USING AUTOMATED PASSENGER COUNTER DATA TO UNDERSTAND RIDERSHIP CHANGE ON A ZONAL BASIS

A Dissertation
Presented to
The Academic Faculty

by

Sanskruti Joshi

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
[May 2019]

USING AUTOMATED PASSENGER COUNTER DATA TO UNDERSTAND RIDERSHIP CHANGE ON A ZONAL BASIS

Approved by:

Dr. Kari Watkins, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Pascal Van Hentenryck
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Simon Berrebi
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: April 25, 2019

ACKNOWLEDGEMENTS

I would like to express my immense gratitude towards Dr. Kari Watkins without whom this thesis would have been impossible to work on. She accepted me in her team when I had no experience working with such enormous data but just loved transit. It was in her transit class when I got inspired to work with her because of her enthusiasm, passion and dedication towards transit. She has shown great confidence in me and without her knowledge and guidance, it would have been difficult to write this thesis.

I would also like to thank Dr. Simon Berrebi who has guided me through this entire research. He has been extremely patient with me when I started working with me and guided me to learn new skills that I can say have helped me a lot not just for this thesis but for my time at graduate school. He has always answered all my questions from really silly ones to some very serious doubts. His comments have helped me shape this thesis in a much better and simpler way. Dr. Pascal Van Hentenryck has been very supportive and available to review and give his valuable comments. His views outside of transit world helped improve this thesis from a whole new perspective.

I would like to thank my family for always supporting me and giving me strength to keep moving forward. My sister, Vyoma has stood with me in my tough times and I would not have made it without her.

Lastly, I would like to thank Sanket who has always made sure that I stay happy and cheerful and keep working in a healthy environment. I am looking forward our new beginning!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AFC	Automated Fare Collection
APC	Automated Passenger Counter
AVM	Automatic Vehicle Monitoring
AVL	Automatic Vehicle Location
FDOT	Florida Department of Transportation
GPS	Global Positioning System
GTFS	General Transit Feed Specification
MARTA	Metropolitan Atlanta Rapid Transit Authority
MDT	Miami-Dade County Transit
NCTR	National Center for Transit Research
NTD	National Transit database
TCRP	Transit Cooperative Research Program
TriMet	Tri-County Metropolitan Transportation District of Oregon

SUMMARY

Transit ridership is down across all modes, and bus ridership is at its lowest level since 1965 (Dickens, 2018). While bus ridership has been decreasing, there has been an increase in vehicle miles traveled, which is alarming for cities as it creates externalities such as traffic congestion, pollution and more traffic fatalities (Federal Highway Administration, 2018). Previous research at the city-level has not yet answered the question of why transit ridership has been changing given the relative heterogeneity in cities and more disaggregate zonal-level research is needed.

This study aims at understanding various issues or gaps in working with Automated Passenger Counter (APC) data and suggests a step-by-step process of cleaning the data in a format that can be used to answer research questions regarding transit ridership on a zonal-basis. To check discrepancies in APC data, this study uses General Transit Feed Specification (GTFS) which is a standard format for transit information, including schedules, stops, number of trips, fares and location of stops. APC was compared with GTFS data to check for gaps, such as missing trips, naming convention of routes and locations of stops. By using GTFS data for checking and cleaning, stops with missing trips were weighted and outliers were removed from the dataset. Apart from removing outliers, GTFS was also used to match up route names in APC similar to that in GTFS to streamline the process of comparing the two datasets. This research provides transit agencies with a methodology to check, clean and store APC data in a standardized format. Best practice of APC data management enables more in-depth analysis of ridership trends at all levels of aggregation.

CHAPTER 1. INTRODUCTION

There have been many technological advances in public transportation over more than 40 years. These include technologies such as GPS-based Automatic Vehicle Location (AVL), Automated Passenger Counters (APC) in the form of infrared beams and treadle mats, and web-applications and mobile apps to give updated information to passengers.

The use of Automated Passenger Counter devices dates back to the late 1970s and early 1980s when a few North American transit agencies started using them on their buses (Attanucci & Vozzolo, 1983). The Toronto Transit Commission installed 100 self-designed dual-beam counters and 16 signposts as a part of their Automatic Vehicle Monitoring (AVM) surveillance in 1976 Seattle installed treadle-mats for their APC systems in 1978. In 1979, California Department of Transportation (Caltrans) started using multiple-beam counters as a demonstration for five smaller transit agencies in California. TriMet installed dual-beam APC units on their fleet in 1982. Central Ohio Transportation Authority in Columbus started using dual-beam APC units in 1982. The Metropolitan Transit Commission of Minneapolis-St. Paul purchased dual-beam counters in 1979. Each of these agencies also carried out regular checks of their APC systems with manual ride-checks in order to make sure that the accuracy levels of the data collected remains within acceptable confidence levels (Attanucci & Vozzolo, 1983)

With advancements and research in the use of automated data collection technologies, more and more transit agencies have started using APC devices. According to TCRP Synthesis 77 on Passenger Counting Systems, the following were major findings regarding APC practice in 2008 (Boyle, 2009):

- Top reasons for not using APC systems were cost and low priority of this system at the transit agency.
- The majority of the transit agencies from the survey conducted under TCRP Synthesis 77 reported that APC is not installed on all of their fleet. This is a common practice to install APCs only on a portion of buses and then rotate them among all the routes.
- Most of the transit agencies are very satisfied or somewhat satisfied with the performance of APC systems in terms of counting passengers.
- About 55 percent of transit agencies use a vendor for developing data processing and reporting software (TCRP 77).
- An organizational benefit due to implementation of APC systems is the positive change among various departments at transit agencies in terms of communication and coordination.

Another important data standard that is very new compared to APC, but has developed very quickly over the past decade, is General Transit Feed Specification. GTFS formerly known as Google Transit Feed Specification was started by few employees of Google as a part of a "20 percent time" program launched by Google to let their employees take up any project or activity that they liked for twenty percent of their working time to foster creativity and innovation in the organization. GTFS was launched in 2005 in collaboration with TriMet. Transit data feeds are complex due to presence of both temporal and spatial data (Roth, 2010). Hence GTFS was developed as a relational database to handle these datasets and use it in Google's Transit Trip Planner. Any transit agency who maintains GTFS files is eligible to send its data to Google for free inclusion on Google

Transit Trip Planner app. To generate GTFS data, transit agencies can use in-house expertise or outsource it to various vendors in the market.

GTFS has undergone changes in the past as an open data source with an active user community that contributes to its development. As of April 2019, there are 21 open proposals on changes in GTFS (Google Open Source, 2019b) whereas 16 proposals are in their early stage and are listed as issues on GitHub (Google Open Source, 2019a). Changes that occur in GTFS are monitored by the user community and Google. Uses of GTFS includes service planning, mapping, schedule information, accessibility information and visualizations based on spatial and temporal information from the same source. There have been advancements to GTFS such as General Bikeshare Feed Specification (GBFS) and GTFS- ride. The purpose of GBFS is to provide bikeshare information in a city, whereas GTFS-ride is being developed to include ridership information files in addition to standard GTFS files.

With the usefulness of APC and GTFS datasets in mind, this study takes a deeper dive into these two datasets for four transit agencies in the US. The organization of this study is as follows. Chapter 2 talks about existing literature on APC and GTFS and how these datasets have been processed and validated over time make sure that the accuracy levels are sufficient enough to use for different research, planning and decision-making purposes. Chapter 3 is about the structure of APC data of each transit agency and its limitations or issues. Chapter 4 gives a detailed account of the methods used for processing of APC data, issues faced during the process and steps taken to mitigate them. Chapter 5 concludes this study with summaries on data for each agency and few key takeaways to be kept in mind while using stop-level APC data.

CHAPTER 2. LITERATURE REVIEW

This section aims at studying various research projects that have been conducted related to Automated Passenger Counters (APC) and General Transit Feed Specification (GTFS). Various studies presented below include early use of APC data until its adoption for National Transit Database (NTD) reporting. These studies talk about the usefulness of APC, its drawbacks, and methods that have been suggested by various researchers for making sure that the accuracy of data collected is maintained. GTFS is a relatively newer than APC but there have been some studies on its development and its use in the market.

2.1 Studies on Automated Passenger Counters

Attanucci and Vozzolo (1983) studied early development of APC technologies and its comparison with manual ride-checks. The authors defined a 4-step process for using APC data.

1. Collect data using APC devices on transit vehicles.
2. Record and store data
3. Transfer data to the central units or facility
4. Report and analyze collected data.

An important aspect of this study was the comparison of various APC technologies and their accuracy levels in reporting ridership counts compared to manual ride-checks. Almost all the agencies had ridership data collection accuracy ranging from 90 to 95 percent. Another observation suggested that APC devices tend to undercount the ridership data and hence most agencies used correction factors to account for undercounting. Along

with data accuracy, APC is more cost-effective than manual ride-checks due to reasons such as less data turnaround time compared to manual ride-checks (which often take up to 1 year) and increase in the information that can be collected using APC compared to manual ride-checks (Attanucci & Vozzolo, 1983).

