massDOT
Massachusetts Department of Transportation

October 2024
Report No. 24-065

Measuring Fare Payment Compliance on MBTA Buses

Principal Investigator (s)
Dr. Eric Gonzales
Dr. Song Gao

University of Massachusetts Amherst

Research and Technology Transfer Section
MassDOT Office of Transportation Planning

U.S. Department of Transportation
Federal Highway Administration

[This blank, unnumbered page will be the back of your front cover]

Technical Report Document Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
24-065		

4. Title and Subtitle	5. Report Date
Measuring Fare Payment Compliance on MBTA Buses	October 2024
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Eric J. Gonzales, Song Gao, Mahdi Azhdari	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
University of Massachusetts Amherst 130 Natural Resources Way, Amherst, MA 01003	11. Contract or Grant No.

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Massachusetts Department of Transportation Office of Transportation Planning Ten Park Plaza, Suite 4150, Boston, MA 02116	Final Report -October 2024 (May 2023-October 2024)
	14. Sponsoring Agency Code n/a

15. Supplementary Notes
Project Champion - Sefira Bell-Masterson, MBTA

16. Abstract

This study presents a method to estimate fare payment compliance within the Massachusetts Bay Transportation Authority bus system using only automatically collected data from the Automated Fare Collection and Automated Passenger Counting systems. Data from over 787,000 passenger boardings were used to identify patterns of fare system non-interaction based on the difference between observed boardings and observed fare transactions. An estimated 22% of bus passengers do not interact with the fare system, but these non-interactions do not all result in lost fare. Some passengers are eligible for discounts and others who hold valid passes or are exempt from fares do not owe any payment at the farebox. Using average fare payments per observed transaction as an estimate of the lost revenue per non-interaction, revenue loss in the bus system is estimated to be between $6.0 million and $7.4 million for 2019. The number of non-interactions scales with ridership so the busiest parts of the system and times of day (midday and PM peak) also have the highest number of non-interactions and estimated lost revenues. There is no spatial pattern in the non-interaction rate. The proposed method allows fare evasion and lost revenues to be systematically tracked.

17. Key Word	18. Distribution Statement
public transportation; fare payment compliance; fare evasion; automated transit data	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
unclassified	unclassified	107	n/a

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

This page left blank intentionally.

Measuring Fare Payment Compliance on MBTA Buses and Light Rail

Final Report

Prepared By:
Eric J. Gonzales, Ph.D.
Principal Investigator

Song Gao, Ph.D.
Co-Principal Investigator

Mahdi Azhdari
Graduate Student Researcher

University of Massachusetts Amherst
130 Natural Resources Way, Amherst, MA 01003

Prepared For:

Massachusetts Department of Transportation Office of Transportation Planning
Ten Park Plaza, Suite 4150, Boston, MA 02116

December 2024

This page left blank intentionally.

Acknowledgments

Prepared in cooperation with the Massachusetts Department of Transportation, Office of Transportation Planning, and the United States Department of Transportation, Federal Highway Administration.

The Project Team would like to acknowledge the efforts of Sefira Bell-Masterson and David Churella (Project Champions) and Nicholas Zavolas (Project Manager).

Disclaimer

This page left blank intentionally

Table of Contents

List of Tables

This page left blank intentionally.

List of Figures

This page left blank intentionally.

List of Acronyms

Acronym	Expansion
AFC	Automated Fare Collection
APC	Automated Passenger Counter
CTPS	Central Transportation Planning Staff
CTrain	Calgary Train
FHWA	Federal Highway Administration
GIS	Geographic Information System
GPS	Global Positioning System
IQR	Interquartile Range
MAE	Mean Absolute Error
MassDOT	Massachusetts Department of Transportation
MBTA	Massachusetts Bay Transportation Authority
MSE	Mean Squared Error
NI	Non-interaction
NTD	National Transit Database
OD	Origin-Destination
ODX	Origin-Destination-Transfer
OLS	Ordinary Least Squares
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
SFMTA	San Francisco Municipal Transportation Agency
SPR	State Planning and Research
SQL	Structured Query Language
SSE	Sum of Squared Errors

This page is left blank intentionally.

1.0 Introduction

This study of Fare Payment Compliance on MBTA Transit was undertaken as part of the Massachusetts Department of Transportation (MassDOT) Research Program. This program is funded with Federal Highway Administration (FHWA) State Planning and Research (SPR) funds. Through this program, applied research is conducted on topics of importance to the Commonwealth of Massachusetts transportation agencies.

Fare evasion reduces needed revenues to the Massachusetts Bay Transportation Authority (MBTA). It is important to monitor fare payment patterns and the effect of policies and technologies on the MBTA's ability to collect fares. In addition to passengers who pay a fare upon boarding, the MBTA's policy is for passholders to validate their fare media at faregates or fareboxes even if no fare amount is deducted in that transaction. Failure to interact with the fare system may be due to evasion, in which case a fare should have been collected, or due to an eligible exemption for various groups of riders who are allowed to travel throughout the MBTA system for free, such as riders who are blind, children under 12, MBTA employees, contractors and retirees, and active members of the military. Although fare gates provide a systematic measure of various types of fare non-payment for the MBTA heavy rail system, it is more difficult to measure different types of non-payment on buses or light rail vehicles in which fare collection is at the discretion of the operator or motor person, and practices such as waiving passengers on or visually validating pass products are common. Conventional practice is to manually observe passengers boarding vehicles in order to estimate the number of passengers that evade fares versus those who are exempt. However, manual observations are costly to perform and therefore conducted infrequently, so there is a need for methods to track fare payment compliance over time and identify where and when manual checks are most valuable.

1.1 Project Overview

The proposed research method is to start with data that the MBTA already collects in a continuous and comprehensive manner. These data include records from automated fare collection (AFC) and from automated passenger counters (APC) used on buses and light rail vehicles in the fleet. APC technologies include devices mounted on doors to count the number of passengers boarding and alighting at each stop thereby providing an estimate of vehicle occupancy. AFC technologies include the fareboxes where riders use cash or tap farecards to pay fares. These technologies are deployed across the MBTA bus fleet. However, the light rail fleet includes only a handful of vehicles that are equipped with APC. Comparing AFC and APC records provides a coarse measure of the rate of fare collection across the bus system. These data are centrally recorded in the MBTA Research Database along with inferred origin-destination patterns from the origin-destination-transfer (ODX) model.

Existing records from infrequent manual observations of fare non-payment provide a more detailed look at fare payment patterns at the locations and times of observation, because a manual observer can see how many of the passengers boarding a vehicle either show a pass

or are clearly eligible for free travel as compared to those riders that appear to evade fares. By itself, this observational data can reveal some of the variation in fare payment behaviors across different parts of the system (e.g., bus vs. light rail, different routes, different times of day).

When compared with the results from AFC, APC, and ODX data, manual observations could also provide insight about the errors or level of uncertainty in estimates from automated methods. An analysis of outliers in the raw data and in the relative counts of AFC transactions and APC passenger boardings is used to filter data that appears to be in error. The product of this part of the research project is a replicable method for data processing and analysis to estimate the number of passengers that do not pay fares and the number of riders who are evading fares using continuously collected data from AFC, APC, inferred data from ODX, and any previous manual counts of fare payment activities.

Building on the analysis described above, patterns in the AFC and APC data that are correlated with higher rates of error or uncertainty are identified for data filtering and processing. Furthermore, a spatial and temporal analysis has been conducted to identify the locations and times when fare system non-interactions are most frequent and when revenue losses are estimated to be greatest. For example, bus stops, routes, neighborhoods, and times of day may be associated with higher rates of fare non-interaction, higher revenue loss, or greater uncertainty in fare evasion estimates. This analysis is used to identify where and when manual spot checks would provide the most value in terms of improving the accuracy in estimating fare non-payment and fare evasion. The product of this analysis is a method to identify where and when additional counts should be collected based on existing data and patterns that may be observed from the continuously collected data sources.

This project comes at an opportune time as the MBTA implements new fare collection technology and policies across the system. The Fare Transformation project includes a shift from legacy farecards and the ability to pay cash onboard to a tappable payment system that will enable riders to board through any door on buses on and light rail vehicles. As changes are made to the ways that fares are collected, it is important to have consistent and replicable methods to provide timely information on fare evasion to decision-makers. This will also be useful for making decisions around fare policies and fare engagement staffing.

1.2 Study Objectives

This project has two main objectives:

1. To use existing data sources to estimate the rates of fare payment compliance on MBTA buses and light rail services.
2. To develop a method to identify when and where manual spot checks of fare payment/evasion behaviors are most valuable.

This report presents a method for data processing and analysis to estimate the number of passengers that do not pay fares and the lost revenues due to non-interactions with the fare

payment system. Passengers that do not interact with the fare payment system include passengers that are exempt from fares (who are not required to interact with the fare system), those that hold valid passes or are transferring (who are supposed to interact with the fare system, but whose fare transaction does not involve collection of any money), and passengers that owe a fare and are therefore evading that fare amount.

The proposed method makes use of continuously collected data from AFC and APC. The report also presents a method to analyze the data to identify where and when the MBTA may consider conducting additional observational fare counts based on existing data and fare compliance patterns that may be observed from the continuously collected data sources.

This page left blank intentionally.

2.0 Research Methodology

The research approach for this study consists of four main components: a literature review to document the existing research on fare evasion and the use of technology to measure fare payment compliance, a review of previous manual observations of fare payment compliance on MBTA vehicles, an analysis of where and when there are non-interactions between the passengers and the fare payment system, and an analysis of the potential lost revenues associated with non-interactions.

The research methods developed in this study are focused primarily on leveraging the data that is continuously collected by the AFC and APC systems, which provide extensive coverage of the MBTA bus services. The results of these analyses are presented in Section 3.0. Implications for where and when to conduct additional manual observations, either for the purpose of improving the accuracy of lost revenue estimates or improving fare payment compliance, are discussed in Section 4.0.

2.1 Literature Review

Fare evasion is a problem in transit systems around the world when passengers do not pay the fare to use the transit system. The consequence is a loss of revenue for transit operators. This section presents a review of recent literature on fare evasion and non-interaction with fare payment systems. Although much of the literature is focused on the behavioral aspects of who is evading fares and why, this study focuses more attention on the methods for measuring fare payment and non-interactions as they relate to the problem of estimating fare compliance rates more generally.

2.1.1 Defining Fare Evasion and Fare Non-Interaction

Fare evasion occurs when a passenger lacks a valid or correct ticket, posing a threat to the finances of transit authorities or public transport companies. This issue has interdisciplinary implications related to travel demand, transport economics, and optimization of inspection programs [1]. There is not a single definition of fare evasion in the literature. Instead, authors define it based on the specific scope of their research and sometimes define subsets of evaders in different categories.

Barabino et al. [2] defines *fare evasion* as “the non-violent act of traveling on public transport in disregard of the law or regulation or contract, having deliberately not purchased, not validated or not correctly adopted the required travel ticket.” According to this explanation, fare evasion includes various behaviors:

1. **Freeloading**: Passengers travel without buying a ticket at all.
2. **Overriding**: Passengers cross multiple transit zones by paying only the basic fare.

3. **Elusion**: Passengers travel without validating their ticket, utilize fake tickets, or misuse existing media, among other unauthorized methods.

Keuchel and Laurenz [3] present another definition of fare evasion based on three groups: passengers traveling without a ticket, those with an invalid ticket, or individuals who have forgotten their tickets. Passengers who forget their tickets might be viewed as engaging in fare evasion from a fare enforcement perspective, leading to potential warnings or citations. However, these incidents do not always lead to a decrease in fare revenue for the transit agency, because the correct pass or ticket may have been purchased and is just not available for confirmation upon inspection.

In an investigation of the public transit system in Lyon, France, researchers classified fare evasion into two distinct groups based on whether it led to a loss in revenue. If a passenger forgets to tap when switching between vehicles or modes during the transfer time, it is considered a fare irregularity without loss of revenue. However, boarding for the first time without tapping under a pay-as-you-go fare system is seen as fare evasion as it also results in a loss of revenue [4,5].

It is important for transit agencies to differentiate between fare evasion leading to revenue loss and fare irregularity without financial impact. In both cases, a passenger did not interact with the fare payment or inspection system, which is referred to in this study as *non-interaction*. A non-interaction is an indication of the potential for lost revenue, but there is only a financial impact on the transit agency in the case of evasion. In the past, transit agencies primarily defined fare evasion and citation issuance based on cases causing direct revenue loss. However, distinguishing between these categories has become more complex due to the adoption of policies such as fare capping, which replaces passes with pay-as-you-go systems. These systems utilize data from the AFC system to establish pricing for pass programs or to allocate revenue sharing in multiagency transit systems [5].

2.1.2 Technologies for Collecting and Enforcing Fares

Fare enforcement methods are typically divided into three categories: onboard fare collection, gated stations or stops, and proof-of-payment systems.

2.1.2.1 Onboard Fare Collection

Onboard fares are collected when passengers pay their fare directly on the transit vehicle using various methods, including depositing cash into a farebox, tapping a smart card on an onboard validator, validating a mobile ticket electronically, swiping a magnetic stripe ticket, or presenting a valid paper ticket. The fare collection process is usually supervised by the operator, who observes passengers paying as they board the vehicle only through the front door.

2.1.2.2 Gated Stations or Stops

At gated stations, passengers must pay their fares at the fare gates before accessing the platform. The fare gates serve to enforce payment, allowing entry only after the fare is paid. This method is commonly used in stations with limited entry points, necessitating fewer

physical structures to prevent fare evasion. Additionally, transit agencies may staff stations with sworn peace officers or authorized civilian staff at fare gates and emergency exits. These personnel monitor fare payments and have the authority to issue citations if they witness a passenger entering a station or stop without paying, for example by hopping over a faregate [5].

2.1.2.3 Proof-of-Payment Systems

In proof-of-payment systems, passengers must legally purchase and validate a ticket before using the service, but there are no physical barriers to enforce this requirement. Payment validation does not occur instantly, and the effects of this delay can differ depending on the control system in place [6]. In most proof-of-payment systems, passengers need to buy or validate their tickets off-board before getting on the vehicle. They must keep their proof of payment with them for the whole journey to be able to confirm that a valid fare was paid in case of inspection. Although these systems usually involve off-board fare collection, some have onboard fare collection. This happens in situations where all-door boarding is allowed but it is not practical to have off-board fare collection at every stop due to costs [5].

2.1.3 Measuring Fare Non-Interaction and Evasion

Many studies have emphasized the importance of measuring and understanding the determinants of fare evasion, including factors like time of day, crowding, as well as demographic and socio-economic characteristics of communities served. Recent technological innovations, such as smart cards, provide new data sources to measure and estimate fare evasion rates more systematically.

Researchers investigating fare evasion in public transportation systems employ a variety of data sources and collection methods, broadly categorized into three main types: survey data, inspection data, and technology data.

2.1.3.1 Survey

Researchers have employed various surveys to study fare evasion for diverse objectives. Many researchers have utilized survey data to assess fare evasion in public transportation systems with a focus on fare evasion behaviors and socio-economic characteristics. For example, Lee [7] conducted a study of bus and light rail passengers on the San Francisco Municipal Transportation Agency (SFMTA) Muni system, employing an intercept survey to interview nearly 41,000 passengers during 1,141 transit vehicle trips. The findings revealed that at least 9.5% of the surveyed riders did not possess valid proof-of-payment. In another survey of the Calgary Train (CTrain) system, Hansen et al. [8] analyzed 33,499 survey cases and identified a fare evasion rate of 4.5%. They also found that regular fare checks by peace officers play a protective role, ensuring a well-ordered environment that, in turn, discourages both crime and social disorder. In a survey conducted on 110 one-way transit trips of Green Line light rail trains in Boston, Massachusetts, researchers estimated that around 22% of the 1,532 passengers that boarded at the rear doors evaded the fare [9]. Additionally, the study revealed that fare evasion rates were higher during the afternoon period.

2.1.3.2 Inspection

Researchers have also used inspection data to measure fare evasion in public transportation systems. Many parameters from inspection data can be used to quantify fare evasion. Common data include passenger warning (those who accidentally evade fare and/or are forgiven for their first offense), passenger citations (those fined for evasion), and inspected passengers (those with valid tickets checked by authorities). These factors aid in determining the *fare evasion rate*, which is the proportion of fare evaders to passengers who are inspected within a specified time period. However, there is disagreement among agencies and researchers about how to define "evaders" when calculating the fare evasion rate. Some include both warnings and citations issued, some only count citations, and a few also attempt to account for passengers who escape when inspectors board a vehicle. The variations in approach could introduce bias when comparing fare evasion rates across agencies or research studies. [2,10]

2.1.3.3 Technology

Smart cards have a built-in integrated circuit, often called a chip, where fare details are stored electronically. Unlike paper tickets, which are thrown away after they expire, smart cards can be recharged with a new fare. Even though smart cards represent the newest technology for ticketing, they do not stop passengers without a valid ticket from entering proof-of-payment systems [11]. A passenger can also potentially evade fares in gated systems by purchasing the wrong ticket type; for example, purchasing a discounted ticket when the full fare is owed.

A study of fare evasion in Montreal, Canada, determined that the connection of smart cards with APC data allows for the measurement of fare evasion in real time [12]. When all passengers pay the fare, the ratio between validations counted by the AFC and boardings counted by the APC approaches 1. Using a GIS-based map helps identify key network points for enforcing checks when this ratio falls below 1.

Despite the proliferation of technologies for collecting fares, counting passengers, and tracking vehicles, there has been limited research on how these sources can be used to estimate the rate of fare irregularities, fare equipment non-interactions, and fare evasions. Egu and Bonnel [4] introduced a classification of fare irregularities and quantitatively defined two fare irregularity rates. The relevant fare collection data is listed in Table 2.1 [4], with the following relationships between the sets: $V \subset \Omega$, $C_t \subset \Omega$, $C_p \subset C_t$, and $C_{pp} \subset C_p$.

Table 2.1 Fare Collection Data Parameters

Symbol	Description	Source of Data
$\boldsymbol{\Omega}$	Set of all boardings	Automated Passenger Counters
$\boldsymbol{V}$	Set of all boardings with a fare transaction	Automatic Fare Collection
$\boldsymbol{C_t}$	Set of all boardings with a fare inspection	Fare Inspection System
$\boldsymbol{C_p}$	Set of all boardings with a fare inspection resulting in an irregularity	Fare Inspection System
$\boldsymbol{C_{pp}}$	Set of all boardings with a fare inspection resulting in an irregularity with lost revenue	Fare Inspection System

Considering only the set of boardings with a fare inspection, the total irregularity rate can be computed as

$$C_i = \frac{|C_p|}{C_t} \tag{1}$$

where the denominator, C_t, is a subset of Ω, whose size depends on the intensity of inspection. Since not all fare irregularities are associated with a loss of revenue, a second measure for the rate of fare evasion with loss of revenue is defined as

$$C_r = \frac{|C_{pp}|}{C_t} \tag{2}$$

using C_{pp} as the numerator. Both rates are measures that are based on the intensity of fare inspections, which require personnel to check if fares were paid.

Considering the complete universe of passenger boardings, AFC data and APC data can be used to define the following ratio

$$V_i = 100 - \frac{|\mathrm{V}|}{|\Omega|} \tag{3}$$

which can be interpreted as the rate of fare non-validation or non-interaction; i.e., the percentage of boardings without a corresponding fare transaction. Unlike the measures that depend on inspections by personnel, V_i can be calculated automatically wherever and whenever AFC and APC data are available.

Egu and Bonnel [4] applied these measures to estimate fare evasions in the context of the public transport network in Lyon, France. The findings of their study imply that using fare inspection logs may have notable limitations for accurately gauging the extent of fare evasion, in part because the inspection rate is low, roughly 1.3% in Lyon. The authors propose that exploring the combination of APC and AFC transactions holds more promise for future research.

A more recent research study to investigate fare non-interactions was conducted within the tram network in Melbourne, Australia. The study utilized extensive automatically gathered data, which included information from APC and AFC systems. This data was employed to assess the prevalence of fare non-interactions. Their results indicated that fare non-interactions were less common at stops near train stations, educational facilities, frequently inspected stops, and during peak hours. Conversely, fare non-interactions were more prevalent at stops with high boarding flows, crowded services, during late evening periods, and on weekends. [13]

A summary of recent research to measure and estimate fare evasion rates is presented in Table 2.2. A common pattern is that studies based only on automatically collected data estimate either the number of evasions or the number of fare interactions. The studies that seek to go a step further to estimate the severity of the non-interaction, for example to estimate the magnitude of revenue loss, all include at least some survey or inspection data. As indicated in the rate defined by [4] in Equations (1–3), inspection data is needed to distinguish between the types of fare irregularities that do or do not result in lost revenue.

2.2 Analysis of Manual Observations

The analysis in this study begins with a review of the insights from available data from manual observations on MBTA light rail and buses. The purpose of looking at manual observations is that the direct inspections of passenger fare payments provide a detailed perspective of fare payment compliance. It turns out there have not been recent or systematic surveys or manual counts of fare payment behaviors on light rail and buses. The most recent reports date back to 2016, which was before the start of the MBTA's Fare Transformation project. The most relevant data for the MBTA is the Green Line rear door boarding study [9] and the systemwide fare compliance data compiled by the Central Transportation Planning Staff (CTPS) [14].

2.3 Automatically Collected Data

Comprehensive datasets of continuously recorded data include Automatic Fare Collection (AFC), Automatic Vehicle Location (AVL), and Automatic Passenger Count (APC) data. These are archived and available through the MBTA Research Database, which is an archive of several datasets that includes transit route, schedule, and stop information. Some of these data are linked and processed to provide estimates of origin-destination patterns across the systems based on the origin-destination-transfer (ODX) model. The general structure of the approach is shown in Figure 2.1, with data from APC and AFC being jointly processed to analysis non-interactions and lost revenues.

The relevant data for this study are from three data tables: AFC.faretransactions, ODX.odx, and APC.stops. Each table contains the following relevant data fields.

Table 2.2 Studies on Fare Evasion Measurement

Citation	City	Data Source	Study Focus	Target Value
Barabino, Di Francesco, Ventura [15]	Cagliari, Italy	Inspection and Passenger Survey	Developed a formal framework for evaluating fare evasion frequency and severity	Fare evasion frequency as number of evasions in time period; Fare evasion severity as consequences of a fare (revenue loss)
Yin, Nassir, Leong, Tanin, Sarvi [13]	Melbourne, Australia	APC and AFC	Analysis of fare non-interactions in the tram network and the factors influencing them	Number of fare evasions
Munizaga, Gschwender, Gallegos [16]	Santiago, Chile	Smart Card Transactions	The study focus is the development and application of a method to incorporate fare evasion correction factors to public transport OD matrices obtained from AFC and GPS data	Partial evasion (during a bus trip stage prior to a Metro trip stage); Total evasion (during all bus-only trip stages)
Egu and Bonnel [4]	Lyon, France	APC, AFC, and Fare Inspections	Introducing two classifications for fare irregularities based on loss of revenue using APC and AFC and inspection data	Fare irregularity with and without revenue loss rates
Sánchez-Martínez [17]	Boston, USA	Smart card Transactions	A stochastic model and framework to estimate fare non-interaction through analysis of travel patterns seen in disaggregate fare transaction data	Average daily non-interaction at various stops; The probability of non-interaction with and without fare evasion
Prokosch and Gartsman [9]	Boston, USA	Observed Passenger Boardings	Observing passengers boarding through all rear doors during one-way trips on two-car Green Line trains to estimate the fare evasion rate	Annual fare evasion rate and lost revenue
Pourmonet, Bassetto, and Trépanier [12]	Montréal, Canada	Smart Card and APC	Linking smart cards and APC for analysis of fare evasion	Ratio between validations and boardings

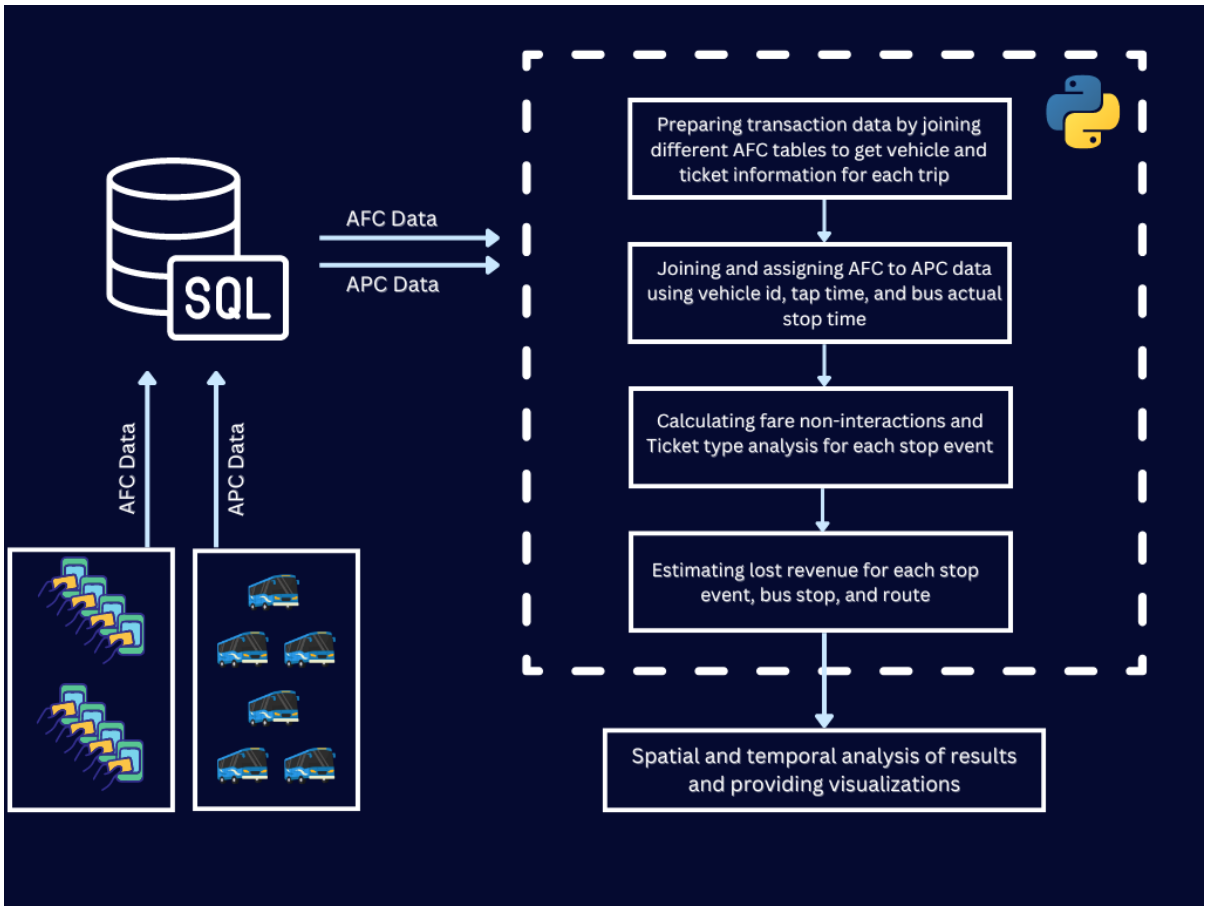

Figure 2.1 Flowchart of Data and Analysis of Non-Interactions

2.3.1 Automatic Fare Collection (AFC)

Automatic fare collection data is collected from the fare collection system at station fare gates and on-board fareboxes on buses and light rail vehicles. The AFC records are associated with cash payments and events in which Charlie Cards or Charlie Tickets are used to load value, pay a fare, or validate a pass. The data is partitioned by month and year and includes records of Charlie Card transactions from individual fare cards as well as passes. The specific data table used is AFC.faretransactions.

AFC.faretransactions contains a record for every fare transaction.

