Leveraging Smart Card Intelligence to Improve Transit Planning: A Case Study of PRESTO in the Greater Toronto and Hamilton Region

By

Tony Zhuang

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science

Department of Civil & Mineral Engineering
University of Toronto

© Copyright by Yan Zhuang 2018
Leveraging Smart Card Intelligence to Improve Transit Planning: A Case Study of PRESTO in the Greater Toronto and Hamilton Region

Tony Zhuang

Master of Applied Science
Department of Civil & Mineral Engineering
University of Toronto
2018

Abstract

While smart cards are primarily designed to automate fare collection, the transaction records contain rich intelligence. This thesis demonstrates the capabilities of smart card analytics in the Greater Toronto and Hamilton Area (GTHA), using data from the PRESTO smart card system. The case studies involve Hamilton Street Railway (HSR), GO Transit (GO), Burlington Transit, and the Toronto Transit Commission (TTC). Methods are demonstrated for: geocoding PRESTO transactions, reconstructing trips, visualizing passenger flows, spatial clustering of passenger groups, computing route level transfer volume matrices, and quantifying transfer delays. The results underscore PRESTO’s superior capability in providing longitudinal transit demand profiles, which can be leveraged to develop enhanced Key Performance Indicators (KPIs) that measure the transit travel experience from the lens of passengers. As PRESTO cards will become ubiquitous in the GTHA, this study addresses the pressing need to demonstrate the value of PRESTO analysis to improve the practice of transit planning.
Acknowledgements

First, to my co-supervisors, Professor Eric J. Miller and Professor Amer Shalaby, it has been a real pleasure being your student. My sincere gratitude goes out to both of you for the ongoing support in the last two years. I still remember that we arrived at this thesis topic of PRESTO analytics within minutes of our very first meeting back in September 2016. I was immediately intrigued and have never regretted since. Even though the last two year has not always been smooth sailing, never for a moment have I not enjoyed working on this thesis.

I would like to thank my former Supervisor of Planning and Schedule at Burlington Transit, Garth Rowland, for agreeing to provide me with full access to Burlington Transit’s data that kick-started the analytics of this thesis. My gratitude also goes out to Steve Lucas, Transit Planner/Analyst at Burlington Transit, for helping me extract and interpret the data. I would also like to thank TTS 2.0, OGS, NSERC, Trapeze, OCE, and SOSCIP for funding this research, as well as Burlington Transit, Metrolinx, and Hamilton Street Railway for providing the data.

To my close friends and what I like to refer to as the elite GIS strike team from the University of Waterloo, Christina Zhang, Elkan Wan, Andus Chen, and Alvin Fan, thank you for all your support and for always being a phone call (or a Facebook message) away to offer me assistance whenever I ran into difficulties. To my parents as well, thank you for all your support in the past two years.

Finally, to the University of Toronto’s StarCraft II Team, it has been a real pleasure being your team lead for the past 2 years. I truly enjoyed sharing our passion for the game together and I am grateful for the support from you all to re-establish the team. During times of loneliness and anxiety, I am very grateful to be accompanied virtually via voice chats and in person.
# Table of Contents

Acknowledgements ........................................................................................................... iii
Table of Contents ............................................................................................................... iv
List of Tables ....................................................................................................................... v
List of Figures ...................................................................................................................... vi
1 Introduction ..................................................................................................................... 1
   1.1 Background and Motivation ...................................................................................... 1
   1.2 Research Questions and Objectives ......................................................................... 1
   1.3 Overview of Smart Card Potentials and Past Research ........................................... 2
2 Data Description and Case Study Overview ................................................................... 4
   2.1 PRESTO Smart Card System Overview ................................................................... 4
   2.2 Hamilton Street Railway (HSR) ............................................................................... 6
   2.3 GO Transit (GO) ...................................................................................................... 6
   2.4 Burlington Transit ................................................................................................... 7
   2.5 Toronto Transit Commission (TTC) ......................................................................... 7
3 Estimating and Visualizing Passenger Flows ................................................................. 8
   3.1 Geocoding PRESTO Transactions with Historical CAD-AVL Data ....................... 8
   3.2 Trip Reconstruction – Interpolating Alighting Location .......................................... 12
   3.3 Development of a New Zonal System for Public Transit O-D Matrices ................... 15
   3.4 Use of Fare Zones to Aggregate O-Ds and Visualize Passenger Movements .......... 18
4 Enhancing Key Performance Indicators with User-Based Metrics ................................. 20
   4.1 Computing Transfer Volume Matrix ........................................................................ 20
   4.2 Informing Service Span Changes with Assessment of User Impacts ....................... 24
   4.3 Application of Spatial Clustering to Visualize Travel Patterns ................................. 28
   4.4 Quantifying Travel Time Variations and Transfer Delays ....................................... 36
5 Conclusion, Reflections, and Future Opportunities with PRESTO ................................. 44
   5.1 Summary of Main Findings ...................................................................................... 44
   5.2 Reflection on Software Tools .................................................................................. 46
   5.3 Reflection on Smart Card Data’s Strategic Importance ........................................... 46
   5.4 Future Research Opportunities with PRESTO ........................................................ 48
6 References ....................................................................................................................... 49
List of Tables

Table 5-1: Trip reconstruction results ........................................................................................................14
Table 6-1: Transfer volume to and from Route 1 and Route 101, Sept 2017 to Dec 2017 .............23
Table 6-2: Boarding and unique rider counts before and after 9:00PM .............................................27
Table 6-3: Profiles of riders before and after 9:00PM ...........................................................................27
Table 6-4: Sample of classified boardings by home transit agency and TTC Stop ID .................29
List of Figures

Figure 2-1: Municipalities in the Greater Toronto and Hamilton Area (GTHA) .........................4
Figure 3-1: Geocoding PRESTO transactions using historical AVL data.................................10
Figure 3-2: Map of geocoded HSR PRESTO boarding..........................................................11
Figure 3-3: Near analysis concept ............................................................................................13
Figure 3-4: Boundaries of Traffic Analysis Zones (TAZs) along arterial roads in Hamilton ......15
Figure 3-5: Consolidation of timing points within 200 metres of each other .........................16
Figure 3-6: Transit Analysis Zones for HSR service area .......................................................17
Figure 3-7: GO Transit one-way O-D flows grouped by fare zones........................................19
Figure 4-1: Route 1 and Route 101 Express, Burlington Transit.............................................21
Figure 4-2: Existing Burlington Transit service span ...............................................................25
Figure 4-3: Proposed coverage service span concept ...............................................................25
Figure 4-4: Proposed midpoint and ridership service span concept .......................................26
Figure 4-5: Clusters of Mi-way riders in Toronto.................................................................30
Figure 4-6: Clusters of Brampton Transit riders in Toronto..................................................31
Figure 4-7: Clusters of Durham Region Transit (DRT) riders in Toronto...............................32
Figure 4-8: Clusters of York Region Transit (YRT) riders in Toronto ....................................33
Figure 4-9: Clusters of Oakville Transit riders in Toronto......................................................34
Figure 4-10: Clusters of Burlington Transit riders in Toronto................................................34
Figure 4-11: Clusters of Hamilton Street Railway (HSR) riders in Toronto............................35
Figure 4-12: Summary of Transfer Time Estimation...............................................................37
Figure 4-13: Boxplots of estimated transfer time by the frequency of use (Route 85)..............39
Figure 4-14: Median bus to bus transfer times by stops, TTC, September 2017......................40
Figure 4-15: Histogram of median transfer times by Stops, TTC, September 2017.................40
Figure 4-16: Hot spot analysis of median transfer times .........................................................41
Figure 4-17: Hot spot analysis of major non-barrier-free bus-to-bus transfer points ...............42
1 Introduction

1.1 Background and Motivation

In the Greater Toronto and Hamilton Area (GTHA), urban sprawl has largely prevailed since World War II. For decades, the private automobile seems to be the ideal means of travel for mobility. However, as the region continues to grow, road capacity becomes increasingly scarce. Over the past decade, the Ontario government has adopted a series of land use policies to contain sprawl and promote intensification (Ministry of Municipal Affairs, 2018). In addition, considerable transit investments were made with the adoption of many new technologies. These new technologies include Computer Aided Dispatch and Automatic Vehicle Locations (CAD-AVL), Automatic Passenger Counts (APC), and Automatic Fare Collections (AFC). The adoption of these new technologies creates an unprecedented amount of data that can be tapped for analysis.

Smart cards are devices designed to store and process data (Hickman, 2017). As a form of AFC, smart cards have the size of credit cards, which make them suitable for many applications that involve identification, authorization, and payment. The implementation of smart cards in transit systems has been primarily for fare collection. However, smart card data contains rich intelligence that can be leveraged by planners to inform decision making. PRESTO is a contactless smart card system using Radio-frequency identification (RFID) technology, developed by the firm Accenture, for Metrolinx to automate fare collection and advance fare integration in the GTHA (Metrolinx, 2018a). With the only exception of Milton Transit, every single transit agency in the GTHA uses PRESTO, and all transit agencies with PRESTO have eliminated, or are planning to eliminate, paper tickets and physical passes. Therefore, in the next few years, the overwhelming majority of transit trips will be taken using PRESTO cards. The objective of this thesis is to demonstrate the capabilities of PRESTO analytics in the GTHA to improve transit planning.