USDOT funded research to evaluate the feasibility of implementing APC for Lane Transit District (Hodges, 1988). For this, they evaluated various transit agencies already using APC technologies for data collection. An interesting observation made was that there was no significant difference between accuracy of APC and manual ride-check data when the manual-ride checkers were aware that their data would be checked with APC. Most of the transit agencies had reliability issues with the hardware component of APC, but in longer term, the benefits of APC out-numbered the issues with the hardware. Some benefits included high turnaround time of data, more parameters through automated counts and flexibility to collect data on any route and for any time period that would potentially help in making decisions on changing services.

Kimpel et. al. conducted a study on checking the accuracy and precision of APC data by comparing it to data obtained from video cameras and developing sampling methodologies for annual NTD reporting. This research was focused on using statistical analysis to verify accuracy and precision of APC data compared to video cameras. Results showed that APC data had accurate boardings when compared to cameras. Another important finding was overestimation of passenger loads due to restrictions on allowing negative load values that aggregated over time to overestimate the loads. By allowing negative load values, using raw boarding and alighting data and by setting zero values at places where the layover was long enough to assume that the buses would be empty,

corrected passenger loads were obtained. For NTD reporting, using APC was found to be more accurate than manual ride-checks. Moreover, APC allowed larger sample sizes at lower costs compared to manual data collection methods (Kimpel, Strathman, Griffin, Callas, & Gerhart, 2003).

Strathman et. al. studied reasons that have been a hindrance to using APC to their full capabilities by using five agencies as case studies. The main reasons found from their study include issues with the basic data storage structure, accuracy in recorded data, controlling drift and lack of balancing algorithms. They performed different statistical tests for each of the identified issues and suggested some steps to overcome them. They also suggested a balancing algorithm for correcting passenger loads (Furth, Strathman, & Hemily, 2005).

A potential use of APC data that has been increasing is its use in reporting ridership for the National Transit Database (Chu, 2010). Before using APC data for NTD reporting, transit agencies have to submit a benchmarking plan and a maintenance plan for APC. For the first year of use of APC data for NTD reporting, the numbers have to be checked by manual counts to make sure that there are no errors. If 100% data is not obtained from APC, then adjustment factors should be used for missed data. Similarly, for errors in the dataset, adjustment factors should be used. If agencies do not have 100% APC penetration in their fleet, alternatives should be considered like using manual counts or random sampling.

A study by Metropolitan Council of Minneapolis-St. Paul metropolitan region discussed auditing the accuracy of APC data on Metro's Blue and Green line Light Rail

Vehicles (LRVs) (Minnesota Metropolitan Council, 2014). They conducted their audit in two phases and found out that APC data was within the warrantied accuracy numbers by the manufacturer. TriMet also studied their APC data and reported that the ridership data was statistically valid and there was no need for ride checkers. However, it was noted that the level of accuracy was obtained by constant efforts over a year to fix all the issues with their APC databases (Minnesota Metropolitan Council, 2014).

Whitmore et. al. (2018) studied the use of Automatic Vehicle Location (AVL) data and Automated Passenger Counter (APC) data to understand the performance of transit systems. AVL-APC devices were mounted in the transit vehicle. AVL devices would start recording trip details such as trip number, time of day, date, vehicle number, arrival and departure times at stop and timepoints. APC data recorded information such as number of passengers boarding and alighting the vehicle, and passenger load, i.e., number of passengers on a bus. Raw data collected from AVL-APC systems was put through a quality check process in order to get rid of erroneous data like missing records and illogical load values such as negative load values. Stop locations were used from a GTFS dataset for the case studies. Processed AVL-APC data was then used for analysis such as on-time performance, bus bunching levels, stop-skipping frequency and bus crowding levels (Pi, Egge, Whitmore, Silbermann, & Qian, 2018).

2.2 Studies on GTFS

A study was carried out by National Center for Transit Research funded by FDOT about the evolution of GTFS, its potential benefits and its future use (National Center for Transit Research, 2011). This research highlighted important characteristics of GTFS that

has made it easy to use in the transit industry. Due to its standard format and continuous involvement of developers, its integration with various mobile apps and web-based applications have become easy. They created a web-application that used both GTFS and APC datasets to visualize ridership trends and changes online. It is difficult to have a standardized method to use both datasets as different transit agencies have different APC formats. Other potential datasets that can be used with GTFS are AVL and AFC systems. A more standardized format for each of the data sources across all transit agencies could make the flow of information more efficient and quicker.

Related to this, a standard called GTFS-ride is in its early development stage. The idea behind developing this standard is to make ridership data more accessible to people. A survey questionnaire was prepared and sent out to 147 transit agencies in Oregon and 6 non-Oregon transit agencies to get their responses on ridership data collection methods, organization of the data and its use by transit agencies (Porter et. al., 2018). From the results of the survey and after careful considerations of all the recommendations made by various transit agencies, a standard format for GTFS-ride was developed. The standard comprises of five files which includes board_alight.txt, ridership.txt, rider_trip.txt, trip_capacity.txt and ride_feed_info.txt of which only ride_feed_info.txt is a required file while the rest are optional files(Carleton, Hoover, Fields, Barnes, & Porter, 2019).

GTFS-ride has been gaining attention with various stakeholders outside of Oregon state have started to participate in its development process. These include King County Metro, Massachusetts Bay Transportation Authority, Metro Transit, Remix, Trillium Solutions Inc., Transport for London, TriMet, Urban Labs LLC, Volpe Transportation Systems Center and Blacksburg Transit (Carleton et al., 2019). Due to participation from

transit agencies, transportation consultants and software vendors, it is expected that this standard will be able to provide open ridership data outside of transit agencies. To demonstrate the process for developing GTFS-ride feed, a study on ridership data for three transit agencies was carried out. Each agency had a different format of storing their historical ridership data and hence new processes were developed for each agency to get data in GTFS-ride's standard format. Lessons learned from the study of the three transit agencies establishes that every transit agency needs its own processes to produce ridership data in GTFS-ride format. Some of the agencies have resources in place to produce GTFS-ride feeds while some might have to go through some manual processes in order to produce GTFS-ride feeds.

With the development of GTFS-ride in future, there will be more opportunities for software vendors to develop new products for processing ridership data from transit agencies to prepare GTFS-ride datasets. Transit agencies will have opportunities to better analyze their data in order to improve their services which can also improve transit ridership for the agencies.

CHAPTER 3. DATA AND ITS LIMITATIONS

The scope of this study involves 4 transit agencies in the US. This section is about the automated passenger data obtained from these transit agencies and how each agency structures their APC data. Every agency has its own data limitations which will be explained further in the context with each agency.

General Transit Feed Specification (GTFS) data has also been used in this study. In most of the cases, GTFS is very structured and follows the standard format. But there are few issues with this data as well. This chapter also talks about these issues in the context of each agency. GTFS was obtained from two of the third-party websites called Transit Feeds and GTFS Exchange. While downloading data, it is important to look into calender.txt or calender_dates.txt files to make sure that the files being downloaded correspond to the same time-period as needed for a study. The structure of GTFS is shown in Figure 1.

GTFS has five required files – agency.txt, routes.txt, stop_times.txt, stops.txt and trips.txt. These files contain following information:

- Agency.txt: Information about name of the transit agency, its website URL, time zone of the agency and contact details
- Routes.txt: Each route in the system and its associated route_ids and route names
- Stop_times.txt: Arrival and departure times for each trip at individual stops.
- Trips.txt: Each trip for individual routes

- Stops.txt: Stop locations where transit vehicle picks up and drops off the passengers