- `faretransaction_key`: a unique identifier for each fare transaction
- `deviceid`: (matches `vehicle` in other tables) a unique identifier for the farebox on each bus or light rail vehicle
- `trxtime`: (matches `tap_time` in other tables) the timestamp with date and time of the fare transaction

- `card`: the card/ticket serial number from the AFC system, which allows for multiple payments by the same passenger to be linked. Cash fares are listed as "cash"

- `tickettypeid`: the type of fare interaction that occurred; e.g., fare deduction, validation, value top-up (possible at vending machines)

- `amount`: the amount paid in the fare transaction in cents

- `bookcanc`: a binary variable indicating whether a fare transaction is a fare deduction or validation (=1) or voiding a previous transaction (=-1)

From this raw data, the counts of passengers that pay fares upon boarding each vehicle are recorded. The device ID and timestamp of transaction can be used to link fare transactions with the location of the vehicle when the fare was paid. As all MBTA systems charge fare only on entry, there are no fare data associated with exits.

2.3.2 Origin-Destination-Transfer Model (ODX)

A model to link trip records and infer origin-destination and transfer patterns in the system has been developed to populate a database of ODX records. Inference models based on farecard data have been improved over the years. The model identifies records from AFC that can be linked to infer transfers or return trip patterns. Several steps are needed to infer destinations and transfer locations for rail passengers, because movements behind faregates are not tracked.

ODX.odx contains trip records with inferred destinations based on a model of origin-destination patterns that is developed from farecard data. The ODX.odx table is useful because it includes a processed version of the fare transaction records that keeps only transactions associated with payment or validation of a fare. As a result, the ODX provide a count for fare transactions that is likely a better representation of actual number of passengers that pay fares at the farebox than the raw AFC.faretransaction data. Another benefit is that the ODX.odx table links the transactions with a route, so that all records associated with a specific route can be directly queried from the database.

- `card`: the card/ticket serial number from the AFC system, which allows for multiple payments by the same passenger to be linked

- `vehicle`: a unique identifier for the farebox on each bus or light rail vehicle

- `tap_time`: the timestamp with date and time of the fare transaction

- `faretransactionkey`: a unique identifier for each fare transaction

- `origin`: (matches `stopid` in other tables) the unique identifier from the GTFS table for the stop where the transaction occurred

- `route`: the route on which the vehicle is operating at the time of the transaction

2.3.3 Automatic Passenger Counter (APC)

Automatic passenger counters (APC) are devices that count the number of passengers boarding and alighting each vehicle. APC devices are not in widespread deployment on MBTA rail vehicles but are now nearly universal on the fleet of buses.

APC.stops includes the counts or boarding and alighting passengers from each vehicle stop event.

- `stopid`: the unique identifier from the GTFS table for the stop where passengers were counted
- `stopname`: the written name of the bus stop, typically nearest crossing streets
- `actstoptime`: the timestamp with date and time of the actual bus stop event, which appears to be the time when the doors open, allowing boarding to begin
- `psgron`: number of passengers counted boarding
- `psgroff`: number of passengers counted alighting
- `psgrload`: estimated number of passengers onboard based on the difference of the cumulative sum of `psgron` and `psgroff`
- `route`: the route on which the vehicle is operating at the time of the stop
- `lat` & `long`: latitude and longitude of stop location
- `bus`: (matches `vehicle` in other tables) a unique identifier for the vehicle on which the APC is located

This raw data provides a measure of the number of people who board a vehicle. Since the vehicle ID in the APC data and the device ID in the fare transaction data can be joined, it is possible to group fare transactions associated with same vehicle stop based on the time stamp.

2.3.4 Data Availability

Although measurements from AFC and APC are associated with errors, the AFC data provides an estimate of the number of fare transactions, which can be compared with the estimate of total boarding passengers from APC to calculate a value of V_i. For the purposes of this project, it is important to identify how many of the vehicles have complete AFC and APC data, because both need to be matched to estimate fare non-interactions. Figure 2.2 shows a time series of the number of unique vehicle IDs for buses in the ODX records and

the APC records within the MBTA Research Database. Over time, the number of vehicles in the fleet that are equipped with APC devices has steadily increased relative to the total number of buses in operation, so now most of the bus fleet is equipped with the technology to allow fare non-interactions to be estimated automatically.

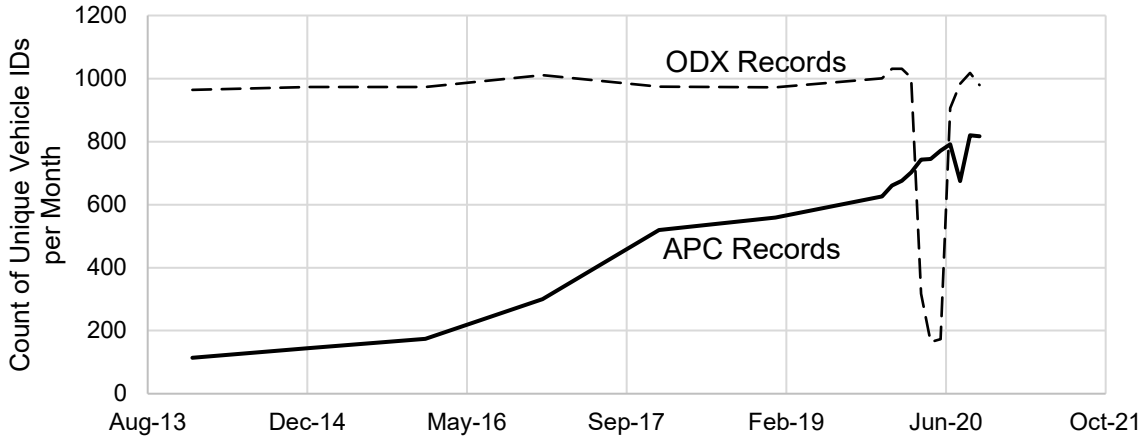

Figure 2.2 Buses in ODX and APC Databases

The MBTA research database only provides APC records through 2020. Data during the reduced service period associated with the beginning of the COVID-19 pandemic also has some discrepancies, with the number of vehicle IDs in ODX records dropping significantly before rebounding in the last months of the year. For the purposes of fare evasion analyses, data are considered from Fall 2019, because travel behaviors and data during 2020 are not indicative of patterns before and after that period.

2.4 Linking Data to Estimate Fare Systems Non-Interactions

For a scope of analysis (e.g., a single vehicle, a route, the whole system, etc. for an hour, a day, etc.), the relevant records can be extracted from the MBTA Research Database and joined to estimate the number of fare system non-interactions and the rate of non-interactions per boarding passenger.

2.4.1 Defining Non-Interactions and Non-Interaction Rate

This study makes use of the method presented in Egu and Bonnell [4] to estimate non-interactions with fare payment and collection systems during stop events. They propose using automatically collected data to define two quantities:

- Ω, the set of all passenger boardings (observed from APC installed on vehicles)
- V, the set of all passenger boardings with recorded fare transactions (observed from the AFC system, based on farebox interactions)

For a defined scope (e.g., set of vehicles and time period), the number of fare system non-interactions, C, is the difference between boardings and fare interactions.

$$C = \Omega - V \tag{4}$$

The finest granularity for which this measure is meaningful is a single vehicle stop event; i.e., a vehicle opening doors at a stop to allow passengers to board. In this instance, the APC counts the number of passengers that enter through each door, the sum of which is Ω. Each passenger that interacts with the farebox by tapping a pass, tapping a farecard, paying with cash, or being recorded by the operator, is counted toward V. The difference is the number of passengers that are not accounted for in the fare collection system.

Another useful measure of fare payment compliance is the non-interaction rate, R, which is the proportion of boarding passengers that do not interact with the fare payment system.

$$R = 1 - \frac{V}{\Omega} \tag{5}$$

Like the fare non-interaction count, the finest meaningful granularity for non-interaction rate is at the level of a single vehicle stop event. However, it is also useful to look at patterns across aggregated transit operations for a bus stop, route, system, time of day, etc.

In an ideal system in which APC and AFC measures of passenger movements and fare payments are without errors and in which all passengers interact with the fare collection system, there would be no non-interactions (i.e., $C = 0$ and $R = 0$). In reality, there may be some errors associated with data collected by APC and AFC systems. Likewise, some passengers may not interact with fare collection equipment, perhaps due to technical malfunction, accidental failure to tap or show a pass, or deliberate fare evasion. As measures of fare payment compliance, the non-interaction count is an indicator of the magnitude of fares unpaid, and the non-interaction rate is an indicator of the probability that a passenger does not pay a fare.

2.4.2 Process for Estimating Non-Interactions

The process for estimating fare non-interactions is guided by equations (4) and (5), making use of the data described in Section 2.4.1. The overall process is summarized graphically in Figure 2.3, which shows that data from APC.stops and AFC.faretransactions are joined in an analysis procedure to count the number of fare system non-interactions, which can also be reported as a rate of non-interactions.

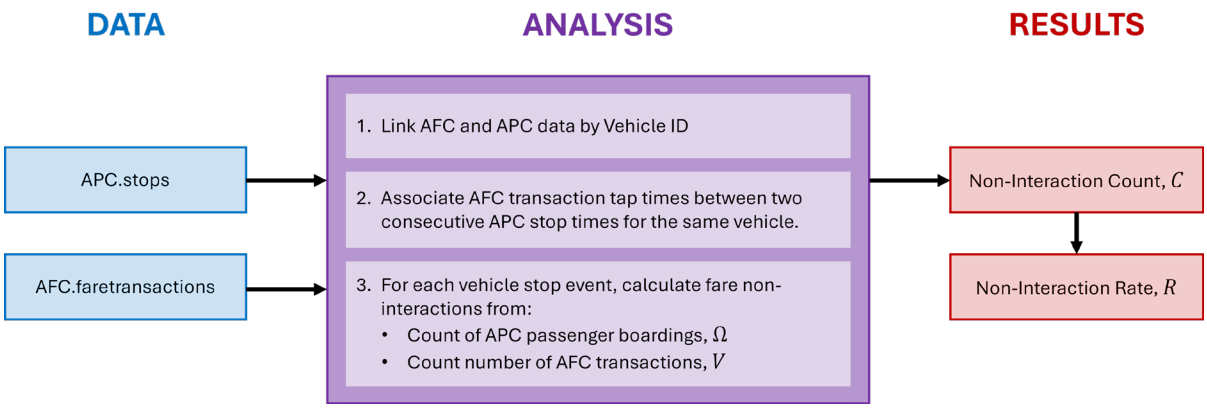

Figure 2.3 Data Analysis Process for Non-Interactions

Although the column titles differ somewhat across tables, their common meaning allows the data to be linked as illustrated in Figure 2.4. In the figure, solid arrows indicate common fields that can be used as a direct join. The dashed arrow from ODX.odx.tap_time to APC.stops.actstoptime indicates two timestamps that can be directly compared to link counts of fare transactions and counts of passenger boardings for individual stop events.

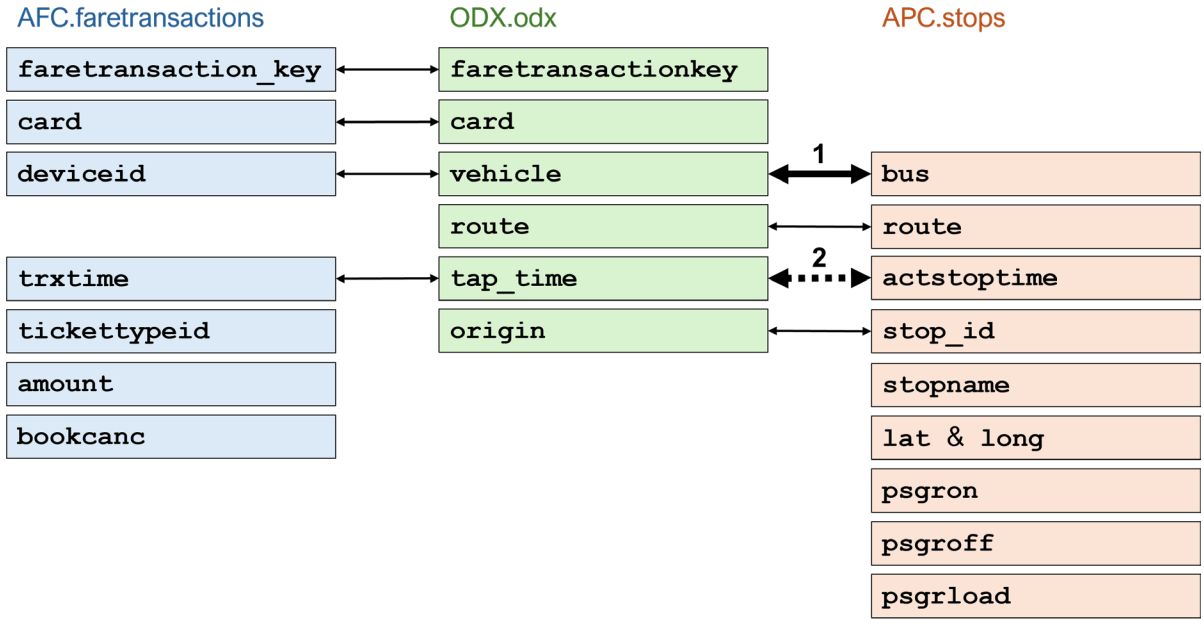

Figure 2.4 Data table linkages from MBTA Research Database

The method is to estimate values of non-interaction count, C, and non-interaction rate, R, for each vehicle stop event within the scope of analysis. The process uses the following steps:

1. Using SQL queries, data are extracted for the vehicles/routes and time period of analysis from AFC.transactions, ODX.odx, and APC.stops.

2. Records of vehicle stop events from APC.stops are joined with records of fare transactions from ODX.odx by the fields APC.stops.bus and ODX.odx.vehicle, which contain the vehicle ID number. This join is marked with "1" in Figure 2.4.

3. Records are sorted by vehicle and timestamp. For a given vehicle, the set of fare transactions that occur between two consecutive stop events, based on the timestamp, are associated with the preceding vehicle stop event. This linkage is marked with "2" in Figure 2.4. An example for a single vehicle is shown in Table 2.3 for a 5-minute period from 9:10–9:15 a.m. on September 18, 2019.

Table 2.3 APC and ODX Data for Vehicle 1792, Route 1, 9/18/2019, 9:10–9:15 a.m.

APC.stops .stopname	**APC.stops** .actstoptime	**APC.stops** .psgron, Ω	**ODX.odx** .trxtime	**ODX.odx** V	**NI Count** C	**NI Rate** R
Mt Auburn@Dewolfe	9:10:42	1	9:10:42	1	0	0
Mt Auburn@Putnam	9:11:51	6	9:11:53	5	1	0.167
			9:12:00			
			9:12:37			
			9:12:42			
			9:12:44			
Mass Ave@Bay	9:13:31	1	9:13:36	1	0	0
Mass Ave@Hancock	9:14:04	3	9:14:05	3	0	0
			9:14:06			
			9:14:08			
Total	—	11	—	10	1	0.091

4. For each vehicle stop event, the reported number of passenger boardings from the APC data are the Ω values, and the count of associated fare transactions are the V values. Table 2.3 shows these values for a sequence of 4 vehicle stop events that occur on a single vehicle in a 5-minute period. The values can also be aggregated for a time period of interest, as shown in the row of totals.

5. For each vehicle stop event, the non-interaction count (C) and non-interaction rate (R) are calculated using equations (4) and (5). Table 2.3 shows that for the 4 observed vehicle stop events in the sample, there was only 1 non-interaction at 1 stop (Mt. Auburn Street @ Putnam Street), which was a non-interaction rate of 0.167 (16.7%) for that stop event. Over the 5-minute sample, the aggregated data showed an average non-interaction rate of 0.091 (9.1%).

This method has been implemented using code in R to process extracted data from the MBTA Research Database for the entire bus system, which is used to calculate the fare non-interaction rate for the entire bus system. Since the granularity of the data is vehicle stop events, these records can be sorted and analyzed by location and time with various levels of aggregation.

2.5 Method for Identifying Data Outliers

An important part of the data analysis is to identify potential errors in the counts of boarding passengers from APC devices and any errors from malfunctioning AFC devices, because errors in measurements contribute to errors in estimated fare system non-interactions. Without additional validation data, the method for identifying errors relies on identifying outliers. Vehicles that have either outlying APC counts or outlying AFC records are flagged as unreliable data sources so that they can be removed from the analysis of non-interaction counts and non-interaction rates.

2.5.1 Assumptions About Data Sources

The calculations of fare non-interactions are based on two basic data sources: APC records and AFC records. Since these data are collected using devices in the real world, they are subject to errors. The following assumptions are made to identify and correct potential errors in the data.

1. *AFC records are more accurate than APC records.* Fare transactions are tracked as individual events, with data on fare payment medium (i.e., Charlie Card serial number) and amount of fare collected. It is assumed that all fare system interactions that actually occur are included in the AFC tables, because these transactions are tracked as part of the MBTA's revenue system. On the other hand, APC records are passenger counts based on passive detectors mounted onboard vehicles. APC records are known to be prone to errors, especially in crowded conditions when it may be difficult to detect the difference between two passengers that board in very close proximity [18]. Furthermore, APC detectors are susceptible to errors if blocked by a standing passenger or large items, such as strollers or luggage.

2. *Passenger boardings must be at least as great as fare transactions* ($\Omega \geq V$). It stands to reason that the number of fare transactions associated with a vehicle stop event should not exceed the number of passengers that board, because no passenger is required to pay more than once. Most passengers interact with the fare payment system by tapping a card or inserting cash into the farebox upon boarding. This assumption may be violated in either of the following cases:

    a. A passenger interacts with the fare payment system after a subsequent stop from where they boarded.

    b. The number of boarding passengers is undercounted by the APC.

In case (a), a passenger would be counted as non-interacting at the stop they boarded but then would be counted as a fare transaction mistakenly associated with a subsequent stop, perhaps because they were looking for their farecard or money while riding. This could result in a situation that the subsequent stop has more fare transactions than boardings, which would be flagged as an error, or this could lead to undercounting the number of non-interactions, which would be difficult to detect. Either way, an aggregated count of fare transactions for the route or system would include the correct total, so there would be no aggregated error in the non-interaction count or rate. In case (b), an uncounted boarding passenger would introduce an error in the calculated non-interaction count and non-interaction rate.

The analysis of outliers described in the following subsections is designed to identify observations that are either excessive in magnitude or contradict assumption 2. The resulting process ensures that the non-interaction counts and rates are never negative, which would be impossible. A more sophisticated analysis of APC count errors would require more detailed data on APC accuracy, especially because the relationship between accuracy and number of boarding passengers is non-linear [18].

2.5.2 Quality Control Counts

An initial quality control check is included in the APC.stops data, which is a comparison of the daily total of boarding and alighting passengers counted by each vehicle's APC devices. If all APC counts are correct, the total number of boardings and alighting passengers should be equal, because vehicles leave and return to the garage with no passengers onboard. The APC.stops.QC_count value reports the percent difference between the two counts. MBTA's practice is to ignore records from vehicles on days with QC_count exceeding 20%. The same threshold is adopted for this study, with the corresponding records being removed from the analysis.

Recognizing that the numbers of boarding and alighting passengers counted by APC can differ, the MBTA rebalances counts so that the number of passengers onboard each bus is recalibrated to 0 at the end of each line. The rebalancing process allocates missing observations proportionally (in whole numbers) to the stop events with the greatest counts. The rebalanced APC counts are used for the analysis in this project.

2.5.3 Definition of Outliers

The interquartile range (IQR) method is used to identify outliers based on the spread of observed values in the data set. Data values are sorted from least to greatest to identify the first quartile (25th percentile), Q_1, and the third quartile (75th percentile), Q_3. The IQR is the difference of these values.

$$IQR = Q_3 - Q_1 \tag{6}$$

Data values are flagged as outliers if they fall below a lower bound, $Q_1 - 1.5IQR$, or above an upper bound, $Q_3 + 1.5IQR$. The IQR method provides a quantitative and consistent way to identify outliers in the data as part of a cleaning process for data analysis and modeling.

On a box-and-whisker plot, a box is drawn from Q_1 to Q_3 to indicate the IQR. Whiskers are then drawn to the lower and upper bounds. A line through the box indicates the median (50^{th} percentile). Any individual outliers are then plotted individually. This type of plot provides a visualization for the spread of values in a data set [19].

2.5.4 Outliers in APC Counts

Each passenger that boards a transit vehicle should interact with the farebox once: either to pay the required fare or to confirm possession of a valid pass or exemption from fare. Therefore, the number of passengers that interact with the farebox on a transit vehicle should not exceed the number of passengers that board the vehicle. In reality, it is unlikely for a farebox to record fare transactions that do not exist, because each record is connected with detailed fare payment information such as the specific farecard or pass used and the amount paid. APC devices, however, are subject to counting errors, because the technology used to detect passengers can make errors in distinguishing between multiple individuals, especially in crowded conditions. As a result, APC counts are susceptible to both positive and negative errors, although they are more likely to undercount passenger boardings [18]. In some cases, undercounting by a small number of passengers may be hard to detect, but aggregating observations by vehicle can reveal devices that are consistently in error.

In this study, APC devices are identified as outliers based on the comparison of observed APC boardings, Ω_o, and counted fare transactions, V, for each stop event. Then, stop events are aggregated by vehicle ID to identify devices that can be considered outliers based on how frequently passengers are undercounted. This is a two-step process:

1. For each stop event, difference between the AFC and APC counts is calculated

$$\Delta = \Omega_o - V. \tag{7}$$

   The stop event is flagged as an *APC error* if $\Delta < 0$ or, equivalently, $V > \Omega_o$, because it is not possible for more fares to be paid than passengers boarding. This condition suggests that the APC device must have undercounted passengers.

2. Stop event data is aggregated by vehicle ID, and an *APC error ratio* is calculated as the ratio of the number of APC errors to the total number of stop events observed. An analysis of the error ratios is done to determine which vehicles can be counted as outliers based on how often a stop event counts as an APC error.

Figure 2.5 shows that the Silver Line has much larger and more varied APC error ratios compared to other bus routes. When all of the data is included, the high values from the Silver Line significantly impact the overall distribution, leading to a higher upper bound and more outliers. When the Silver Line data are excluded, the error ratios are lower and more consistent. The Silver Line differs from other bus routes in that no fares are collected at Logan Airport stops (Silver Line route 1), and underground stations are controlled by faregates for parts of Silver Line routes 1, 2, and 3. The different operating conditions appear to somewhat affect the APC error ratio and to significantly affect the calculated non-interaction rate (described in the next subsection), so Silver Line data is separated from the

other bus data to improve the accuracy of this analysis. Consequently, records related to vehicles identified as outliers in this analysis are removed from the data. Eliminating undercounted observations from APC devices that appear to be consistently reporting errors is intended to remove biased observations and will lead to increased estimates of fare non-interaction counts.

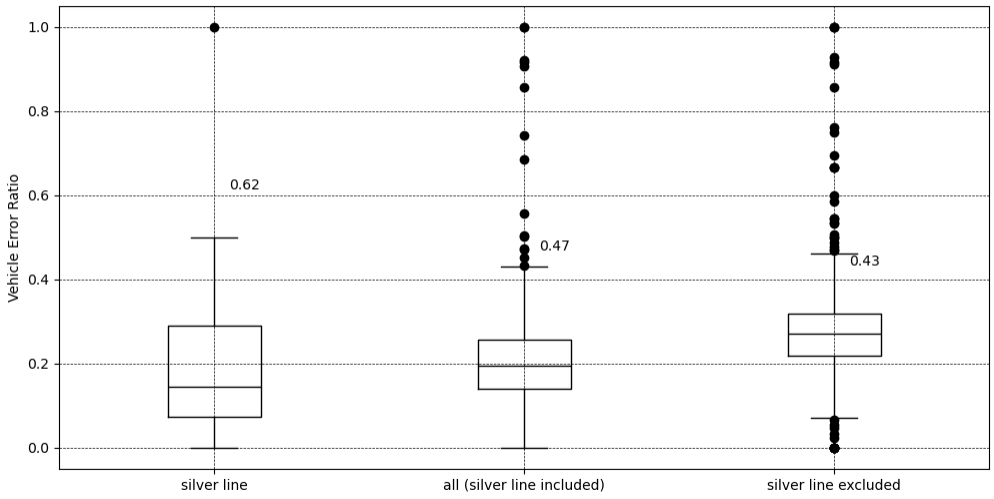

Figure 2.5 APC Error Ratios for All Data, Silver Line Excluded, and Only Silver Line

In this analysis process, APC devices that overcount passengers are not considered because it less likely that the APC observation is less than the AFC count. This is an error that is much harder to identify but is included in the analysis of non-interaction counts, described below.

2.5.5 Outliers in Non-Interaction Counts

The analysis of APC counts identifies observations of negative non-interaction counts as APC errors. The other type of error to identify is excessively high non-interaction counts. Conceptually, this error is harder to identify because it is physically possible for many passengers to board a vehicle and not interact with the AFC system. An overestimate of non-interaction can occur if the APC overcounts the number of boarding passengers and/or the AFC device under-records actual attempts to interact with the AFC system. From a single data point, it is not possible to know whether a high non-interaction count (high $\Delta = \Omega_o - V$) represents a large number of actual fare non-interactions by passengers or malfunctioning equipment.

Again, outliers are identified based on the comparison of observed APC boardings, Ω_o, and counted fare transactions, V, for each stop event. Then, stop events are aggregated by vehicle ID to quantify the non-interaction ratio for the vehicle. This is a two-step process:

1. For each stop event, the difference between the AFC and APC counts is calculated for Δ as given by Equation (7).

2. Stop event data is aggregated by vehicle ID to calculate the non-interaction rate

$$R = 1 - \frac{V}{\Omega_o} \tag{8}$$

for each vehicle. An analysis of the non-interaction rates is done to determine which vehicles can be counted as outliers based on the magnitudes of these estimates.

Figure 2.6 shows the box-and-whisker plot for the non-interaction rate (NI_rate) for three groups: the Silver Line, all data (including the Silver Line), and all data excluding the Silver Line. Vehicles with an NI_rate above the upper limit are seen as outliers and are removed from our analysis. The Silver Line has much higher and more variable NI_rate values than other bus routes, because no fares are collected onboard at Logan Airport and underground gated stations. For the combined data, the upper limit is 0.43, and it is the same for the data excluding the Silver Line. These limits is used to identify and remove vehicles that might be overcounting non-interactions due to errors in the APC or AFC systems.

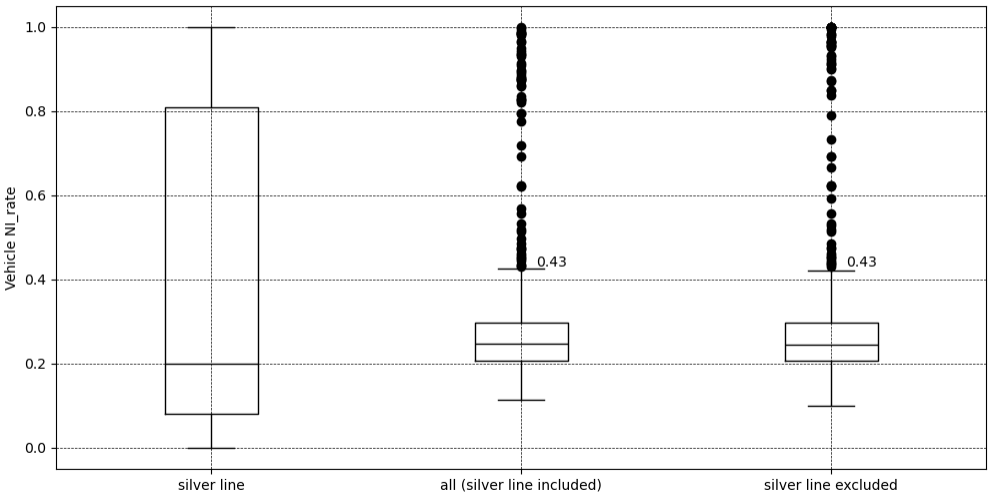

Figure 2.6 Non-interaction Rates by Vehicles for All Data, Silver Line Excluded, and Only Silver Line

Removing these outliers is intended to remove biased data, leading to more accurate estimates of fare non-interactions. Since the Silver Line behaves differently and can skew the overall analysis, it is important to study it separately. As a result, any vehicles identified as outliers in this analysis are removed from the dataset.

