved to help guide the future implementation of a distance-based fare system. UTA managers and staff can negotiate pass and promotional programs respectively with the objective to increase ridership, but even though these solutions solve an immediate problem, do not provide long-term results (Legislative Auditor General, 2014).

2012 Recommendation Summary	Current Audit Findings
UTA Board of Trustees should clarify its fare strategy including: • Target level of discounts for different types of fare passes • Target level of subsidies for different types of services • Target minimum farebox recovery rate.	Board policy has not been updated since 2008, and subsidy levels continue to vary widely by passenger type and mode. However, UTA's farebox recovery rate has increased for the last several years. The board is currently reviewing a distance-based fare program that may address some of these issues.
UTA should continue to develop good passenger data to support informed decisions on fares and fare policy, including obtaining feedback from transit users.	UTA has implemented electronic fare collection (EFC) products for some fare types; however, the system lacks full implementation and user compliance.
UTA should use more frequent surveys or other means to better understand the passenger experience, monitor passenger trips completed, and estimate transit market share.	Additional metrics are needed to help the board with its customer focus, including: • Customer satisfaction measurements • Transit market share measurements.

Figure 50. Summary of Prior Audit Recommendations and Current

Another question the board should review is the disparity of the fare system between those who pay the public rate and those with discounted passes. Available data indicates the public pay riders pay over twice as much as pass program riders. Public pay riders are defined as those who pay fares through publicly available means (cash, monthly pass, Farepay pass, etc.). Pass program riders are those who have discounted passes through membership in a participating organization (Ed pass, Eco pass, etc.). Figure 51 shows the differences between public pay and pass programs based on ridership and fare revenue (Legislative Auditor General, 2014).

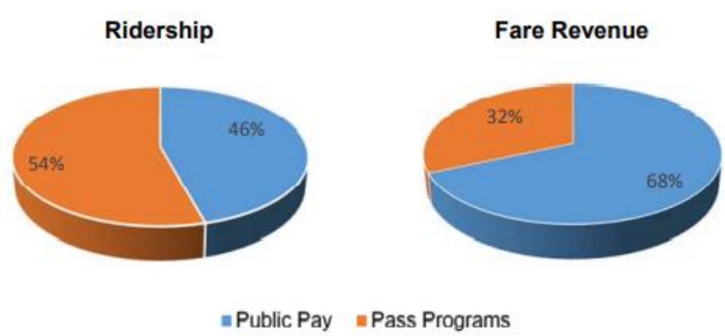


Figure 51. Comparison of 2011 Ridership Data and 2013 Fare Data

UTA was trying to raise its systemwide farebox recovery ratio to 30 percent by 2020 without decreasing ridership. One of the UTA executive stated that the current farebox recovery level was at the highest range it could be based on the market. Lower targets will lead to generate more revenues from the discounted programs, but this situation raises the questions whether too much burden for fare payment is being placed on public riders who are not eligible for the special pass programs (Legislative Auditor General, 2014).

Many discount suggestions were developed by the UTA staff such as (Legislative Auditor General, 2014):

- FAREPAY Card. With a prepaid, reloadable electronic FAREPAY card, riders save up to 20 percent off the base fare through the end of 2014.
- Hive Pass. The Hive Pass is available only to residents of Salt Lake City. Residents can purchase a Hive Pass for one year for $350 dollars, an 85 percent discount off the regular price of $2,376.
- State of Utah Pass Program. State government has transitioned from a standard Eco Pass program to a new type of agreement at an introductory rate. UTA reports state employee ridership has doubled since the agreement's introduction.

The audit also provided information about subsidy level by transit mode. Subsidy levels can be calculated for just operating costs or for total costs. UTA and other transit agencies focus on operating costs which are important and must be used to account for the much heavier capital investment required by rail systems compared to buses. Subsidy rates vary among the three modes. Light rail requires the least subsidy, whether only operating cost or total costs are included, while commuter rail is the most subsidized. Figure 52 shows the changes overtime among the bus, light and commuter rail (Legislative Auditor General, 2014).

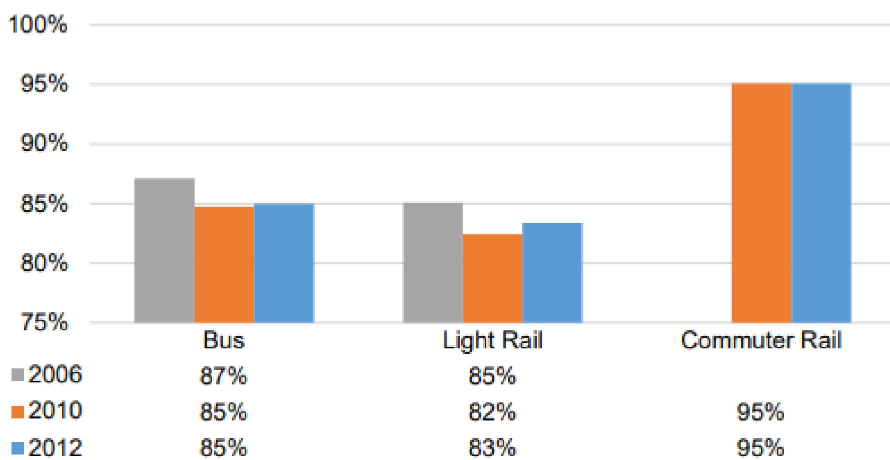


Figure 52. Tax Subsidies of Total Costs by Mode

Although commuter rail is more highly subsidized than other modes; it tends to carry more affluent riders. According to 2011 UTA survey data, most FrontRunner riders earn more than $50,000 per year, whereas the majority of bus riders earn less than $25,000 per year (Legislative Auditor General, 2014).

The board should review the effect of fare policies. Given the large taxpayer subsidy, it seems appropriate for the board to periodically review the overall effect of fare policy on different types of customers. The market approach used tends to favor large groups and choice riders; individuals and transit-dependent riders are less able to command discounts (Office of the Legislative Auditor General, 2014).

Prior audits have highlighted some concerns with the UTA passenger data. UTA did made improvements on their EFC system, but still need to develop a better method to analyze ridership from the collected data. Besides this situation, inconsistencies and limitations make difficult to obtain an accurate analysis. For this reason, the board should use its internal auditor to help implement audit recommendations and improve data practices (Legislative Auditor General, 2014).

The EFC system will help to enhance UTA services by analyzing the ridership patterns. However, after investing $19 million, UTA's EFC technology as currently implemented is still insufficient to monitor ridership, particularly the ridership responsible for the majority of UTA revenues. Figure 30 shows just over half of UTA's riders are not tracked using the EFC system. While the other half of UTA customers is tracked, EFC non-compliance, both at the beginning and the end of customers' trips, reduces the amount of complete electronic customer data to just over one-third of UTA boardings. The EFC system has made improvements in identifying customer behavior but lacks full implementation and compliance (Legislative Auditor General, 2014).

One issue that UTA is facing regarding the collected information is that the departments do not generate and store their data in a uniformly accessible manner across all UTA departments. The information is collected for various purposes; however, there is not any guide on how to categorize and process the data. It was also found that that there is not a central client identification system for organizations with EFC passes which could lead to confusion due to one client could have more than one identification. These data inconsistencies limited the ability to verify client revenues and validate client contracts, lengthening the audit process (Legislative Auditor General, 2014).

The primary goal of the audit report is to emphasize the problems of the UTA to design a new plan to focus on the customer. Two of the measures that should be included in the board performance plan are (Legislative Auditor General, 2014):

- Customer Experience: Board performance metrics are primarily financial; while that data is important, the board should also ask for metrics dealing with customer experience and satisfaction.
- Market Share: Currently, UTA tracks boardings (calling it ridership); while the tracking of boardings is important and used industrywide, it is limited and influenced by the transfer rate. Market share tracks the percentage of travelers using transit and is not altered by changes in transfer rates.

The study reports that UTA dashboard is more focused un any finance related measure than customer satisfaction and system monitoring. For that reason, the Board developed a strategic plan oriented to improve customer service and increase ridership by consistently measure and report customer satisfaction and provide a regular and consistent transit market-share information to the Board, respectively (Legislative Auditor General, 2014).

Four important recommendations were provided to the UTA Board of Trustees based on the audit results to improve transit service like (Legislative Auditor General, 2014):

1. Periodically review fare policy implementation. The review should include analyzing taxpayer subsidies provided to different customer groups and service modes as well as integrating public and stakeholder feedback.
2. Improve data practices by making better use of its internal auditor to periodically review and validate information it receives.
3. Direct UTA staff to provide them with regular and consistent customer feedback metrics.
4. Direct UTA staff to begin providing them with regular and consistent transit market-share information.

3.3.2 Transit Applications

3.3.2.1 Improving Efficiency and Reliability of Bus Rapid Transit

Significant changes are being made to the transportation system to enhance it for future years. As it is predicted, the population will grow, and the necessity for transit modes will be expected to increase. For that reason, measurements need to be taken to cover the needs and demands of the passengers. The Utah Transit Authority (UTA) which is the primary public transit provider in Salt Lake City is evaluating the possibility of implementing new strategies to improve their service. In order to accomplish this objective, Bus Rapid Transit (BRT) is being analyzed due to being an innovative, high-capacity, lower-cost public transit solution that can significantly improve mobility and efficiency of the transportation system. Two operational strategies for improving the efficiency and reliability of BRT system will be studied through this paper (Liu et al., 2018):

- Transit Signal Priority (TSP) benefits and impact based on GPS and by using a g microscopic simulation.
- Fare collection methods advantages by determining the contributing factors of dwell time (DT)

BRT is usually defined as an integrated system with a strong, transit-oriented identity, which consists of running ways (very often exclusive lanes), especially designed rail-like stations, high-capacity low-floor vehicles, improved services, and state-of-the-art Intelligent Transportation Systems (ITS). This type of transit mode is designed to improve the efficiency of the transportation system by increasing its reliability and ridership (Liu et al., 2018).

The chosen case study was the Utah Transit Authority (UTA) in Salt Lake, Utah. The UTA's system consists of a commuter rail (FrontRunner), light rail transit (LRT - TRAX), bus rapid transit (BRT - MAX), streetcar, bus (local, express, special purpose), and paratransit modes. Due to being a low cost, reliable, and comfortable transit mode, the BRT is gaining excellent reviews from its riders. As of today, there is one BRT deployment, the 35MAX BRT route, but more BRT lines are evaluated to be implemented in the coming years (Liu et al., 2018).

As stated in *Improving Efficiency and Reliability of Bus Rapid Transit* report (Liu et al., 2018), TSP is an operational strategy that facilitates the movement of in-service transit vehicles through intersections controlled by traffic signals. Meaning the green light will be extended or the red light will be reduced to improve the time cycle of a traffic light to enhance the travel times duration and schedule adherence. With the use of the GPS, it is considered that the TSP will impact positively on traffic operations although it is still a new technology to be evaluated in depth.