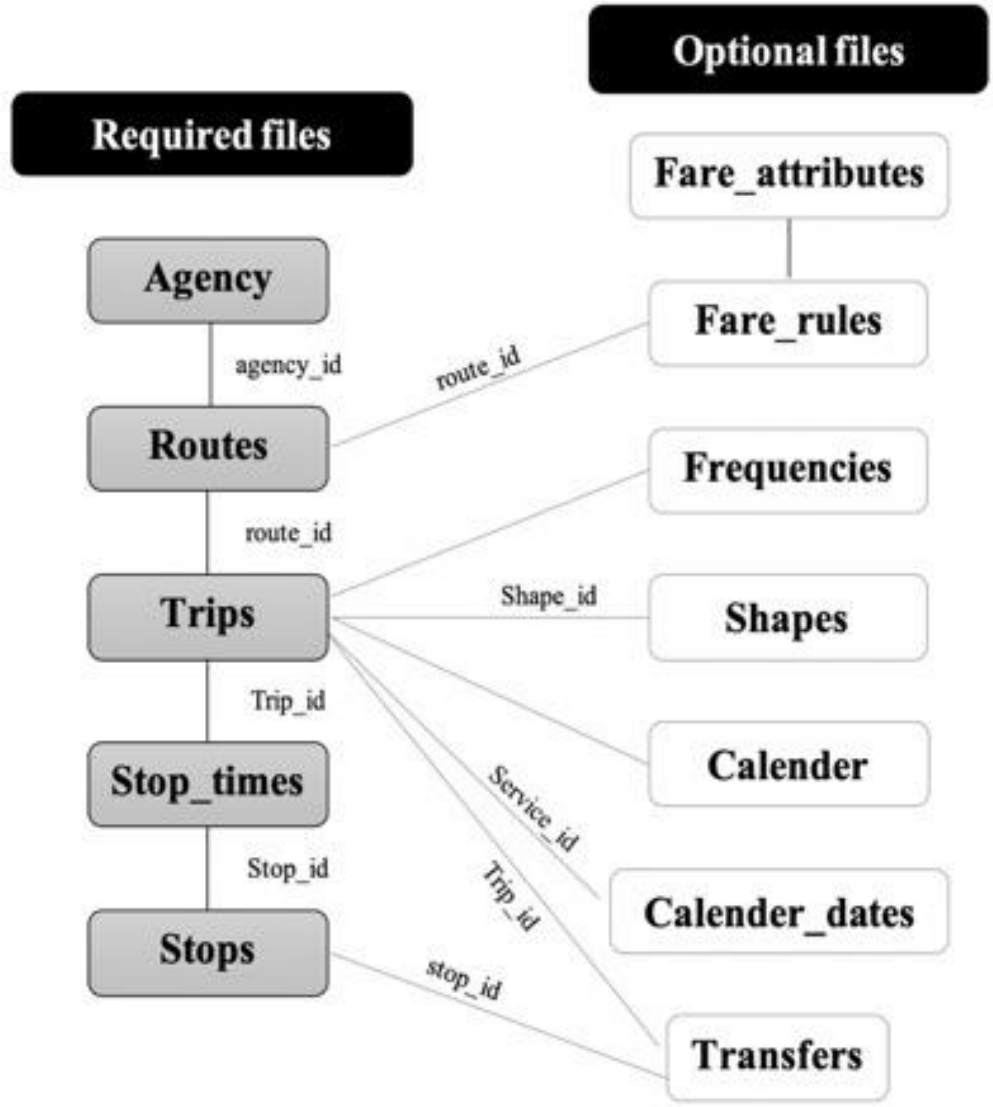

Figure 1 GTFS standard format and relation between files

3.1 Metropolitan Atlanta Rapid Transit Authority, Atlanta, Georgia

The Metropolitan Atlanta Rapid Transit Authority (MARTA) is the primary public transit agency of Atlanta in the state of Georgia. It operates in Fulton, Clayton and Dekalb counties and city of Atlanta (Office of Transit System Planning, 2017). MARTA operates a bus network as well as a rail network. The rail network has a service span of up to 21 hours. MARTA also operates a paratransit service called Mobility in compliance with Americans with Disabilities Act (ADA) for people who cannot use MARTA's regular services for navigation. As of June 2016, MARTA's service covers up to 1311 miles which includes 338 railcars, 101 bus routes with a fleet of 570 buses and 211 Paratransit vans (Office of Transit System Planning, 2017).

MARTA's automated passenger counter data is at stop-route-trip level which means that each trip corresponds to a combination of route, stop and direction where the trip has been made. Since the data has different types of days, it helps in carrying out analysis for weekdays and weekend ridership trends separately. For MARTA, the data was received for 5 years from 2014 to 2018. Appendix A.1 gives information about MARTA's metadata.

3.1.1 Data Limitations

Automated passenger counter data generally has information about number of samples that corresponds to the average boardings and alightings for each row in the data, i.e., for each trip for a combination of stop, route and direction there will be a column that gives information about the number of samples used to compute the averaged values for boardings and alightings. This data is missing from MARTA's APC which should ideally

be included in a good APC dataset to make sure that the sample size for each row in the data is sufficient to be considered for analysis.

There is no column that gives information about the year and month for which the data was collected. The only information available about the year and the month or the markup period is from the file name. Due to this, while using the data, year and month have to be added manually for each time period which is a cumbersome task in the long run for multiple time periods. Even though this is a trivial issue, it might become tedious to add it manually for large datasets. This issue remains the same for all four agencies discussed here.

Trip start time is only useful since it gives information for each trip for each stop for each route. MARTA's APC doesn't have information related to the trip number but information about trip start time is sufficient for knowing the number of trips for a stop, route and direction combination unless a research involves matching up trip_ids from APC with other data sources like GTFS.

There is no information available for stop location. Hence if spatial analysis is to be carried out using this dataset, additional location data will have to be added to the data. It is advisable to have two columns for latitude and longitude for each stop in the system.

MARTA's APC data has directions namely Eastbound, Westbound, Northbound and Southbound instead of numbers 0 and 1 for opposite directions on a route. If APC data is used as a standalone source, this is not an issue but if it has to joined with GTFS using direction as one of the common fields between tables, this creates an issue of non-matching values due to different naming conventions.

3.2 Metro Transit, Minneapolis-St. Paul, Minnesota

Metro Transit is the primary transit agency of the Twin cities of Minneapolis and St. Paul in the state of Minnesota. It has a network of buses, light rail and commuter trains. It serves an area of 907 sq. mi. with 130 routes (including Maple Grove Transit operated by Metro Transit) which is made up of 55 urban local service, 63 express, 9 suburban local, 2 light rail and 1 commuter rail routes (Metro Transit, 2019).

Metro Transit's automated passenger counter data is at stop-route-trip level which means that each trip corresponds to a stop and route number. For Metro Transit, APC data was received for years starting from 2008 through 2017. Appendix A.2 gives information about Metro Transit's metadata.

3.2.1 Data Limitations

There is no information about stop locations, i.e., information about latitude and longitude of the stops is missing. This information is useful for spatial analysis of ridership trends. Due to lack of this information, other sources (such as GTFS) need to be used to add location data to the APC dataset.

Metro Transit's data has direction as names like East, West, North and South. This is an issue when matching up with GTFS because GTFS uses 0 and 1 as direction_ids.

3.3 TriMet, Portland, Oregon

TriMet is the primary transit service provider in Portland, Oregon. It has a network of buses, light rail, commuter rail and LIFT paratransit service. TriMet serves an area of

533 sq. mi. which includes 79 bus lines, 5 MAX light rail lines, 3 commuter trains and 253 LIFT buses (TriMet, 2018).

TriMet's automated passenger counter data is at stop-route-trip level which means that each trip corresponds to a combination of route, stop and direction where the trip has been made. For TriMet, the data was received for years starting from 2007 through 2017. TriMet's APC contains information about the type of day which makes weekdays and weekend analysis possible. Appendix A.3 gives information about TriMet's metadata.

3.3.1 Data Limitations

There were no major data limitations with TriMet. The only issue was with the naming of their trip numbers. It is different from that in GTFS and hence would cause an issue if two tables from APC and GTFS are to be joined using trip_id as a common field. For the scope of this study, this was not an issue since trip_id was not required to be used as a common field.

3.4 Miami-Dade Transit, Miami, Florida

Department of Transportation and Public Works (DTPW) is responsible for providing transit services in Miami-Dade County region. It operates in a service area of about 306 sq. mi. DTPW provides four major modes of transit – bus, heavy rail, automated people mover (APM) and demand responsive service (for people with disabilities). It has a total of 95 bus routes, 2 heavy rail routes and 3 APM routes (Miami-Dade Transit, 2018).

MDT's automated passenger counter data is at stop-route-trip level which means that each trip corresponds to a combination of route, stop and direction where the trip has

been made. For MDT, APC data was received for 5 years from 2013 to 2017. Since the data has different types of days, it helps in carrying out analysis for weekdays and weekend ridership trends separately. Appendix A.4 gives information about MDT's metadata.

3.4.1 Data Limitations

The information about stop location has been corrected for any errors but the precision of those numbers is lesser than that found in stops.txt in GTFS. Hence, it will be advisable to use stop location information from stops.txt from GTFS instead of APC.

CHAPTER 4. METHODOLOGY: DATA CLEANING AND PROCESSING

This section explains different methods that were used to process APC and GTFS datasets. Every agency had similarities as well as differences in their automated passenger counter data. Hence, for each agency there were variations in the data cleaning process in order to get the resulting tables in a standard format. Each sub-section follows a step-wise process for checking, processing and cleaning the datasets. Each sub-section also talks about the steps taken in order to merge GTFS and APC either to obtain certain information from GTFS or to check APC data for any errors.

The flowchart below (Figure 2) gives a general framework of the methodology used for each agency.