1.2 Research Questions and Objectives

The broad research question of this thesis is:

- As PRESTO cards are about to become ubiquitous in the GTHA, what can we do with this huge volume of user data that PRESTO generates?
This study presents a toolbox of methods to process and analyze PRESTO data that could enhance transit service planning. After an overview of the PRESTO data and study areas (Chapter 2), the study is grouped into two parts: estimating and visualizing passenger behaviour (Chapter 3) and improving transit analysis with user-based metrics (Chapter 4).

Specific research objectives involve:

- Geocoding PRESTO card transactions using historical vehicle locations (Section 3.1)
- Interpolating alighting locations (Section 3.2)
- Delineating a new zonal system for transit origins and destinations (O-Ds) (Section 3.3)
- Aggregating passenger flows with fare zones (Section 3.4)
- Computing transfer matrix (Section 4.1)
- Assessing user base impacts of changes in service span (Section 4.2)
- Applying spatial clustering to visualize travel patterns (Section 4.3)
- Quantifying transfer delays and assessing transfer quality (Section 4.4)

Considering that PRESTO is used by 9 local transit systems in the GTHA, analyzing every transit agency’s PRESTO data in depth would make the research scope prohibitively large. Therefore, this thesis focuses on four transit agencies: Hamilton Street Railway (HSR), Burlington Transit, GO Transit, and the Toronto Transit Commission (TTC). Where applicable, this thesis demonstrates more streamlined processes and methods that make use of GIS software. This thesis also evaluates the existing state of practice in transit service planning and identify the additional insights we could extract from PRESTO. Improvements over more traditional data sources, such as farebox, AVL, and APC, are highlighted.

1.3 Overview of Smart Card Potentials and Past Research

Unlike records from conventional farebox units, passenger cordon counts, and automated passenger counts, smart cards allow for the identification of individual transit users. Planners and analysts can track the activities of transit riders throughout a period of time. As summarized by Pelletier, Trépanier, & Morency (2011), existing smart card research can be categorized into: 1) strategic studies, 2) tactical studies, and 3) operational studies.
First, strategic level studies involve the understanding of passenger behaviour for long-range planning initiatives such as forecasting and marketing. This level of research is focused on user classification, as personal information on smart card users is generally not available. One example is the analysis of fare loyalty programs, where the usage pattern of riders is analyzed to make recommendations on fare policy (Trepanier & Morency, 2010).

Second, tactical level studies refer to service adjustments and network designs where the travel patterns are analyzed to adjust transit service. Examples in this category include the estimation of origin and destination (O-D) matrix (Alsger, Mesbah, Ferreira, & Safi, 2015) and transfer journey analysis (Gordon, Koutsopoulos, Wilson, & Attanucci, 2013). In some transit systems, passengers “tap on” as they board the transit vehicle, and “tap off” at their destination stop. This allows the smart card transaction to record both the origin and destination locations of the passenger. Even for bus systems that only require a “tap on,” the destination location can be estimated using the passenger’s travel history (Trépanier, Tranchant, & Chapleau, 2007).

Third, operational studies use Key Performance Indicators (KPIs) to evaluate the transit network, where passenger experience and behaviour are broken down in detail, which could include on-time performance, schedule adherence, and transfer wait times. For example, Uniman, Attanucci, Mishalani, & Wilson (2010) explore the potential of using smart card data to quantify the reliability of London Underground as experienced by passengers, Jang (2010) demonstrates methods to query travel times and transfer times in Seoul, South Korea, and Zhao, Wang, Woodburn, & Ryerson (2017) provide methods for identifying poor performing transfer nodes between subways and buses in Nanjing, China.

Overall, however, most existing smart card studies tend to involve rather complex mathematical models and equations that may be difficult for transit agencies to apply. As articulated in Hickman (2017), “there remains a pressing need to demonstrate the value of smart card analytics to improve the practice of transit planning.” This study attempts to address the need to demonstrate such value, through a toolbox of practical methods that transit agencies can apply. At the time of writing, there has yet been a systematic study of PRESTO data in the GTHA.
2 Data Description and Case Study Overview

2.1 PRESTO Smart Card System Overview

The Greater Toronto and Hamilton Area (GTHA) includes the City of Toronto, York Region, Durham Region, Peel Region, Halton Region, and the City of Hamilton (Figure 2-1).

Figure 2-1: Municipalities in the Greater Toronto and Hamilton Area (GTHA)\(^1\)

There is a total of 9 local transit agencies: Toronto Transit Commission (TTC), York Region Transit (YRT), Durham Region Transit (DRT), Brampton Transit, Mississauga Transit (MiWay), Milton Transit, Oakville Transit, Burlington Transit, and Hamilton Street Railway (HSR).

\(^1\) Base Map Sources: Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community
Regional transit is operated by Metrolinx through GO Transit (GO). Metrolinx was created in 2006 as an arms-length Crown agency to manage and integrate public transit in the GTHA. One of Metrolinx’s mandates is to implement the PRESTO smart card system across the GTHA transit agencies (Metrolinx, 2018a).

This study uses PRESTO usage transaction data provided by Metrolinx for the period between September 1, 2017 and December 31, 2017, covering all transit agencies in the GTHA that uses PRESTO. Each record in the dataset represents a usage transaction (i.e. a tap). For each transaction record, the dataset contains several attributes including:

- Farecard unique ID
- Date and timestamp of the transaction (in minutes, or in seconds for TTC)
- Farecard transaction sequence (a number which counts from 0 to 4095 and then resets)
- Service provider (i.e., transit agency code)
- PRESTO device unique ID and type
- Vehicle unique ID (only applicable to bus trips)
- Route number (only applicable to bus trips)
- Transaction location (location where the transaction took place)
- Tap type (i.e., tap on, tap off, default tap on)
- Card type (i.e., adult, student, child)
- Transaction status (i.e., 0 = valid, 1 = void, 2 = invalid);
- Farecard balance; and
- Payment or payment amount.

For the 905 transit agencies2, the transaction location field of each PRESTO transaction is derived from a time-series that tags the transaction to the scheduled upstream timing point – assuming that the vehicle is 100% on time. At the time of this writing, only GO Transit and TTC’s PRESTO units are GPS-enabled, which means that the PRESTO transactions from 905 transit agencies need to be geocoded using historical CAD-AVL data to obtain their boarding locations. In addition, only GO Transit currently uses a zonal fare system, which necessitates passengers to perform a tap

---

2 “905 transit agencies” refers to neighbouring transit agencies around the City of Toronto, which include York Region Transit (YRT), Durham Region Transit (DRT), Brampton Transit, Mississauga Transit (MiWay), Milton Transit, Oakville Transit, Burlington Transit, and Hamilton Street Railway (HSR)
off when alighting a transit vehicle. All other transit agencies in the GTHA uses a flat fare, so only a tap on is required when the passenger boards a vehicle. Without a tap off requirement, the destination of each trip is unknown.

2.2 Hamilton Street Railway (HSR)

The City of Hamilton is a single-tier municipality located southwest of Toronto, with a population of 536,917, the ninth largest in Canada. Hamilton Street Railway (HSR) is the local transit agency, operating a fleet size of 251 buses, 34 routes, and an annual ridership of 21.49 million revenue passengers in 2016 (Canadian Urban Transit Association, 2017). At the time of writing, the regular adult fare is $2.30, the lowest in the GTHA and the fare is reduced to $0.70 if passengers transfer to and from GO Transit Services. Passengers can connect for free to Burlington Transit without paying a second fare, and around 20% of Burlington Transit’s riders come from Hamilton, mainly through the Plains Road corridor (Jarret Walker + Associates, 2017). According to data provided by HSR staff, HSR’s PRESTO adoption rate is 36% for September 2017. Hamilton is home to McMaster University and Mohawk College, each with over 30,000 in student population and both institutions have a U-Pass program that requires every student to purchase a discounted transit pass as part of their student fees (McMaster, 2018; Mohawk, 2018). At the time of this study, only McMaster University students’ U-Pass are programmed onto PRESTO cards, while Mohawk College’s U-Pass takes the form of a sticker affixed on their student cards. The HSR case study demonstrates methods to geocode PRESTO transactions using historical vehicle locations and reconstruct trips.

2.3 GO Transit (GO)

Government of Ontario Transit (GO) is a regional transit system operated by the arms-length provincial agency, Metrolinx. GO operates a radial network of commuter rail and regional bus service that converges towards Union Station, Toronto. GO uses a zonal fare system where fares vary with respect to the number of zones the passenger has travelled, which necessitates a tap off when passengers alight a transit vehicle. The tap-off requirement allows GO to have very complete O-D information of its passengers. GO also has the highest PRESTO adoption rate of 84.8% (Metrolinx, 2017). In this study, GO is used to demonstrate methods to aggregate and visualize O-D flows by fare zones.
2.4 Burlington Transit

The City of Burlington is a lower-tier municipality under Halton Region, with a population of 183,314, located west of Oakville and east of Hamilton. Burlington Transit is the second smallest transit agency in the GTHA, with an annual ridership of 1.90 million in 2017. According to transit staff at Burlington, Burlington Transit’s PRESTO adoption rate is 73% in 2017, as monthly passes are now only available via PRESTO and the price of paper tickets have become more expensive than PRESTO fares. The PRESTO adoption rate is expected to increase further with the planned elimination of paper tickets in the coming years. This study leverages Burlington Transit’s high adoption rate and uses Burlington Transit’s PRESTO data to compute route-level transfer matrices and user-based analysis of service span adjustments.