Bus operating time will affect considerable the efficiency of the transit service, and it is composed of two main components: time spent between stops (running time) and time spent at stops [dwell time (DT)]. Understanding the variability of these components, will provide a better insight to transit planners to design a more effective system (Liu et al., 2018).

DT is defined as the sum of passenger service time, boarding lost time, and door opening and closing time. Passenger service time is considered the main component affecting the DT because it can be influenced by passenger demand, fare payments, vehicle configuration, passenger load, door usage, platform configuration, atypical passenger boarding, passenger age, time of the day, and fare payment issues. To develop a DT model, researchers are using the data collected by automatic systems such as automatic passenger counter (APC), automatic vehicle location (AVL), and automatic fare counting (AFC) because of their accuracy and datasets sizes. The classic linear regression model with ordinary least squares (OLS) estimation has been widely used as the means for modeling bus DT. In multivariate DT models, several different factors are considered, such as fare payment methods, crowding effect, and passenger population, and it is expressed with the formula (Liu et al., 2018):

$$DT = \sum_{i=0}^{N} \alpha_i * B_i + \sum_{j=1}^{2} \beta_j * A_j + C$$

Where

- α_i is the average boarding time per passenger using the i^{th} fare payment method.
- B_i is the corresponding number of boarding passengers.
- N is the number of available fare payment methods.
- β_j is the average alighting time per passenger using the j^{th} door.
- A_j is the number of alighting passengers through the j^{th} door.

- C is time spent for door opening and closing (dead-time).

Various scenarios were studied using the DT model showing that different payment structures generated changes in DT, so developing strategies to improve the fare payments, will positively impact the DT magnitudes and variation.

The 35M MAX BRT (Figure 31) was the first BRT line in Salt Lake City, with UTA starting its operation in 2008. The bus runs a 10.8-mile distance on the 3300 S/3500 S corridor connecting the suburban town of Magna, Utah, and the light rail station at Millcreek. Because of the good result the BRT is showing, several lines are planned to be implemented by 2030. The 35M BRT uses low-floor, three-door buses that are 40 feet in length. The buses have 28 seats and 32 standees, for a total capacity of 60 passengers. The 35M BRT buses run on 15-minute headways on weekdays and 30-minute headways every Saturday. The buses use, on average, about 45% of their capacity during peak hours. Several fare payment options are available, including onboard cash payments with exact change into the fare box, electronic fare payments with a smart card, prepaid tickets purchased from a ticket vending machine (TVM), and transfer tickets. The 35M BRT drivers are instructed to use all doors for boarding and alighting, and no fare inspection is required. All three doors are equipped with smart card readers (tap on and off), and only passengers who wish to use onboard cash payments are required to board at the front door. A TVM is located at every BRT station, and no fare validation on boarding is necessary. The direct benefit of TVMs is the reduction in DT and operational (Liu et al., 2018).

The collected data used to model the DT and the fare payment structure quantitatively was from May 2014. The APC records included a total of 34,937 observations for 28 stops and provided information on travel direction, station ID and location, bus departure and arrival time, DT, number of boarding and alighting passengers, and station spacing at every station. AFC records included a total of 24,121 observations, with each entry representing individual passenger tap-on (boarding) or tap-off (alighting) at a specific date, time, and station. AFC and APC data were post-processed for matching based on the following criteria (Liu et al., 2018):

a) AFC and APC records should have the same date.
b) the same station ID
c) the time stamp difference between the matching records is less than two minutes (to accommodate any measurement error).

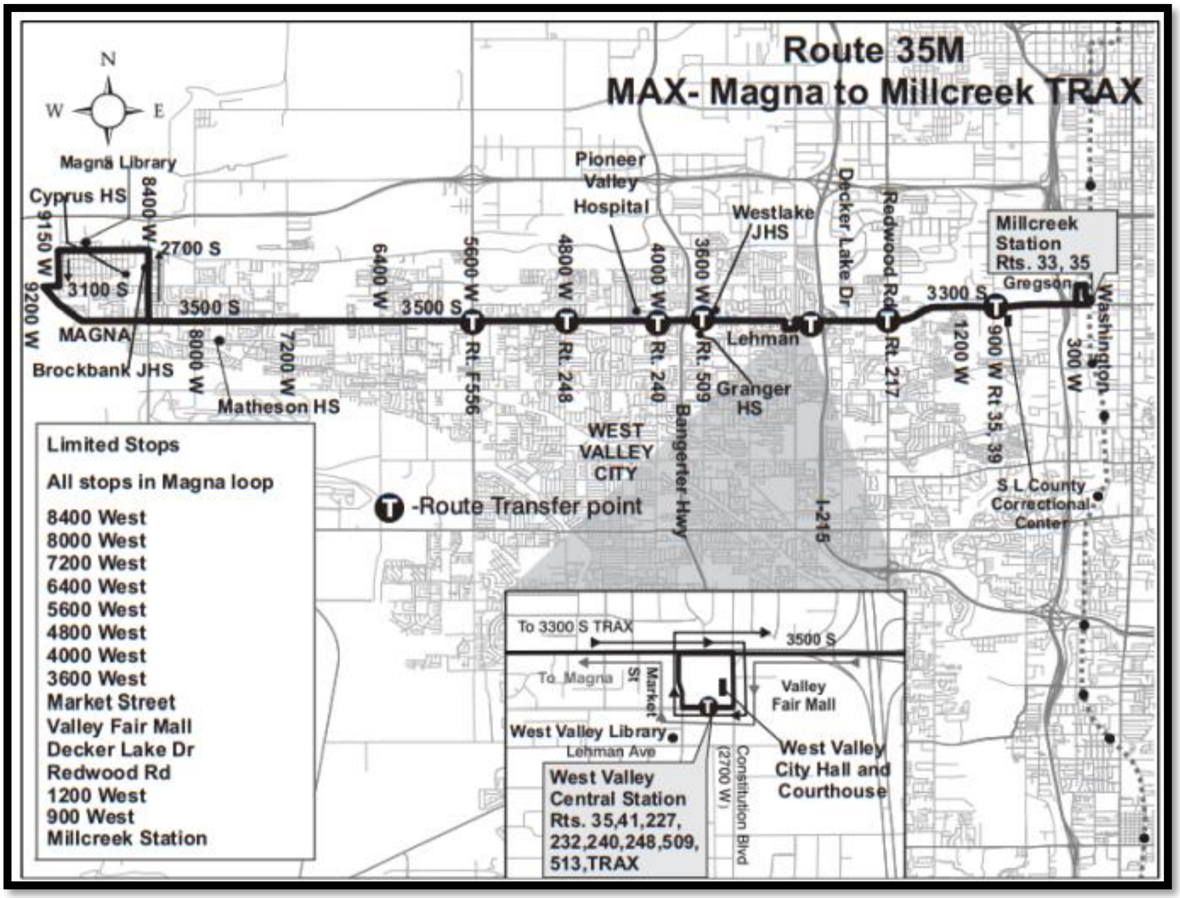


Figure 53. BRT 35M Route Layout

Genetic Algorithm (GA) for Determining Behavior-Controlled DT

Bus DT is generally influenced by the number of passengers boarding, alighting, or atypical passenger activities. The consensus of DT modeling distinguishes between sequential and simultaneous boarding and alighting. Sequential model assumes passenger activities (boarding and alighting) occur subsequently, and the DT is modeled as (Liu et al., 2018):

$$DT = \sum_{j=1}^{A} t_j^a + \sum_{j=1}^{B} t_j^b + deadtime$$

Where

- A and B represent the number of alighting and boarding passengers, respectively.
- t_j^a and t_j^b are the times that passenger *j* takes to alight or board.
- dead-time is the time needed to open and close the doors.

In case of simultaneous boarding and alighting, where boarding and alighting occur at the same time, DT is modeled as (Liu et al., 2018):

$$DT = \max\left\{\sum_{j=1}^{A} t_j^a, \sum_{j=1}^{B} t_j^b, Atypical\right\} + deadtime$$

This model indicated the collected data needed to be divided into three separate classes: boarding controlled (BC), alighting controlled (AC), and atypical situations. GA uses a random number generator to populate alternatives at each iteration and the process continues until an optimal solution is reached (Liu et al., 2018).

DT Modeling and Fare Payment Structure Analysis

Atypical scenarios were excluded from the DT modeling due to possible irreproducible features that it might have. Therefore, only BC-related d APC observations were used in the analysis and modeling of DT to obtain accurate results. Multivariate regression was adopted for DT modeling in this study. DT was represented as a linear collective function of independent variables, expressed as (Liu et al., 2018):

$$DT_i = \beta * X_i + \varepsilon_i$$

where

- DT_i is DT for the i^{th} observation,
- β is a vector of estimable coefficients, associated with each righthand-side variable,
- X_i is the vector of measurable characteristics that determine DT for the i^{th} observation,
- ε_i is the disturbance term.

From the DT models, the following interpretation were made (Liu et al., 2018):

<u>Boarding</u>

The average estimated boarding time for EFC users was around 5.2 seconds/passenger that was much longer than the suggested time (2.75 seconds). Two possible reasons will be related with this difference: the tap on/tap off EFC reader on the UTA fleet has slower refresh rates compared with the common smart card reader systems used in the U.S.; and a significant portion of EFC users delayed their boarding process by searching for the card.

The average boarding time for passengers who use prepaid passes, transfer tickets, or fare evasion was about 1.8 seconds/passenger. While the average boarding time for passengers who paid their fare by cash was about 6.9 seconds/passenger. This difference could be due to passengers not having the cash ready before boarding the bus.

Alighting

The average alighting time for EFC users and non-EFC users was 2.0 seconds/passenger, and 1.6 seconds/passenger, respectively.

Stop Characteristics

Studying the stop characteristics like stop placement, design, and built environment it is critical because it negatively impact the DT. For example, the stops at malls, near hospital and transfer stops take more time than a single stop along the route. This could be due to the passenger's characteristics from those places, but it is important to know which factors influence either positively or negatively the DT.

Dead-Time

The door is open/closed at least once at each stop (that has a DT). Thus, the minimum estimated dead time was 3.5 seconds.

Miscellaneous Factors

Miscellaneous variables that could affect DT, such as time-of-day, day-of-week, and crowding effects, were also explored. Day-of-week yielded statistically insignificant effects on DT according to the model; however, it does predict that weekends are prone to have longer expected DTs.

Through this study, it could be appreciated the effectiveness and impacts of GPS-based TSP strategies with mixed-traffic BRT upgrades, and they are summarized below (Liu et al., 2018):

- GPS-based TSP strategies can provide transit delay reductions and travel time savings as effective as those of traditional TSP tools with fixed or distance-limited detection. However, the GPS-based TSP strategy is more advanced because of its flexibility in setting and adjusting detection–activation distances, its extensible features, and its relatively lower equipment costs (when most of the transit vehicles have been equipped with GPS).

- In a mixed-traffic BRT system (optimization of bus stop locations and provision for queue jump lanes) with unconditional GPS-based TSP strategy, CTSP strategy considering bus occupancy, and CTSP2 strategy considering bus occupancy and schedule adherence, the total reduction in peak hour transit delay can be, respectively, 13%, 13%, and 3% compared with BRT alone; and the total savings in peak hour transit travel time in those same three scenarios can be, respectively, up to 9%, 7%, and 3% compared with BRT alone.