The fare non-interaction rate represents that ratio of passengers that do not interact with the AFC system, which can also be interpreted as the probability of non-interaction. This value is

calculated based on the count of fare non-interactions is calculated for each stop event. The difference between raw count of observed boardings from APC devices and the number of fare transactions recorded by the AFC system at each stop event is the Δ value defined in equation (2). An adjusted count of boardings, Ω, is defined by

$$\Omega = \max\{\Omega_o, V\} \tag{9}$$

to correct the APC errors that are not associated with the vehicle outliers. The effect is that negative values of Δ become 0.

The corrected APC counts of boarding passengers and the counts of AFC system records can be aggregated at different scales to identify the highest rates of fare non-interactions. At the level of individual stops, the Ω is the sum of corrected APC counts and V is the sum of AFC records for all observed stop events associated with each bus stop. Then, the non-interaction rate, R, is calculated as in equation (8).

2.6 Estimating Lost Revenues

The relationship between fare non-interactions and lost revenue is complex. Simply multiplying the number of non-interactions (C) by the bus fare would imply that every non-interacting passenger should have paid a full fare. Passengers paying with cash or stored value Charlie Cards must interact with the farebox for their fare to be collected; non-interactions with these passengers represent the lost revenue of a fare. However, there are many situations in which the fare system does not collect any money from a boarding passenger. Passengers that transfer from rail or a bus with equal or greater fare do not pay an additional fare. Many passengers use weekly or monthly passes that allow unlimited travel without additional amounts deducted per trip. Even if transferring passengers or passholders do not interact with the farebox, there is no lost revenue, because the contact with the AFC system is only to create a record that valid fare media was used. There are also passengers that are exempt from paying fares altogether, including MBTA employees and children under 12 years of age, and these exempt passengers do not interact with the fare system at all.

To calculate the lost revenues associated with non-interacting passengers, the following questions need to be answered:

1) How many of the non-interacting passengers owe a fare at the farebox, and what amount of fare do they owe?

2) How many of the non-interacting passengers do not owe additional fare because they are transferring or hold a valid pass?

3) How many of the non-interacting passengers are exempt from fares and therefore do not interact with the fare system in any case?

Definitive answers to these questions would require manual inspections or surveys of passengers, because the automated systems do not collect this information.

Without additional observations, lost revenues can only be estimated. The possible range of lost revenues is large:

- **Best-Case Scenario**: All non-interacting passengers are exempt from fares, hold a valid pass, or making a valid transfer, all of which require no additional fare payment. In this case, no revenues would be lost at all.

- **Worst-Case Scenario**: All non-interacting passengers are evading full fares, so the lost revenues would equal the non-interaction rate multiplied by the full fare ($1.70 on local buses, more on express services).

The real lost revenues are more likely somewhere in between. Using only AFC and APC data, lost revenues can only be estimated by making some assumptions. For this study, two assumptions are made and considered separately:

1) The composition of fare transactions for interacting passengers is representative of the composition of non-interacting passengers. This assumption implies that calculating the average amount of fare paid during each observed transaction is representative of the lost revenue associated with each non-interaction.

2) Children are a significant portion of the fare-exempt passengers. The estimated percentage of bus riders that are children is subtracted from the non-interaction rate, because these riders are known to pay no fare.

The method proposed in this study for estimating the lost revenues associated with non-interacting passengers is to make use of the observed fare payments for interacting passengers in accordance with the first assumption. Figure 2.7 summarizes the process for estimating the lost revenues based on fare transaction data and the non-interaction rate.

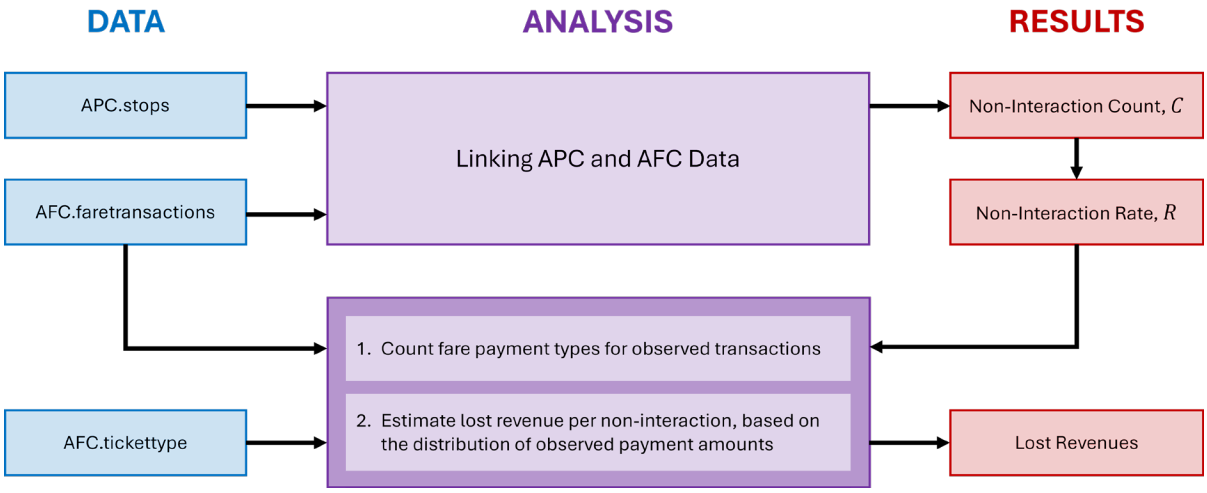

Figure 2.7 Data Analysis Process for Lost Revenue

There are some caveats associated with the first assumption, because different circumstances may lead to passengers being more or less likely to interact with the fare system. Examples of errors associated with this assumption include:

- Passengers holding valid passes or making a valid transfer would have less incentive to avoid a fare interaction than a passenger who must pay a fare. This would make actual lost revenues greater.

- Although technically against MBTA policy, operators may waive on passengers or open back doors at stops with significant numbers of transferring passengers (e.g., at rail stations where many passengers transfer onto buses), because these transfer passengers do not owe additional fare. This would make actual lost revenues lower.

These errors can only be quantified by conducting manual inspections or surveys to compare the composition of non-interacting passengers with fare transaction data. Furthermore, the second assumption relies on having data about the number of passengers that are exempt from paying fares, but these passengers do not interact with the fare system and could only be observed manually.

The proposed analysis of lost revenues includes two parts. First, the fare types among observed transactions are analyzed to show the composition of the interacting passengers. This can be compared against previous manual counts from the 2017 Green Line rear door boarding study [9] to assess the validity of the first assumption. Second, the average amount of fare collected with each observed transaction is calculated to provide an estimate of the lost revenue per non-interacting passenger. This average includes no fare collected from passholders and transferring passengers, any discounts that eligible passengers receive, and the actual full fare paid.

2.6.1 Fare Payment Types

The fare payment types for each transaction are recorded in the AFC database. There are 60 different fare payment types, which can be categorized by their impact on revenue. Passengers that hold valid passes do not pay any additional fare upon boarding a bus. Passengers that use a stored-value card may pay a full fare or be eligible for a free transfer.

The following types of passengers are eligible to pay 50% reduced fares as of 2019:

- People with disabilities and Medicare cardholders

- People 65 and older

- Middle and High School students who attend a school in the MBTA's Student Pass Program

- People aged 18-25 with low income

The following passengers aways ride free:

- MBTA employees
- Children 11 years old and younger
- People who are legally blind
- Uniformed military personnel
- Police and firefighters
- Government officials

Aggregated data for observed fare transactions on buses are presented in Appendix A.

Since there are no data are associated with the fare non-interactions, it is not possible to know the exact composition of passengers that do not interact with the AFC system. The distribution of fare payment types can be compared with the observations of rear-door boarding passengers from the 2017 Green Line study [9] to assess the likelihood if deviations that would contribute to errors in lost revenue estimates.

2.6.2 Estimated Lost Revenues per Non-Interaction

Although some fare payment types are always associated with no additional fares collected (e.g., valid passes), many fare types are associated with a range of fare amounts depending on the type of service (e.g., local bus fare $1.70, express bus service $4.25), transfers, or other discounts. Since the amount of fare collected with each transaction is included in the AFC.faretransactions database, the average fare collected from interacting customers can be calculated for any set of fare transactions.

In this study, the composition of fares among observed transactions is assumed to be representative of the composition of non-interacted passengers at each bus stop, route, or time period. Within each period of analysis (i.e., day, time period, location), the average amount of fare collected from each interacting passenger is an estimate of the lost revenue per non-interacting passenger. In aggregating transactions by location and time of day, this method accounts for variations in the locations and times where a greater share of passengers may be using passes or transferring from other modes.

This analysis comes with the caveat that is possible that actual lost revenues are greater if fare evaders who should be paying a fare are over-represented in the population of non-interacting passengers. Likewise, it is possible that lost revenues are lower if non-interacting passengers include a greater share of pass holders and exempt riders who do not owe any additional fare upon boarding.

Furthermore, this method does not account for the number of exempt riders who never interact with the fare system. This group includes children under 12 years of age, for whom

there is no specific data about spatial and temporal distributions of ridership. The American Public Transportation Association (APTA) reports that 4% of bus riders nationally are children under the age of 14 [20]. Using this as a coarse estimate for the percentage of exempt bus riders on the MBTA, the non-interaction rate may be 4 percentage points lower than the estimate from APC and AFC data alone. Accounting for this would provide a lower estimate of lost revenues, but it is not specific to location, time, or even the MBTA.

Without specific fare inspection data to compare against AFC records, these analyses provide a reasonable range of estimates of lost fares per non-interacting passenger.

2.7 Modeling Non-Interactions and Lost Revenue

The processes described in the preceding sections are used to estimate fare system non-interactions and corresponding lost revenues from the data collected automatically by AFC and APC systems. These observations can be aggregated to different spatial and temporal scales depending on the question of interest. For example, aggregating all observations across all stop events provides systemwide measures of non-interactions and lost revenue. Data can also be aggregated at the level of individual stop locations, routes, or neighborhoods. To gain insights about the factors that determine non-interactions and revenue losses, or to make predictions of these values, it is useful to create models that relate characteristics of the system with estimated values of non-interactions and lost revenues.

2.7.1 Data Aggregation

In this study, the goal for modeling fare non-interactions and lost revenues is to understand how these values vary across different locations in the city by time of day and day of the week. Understanding these variations provides insight about where and when lost revenues are most frequently occurring, which can be used to prioritize resources for collecting additional manual observations, implementing targeted enforcement, or planning other interventions to improve fare payment compliance.

At a spatial scale, individual transit stops often have too few passengers to make meaningful estimates of fare non-interactions or characterize fare transaction data. In this study, data is aggregated to a grid of square zones, each covering 800,000 square meters. Census tract zoning was considered but not ultimately used, because many transit stops are located along major streets that often serve as boundaries for these tracts. This would concentrate much of the data on the boundaries and split observations for opposing directions on the same route. Furthermore, many census tracts are small to adequately aggregate observations for analysis. The chosen grid system offered a more appropriate structure for analyzing patterns of fare non-interactions and the estimated lost revenues.

To account for variation by time of day, data is aggregated into 5 time periods that characterize different prevailing travel patterns and passenger demographics. These time

periods are an aggregation of the weekday time periods defined in the MBTA Service Delivery Policy [26]:

1) **AM Peak (5:00 a.m.–9:00 a.m.)**: The morning peak is dominated by commuters travelling to work, many of whom are passholders. The system is busy, with the highest rates of passenger boardings during this time.

2) **Midday (9:00 a.m.–1:30pm)**: The middle period of the day served lower demand that is more varied in composition.

3) **Midday School (1:30 p.m.–4:00pm)**: Demands increase in the early afternoon with a significant number of school-aged students using the system after schools release for the day.

4) **PM Peak (4:00 p.m.–6:30pm)**: The evening peak is similar to the morning in that there are a significant number of commuters using the service, but these users are mixed with many other trip purposes. The system is busy and the passenger composition is varied.

5) **Evening (6:30 p.m.–11:59pm)**: The evening hours are characterized by diminishing demand.

2.7.2 Ordinary Least Squares Regression Model

The Ordinary Least Squares (OLS) regression model was applied to predict revenue loss per hour for each zone, broken down by day and time period. OLS is a widely used technique in linear regression that examines the relationship between the outcome and one or more input variables. It works by minimizing the difference between the actual and predicted values. The approach assumes a linear relationship between variables like weekday, time category, and APC_count and the target variable. Coefficients for each explanatory variable are calculated to create a formula that estimates revenue loss based on these inputs.

Each predictor in the OLS model is assigned a coefficient that reflects both the strength and direction of its relationship with the dependent variable. A positive coefficient means an increase in that feature is associated with a higher revenue loss, while a negative coefficient indicates the opposite. The OLS method focuses on minimizing the sum of squared errors (SSE), ensuring the model captures the best linear fit for the data. The simplicity and interpretability of OLS make it an ideal baseline for understanding which factors most strongly influence revenue loss, and the model's performance is assessed using metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to quantify prediction accuracy.

2.7.3 Neural Network Machine Learning Model

A neural network was used to predict revenue loss in each zone per day per hour. Neural networks are effective models for capturing patterns and complex relationships within data. This model was selected because it is capable of handling both simple and non-linear relationships between the input features and the target variable [21].

2.7.3.1 Model Structure

- **Input Layer:** The input layer includes the variables used for making predictions, such as the categorical data like weekday and time category, as well as the numerical variables such as APC_count. To allow the model to understand and use the categorical data, we transformed them into numerical format using One-Hot Encoding, which helps the neural network process this information more efficiently [22].

- **Hidden Layers:** The model consists of several hidden layers, where input data is processed through transformations. Each hidden layer uses activation functions like ReLU (Rectified Linear Unit) to introduce non-linearity, helping the model detect complex patterns. ReLU is a popular choice because it avoids issues like the vanishing gradient problem. We tested both a two-layer and a five-layer model, but since there was no significant difference in performance, we chose to proceed with the simpler two-layer model to reduce complexity in the model.

- **Output Layer:** The output layer of the neural network contains a single neuron. This neuron provides the final prediction of revenue loss per hour per day in different zones, which is a continuous value, making it suitable for this regression task. The activation function in this layer is linear, as no further transformation is needed to output a continuous number.

- **Neurons in Hidden Layers:** Neurons are the building blocks of a neural network. In each hidden layer, they take input from the layer before, combine these inputs by applying weights, and then pass the result through an activation function, like ReLU. The number of neurons plays a crucial role in how well the model can identify patterns in the data. More neurons can help the model pick up on more complex relationships, but having too many can also make the model overly complicated and prone to overfitting.

- **Neurons in Output Layer:** The output layer contains a single neuron, as we are predicting a single continuous value—revenue loss per hour per day per zone. This neuron outputs the final prediction after processing the input through the hidden layers.

- **Optimization and Loss Function:** For training, we used the Adam optimizer, which adapts the learning rate throughout the process to help the model learn effectively. The model's accuracy was assessed using the Mean Squared Error (MSE) loss function, which calculates how far the predictions are from the true values. MSE

places a heavier penalty on larger errors, pushing the model to reduce significant prediction mistakes [23].

2.7.3.2 Permutation Feature Importance

Breiman [24] proposed permutation feature importance within the framework of random forests. This approach evaluates the influence of individual features on a model's predictions by analyzing their impact on the model's performance. A feature's importance is determined by measuring how much the prediction error increases after its values are shuffled. If shuffling results in a noticeable increase in error, the model depends on that feature, making it important. Conversely, if the error remains largely unchanged, the feature has little impact on the model's predictions.

Breiman's method involves breaking the link between a specific feature and the target variable to see how much the model's accuracy is affected. By shuffling the values of a feature, we essentially remove its influence on the model's predictions while leaving the other features unchanged. The drop in performance that follows tells us how important that feature was to the model's ability to make accurate predictions. This method was popularized in machine learning through its inclusion in the random forest framework, but it has since become a standard technique in model interpretation [24,25].

This page left blank intentionally.

3.0 Results

The results of this research are presented in three main parts. First, a summary of findings from the previous studies that involved manual observations of fare non-interactions are described. Second, the results of the analysis of non-interactions and estimated lost revenues from the automatically collected data are presented. Third, a comparison of regression and machine learning models provides some insights about the factors that affect these values.

3.1 Insights from Manual Observations

This section includes three parts. First, a summary of the fare compliance data and patterns from the Green Line rear door boarding study [9]. Second, a summary of more systemwide fare compliance data compiled by the Central Transportation Planning Staff (CTPS) [14]. Finally, a brief discussion is provided for how the limitations and insights from these studies guided the study of automatically collected data.

3.1.1 Observations of Rear-Door Boarding on Green Line (Light Rail)

A detailed study of fare payments on the Green Line was conducted in 2016, with findings were presented in the 2017 Annual Meeting of the Transportation Research Board [9]. The study focused on the behavior of passengers who board through the rear doors on light rail vehicles, with the goal of estimating lost revenues from passengers who do not pay the correct fare.

Detailed boarding and fare inspection data was collected for 110 single-direction trips on the Green Line surface branches in April and May 2016. Data was collected by research staff working alongside fare inspectors in two types of configurations:

- On *fare-inspection* cars, the team of researchers and fare-inspectors recorded the number of passengers boarding through rear doors, inspection counts of passes validated with a portable device, noting the number of passengers who pre-paid, held passes, did not hold passes, or refused to comply with inspection.

- On *normal-entry* cars, the team consisted of only researchers who collected observations and documented the number of passengers that boarded through rear doors and the number of those passengers who proceeded to the farebox after boarding.

A limitation of manually counted fare compliance data is that the process is complicated by the multiple methods of fare payment, not all of which can be confirmed in real-time by an inspector. For passengers that board through a rear-door, there are several possible cases:

1) The passenger may proceed once onboard to pay at the farebox, which is assumed to be a compliant fare.

2) The passenger may hold a valid time-limited pass, which can be verified by the fare inspector visually or with a handheld device.

3) The passenger may hold a stored value smart card (Charlie Card), which can be verified with a handheld device.

4) The passenger may hold a stored value magstripe card (Charlie Ticket), which cannot be verified in the field, because the time of a fare transaction is recorded by the farebox.

5) The passenger may have paid cash, in which there is no verifiable record of the transaction.

The 1,577 rear-door boarding passengers who were inspected through this study were categorized as follows:

- 69% held valid time-limited passes
- 7.5% had pre-paid their fares
- 1.3% refused to respond to fare inspectors, and appear to be intentional fare evaders
- The remaining roughly 22% of the rear-boarding passengers had no verifiable record of payment, and therefore represent potential fare evaders.

A useful aspect of the Green Line study was that data was collected across multiple branches of the line and times of day, so that the variation in observed fare payment compliance could be assessed. Prokosch and Gartsman [9] present the following observations regarding the relative prevalence of fare non-compliance on the Green Line:

1) Lost revenue is concentrated in the AM Peak in the inbound direction. This is consistent with the fact that the largest numbers of passengers boarding at surface stops are during the AM Peak for trains headed toward downtown Boston. In the PM Peak, most passengers are boarding trains from gate-controlled underground stations in the city center.

2) Total lost revenue in the mid-afternoon and PM Peak is similar to the AM Peak. Although the evasion rate is lower in afternoon and evening hours, the high volume of travel is spread across more hours. The net effect is that the estimated revenue losses are similar to the AM Peak hours.

3) The fare evasion rate per passenger is roughly double per PM rear boarding compared to the AM. Although there are many more rear door boardings in the AM Peak, the passengers are mostly commuters who hold passes.

4) There is high correlation between crowding, both in terms of passengers boarding and the number of passenger onboard vehicles, and fare evasion, with the data showing an exponentially increasing relationship. In more crowded vehicles, it is physically more difficult to move toward the farebox to pay once boarded.

5) The behaviors of Green Line operators have some effect on the numbers of passengers boarding through rear doors and on the rates of fare evasion. Since operators juggle two objectives: maximize fare revenues collected and maintain schedule adherence, the behaviors of drivers in terms of which doors they open, when they open doors, and what announcements they make all vary by time of day and level of crowding.

3.1.2 Systemwide Analysis of Fare Compliance from CTPS Study

A systemwide report on fare compliance was conducted by CTPS in 2016 [14]. The CTPS study focused largely on fraudulent interactions with faregates, which are relevant only to heavy rail services and the parts of the Green Line that operate in faregate-controlled stations in downtown Boston. The report includes summaries of findings from three relevant data sources: National Transit Database (NTD) non-interaction survey, fare enforcement citations, and short fares recorded in the automatic fare collection (AFC) database.

The NTD non-interaction survey consists of one-hour spot checks by tabulators who observe how passengers appear to be paying fares. The data collection is reported to be sparse but to provide a global overview of how fares are being paid. Table 3.1 shows the measured rate of fare non-payments across the bus/trolleybus garages and light rail lines based on the NTD non-interaction surveys.

The reported findings from the NTD non-interaction survey include:

1) The percentage of passengers that do not interact with a farebox is reported to be 0.9 percent ± 0.1 percent overall.

2) Passengers that pay a short fare are observed as in the NTD non-interaction data as paying cash.

3) Operators were not observed to use the short fare button on the farebox to count fare evaders, so short fare data from the AFC system does not include passengers that fail to interact with the farebox in any way.

Fare enforcement citation data are reported for state fiscal years 2012, 2013, and 2014. The vast majority of the citations are issued at gate-controlled stations. Across the years reported, there were an average of 4,492 citations issued annually systemwide. Of these, an average of 137 citations per year were issued on the four surface branches of the Green Line, 1 citation per year on the Mattapan line, and 4.6 citations per year were issued on buses. The very low number of citations on buses is an indication that transit police are only called to buses for serious incidents, and fare compliance on buses is not otherwise systematically enforced.

Table 3.1 Fare Nonpayment Rates NTD Survey, 2010-2015 [14]

Garage or Line	Did Not Pay (%)	Rear Door Boarding (%)
All Bus Garages	0.9 ± 0.1	—
Lynn	1.2 ± 0.4	—
Southampton	1.2 ± 0.2	—
Cabot	1.0 ± 0.2	—
Arborway	0.9 ± 0.3	—
Fellsway	0.8 ± 0.2	—
Somerville	0.6 ± 0.4	—
Charlestown	0.4 ± 0.2	—
Albany	0.2 ± 0.2	—
Quincy	0.2 ± 0.2	—
North Cambridge	0.6 ± 0.1	—
Surface Green Line	1.1 ± 0.1	9.3 ± 0.3
Branch C	1.3 ± 0.2	7.2 ± 0.5
Branch E	1.2 ± 0.3	8.2 ± 0.7
Branch D	1.1 ± 0.2	12.4 ± 0.5
Branch B	0.9 ± 0.1	7.6 ± 0.4
Mattapan Line	12.5 ± 0.5	—

A more comprehensive view of fare compliance based on short fares is presented in the CTPS study. Although not based on direct manual surveys, the short fares are recorded in the AFC system when the operator pushes the short fare button on the farebox to indicate that the incorrect fare was paid. This can occur if a passenger does not insert enough cash or does not have sufficient value on a stored value card. As noted above, operators were not observed to use the short fare button to count fare evaders who did not interact with the farebox at all. Short fares indicate a revenue loss but are not necessarily a measure of intentional fare evasion. Some of the insights and findings from the short fare data include:

1) Short fares comprised roughly 2% of all farebox transactions in state fiscal year 2014.

2) Roughly 30-40% of short fares are associated with $0 paid.

3) The percentage of transactions categorized as short fares varies by route and is tabulated by garage. The rate of short fares by garage is correlated with demographics. Higher percent of riders living in households earning less than $30,000 has the strongest with more short fares. Higher percentage of riders who identify themselves as a minority is also correlated with more short fares.

4) The rate of short fares is greatest in midday and evening hours. It is lowest during commuting hours.

5) The rate of short fares is greatest in summer months (June, July, August), and higher on weekends than on weekdays.

Of the 15 routes with the highest rates of short fares, 11 are associated with the Lynn garage. The highest numbers of short fares are associated with heavily traveled routes, including many Key Bus Routes.

3.1.3 Insights for Analysis of Automatically Collected Data

Based on the insights summarized above, the analysis focuses on patterns in fare payment and non-interaction in the following ways:

1) Distinguish between passengers who do not interact with the AFC at all, some of whom are evading fares but some of whom may hold passes or be eligible for fare exemption (e.g., children), and passengers who are counted as short fares, who are also associated with a revenue loss.

2) Evaluate variability by:

    a. Time of Day

    b. Day of Week

    c. Location (stop location, route)

    d. Number of passengers boarding

3.2 Fare System Non-Interactions

Data for 4 weekdays were picked over a span of 4 weeks (one weekday from each week) to ensure good coverage throughout the week. The selected dates are Wednesday, September 18; Thursday, September 26; Friday, October 4; and Monday, October 7, 2019. This approach provided a balanced view of transit operations throughout the selected period.

This study comprehensively analyzes records related to 322,505 stop events made by 557 vehicles at 7,014 bus stops throughout the Greater Boston area serviced by the MBTA. These stop events include detailed information about passenger boardings and alightings, fare transactions, and vehicle movements. This dataset provides valuable insights into the patterns of transit usage, fare compliance, and operational efficiency. The map in Figure 3.1 illustrates the geographical distribution of the bus stops covered in this study, highlighting the extensive reach of the MBTA's bus network within the Greater Boston area (an interactive version of the map is available online at https://umass-amherst.maps.arcgis.com/apps/instant/basic/index.html?appid=307b744e12af431a843798146bc40ca7). Bus stops within 100 meters of a rapid transit station have a high potential for use by transfer riders and are shown in red.

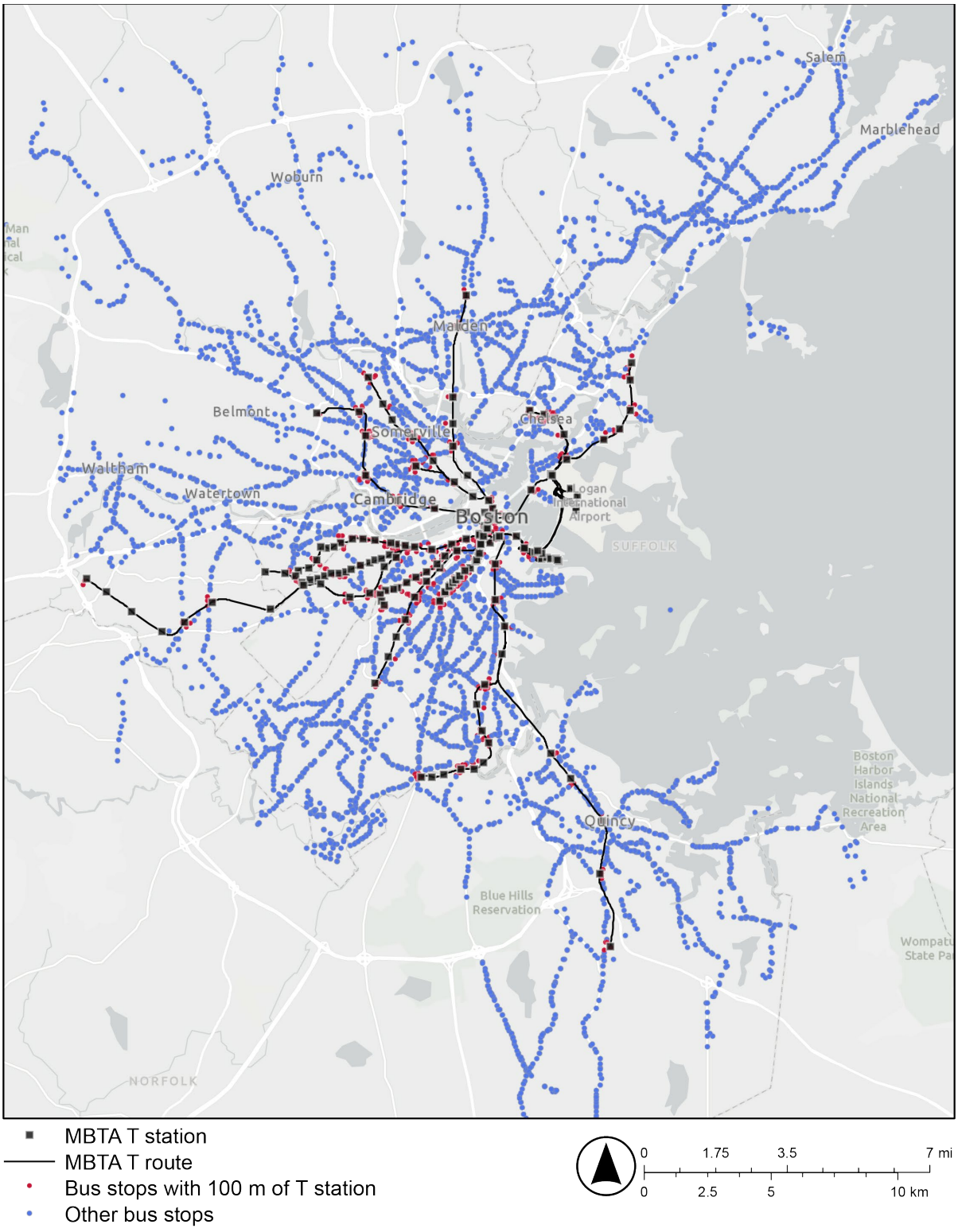

Figure 3.1 MBTA Bus Stop Locations

3.2.1 Identification of Outliers

The sample of data that is used for the following section consists of records of all bus stop events from four days: 2019; Wednesday, September 18, 2019; Thursday, September 26, 2019; Friday, October 4, 2019; and Monday, October 7, 2019. The first column of Table 3.2 shows the how many of the 554 vehicles in the data set are identified as outliers using the processes described above. The second and third columns of Table 3.2 show the corresponding numbers of stop events and estimated passenger boardings associated with these vehicles. In total, 17% of vehicles are identified as outliers.