2.5 Toronto Transit Commission (TTC)

The Toronto Transit Commission (TTC) is the largest transit agency in the GTHA, carrying an annual ridership of over 538 million with a network of buses, streetcars, and subways. As the last transit agency to adopt PRESTO, the TTC completed the installation of PRESTO devices in late 2016. In 2017, the PRESTO adoption rate was 14.5% with tokens and monthly passes still in circulation at the time of writing; however, tokens and monthly passes will all be phased out in two years (TTC, 2018). Unlike 905 transit agencies, the TTC’s PRESTO devices are GPS-enabled. Therefore, each transaction contains geographic locations. The TTC case study uses hot spot analysis to visualize the travel patterns of riders from neighbouring transit agencies and analyze transfer delays.
3 Estimating and Visualizing Passenger Flows

Smart card usage records are a passive data source for passenger travel. Besides the primary function of fare collection, smart card transactions produce detailed transaction data that could be used in different transportation planning applications (Park, Kim, & Lim, 2008). One of the most common applications is to understand public transport users’ origins and destinations (O-Ds), and their utilization of the transit networks (Reddy, Lu, Kumar, Bashmakov, & Rudenko, 2009). The readiness of the smart card transactions, however, depends both on how integrated the fare collection system is with the Computer Aided Dispatch / Automatic Vehicle Location (CAD-AVL) system and whether the transit system requires passengers to tap-off when alighting a transit vehicle. For transit agencies where the smart card devices are not GPS-enabled, the smart card transactions need to be geocoded using historical vehicle location data from CAD-AVL systems. With geographic coordinates, transaction timestamps, and the ability to track individual passengers through the unique card numbers, it is then possible to infer the location of alighting for tap-on only systems. Then, the resultant O-Ds need to be aggregated into a suitable zonal system to summarize travel patterns.

Using Hamilton as the case study, this section demonstrates methods to geocode PRESTO smart card transactions using historical CAD-AVL data and estimate alighting locations. Since traditional Traffic Analysis Zones (TAZs) tend to split along major arterial roads, a novel zone system is created to aggregate individual PRESTO O-Ds to ensure that the inbound stops and outbound stops of a transit route are not split into two different zones. Furthermore, this section presents a case study of GO Transit, to test the use of fare zones to visualize O-Ds.

3.1 Geocoding PRESTO Transactions with Historical CAD-AVL Data

Due to the lack of accurate GPS locations of PRESTO transactions, the PRESTO transactions need to be first geocoded using historical CAD-AVL data. The HSR case study involves PRESTO transactions from September 18, 2017, to September 29, 2017, as well as historical CAD-AVL data that covers the same period as the PRESTO transactions. Each PRESTO transaction contains a field called “VehicleID,” which is a tag permanently attached to each physical PRESTO unit. The PRESTO unit, and by extension the PRESTO VehicleID field, stays with the bus that it is installed on and is associated with the Bus ID until the specific PRESTO Unit is removed and
installed onto another bus. Metrolinx provided a vehicle mapping table that matches each PRESTO device numbers to the HSR bus numbers the unit is installed on. This enables the capability to look up the geographic location of each PRESTO transaction using historical CAD-AVL data.

Each row in the CAD-AVL data represents one arrival and departure event of a bus at a specific stop, for a specific trip. The historical CAD-AVL data includes the following fields:

- Bus ID
- Scheduled Arrival Time
- Actual Arrival Time
- Scheduled Departure Time
- Actual Departure Time
- Route Number
- Trip Number
- Stop Number

The geocoding process is as follows:

1. Through the vehicle mapping table, each PRESTO transaction can be associated with a physical vehicle in the bus fleet of HSR.
2. Both the PRESTO transactions and CAD-AVL data were imported into a SQLite database, and then a range-based table join is performed using the PRESTO transaction time (TrxTime) to the Actual Arrive Time of HSR vehicle.
3. If the PRESTO transaction time is between the Actual Arrival Time and one minute after the Actual Arrive Time of a specific bus stop, then that specific PRESTO transaction will be geocoded to that bus stop. Departure times are not used to account for passengers who may tap the PRESTO card after the door is closed.
4. As TrxTime is only precise to the minute, while CAD-AVL data is precise to the seconds, there are instances where each PRESTO transaction is matched to multiple bus arrival events. In the event of one to many matches, one event is selected at random with the rest discarded.
Figure 3-1 provides a summary of the geocoding process. Using this method, 90% of the PRESTO transactions can be geocoded, which yields a total of 425,892 PRESTO transactions. Figure 3-2 maps the results spatially across Hamilton. In line with the expectation that there would be significant transit demand around post-secondary institutions, there is a considerable concentration of boarding at McMaster University and Downtown Hamilton. What is missing, however, is Mohawk College. At the time of this study, students at Mohawk College have yet to move their U-Pass program onto the PRESTO cards.

**Figure 3-1: Geocoding PRESTO transactions using historical AVL data**

<table>
<thead>
<tr>
<th>PRESTO fields include:</th>
<th>AVL fields include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Unique Card ID</td>
<td>• AVL Vehicle Number</td>
</tr>
<tr>
<td>• PRESTO Vehicle ID</td>
<td>• Route Number</td>
</tr>
<tr>
<td>• AVL Vehicle Number (looked up from a table provided by Metrolinx)</td>
<td>• Trip Number</td>
</tr>
<tr>
<td>• Transaction Date &amp; Time</td>
<td>• Stop Location</td>
</tr>
<tr>
<td></td>
<td>• Actual Arrival Time</td>
</tr>
<tr>
<td></td>
<td>• Actual Arrival Time + 1 mins</td>
</tr>
<tr>
<td></td>
<td>• Actual Departure Time</td>
</tr>
<tr>
<td></td>
<td>• Scheduled Arrival Time</td>
</tr>
<tr>
<td></td>
<td>• Scheduled Departure Time</td>
</tr>
</tbody>
</table>
Figure 3-2: Map of geocoded HSR PRESTO boarding

3 Base Map Sources: Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community
3.2 Trip Reconstruction – Interpolating Alighting Location

After each transaction is geocoded, additional logic is required for estimating the destination of passengers’ journeys. Several researchers have used a few assumptions to estimate users’ travel patterns and produce origin and destination matrices for tap-on only systems. For example, Barry, Newhouser, Rahbee, and Sayeda (2002) presents one of the first studies for tap-on only systems that focused on estimating station-to-station origin and destination (O-D) tables for the New York City subway system. In their study, they introduced a methodology that is based on two assumptions to infer a destination station for each user’s boarding. These assumptions are inspired by the continuity of daily travel as well as the circularity of daily journey chains. More specifically, the first assumption is that in most of the cases the destination of a passenger journey is the origin of the next journey. The second assumption is that in most of the cases, the final daily destination of a passenger is the first daily origin. The aforementioned two assumptions were extensively used by researchers afterwards throughout the literature. For example, Pelletier et al., (2011) used a similar approach to estimate bus service users’ destinations in Gatineau, Québec, along with a few other researchers (Devillaine, Munizaga, & Trépanier, 2012; Munizaga et al., 2014; Nunes et al., 2016).

For this study of HSR, only passengers who made more than one trip per day on HSR is considered. If the passenger only made one boarding in HSR on a given day, the trip is classified as an unlinked trip and excluded from the study. Hence, as PRESTO adoption rises, the share of unlinked trips is expected to fall drastically. While He & Trépanier (2015) proposed a method to estimate the alighting location of unlinked trips through looking up a similar transaction with similar spatial and temporal profiles, the method can only account for a small fraction of the overall sample and have not been repeated by any other study. Consequently, the study performs trip reconstruction using the geocoded PRESTO transactions for linked trips only, with the primary intent of providing a proof of concept in anticipation of the ubiquitous adoption of PRESTO. Two main assumptions are applied, similar to Trépanier, Tranchant, & Chapleau, (2007b). The two assumptions are:

1. If the transaction is not the last boarding during the day, the stop on route 1 that is closest to the tap-on location of route 2 is assumed to be the alighting location.
2. If the boarding is the last one during the day, the first location of the first boarding is used as the alighting location.

In addition, the study proposes a streamlined spatial validation process to check whether the rider’s subsequent boarding is within a reasonable distance from previous transit route the passenger is alighting from. This spatial validation involves:

1. Drawing a straight line between the tap-on location to the closest point along the route of the previous transaction to obtain an “as the crow flies” distance. This was performed using the NeighbourFinder tool in FME.
2. If the distance is less than 500 metres, the interpolated alighting location is accepted.

The spatial validation uses the NeighbourFinder tool in FME. The concept of the NeighbourFinder tool is the same as the Near tool in ArcGIS, where the tool calculates the distance and proximity information between the input features that are closest to another layer (ESRI, 2016). The data used contain the route geometry of all transit lines in the system as the near features, and points for all transit stops as the input features. Then a “point to line” proximity calculation was performed to obtain an “as the crow flies” distance from the input features (i.e. transit stops) to the near features (i.e. transit routes). This produces a table of stop to transit line distances that can be joined to the result of the trip reconstruction to test whether the subsequent boarding location is within a reasonable distance from the transit route that the passenger used in the previous trip.