- The average non-transit travel time along the study corridor had no significant differences between a base scenario (existing conditions), a scenario with TSP strategies added, and a scenario with BRT upgrades made.

- In a regular bus system, with traditional TSP and GPS-based TSP strategies, the total delays for peak hour side-street traffic increased by 6% compared with existing conditions.

- In a mixed-traffic BRT system (optimization of bus stop locations and provision for queue jump lanes), with strategies of unconditional GPS-based TSP, CTSP considering bus occupancy, and CTSP2 considering bus occupancy and schedule adherence, the total delays for peak hour side-street traffic increased, respectively, by 3%, 2%, and 1%, compared with BRT alone.

- TSP strategy and BRT can increase the average queue lengths on critical side streets, but in general, the impact on queue lengths was minor. No significant impacts on the average number of stops on side streets were seen.

Besides the results from the development of these models, it is important to have in consideration that they were built using specific conditions of the transit route in Salt Lake County, therefore the results obtained cannot cover the situation in other states. Geographical features and technology capabilities should be fully considered when GPS-based systems are being implemented and operated due to the probability of being inaccurate in densely built environments because of the urban-canyon effect (Liu et al., 2018).

Understanding the factors affecting the DT it is important when modeling an effective approach of the transit system. These factors can be supported with the availability of APC data in most transit systems. Even though the APC and AFC datasets provide a substantial sample of information, it does not include the fare transactions that do not have electronic footage, which still account for a large portion of the fare payment structure in most transit systems. This analysis results suggest considering the DT as an optimization problem to resolve the gap in fare payment structure estimation (Liu et al., 2018).

The DT model could be adapted to any transit system with access to APC/AFC database to identify certain conditions influencing the DT extent. Even though this model provides valuable insights about the DT variations, it could still be enhanced to provide guidelines on fare evasion estimation, TVM cost-benefit analysis, and instructional guidance to facilitate a smooth boarding/alighting process leading to improve the transit efficiency, DT magnitude and be potentially useful to future BRT projects (Liu et al., 2018).

3.3.2.2 Fare Payment Structure and Dwell Time

Transit Capacity and Quality of Service Manual (TCQSM) defines dwell time (DT) as the sum of passenger service time, boarding lost time, and door opening and closing time. DT is an important bus operational component meaning that any variability can alter the quality and functionality of the transit service. As stated in *Fare Payment Structure and Dwell Time* report (Fayyaz & Liu, 2015), this paper studies a method to determine the impact of detailed fare payment installed on Utah Transit Authority BRT line 35 M.

To complete the analysis, data needs to be gathered from the automated fare collection (AFC) systems. It is important to know that as per ridership increases, passenger service time will increase leading to increase in DT, bus run time variation, and decreasing operation performance. Transit planners needs to have knowledge of how these factors influence the DT to have a more accurate planning process (Fayyaz & Liu, 2015).

Passenger service time is understood to be the largest component of DT and influenced by passenger demand, fare payments, vehicle configuration, passenger load, door usage and platform configuration. There are also other factors that can influence the effectiveness of the DT such as: fare type, bus design, stop design, atypical passenger boarding, passenger age, time of the day and fare payment issues (Fayyaz & Liu, 2015).

Prior researched on DT were conducted using manually collected data, but the labor and time were intensive, and the sample is small. However, using automatically collected data, some information could be lost for non-electronic fare payment methods like cash payment and prepaid paper tickets, but the sample is considerable with no extra cost if the automated systems were already installed on the vehicles (Fayyaz & Liu, 2015).

For this study, the data used was collected automatically through the Automatic Passenger Counter (APC) and AFC systems installed on UTA buses for bus rapid transit (BRT) line 35M, and manually for limited time. Automatic observations were collected from UTA bus route 35M during the entire month of May 2014 with a total of 65536 observations for 28 stops. AFC and APC data were then matched based on the following conditions (Fayyaz & Liu, 2015):

a) AFC and APC records have the same date.
b) They use the same station.
c) The time difference between records is less than 2 minutes (to consider any measurement error).

If there were two matched lines, the AFC record was added to the APC location. Manual collected data consist of 120 observations during 7 to 9 AM, 11 AM to 1 PM and 4 to 6 PM. Three categories were taken into consideration of the main factors affecting the DT: number of passenger boarding (BC), number of passenger alighting (AC), or atypical activity, and it was determined that three models were needed to have a better understanding of the impacts that could have on the DT. There were also constrains used when preliminary filtration was applying to the collected information like Dwell time more than 3 minutes or Dwell time divided by the number of boarding passengers be less than 1 second. Terminal observations were excluded because it included the drivers' layover time (Fayyaz & Liu, 2015).

BRT 35M buses have low floors, are equipped with smart card readers at all doors, have 28 seats (total capacity is 60 passengers), and have three doors. All three doors are available for both boarding and alighting, but passengers who want to buy their ticket from the driver (exact change into the fare box) are limited to board from the front door. The summary of statistics of the data are presented in Table 9 (Fayyaz & Liu, 2015).

Table 9. Summary Statistics

Variable	Obs.	Mean	Std. Deviation	Min	Max	Sum
DT	7725	17.055	13.270	4.8	169.8	-
Weekend	7725	0.084	0.277	0	1	-
B-EFC	7725	0.263	0.610	0	8	2033
B-CTVM	7725	2.689	2.282	0	18	20770
B-TVM[a]	7725	1.743	2.163	0	18	13467
B-Cash[a]	7725	0.935	1.605	0	12	7225
A-EFC	7725	0.051	0.235	0	3	396
A-CTVM	7725	0.917	1.443	0	16	7087
Door-Cycle	7725	1.288	0.496	0	5	-
Fair-Mall stop (Magna direction)	7725	0.083	0.276	0	1	-
3575 W stop	7725	0.035	0.185	0	1	-
3955 W stop	7725	0.059	0.235	0	1	-
Fair-Mall stop (TRAX direction)	7725	0.035	0.183	0	1	-
1685 W stop	7725	0.049	0.215	0	1	-

[a] B-Cash and B-TVM's summary statistics are based on estimated results of model estimation and testing section.

Variables

- DT: dependent variable that measures the time (in seconds) between door open and close.
- BT: Total number of boarding passengers.
- AT: Total number of alighting passengers.
- B-EFC: independent variable for the number of passengers boarding using electronic fare payment.
- A-EFC: independent variable for the number of passengers alighting using electronic fare payment.
- B-TVM*: variable for the number of passengers boarding using tickets (bought from ticket vending machines) or passengers who do not pay. Note that the boarding and alighting time for the TVM and NonPayers are almost the same as there is no visual inspection.
- B-Cash*: variable for the number of passengers boarding who buy ticket from bus driver. Note that cash payers are limited to board from the front door.
- B-CTVM: Sum of B-TVM and B-Cash within the dwell time observation. This variable is equal to BT – BEFC.
- A-CTVM: independent variable for the number of passengers alighting who paid fare by TVM or Cash. There is no inspection or transaction related activity associated with Cash and TVM payers while alighting. This variable is equal to AUT – A-EFC.

To find the best model specification, independent variable should be included as many as possible so that λ's value become closer to 1. Table 10 shows the result for the B-CTVM Model and B-TVM & B-Cash Model (Fayyaz & Liu, 2015)

Table 10. Model Results

Model	B-CTVM Model		B-TVM & B-Cash Model	
DT	Coefficient	t-stat	Coefficient	t-stat
Weekend	1.411	4.04	1.087	6.30
B-EFC	4.992	30.64	5.279	65.57
B-CTVM	3.329	71.46		
B-TVM			1.803	73.03
B-Cash			6.917	211.68
A-EFC	2.623	6.28	2.020	9.79
A-CTVM	1.741	23.14	1.611	43.40
Door-Cycle	1.580	8.06	1.509	15.60
Fair-Mall stop indicator (Magna dir.)	2.478	6.36	2.116	11.00
3575 W stop indicator	-2.598	-4.95	-2.588	-9.98
3955 W stop indicator	2.222	5.38	1.617	7.92
Fair-Mall stop indicator (Trax dir)	3.766	6.72	3.432	12.40
1685 W stop indicator	2.479	5.44	2.287	10.15
Constant	2.411	8.32	2.026	14.15
Adjusted R-Squared =	0.5937		0.9009	

Variables

- Weekend: indicator variable that shows whether the observation is collected on weekend (equal 1) or on weekdays (equal to zero).
- Door-Cycle: variable that shows how many time bus doors were opened and closed in the observation.
- Stop-[station name]: indicator variables show the station name that the observation was collected.

To obtain better results, three recommendations are suggested when validating the results (Fayyaz & Liu, 2015):

- First test checks the consistency of coefficients of common variables in two models,
- Second test check the estimated B-Cash and B-CTVM with the manually collected data,
- Third test check the effect of estimation error in coefficients.

Overall, the B-TVM & B-Cash Model shows a good statistical fit with an adjusted R-squared of 0.90. It could be appreciated the weekend results on DT are a second larger than the weekdays. This could be due to passengers are not in hurry during the weekend when they are taking the bust resulting in spending more time when board or alight. The model predicts 1.5 seconds for each time that a bus door-cycle. Dead-time can be estimated as sum of the constant term and door-cycle time, so $Dead\ time = 1.5 * doorcycle + 2\ seconds$ (Fayyaz & Liu, 2015).

Average boarding time for passengers who use on-board fare collection methods [i.e., electronic fee collection (EFC) and Cash] are much larger than off-board fare collection method (i.e., TVM). It was estimated the average boarding time for EFC, Cash and TVM could be 5.3, 6.9, and 1.8 seconds, respectively. This situation is due to card readers installed on busses take about 1 second to read the card with respect to EFC and Cash customers which take long to process those transactions (Fayyaz & Liu, 2015).

Average alighting time for TVM and Cash payers are the same as they do not have any other transaction when leave the bus. However, EFC payers have check-out transaction time. Average alighting time for passenger who use TVM or Cash is estimated to be about 1.6 seconds, and 2 seconds for EFC users (Fayyaz & Liu, 2015).

It was also observed that for Fair-mall stop is estimated to be on average 2.1 and 3.4 seconds (for each direction) longer than other stops. This situation could be cause due to passengers carrying shopping bags when boarding the bus incrementing the boarding time. The same situation could happen with 3955 W and 1685 W stops which have an average time of 2.2 and 2.5 seconds (respectively) longer than other stops. 3955 W is placed in front of the hospital, so longer boarding times are expected due to sick passengers. 1685 W is a transfer stop between bus route 35M and bus route 217 leading to longer boarding time due to drivers waiting for passengers to complete their transfers (Fayyaz & Liu, 2015).