Table 3.2. Share of Data Identified as Outliers

Outliers	Vehicles	Stop Events	AFC Transactions
Total Count	554	275,940	874,497
Outliers, APC Errors	26	7,408	21,452
Outliers, Non-Interaction Rate	59	11,308	17,001
QC_Counts > 20	11	9,112	23,416
All Outliers	96	27,828	61,869
All Outliers as % of Total	17.3%	10.1%	7.1%

3.2.2 Non-Interactions by Bus Stop

Figure 3.2 shows a map of the fare non-interaction counts and rates by bus stop location (an interactive version of the map is available online at https://umass-amherst.maps.arcgis.com/apps/instant/basic/index.html?appid=307b744e12af431a843798146bc40ca7). The size of each circle represents the count of non-interactions and color represents with non-interaction rate, with darker blue indicating locations with higher rates. The non-interaction counts are highly correlated with ridership, with the largest circles appearing at rapid transit and Silver Line stations with large numbers of passengers transferring between rail and bus or between buses. The non-interaction rate, however, is much more varied and shows no clear spatial pattern. Most of the stops with high non-interaction rates served few passengers, so a single non-interaction represents a relatively high percentage of a small number of total boarding passengers.

Another way to look at this non-interaction data is in a ranked list of the bus stops with the highest non-interaction counts. Table 3.3 shows the top 20 bus stops overall. It is clear from the bus stop names that these stops are almost located at locations with high numbers of transferring passengers. Nubian (formerly Dudley) tops the list, followed by rapid transit stations throughout the network. Table 3.4 shows the top 20 bus stops after those within 100 meters of a rapid transit station have been removed. These are stops that are mostly located in the southern neighborhoods of Boston where bus ridership is high.

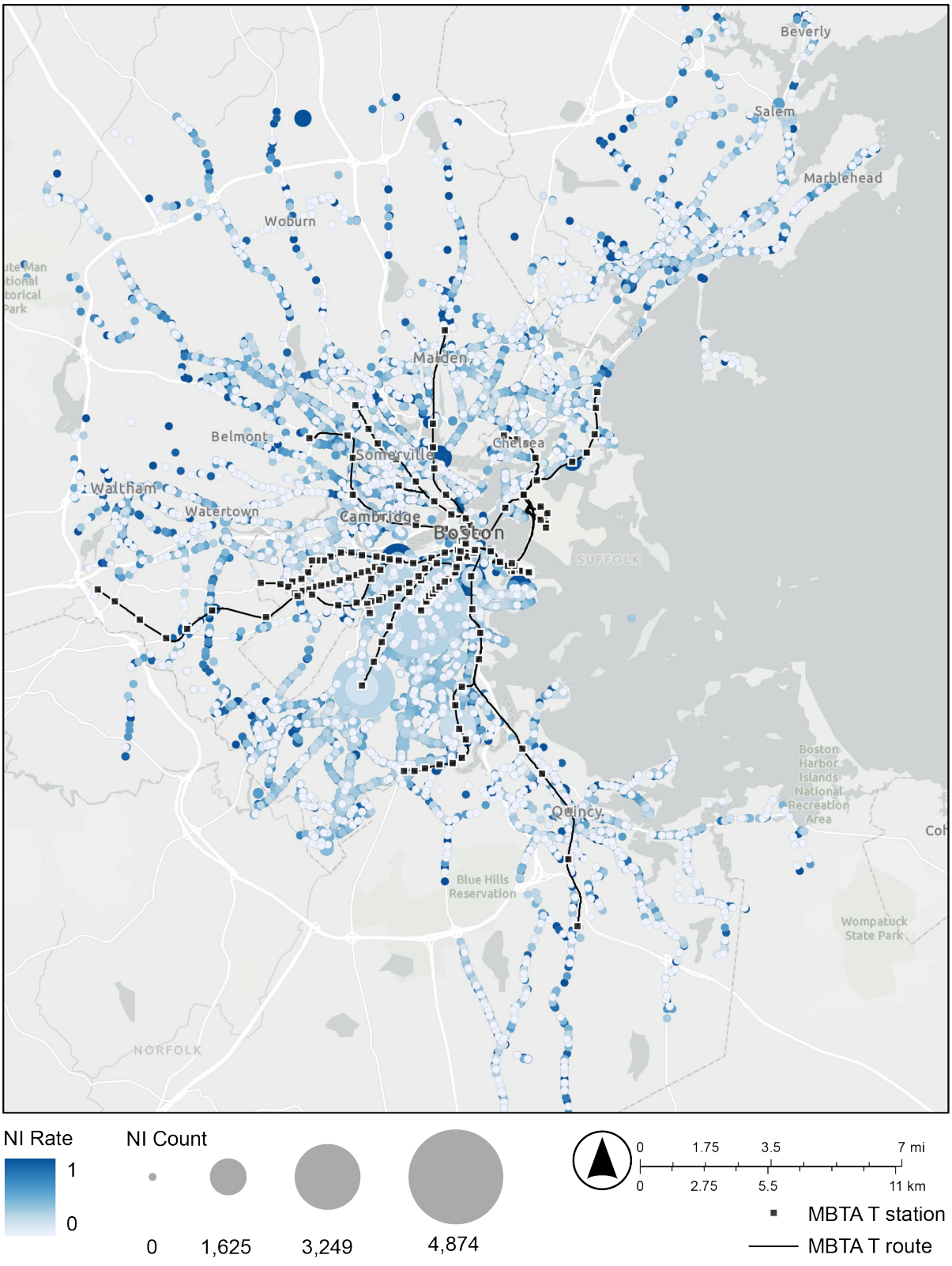

Figure 3.2 Non-interaction Counts and Rates by Stop Location

Table 3.3 Top 20 Bus Stops by Non-Interaction Count

Rank	Stop Name	APC Count	NI Count	NI Rate
1	DUDLEY STATION	29,806	4,874	0.164
2	RUGGLES LOWER BUSWAY - LANE	18,965	3,199	0.169
3	FOREST HILLS STATION LOWER B	19,426	3,079	0.158
4	ASHMONT BUSWAY	13,564	1,605	0.118
5	FOREST HILLS STATION UPPER B	15,501	1,480	0.095
6	KENMORE STATION BUSWAY	1,469	1,451	0.988
7	BROADWAY STATION - RED LIN	1,569	1,346	0.858
8	MALCOLM X BLVD @ KING ST	4,667	1,069	0.229
9	JACKSON SQUARE BUSWAY	6,994	958	0.137
10	CITY POINT BUS TERMINAL	1,342	936	0.697
11	HARVARD SQ @ GARDEN ST - DAW	2,177	925	0.425
12	ANDREW STATION BUSWAY	6,735	888	0.132
13	250 DORCHESTER AVENUE	895	873	0.975
14	MASSACHUSETTS AVE @ ALBANY S	4,927	852	0.173
15	AVE LOUIS PASTEUR @ LONGWOOD	2,039	823	0.404
16	MATTAPAN SOUTH BUSWAY	3,265	803	0.246
17	HAYMARKET BUSWAY	5,058	781	0.154
18	SULLIVAN STATION BUSWAY - BE	6,597	764	0.116
19	1624 BLUE HILL AVE @ MATTAPA	3,509	701	0.200
20	W BROADWAY @ BROADWAY STATIO	4,761	686	0.144

Table 3.4 Top 20 Bus Stops over 100 m from Rapid Transit by Non-Interaction Count

Rank	Stop Name	APC Count	NI Count	NI Rate
1	CITY POINT BUS TERMINAL	1,342	936	0.697
2	250 DORCHESTER AVENUE	895	873	0.975
3	AVE LOUIS PASTEUR @ LONGWOOD	2,039	823	0.404
4	AVE LOUIS PASTEUR @ THE FENW	1,270	623	0.491
5	HUMBOLDT AVE @ TOWNSEND ST	1,101	606	0.550
6	WARREN ST @ TOWNSEND ST	1,968	564	0.287
7	BLUE HILL AVE @ ELLINGTON ST	2,217	530	0.239
8	COLUMBUS AVE @ WALNUT AVE	1,537	507	0.330
9	BLUE HILL AVE @ MORTON ST	2,110	471	0.223
10	MASSACHUSETTS AVE @ HARRISON	3,301	468	0.142
11	HYDE PARK AVE @ OAK STREET	3,156	441	0.140
12	MALCOLM X BLVD @ SHAWMUT AVE	2,130	441	0.207
13	MALCOLM X BLVD OPP O BRYANT	1,326	440	0.332
14	MALCOLM X BLVD @ O BRYANT HS	1,252	434	0.347
15	WASHINGTON ST @ COLUMBIA RD	1,916	404	0.211
16	COLUMBIA RD @ WASHINGTON ST	1,837	399	0.217
17	WASHINGTON STREET @ FOUR COR	1,490	394	0.264
18	WARREN ST @ QUINCY ST	1,994	390	0.196
19	MORTON ST @ BLUE HILL AVE	2,287	386	0.169
20	WARREN ST @ SUNDERLAND ST	1,717	381	0.222

3.2.3 Non-Interactions by Route

Perhaps a more practical way to use this information is to consider the non-interaction rates aggregated to bus routes, because manual counts or fare inspections are likely to happen on-board the vehicles. Figure 3.3 shows a map of the routes with the width of each line indicating the non-interaction count for the route and darker color indicating the non-interaction rate (an interactive version of the map is available online at https://umass-amherst.maps.arcgis.com/apps/instant/basic/index.html?appid=307b744e12af431a843798146bc40ca7). Similar to the bus stop data, the routes with the highest non-interaction count are those that run through the central and southern parts of the city. The high non-interaction rates appear on routes in outlying neighborhoods.

These patterns are confirmed by the ranking of top 20 bus routes by non-interaction count and non-interaction rate shown in Table 3.5 and Table 3.6. Non-interaction counts are strongly correlated with ridership, but non-interaction rate does not show the same pattern, and some of the higher non-interaction rates are associated with routes that have relatively low ridership in total.

3.2.4 Non-Interactions by Time

The fare system non-interaction rate changes over time. An initial analysis of non-interactions by day of the week shows that each weekday is roughly similar. Table 3.7 shows the results of the analysis over four weekdays to reveal that out of approximately 787,000 passengers that boarded buses, nearly 173,000 passengers did not interact with the fare system. This indicates a fare non-interaction rate of 22%. The non-interaction rate ranged from 21% to 23% on different weekdays, showing a consistent pattern across the days.

To better understand how fare non-interactions and passenger boardings change throughout the day, the day has been broken into key time periods based on the time periods defined in the MBTA Service Delivery Policy [26]. Table 3.8 shows the values of APC count, NI count, and NI rate summarized by the timer period. More detailed time series of APC count in Figure 3.4 and NI rate are presented in Figure 3.5.

- **AM Peak (5:00 to 9:00):** During the AM peak period, passenger boardings steadily increase, with a significant spike between 8:00 and 9:00, showing the morning rush hour when people commute to work or school. The fare non-interaction rate also remains relatively stable in the early hours but shows a noticeable increase during the peak boarding times. This correlation indicates that higher passenger volumes during the morning peak are associated with increased fare non-interactions.

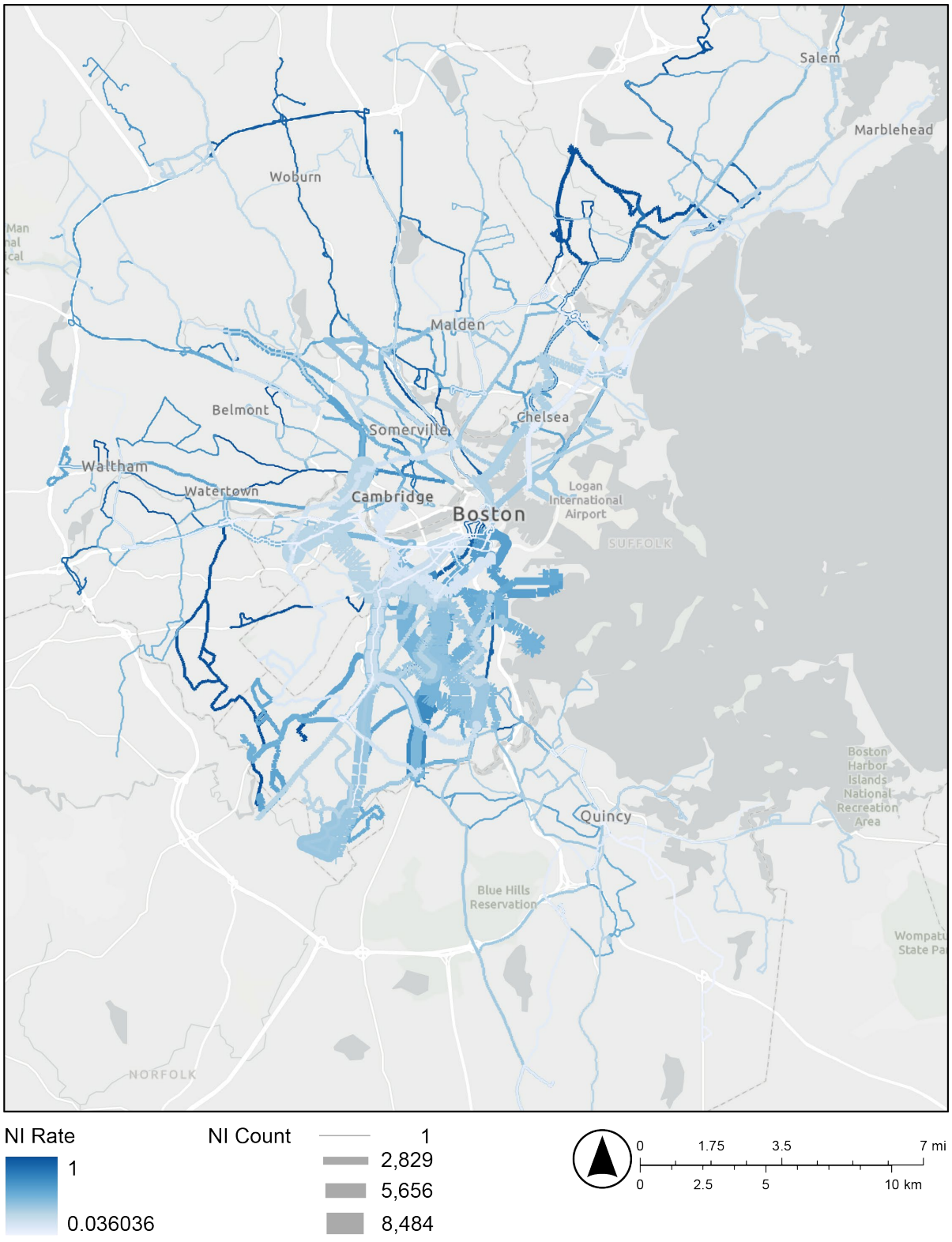

Figure 3.3 Non-interaction Counts and Rates by Route

Table 3.5 Top 20 Bus Routes by Non-Interaction Count

Rank	Route	APC Count	NI Count	NI Rate
1	66	45,674	8,484	0.186
2	1	46,023	7,998	0.174
3	23	37,989	7,898	0.208
4	28	26,997	7,014	0.260
5	22	31,448	6,955	0.221
6	32	27,008	5,478	0.203
7	16	21,747	4,858	0.223
8	15	22,734	4,563	0.201
9	39	25,656	4,429	0.173
10	9	23,187	3,944	0.170
11	31	16,354	3,573	0.218
12	111	16,783	3,197	0.190
13	11	13,091	3,153	0.241
14	44	13,184	3,080	0.234
15	47	19,230	3,072	0.160
16	19	13,353	2,846	0.213
17	45	11,980	2,837	0.237
18	8	12,729	2,792	0.219
19	7	13,223	2,731	0.207
20	10	11,158	2,581	0.231

Table 3.6 Top 20 Bus Routes by Non-Interaction Rate

Rank	Route	APC Count	NI Count	NI Rate
1	429	2110	960	0.455
2	71	350	152	0.434
3	90	171	71	0.415
4	73	241	95	0.394
5	38	2815	1067	0.379
6	78	105	36	0.343
7	556	284	96	0.338
8	436	891	300	0.337
9	95	701	224	0.320
10	60	746	233	0.312
11	52	2233	697	0.312
12	430	335	103	0.307
13	201	526	161	0.306
14	352	457	137	0.300
15	43	3905	1155	0.296
16	18	1936	571	0.295
17	96	217	64	0.295
18	92	620	182	0.294
19	134	606	177	0.292
20	558	383	111	0.290

Table 3.7 Non-interaction and Automatic Passenger Count Data by Weekday

Day of Week	APC count, Ω	NI count, C	NI rate, R
Monday	198,654	45,661	0.23
Wednesday	201,970	42,524	0.21
Thursday	198,586	41,731	0.21
Friday	187,795	42,840	0.23
All Days	787,005	172,756	0.22

Table 3.8 Non-Interaction and Automatic Passenger Count Data by Time Period

Time Period	APC count/hr, Ω	NI count/hr, C	NI rate, R
AM Peak	12,910	2,398	0.19
Midday	8,611	1,861	0.22
Midday School	13,041	2,920	0.22
PM Peak	14,627	3,194	0.22
Evening	5,588	1,588	0.28
All Time Periods	10,090	2,215	0.22

- **Midday (9:00 to 13:30):** The midday period sees a noticeable drop in passenger boardings, which then remains steady. This suggests that fewer people are traveling during these hours, likely because of work or school commitments. Meanwhile, the fare non-interaction rate shows some fluctuations but generally stays around the same level as the morning peak. This consistency implies that the pattern of fare non-interactions doesn't change much during midday hours.

- **Midday School (13:30 to 16:00):** During the midday-school period, the APC count graph shows a slight increase in boardings, which probably relates to school dismissals and early afternoon activities. The fare non-interaction rate graph continues to fluctuate slightly but shows a gradual increase towards the later part of the afternoon. This suggests that fare non-interactions might be influenced by the increased travel of school-related passengers who might be more likely to interact with fareboxes.

- **PM Peak (16:00 to 18:30):** The late afternoon and evening period show a significant rise in passenger boardings, peaking around 5:00 PM. This matches the evening commute as people return home. The fare non-interaction rate, having heald steady through the afternoon begins to increase as the PM peak subsides. As the evening rush hour ends, passengers in the later evening hours are more likely to not interact with the fare system even as total ridership falls.

- **Evening (18:30 to 23:59):** In the evening period, the APC count graph shows a steady decline in boardings as passenger activity decreases towards the end of the

day. In contrast, the fare non-interaction rate keeps rising, reaching its highest observed levels at the end of the evening. This suggests that although fewer passengers board buses in the evening, a higher proportion of them are not interacting with fare boxes.

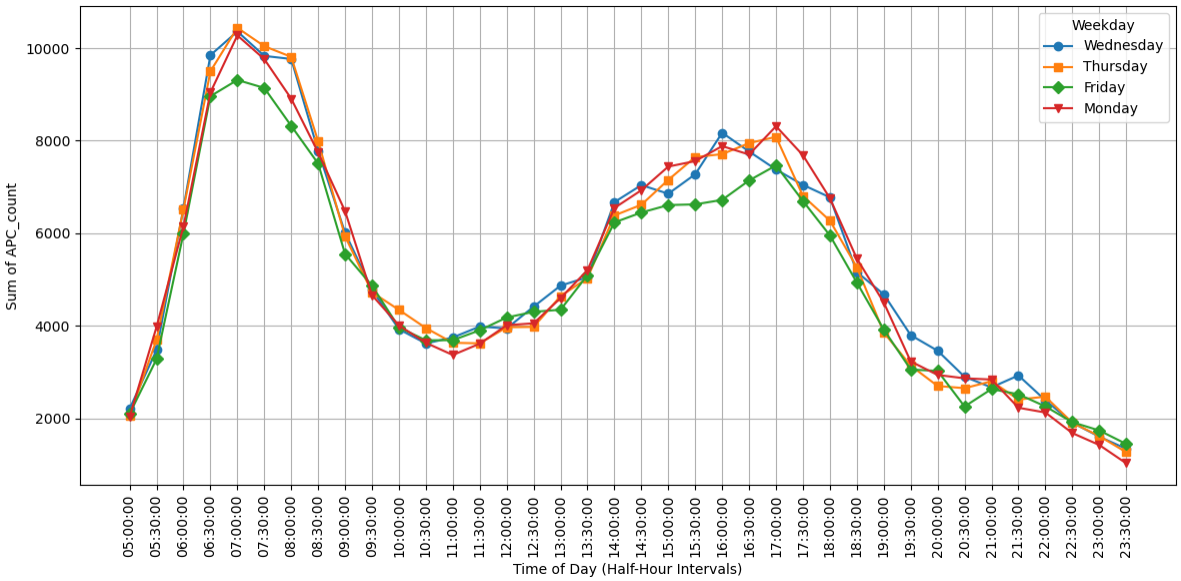

Figure 3.4 Time series of total passengers observed boarding in system (APC_count)

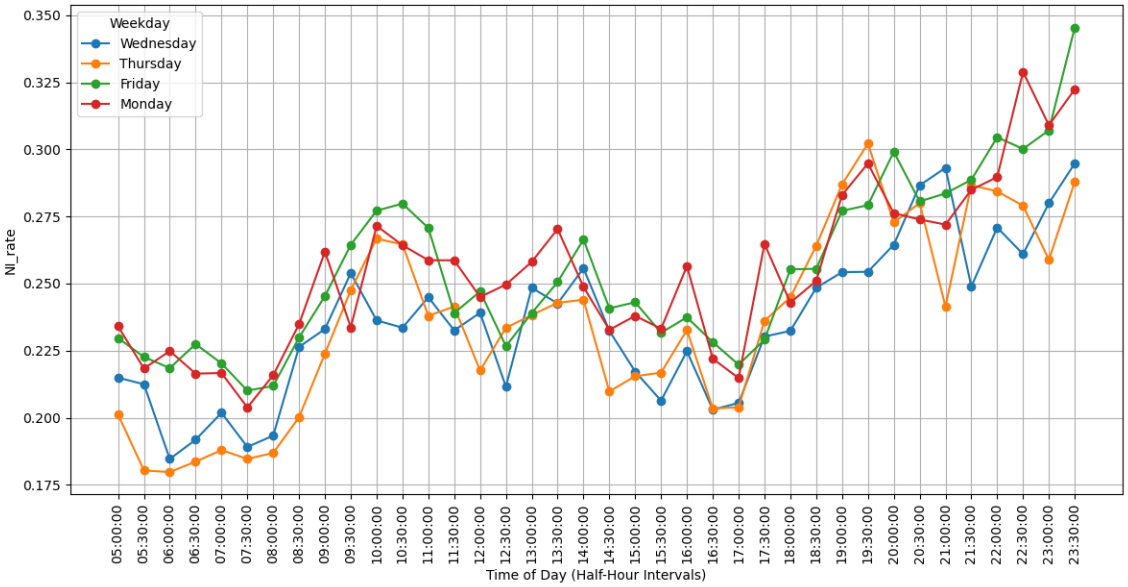

Figure 3.5 Time series of system-wide non-interaction rate (NI_rate)

The difference between the patterns of non-interaction rate and the total bus ridership is significant because it means that the fare non-interaction rate is not a static feature of the bus system. People in the afternoon and evening are more likely to not interact with the fare

payment system, and the non-interaction rate shows an increasing trend until the end of the day. This pattern is consistent with findings from a previous study of the Green Line [9].

3.2.5 Non-Interactions by Location and Time

The data on non-interactions can be combined and analyzed by both time and location to see how patterns vary by location and how these spatial differences change over time. Although it is possible conduct this analysis at varying spatial resolutions, the goal of this part of the study is to break up the Boston Metropolitan Region into zones, within which stop event data is aggregated to by time of day to count non-interactions, estimate a zone- and time-specific counts of non-interactions per hour and non-interaction rates per boarding passenger.

For this purpose, a grid system with square zones, each covering 800,000 square meters (approximately 0.3 square miles), was used for zoning. Census tracts were considered but not selected because many bus stops are located along major streets that frequently serve as boundaries for these tracts. This could have caused much of the data to be concentrated on the boundaries, which would affect capturing geographical insights. Furthermore, the census tracts were deemed too small to adequately capture geographical patterns. The chosen grid system offered a more appropriate structure for analyzing fare compliance patterns. The stop event data was spatially aggregated to the grid and temporally aggregated by day of week and time period of the day. Therefore, each square is associated with 20 aggregated data points for passenger boarding count, non-interaction count, estimated revenue loss per passenger, and the resulting estimated revenue loss per hour.

Appendix B includes maps of the non-interaction count per hour in each time period (interactive versions of the maps are available online at https://arcg.is/0XqODi0). The count of non-interactions is highly correlated with bus ridership, so the locations with the highest counts of non-interactions are the locations with the greatest numbers of bus passengers. In the morning peak, values are more distributed, in large part because passengers board buses in residential neighborhoods that are spread across the region. The non-interaction counts are more concentrated in the center of the region during the Midday School and PM Peak time periods, when more passengers are boarding vehicles near the central business district or at major transfer stations.

Appendix C includes maps of the non-interaction rate in each time period. In all time periods, there is no clear spatial pattern for the non-interaction rate. This means that the likelihood of a passenger not interacting with the fare system is random and not correlated with any specific parts of the city. The few very high values of non-interaction rates appear on the outlying parts of the city, where passenger ridership is low, so even a single non-interaction may make up a significant portion of the total observed passenger boarding.

The main take-away from these maps is that the patterns of passenger non-interactions are closely aligned with the locations and times that passengers are boarding vehicles. Although the aggregated data by time of day shows that non-interaction rates are affected by the time of day, the spatial variation appears to be random.

3.3 Lost Revenues from Fare Non-Interactions

As described in Section 2.6, the revenue losses associated with fare system non-interactions are not as simple as multiplying the fare by the count of non-interactions. Assumptions must be made about the composition of riders among the non-interacting population. A best-case estimate is that no revenues are lost if all passengers are exempt or hold valid passes. A worst-case scenario is that all non-interactions are associated with evasions of full fares. A middle case is to assume that the composition of non-interacting passengers matches those that are observed, then the average amount collected per fare transaction can be used to estimate the lost revenues due to fare non-interactions.

3.3.1 Observed Fare Payment Types

The observed fare payment types for the bus records used in this study are presented in Appendix A. These fare payment types are grouped by type to allow a comparison with the observations from the Green Line rear door boarding study [9]. The left column of Table 3.9 presents the aggregated fare payment types for passes, other pre-paid or exempt fares, and passengers whose fare transaction is associated with an amount paid. In the Green Line study passengers that boarded through the rear doors were inspected to identify the status of fare payment. The right column of the table presents the observed percentage of rear-boarding passengers that held valid passes, those with pre-paid fare, and the remaining percentage which are the potential fare evaders.