Figure 3-3: Near analysis concept

Among the geocoded HSR PRESTO transactions, 161,009 PRESTO trips are unlinked, representing 37.81% of the geocoded PRESTO transactions. Conversing with HSR staff, the high

4 Diagrams from ESRI (2016)
number of unlinked PRESTO trips is a result of the under-representation of frequent HSR riders in the current pool of HSR PRESTO users, as monthly passes and paper tickets are still in circulation. In addition, Mohawk College students have yet to adopt PRESTO for their U-Pass program. Finally, it is important to remember that the PRESTO transactions had to be geocoded using historical CAD-AVL data, which resulted in discarding 10% of the transactions, and increasing the number of unlinked trips. There is currently no means to verify the accuracy of the Vehicle Mapping Table used to link the PRESTO transactions to the historical AVL data.

Nevertheless, among the linked transactions where the same card made more than 1 transaction per day, 86.41% of the estimated alighting areas are within 500m from the previous bus route, resulting in 228,873 O-D pairs for the study period (Table 3-1). This is a reasonably good result, and the algorithm can be readily repeated using the Near tool in ArcGIS, in addition to the NeighbourFinder tool in FME that this study has used. In the future, the low HSR PRESTO adoption rate will not be an issue, when paper tickets are phased out and transit passes are only available on PRESTO. The upcoming PRESTO device refresh is also expected to include GPS support that could geotag each transaction, which may eliminate the need for geocoding. While it is true that at this stage, trip reconstruction cannot produce a reliable O-D for HSR due to the low PRESTO adoption rate, future upgrades to the PRESTO system bears very promising potential.

Table 3-1: Trip reconstruction results

<table>
<thead>
<tr>
<th>Distance from Previous Route</th>
<th># of Transactions</th>
<th>% to Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 500m</td>
<td>228,873</td>
<td>86.41%</td>
</tr>
<tr>
<td>500 to 1000m</td>
<td>5912</td>
<td>2.23%</td>
</tr>
<tr>
<td>1000 to 2000m</td>
<td>4842</td>
<td>1.83%</td>
</tr>
<tr>
<td>2000 to 3000m</td>
<td>2236</td>
<td>0.84%</td>
</tr>
</tbody>
</table>
3.3 Development of a New Zonal System for Public Transit O-D Matrices

To analyze travel flows, transportation planning models traditionally use Traffic Analysis Zones (TAZ) to aggregate O-Ds. Unlike the O-Ds obtained from household travel surveys, where the assignment to TAZs are direct, mapping smart card O-Ds cannot be easily done as traditional Traffic Analysis Zones tend to have boundaries along major arterial roads (Figure 3-4). Tamblay, Galilea, Iglesias, Raveau, & Munoz, (2016) proposes a fractional assignment of stops to TAZs using a logit model, with socioeconomic, land use and access surveys as inputs. However, this method is likely too complex for transit agencies to adopt and it relies on traditional travel demand survey where sample size is spatially uneven (Tamblay et al., 2016). Meanwhile, traditional TAZs cannot be effectively applied for more localized route level analysis, due to the splitting of inbound and outbound trips into different zones. There is the need, therefore, to develop an appropriate unit of aggregation that is suitable for transit analysis.

Figure 3-4: Boundaries of Traffic Analysis Zones (TAZs) along arterial roads in Hamilton

---

5 Base map sources: Esri, DeLorme, HERE, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), MapmyIndia, Tomtom; Traffic Analysis Zones sourced from the Transportation Tomorrow Survey
This study proposes a new zonal system that splits the transit service area around major bus stops of bus routes. The new zones are delineated around timing points of bus routes in HSR and will not split along major arterial roads. Timing points are bus stop locations where schedule adherence is measured; these stops also tend to be located near major intersections.

The method will use the Integrate tool in ArcGIS to consolidate timing points that are within 200 metres of each other so that each major intersection would only have one consolidated timing point. The concept of ArcGIS’ “Integrate” tool is illustrated in Figure 3-5. Next, the consolidated timing points are fed into the Network Analyst Extension in ArcGIS, where a base network is built using OpenStreetMap to divide the city into non-overlapping catchment areas around the consolidated timing points, using the “Create Service Area” tool.

**Figure 3-5: Consolidation of timing points within 200 metres of each other**

The procedures to create the new zone system is as follows:

1. All timing points (also known as nodes) for HSR are retrieved and mapped in ArcGIS.
2. The “Integrate” tool is then used to consolidate all timing points that are within 200m of each other – creating essentially one “super stop” on each major intersection.
3. A network dataset is created using OpenStreetMap, in ArcGIS.
4. Using Network Analyst, non-overlapping network buffers are created for all consolidated timing points in Hamilton to delineate the new zones, using the “Create Service Area” tool.
5. The transit service area polygon is then used to clip the resultant polygons to prevent the zones to extend beyond an agency’s service area, with some additional manual polygon editing to remove rough edges.

The results produce a series of new zones that do not split along major arterial roads (Figure 3-6). For denser downtown areas, the timing points naturally become denser, creating smaller zones,
while in more suburban areas, the relative scarcity of timing points create bigger zones. This new “Transit Analysis Zone” system can potentially be used to aggregate O-Ds created from trip construction of PRESTO transactions or re-aggregate O-Ds from travel surveys.

**Figure 3-6: Transit Analysis Zones for HSR service area**

---

6 Base Map Sources: Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community
3.4 Use of Fare Zones toAggregate O-Ds and Visualize Passenger Movements

Another case study uses GO Transit (GO) to convert PRESTO transactions into O-Ds. Unlike every other transit agency in the GTHA, GO uses a zonal fare system that requires passengers to tap off when alighting. This creates a convenient and simple medium to summarize and map the O-D flows of passengers. However, the transaction locations cannot be directly linked to the Stop IDs of GO’s publicly available GTFS files, so geocoding is done on the list of transaction locations (TrxLocationIDs) using the descriptions from the LocationName fields that are associated with each TrxLocationID. On the stops.txt for GO’s GTFS files, each stop contains a zone number that specifies the fare zone that the stop belongs to. The mean centres\(^7\) of each fare zones are calculated to help visualize the results.

For GO Train passengers, there is the option to set a default trip on their PRESTO cards between two stations, which would only require the passenger to tap on at the default origin station and eliminates the need for the passenger to tap off at the destination station. For tap on transactions with default trip, the origin and destination locations are provided by the tap on transaction. For tap on trips without default destination, the destination and origin locations are provided by the tap off transaction. For this case study, two weeks of PRESTO data spanning from September 1, 2017 to September 15, 2017 are used. The procedural summary is as follows:

1. Manual cleanup of transaction location description is performed to ensure that the addresses can be read correctly by the Google Geocoder.
2. Then, the list of location names is fed into the Google Geocoder via a Google API.
3. To summarize the data, the mean centre of all GO Transit fare zones was calculated from the publicly available GTFS files, using the Mean Center tool in ArcGIS.
4. Tap on with default trip and tap off transactions are included in the final transaction lists.
5. Tap on transactions without default trip are removed because the subsequent tap off transaction contains the tap on location to calculate the final fare.

\(^7\) The Mean Centre is given by:

\[
\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}
\]

\(x_i\) and \(y_i\) = the coordinates for feature \(i\)

\(n = \text{total number of features}\)
6. Each PRESTO transaction is then associated to their respective fare zones.

7. Next, the O-Ds are grouped by fare zone IDs.

8. Finally, lines are generated between the mean centres of fare zones to visualize the O-Ds.

Figure 3-7 presents a map of the O-Ds aggregated by fare zones. In line with the expectation that the majority of GO’s ridership are on its commuter rail network, the red and orange lines on the map are all originating along GO’s train lines. The further one moves away from the municipalities neighbouring Toronto, the thinner the ridership is as those areas are only serviced by buses.

**Figure 3-7: GO Transit one-way O-D flows grouped by fare zones**

---

8 Basemap Sources: Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community
4 Enhancing Key Performance Indicators with User-Based Metrics

Unlike data from the traditional farebox, Automatic Vehicle Locations (AVL), and Automatic Passenger Counts (APC), smart card can track individual passengers over a specified period of time; this allows for the detailed examination of individual travel patterns and the analysis of service quality from the customers’ perspective. The commonly used key performance indicators (KPIs) in transit planning tend to focus on average performance and route level summaries. The reliance on route-based measures is largely a result of data constraints because it is difficult to track the detailed journeys and experience of individual passengers in the transit network without smart card data. This chapter demonstrates improvements to common transit Key Performance Indicators (KPIs) employing the use of PRESTO data, focusing on ridership and transfers. Burlington Transit and the TTC will be used for case studies. Four areas of applications are presented, each with increasing level of disaggregation. Burlington Transit is used as the case study to demonstrate methods to compute transfer matrix and assess user impacts of service change proposals, while the TTC case study presents visualizations of travel patterns and analyses transfer quality.

4.1 Computing Transfer Volume Matrix

Ridership and productivity are two key KPIs to evaluate the performance of a transit route from the perspective of passenger demand. Ridership is measured as the number of passenger trips or linked trips, as a one-way ride from origin to destination, excluding transfers (Canadian Urban Transit Association, 2017). Productivity is measured by the number of boarding per service hour, which controls for the level of service provided on the route (Canadian Urban Transit Association, 2017). However, public transit functions as a network, and especially for high ridership transit systems where transit routes form a grid system, a considerable portion of riders must travel on more than one route in order to reach their destinations. Therefore, there are substantial ridership interactions between routes, where PRESTO’s capability to keep track of the activities of individual users enables the opportunity to compute transfer matrices that quantify the route to route transfer volumes.