A new method was developed to estimate the number of passengers boarding that have no AFC record associated with them and use different fare payment method. The advantage of this method is that it required automated data collection which is collected through APC and AFC systems and used a large sample dataset. The model result can be used to predict dwell times for future year with increased ridership. For future research, we recommend to further explore the impact of built environment, socioeconomic parameters of passenger, and stop design model on bus dwell time (Fayyaz & Liu, 2015).

3.3.2.3 Assessing Social Equity

Evaluating the social equity by transit planners when designing a new route, it is required by federal law. Before any changes in the fare system, this inequality needs to be analyzed to understand the impacts it could have when implementing a new fare payment system. Some of the categories to have in consideration when conduction the evaluation is the race, color, or national origin. There is a lack of guidelines of how to address such discrimination in the studied area, therefore, the goal of this study is to develop and apply a new method for assessing social equity impacts of distance-based public transit fares (Farber et al., 2014).

Due to having few tools for transit planners to study this criteria, social equity has been underrepresented in transportation planning research and practice. And this is an important situation to be evaluated to have a balance with fairness like the one between the environmental and economic concerns. As stated in *Assessing social equity in distance based transit fares using* *a model of travel behavior* study, this research intends to address this gap by developing a Geographic Information System-based Decision Support System (GIS-DSS) for evaluating the social equity impacts of transit fare policy. GIS-DSSs assist users with evaluating the costs and benefits of hypothetical solutions to inherently spatial problems and they are recommended by the Federal Transit Administration for analyzing the equity impacts of proposed changes in transit route and fare structures.

The case study chosen to implement this method was the Utah Transit Authority (UTA) which is a large state-authorized transit operator in the United States, serving to a population of 1.8 million people with a fleet of busses, vanpools, and light and commuter rail locomotives. The UTA charges a local-service flat-rate fare for one- and two-way trips regardless of distance travelled on the transit system, but it is considering a distance-based fare structure to increase their ridership (for shorter distance travelers) and greater levels of fare-box revenue overall (Farber et al., 2014).

The data used was from the Utah Household Travel Survey (UHTS) conducted in the spring of 2012. The one-day trip survey recorded 101,404 trips taken by 27,046 individuals living in 9155 households across the state of Utah. This provides a spatially dense sample of 6238 households and 16,071 individuals considered to be within the operating district of the transit authority. The collected information about individual trips, the trip-makers, and households is portrayed in Table 11 showing social inequities in transit trip generation and distance travelled in the Wasatch Front (Farber et al., 2014).

From Table 11, it can be concluded that distance-based fares benefit the lower socioeconomic passengers because of the number of short distances trips they made. For example, Hispanics and non-white respondents are 70% more likely to ride transit in comparison to non-Hispanics and whites, but whites travelled about 45% greater distances on transit than their non-white counterparts. Students who are employed for more than 25 hours per week have the highest penetration rates of transit use, but fully employed transit riders travelled 40–65% longer distances compared to others (Farber et al., 2014).

The analysis before shows a common pattern among different socioeconomics groups. But to implement a model that highlights the impact of each dimension, various variables need to be declared. Probability functions for each dependent variable and a correlation structure between the random components affecting each decision process will be accounted for the derivation of an ordinal/continuous model that will be used to simultaneously estimate the joint decision of how many transit trips to take (discrete-ordered), and how far to travel by transit in a day (continuous). Relevant factors related with the social equality were taken into consideration such as: low income, elderly, low education, employment status, race, and others using distance to the Central Business District (CBD) and a polynomial function of the household spatial coordinates and a set of recent transportation studies (Farber et al., 2014).

Table 11. Transit Ridership, Trip, Distance Travelled for Household Characteristics.

	Ridership percentage	Trips	Distance travelled (miles)
Household income < 0.014,0.002>			
No answer	1.91	1.73	20.09
Under $35,000	5.17	1.94	13.25
$35,000–$49,999	2.94	1.88	19.14
$50,000–$99,999	2.24	1.85	20.56
$100,000 or more	2.17	1.86	24.51
Hispanic < 0.038,0.348>			
Yes	4.38	1.89	17.39
No	2.60	1.88	19.60
Prefer not to answer	3.02	1.64	5.34
Race < 0.000,0.018>			
White or Caucasian	2.50	1.89	19.93
All other	4.61	1.77	14.25
Age < 0.110,0.944>			
18–24 years old	6.87	1.82	17.47
25–34 years old	4.07	1.88	17.51
35–44 years old	3.37	1.84	24.19
45–54 years old	3.49	1.87	19.73
55–64 years old	3.13	1.88	17.82
>65 years old	1.59	2.08	13.91
Employment < 0.000,0.000>			
Employed full-time	4.58	1.84	22.43
Employed part-time	3.05	1.84	13.61
Student, not employed or employed less than 25 h/week	6.88	1.76	15.96
Student, employed 25+ h/week	10.79	1.91	14.61
All other	1.36	2.00	13.88
Education < 0.017,0.005>			
High school or less	3.73	1.85	13.08
Some college/vocational/associates	3.40	1.87	18.56
Bachelors	2.90	1.80	19.34
Grad./post grad.	5.26	1.95	21.88
Licensed < 0.000,0.000>			
Yes	3.14	1.86	20.39
No	13.67	1.96	11.34
Limited mobility < 0.012,0.290>			
Yes	7.05	2.09	14.99
No	3.51	1.86	19.16
Prefer not to answer	6.90	2.00	20.75
Number of vehicles < 0.000,0.000>			
Zero vehicle household	27.34	2.26	7.42
1 vehicle household	5.40	1.78	14.04
2 vehicle household	1.87	1.89	23.70
3+ vehicle household	1.99	1.81	23.21
Home ownership < 0.000,0.000>			
Rent	5.56	1.86	11.66
Own	2.17	1.87	22.77
Other	0.77	2.00	17.61
Number of years (Residence)<0.158,0.000>			
Less than 1 year	4.44	1.97	16.23
1–5 years	2.58	1.88	19.21
More than 5 years	2.42	1.83	19.96
Place type (Self-reported)<0.000,0.000>			
City, downtown with a mix of offices, apartments and shops	7.44	1.72	7.41
City, residential neighborhood	2.82	1.94	13.72
Suburban neighborhood, with a mix of houses, shops and businesses	3.15	1.91	22.28
Suburban neighborhood, with houses only	2.09	1.83	22.95
Other	1.61	1.78	34.41
Residence type < 0.000,0.000>			
Single-family house (detached house)	2.17	1.88	22.82
Building with 4 or more apartments or condos	6.17	1.92	11.38
Other	3.34	1.74	13.89

From the obtained results when the ordinal/continuous model was implemented, it was shown that factors associated with taking more trips include (Farber et al., 2014):

- being 18–24 years old
- living in a household with retirees for household heads
- being a student that is employed for more than 25 hours per week
- being highly educated
- living in a zero-vehicle household
- living in larger households
- living in a suburban neighborhood that maintains a high mix of land uses

And factors associated with taking lees trips include:

- being younger than 18
- being self-employed,
- being female
- living in households with 2 or more vehicles
- living in households that are neither rented nor owned.
- living in households with many children's bicycles
- living in households that are farther away from long-distance commuter rail

Many coefficients were insignificant on their own, but became significant in the presence of one of the spatial expansion terms such as:

- being aged older than 65
- having a mobility limitation
- being unemployed or retired
- being Hispanic
- not having a driver's license
- being low-income and living in a rented house

Understanding the impact of these factors on trip generations and distance traveled will improve to have a better equity of distance-based fares approach. The model can also be used to the expected change in fares paid by individuals of a specific demographic profile, differentiated by residential location (Farber et al., 2014).

The development of this model is addressed to improve the fare prices of the public transit based on the distance traveled benefiting the low socioeconomic groups which it is the most dependent of these type of transportation modes. For this reason, it is important to conduct a transit study of different low-income neighborhoods to have a better insight of their situation and how the new fare system based on traveled distance will impact (Farber et al., 2014).

However, the research has some limitations listed below along with the limitation of implementing the model to other areas with different demographics (Farber et al., 2014).

- First, the Utah Household Travel Survey provides only a limited one-day snapshot of travel for each responding household.
- Second, the trip records are geocoded at trip-ends, but detailed routing information is not captured by the web-based survey instrument. This means that actual distances travelled along the transit network are not known and must be imputed based on shortest paths or otherwise approximated using common-sense.
- Third, the trip generation and distance models were calibrated without a price variable, implicitly assuming that travelers are not sensitive to price.
- Fourth, only a single type of distance-based fare structure in this research was explored, one that begins with a fixed cost per trip and increases linearly with distance.
- Firth, a follow-up social equity analysis of distance-based fares would benefit from the incorporation of estimated and projected population characteristics.

The research aims to improve the social equity and increase public ridership to decrease gas emission provoked by the big number of single passenger vehicles on the road. The developed model is a new methodology that focuses on the impact of distance-based fares in a determined socioeconomic group (Farber et al., 2014).

3.3.2.4 Effects of Transit Operations on Emissions

The environment is affected by many modes of pollution, but one of the most significant sources is the transportation. For this reason, it is important to understand how it is affected and analyze ways to improve this situation. Many of the transit agencies are updating their transportation system to electronic ones and creating methods to study the customers patterns behaviors to design and implement better transit services that complies with customers' demands and reduce gas emissions. The objective of this research is to analyze the public transit impact on mobility-related pollutant emissions, for an entire metropolitan region in northern Utah (Mendoza et al., 2019).

As stated in *Modeling net effects of transit operations on vehicle miles traveled, fuel consumption, carbon dioxide, and criteria air pollutant emissions in a mid-size US metro area: findings from Salt Lake City, UT publication* (Mendoza et al., 2019), 2016 EFC and APC data were analyzed along with service schedules and routes from General Transit Feed Specification (GTFS) data of the Utah Transit Authority (UTA) to estimate the pollution impact in the region by accounting for vehicle miles traveled, gasoline gallons equivalent of fuel consumed, and multiple pollutant species emitted. It was stated that buses, light rail and commuter rail contribute approximately 1.5% of the road emissions. However there has been significant reduction of nitrogen oxides (NO_x), fine particulate matter ($PM_{2.5}$), and sulfur oxides (SO_x) emissions when the bus fleet upgraded to 2010 model and newer, diesel and compressed natural gas (CNG) buses, as well as modeling an envisioned change to Tier 3 locomotives for the commuter rail system.

The evolution of the transportation technology has captured the interest in analyzing large-scale properties of human mobility to examine how the pollution it is affected due to people mobility modes. Since the research depends on the existence of large volumes of spatial and temporal data, some systems were believed that could provide valuable information about the passengers such as (Mendoza et al., 2019):

- Electronic fare collection (EFC) systems which generate fine-grained data on transit ridership that open possibilities for analyzing transit-rider travel behavior in new ways.
- Automated passenger counter (APC) devices that use infrared lights on vehicle doorways to track passenger boarding and alighting.
- General Transit Feed Specification (GTFS) which is an open data specification for representing movement in time and space of scheduled transit services for a given transit operator.