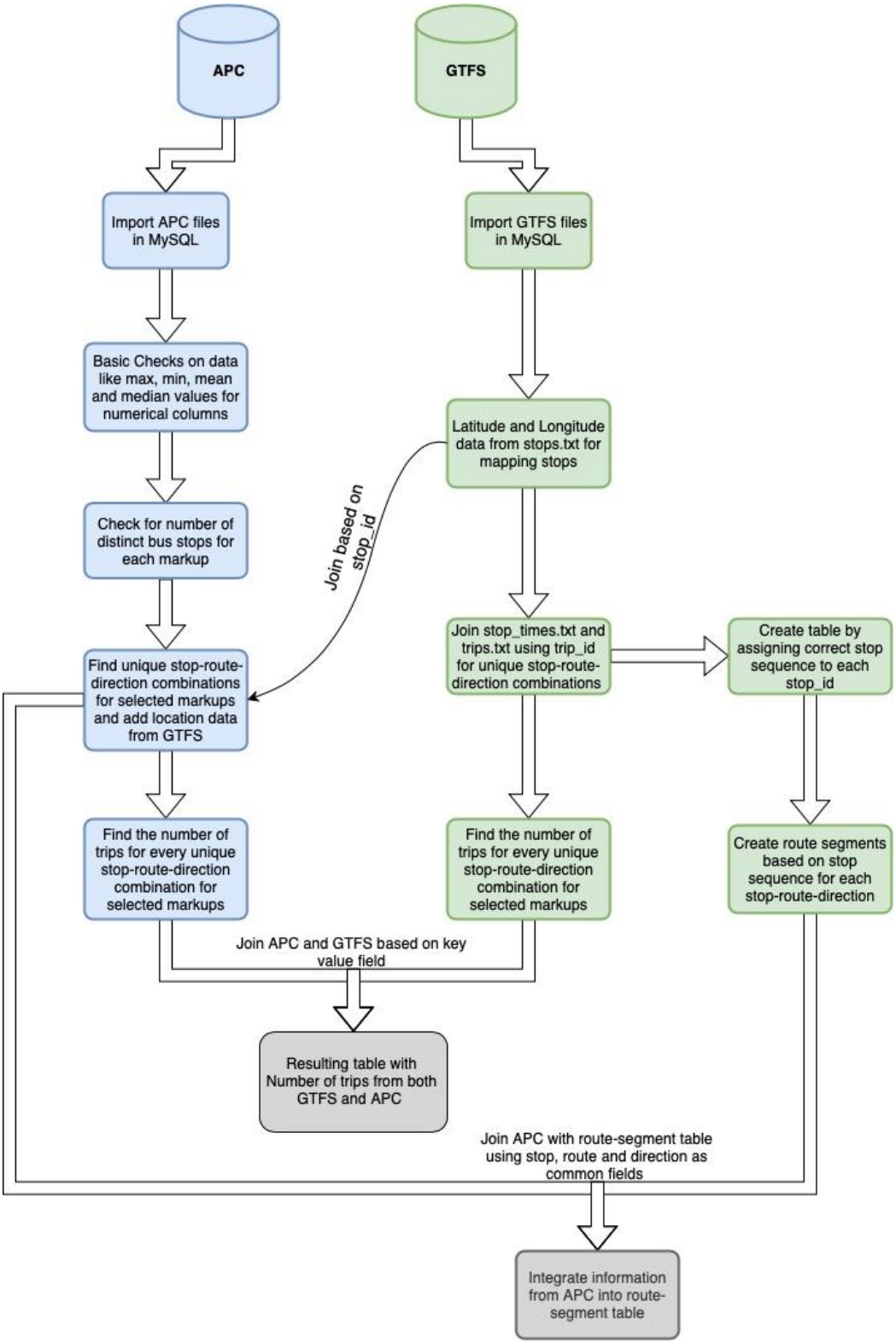

Figure 2 Flowchart showing the methodology

4.1 General Methodology

Before performing any checks, APC and GTFS datasets were imported in an SQL database. During the import, it was critical to check the datatype (whether its numeric or character) of each column in the raw data. Since there was no information about the year and month in the data itself, those two columns were manually added during the import process for both APC and GTFS datasets.

4.1.1 Initial Checks

The first step in the process of checking the data was to perform initial checks to understand the data. The following checks were performed on the tables that were imported in SQL:

Step 1. Total number of rows in the data to know the size of the data.

Step 2. For each column, number of unique values and list the unique values if less than 10 was counted. This helps in knowing unique values for columns like day_type, year and direction.

Step 3. For each column, the number of null or empty values was counted. This helps in knowing which column has missing values in the dataset especially in case of columns like ons, offs and stop_id that contains critical information and should not have null values.

Step 4. For each year and month, i.e. for each markup period, the number of unique stops were counted. This step helps in understanding any major change in the number of bus stops and whether that happened due to missing data or there was an actual

change made by the transit agency. In case of MARTA's APC, there were no such errors in the data.

Step 5. For numeric columns like ons, offs, and Ons&Offs, mean, median, maximum and minimum values were calculated to know the spread of the data.

Next steps involve getting number of trips for unique combinations of stop, route and direction (SRD) for selected markups for further analysis. The number of trips for each SRD are then validated using GTFS data for same time periods. This step helps in getting rid of any SRD combinations that have huge differences in number of trips. Hence, this is a crucial step in cleaning the data for stop-level analysis.

4.1.2 Processing of APC

To obtain unique SRD combinations in APC, following steps were taken on new table that only contained weekdays and selected markups (Let's call it T1):

Step 1. To get stop-route-direction (SRD) combinations present in both markup periods, I created a new temporary table (temp1) from T1 by applying group by condition in SQL query that would give resulting table that had unique stop, route, direction and markup combinations. Thereafter, I created another temporary table (temp2) from temp1 with a group by condition on only stop, route and direction columns with a count condition that creates a new column that gives the count of occurrence of that SRD combination in the table. If the count is 1 it means that the combination occurs only once whereas if the count is 2, it means that the stop-route combination is present in both the markups.

Step 2. Another table(temp3) was created with counts equal to 2 from temp2 in the previous step. A new column was added to temp3 and T1, that acted as a common key field to join temp3 with T1 and with GTFS in later steps. This key field should be a combination of stop, route and direction columns in a format of stopid_routeid_directionid.

Step 3. I wrote an SQL script to join temp3 and T1 that added trips from T1 to temp3 based on the common key value field. This step ensures that only trips included in the analysis correspond to constant SRD combinations.

Step 4. A final table was created (apc_trips) by grouping by stop-route-direction combinations on temp3 and adding a count condition on trips column to get a new column with number of trips for each stop-route combination.

4.1.3 Processing of GTFS

To obtain unique SRD combinations in GTFS, following steps were taken:

Step 1. In GTFS, no single table has information about stop, route, direction and trips. Hence, to obtain all the information in one single table two tables namely trips.txt and stop_times.txt were used.

Step 2. Service_id should be checked in GTFS from calender.txt or calender_dates.txt to make sure that only weekdays are used. This can be different depending on the scope of a study.

Step 3. Same as APC, data for only the required markups was used to obtain unique stop-route combinations. First step was to join stop_times.txt and trips.txt based on

trip_id, year and month as a common field with service_id that corresponded to weekdays only. For convenience, call this table T2.

Step 4. Next, I created two tables (let's call them gtfs1 and gtfs2) with two different markup periods to make joins with APC quicker and two separate tables of two markups makes it easier to visualize different time periods individually. These tables should be created by applying a grouping condition on stop, route and direction for each markup along with a count condition on trip_id column that will create a new column with number of trips corresponding to each stop-route-direction combination.

Step 5. The resulting tables gtfs1 and gtfs2 will have unique stop-route-direction combinations. Add a key field column to each of the tables similar to APC in a format of stopid_routeid_directionid.

Once apc_trips, gtfs1 and gtfs2 are obtained, I joined apc_trips once with gtfs1 and then with gtfs2 based on common key field with a condition on apc_trips for each markup. Hence two new tables were obtained for each markup that had information for number of trips both in APC and GTFS corresponding to each stop-route combination. A new column was added to both the tables to compute difference between number of trips in GTFS and APC. The resulting table can be exported as comma separated file and used for visualization of difference in number of trips between GTFS and APC for SRD combinations.

4.1.4 Processes for creating route-segments

For route-segment level analysis, i.e., making route segments by clustering stops based on their stop sequence, following steps were taken to create route segments using

GTFS and APC data. The idea here was to create route segments with 7 to 14 stops. Every segment starting from the first stop on a route for each direction had 7 stops until the last set of stops was reached that had more than 7 stops but less than 14. All those stops were assigned a segment. For general purposes, number of stops to form a segment should be decided based on the scope of the study. To obtain route-segments, the following steps were taken:

Step 1. Tables stop_times and trips were used from GTFS to create route segments. These tables were joined based on trip_id as a common field (Let's call it T3).

Step 2. A new table was created from T3 by grouping route_id, stop_id, direction_id and stop_sequence along with a count condition on stop_sequence that will create a new column with the count of occurrence of that stop_sequence for each SRD combination.

Step 3. A new table was created from the table in step 2 by grouping on stop, route and direction and applying a maximum condition on stop_sequence which gives resulting table with stop sequence that occurs for maximum number of times for each SRD.

Step 4. A new field was then created with a key field in format of stopid_routeid_directionid. This final table had stop-route-direction with corrected stop_sequence and a key field (let's call this table gtfs_seg).

Step 5. A table (let's call it apc_seg) was created with constant SRD combinations from APC by following the steps that were used to obtain unique SRD combinations for APC data in 4.1.2.

Step 6. Thereafter, I joined tables gtfs_seg and apc_seg in order to get boardings and alightings information for each stop-route combination. This will be the final table for route-segments.

Step 7. The final step involves an "IF" condition that assigns same segment number starting from first stop_id on a route and direction until it reaches 7^{th} stop. If next iteration has 7 stops, it will increment the segment number by 1 and keep checking till more than 7 but less than 14 stops are remaining. Once that happens, the query assigns the last route segment number to all those stops. In general, the count of stops should be set up according to the upper (here 14) and lower (here 7) limit on number of stops in a segment.