In late 2016, in support of Burlington’s urban intensification policy, the consulting firm Jarret Walker + Associates was commissioned by Burlington Transit to review their existing service and
propose options to redesign the network to increase transit ridership (Jarrett Walker + Associates, 2017). The scope of the study did not include any analysis of PRESTO data. Therefore, the Jarrett Walker + Associates (2017) study is used as the baseline for the existing state of practice in transit service review to highlight the additional insight PRESTO could provide.

One of the main proposals in Jarrett Walker + Associates (2017) is the elimination of Route 101 Express. In Burlington, the busiest corridor is served by Route 1 and Route 101 Express (Figure 4-1). Both routes operate along the same Plains Road corridor from Burlington GO Station to Downtown Hamilton, representing over 20% of Burlington Transit’s ridership. Route 101 operates as an express route that skips two-thirds of Route 1’s stop. Four months of PRESTO transactions spanning from September 1, 2017 to December 31, 2017 are used to compute transfer matrices.

**Figure 4-1: Route 1 and Route 101 Express, Burlington Transit**

![Route 1 and Route 101 Express, Burlington Transit](image)

Route 1 operates every 30 mins all day until late night, while Route 101 operates every 15 mins during the peak, 30 mins during the mid-day, and no service past the PM peak. The travel time from Hamilton Downtown to Burlington GO Station is 10 mins faster by the Route 101 Express. The analysis from Jarrett Walker + Associates (2017) emphasizes the need for consistent frequency and criticized the design of two routes operating every 30 mins during the mid-day. The study suggests potentially eliminating the Route 101 Express and instead, run Route 1 every 15

---

9 Map from Burlington Transit (2017)
mins all day, allowing riders to have more freedom to begin their trips spontaneously during the midday.

The analysis, however, did not consider transfer volume. Without an analysis using PRESTO transactions, the elimination of Route 101 Express may have unforeseen effects on other routes in Burlington Transit that connect with the Route 101 Express. The analysis in this section will address the gap, through the computation of a transfer matrix for Route 1 and Route 101 to provide a more comprehensive demand profile of Burlington Transit’s busiest corridor. The procedure to compute route level transfer matrix is as follows:

1. For each PRESTO transaction, identify transfer transactions through the Transaction Type field (TrxTypeID). When TrxTypeID = “16”, the transaction is flagged as a transfer where the passenger does not pay a second fare.
2. For each identified transfer transaction, add a field named “From Route” and set the field equal to the route number of the previous transaction, given by the LineID field.
3. For each identified transfer transaction, add a field named “To Route” and set the field equal to the LineID of the current transaction.
4. All the transfer transactions are then filtered out to a new spreadsheet and pivot tables are applied in Excel to summarize the data into a transfer matrix.

Table 4-1 summarizes the transfer profile for two of Burlington Transit’s busiest routes, the Route 1 and Route 101 Express. The results show that Route 10, Route 21, and Route 81 are the top 3 routes that Route 101 passengers transfer to, while Route 10, Route 21 and Route 3 are the top 3 routes that Route 1 passengers transfer to. Meanwhile, about 14.5% and 6.2% of users on Route 1 and Route 101 respectively take advantage of the 2-hour windows to complete more than 1 trip within one fare. Considering that Route 21 operates along Fairview Street, which is a continuation of Plains Road east of Brant Street in Burlington, this insight could be used, for example, to support service changes to implement an interline of Route 21 with Route 1/101.
Table 4-1: Transfer volume from Route 1 and Route 101, Sept 2017 to Dec 2017

<table>
<thead>
<tr>
<th>To Route</th>
<th>Route 1 (Local)</th>
<th>Route 101 (Express)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.6%</td>
<td>4.5%</td>
<td>8.5%</td>
</tr>
<tr>
<td>101</td>
<td>7.0%</td>
<td>6.3%</td>
<td>6.6%</td>
</tr>
<tr>
<td>2</td>
<td>6.6%</td>
<td>4.6%</td>
<td>5.4%</td>
</tr>
<tr>
<td>3</td>
<td>8.4%</td>
<td>4.8%</td>
<td>6.3%</td>
</tr>
<tr>
<td>5</td>
<td>2.5%</td>
<td>5.5%</td>
<td>4.3%</td>
</tr>
<tr>
<td>6</td>
<td>7.9%</td>
<td>6.8%</td>
<td>7.2%</td>
</tr>
<tr>
<td>10</td>
<td>16.5%</td>
<td>14.0%</td>
<td>15.0%</td>
</tr>
<tr>
<td>12</td>
<td>7.5%</td>
<td>6.4%</td>
<td>6.8%</td>
</tr>
<tr>
<td>21</td>
<td>12.6%</td>
<td>15.4%</td>
<td>14.3%</td>
</tr>
<tr>
<td>25</td>
<td>8.4%</td>
<td>7.7%</td>
<td>8.0%</td>
</tr>
<tr>
<td>50</td>
<td>0.6%</td>
<td>0.0%</td>
<td>0.2%</td>
</tr>
<tr>
<td>51</td>
<td>0.2%</td>
<td>0.0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>52</td>
<td>0.2%</td>
<td>0.0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>80</td>
<td>2.4%</td>
<td>8.4%</td>
<td>6.1%</td>
</tr>
<tr>
<td>81</td>
<td>1.8%</td>
<td>15.2%</td>
<td>9.9%</td>
</tr>
<tr>
<td>87</td>
<td>2.5%</td>
<td>0.1%</td>
<td>1.1%</td>
</tr>
<tr>
<td>300</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
### 4.2 Informing Service Span Changes with Assessment of User Impacts

Another key part of a transit service review is service span, which measures the temporal coverage of transit service. Some transit routes may have high service span which starts service earlier in the day and ends later at night. It is true that the AM Peak and PM Peak periods are where the bulk of the ridership happens, but if a city is trying to promote intensification and reduce the mode share of the automobile, transit must be available for a more diverse range of trip purposes than just the morning and afternoon commutes. In order for transit to become a competitive mode choice, the transit agency needs to focus on the midday, evening, and weekend services, in addition to the weekday peaks.

PRESTO provides the capability to compute the number of unique riders over a given period of time, which provides an additional dimension of insight into the percentage of the user base that could be impacted by a proposed service change. In the same study by Jarrett Walker + Associates (2017), options were presented to re-design the entire transit network with the intent to concentrate resources towards busier times and higher demand corridors. Three potential concepts were presented for discussion, with each concept representing the varying extent of concentrating resources towards busier time periods and restructuring routes to form a grid pattern. However, a hard constraint was imposed to only review the existing service hours and budget.

Figure 4–2 shows the existing service spans. Figure 4–3 shows the “coverage” concept where resources are split evenly between high demand and low demand areas. Figure 4–4 shows the other two concepts: one where 70% of resources are dedicated to high demand areas at the top, and one where 90% of resources are dedicated to high demand areas at the bottom.
Figure 4-2: Existing Burlington Transit service span

Figure 4-3: Proposed coverage service span concept

*50% ridership and 50% coverage

10 Diagram from Jarrett Walker + Associates (2017)
11 Diagram from Jarrett Walker + Associates (2017)
Figure 4-4: Proposed midpoint and ridership service span concept

*70% ridership and 30% coverage concept (top) and 90% ridership and 10% coverage concept (bottom)

What the three concepts have in common, are significant reductions in service span, where services would end after around 9:00PM to 10:00PM at night, in order to enable higher frequencies. As boarding after 9:00PM only accounts for a small share of total ridership, this may seem rational, and ridership gains should materialize from a business perspective. A user-based analysis through PRESTO, however, could provide a more fulsome picture.

With PRESTO transaction data, it is possible to ask the question:

- What would be the user impacts, if services end after 9:00PM?

Using the same set of PRESTO transaction data from September 2017 to December 2017, the number of transactions and number of unique users were counted after 9:00PM for the 4 months study period, with results presented in Error! Reference source not found. Furthermore, the average and median total boardings per unique users before and after 9:00PM are calculated over the span of the study period from Sept 2017 to Dec 2017, and results are summarized in Table 4-3.

12 Diagram from Jarrett Walker + Associates (2017)
Table 4-2: Boarding and unique rider counts before and after 9:00PM

<table>
<thead>
<tr>
<th>Rider Categories</th>
<th>Total Boarding</th>
<th>Total # of Unique Riders</th>
<th>% to Total Boarding</th>
<th>% to Total Riders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riders After 9PM</td>
<td>31677</td>
<td>8034</td>
<td>4.5%</td>
<td>35.4%</td>
</tr>
<tr>
<td>Riders Before 9PM</td>
<td>671190</td>
<td>21970</td>
<td>95.5%</td>
<td>96.9%</td>
</tr>
<tr>
<td>All Riders</td>
<td>702867</td>
<td>22675</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 4-3: Profiles of riders before and after 9:00PM

<table>
<thead>
<tr>
<th>Rider Categories</th>
<th>Average Total Boarding (Sept to Dec 2017)</th>
<th>Median Total Boarding (Sept to Dec 2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riders After 9PM</td>
<td>57</td>
<td>26</td>
</tr>
<tr>
<td>Riders Before 9PM</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>All Riders</td>
<td>31</td>
<td>6</td>
</tr>
</tbody>
</table>

The findings show that late night boarding represents only 5% of Burlington Transit’s PRESTO transactions, yet 35.4% of the total user base would be affected. Furthermore, the passenger profiles (Table 4-3) show that the average total number of boarding made over the study period of those users who use services at night (i.e. after 9PM), at 57 trips over 4 months, is much higher than the average boarding of those users who use services before 9PM, at 32 trips over 4 months. The median total boarding of users after 9:00PM is 26, compared to the system-wide median of 6 boarding.