Although the spatial granularity of econometric analyses is limited, the high spatial and temporal resolution used, in conjunction with transit service disaggregation, enables a more detailed analysis at the sub city level. The capacity of econometric models to extrapolate from samples to populations is linked to the coarse resolution of these studies, where space, time, and individuals are combined to produce aggregate results. In addition, the increased resolution helps the quantification of the air quality and pollutant exposure in denser urban areas where vehicle emissions are dominant (Mendoza et al., 2019).

Transportation is a daily activity performed by people. Nowadays exist several modes of transit which facilitate the movement of people from one place to another. Even though it provides enormous benefits to the population, it also creates environmental problems including congestion and degraded air quality. Public transit could help in reducing the air pollution and ease the traffic congestion but studying the transit impact on urban air quality is a complex topic due to being the result not only of the transit operations but also the unobserved activities generated by other modes (Mendoza et al., 2019).

The Utah Transit Authority (UTA) operates transit services along the Wasatch Front of northern Utah, a city with a current population of 1.8 M people that is growing rapidly and expected to reach a total population of 2.5 M within the next 30 years. UTA provides public transit throughout a 7-county area, most of which has significant air quality concerns and is qualified as areas of nonattainment and maintenance for multiple pollutants (Mendoza et al., 2019).

With the implementation of the new electronic technology like electronic fare collection (EFC) systems and automated passenger counter (APC) devices, an extensive study could be completed to understand the passenger travel behavior to direct resources and design better schedules and routes where it is needed. That could be one way to determine the transit impact to air pollutant by observing where there are more vehicles movement. Public transportation can help to improve the environmental issues by providing a method to mitigate not only air quality pollutants, but also CO_2 and other greenhouse gas emissions from personal vehicles (Mendoza et al., 2019).

The spatial ridership data provide by EFC systems contains only numeric identifiers for route traveled and origin/destination stations. To estimate avoided emission, the GTFS feeds was used to trace each EFC trip trajectory. GTFS feeds include route information including stop time and location and can be used to generate GIS features depicting the geometry of each transit route. The stop sequencing information in the GTFS data allows to look up an arbitrary trip origin stop on an arbitrary route and reconstruct the set of stops passed and the path through space traversed to the corresponding trip destination stop (Mendoza et al., 2019).

This report estimates both direct and indirect emissions by allocating the direct emissions spatially and temporally using the route information based on GTFS feeds. To perform the estimation, the UTA service was grouped into three classes according to emission profiles. Buses, Express Buses, Ski Bus, and Park City Bus were grouped as the Bus class, TRAX and Streetcar were grouped as Light Rail, and the FrontRunner constituted the Commuter Rail (Mendoza et al., 2019).

Three assumptions were made for estimating avoided emissions for avoided auto trips that would take place (Mendoza et al., 2019):

1. It was assumed each transit trip represents an avoided auto trip that would have been made in a single passenger vehicle had it not used a UTA service.
2. It was assumed that the avoided auto trip would have followed the same route as the UTA service trip observed in the EFC data.
3. As not all riders pay their fare via EFC card, we assume that the spatial and temporal distribution of trips by cash fare-payers (and passengers who avoid paying fares altogether) is proportionate with trips by EFC fare-payers.

The UTA's EFC system requires to tap on and off the rider's cars every time they board or alight from buses. This collected data allows the system to compute the trip length and charges the cardholder variably by length of trip; if no tap-off is recorded, the cardholder is charged for the maximum trip possible on that route. UTA also have the infrared automated passenger counter (APC) devices on the vehicles allowing to detect interruptions of an infrared beam across vehicle doorways, thus counting boardings and alightings at every stop without any input from riders. With the APC data, the passenger route could be reconstructed by analyzing every stop-to-stop segment, and from these to calculate total system-wide passenger miles travelled, but they do not support inference of the point-to-point path followed by any single rider (Mendoza et al., 2019).

- For bus services, the EFC card reader is on board the bus, thus allowing the EFC information system to record the card unique ID, the route, plus the unique stop ID and time of boarding or alighting.

– For commuter rail and light rail services the card reader is on a pedestal at the train station, and therefore records the location (via unique station ID) and time of boarding or alighting, but not(explicitly) the route used.

The analysis results were divided into four components (Mendoza et al., 2019):

1. Total emissions: This analysis focuses on a system-wide overview without any temporal or spatial disaggregation. Emissions from a newer, upgraded fleet ('Sensitivity Test') show that the greatest benefit would be in bus emissions, where emissions of most species (including $PM_{2.5}$ and NO_x) would be reduced by approximately 50%, while NMHC emissions would be reduced by over 75%. The slight increase in fuel consumption of the new fleet is due to larger numbers of CNG buses which are slightly less fuel-efficient than diesel. Figure 54 compares the avoided VMT, fuel consumption, and emissions from the use of UTA's services against the on road sector VMT, fuel consumption, and emissions for the seven-county UTA service region. Approximately 1.5% of total on road emissions for the service area are offset due to the use of transit. Within Utah's Wasatch Front, on road transportation account for between 25%-50% of pollutant emissions. Therefore, any reduction that can be achieved in this sector is significant in the overall emissions budget.

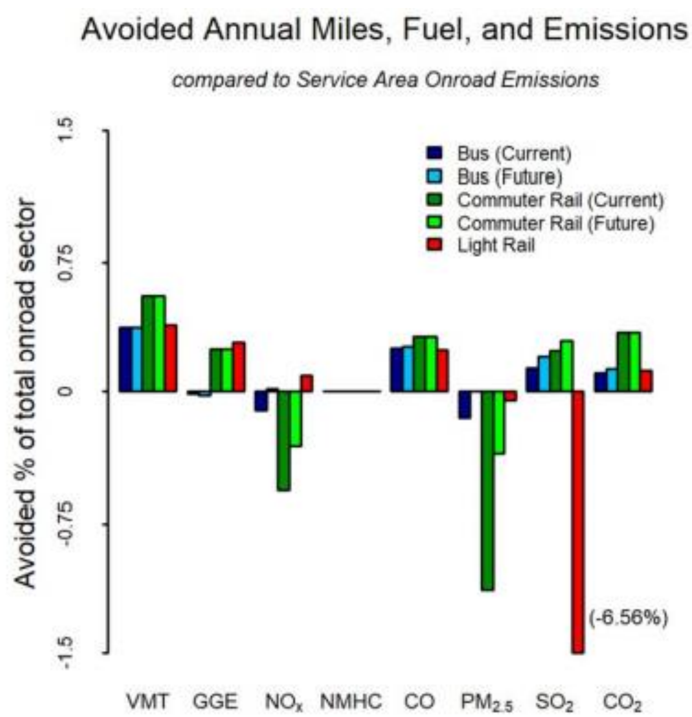


Figure 54. Percent Net Avoided VMT, Fuel, and Emissions

2. Spatial patterns: This analysis focuses on annual spatial data. The spatial pattern of average annual avoided travel for all services. The largest amount of avoided travel due to bus's

use is found in the most frequently traveled routes, which are generally located in the central parts of cities.

3. Temporal patterns: This analysis focuses on monthly and hourly data without incorporating spatial information. The light rail is considered as a net reducer of all emissions given that the induced emissions used to generate the electricity for its use are outside the airshed. Due to the significantly larger amount of $PM_{2.5}$ and NO_x that buses and commuter rail emit per vehicle when compared to cars, the effect of avoided trips is minimal.
4. Spatiotemporal patterns: This analysis explores both the space and time disaggregation, leveraging the power of the high spatiotemporal dimension available in our datasets and methodology.

This study provides an analysis of the transit system emissions and avoided emissions due to buses, light rail, and commuter rail use, allowing policy makers and transit planners to address unobservable impacts to net emissions stemming from service changes or ridership changes. Differences were found among the services provided by the UTA associated with avoided miles of travel, fuel consumption, and emissions. The morning and afternoon commute hours result in the largest amount of avoided travel and emissions. Net avoided benefits from the buses were stronger in the central parts of the cities and inter-city connections routes. The commuter rail had a similar effect throughout its route while the light rail ridership and offset emissions, was highest in the central trunk line and on the route connecting the University of Utah to downtown Salt Lake City. The light rail was a net emissions reducer within the UTA services airsheds. Upgrading buses to newer diesel and CNG would turn the bus system from a net contributor of additional NO_x and $PM_{2.5}$ emissions to a net reducer of these pollutants due to avoided car trips (Mendoza et al., 2019).

Even though this analysis forecast valuable information about how the public service could positively influence the environmental pollution, some limitations were found such as (Mendoza et al., 2019):

- All the trip segments were binned to the hour associated with the route's first stop, which may cause a small discrepancy between the times associated with each trip segment.
- Avoided auto trips are represented solely by EFC trips which only account for approximately half of UTA's total trips. This has the potential to bias overall trip volumes both spatially and temporally because the non-EFC passengers could have considerably different transit travel behavior compared to EFC users.
- Since avoided trips were spatially located following the same route as the transit equivalent of the EFC trip, our methodology has the potential to overestimate some avoided emissions since the most efficient or shortest route for an automobile traveler may not be the one taken by the bus.

There are important studies to complete about the public transportation impact to air quality. Public transit can help reduce the greenhouse effect, because it reduces the number of single occupancy vehicles on the road. Further, vehicles that are poorly maintained and in inappropriate conditions could contribute negatively to the atmospheric pollution. This study suggests for future research to investigate if the lack of people at waiting stops or walking could be induced due to the avoidance of exposure to certain degraded air quality environments, and also the impact of rider behavior to gasoline prices because during the recession of 2008, it was found a reduction in vehicle miles traveled by personal vehicles (Mendoza et al., 2019).

4 CONCLUSIONS AND RECOMMENDATION

Automated Fare Collection (AFC) systems, also referred to as Automatic Fare Collection or Electronic Fare Payment (EFP) systems, have been widely adopted by transit agencies around the world. They do not only provide time savings, convenience, and other advantages in revenue collection, but also produce massive, continuous, and anonymized digital records of fare transactions. The data presents unprecedented opportunities for transit planners and researchers for capturing and analyzing passenger travel patterns in addition to producing general ridership and revenue information. AFC fare media provide comprehensive and exhaustive data with precise spatial and temporal information than traditional household travel surveys or on-board surveys could ever provide.

AFC systems provide a bias-free and low-cost approach to obtaining long-term continuous observations of transit passenger travel behavior. This makes it possible for the discovery of regularities and dynamics in passenger travel and can assess users' response to changes in the transit system (e.g., fares, routes, and scheduling modifications). It can also help understand the effects of external factors such as demographic and social trends, technology, and mobility service innovations. However, despite the advantages of the data provided by AFC systems, it is not entirely clear how AFC data could be best utilized to efficiently support operational analysis, inform transit service planning and scheduling, and facilitate long-term planning for public transportation services.

4.1 Lessons Learned

In the previous sections, a comprehensive literature review and case studies of existing AFC systems were presented. The findings from these tasks present a wide variety of relevant topics. The topics identified in the Literature Review and Case Studies cover different areas of public transportation. They provide many examples on how transit agencies can benefit from the efficient use of AFC data to better understand the service needs of the transit users. In addition, combining AFC data with other transit ITS datasets can provide additional information that can be very useful for transit planning and operations and to assist with the overall decision-making process.