Sections 4.2, 4.3, 4.4 and 4.5 are about specific issues faced while performing steps under sections 4.1.1, 4.1.2, 4.1.3 and 4.1.4. Wherever an issue was encountered, its solution and the step involved has been mentioned. Each sub-section also contains results of the checks wherever necessary.

4.2 Metropolitan Atlanta Rapid Transit Authority, Atlanta, Georgia

Results from initial check shows that MARTA's APC had 13,258,713 rows. For numeric columns like ons, offs, and Ons&Offs, mean, median, maximum and minimum values were calculated to know the spread of the data. Its results are presented below.

Column Name	Min	Max	Average	Median
Ons	0	60	0.299	0
Offs	0	67	0.298	0
OnsOffs	0	67	0.621	0

4.2.1 Issues with data and its solutions

The following are issues faced during processing of APC and GTFS for MARTA and solutions adopted to overcome them.

Issue 1. MARTA's APC has names for directions whereas GTFS has numbers 0 and 1. Moreover, there is no information in GTFS that can be extracted to get the directions in names instead of numbers.

Solution. Since there was an issue with direction names, before obtaining unique SRD combinations, each markup was checked for all stop-route combinations that had two directions instead of just one. There were only 2% stop-route combinations that had two directions. Those combinations were removed from the process of finding unique SRD combinations. By doing this, further processing of APC did not involve using direction_id. All the steps performed in 4.1.2 will remain the same with a change of using stop-route combinations instead of SRDs. It is important to make sure that the removed stops are not the stops with high ridership. If so, each direction needs to be manually identified to be used for obtaining SRDs.

Issue 2. Another issue that was encountered was with naming of stop_ids in APC. For 2014, APC's stop_ids corresponded to stop_id whereas for 2018 APC's stop_ids corresponded to stop_code in stops.txt in GTFS.

Solution. An additional step was taken to add a final stop_id column in APC by matching up with GTFS for each markup. Rest all the steps from 4.1 remained the same.

Issue 3. Moreover, route number in APC corresponded to route_short_name in routes.txt in GTFS. Hence, these are important issues to be checked for before joining the two datasets and finding unique SRDs.

Solution. Route_ids corresponding to route_short_name was added from routes.txt table from GTFS to MARTA's APC.

The results obtained as an output from steps performed in 4.1.2 and 4.1.3 are tabulated below. The Table 1 shows percentage of stop-route combinations that have zero, ±1, ±2, ±3, ±4 and more. Figure 3 shows comparison of years 2014 and 2018 for percentage of missing trips in APC or GTFS. Since percentages of stop-route combinations with difference in number of trips are very low, the figure shows values for "0" difference off the top of the chart in order to scale up the other values. A positive difference means that there are more GTFS trips and APC and vice versa.

Table 1 Percentages of SR combinations for APC-GTFS trip comparison for MARTA

year	2018	2014
less than -4	0.43 %	1.56 %
-4	0.00 %	0.00 %
-3	0.02 %	0.75 %
-2	2.69 %	0.54 %
-1	3.84 %	4.11 %
0	88.54 %	88.50 %
1	2.69 %	1.73 %
2	1.32 %	0.64 %
3	0.00 %	0.62 %
4	0.02 %	0.00 %
more than 4	0.02 %	0.02 %
Null Values	0.43 %	1.53 %
Total	100.00 %	100.00 %

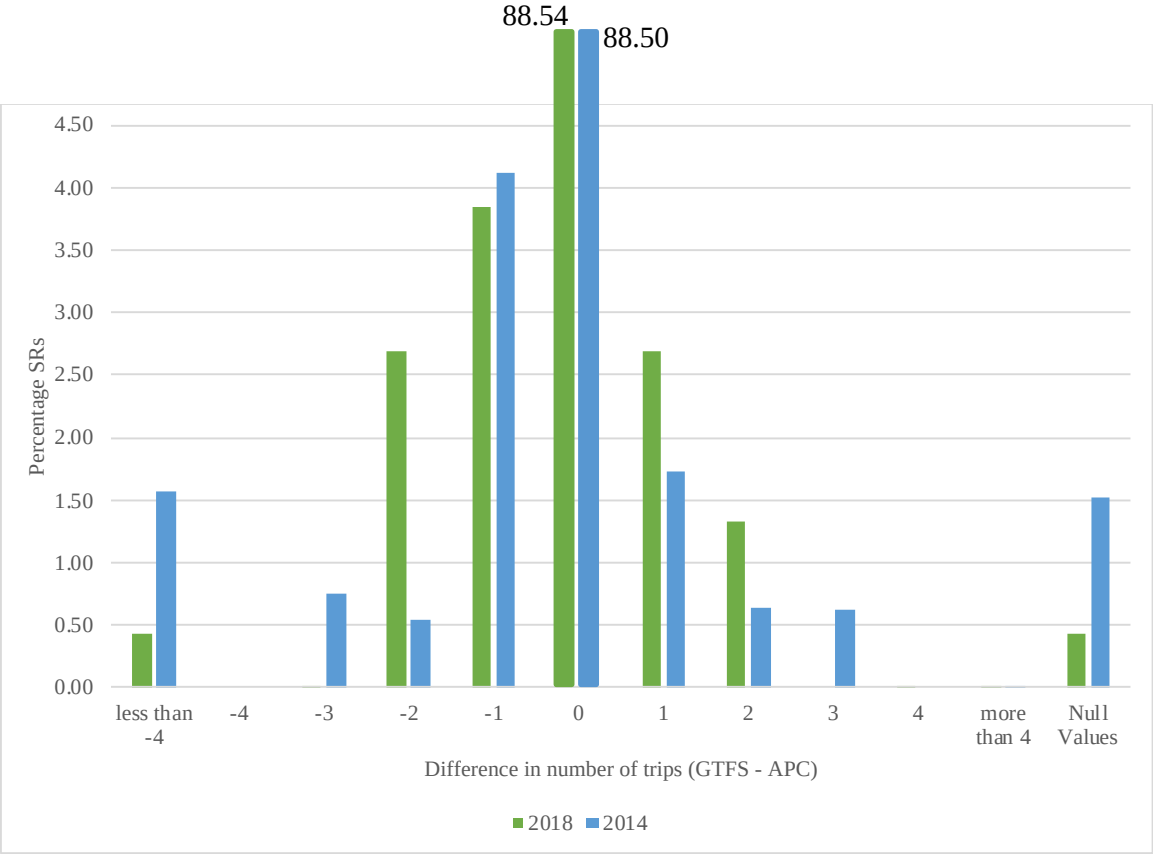

Figure 3 Graphs showing percentages of SR combinations for APC-GTFS trip comparison for 2014 and 2018 for MARTA. The figure shows values for "0" difference off the top of the chart in order to scale up the other values.

4.3 Metro Transit, Minneapolis-St. Paul, Minnesota

Results from initial check shows that Metro Transit's APC had 32,177,288 rows. For numeric columns like On_avg, Off_avg and rte_level_trip_obs, mean, median, maximum and minimum values were calculated to know the spread of the data.

Column	Min	Max	Avg	Median
On_avg	0	80	0.468	0.03
Off_avg	0	78	0.4682	0.08
Rte_level_trip_obs	0	280	27.1368	12

4.3.1 Issues with data and its solutions

Issue 1. Metro's APC has direction as names instead of numbers 0 and 1 which is a standard practice in GTFS. Due to this, there was a challenge in matching up data from GTFS and APC.

Solution. In trips.txt in GTFS there is a column called trip_headsign. For Metro's GTFS, this column had direction name included along with the name of the route. Hence, direction was extracted from trip_headsign using a Wildcard condition that can identify a word from a string of words and assign names to direction_id column instead of default numbers 0 and 1. Hence direction in GTFS was changed to East, West, North and South to match with APC.

The results obtained from 4.1.2 and 4.1.3 are tabulated below. The Table 2 shows percentage of stop-route-direction combinations that have zero, ±1, ±2, ±3, ±4 and more. Figure 4 shows comparison of years 2012 and 2017 for percentage of missing trips in APC or GTFS. Since percentages with difference in number of trips are very low, the figure shows values for "0" difference off the top of the chart in order to scale up the other values. A positive difference means that there are more GTFS trips and APC and vice versa.