Hence, reducing service at night would end up affecting Burlington Transit’s most frequent users. Without PRESTO data, however, such analysis would not be possible, and network optimizations that do not account for the impact on individual users would not result in ridership increases. This simple method can be applied on a route to route basis to evaluate the impact on the number of unique riders, and based on those affected riders’ travel patterns, inform route design choices.
4.3 Application of Spatial Clustering to Visualize Travel Patterns

One of the unique characteristics of PRESTO is that it is used by 10 transit agencies in the GTHA. While PRESTO devices of 905 agencies are not all GPS-enabled yet, each usage transaction does contain a Service Provider field that documents the transit agency that the passenger travelled on. Since the PRESTO smart card system is in use across the GTHA, there is a unique opportunity to classify passengers based on their origin municipality. Through the Service Provider field included in the PRESTO data, it is possible to build a rider profile based on the number of boardings made on different transit agencies across the study period and analyze the travel patterns through spatial clustering. This case study of rider classification and clustering involves PRESTO data from every transit agency in the GTHA over a period of 4 months from September 2017 to December 2017. The dataset contains over 99 million rows of PRESTO transactions, which effectively makes this case study the most data-intensive exercise in this thesis.

The travel patterns of riders are illustrated on the TTC network, since their PRESTO devices are GPS-enabled, where each PRESTO transaction on the TTC is geotagged. It is true that TTC has a low PRESTO adoption rate of only 14.5% in late 2017, whereas the 905 transit agencies all have much higher PRESTO adoption rates (TTC, 2018; Metrolinx, 2017). Hence, while the TTC’s PRESTO adoption rate is low, the combination of neighbouring transit agencies’ high PRESTO adoption rate and TTC’s PRESTO system that geotags every transaction presents a unique opportunity to apply spatial clustering techniques to visualize the travel patterns of riders from different neighbouring municipalities.

Leveraging the 4 months of PRESTO of data, the usage history of each PRESTO card can be analyzed to extract the home transit agency of each rider. This study uses the first transaction of each day during the study period to classify riders. For example, if a PRESTO card has first-day transactions on the TTC and Brampton Transit, but shows that there are 50 first usage transactions on Brampton Transit and only 10 first usage transactions on the TTC, then this particular PRESTO card would be classified as a card from Brampton. The procedures are summarized below:

1. An SQLite database was created to store the 4 months of PRESTO transaction data, containing over 99 million rows.
2. For each unique card ID on each separate date, the SQL query identifies the first usage transaction and record the transit agency that the passenger travelled on – with regional GO Transit excluded.

3. For each unique Card ID, the local transit agency with the greatest number of first usage transactions is set as the home agency.

4. TTC’s PRESTO transaction data is then grouped at the stop level by home agencies (see Figure 4-4 for example); each column provides the sum of boardings by riders from different transit agencies.

5. Finally, the Getis-Ord Statistics are then applied to the list of TTC bus stops layer to identify spatial hot spots of travellers from different transit agencies. The analysis is performed in ArcGIS using the “Hot Spot Analysis” tool with inverse distance as the spatial relationship.

6. The hot spots of passengers from different neighbouring agencies are mapped through a series of maps, one per neighbouring transit agency.

Table 4-4: Sample of classified boardings by the home transit agency and TTC Stop ID

<table>
<thead>
<tr>
<th>TTC StopID</th>
<th>Durham Region Transit</th>
<th>Brampton Transit</th>
<th>Burlington Transit</th>
<th>Hamilton Street Rail</th>
<th>Mi-Way</th>
<th>Oakville</th>
<th>York Region Transit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>41</td>
</tr>
</tbody>
</table>

Getis-Ord Statistics is given by:

\[ G_t^* = \frac{\sum_{j=1}^{n} w_{t,j} x_j - \bar{X} \sum_{j=1}^{n} w_{t,j}}{s} \], \quad \bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2};

\[ x_j = \text{attribute value for feature } j \]
\[ w_{t,j} = \text{spatial weight between feature } i \text{ and } j \]
\[ n = \text{total number of features} \]


Shapefiles of transit routes and stops used for all maps in this section are generated from publicly available GTFS files in March 2018, obtained from [https://transitfeeds.com](https://transitfeeds.com).
Figure 4-5 shows the clusters of Mississauga transit riders on the TTC. The clusters can be found in the Union Station area in the south, representing riders that connect onto the TTC via the GO Trains, and along the Bloor Subway lines where riders likely connect onto the TTC at Kipling Subway stations in the west. Few hot spots can be spotted, however, in the north and east part of the TTC network.

Figure 4-5: Clusters of Mi-way riders in Toronto

![Cluster Map](image.png)

Figure 4-6 presents the clusters of Brampton transit riders in Toronto. The patterns show clusters around the Union Station area in the south where riders connect from GO Trains. Clusters can also be found in the airport areas in the north-west and York University in the north where riders connect from Brampton Transit’s express bus routes (Brampton Zum). Interestingly, there are also more hot spots scattered across Toronto. Brampton Transit is currently soaring in ridership, growing by 18% from 2016 to 2017 (City of Brampton, 2018). From 2009 to 2017, transit ridership grew by 122.8% while population only increased by 22.1% (City of Brampton, 2018). One of the main reasons cited for Brampton Transit’s success is that, compared to its peer transit agencies, Brampton Transit recognizes the importance of network effects and the integrated nature
of the Toronto Region (Marshall, 2018). The regional view is reflected in Brampton’s network design, where particular attention is paid to ensure connections to major transit hubs and post-secondary institutions, even if it requires extending routes well beyond Brampton’s municipal boundary. The focus on integrating the transit network with the TTC results is more diversity of destinations for Brampton Transit riders in Toronto, which can be attributed to the more scattered hotspots of Brampton Transit. Compared to Mississauga, it appears that the destination of Brampton’s riders is more diverse, despite Mississauga being geographically closer to Toronto.

**Figure 4-6: Clusters of Brampton Transit riders in Toronto**
Figure 4-7 maps the clusters of Durham Region Transit (DRT) riders in Toronto. Similar to Mississauga, clusters are mostly found in Downtown Toronto where riders connect from GO Trains and from the east, where riders connect directly from DRT routes onto the TTC network. Substantial concentrations can be found in the Scarborough area near the locations of Centennial College and University of Toronto Scarborough Campus.

**Figure 4-7: Clusters of Durham Region Transit (DRT) riders in Toronto**
Figure 4-8 maps the clusters of York Region Transit (YRT) riders in Toronto. Compared to Mississauga and Durham Region, the hot spots in Downtown Toronto are weaker. Meanwhile, the hot spots in the North York Centre Station area in the north are stronger. This reflects the fact that York Region is geographically closer to the North York Centre area and a considerable portion of YRT’s resources are concentrated towards feeding TTC’s subway stations in the north.

**Figure 4-8: Clusters of York Region Transit (YRT) riders in Toronto**

Finally, for Oakville (Figure 4-9), Burlington (Figure 4-10), and Hamilton (Figure 4-11) transit riders, because their home transit agencies do not have any bus routes that directly connect with the TTC’s network, the clusters of those three systems are heavily concentrated in Downtown Toronto, where riders connect from the GO Trains at Union Station.
Figure 4-9: Clusters of Oakville Transit riders in Toronto

Figure 4-10: Clusters of Burlington Transit riders in Toronto
Overall, the patterns reflect the varying degree of connectedness between the neighbouring transit agencies and the TTC. The further away the transit agency, the more concentration the hot spots are around Union Station. As the PRESTO adoption rate improves for the TTC, the classification method can be adapted to visualize the riders of specific transit routes. As O-D information may seem abstract and hard to understand for those who are not modellers, the outputs of the methods proposed in this section produce simpler, easier to understand visuals that can help facilitate the conversation during a transit service review.
4.4 Quantifying Travel Time Variations and Transfer Delays

Traditionally, user experiences are measured using customer surveys which provide cross-sectional snapshots of customer perceptions. Smart card data, with its large volume and continuous nature, could be more spatially and temporally representative compared to traditional customer satisfaction surveys (El-Geneidy, Horning, & Krizek, 2011). The existing literature is rich with transit indicators. Furth & Muller (2006) observe that extreme scenarios tend to exert more influence over one’s recollection of the experience. Generally speaking, travellers are more concerned with travel time reliability than travel time, because a reliable travel time enables better trip planning (Bates, Polak, Jones, & Cook, 2001). Planning trips that involve unreliable service is more stressful, as the experience of waiting for such a service at the bus stop induces anxiety. There is, however, a paucity of research on using smart card data to assess transfers. Jang (2010) demonstrated simple queries that could highlight nodes with high transfer volumes and transfer times in Seoul, South Korea. Zhao et al. (2017) used smart card data from Nanjing, China to identify high priority transfer nodes with poor performance that need review, but the study only focused on the subway to bus transfers, as passengers alighting from buses do not perform a tap off. Both Jang (2010) and Zhao et al. (2017) have case studies where tap off is required. This section will present methods to measure transfer delays experienced by users on TTC’s high frequency, tap-on only bus network.