The following items highlight the key topics that were identified by the research team, as they relate to this project:

The Age of Big Data. Because of the increasing demand for high-quality data, agencies should have the necessary resources to perform data analysis and create visualizations of the data for a better communication and comprehension with the stakeholders regarding the transit system condition, emphasizing the significance of capital investments. The following are some of the tools that can be used to take advantage of the data and present the information to stakeholders:

- Business Intelligence and Analytics. Business Intelligence (BI) is the term applied to the ability of an organization to collect, maintain, and organize data. The BI technologies can provide historical, current, and predictive information on business operations by transforming raw data into meaningful and useful information, which can be used to inform more effective strategic, tactical, and operational insights and decision-making.

- Data Visualization. Visualizations could provide a better understanding of the collected data by using pictures, images, or animations depending on the best way to communicate with an intended audience.

Using Integrated Electronic Data for Service Planning. With the collection of data from transit ITS systems, transit agencies should have the necessary information to improve the service in an efficient way. For instance, the information about riders boarding and alighting can be used to estimate the impact of passengers on route re-designs. Replacing a manual collecting process with an electronic one can enhance the reliability and accuracy of the information collected. It can also provide a better picture of the condition of the transit system. Therefore, electronic data can be used to allocate the necessary resources to efficiently manage the transit system and allow performing innovative analysis, improving the system, and enhancing the quality of the transit services.

Visualizing Transportation Networks. Using different types of visualization can assist transit professionals with the improvement of the transit system. Visualizations can serve as a window to the data, assisting a particular audience with focusing on the right amount of information for their needs. Therefore, target audiences such as transportation consumers, transportation administrators, and civic interest groups can benefit from different transportation visualizations methods.

Automated Data to Improve Decision Making. Transit agencies can organize and use the existing AFC data to better understand the potential impacts of fare options on ridership and revenue. Better information about AFC data and the implications of fare changes enables transit agencies to support a more robust public debate, a higher degree of accountability, and ultimately a wiser decision making.

AFC Data for Market Research and Demand Modeling. The implementation of Automated Fare Collection (AFC) systems provides new opportunities for improving the transportation services. Data can be mined to create inputs to operations planning and demand forecasting models. The information provided by these systems also allow the agencies to formulate questions about fares that are related to the transit services like: how are current fare products being purchased and used? and how have those patterns have changed over time? In addition, data can be used for demand modeling. The key considerations for selecting an appropriate model include the availability of historical data, the attributes that define the current and proposed fare products, and the availability of origin-destination-transfer information. Lastly, AFC data can help assess the performance of the system using market segmentation to better take into consideration the needs of different segments of the population.

Transit Equity. The collected data from the AFC systems open many possibilities for transit planners on developing different methods that can facilitate the prediction of people origins, destinations, and transfers as part of the transit trips. An origin-destination (OD) prediction can be used in the analysis and reporting of agencies' social goals, such as the provision of equitable service regardless of race, national origin, or ethnicity, which is federally required in the USA by Title VI of the Civil Rights Act of 1964. Title VI prevents agencies receiving federal funding from having a disparate impact with regards to race, ethnicity, or national origin. In complying with this law, transit agencies must report regularly on how their service is provided to populations with different demographics. Fare equity is a topic that will help planners and policymakers design a better infrastructure that can address people needs and demands.

Transit Fares for Disadvantaged Populations. It is expected by 2050 a significant growth in the globally elderly population. This projection will not only mean that the current lifespan will be extended, but it will bring as a consequence lots of changes. Therefore, transportation services need to be implemented to be able to cover the requests and demands that are expected to greatly increase. For this reason, transit agencies must take actions to provide better transit and affordable options, in particular regarding disadvantaged segments of the population.

Public transportation is a key component of urban transportation solutions that help mitigate congestion, reduce vehicle emissions, and promote sustainable growth. Therefore, specialized tools are needed to bring new insights on passenger behavior and demand characteristics of public transportation system by taking advantage of the massive continuous digital records of boarding logs from the data from the different transit systems, in particular the AFC system.

The selected work in this research study also included relevant North American and International experiences in the use of AFC data. The research identified in the relevant experiences section present additional information on the potential use of AFC data. The research present innovative uses of the data and results from research studies. In this area, it is worthwhile mentioning the work of international research investigators, as there is a vast amount of information that covers a variety of useful topics relevant to the use of AFC data. Although there is a wide variety of topics to cover, the areas captured in this research study include some of the most investigated areas. This includes selected research of the following topics:

- Understanding Fare Increases
- Trip Purpose Inference
- Fare Structure and Social Vulnerability
- Prioritizing Bus Schedule Revisions
- Carbon Emissions
- Jobs and Housing Relationships
- Commuting Patterns
- Extracting Boarding Information

- Transport Information Services
- Fare Collection Interoperability
- Policy and Planning
- Transit Assignment Modeling
- Analysis of Transit Service Performance.

Overall, the topics outlined in this research represent a good depiction of the available topics on the use of AFC data. The work in this research can help understand the challenges and opportunities researchers and practitioners face with the efficient collection, handling, and applications of the data from fare collection systems. Besides the identified topics for the potential use of AFC data, below are some observations from this study:

- AFC systems from transit agencies generate large amounts of data as part of their daily operations.
- The use of data can provide valuable insights to improve efficiencies in the delivery of public transportation services. However, the collected data are not being fully utilized in the planning and decision-making process.
- Combining AFC data with other ITS data presents many opportunities. Transit agencies can benefit from using enhanced datasets with information from the ITS systems. Therefore, there is a need for database systems that can take advantage of data from the all the existing ITS systems. In addition, to effectively and efficiently use the data from these systems, there is a need of sophisticated optimization models and decision support tools.
- The work in the advanced use of AFC data has been more in the academic research context. So, there is a need for technology transfer mechanisms for the research to be effectively transferred to practitioners for their use.
- There are many opportunities for using the data to improve the transportation services, but for this to materialize it requires long-term investment in infrastructure and talent. Further, there is a need of a roadmap that may require the support of County, State, or Federal agencies. In addition, vendors and developers of these system may need good incentives to develop tools or systems that can be costly.

After the extensive Review of the Literature and the Case Studies, it is clear that there are a lot of research material in this area, where specific methodologies and tools have been developed mainly to support the planning and operations of transit agencies. However, there is still a need of an efficient, comprehensive, and user-friendly system to fully take advantage of AFC data. An analytic and visualization tool that can mine the AFC data in combination with other datasets such as GTFS, census, and land use can be used to analyze passenger behavior and the demand characteristics of public transportation. This can also help detect patterns and trends and understand the general dynamics in passenger travel.

4.2 Proposed Data Mining Tool

This section presents a framework for the development of a data mining tool. The proposed tool can be useful for general data analysis, but it can also be used for granular data analysis at a higher level of detail. The proposed tool can extend the utility of AFC systems and help transit agencies and planning organizations to have the necessary data and information to plan and deliver a more efficient and equitable transit service.

The development of a data mining tool can provide the data analytic capabilities needed to support the operation analysis, service planning activities, and long-term planning strategies. It can also provide a means for the agency to observe and monitor system performance on a continuous and long-term basis, which makes it possible to discover regularities and dynamics in passenger travel, and to assess user responses to changes in the transit system (e.g., fare, route, scheduling, etc.), as well as to explore external factors such as demographic and social trends, connected and automated vehicles (CAV) technologies, and mobility service innovations. This can lead to better understanding of system dynamics and informed investment and policy decisions that serve the mobility needs, enhance system efficiency, and promote sustainable growth.

The following table and figures provide information of the framework proposed by the research team. Table 12 provides general information of the different data sources, based on information collected from the previous tasks and additional internet search conducted by the research team. Note that some of this information may vary depending on the vendor and the used AFC system technology.

Figure 55 illustrates the data elements of the proposed data mining tool. On the left-hand side, the proposed data sources are listed such as AFC, GTFS, census, and land use data. The middle of the diagram presents the data that will be stored in the database system, including the raw data, summarized data, and data ready for mining. Each dataset will have particular uses. For instance, summarized data can be used to quickly generate specific reports and mining data can be used for more sophisticated analysis. On the right-hand side, there are the different tools that the Business Intelligence (BI) Users will be able to use. This includes generating reports, conducting data analysis, mining the data, and displaying visualization components like charts or maps.

Similarly, Figure 56 depicts the AFC Data Mining System. This presents a cloud-based system that will be able to be accessed by different devices like desktops, laptops, tablets, or even smartphones. The software application in the data mining system will be the Graphic User Interface that will connect with the data mining database to allow users perform different functions for the visualization of data, creation of query statements, generation of reports, or exporting data in different file formats (e.g., csv, shapefiles, or pdf).

Table 12. Data Sources

Automated Fare Collection (AFC) Data	○ Card ID ○ Fare type ○ Transaction ID ○ Transaction Sequence Number ○ Route number ○ Direction ○ Datetime stamp ○ Boarding time ○ Alighting time ○ Boarding location ○ Alighting location ○ Transfers Stations
Parcel level land use data	○ Land use classification (residential, transportation, institutional and public buildings, commercial, and industrial)
Census data	○ Demographic information (age, race, ethnicity, gender, marital status, income, family characteristics, household composition and size
General transit feed specification (GTFS)	○ Static component that contains public transportation schedules and associated geographic information ○ Real-time component that contains arrival or departure predictions, vehicle positions, and service advisories