Table 2 Percentages of SRD combinations for APC-GTFS trip comparison for Metro Transit

year	2017	2012
less than -4	0.30 %	0.50 %
-4	0.00 %	0.20 %
-3	0.50 %	1.01 %
-2	0.85 %	2.37 %
-1	4.03 %	9.62 %
0	93.60 %	82.03 %
1	0.00 %	3.60 %
2	0.00 %	0.01 %
3	0.00 %	0.62 %
4	0.03 %	0.00 %
more than 4	0.39 %	0.01 %
Null Values	0.32 %	0.03 %
Total	100 %	100 %

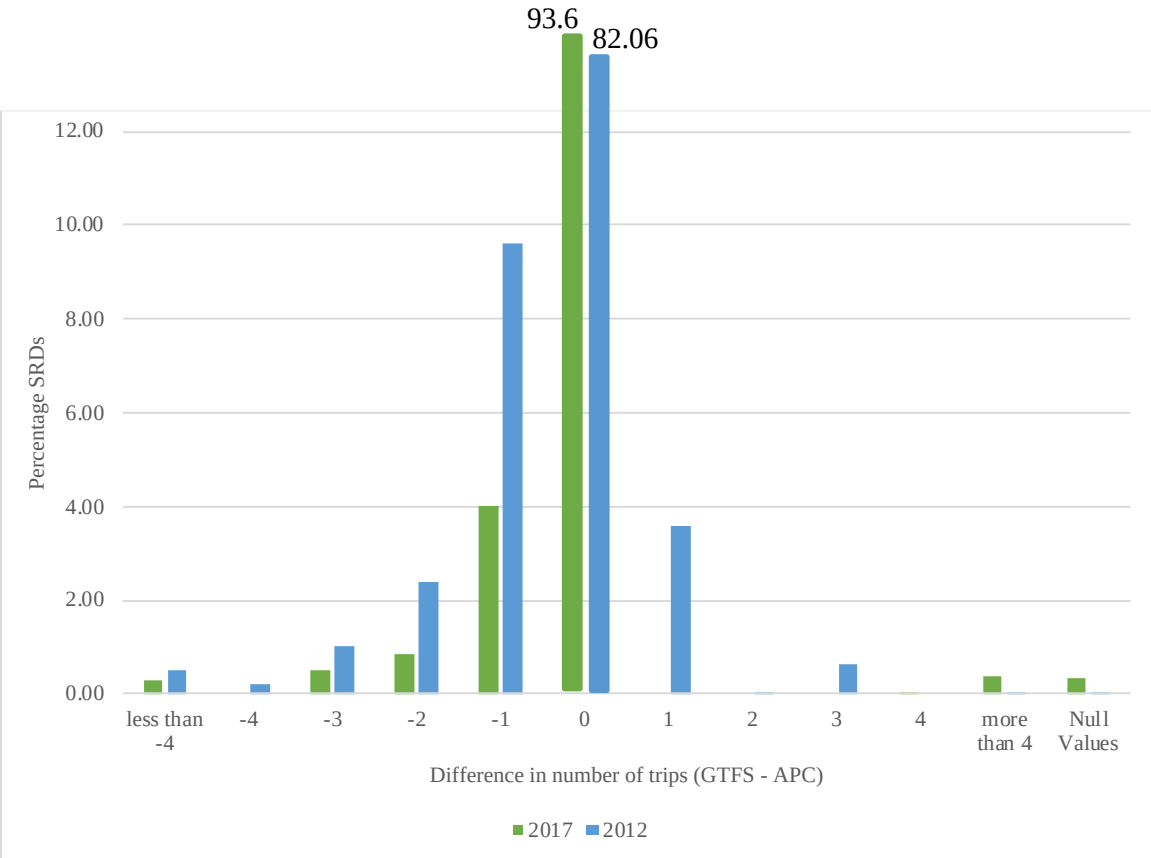

Figure 4 Graphs showing percentages of SRD combinations for APC-GTFS trip comparison for 2012 and 2017 for Metro Transit. The figure shows values for "0" difference off the top of the chart in order to scale up the other values.

4.4 TriMet, Portland, Oregon

Results from initial check shows TriMet's APC had 28,544,818 rows. For numeric columns like ons, offs, estimated_load and trip_obs, mean, median, maximum and minimum values were calculated to know the spread of the data.

Column	max	min	median	avg
ons	390	-36.8	0.18	0.85
offs	250	0	0.2	0.86
estimated_load	322	-148.34	9.55	12.21
trip_obs	69	0	12.0	25.17

4.4.1 Issues with data and its solution

Issue 1. Trip_id in APC had shortened version then that in GTFS. Even though a direct match with trip_id was not used, if GTFS and APC needs to be joined using trip_id, this can cause an issue with the merging of the data.

Solution. It was observed from GTFS's trips.txt that trip_id is a combination of route_id, direction_id, service_id and trip_number in GTFS. Hence a new column was created in APC to concatenate these columns in the specified sequence to obtain trip_id same as GTFS.

The results obtained from 4.1.2 and 4.1.3 are tabulated below. The Table 3 shows percentage of stop-route-direction combinations that have zero, ±1, ±2, ±3, ±4 and more. Figure 5 shows comparison of years 2012 and 2017 for percentage of missing trips in APC or GTFS. Since percentages with difference in number of trips are very low, the figure shows values for "0" difference off the top of the chart in order to scale up the other values. A positive difference means that there are more GTFS trips and APC and vice versa.

Table 3 Percentages of SRD combinations for APC-GTFS trip comparison for TriMet

year	2017	2012
less than -4	0.00 %	0.00 %
-4	0.00 %	0.00 %
-3	0.00 %	0.00 %
-2	0.00 %	0.00 %
-1	0.45 %	0.00 %
0	98.75 %	93.46 %
1	0.00 %	0.91 %
2	0.00 %	0.03 %
3	0.01 %	0.03 %
4	0.00 %	0.01 %
more than 4	0.16 %	4.92 %
Null Values	0.63 %	0.63 %
Total	100.00 %	100.00 %

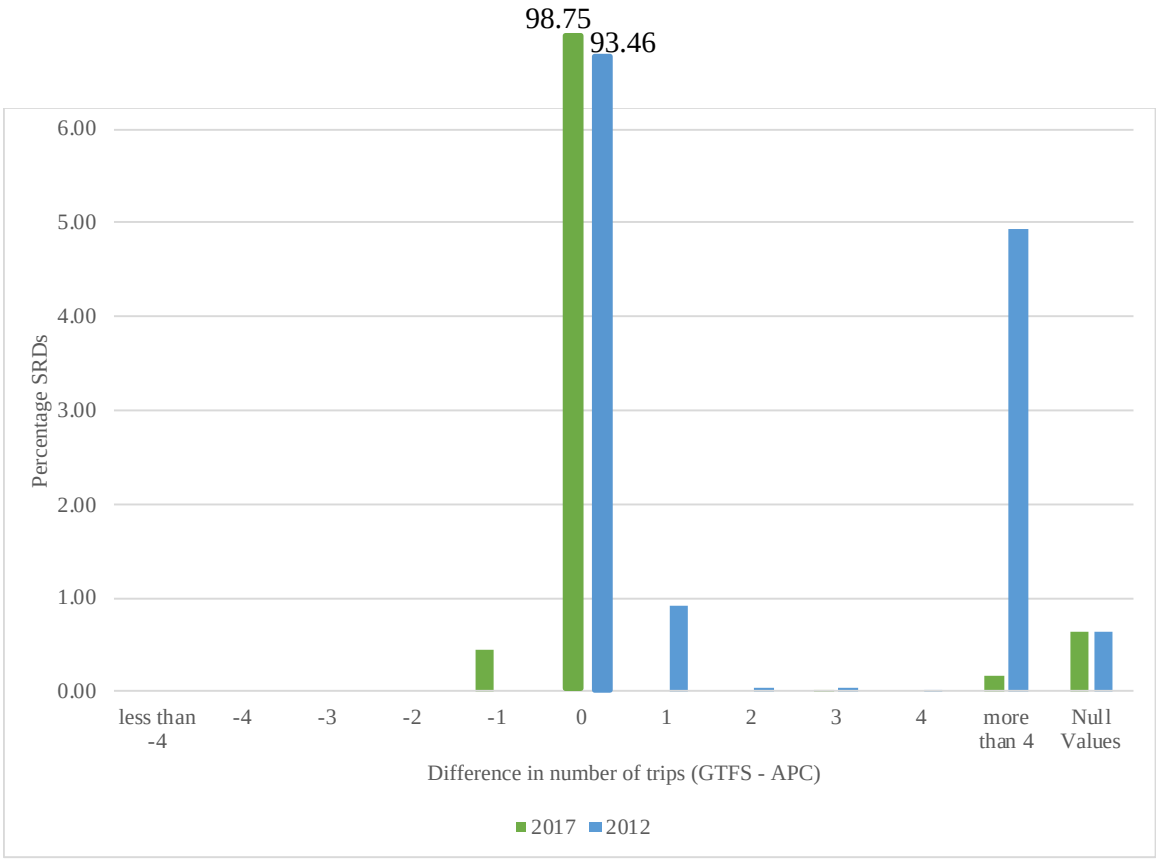

Figure 5 Graphs showing percentages of SRD combinations for APC-GTFS trip comparison for 2012 and 2017 for TriMet. The figure shows values for "0" difference off the top of the chart in order to scale up the other values.

4.5 Miami-Dade Transit, Miami, Florida

Results from initial check shows that MDT's APC had 8,667,932 rows. For numeric columns like QSTOP_AVG_ON, QSTOP_AVG_OFF and N_TRIP_SAMPLES, mean, median, maximum and minimum values were calculated to know the spread of the data.