The main issue with average performance and route level KPIs is that they do not measure the experience through the lens of customers. In transit agencies where routes form a high-frequency grid network, most transit journeys would require at least one connection, also known as a transfer. Today, with the availability of real-time bus arrival information through CAD-AVL systems, passengers can minimize their wait times at their origin stops. However, passengers cannot effectively control their arrival times at their transfer locations, as transferring passengers have no choice but to wait for the connecting vehicle captively. Currently, according to Toronto Transit Commission (2017), TTC’s current service standards use On-Time Performance (OTP), measured as the ratio of the number of trips that are deemed “on time” to the total number of trips, for lower frequency routes. For high-frequency routes, headway regularity measures the ratio of the number of trips with acceptable headways over the total number of trips. Meanwhile, the performance of connections is currently not included in TTC’s service standards, as it is difficult
to quantify the number of transfer delays experienced by the customers without smart card technology that tracks passenger activities like what PRESTO can ultimately provide. The adoption of PRESTO presents the opportunity to identify poor performing transfer nodes.

For this study, because historical AVL data is not available, the actual wait times between alighting the first vehicle and waiting for the arrival of the connecting vehicle cannot be computed. However, it is possible to estimate the transfer delays through an average operating speed. Finally, the TTC operates a very integrated system between its surface routes and subways. Many major subway stations allow barrier-free transfers inside the subway stations. Hence, this study will only include bus to bus transfers that are not barrier-free, and transfer times will be capped at 30 mins to filter out instances where the passenger may have engaged in activity near the transfer location, similar to Zhao et al. (2017). A list of TTC stops with geo-coordinates are obtained from the stops.txt file of TTC’s GTFS files. The stops.txt file contains fields that include stop name, stop code, latitude, and longitude. The stop code is an integer field that provides a unique identifier to each stop that can be readily joined to the transaction location (TrxLocationID) field of TTC’s PRESTO usage transactions. The procedures for transfer quality analysis are as followed:

1. For each transfer transaction time, the script calculates the difference from the transfer tap’s transaction time to the previous transaction time.
2. For each transfer transaction location, the script calculates the Euclidian distance to the previous transaction location (Figure 4-12).
3. Then, a filter is applied only to include a bus to bus transfers and the measured “as the crow flies” distance is divided by an average speed of 20 km/h to estimate in-vehicle time.
4. The estimated travel time is then subtracted from the difference between the first and second transaction time calculated from step 1, to yield an estimated transfer time.
5. Instances with estimated transfer time exceeding 30 mins are filtered out.

Figure 4-12: Summary of Transfer Time Estimation

\[
\text{Estimated In-Vehicle Travel Time} = \frac{\text{Distance from Location 1 to Location 2}}{20 \text{ Kilometres / Hour}} \\
\text{Transfer Time} = \text{(Transfer Transaction Time)} - \text{(1st Transaction Time)} - \text{(Estimated In-Vehicle Travel Time)}
\]
To analyze the consistency of transfer time experienced:

6. The output after step 5 is grouped by the PRESTO card IDs to measure the consistency of transfer time experienced by passengers.

7. Passengers are then classified into 3 bins by the number of times the passenger have made the transfer between the same combination of routes. The three bins are: 2 to 5 times, 6 to 10 times, and 11 to 15 times.

8. Box plots are created to visualize the spread of transfer times, group into the 3 bins of users created in Step 7.

To analyze transfer time spatially:

9. The same output after step 5 is grouped by transaction locations to visualize the transfer delays spatially across the network. Locations are required to have at least 100 transfer observations to be included.

10. The median transfer time is calculated for each transfer location (i.e. stop ID).

11. For the median transfer times by stops, the Getis-Ord Statistics is used to identify hot spots of high transfer times and volume with inverse distance as the spatial relationship.

12. The results are mapped.

Figure 4-13 presents box plots of transfer times experienced, filtered to only include passengers that connect onto Route 85, Sheppard East bus from 9 perpendicular routes that intersect Route 85. The box plots are grouped by the number of times the individual passengers have made the connection onto Route 85 in September 2017. The riders who make the trip less frequently experience more instances of outliers and a larger spread in transfer times, compared to riders who make the same trip more frequently. This shows that frequent riders tend to experience a more reliable service.
Figure 4-13: Boxplots of estimated transfer time by the frequency of use (Route 85)

Figure 4-14 maps the transfer locations with at least 100 transfer observations in the study period. Figure 4-15 graphs the distribution of median transfer times by locations and the distribution of median transfer time is a normal distribution with a left skew. Figure 4-16 shows the Getis-Ord Statistics result. Hot spots of high transfer times, marked by red dots, should be reviewed.

Interestingly, there are some clusters of locations with high transfer times near York University and the centre part of Eglinton Avenue. The cause may be due to construction activities, as the study period takes place during the 3 months prior to the opening of the Spadina Subway Extension that would have reached York University, and the Eglinton Crosstown LRT is also under construction with lots of construction activities near the two subway lines. One outlier can be spotted, Jane Station, where despite being a subway station, there is no barrier-free transfer and the median transfer time for bus-to-bus transfers is high.
Figure 4-14: Median bus to bus transfer times by stops, TTC, September 2017

Figure 4-15: Histogram of median transfer times by Stops, TTC, September 2017
The transfer nodes can also be weighted by transfer volume, in addition to the median transfer time. Figure 4-17 shows the result of the Getis-Ord statistics for transfer volumes. It can be seen that the hot spots of high transfer volumes are generally near the hot spots of high median transfer time. A more fine-grained measure that compares the relative volume of the transfer nodes that require review, however, would require a higher PRESTO adoption rate. At this point, considering that 45% of TTC’s ridership is from monthly pass holders (i.e. frequent users), and monthly passes were not yet on PRESTO in 2017 for the vast majority of users, there is no reliable transfer volume that is accurate enough for a node to node comparison of transfer volume (Toronto Transit Commission, 2018).
Overall, the transfer experience can be inconsistent on the TTC network. Understanding the specific causes of each node will require detailed examination of historical AVL data, which this study, unfortunately, do not have access to. Nevertheless, reviewing the TTC’s service standards can offer some high-level insights. The inconsistent transfer performance could be a result of TTC’s lesser focus on mid-route On Time Performance (OTP). The service standard only requires 60% of trips to arrive within 1 minute early and 5 minutes late. Meanwhile, 90% of trips are required to depart within 1 minute early and 5 minutes from the origin time points (Toronto Transit Commission, 2017). Since most surface bus routes terminate at subway stations, the emphasis on origin time point suggests that although the TTC’s bus network forms a grid geometrically, the operating standard resembles attributes of a radial network where the bus routes have a clear primary focus of feeding passengers towards subway stations. Since a considerable portion of the bus to bus transfers on TTC’s high-frequency grid happen at intersections, away
from origin time points, the low arrival OTP targets from the service standards is contributing to
the inconsistent transfer experience of passengers on the bus network.

For future studies, it would be an interesting investigation to obtain more recent PRESTO
transaction data to investigate the effect on passenger transfer experience, after the opening of
York University subway station and the Eglinton Crosstown LRT. In-vehicle travel time
estimation can also be improved through the use of a schedule-based transit assignment model or
historical AVL data, rather than assuming an average operating speed. This case study also only
assessed the hot spots of high transfer times spatially. When the PRESTO adoption of the TTC
improves, space-time analysis\textsuperscript{16} or multi-variate spatial clustering\textsuperscript{17} can be applied to assess the
hot spots of high transfer times both spatially and temporally, along with any other operational
variable from the CAD-AVL system (e.g. passenger loads, on time performance, average
operating speed).

As PRESTO transaction is a continuous data source, a useful analogy to guide future research is to
treat each bus stop like rain gauges, where measures are being recorded for as long as the
equipment is working. Hence, it is my view that the rich pool of spatial analysis methods that were
originally purposed to analyze climate and weather events could potentially be adapted to analyze
transfer experience.

\textsuperscript{16} Further details on space-time analysis can be found here: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/space-time-analysis.htm
\textsuperscript{17} Further details on multi-variate spatial clustering can be found here: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-spatially-constrained-multivariate-clustering-works.htm
5 Conclusion, Reflections, and Future Opportunities with PRESTO

5.1 Summary of Main Findings

This thesis has demonstrated a series of potential applications of smart card analytics through a case study of PRESTO. The study involved four months of PRESTO usage transactions spanning from September 1, to December 31, 2017. Four transit agencies, Hamilton Street Railway (HSR), GO Transit (GO), Burlington Transit, and the Toronto Transit Commission (TTC), were used as case studies to present a toolbox of PRESTO analytics. Eight applications have been demonstrated, and the major findings are summarized below.

Section 3.1 shows that PRESTO transactions could be geocoded using historical vehicle location data, provided that a vehicle mapping table is available that documents where each individual PRESTO unit is installed during the study period. This mapping table is usually maintained by the maintenance department, and the update frequency will vary. Hence, it is important to keep in mind that repairs happen on an ongoing basis and the PRESTO units may be removed from buses that require extensive repairs. The longer the study period, the more risks of inaccuracy with the vehicle mapping table. This process is also useful in general to enrich PRESTO transactions with additional data fields from the CAD-AVL system, even after the expected PRESTO device refresh that is expected geotag each transaction.