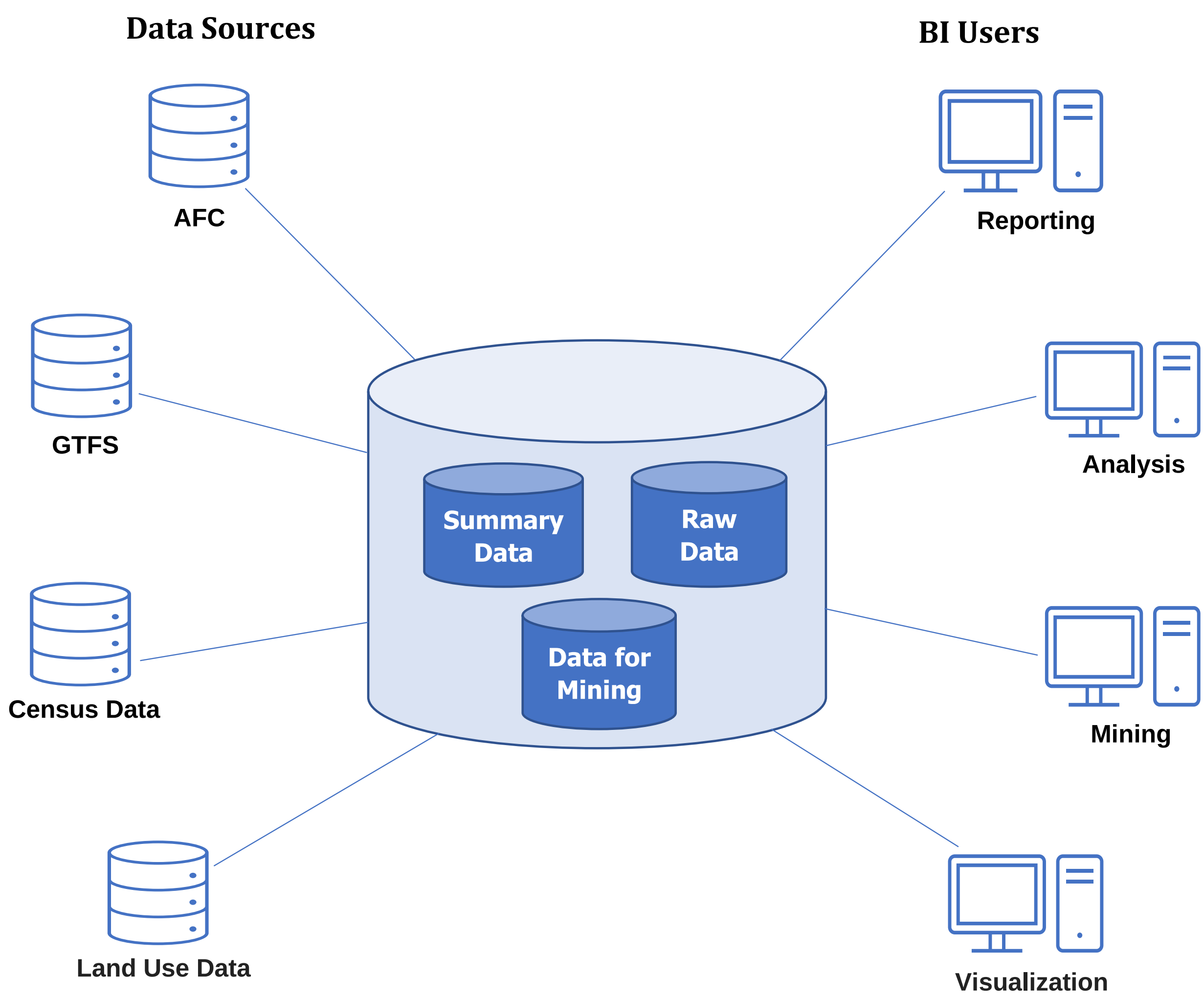


Figure 55. AFC Mining Data Elements

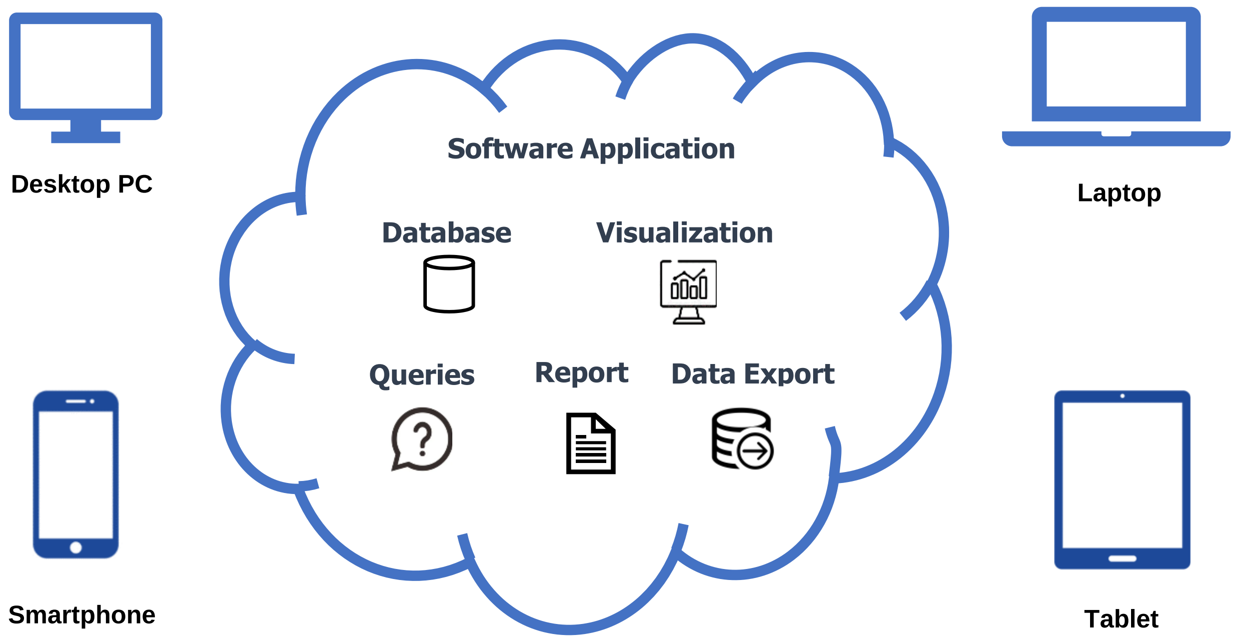


Figure 56. AFC Data Mining System Prototype

The proposed data mining tool will provide visual analytics of the transit system performance based on the AFC, GTFS, census, and land use data. The tool can support transit planning from operational, tactical, and strategical levels. If developed, the web-based system will incorporate data integration, visualization, analysis, and query capabilities to demonstrate a variety of potential applications of AFC data in support of transit planning and operations activities. Sophisticated data mining and visualization techniques will be employed to help identify behavioral patterns and examine the regularities and dynamics in passenger travel that can be used in the overall decision making process.

5 FUTURE RESEARCH

With the growing and aging of the population, the increasing traffic congestion, and the rapidly evolving technologies and mobility services, there is a pressing need to capture and monitor the trends and impacts on the transit system, as well as gain a better understanding of the spatial and temporal dynamics of transit demand. The massive and continuous information provided by the AFC data provide great potential to help examine the passenger behavior and demand characteristics of public transportation to support service planning as well as long-term strategy development.

Transit planners and transportation organizations still have limited instruments to obtain transit data. In general, agencies continue to obtain certain data through household travel survey data or transit on-board surveys, which are costly and time consuming. On the other hand, the massive and continuous boarding information provided by the automated fare collection (AFC) systems present great potentials as a bias-free and low-cost approach. Most transit agencies in the U.S. collect data from the electronic farebox and other systems. However, there is a lack of effective tools or mechanisms that allow for the efficient extraction and optimum use of transit data and information. So, this is an area that needs particular attention.

In addition, since the amount of data collected by the transit agencies continues to grow, the development of sophisticated techniques is key for taking full advantage of the available data. More research studies are needed on how to effectively manage and utilize large amounts of data and how that can be used by transit agencies and decision-makers. The development of new techniques or the technology transfer of previous research can benefit not only transit agencies, but also the whole transportation sector. Therefore, research is needed to develop computerized systems, sophisticated techniques, and innovative tools to help planners and researchers take full advantage of the data to assess the travel characteristics and behaviors of the users of the transit system.

Lastly, more work is needed to identify the role of Transit ITS vendors and software developers in the development of systems and tools. Perhaps, government agencies and decision makers need to look at the possibility of encouraging vendors and developers to create technological systems that combine ITS data with other datasets such demographic data, parcel data, census data, and other related datasets for the use by transportation and transit agencies. This is an area that needs to be explored to take full advantage of the readily available datasets that can be used for improving transit service that can promote the use of public transportation.

REFERENCES

Agard, B., Morency, C., & Trépanier, M. (2016, April 21). MINING PUBLIC TRANSPORT USER BEHAVIOUR FROM SMART CARD DATA. Retrieved from https://www.sciencedirect.com/science/article/pii/S1474667015359310.

Ali, A., Kim, J., & Lee, S. (2015). Travel behavior analysis using smart card data. KSCE Journal of Civil Engineering, 20(4), 1532-1539. doi:10.1007/s12205-015-1694-0.

Allen, W. E., Murray, R. D., & Acumen Building Enterprise, Inc. (2016). Implementation of Smart Card Automatic Fare Collection (AFC) Technology in Small Transit Agencies for Standards Development. Washington, D.C: The National Academies of Sciences, Engineering, and Medicine. Retrieved August 12, 2020, from http://onlinepubs.trb.org/onlinepubs/IDEA/FinalReports/Transit/Transit79.pdf.

Barry, J., Freimer, R., & Slavin, H. (2009). Use of Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. Transportation Research Record: Journal of the Transportation Research Board, 2112(1), 53-61. doi:10.3141/2112-07.

Casey R. and J. Collura. (1994). Advanced Public Transportation Systems: Evaluation Guidelines. Federal Transit Administration, Washington, D.C.

Casey, Robert F., Labell Lawrence N., Moniz, Leisa, Royal Jackson W., Sheehan Michael, Sheehan, Terry, Brown, Alex, Foy, Malinda, Zirker, Margaret, Schweiger, Carol L., Marks, Buck, Kaplan, Bruce, and Parker, Doug. (2000). *Advanced Public Transportation Systems: The State of the Art Update 2000*. Transit Administration Office of Research, Demonstration, and Innovation, Washington, D.C.

Cevallos, F. (2007). APTS Data Archiving and Mining System (ADAMS). Florida Department of Transportation.

Cevallos F. and X. Wang. (2008) ADAMS: A Data Archiving and Mining System for Transit Service Improvements, Transportation Research Record No. 2063, pp. 43-51.

Cevallos, F. and X. Wang. (2010a). Florida Advanced Public Transit Systems Program. Florida Department of Transportation.

Cevallos, F., X. Wang, and J. Skinner. (2010b). Using Data from an Electronic Fare Collection System to Identify the Travel Behavior of Seniors and People with Disabilities. Federal Transit Administration, U.S. Department of Transportation.

Chen, Z., & Fan, W. (2018). Extracting bus transit boarding stop information using smart card transaction data. Journal of Modern Transportation Volume, 26, 209-219. ttps://doi.org/10.1007/s40534-018-0165-y.

Cheon, S., Lee, C., & Shin, S. (2019). Data-driven stochastic transit assignment modeling using an automatic fare collection system. Transportation Research Part C: Emerging Technologies,Vol. 98, pp. 239-254, Rep. doi: https://doi.org/10.1016/j.trc.2018.09.011.

Coleman, M., Tarte, L., Chau, S., Levine, B., & Reddy, A. (2018). A Data-Driven Approach to Prioritizing Bus Schedule Revisions at New York City Transit. Transportation Research Record: Journal of the Transportation Research Board, 2672(8), 86-95. doi:10.1177/0361198118796717.

Danifo, P., Gradman, J., Pepple, S., & Uhlenhopp, P. (2014). Visualizing Transportation Networks: MTA Case Study (Rep.). New York, NY. Retrieved September 16, 2020, from https://www.researchgate.net/publication/270451268_Visualizing_Transportation_Networks_MTA_Case_Study.

Dickens, M. and Hughes-Cromwick, M. (2019, February). Leveraging Big Data in the Public Transportation Industry. Retrieved August 31, 2020, from The American Public Transportation Association (APTA) website: https://www.apta.com/wp-content/uploads/Big-Data-Policy-Brief.pdf.