Table 3.9 Fare Payment Types in AFC.faretransactions and Green Line Study

Fare Payment Type	AFC Records (%)	Rear Boarding (%)
Passholder	65.9	69.0
Other Prepaid/Exempt Fare	7.6	7.5
Full of Discount Fare	26.5	23.5

It is not possible to know the true distribution of fare payment types among non-interacting bus passengers without collecting manual observations. However, the distribution of values in the AFC transaction records is similar to the observation of the rear-boarding passengers from the Green Line study. Therefore, the estimate from observed AFC transactions is at least a plausible estimate of the composition of non-interacting customers.

3.3.2 Lost Revenues by Bus Stop

The estimated lost revenues are associated with the fare payment types, so transfers are relevant. Figure 3.6 shows a map of stops with size for the number of transfers and color for the rate or ratio of recorded transfers to total AFC transactions (https://umass-amherst.maps.arcgis.com/apps/instant/basic/index.html?appid=307b744e12af431a843798146bc40ca7).

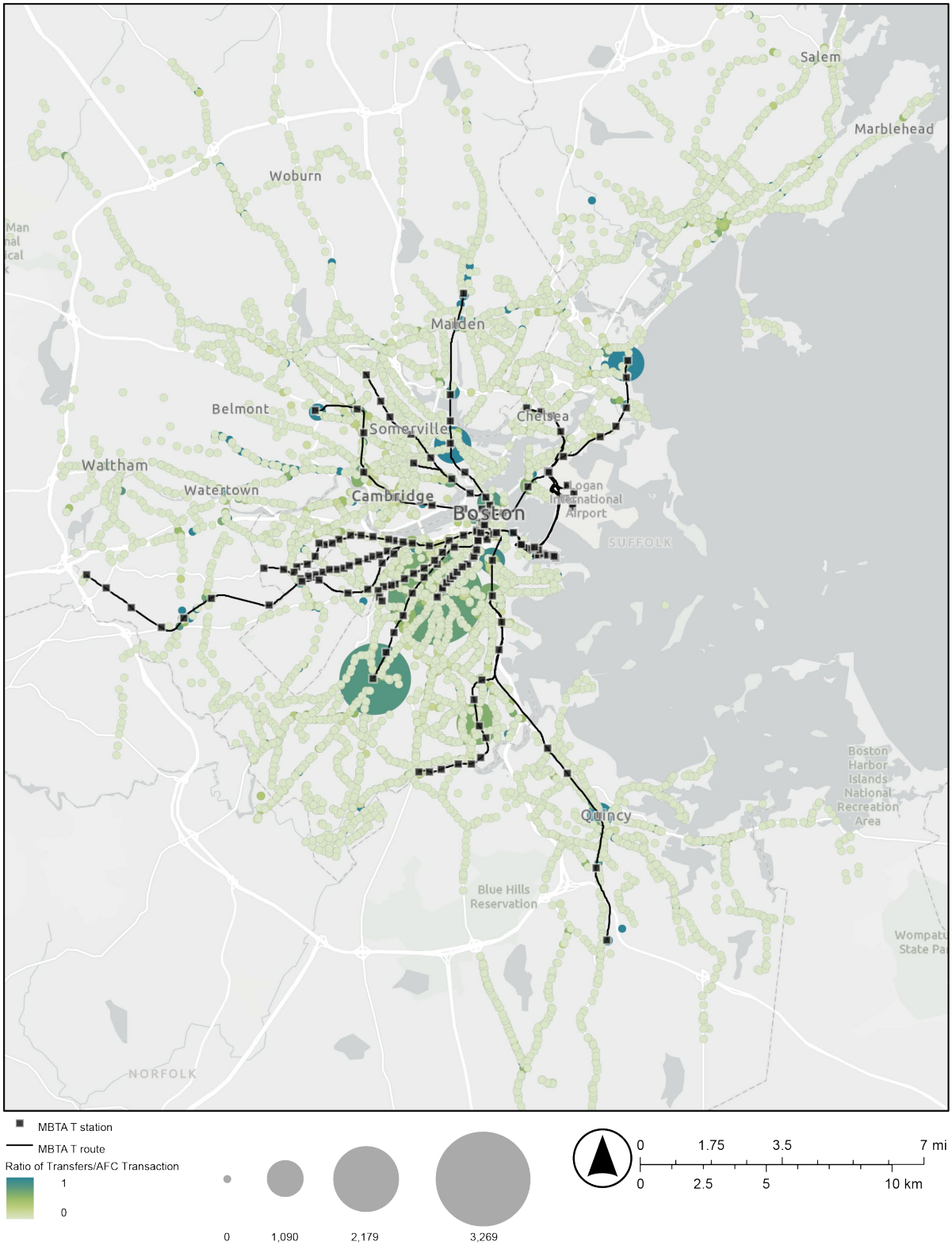

Figure 3.6 Transfer Count and Rate by Bus Stop

Clearly, most of the bus stops in MBTA system are associated with relatively low numbers of transfer and a low ratio of transferring passengers. A few stops at key locations where the rapid transit network intersects with the bus have very large numbers of transferring passengers and those represent a very high proportion of the fare system interactions. These stops are stations such as Nubian, Ruggles, Forest Hills, all of which topped the list of stops with the highest non-interaction counts.

Using the AFC fare payment type data, the estimated total lost revenues per hour and the lost revenues per non-interacting passenger are mapped in Figure 3.7 (https://umass-amherst.maps.arcgis.com/apps/instant/basic/index.html?appid=307b744e12af431a843798146bc40ca7). The map shows that largest lost revenues per hour (circle size) are at the same locations that have the highest non-interaction count in Figure 3.2. The top 20 bus stops by estimated lost revenues per hour are listed in Table 3.10.

Table 3.10 Top 20 Bus Stops by Estimated Lost Revenue per Hour

Rank	Stop Name	NI count/hr	$/NI	Lost $/hr
1	DUDLEY STATION	64.1	0.19	12.50
2	RUGGLES LOWER BUSWAY - LANE	42.1	0.16	6.64
3	KENMORE STATION BUSWAY	19.1	0.28	5.41
4	CITY POINT BUS TERMINAL	12.3	0.44	5.38
5	PARK DR @ FENWAY STA	3.2	1.70	5.37
6	MASSACHUSETTS AVE @ ALBANY S	11.2	0.41	4.61
7	BROADWAY STATION - RED LIN	17.7	0.24	4.31
8	FOREST HILLS STATION LOWER B	40.5	0.09	3.82
9	ASHMONT BUSWAY	21.1	0.16	3.35
10	HAYMARKET BUSWAY	10.3	0.31	3.18
11	HARVARD SQ @ GARDEN ST - DAW	12.2	0.24	2.98
12	MALCOLM X BLVD @ KING ST	14.1	0.21	2.91
13	SUMMER ST @ SOUTH STATION -	8.1	0.36	2.90
14	LYNN COMMUTER RAIL BUSWAY	4.2	0.67	2.83
15	1624 BLUE HILL AVE @ MATTAPA	9.2	0.31	2.82
16	SALEM COMMUTER RAIL STATION	3.1	0.91	2.80
17	HUMBOLDT AVE @ TOWNSEND ST	8.0	0.33	2.61
18	MASSACHUSETTS AVE @ HARRISON	6.2	0.40	2.49
19	AVE LOUIS PASTEUR @ THE FENW	8.2	0.29	2.34
20	MASSACHUSETTS AVE @ JOHNSTON	5.3	0.42	2.26

Where the trends differ is that estimated lost revenues per passenger exhibit a distinct spatial pattern, whereas the non-interaction rate appears more or less random. At the largest transfer points and throughout the central neighborhoods of Boston, the lost revenues per person are at or below the average. At transfer stops, the low estimate of lost revenue per non-interaction can be attributed to the large numbers of passengers transferring from rail or other bus lines. Operators may also be more likely to wave on passengers or allow back door boarding at these locations, because they know so many of the passengers do not need to pay an additional fare.

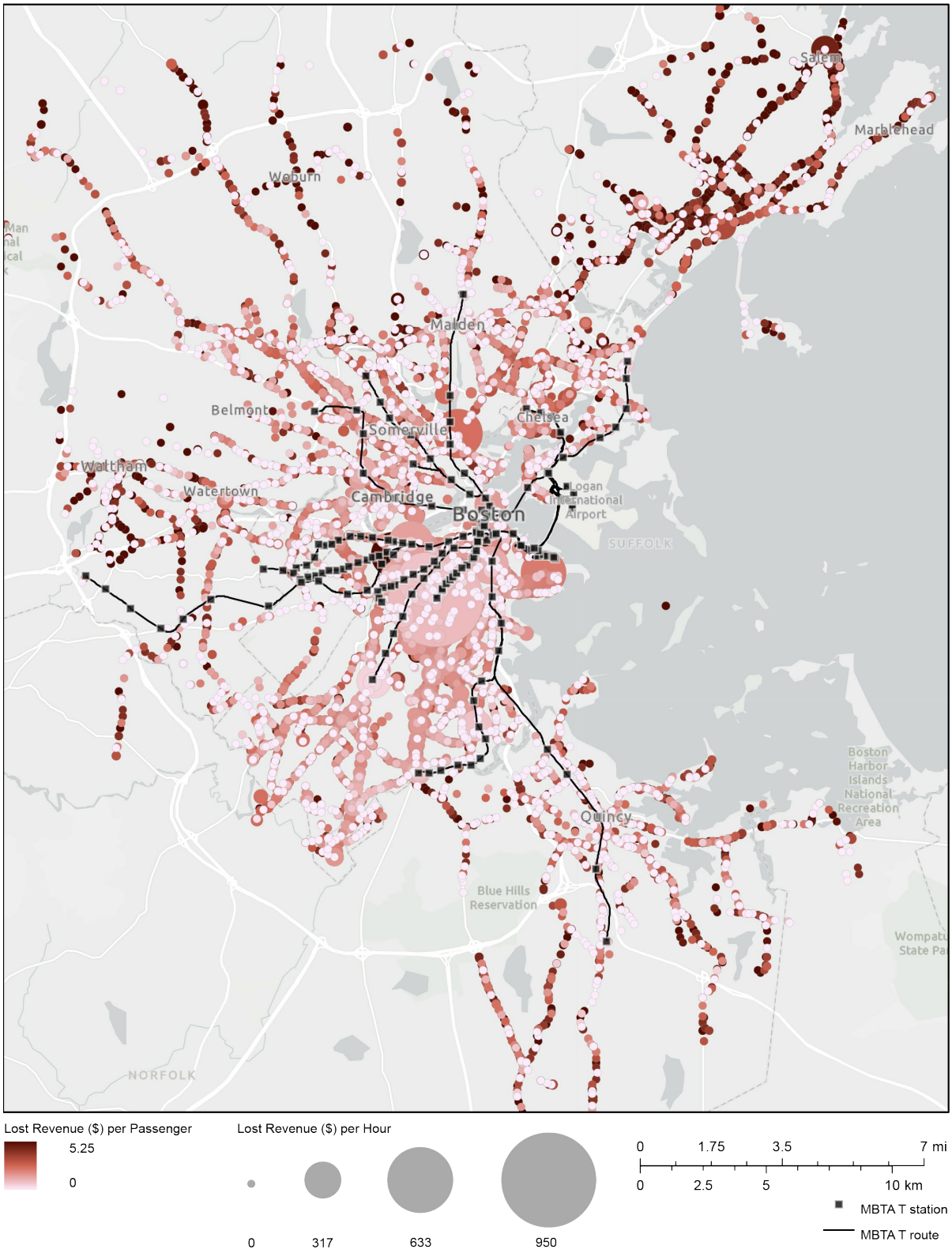

Figure 3.7 Lost Revenues per Hour and per Non-Interaction by Bus Stop

In the outlying communities, the estimated lost revenue per non-interaction is distinctly higher even though fewer passengers are not interacting. At the edges of the MBTA bus system, there are more routes that charge higher express fares and there are fewer points where passengers can transfer from one vehicle to another.

3.3.3 Lost Revenues by Route

It is also possible to look at the estimated revenue losses by route. These are mapped in Figure 3.8 with the line width indicating the total revenue losses per hour on the route and the color indicating the estimated lost revenue per non-interaction, which is the average fare payment associated with an AFC transaction (https://umass-amherst.maps.arcgis.com/apps/instant/basic/index.html?appid=307b744e12af431a843798146bc40ca7). Like the bus stop data, the estimated lost revenue per passenger is low on routes within the city center. However, estimated revenue losses per hour are higher because the ridership and associated count of non-interactions is highest in the center. The highest estimated revenue losses per passenger are on the express routes that serve the edges of the MBTA service area. Notably, there are several express routes on the north side of the region. These patterns are confirmed in the ranked lists of routes by estimated revenue loss per hour (Table 3.11) and estimated revenue loss per non-interaction (Table 3.12).

3.3.4 Lost Revenues by Time

Table 3.13 shows how the estimated lost revenues per non-interaction vary by time of day. The estimated lost revenue per non-interaction ($/NI) is simply the average amount collected per observed fare transaction in the corresponding time period. Multiplying the non-interaction count per hour by the corresponding lost revenue per non-interaction provides an estimate of the systemwide revenue losses per hour of the day.

The rate of passenger boardings is high during the AM Peak (5:00 a.m.–9:00 a.m.). However, the non-interaction rate is lowest during this time (see Table 3.13) and average amount per fare transaction is also low, at $0.30 per passenger. This is likely due to the large share of AM Peak ridership by monthly passholders that commute. As a result, the estimated rate of total lost revenues is lower in the AM Peak than any other time period except for the evening hours. In contrast, the Midday period (9:00 a.m.–1:30 p.m.) has fewer boarding passengers but the highest average amount of fare per transaction, at $0.43. More passengers during this time period are paying full fares. Therefore, estimated revenue losses are greater in the Midday hours than in the AM Peak.

The Midday School (1:30 p.m.–4:00 p.m.) and PM Peak (4:00 p.m.–6:30 p.m.) periods show sustained high rates of passenger demand with corresponding high numbers of passenger non-interactions. Even though average amounts per fare transaction drop back down to $0.33 and $0.30, respectively, these time periods exhibit the highest estimated lost revenues. The Evening (6:30 p.m.–11:59 p.m.) period is characterized by the lowest passenger volumes of the day. Despite the non-interaction rate being highest in the Evening (see Table 3.13), the low total number of non-interactions makes the total revenue losses are lowest during these hours.

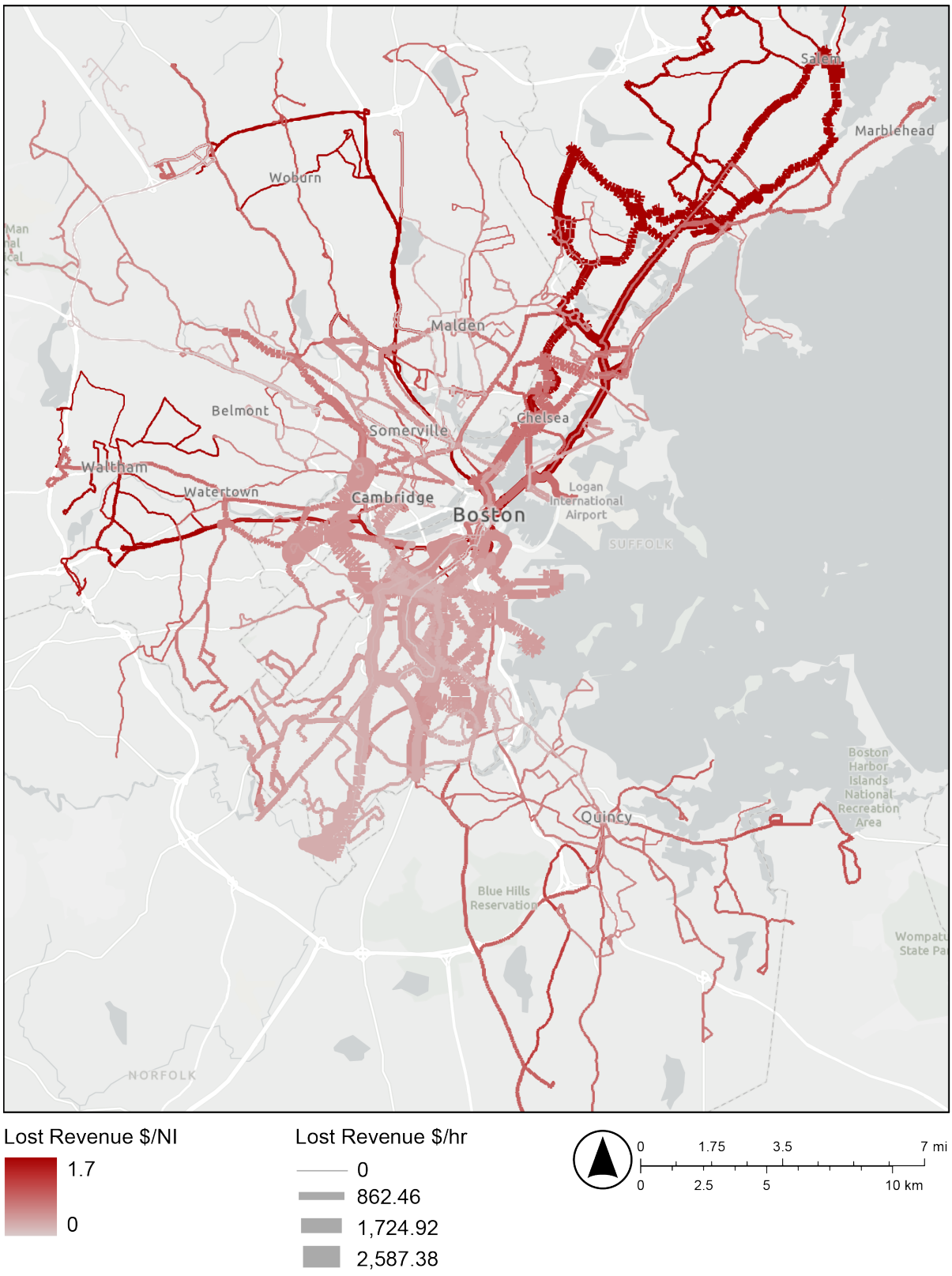

Figure 3.8 Lost Revenues per Hour and per Non-Interaction by Route

Table 3.11 Top 20 Bus Routes by Estimated Lost Revenue per Hour

Rank	Route	NI count/hr	$/NI	Lost $/hr
1	1	105.2	0.36	38.31
2	66	111.6	0.30	34.04
3	23	103.9	0.28	29.46
4	28	92.3	0.31	28.42
5	22	91.5	0.28	25.20
6	39	58.3	0.31	18.35
7	16	63.9	0.28	18.20
8	111	42.1	0.42	17.81
9	9	51.9	0.33	16.93
10	32	72.1	0.23	16.83
11	15	60.0	0.27	16.18
12	7	35.9	0.41	14.61
13	455	16.2	0.82	13.21
14	11	41.5	0.32	13.18
15	429	12.6	1.02	12.89
16	426	11.0	1.12	12.41
17	77	29.2	0.40	11.64
18	31	47.0	0.24	11.22
19	450	7.9	1.38	10.93
20	47	40.4	0.27	10.89

Table 3.12 Top 20 Bus Routes by Estimated Lost Revenue per Non-Interaction

Rank	Route	NI count/hr	$/NI	Lost $/hr
1	354	0.6	1.73	0.98
2	170	0.1	1.71	0.20
3	505	2.1	1.58	3.27
4	325	1.4	1.39	2.01
5	450	7.9	1.38	10.93
6	434	0.2	1.31	0.24
7	554	0.9	1.23	1.16
8	456	2.2	1.16	2.55
9	553	1.7	1.16	1.96
10	426	11.0	1.12	12.41
11	504	1.6	1.11	1.83
12	465	1.9	1.08	2.03
13	352	1.8	1.08	1.94
14	435	2.8	1.06	2.97
15	326	2.7	1.04	2.78
16	429	12.6	1.02	12.89
17	436	3.9	0.96	3.77
18	451	1.0	0.93	0.90
19	556	1.3	0.85	1.08
20	503	0.2	0.84	0.17

Table 3.13 Estimated Lost Revenue by Time of Day

Time Period	APC count/hr	NI count/hr	$/NI	Lost $/hr
AM Peak	12,910	2,398	$0.30	$731
Midday	8,611	1,861	$0.43	$797
Midday School	13,041	2,920	$0.33	$954
PM Peak	14,627	3,194	$0.30	$972
Evening	5,588	1,588	$0.32	$509
All Time Periods	10,090	2,215	$0.33	$742

3.3.5 Lost Revenues by Time and Location

The same grid system that is used for the spatial and temporal aggregation of non-interaction data (described in Section 3.2.4) is used for analysis of lost revenue. There are two values that are pertinent to the discussion of lost revenues. First, is the average fare collected per AFC transaction, which is the estimate for the lost revenue per non-interaction. Then, this location- and time-specific amount of lost revenue per non-interaction can be multiplied by the count of non-interactions per hour to estimate the total lost revenue per hour.

Appendix D includes maps of the average dollar amount of fare collected for AFC transaction records in each period. The maps show that fare amounts tend to be greater in outlying communities than in the center of the region. This is consistent with the fact that for transferring passengers, bus fares are paid at the farebox when the bus is the first mode used (as would be the case if boarding in an outlying neighborhood). Passengers that transfer to buses from the rail system have paid their fares at the rail system faregates and then have free transfer when boarding the bus. Stations with high numbers of transferring passengers include Ruggles, Forest Hills, and Nubian. Furthermore, express buses provide service in outlying communities, and fares for those service are higher than for local buses. The average fare amounts are also lowest during the AM Peak time period, when the travel demand is dominated by commuters, many of whom use monthly passes.

Appendix E also includes maps of the estimated total lost revenues per hour in each square of the grid. This value is the product of the average fare per transaction and the non-interaction count per hour. The spatial distribution of lost revenues is similar to the distribution of non-interaction counts, because there is more spatial variability in non-interactions (due to variability in total ridership) than the spatial variability of average fare amounts. The general trend is that revenue losses are more spatially dispersed in the AM Peak when passenger boardings are also spatial dispersed across the region. In the later time periods of the day passenger boardings are more concentrated in the city center, as are the estimated lost revenues.

3.3.6 Estimated Total Lost Revenues

Overall, the average amount per observed fare transaction is $0.33, which is an estimate of the average lost revenue per non-interaction. The average non-interaction rate is estimated as

0.22, based on the comparison of AFC and APC counts. Applying this to annual data for MBTA buses can provide an estimate of the possible magnitude of lost revenues from non-interacting bus passengers. The 2019 NTD Agency Profile reports 100,252,985 unlinked bus trips. Using the NI rate and average fare amount identified in this project, this would represent roughly $7.4 million in lost revenues in 2019.

An important caveat is that exempt passengers are not typically observed in the fare transaction data, especially children under the age of 12. Without manual observations, there is no data on how many of the non-interactions are these passengers. Although it is a coarse measure, APTA provides an estimate that 4% of bus riders nationally are children under the age of 14 [26]. Since none of the exempt passengers owe a fare, and exemptions make up less than 0.5% of the observed fare transactions, an alternative estimate of lost revenue can be made by reducing the number of estimated fare non-interaction to 0.18. This provides a lower estimate of lost revenues that reflects the fact that there may be a significant number of exempt riders getting counted by APC devices. For 2019, this lowers the estimated lost revenues to $6.0 million.

Since the COVID pandemic, transit ridership has dropped. The most recent NTD Agency Profile for 2023 reports 79,487,957 unlinked bus trips. If the make-up of riders has not changed since COVID, the logic described above would correspond to lost revenues in the range from $4.8 million to $5.8 million.

3.4 Modeling Lost Revenues

With the data aggregated by space and time as described in Section 3.3.3, it is possible to develop models to understand and predict the factors that drive the rate of lost revenues per hour in different locations. As described in Section 2.7, the models developed in this study are to estimate lost revenues per hour as the dependent variable. There are two ways that models can be useful. One is to understand the quantitative relationship between explanatory factors and the dependent variable, and regression techniques are well suited for this purpose. An OLS regression model is estimated to identify coefficient values that can be interpreted to guide policy decisions. The other purpose of models is to make accurate predictions of the lost revenue based on anticipated conditions or where observations to calculate revenue losses are not available. A well-performing regression model can be suitable for predictions, but it is often possible to obtain better performance with more sophisticated machine learning techniques.

In this section, results of an OLS regression analysis are presented and then compared with the results of a neural network machine learning model. For both models, the data segmented into two parts: 70% of the data was randomly sampled to form a training set, used to estimate the models; and the remaining 30% of the data formed a testing set, used to evaluate the accuracy of predicted values.

3.4.1 OLS Regression Model

A linear regression model was estimated to estimate lost revenue per hour in each zone based on the count of observed boarding passengers per hour within each zone as measured by APC, controlling for the day of the week, and time period of the day. Each data point corresponds to the estimated lost revenue per hour in a zone (as shown on the maps in Appendix B-E) for a time period and day.

The day of the week is a categorical variable with possible values of Monday, Wednesday, Thursday, and Friday. These are transformed into three dummy variables (Monday, Wednesday, and Thursday) to compare revenue losses per hour against Friday as a reference. The time period is also a categorical variable with possible values of AM Peak, Midday, Midday School, PM Peak, and Evening. These are transformed into four dummy variables (Midday, Midday School, PM Peak, and Evening) to compare against the AM Peak as a reference. Each dummy variable takes a value of 1 if an observation is in the corresponding day or time period, and it is 0 otherwise.

Each coefficient of the OLS model (shown in Table 3.14) represents the effect of a unit change in explanatory variable on the estimated lost revenue per hour in a zone. The coefficient for APC count/hour can be interpreted as the marginal lost revenue per boarding passenger as measured by the APC device. The value of $0.064/APC count is slightly less than the product of the average non-interaction rate (0.22 from Table 3.4) and the estimate lost revenue per non-interaction ($0.33 from Table 3.6), which is $0.073/boarding. The difference is attributable to the fact that the intercept and the coefficients for the four time period dummy variables are all positive, and therefor explain part of the lost revenues.

Table 3.14 OLS Model Coefficients to Estimate Lost Revenue per Hour

Feature	Coefficient	P-value	Std. Error
Intercept	0.323	0.0019	0.104
APC Count/Hour	0.064	0.0000	0.001
Monday	−0.008	0.9439	0.109
Thursday	−0.128	0.2422	0.110
Wednesday	−0.131	0.2278	0.108
Midday	1.555	0.0000	0.116
Midday School	1.009	0.0000	0.120
PM Peak	0.718	0.0000	0.117
Evening	0.491	0.0000	0.120

The p-value provides an indication of how likely the variable is to actually have no influence on the dependent variable. A low p-value implies a statistically significant relationship. A common threshold is to consider variables at the 95% significance level, which corresponds to p-values less than 0.05. By this measure, the intercept and coefficients of APC count/hour and time period variables are all statistically significant.

The standard error is a measure of the accuracy of the estimated coefficients with respect to the sample population. The standard error is the square root of the variance of the corresponding coefficient. Although the intercept and time period dummy variables all have statistically significant p-values, the relative magnitude of the standard error compared to the coefficient is much larger than for APC count/hour. This implies that the APC count/hour coefficient is the most accurate. It is also the driving explanatory factor in any zones with higher ridership, because the lost revenues increase linearly with APC count/hour while the other terms are binary.

Although Table 3.13 shows that lost revenues per hour are greatest during the PM Peak, the OLS results show that the coefficient is higher for the Midday and Midday School periods. This means that the greater hourly losses in the PM Peak are explained by higher APC counts during that time period rather than higher lost revenue per non-interaction. All of the time periods have positive coefficients, which indicates that, controlling for APC count, revenue losses are lowest in the AM Peak.

3.4.2 Neural Network Machine Learning Model

The neural network machine learning model was developed for the same set of explanatory variables to predict lost revenues per hour. The neural network model is structured as a set of linear models linking the explanatory and dependent variables through a set of hidden layers containing multiple nodes. A total of 13 neural network configurations were assessed, each with varying numbers of hidden layers and neurons.