Section 3.2 demonstrates trip reconstruction procedures. For PRESTO users that made more than one boarding per day, alighting locations can be interpolated using the subsequent boarding locations in the same day. If the boarding is the last of the day, the first boarding location of the day can be used as the estimated alighting location. The results can be validated spatially by measuring the distance between the subsequent boarding location to the route that the passenger alighted from. The HSR case study shows that 86% of estimated alighting locations are within 500 metres away from the first transit route the passenger was on, which confirms the robustness of the two tour chain assumptions.

Section 3.3 presents a new zonal system that is more suitable for public transit O-Ds. Since traditional Traffic Analysis Zones tend to split along major arterial roads, the inbound O-Ds and outbound O-Ds of the same transit route could be split into different zones. A new zonal system is developed through re-dividing the study area around the transit system’s major stops. For HSR,
the locations of timing point stops were first consolidated to group timing points within 200 metres of each other. A network dataset is then created using ArcGIS Network Analyst extension with OpenStreetMap as the base layer, followed by delineating non-overlapping network buffers around each consolidated timing point. The result is a new “Transit Analysis Zone” system that allows for the direct assignment of transit O-Ds which do not split along major arterial roads.

Section 3.4 shows that for transit agencies that use zonal fare systems, fare zones could be used to aggregate and visualize O-Ds extracted from smart card transactions. The case study of GO Transit demonstrated the use of fare zones to aggregate O-D flows.

Section 4.1 presents methods to compute route level transfer volume matrix using Burlington Transit’s Route 1 and Route 101 as the case study. The query is simple and capitalizes on the PRESTO data’s ability to track individual passengers during the study period.

Section 4.2 uses PRESTO data to evaluate the impact of service span changes on the user base by calculating the number of unique riders using the PRESTO card IDs. The method is applied towards Jarret Walker + Associate (2017)’s three alternative network design concepts for Burlington Transit that propose to end service at around 9:00PM at night in order to allow for higher frequencies during the day. The findings highlighted the fact that while the demand after 9:00PM represents only 4.5% of total boardings, 34.5% of users would be affected and those users tend to be frequent riders.

Section 4.3 classifies PRESTO users according to their home transit agency and applied spatial clustering to visualize the travel patterns of riders from neighbouring transit agencies on the TTC network. Using 4 months of PRESTO transactions from every single transit agency in the GTHA, each PRESTO card ID was matched to a local transit agency using the first transaction of each day during the study period. TTC’s PRESTO transaction data are then summarized by stops, and the Getis-Ord Statistics is used to identify hotspots of rider from neighbouring transit agencies. The findings showed that the travel patterns of Brampton riders are the most diverse, which can be attributed to the considerable amount of resources Brampton transit directs to ensure good quality connections to major transit hubs and post-secondary institutions, even if it requires extending routes well beyond Brampton’s municipal boundary. For other agencies, the further away the transit agency’s network is, the more concentrated the hot spots are around Union Station where passengers connect from GO Trains.
Section 4.4 quantifies bus-to-bus transfer delays on TTC’s high-frequency grid network. For each transfer transaction time, the difference from the transfer time to the previous transaction time (i.e. the difference between two transaction times) is calculated. In addition, the Euclidian distance is measured between the first and second boarding location and an average travel speed of 20 km/h is used to estimate the in-vehicle travel time. The estimated transfer time is then computed by subtracting the total travel time by the in-vehicle travel time. The variability of transfer times can then be measured at the passenger and stop level. The findings show that higher frequency users tend to experience less transfer time variability, and the stop level analysis highlighted locations with median transfer times in excess of 15 mins that should be further reviewed.

Overall, this thesis demonstrates PRESTO’s capability to observe passenger behaviour, enabled through smart card’s ability to track individual passengers spatially and temporally. As PRESTO is poised to become ubiquitous across the GTHA, this study provides a timely toolbox that showcases how PRESTO data can be leveraged to improve the practice of transit planning.

5.2 Reflection on Software Tools

In this thesis, most geoprocessing is completed using Safe Software Feature Manipulation Engine (FME) software that supports over 400 file formats. Meanwhile, ArcGIS is mostly used for mapping and visualization, supported by Excel. This thesis began with just 1 week of PRESTO data from Burlington Transit. Due to Burlington Transit’s small size, the data volume was small. Most analysis and data processing was done with ArcGIS, FME, and sometimes Excel without the need for a SQL database. Halfway through this thesis, when historical AVL was involved in geocoding PRESTO transactions, and the dataset expanded to include 4 months of transaction data from all transit agencies in the GTHA, I had to switch to an SQLite database to complete the table joins, rather than using FME. The increasingly data-driven nature of transit analysis will inevitably require the use of databases – and the days of relying on Excel are numbered.

5.3 Reflection on Smart Card Data’s Strategic Importance

Generally speaking, Canadian transit agencies do not use modelling to inform service planning (Miller, Shalaby, Diab, & Kasraian, 2018). From a transit planning perspective, the salient questions are: how much transit service to provide, when to provide the services, and where to provide the service. In transit terms, the questions concern the routes’ frequency and span as well as the location of transit routes. The answers to these three questions often involve a trade-off
between ridership and coverage, given a fixed municipal budget (Walker, 2011). The ridership and coverage objectives often compete and direct transit planning towards opposite directions. A high ridership-based design would call for more concentration of resources along higher density corridors, while a coverage-based design would ensure that every part of the transit service area is covered by some transit service.

PRESTO provides continuous and longitudinal data of transit demand. Traditional O-D matrices used by travel demand models tend to focus on the “average day” rather than accounting for the experience of transit riders over an extended period of time. The amount of travel demand on transit is contingent on the availability of high-quality service. By limiting the observation to only the passenger flow of one day, the “average day” O-Ds would provide a very limited picture of transit demand. The reason is rather simple. Travel needs vary from day to day. For example, a passenger who mostly rides along high-frequency corridors is very unlikely not to have needs to travel to areas of the city that are lower in density. Similarly, the same passenger who normally rides during the AM and PM peak periods to and from work, when transit frequency is highest, may sometimes need to return home later at night, when the wait time for transit may be longer. Unlike the automobile, where the road infrastructure and capacity are available around the clock, the amount of transit service decreases as time moves away from the AM and PM peaks. However, the reduction in off peak service would result in a considerable increase in wait time, which places limits on the freedom for more spontaneous travel. The concept of freedom enabled by frequency and service span is not experienced by drivers, nor is it accurately reflected in travel models that only involve data from the “average day.” PRESTO data provides the potential to quantify the link between peak and off peak ridership.

If a city is trying to promote intensification and reduce the mode share of the automobile, transit must be available for a more diverse range of trip purposes than just the morning and afternoon commutes. In order for the GTHA to successfully intensify, the mode share of transit must increase. The continuous nature of smart card data illustrates a more fulsome profile of transit demand that emphasizes the importance of a network with high frequency and long service spans.
5.4 Future Research Opportunities with PRESTO

While this thesis has demonstrated a series of analytical capabilities using PRESTO data, there remain other potential areas that were not explored in these case studies. In addition, there are substantial improvements to PRESTO planned in the next few years that will present future research opportunities. The new initiatives and the new opportunities for research include:

- In the next few years, 905 transit agencies will be undergoing a PRESTO device refresh. The new PRESTO units are expected to include GPS capabilities that could geotag each transaction and interface directly with CAD-AVL systems. The improved PRESTO location information will allow for more accurate estimation of O-Ds without the need to geocode the transactions using historical AVL data.

- DeviceID, the unique identifier for the PRESTO unit, is a field associated with each PRESTO usage transaction that has not been explored in this thesis. Major transit stations always have a considerable number of PRESTO units for a variety of functions that include fare payment, card balance checking, loading funds, and purchasing transit passes. If the researcher can obtain the location of each PRESTO unit inside the transit station, the PRESTO transaction records can be used to infer pedestrian movements inside stations and conduct pedestrian movement analysis.

- In the fall of 2018, PRESTO is expected to make available the registered postal code of each PRESTO card for analytics. Right now, it is not possible to directly identify the home area of a transit passenger from the PRESTO transaction records. Considering that most PRESTO transactions are made by registered PRESTO cards, the availability of postal codes will enable the potential to enrich PRESTO data with sociodemographic characteristics of the passenger’s neighbourhood. It would be helpful to understand the extent to which demand varies and how much of the variation can be explained by demographic variables, service variables, and any other external factors. Should these variables be identified, there could be improvements in transit demand modelling.

- Metrolinx recently announced the PRESTO mobile project, which will begin with a PRESTO app, followed by the ability to tap phones against PRESTO readers, and eventually integrate PRESTO into a multi-modal trip planning application (Metrolinx, 2018b). PRESTO mobile could provide additional context for each PRESTO transaction that can be tapped for additional insights.
6 References


Transactions on Intelligent Transportation Systems, 17(1), 133-142. doi:10.1109/TITS.2015.2464335


Trepanier, M., & Morency, C. (2010). Assessing transit loyalty with smart card data. 12th World Conference on Transport Research, Lisbon, Portugal, Retrieved from https://www.researchgate.net/publication/234129247_Assessing_Transit_Loyalty_with_Smart_Card_Data?enrichId=rgrreq-d71db42fd5ec050e999ab64be6eef23-XXX&enrichSource=Y292ZXJQYXdOzlzNDEyOTI0NztBUzozOTMxNTgxMjgwOTUyMzIAMTQ3MDe0NzYxMDU3Nw%3D%3D&el=1_x_2&esc=publicationCoverPdf