Dumas, R. A. (2016). Analyzing transit equity using automatically collected data (Master's thesis, Massachusetts Institute of Technology. Department of Urban Studies and Planning.; Massachusetts Institute of Technology. Department of Civil and Environmental Engineering., 2015). Boston: Massachusetts Institute of Technology. doi: http://hdl.handle.net/1721.1/103650.

Egu, O. and Bonnel, P. Can we estimate accurately fare evasion without a survey? Results from a data comparison approach in Lyon using fare collection data, fare inspection data and counting data. Public Transport 12, 1–26 (2020). https://doi.org/10.1007/s12469-019-00224-x.

Eom, J.K., Song, J.Y. & Moon, D. (2015). Analysis of public transit service performance using transit smart card data in Seoul. KSCE J Civ Eng 19, 1530–1537. https://doi.org/10.1007/s12205-015-0013-0.

Farber, S., Bartholomew, K., Li, X., Páez, A., & Habib, K. M. (2014). Assessing social equity in distance based transit fares using a model of travel behavior. Transportation Research Part A: Policy and Practice, 67, 291-303. doi:10.1016/j.tra.2014.07.013.

Faroqi, H., Mesbah, M., Kim, J., & Tavassoli, A. (2018a, October). A model for measuring activity similarity between public transit passengers using smart card data. Retrieved June 26, 2020, from https://doi.org/10.1016/j.tbs.2018.05.004.

Faroqi, H., Mesbah, M., & Kim, J. (2018b). Applications of transit smart cards beyond a fare collection tool: A literature review. Advances in Transportation Studies, 45, 107-122. doi:10.4399/978255166098.

Fayyaz S. & Liu, C. (2015.). Fare Payment Structure and Dwell Time (Rep.). doi:https://www.westernite.org/annualmeetings/15_Las_Vegas/Papers/8A-Fayyaz.pdf.

First Data Government and Transit Task Force. (2010, May). Transit Payment Systems: A Case for Open Payments. Retrieved January 4, 2021, from https://www.firstdata.com/downloads/thought-leadership/transit-payment-systems_wp.pdf.

Governor's Commission on the Future of Transportation. (2019, April 5). A Vision for the Future of Massachusetts Regional Transit. Retrieved October 28, 2020, from https://www.mass.gov/files/documents/2019/04/17/dot-rta_task_force_report_040519.pdf.

Gschwender, Antonio, Munizaga, Marcela, and Simonetti, Carolina. (2016). Using smart card and GPS data for policy and planning: The case of Transantiago. Research in Transportation Economics. Elsevier.

Haney, E., Corley, T. A., & Forman, B. (2019, August 05). Prioritizing Equitable Growth Through Fare Policy. Retrieved January 27, 2021, from https://massinc.org/research/prioritizing-equitable-growth-through-fare-policy/.

Hanft, J., Iyer, S., Levine, B., & Reddy, A. (2016). Transforming Bus Service Planning Using Integrated Electronic Data Sources at NYC Transit. Journal of Public Transportation, 19(2), 89-108. doi: http://doi.org/10.5038/2375-0901.19.2.6.

Huang, X., & Tan, J. (2014). Understanding spatio-temporal mobility patterns for seniors, child/student and adult using smart card data. ISPRS - International Archives of the

Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-1, 167-172. doi:10.5194/isprsarchives-xl-1-167-2014.

Hwang M., Kemp J., Lerner-Lam E., Neuerburg N., Okunieff P., and Schiavone J. (2006) Advanced Public Transportation Systems: State of the Art Update 2006. Federal Transit Administration, U.S. Department of Transportation, Washington, D.C.

Iliopoulou, Christina and Kepaptsoglou, Konstantinos. (2019). Combining ITS and optimization in public transportation planning: state of the art and future research paths. European Transport Research Review. Springer. https://doi.org/10.1186/s12544-019-0365-5.

International Bank for Reconstruction and Development / The World Bank. (2016, June). Public Transport Automatic Fare Collection Interoperability Assessing Options for Poland. Retrieved July 10, 2020, from https://elibrary.worldbank.org/doi/abs/10.1596/24931.

Lathia, N., Smith, C., Froehlich, J., & Capra, L. (2012). Individuals among commuters: Building personalised transport information services from fare collection systems (5th ed., Vol. 9, pp. 643–664).

Lee, S.G., Hickman, M. Trip purpose inference using automated fare collection data. Public Transp 6, 1–20 (2014). https://doi.org/10.1007/s12469-013-0077-5.

Li, T., Sun, D., Jing, P., & Yang, K. (2018). Smart Card Data Mining of Public Transport Destination: A Literature Review. Information, 9(1), 18. MDPI AG. doi:10.3390/info9010018.

Liu, X., Zlatkovic, M., Porter, R., Fayyaz, K., & Yu, S. (2018, June). Transportation Research Board (Rep.). Retrieved December 28, 2020, from U.S. Department of Transportation, University Transportation Centers Program. website: https://www.ugpti.org/ resources/reports/downloads/mpc18-349.pdf.

Long, Y., & Thill, J. (2015, April 04). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. Retrieved from https://www.sciencedirect.com/science/article/pii/S0198971515000356.

Ma, Xiaolei, Congcong Liu, Huimin Wen, Yunpeng Wang, and Yunpeng Wang. (9 Dec. 2016). Understanding Commuting Patterns Using Transit Smart Card Data. Journal of Transport Geography, vol. 58, pp. 135–145. doi:http://dx.doi.org/10.1016/j.jtrangeo.2016.12.001.

MBTA Fiscal and Management Control Board (2018, June 18). *MBTA Tariff and Statement of Fare and Transfer Rules* (Rep.). Retrieved November 2, 2020, from Massachusetts Bay Transportation Authority website.

Mendoza, D. L., Buchert, M. P., Lin, J. C., & Https://orcid.org/0000-0002-7884-7362, D. L. (2019, August 13). IOPscience. Retrieved January 13, 2021, from https://iopscience.iop.org/article/10.1088/2515-7620/ab3ca7.

MTA Office of the Inspector General (2019, July 2). *NYC Transit Fare Evasion Surveys: OIG Concerns MTA/OIG #2019-12*. Metropolitan Transit Authority, New York.

Legislative Auditor General. (2014, August). A Performance Audit of the Utah Transit Authority (Rep.). Retrieved January 7, 2021, from https://le.utah.gov/audit/12_01rpt.pdf.

Patskanick, T. & D'Ambrosio, L. (2019). Transit Fare Discount Processing Improvements for Disadvantaged Populations (Rep.). Boston, MA: New England University Transportation Center. Retrieved October 19, 2020, from https://rosap.ntl.bts.gov/view/dot/43822.

Pelletier, Marie-Pier, Trepanier, Martin, and Morency, Catherine. (2011). Smart card data use in public transit: A literature review. Transportation Research Part C. Elsevier.

Pincus, K. S. (2014). Analysis of the 2012 Massachusetts Bay Transportation Authority fare increase using automated fare collection data (Master's thesis, Massachusetts Institute of Technology, 2014). Massachusetts: Massachusetts Institute of Technology. doi: http://hdl.handle.net/1721.1/89869.

Reddy, A. & Levine, B. (2016). Using Automated Data for Operations Control at NYCT (Rep.). New York, NY. Retrieved September 4, 2020, from https://web.northeastern.edu/transitdata2016/wp-content/uploads/2016/08/Reddy_Alla.pdf.

Sahin, I., & Altun, Y. (2007). Potential Uses of Electronic Fare Payment Records for Public Transit Agencies [Abstract]. ITE Journal, 77(12), 22-27. Retrieved August 4, 2020.

Shannon, E… & Bellisio, A. (2013, March). The MTA in the Age of Big Data: Transforming the Wealth of MTA Data into Accessible, Meaningful, Visual, Interactive Information. Retrieved October 5, 2020, from https://www.pcac.org/wp-content/uploads/2017/07/MTA-in-the-Age-of-Big-Data-7_25_17.pdf.

Shireman, M. T. (2011). Using automatically collected data for bus service and operations planning (Master's thesis, Marquette University, 2011). Massachusetts: Massachusetts Institute of Technology.

Sikorski, D. (2019, April 2). AFC 2.0 Next generation MBTA fare system. Retrieved November 6, 2020, from Massachusetts Bay Transportation Authority website: https://www.mass.gov/doc/afc-20-next-generation-mbta-fare-system/download.

Smart Card Alliance. (2011, November). Transit and Contactless Open Payments: An Emerging Approach for Fare Collection (Rep.). Retrieved January 20, 2021, from Secure Technology Alliance website: https://www.securetechalliance.org/publications-transit-financial-2011/.

Stuntz, A., Attanucci, J., & Salvucci, F. (2016.). Fare Policy Analysis Using Automated Fare Collection Data: Market Research and Demand Modeling at the MBTA (Rep.). Retrieved October 26, 2020, from Northeastern University website: https://web.northeastern.edu/transitdata2016/wp-content/uploads/2016/08/Stuntz_Andrew.pdf.

Stuntz, A. W. (2018, June). Transit fare policy: Use of automated data to improve incremental decision making. Retrieved November 25, 2020, from https://dspace.mit.edu/handle/1721.1/119275.

TransitCenter. (2018). The Data Transit Riders Want. TransitCenter. Retrieved October 14, 2020, from https://transitcenter.org/wp-content/uploads/2018/12/TC_TransitData_Final_FullLayout_121718-1.pdf.

Trépanier, M., Habib, K. M., & Morency, C. (2012). Are transit users loyal? Revelations from a hazard model based on smart card data. Canadian Journal of Civil Engineering, 39(6), 610-618. doi:10.1139/l2012-048.

The World Bank. International Bank for Reconstruction and Development. (2016, June). Public Transport Automatic Fare Collection Interoperability Assessing Options for Poland. Retrieved July 10, 2020, from https://elibrary.worldbank.org/doi/abs/10.1596/24931.

Vanderwaart, C. (2016). Planning transit networks with origin, destination, and interchange inference (Master's thesis). Boston: Massachusetts Institute of Technology. doi: http://hdl.handle.net/1721.1/103820.

Verbich, David and El-Geneidy, Ahmed. (2017). Public transit fare structure and social vulnerability in Montreal, Canada. Transportation Research Part A. Elsevier.

Welch, Timothy F. and Widita, Alyas. (2019). Big data in public transportation: a review of sources and methods. Transport Reviews, 39:6. https://doi.org/10.1080/01441647.2019.1616849.

Wu, L., Kang, J.E., Chung, Y., and Nikolaev, A. (2020). Inferring Origin-Destination demand and user preferences in a multi-modal travel environment using automated fare collection data. Omega, Elsevier. https://doi.org/10.1016/j.omega.2020.102260.

Zhang W., Chen F., Wang Z., Huang J. and Wang B. (2017). Evaluation of carbon emission reductions promoted by private driving restrictions based on automatic fare collection data in Beijing, China. Journal of the Air & Waste Management Association, 67:11, 1249-1257. DOI: 10.1080/10962247.2017.1320597.