- **1 Hidden Layer:** 8 neurons, 16 neurons, 32 neurons, 64 neurons, 128 neurons
- **2 Hidden Layers**: 64-32 neurons, 128-64 neurons
- **3 Hidden Layers**: 128-64-32 neurons, 64-32-16 neurons
- **4 Hidden Layers**: 128-64-32-16 neurons, 64-32-16-8 neurons
- **5 Hidden Layers**: 256-128-64-32-16 neurons, 128-24-32-16-8 neurons

These configurations allowed for a comprehensive comparison of how increasing the depth and number of neurons in each model affected predictive performance. The performance of each model was evaluated using mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), listed in Table 3.15 in increasing order of model complexity.

After testing models ranging from simple to more complex, the best performing model was a two hidden layer model with 128 and 64 neurons, respectively, as illustrated in Figure 3.9. This indicates that more complex models with more hidden layers and nodes, may overfit the data, reducing their overall accuracy. The two-layer model with 128 and 64 neurons provides a good balance between simplicity and effectiveness, avoiding the drawbacks of excessive complexity while still offering better results than simpler models.

Table 3.15 Comparison of Neural Network Model Structures

Hidden Layers	Neurons	MAE	MSE	RMSE
1	8	1.371313	12.00552	3.464899
1	16	1.413905	12.06506	3.473479
1	32	1.387	11.79714	3.434696
1	64	1.398206	11.86658	3.44479
1	128	1.367637	11.617	3.408371
2	64-32	1.379989	12.17315	3.489005
2	128-64	1.346536	11.03577	3.322014
3	128-64-32	1.385819	12.16248	3.487474
3	64-32-16	1.517075	11.44843	3.383553
4	128-64-32-16	1.40052	12.47552	3.532069
4	64-32-16-8	1.365358	12.00667	3.465064
5	256-128-64-32-16	1.424234	11.90503	3.450367
5	128-64-32-16-8	1.412586	12.36998	3.517098

In a machine learning model, it is not possible to generate a table of coefficients as for a regression. Instead, the permutation feature importance (described in Section 2.7.3.2) provides a score that indicates how important an input variable is for the accuracy of predicting the dependent variable. The feature importance of each explanatory variable is presented in Table 3.16 in decreasing order of importance. The magnitude of importance is more meaningful than the sign. Like the OLS regression, the neural network model shows that the count of boarding passengers is by far the most important determinant of lost revenues. Except for Midday-School, the time period is also consistently more important than day of the week for determining lost revenues.

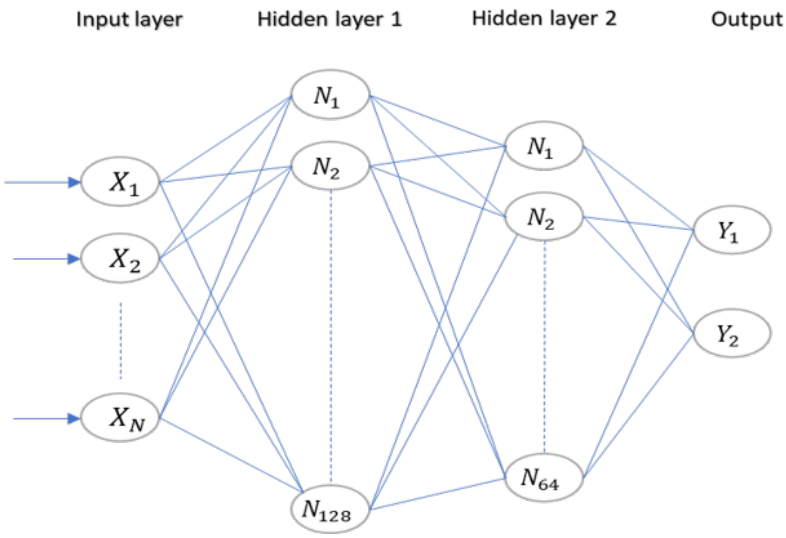

Figure 3.9 Structure of the Neural Network Model

Table 3.16 Feature Importance for Neural Network Model

Feature	Importance
APC Count/Hour	±6.737
PM Peak	±0.599
Evening	±0.107
AM Peak	±0.103
Midday	±0.099
Wednesday	±0.077
Monday	±0.066
Friday	±0.063
Midday School	±0.058
Thursday	±0.031

3.4.3 Comparison of OLS Regression and Neural Network Model Performance

The complexity of a machine learning model is only worthwhile if it offers better predictive capability than a simpler regression. A direct comparison of model accuracy is based on applying the fitted models to the test sample of data and evaluating the MAE, MSE, and RMSE, as shown in Table 3.17. By all three measures, the neural network outperforms the regression, although not by a large margin. This means that the interpretation of model coefficients from the OLS regression are relevant for supporting understanding of the relationship between values, but there are some non-linearities that are better captured by the neural network structure.

Table 3.17 Comparison of Model Performance

Performance Measure	OLS Regression	Neural Network
MAE	1.659796	1.346536
MSE	13.131715	11.03577
RMSE	3.623770	3.322014

Another way to compare model performance is to look at the fit of the data graphically. Figure 3.10 shows for the OLS regression the relationship between predicted values versus actual values. A perfectly accurate model would show all data points aligned on the dashed line with slope 1. Points above the line indicate overpredictions and points below indicate underpredictions. An alternative way to view this comparison is to plot the residuals, shown in Figure 3.11. The residual is the difference between the predicted and actual values, so a perfect model would have all residuals equal to 0. A positive residual is associated with an actual value that exceeds the prediction, and a negative residual is associated with an actual value that is less than the prediction.

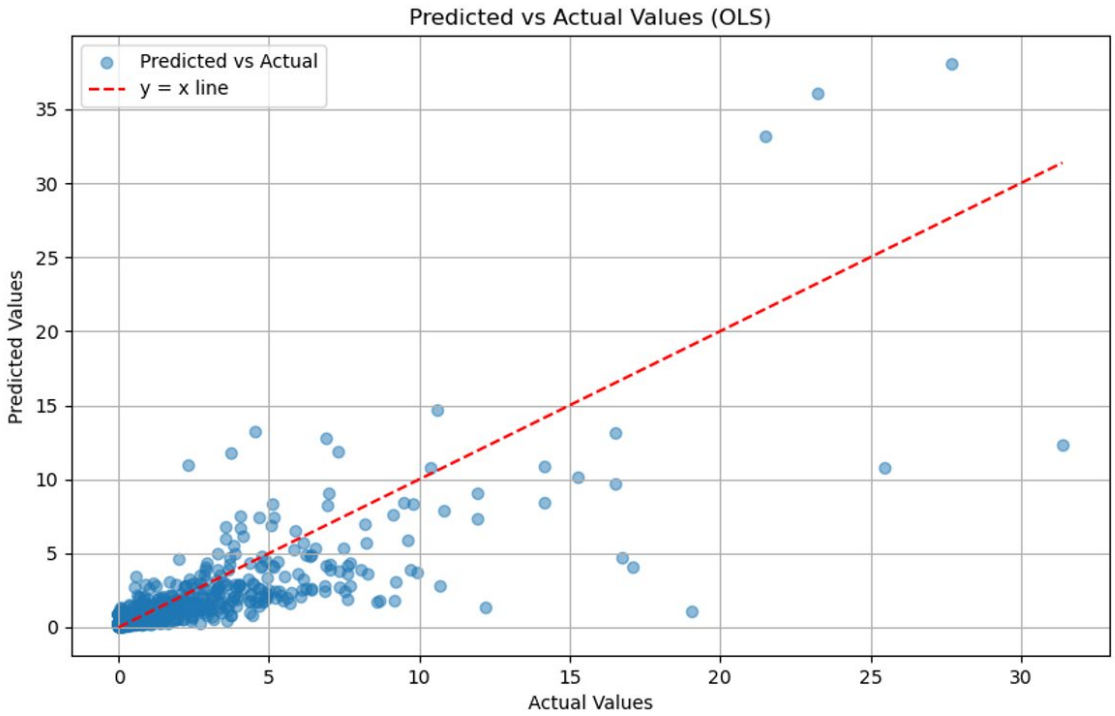

Figure 3.10 Predicted versus observed values using the OLS model

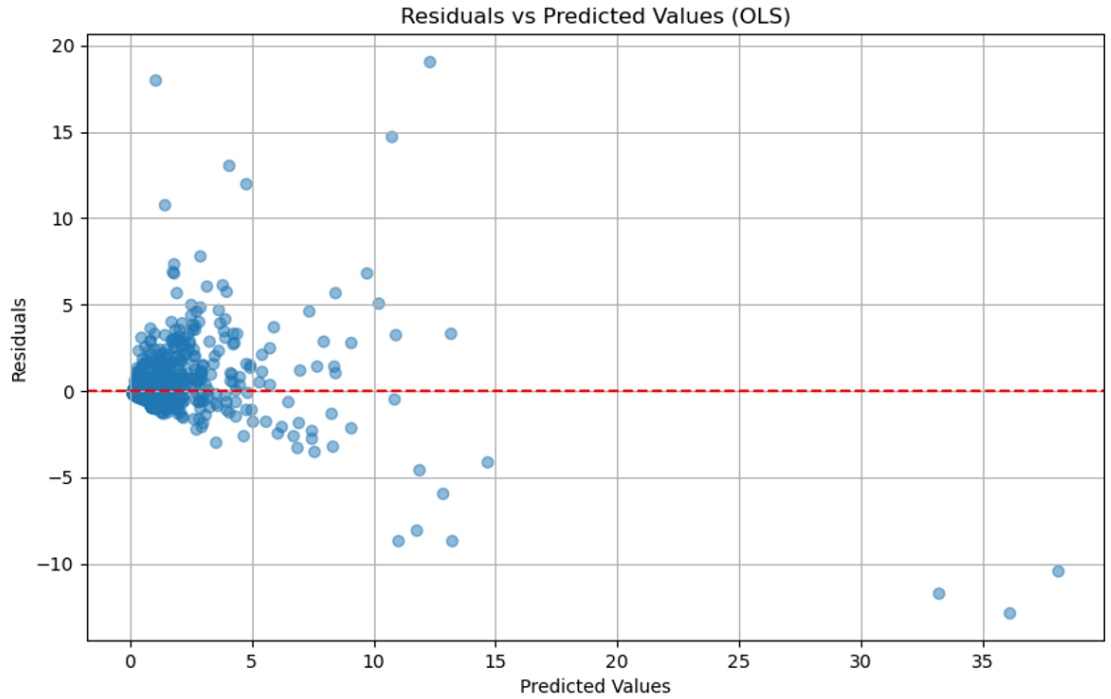

Figure 3.11 Residual error versus predicted values from the OLS model

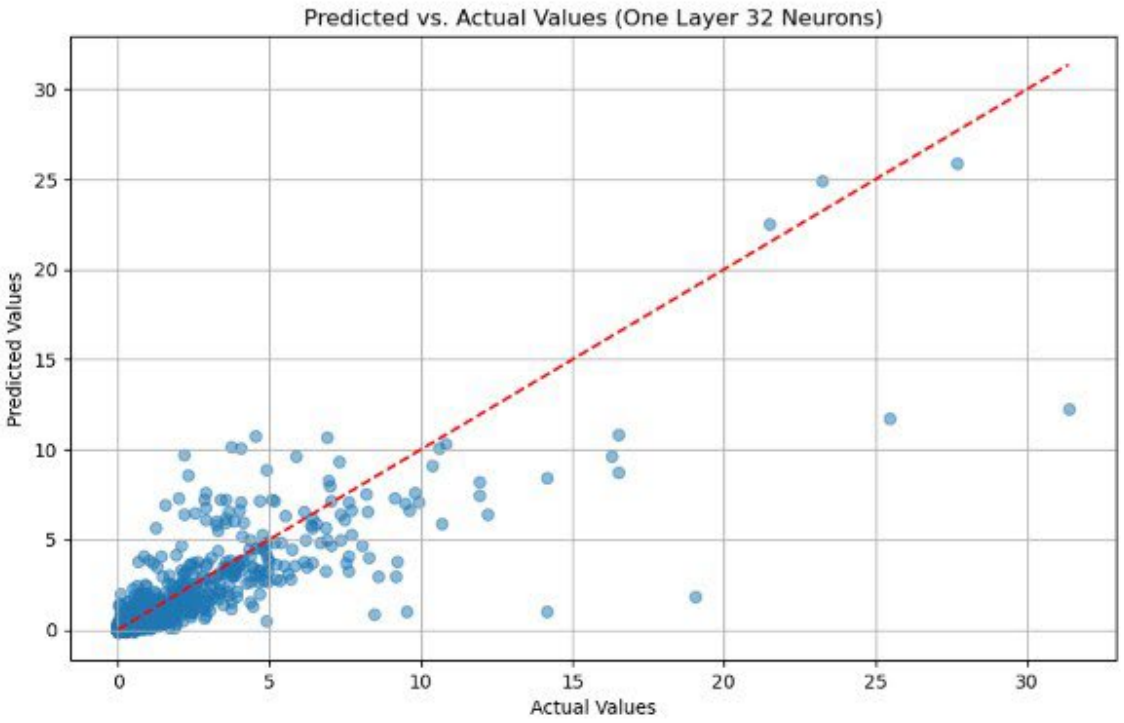

Figure 3.12 Predicted versus observed values using the NN model

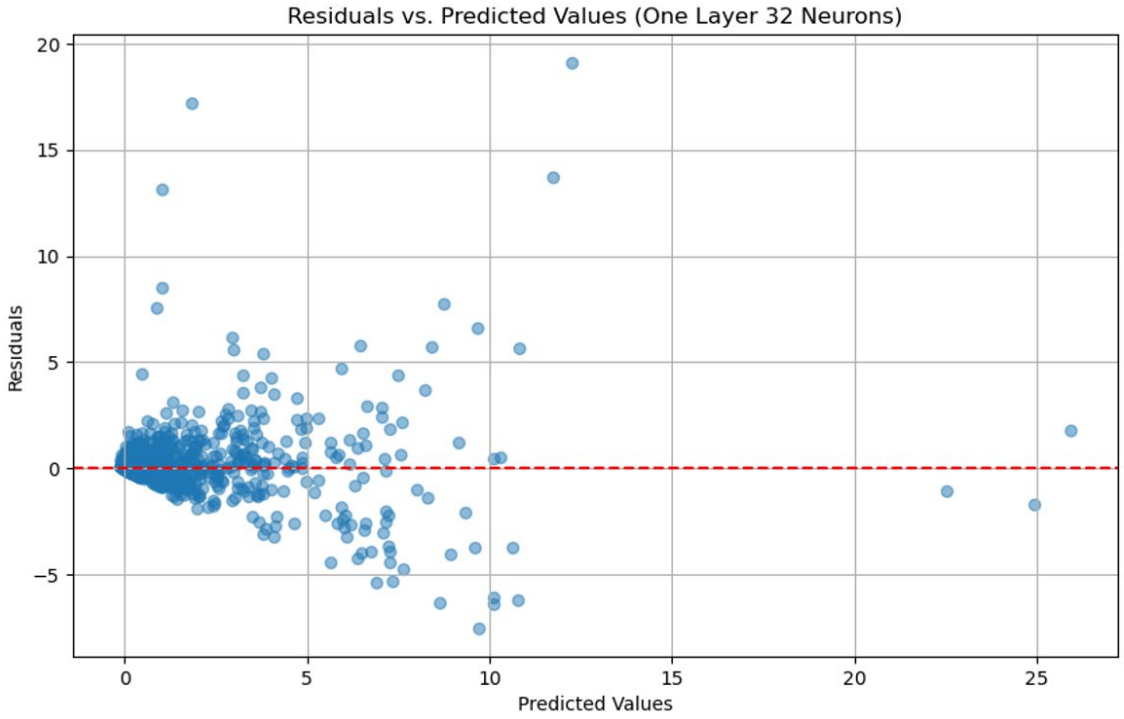

Figure 3.13 Residual error versus predicted values from the NN model

The performance of the neural network model is shown in Figure 3.12 by the relationship between the predicted values versus actual values. There are fewer points above the line of slope 1, which means that fewer observations are over-estimated by the model. The overall scatter of the data is similar error measures in Table 3.17. The residual error plot in Figure 3.13 also shows that the distribution of errors is similar to the OLS regression model.

Although there is similar scatter in the plots for both models, the OLS plots show some systematic underprediction of smaller values. The neural network model shows that residuals are more centered around 0, indicating that the model has less bias even if it is only a little more accurate overall. Since these models perform similarly, the simpler form and interpretable parameters of the OLS regression make it a more useful model for this set of data.

This page left blank intentionally.

4.0 Implementation and Technology Transfer

There are two main ways that this research can be implemented. The first is to use the findings and proposed methods to continue estimating and monitoring non-interactions and potential lost revenues. The second is to use this information to plan targeted deployment of staff for collecting additional data or conducting fare inspections.

4.1 Monitoring Fare Compliance

The proposed methods make use of data that are automatically collected from MBTA buses and that will be collected on newer light rail vehicles equipped with APC. Since the spatial and temporal scale of aggregation is flexible, the proposed method could be used to estimate systemwide average statistics or more detailed reports by location, route, or time. To support this outcome, the codes that were developed for the data processing and analyses in this project will be shared with MBTA staff so that they can continue to be used as new data is continuously coming in.

With the roll-out of new contactless fare payment methods as part of MBTA's Fare Transformation project, the proposed method of comparing fare transaction records with APC observations will remain a valid method to count non-interactions. The need to monitor and review this data may be of increased importance as passengers are no longer expected to interact with a farebox beside the driver upon boarding the vehicle.

4.2 Prioritizing Manual Observations

Collecting manual observations of fare compliance data by passive observation, surveys, or inspection are all labor-intensive activities, and thus costly to conduct. The lack of manual count data in recent years is an indication of the barrier this poses to monitoring fare compliance. As a result, fare system non-interactions, evasions, and lost revenues on parts of the transit system that are not controlled by faregates can only be roughly estimated.

Prioritizing times and locations to conduct additional manual observations depends on the objectives of the study. A focus could be on collecting data to understand patterns of behavior, in which case there is value in sampling across the system and times of day. A particularly valuable focus would be to conduct inspections or surveys of passengers in order to determine quantitative answers to the questions presented in Section 2.6:

1. How many of the non-interacting passengers owe a fare at the farebox, and what amount of fare do they owe?
2. How many of the non-interacting passengers do not owe additional fare because they are transferring or hold a valid pass?

3. How many of the non-interacting passengers are exempt from fares and therefore do not interact with the fare system in any case?

Furthermore, data collected onboard vehicles can be used to quantify the accuracy of APC counts in the context of the MBTA system. With a large enough sample of manual counts of boarding passengers to compare against APC counts, it would be possible to characterize the range of error in those automated counts and the ways that those errors vary with conditions like number of boarding passengers or vehicle crowding.

If the goal is to increase enforcement, the emphasis may be on reducing the rates of non-interactions or reducing the total amount of lost revenues. Perhaps the most straightforward cost-benefit calculation would be to deploy fare enforcement at the busiest locations during the midday and PM peak hours when lost revenues are greatest. In practice this may be implemented by focusing on particular transit stops where personnel can observe at the roadside or on transit routes where personnel can observe onboard the vehicle. The estimated revenue losses per hour reported for bus stops in Table 3.10 and for routes in Table 3.11 show that the magnitudes of revenue losses are not likely high enough to justify sustained enforcement across all times of day at many locations. Spot check could prioritize locations and routes that have notable high non-interaction counts and/or lost revenues.

The models may be useful for predicting where the greatest lost revenues are likely to be occurring, especially if data shows changing trends in passenger boardings. For example, the results of this study suggest that there is not much difference between weekdays, but the midday and PM peak periods are where the most revenue losses are occurring.

5.0 Conclusion

This study measures fare evasion and fare non-interaction within the MBTA bus system by using advanced data collection from the AFC and APC systems. Data from over 787,000 passenger boardings across the MBTA bus system and found that about 22% of these passengers did not interact with the fare system. Although it is unlikely that every non-interacting passenger is evading a full fare, this still suggests that significant revenue losses are associated with fare system non-interactions on buses.

This project makes use of automatically collected data associated with observed fare transactions and automated counts of boarding passengers. Although there are no direct observations of the types of passengers that do not interact with the fare payment system, a straightforward approach is to assume that the composition of non-interacting passengers is similar to the composition of observed transactions in terms of the numbers of passenger that hold valid pass, are making transfers, owe discounted fares, and owe full fares. A comparison of fare types from the AFC transaction records with manual inspections of rear-door boarding passengers from the 2017 Green Line study [9] shows consistency between the two. The average amount associated with a fare transaction on MBTA buses is $0.33, so this provides an estimate of the lost revenue per non-interacting passenger.

In addition to the passengers that are intended to interact with the fare payment system, there are a number of fare exemptions for children under the age of 12, MBTA employees, blind individuals, and active military in uniform, among others. There are very limited records of these passengers in the AFC database, yet these passengers would be counted by APC devices and contribute to the estimated non-interaction rate. Manual observations would be needed to quantify the numbers of exempt passengers that are using MBTA buses. In lieu of that data, APTA's estimate that 4% of bus riders nationally are children under the age of 14 [26] provides at least a rough estimate of the magnitude of this group of riders.

Using the observations from a sample of AFC and APC data collected in 2019, before the COVID-19 pandemic it is estimated that 22% of the 100 million unlinked bus trips in that year were non-interactions. This would translate to lost revenues in the range of $6.0 million - $7.4 million for bus riders alone in 2019. The higher estimate is if all non-interactions are associated with the average $0.33 fare amount per transaction, and the lower estimate accounts for the potential number of exempt children and other passengers.

Aside from seeking to quantify systemwide totals for the number of fare system non-interactions and lost revenues, an important goal of this project was to evaluate the variation by time and location in order to provide insights into patterns of passenger behaviors and where it would be most valuable to collect additional manual observations.

One of the key findings is the significant variation in non-interaction rates and associated revenue losses across different times of the day. The morning peak is associated with high ridership, but relatively low non-interaction rates and low fare amounts per passenger. This is likely because the morning is dominated by regular commuters who hold monthly passes. The midday and afternoon peak hours exhibit high demand and increased non-interactions

which lead to greater lost revenues. The midday, in particular, has much higher average fare amounts among observed transactions, likely because there are relatively fewer passholders and more full-fare paying passengers using buses during those hours. This shows that the value of targeted inspections would be greatest during these midday and afternoon peak hours to gather more specific data when most non-interactions are happening and to mitigate revenue losses.

Additionally, the geographical analysis revealed that the counts of non-interactions and the resulting lost revenues are closely correlated with total ridership. The non-interaction rate appears random across the region at all times of the day. The highest non-interaction rates appear in outlying neighborhoods where ridership is low. In these locations, even a single non-interaction can be a high percentage of a small number of observed boardings.

If the goal of efforts to collect additional manual observations is to maximize the effectiveness of those efforts, inspecting or surveying passengers in the busiest parts of the system and at the busiest times of day is likely to provide the most useful data to address assumptions made in this project due to limitations of the available automated data. Specifically, there would be value in comparing manual boarding counts to APC counts in different parts of the MBTA system in order to quantify the errors associated with those measurements. It would also be valuable to inspect or survey non-interacting passengers in order to determine how the composition of non-interacting passengers compares to the observed fare transactions.

Overall, these findings give transit authorities useful insights for creating more effective strategies to handle fare evasion and recover the resulting revenue losses. By understanding the patterns of non-interaction and revenue losses, especially during peak times and in high-risk areas, transit agencies can better focus their resources on fare inspections and enforcement. This approach will help improve revenue recovery and operational efficiency. Unlike traditional studies that rely on samples collected from surveys or inspection data, our approach uses entirely automatically collected data from all trips. This comprehensive method provides a strong framework for continually monitoring and managing fare compliance in public transit systems.

6.0 References

1. Barabino, B. and S. Salis (2020). Do students, workers, and unemployed passengers respond differently to the intention to evade fares? An empirical research. *Transportation Research Interdisciplinary Perspectives*, 7:100215.

2. Barabino, B., C. Lai and A. Olivo (2020). Fare evasion in public transport systems: A review of the literature. *Public Transport*, 12:27-88.

3. Keuchel, S. and K. Laurenz (2018). The effects of a higher ticket inspection rate in a medium-size public transportation system. *Transportation Research Procedia*, 31:56-66.

4. Egu, O. and P. Bonnel (2020). Can we estimate accurately fare evasion without a survey? Results from a data comparison approach in Lyon using fare collection data, fare inspection data and counting data. *Public Transport*, 12(1):1-26.

5. Wolfgram, L., C. Pollan, K. Hostetter, A. Martin, T. Spencer, S. Rodda and A. Amey (2022). Measuring and Managing Fare Evasion. TCRP Research Report 234. Transportation Research Board: Washington, D.C.

6. Barabino, B., S. Salis and B. Useli (2014). Fare evasion in proof-of-payment transit systems: Deriving the optimum inspection level. *Transportation Research Part B: Methodological*, 70:1-17.

7. Lee, J. (2011). Uncovering San Francisco, California, Muni's proof-of-payment patterns to help reduce fare evasion. *Transportation Research Record*, 2216(1):75-84.

8. Hansen, S., B. Whitelaw and J. D. Leong (2012). Tackling fare evasion on Calgary transit's CTrain system. *Sustaining the Metropolis*, 84.

9. Prokosch, A. and A. Gartsman (2017). All-door boarding without proof-of-payment: Revenue impacts and operational implications. Paper Number 17-06500. *Transportation Research Board Annual Meeting 96th Annual Meeting*, January 8-12. Washington, D.C.

10. Larwin, T. F. (2012). Off-board fare payment using proof-of-payment verification. TCRP Synthesis 96. Transportation Research Board: Washington, D.C.

11. Barabino, B., M. Di Francesco and S. Mozzoni (2014). An offline framework for handling automatic passenger counting raw data. *IEEE Transactions on Intelligent Transportation Systems*. 15(6):2443-2456.

12. Pourmonet, H., S. Bassetto and M. Trépanier (2015). Vers la maîtrise de l'évasion tarifaire dans un réseau de transport collectif. *11e Congrès International De Génie Industriel*. Québec, Canada.

13. Yin, T., N. Nassir, J. Leong, E. Tanin and M. Sarvi (2022). Investigation into public transport fare noninteractions using large scale automatically collected data. *Australasian Transport Research Forum (ATRF), 43rd*. Adelaide, Australia.

14. Andrews, S., A. Demchur, A. Reker, K. Dumas, and K. DeLauri (2016). MBTA Passenger Noninteraction with Automated Fare Collection Equipment. Report. Central Transportation Planning Staff, Boston Region Metropolitan Planning Organization.

15. Barabino, B., M. Di Francesco and R. Ventura (2023). Evaluating fare evasion risk in bus transit networks. *Transportation Research Interdisciplinary Perspectives*, 20, 100854.

16. Munizaga, M. A., A. Gschwender and N. Gallegos (2020). Fare evasion correction for smartcard-based origin-destination matrices. *Transportation Research Part A: Policy and Practice*, 141:307-322.

17. Sánchez-Martínez, G. E. (2017). Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. *Transportation Research Record*, *2652*(1), 1-7.

18. Pronello, C., & Garzón Ruiz, X.R. (2023). Evaluating the performance of video-based automated passenger counting systems in real-world conditions: A comparative study. *Sensors*, 23, 7719.

19. Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

20. Clark, H. (2017). Who Rides Public Transportation. American Public Transportation Association. Available from: https://www.apta.com/research-technical-resources/research-reports/who-rides-public-transportation/

21. Goodfellow, I., *Deep Learning*. 2016, MIT press.

22. Hinton, G.E. Learning distributed representations of concepts. in Proceedings of the eighth annual conference of the cognitive science society. 1986. Amherst, MA.

23. Kingma, D.P., Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

24. Breiman, L., Random forests. Machine learning, 2001. 45: p. 5-32.

25. Molnar, C., Interpretable machine learning. 2020: Lulu.com.

26. Massachusetts Bay Transportation Authority (2021), MBTA Service Delivery Policy 2021. MBTA Fiscal and Management Control Board.

Appendix A. Fare Payment Types

The observed fare payment types, frequency, and average transaction amount are reported for AFC.faretransaction records associated with buses on four weekdays: Wednesday, September 18, 2019; Thursday, September 26, 2019; Friday, October 4, 2019; and Monday, October 7, 2019.