Column Name	Min	Max	Avg	Median
QSTOP_AVG_ON	0	99	0.52446	0.1
QSTOP_AVG_OFF	0	99	0.52157	0.11
N_TRIP_SAMPLES	1	788	14.3068	8.00

4.5.1 Issues with data and its solution

Issue 1. Route number in APC corresponded to route_short_name in GTFS. This created files with no data when GTFS and APC were merged.

Solution. To overcome this issue, route_id was added to Miami's APC data by matching up with route_short_name from routes.txt in GTFS.

The results obtained from 4.1.2 and 4.1.3 are tabulated below. The Table 4 shows percentage of stop-route-direction combinations that have zero, ±1, ±2, ±3, ±4 and more. Figure 6 shows comparison of years 2013 and 2017 for percentage of missing trips in APC or GTFS. Since percentages with difference in number of trips are very low, the figure shows values for "0" difference off the top of the chart in order to scale up the other values. A positive difference means that there are more GTFS trips and APC and vice versa.

Table 4 Percentages of SRD combinations for APC-GTFS trip comparison for MDT

year	2017	2013
less than -4	0.00 %	0.02 %
-4	0.00 %	0.00 %
-3	0.01 %	0.00 %
-2	0.00 %	0.00 %
-1	0.00 %	0.01 %
0	78.73 %	85.48 %
1	6.85 %	7.00 %
2	5.58 %	5.09 %
3	4.28 %	1.94 %
4	0.50 %	0.14 %
more than 4	3.78 %	0.30 %
Null Values	0.28 %	0.00 %
Total	100.00 %	100.00 %

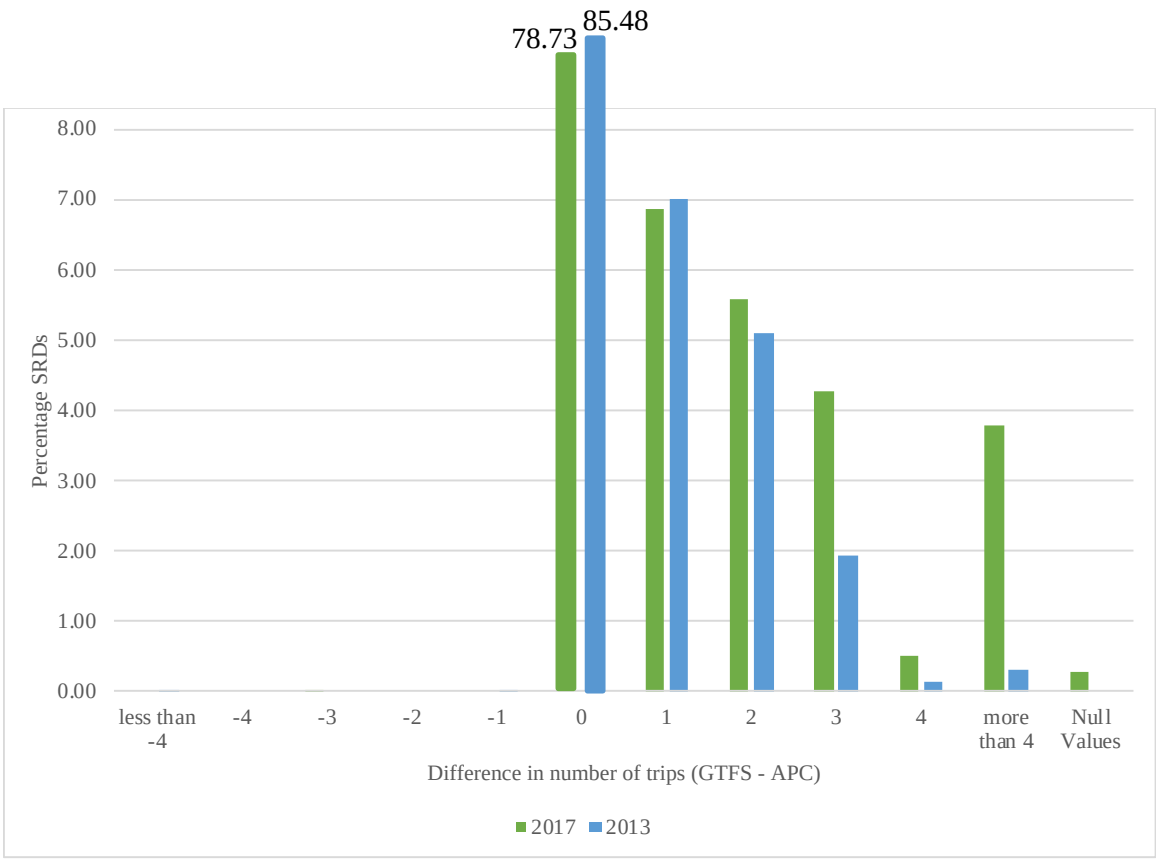

Figure 6 Graphs showing percentages of SRD combinations for APC-GTFS trip comparison for 2013 and 2017 for MDT. The figure shows values for "0" difference off the top of the chart in order to scale up the other values.

CHAPTER 5. RECOMMENDATIONS AND CONCLUSIONS

This chapter summarizes the findings of APC and GTFS datasets of all the transit agencies studied. Each agency had its own limitations with the dataset due to which errors occurred during the data cleaning process discussed in the previous two chapters. From the literature, it is clear that there have been no standard formats for transit agencies to produce their APC to date and hence it makes it difficult to integrate APC with GTFS using a common procedure. Through the processes carried out on APC data and by merging it with GTFS, different types of issues were discovered, and processes were developed to clean the data and make it more useful for further research.

Sometimes, an agency may make changes to their GTFS such as changes in the route_ids for different time periods. For GTFS to be used by researchers or anyone doing analysis across time periods, it is necessary for the agency to report such changes as a part of the metadata so that users can be careful when they do temporal analysis using GTFS. Since GTFS is published by transit agencies who also manage APC data, consistency should be maintained in using the same naming conventions in APC as that in GTFS. For example, it is mandatory to have 0 or 1 in direction_id in GTFS. Transit agencies should try to make sure that the same codes for direction are used in their APC as well or provide information in the metadata about which directions in GTFS correspond to the ones in APC, for example, 0 being Eastbound and 1 being Westbound.

Table 5 summarizes the issues found in the APC and GTFS dataset for each agency. Information that should be added to all APC data is corrected location of stops. Some

agencies had stop location, but it was not calibrated from the erroneous readings as a result of lack of accuracy in GPS devices.

Table 5 Summary of Issues in APC and GTFS by Agency

Agency	Issues
MARTA, Atlanta	• Absence of number of observations indicating sample size • Stop_ids are not consistent in APC and GTFS over time-periods • Route_id in APC correspond to route_short_name in GTFS instead of route_id • Different naming convention of direction_id in APC and GTFS
Metro Transit, Minneapolis-St. Paul	• Different naming convention of direction_id in APC and GTFS • Absence of open data for December 2012 markup period
TriMet, Portland	• Different naming convention of trip_id in APC and GTFS • GTFS dates for markups have overlaps due to error in reporting by the agency
MDT, Miami	• Different naming convention of route_id in APC and GTFS

From GTFS feeds of all agencies, it was observed that transit agencies do not tend to publish most of the optional files mentioned in GTFS guidelines. There is a new standard being introduced called GTFS-ride as mentioned in chapters 1 and 2. GTFS-ride has only one required file, i.e., ride_feed_info.txt which gives information about source and attributes of other ridership files. In this file, there are columns called ride_start_date and ride_end_date representing the start and end date of ridership data. Instead of these fields

being optional, they should be made required in order to know the time-period of the ridership data. Other files added as a part of GTFS-ride include board_alight, trip_capacity, rider_trip and ridership. Each of these files are optional for the feed. From experience with optional files of GTFS, guidelines for GTFS-ride should change and have board_alight as a required file. This file will contain information about boardings and alightings at each stop. Required attributes in this file are trip_id, stop_id, stop_sequence and record_use. Boardings and alightings counts are optional fields. Again, these fields should be marked as required in order to get complete information about passenger movements which is expected from this file. This information can be used to calculate passenger loads and could have a potential use to map out a typical load on a bus for different time-periods in a day.

GTFS-ride should also make trip_capacity as a required file and its fields related to the capacity of bus like seated_capacity, standing_capacity, wheelchair_capacity and bike_capacity should be made required. This information can be used for an interactive user map with color coded icons of transit vehicle that shows availability of capacity of a transit vehicle.

GTFS as a standard has optional files like frequencies.txt and fare_attributes.txt which can give information on headway and fare for different routes in the system. A use of headways and frequencies is on-time performance analysis. But since these files are not optional, they are typically not published by the transit agencies. Since GTFS-ride is in its early stages, lessons learned from GTFS can be used to make sure all essential information is included in the different files and make them required in order to get information in its entirety.

According to TCRP Report 113, *"The transit industry is in the midst of a revolution from being data poor to data rich"* (Hemily, Furth, Strathman, & Muller, 2006). The importance of collecting automated data has been increasing in the transit industry. Clean APC datasets can be used for analysis at higher levels of aggregation as well as disaggregated levels such as at the stop-level. It can be used for demand analysis based on geographic location. For example, trend analysis for boardings can be done based on some specific geographic boundaries or route-segments. Archived APC data can be used to carry out ridership analysis based on time of day or type of day which can be used to make decisions related to service changes in a transit system. Data about number of passengers on a bus can be used to perform passenger crowding analysis which can be used to understand demand on a route. AVL-APC data about headways and passenger crowding can give insights on on-time performance and bus-bunching.