Table A.1 Fare Payment Types in AFC.faretransaction Records

Description	Count	Percent	Avg. Amount
Time Limited Passes			
Monthly Link Pass	314,372	24.4%	$ —
Monthly Link Student 7 Days	148,325	11.5%	$ —
7-Day LinkPass - RPP	115,191	9.0%	$ —
Local Bus Monthly Pass Adult	67,514	5.2%	$ —
7 Day Link Pass active FVM/TOM/RST	58,828	4.6%	$ —
Monthly Link Senior	45,617	3.5%	$ —
Monthly Link T.A.P.	37,729	2.9%	$ —
Inner Express Bus Monthly Pass	21,988	1.7%	$ —
CR Monthly Pass Adult Zone 1a	11,628	0.9%	$ —
Monthly LinkPass Youth	10,007	0.8%	$ —
Outer Express Bus Monthly Pass	4,072	0.3%	$ —
CR Monthly Pass Adult Zone 2	2,251	0.2%	$ —
CR Monthly Pass Adult Zone 1	1,851	0.1%	$ —
CR Monthly Pass Adult Zone 4	1,638	0.1%	$ —
CR Monthly Pass Adult Zone 3	1,637	0.1%	$ —
CR Monthly Pass Adult Zone 6	1,259	0.1%	$ —
CR Monthly Pass Adult Zone 8	956	0.1%	$ —
CR Monthly Pass Adult Zone 7	807	0.1%	$ —
CR Monthly Pass Adult Zone 5	614	0.0%	$ —
** ID w/o SV Retiree	612	0.0%	$ —
1 Day Link Pass active FVM/TOM/RST/FBX	557	0.0%	$ —
1-Day LinkPass - RPP	157	0.0%	$ —
Monthly Commuter Boat Pass	106	0.0%	$ —
CR Monthly IZ Pass 4 Zones	13	0.0%	$ —
CR Monthly IZ Pass 3 Zones	9	0.0%	$ —
CR Monthly IZ Pass 5 Zones	9	0.0%	$ —
CR Monthly IZ Pass 1 Zone	7	0.0%	$ —
CR Monthly IZ Pass 7 Zones	4	0.0%	$ —
CR Monthly IZ Pass 2 Zones	2	0.0%	$ —
CR Monthly IZ Pass 6 Zones	2	0.0%	$ —
Subtotal	847,762	65.9%	—

Transfer Passengers			
Transfer Flex Adult	71,444	5.6%	$ 0.03
Transfer Flex Senior	11,053	0.9%	$ 0.01
Transfer Flex T.A.P.	7,570	0.6%	$ 0.01
Transfer Flex Student	2,367	0.2%	$ 0.01
Transfer Flex Youth	677	0.1%	$ 0.02
Subtotal	93,111	7.2%	—
Exempt Passengers*			
ID without SV Blind 5 yr. Validity	3,670	0.3%	$ —
The RIDE ID	983	0.1%	$ —
Bus Cash Police Fire	62	0.0%	$ —
Public Official Ids w/o SV w Pb	41	0.0%	$ —
Bus Cash Blind Person	35	0.0%	$ —
Subtotal	4,791	0.4%	—
Discounted Fare			
ID with SV Senior	25,921	2.0%	$ 0.90
ID with SV T.A.P. 5 yr. Validity	16,432	1.3%	$ 0.87
Student IDs w SV w 5 day	13,090	1.0%	$ 0.86
Bus Cash Senior	2,398	0.2%	$ 0.86
Permit Senior/TAP 30 days validity	2,002	0.2%	$ 0.65
Youth ID w SV	1,415	0.1%	$ 0.90
Bus Cash Student	1,400	0.1%	$ 0.86
ID with SV T.A.P. 1 yr. Validity	474	0.0%	$ 0.91
ID with SV T.A.P. 4 yr. Validity	168	0.0%	$ 0.87
Bus Cash Retiree	156	0.0%	$ —
ID with SV T.A.P. 3 yr. Validity	106	0.0%	$ 0.89
Bus Cash T.A.P.	72	0.0%	$ 0.85
ID with SV T.A.P. 2 yr. Validity	39	0.0%	$ 0.85
Subtotal	63,673	4.9%	—
Full Fare			
SV Adult (SC)	198,705	15.4%	$ 1.82
Bus Cash Short	38,785	3.0%	$ 0.70
Bus Cash Adult	25,363	2.0%	$ 2.06
SV Adult (MC)	10,663	0.8%	$ 1.78
SV Change Card Adult	3,795	0.3%	$ 1.74
Subtotal	277,311	21.6%	—

*Note that the exemptions do not include any records of children under 12 years of age. This entire category is likely to undercount true values, because exempt passengers are not expected to interact with the fare payment system.

Appendix B. Maps of Non-Interaction Counts

Interactive versions of all maps are available online at https://arcg.is/0XqODi0.

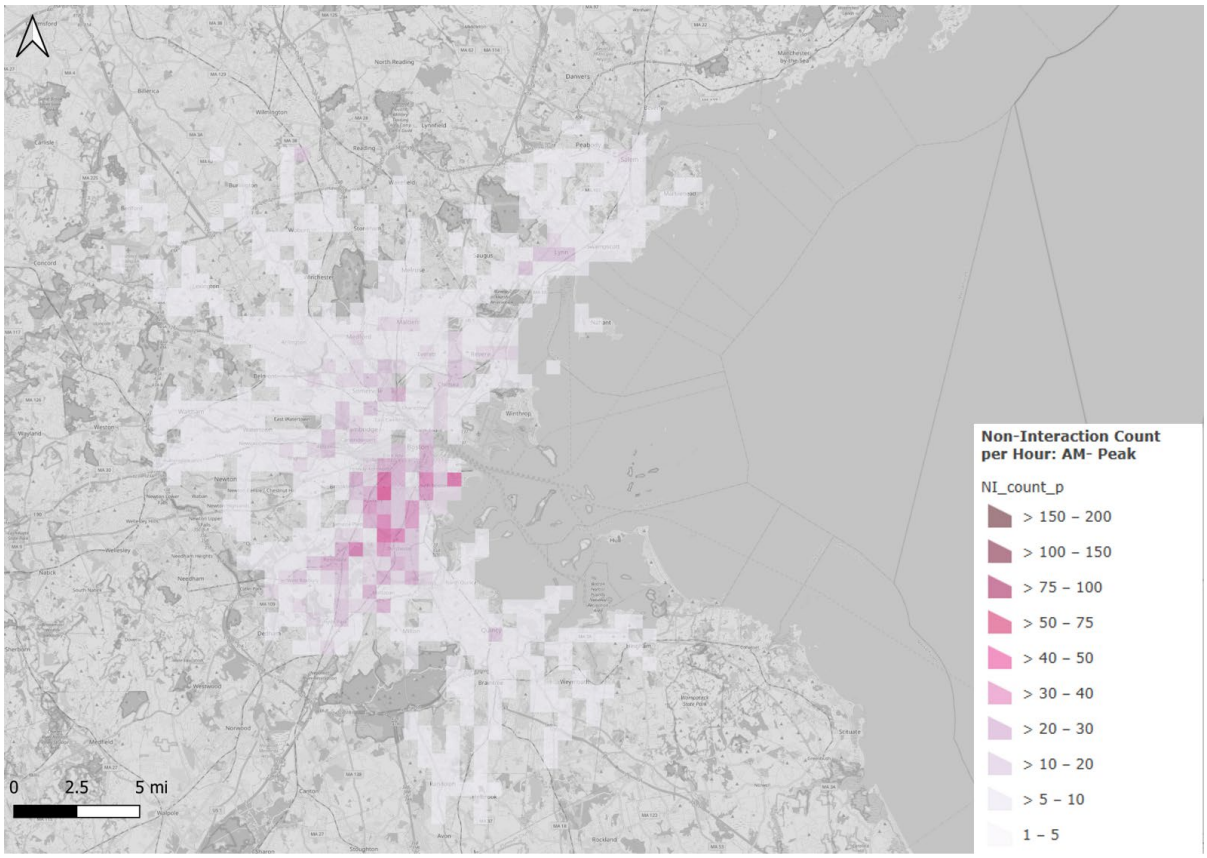

Figure B.1 Non-Interaction Count per Hour in AM Peak (5:30 a.m.–9:00 a.m.)

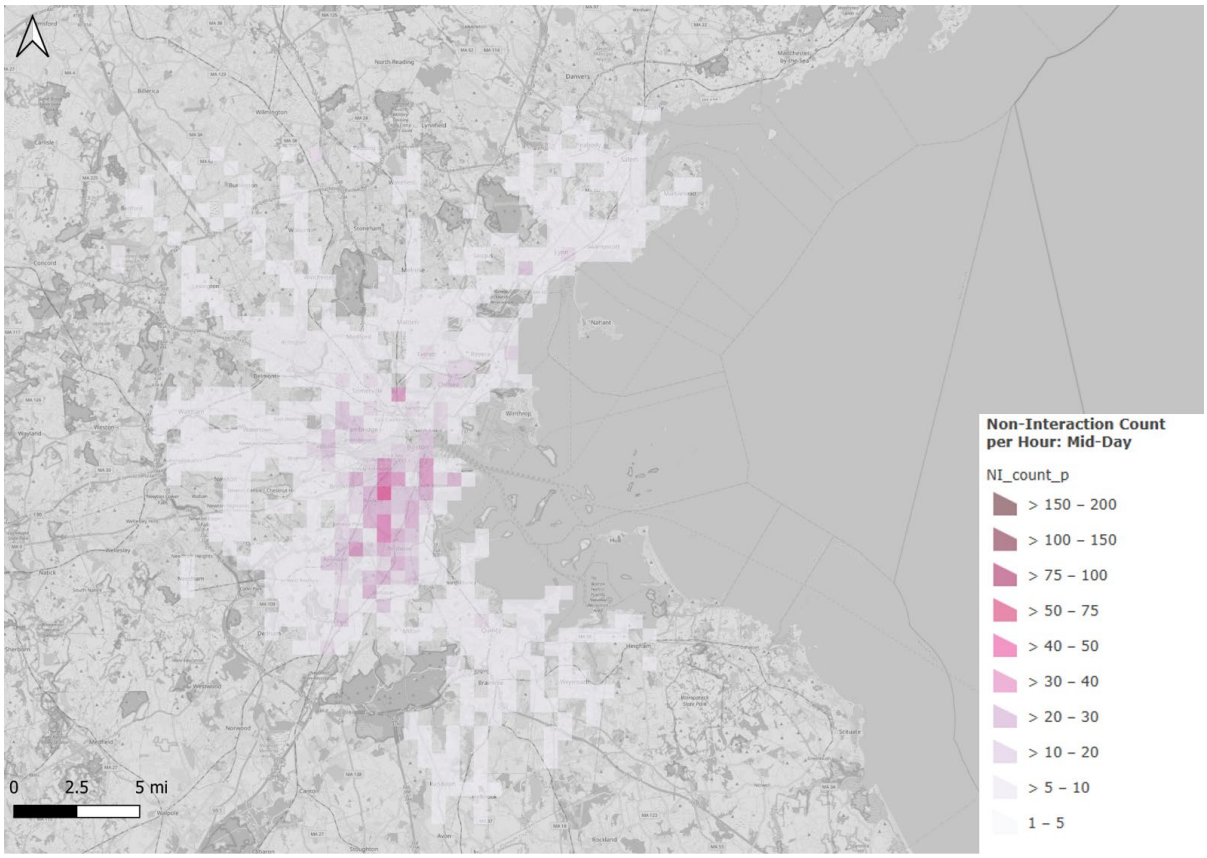

Figure B.2 Non-Interaction Count per Hour in Midday (9:00 a.m.–1:30 p.m.)

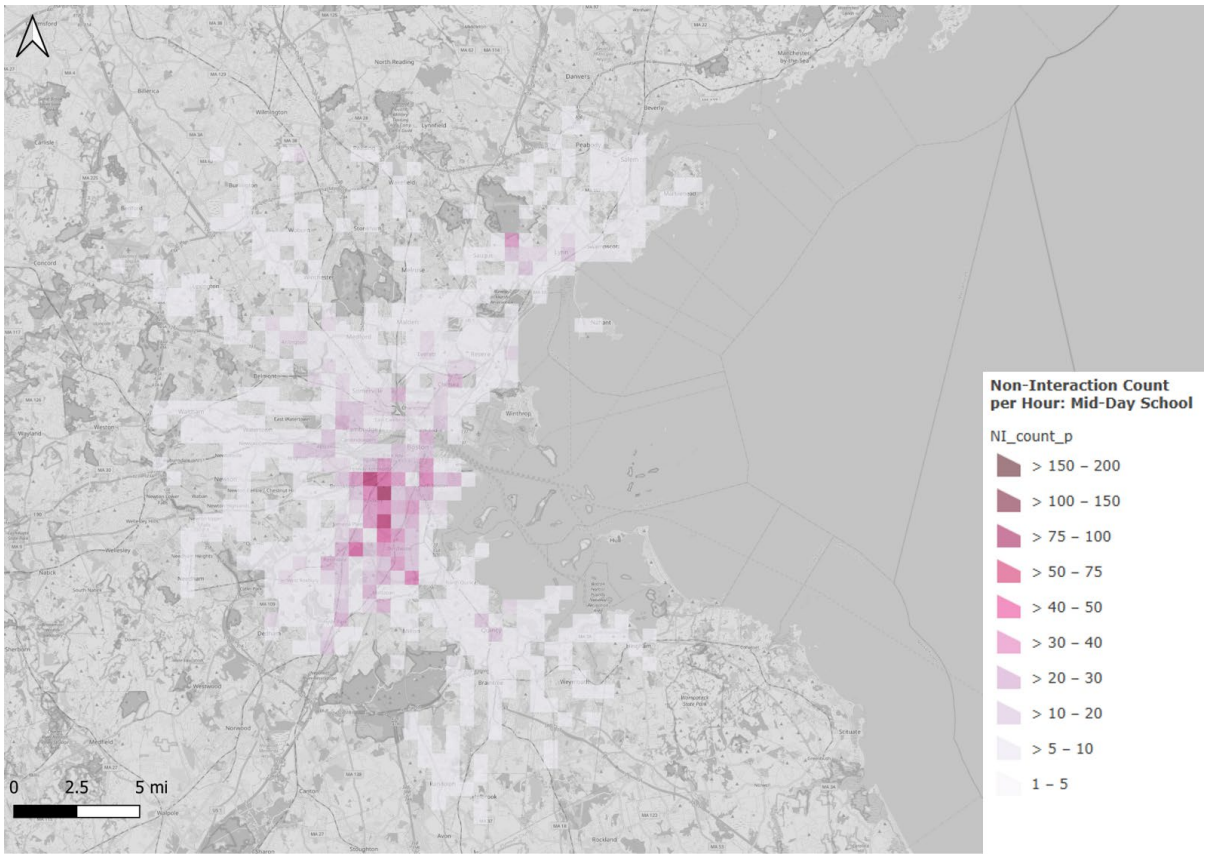

Figure B.3 Non-Interaction Count per Hour in Midday School (1:30 p.m.–4:00 p.m.)

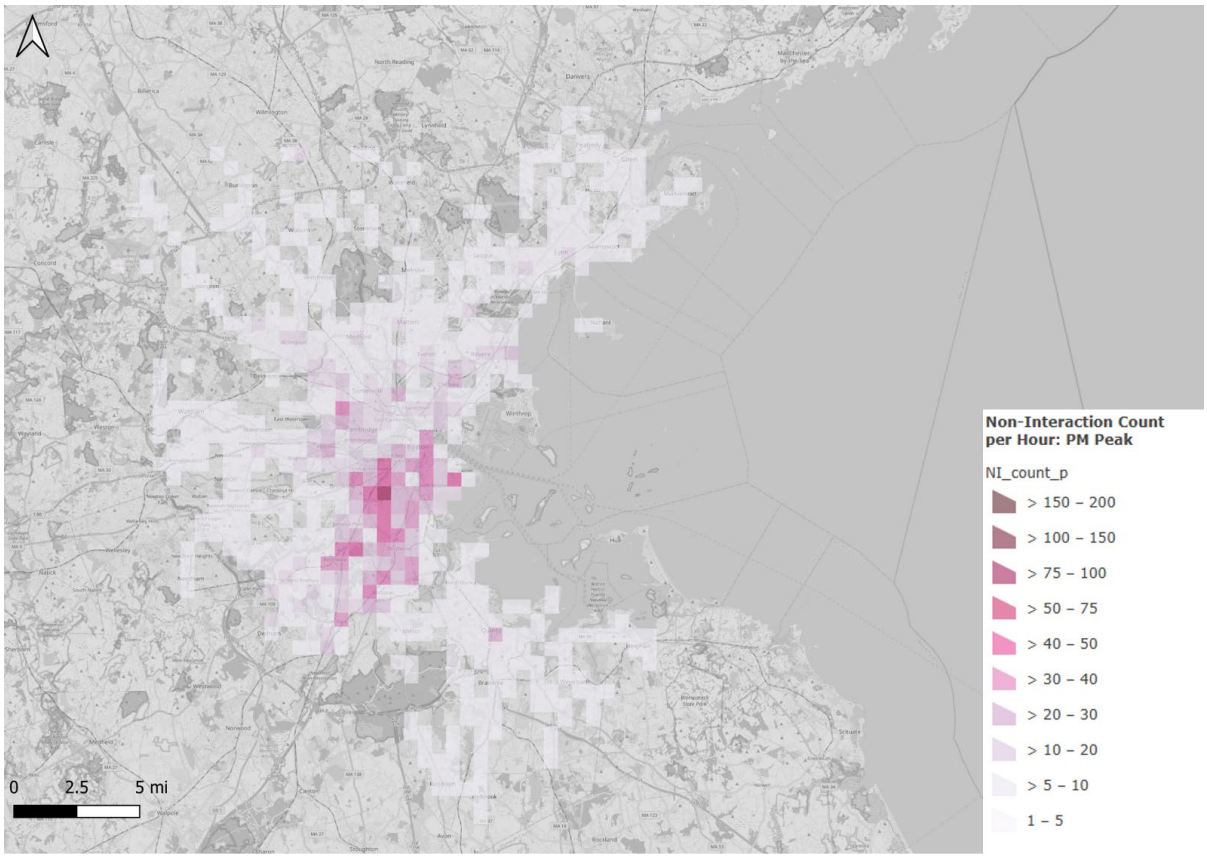

Figure B.4 Non-Interaction Count per Hour in PM Peak (4:00 p.m.–6:30 p.m.)

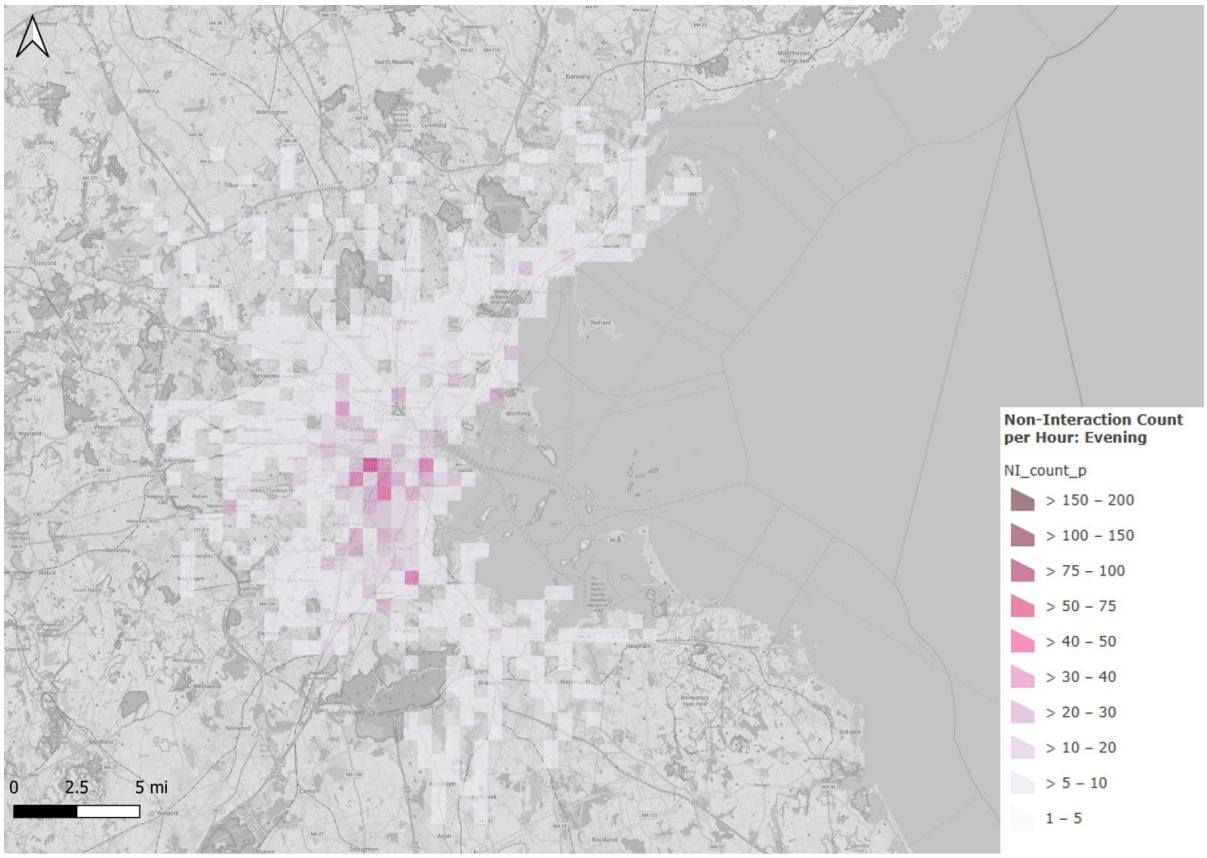

Figure B.5 Non-Interaction Count per Hour in Evening (6:30 p.m.–11:59 p.m.)

Appendix C. Maps of Non-Interaction Rates

Interactive versions of all maps are available online at https://arcg.is/0XqODi0.

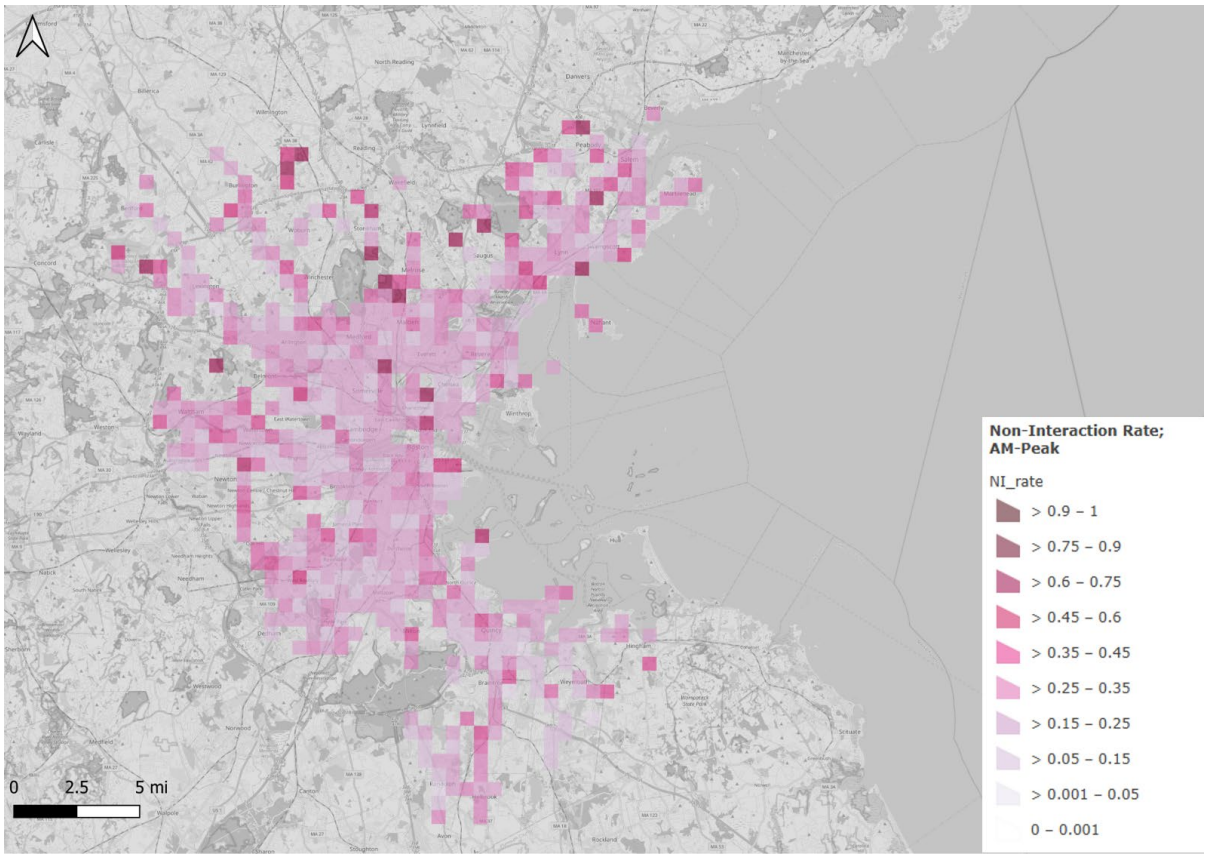

Figure C.6 Non-Interaction Rate in AM Peak (5:30 a.m.–9:00 a.m.)

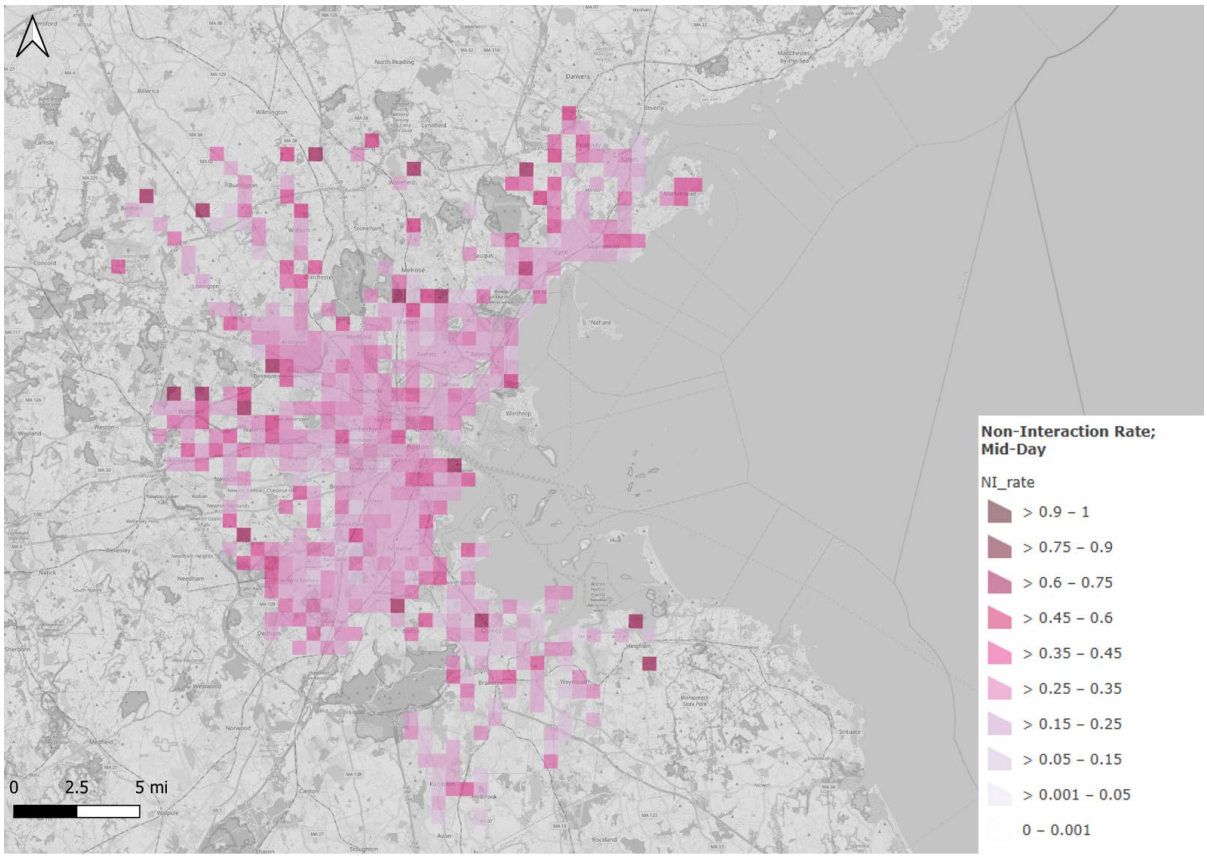

Figure C.7 Non-Interaction Rate in Midday (9:00 a.m.–1:30 p.m.)