Automated passenger counters can be used during special events that can help make service changes during the event in order to accommodate for change in number of passengers in real-time. If a transit agency makes changes in its service (maybe along a route or an overall system-wide service change), automatic collection systems can be very useful in performing before and after studies to check the impact of service change on ridership.

AVL-APC data can be used for various mapping purposes like showing the level of passenger crowding on a transit vehicle based on either historical or real-time data. This information can help passengers decide to board an upcoming vehicle or wait for another vehicle that is less crowded. Transit agencies can use various maps showing ridership along different routes to make people understand ridership trends along different routes. If an

agency makes changes in its service that results in increased ridership along a route, it can be shown on maps to make more informed and visually appealing results for the general public. AVL-APC data can also be used in various performance measures and level of service measures. Such level of service measures can be found in Transit Capacity and Quality of Service Manual (Parsons Brinckerhoff et al., 2013).

We live in a time where good data is very important to make data-driven decisions. In order to understand changes in transit ridership, the most essential component is good quality data to make informed decisions. This will help transit agencies to not just improve services but also to provide users with information on various new advancements in transit technologies and their potential benefits.

APPENDIX A

A.1 MARTA: Description of columns in APC

Column Name	Description
Stop_id	Stop numbers for each stop in the system
Route	Route number for each route in the system
Stop_name	Stop name associated with each stop number
Day_type	Type of day to which the data corresponds - weekday, Saturday or Sunday
Direction	Direction of every trip (for example, Eastbound, Westbound etc.)
Trip_Start_Time	Start time of each trip that the transit vehicle makes
Ons	Number of boardings at each stop for each trip for each route
Offs	Number of alightings at each stop for each trip for each route
Ons & Offs	Addition of ‘Ons’ and ‘Offs’ columns

A.2 Metro Transit: Description of columns in APC

Column Name	Description
Service_id	Type of day. Metro uses ‘Wk’ for weekday, ‘SAT’ for Saturday and ‘SUN’ for Sunday
Line_id	Route number for each route in the system
Line_Direction	Direction of every trip
Trip_number	Trip number associated with each new trip made by the transit vehicle
Stop_sequence	Sequence of a stop for any given trip
Trip_timepoint_time	The time at which transit vehicle passes a timepoint on its route
Site_id	Stop number of each stop
On_avg	Average number of boardings at each stop for each trip for each route
Off_avg	Average number of alightings at each stop for each trip for each route
Rte_level_trip_obs	Number of observations that were taken to get to the average boardings and alightings data

A.3 TriMet: Description of columns in APC

Column Name	Description
summary_begin_date	Begin date of data collection
service_key	Type of day -- ‘W for weekday, ‘S’ for Saturday and ‘U’ for Sunday
route_number	Route number for each route in the system
direction	Direction of every trip
trip_number	Trip number associated with each new trip made by the transit vehicle
stop_time	Time at which transit vehicle rests at a stop
location_id	Stop number of each stop
public_location_description	Intersection location of stop or intersection closest to stop
ons	Average number of boardings at each stop for each trip for each route
offs	Average number of alightings at each stop for each trip for each route
estimated_load	Number of passengers on transit vehicle at the time of a stop
trip_obs	Number of observations that were taken to get to the average boardings and alightings data
x_coord	Latitude of a stop location
y_coord	Longitude of a stop location

A.4 MDT: Description of columns in APC

Column Name	Description
qstopdayofwk	Type of day - ‘1’ for weekdays, ‘2’ for Saturday and ‘3’ for Sunday
qstoproute	Route number for each route in the system
qstopdir	Direction of every trip
qstoptrip	Start time of teach trip made by the transit vehicle
qstopblock	Block number for each transit vehicle and driver
qstopseqstop	Sequence of a stop for any given trip
qstopqstop	Stop number of each stop
qstopname	Stop name of each stop
qstop_avg_on	Average number of boardings at each stop for each trip for each route
qstop_avg_off	Average number of alightings at each stop for each trip for each route
n_trip_samples	Number of observations that were taken to get to the average boardings and alightings data
qstop_avg_lat	Latitude of a stop location
qstop_avg_long	Longitude of a stop location
qstop_calibrated_lat	Corrected latitude of a stop location for errors in reported latitude
qstop_calibrated_long	Corrected longitude of a stop location for errors in reported longitude

REFERENCES

Attanucci, J., & Vozzolo, D. (1983). Assessment of Operational Effectiveness, Accuracy, and Costs of Automatic Passenger Counters. *Transportation Research Record 947*.

Boyle, D. K. (2009). *TCRP Synthesis 77: Passenger Counting Systems*.

Carleton, P., Hoover, S., Fields, B., Barnes, M., & Porter, J. D. (2019). GTFS-ride A UNIFYING STANDARD FOR FIXED-ROUTE RIDERSHIP DATA.

Chu, X. (2010). *A Guidebook for Using Automatic Passenger Counter Data for National Transit Database (NTD) Reporting*: National Center for Transit Research at CUTR, University of South Florida.

Dickens, M. (2018). *Quarterly and Annual Totals by Mode*. Retrieved from https://www.apta.com/resources/statistics/Documents/APTA-Ridership-by-Mode-and-Quarter-1990-Present.xlsx

Federal Highway Administration, F. (2018). U.S. Department of Transportation - Federal Highway Administration. *Travel Monitoring - Traffic Volume Trends.* Retrieved from https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm

Furth, P. G., Strathman, J. G., & Hemily, B. (2005). Making Automatic Passenger Counts Mainstream: Accuracy, Balancing Algorithms, and Data Structures. *Transportation Research Record*(1927(1)), 206-216. Retrieved from https://doi.org/10.1177/0361198105192700124.

Google Open Source. (2019a). Open Issues. Retrieved from https://github.com/google/transit/issues

Google Open Source. (2019b). Open proposals. Retrieved from https://github.com/google/transit/pulls

Hemily, B., Furth, P. G., Strathman, J. G., & Muller, T. H. J. (2006). *TCRP Report 113 Using Archived AVL-APC Data to Improve Transit Performance and Management*(pp. 93). Retrieved from http://www.trb.org/Publications/Blurbs/156999.aspx doi:10.17226/13907

Hodges, C. C. (1988). *Automatic Passenger Counter Systems : The state of the practice*. Retrieved from Washington, D.C.: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjvtpjCptzhAhVPdt8KHajEBj4QFjAAegQIABAC&url=https%3A%2F%2Frosap.ntl.bts.gov%2Fview%2Fdot%2F398%2Fdot_398_DS1.pdf%3F&usg=AOvVaw3emDTQo-HdQzQumMR4i15U

Kimpel, T. J., Strathman, J. G., Griffin, D., Callas, S., & Gerhart, R. L. (2003). Automatic Passenger Counter Evaluation : Implications for National Transit Database Reporting. *Transportation Research Record*(1835). Retrieved from http://dx.doi.org/10.3141/1835-12.

Metro Transit. (2019). Metro Transit 2018 Facts. Retrieved from https://www.metrotransit.org/metro-transit-facts#Rides,%20Miles%20and%20Routes

Miami-Dade Transit. (2018). *TRANSIT DEVELOPMENT PLAN ANNUAL UPDATE*. Retrieved from https://www.miamidade.gov/transit/library/pdfs/misc/2019-tdp-annual-plan.pdf

Minnesota Metropolitan Council. (2014). METRO TRANSIT METRO BLUE/GREEN LINE LIGHT RAIL VEHICLE (LRV) AUTOMATED PASSENGER COUNTER. 11.

National Center for Transit Research. (2011). Expanding the Google Transit Feed Specification to Suport Operations and Planning. 64.

Office of Transit System Planning, M. (2017). *MARTA FY 2017 Service Standards*. Retrieved from https://www.itsmarta.com/uploadedFiles/SERVICE_STANDARDS%20_FY_2017_Final.pdf

Parsons Brinckerhoff, Kittelson & Associates Inc., KFH Group Inc., Texas A&M Transportation Institute, & Arup. (2013). *Transit Capacity and Quality of Service Manual, Third Edition* (978-0-309-28344-1). Retrieved from

Pi, X., Egge, M., Whitmore, J., Silbermann, A., & Qian, Z. (2018). Understanding Transit System Performance Using AVL-APC Data: An Analytics Platform with Case Studies for the Pittsburgh Region. *Journal of Public Transportation, 21*(2), 19-40. doi:https://doi.org/10.5038/2375-0901.21.2.2

Roth, M. (2010). How Google and Portland's TriMet Set the Standard for Open Transit Data. Retrieved from https://sf.streetsblog.org/2010/01/05/how-google-and-portlands-trimet-set-the-standard-for-open-transit-data/

TriMet. (2018). *TriMet At-A-Glance*. Retrieved from https://trimet.org/ataglance/trimet-at-a-glance-2018.pdf