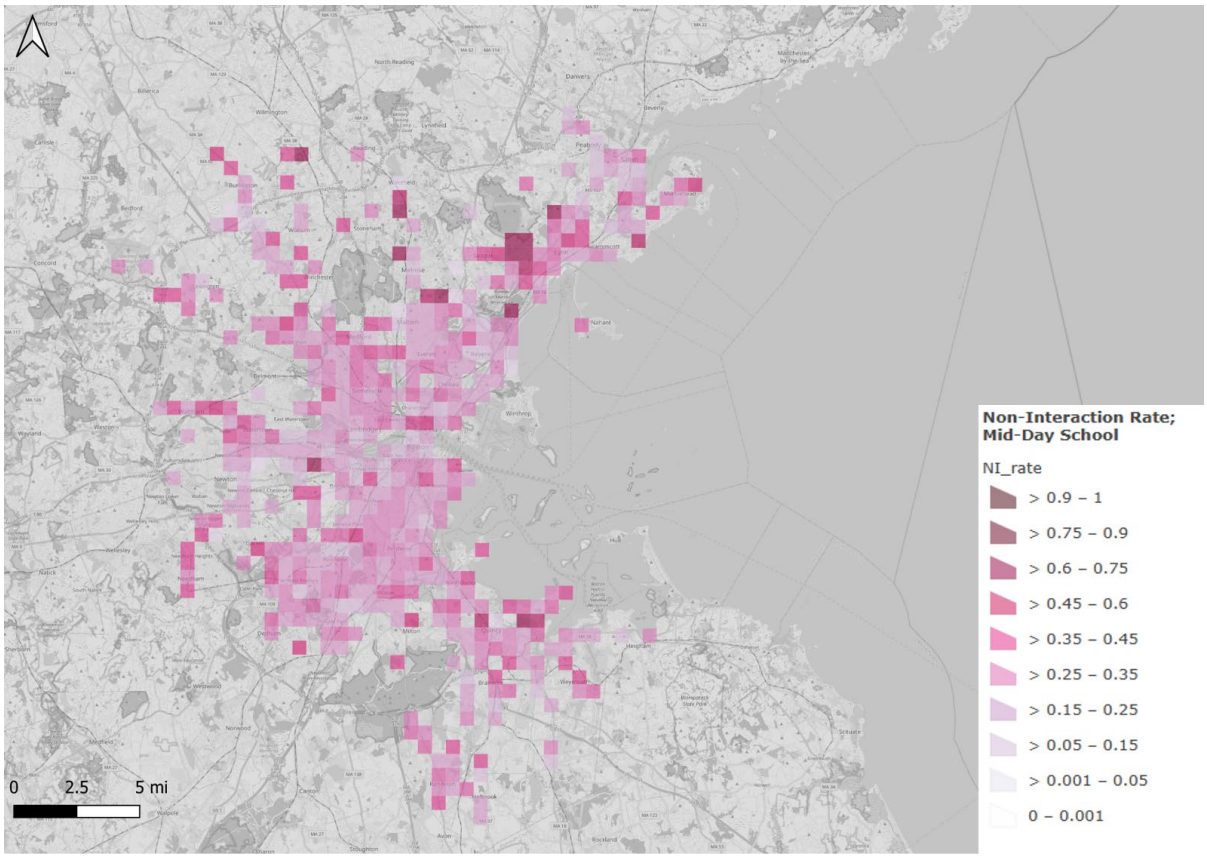

Figure C.8 Non-Interaction Rate in Midday School (1:30 p.m.–4:00 p.m.)

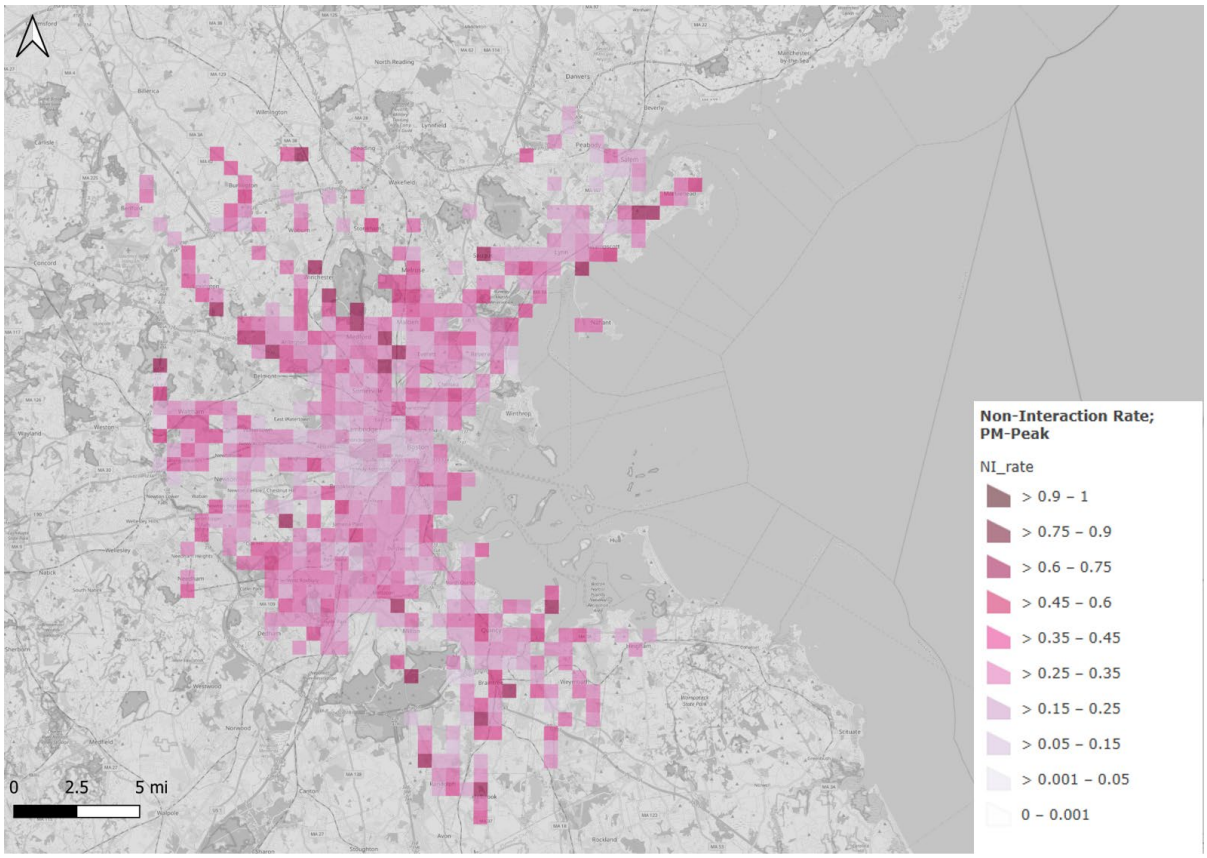

Figure C.9 Non-Interaction Rate in PM Peak (4:00 p.m.–6:30 p.m.)

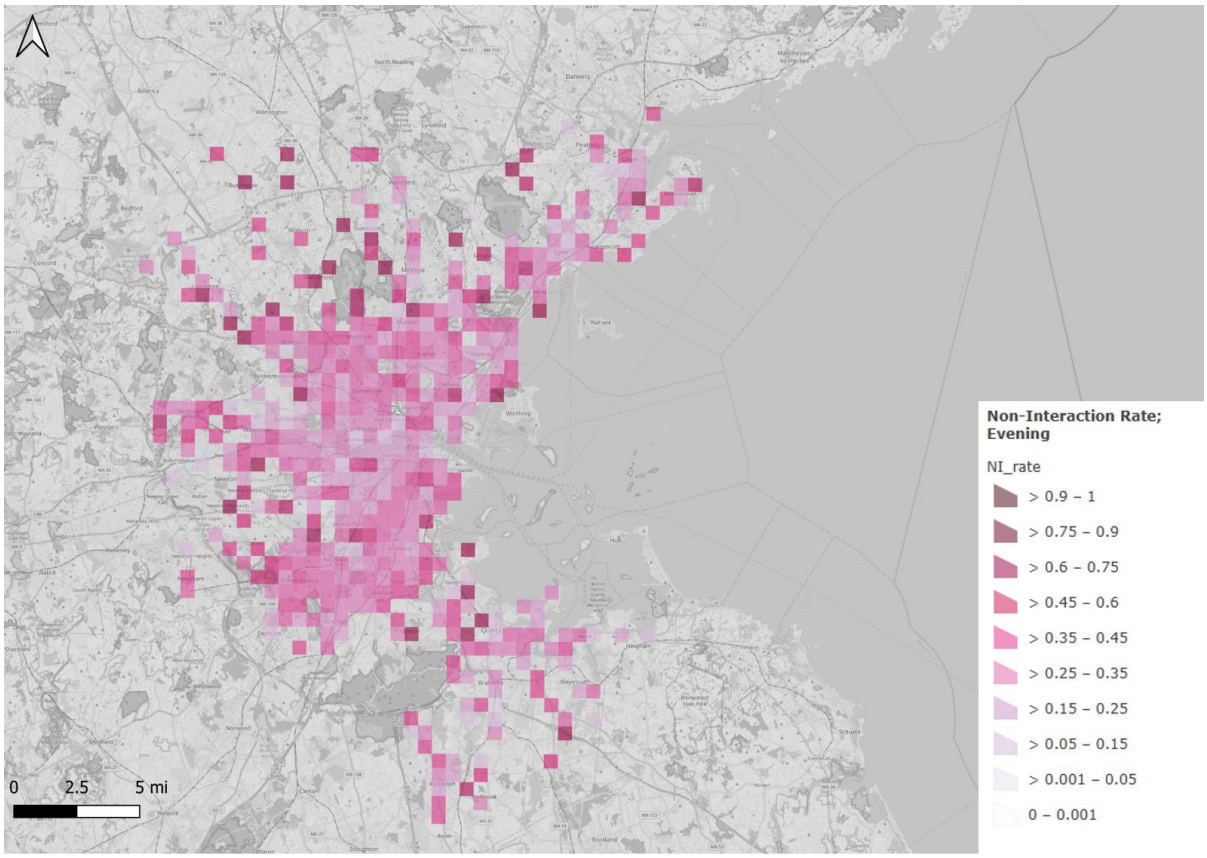

Figure C.10 Non-Interaction Rate in Evening (6:30 p.m.–11:59 p.m.)

Appendix D. Maps of Average Fare Amounts

Interactive versions of all maps are available online at https://arcg.is/0XqODi0.

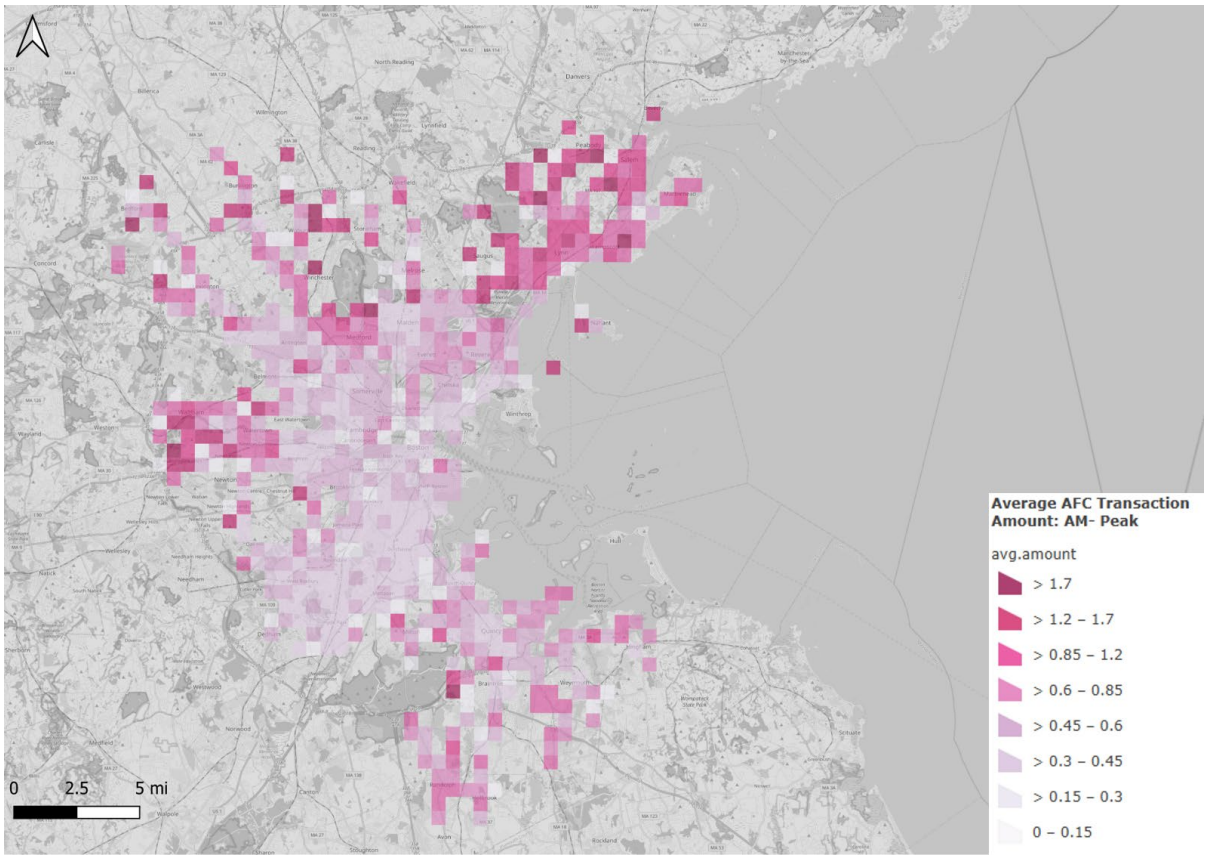

Figure D.11 Average AFC Transaction Amount in AM Peak (5:30 a.m.–9:00 a.m.)

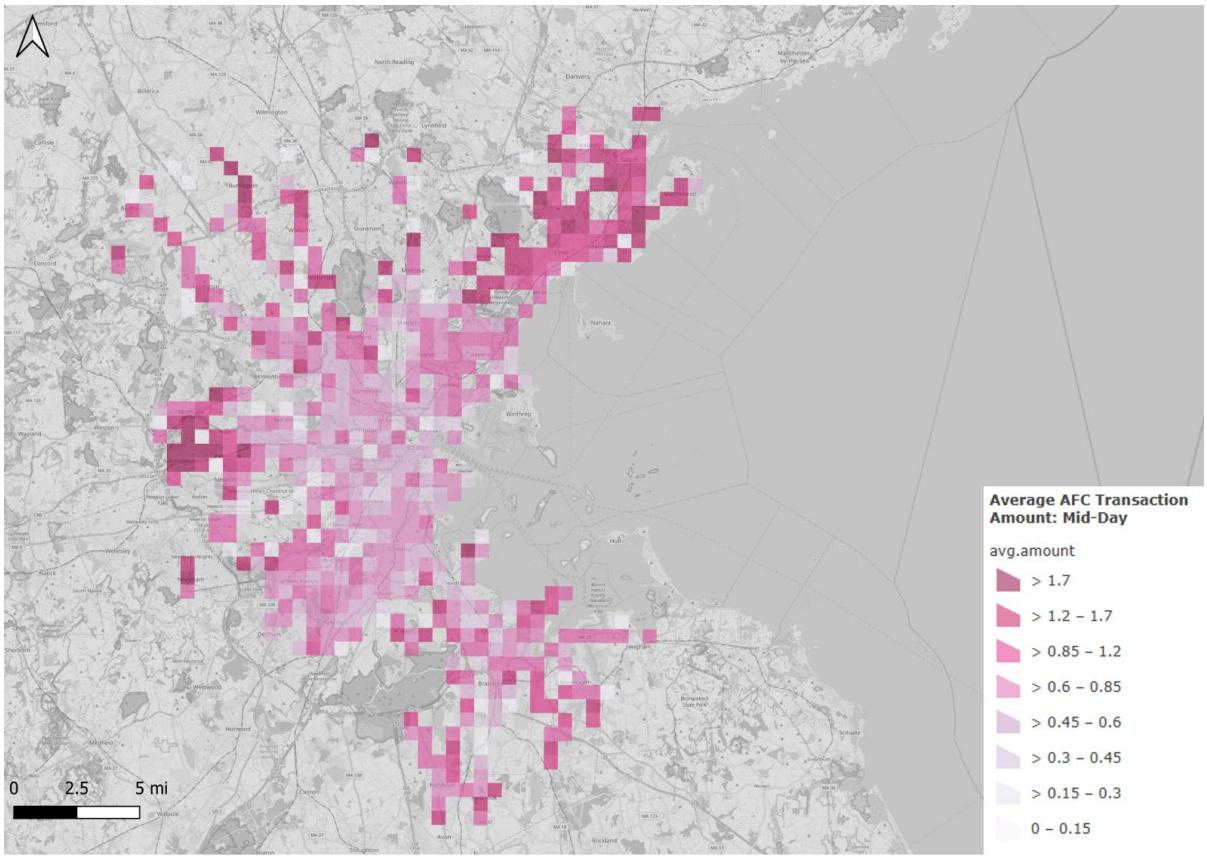

Figure D.12 Average AFC Transaction Amount in Midday (9:00 a.m.–1:30 p.m.)

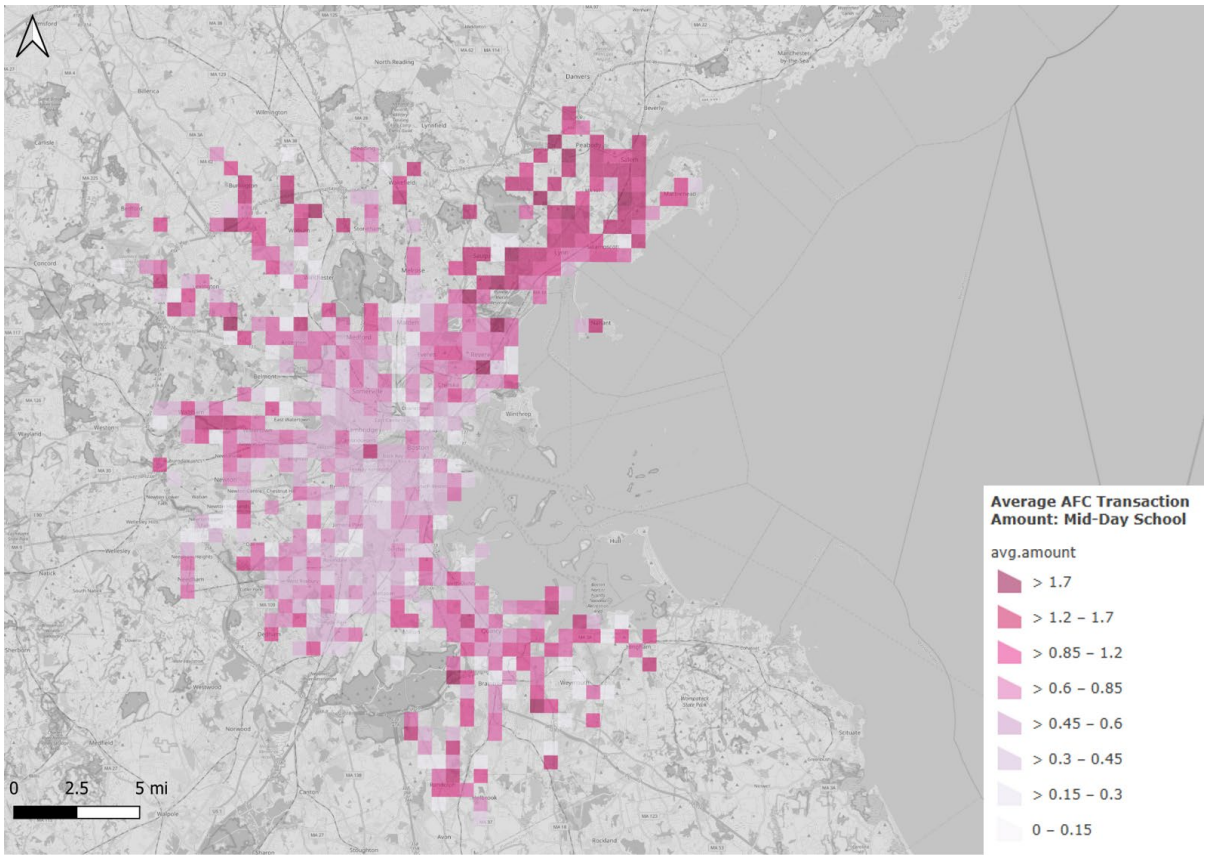

Figure D.13 Average AFC Transaction Amount in Midday School (1:30 p.m.–4:00 p.m.)

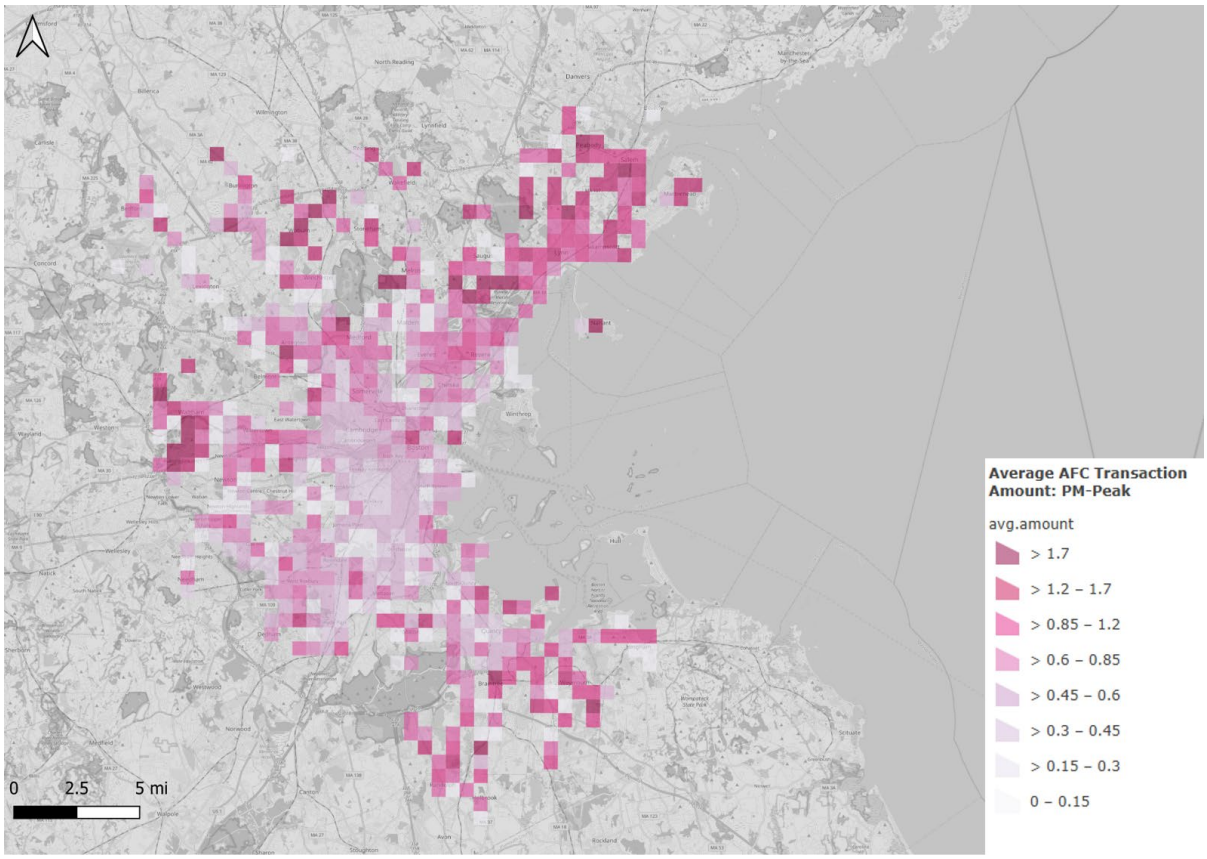

Figure D.14 Average AFC Transaction Amount in PM Peak (4:00 p.m.–6:30 p.m.)

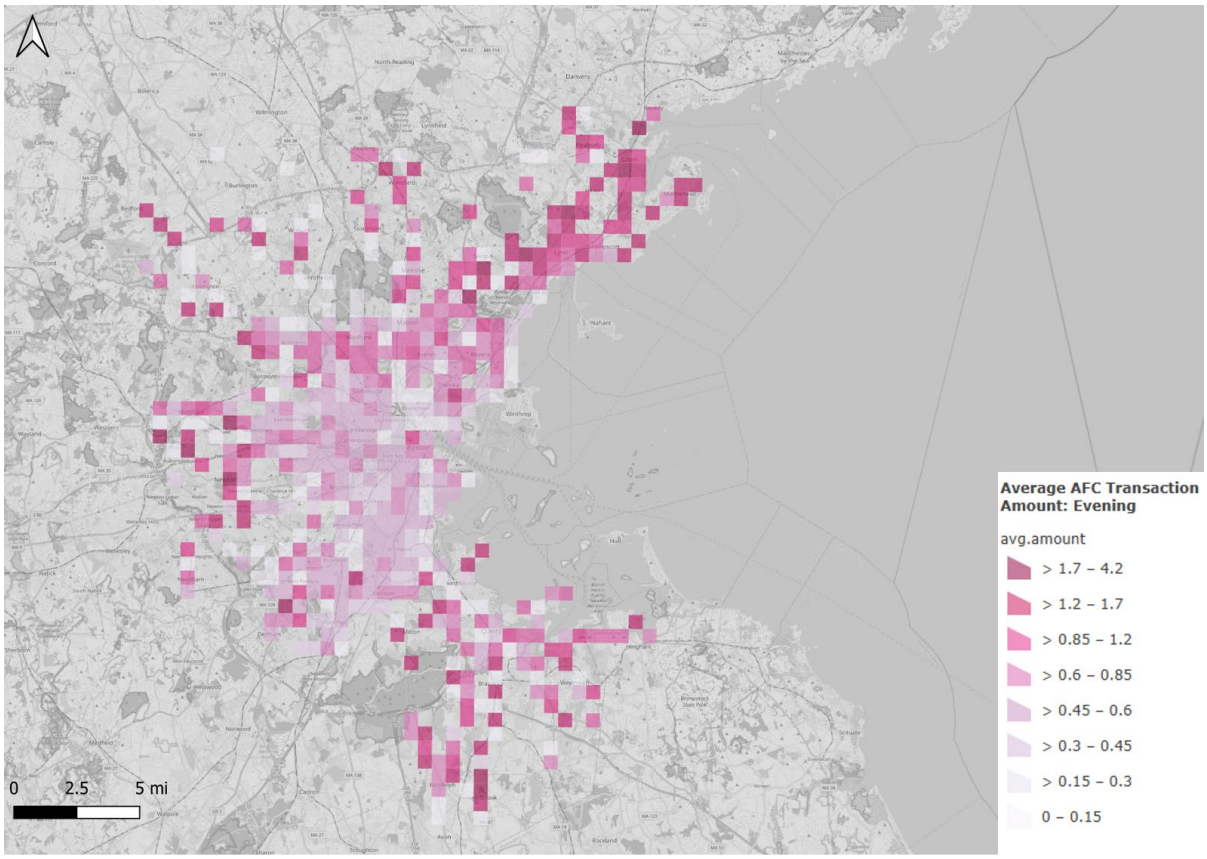

Figure D.15 Average AFC Transaction Amount in Evening (6:30 p.m.–11:59 p.m.)

Appendix E. Maps of Estimated Lost Revenue

Interactive versions of all maps are available online at: https://arcg.is/0XqODi0

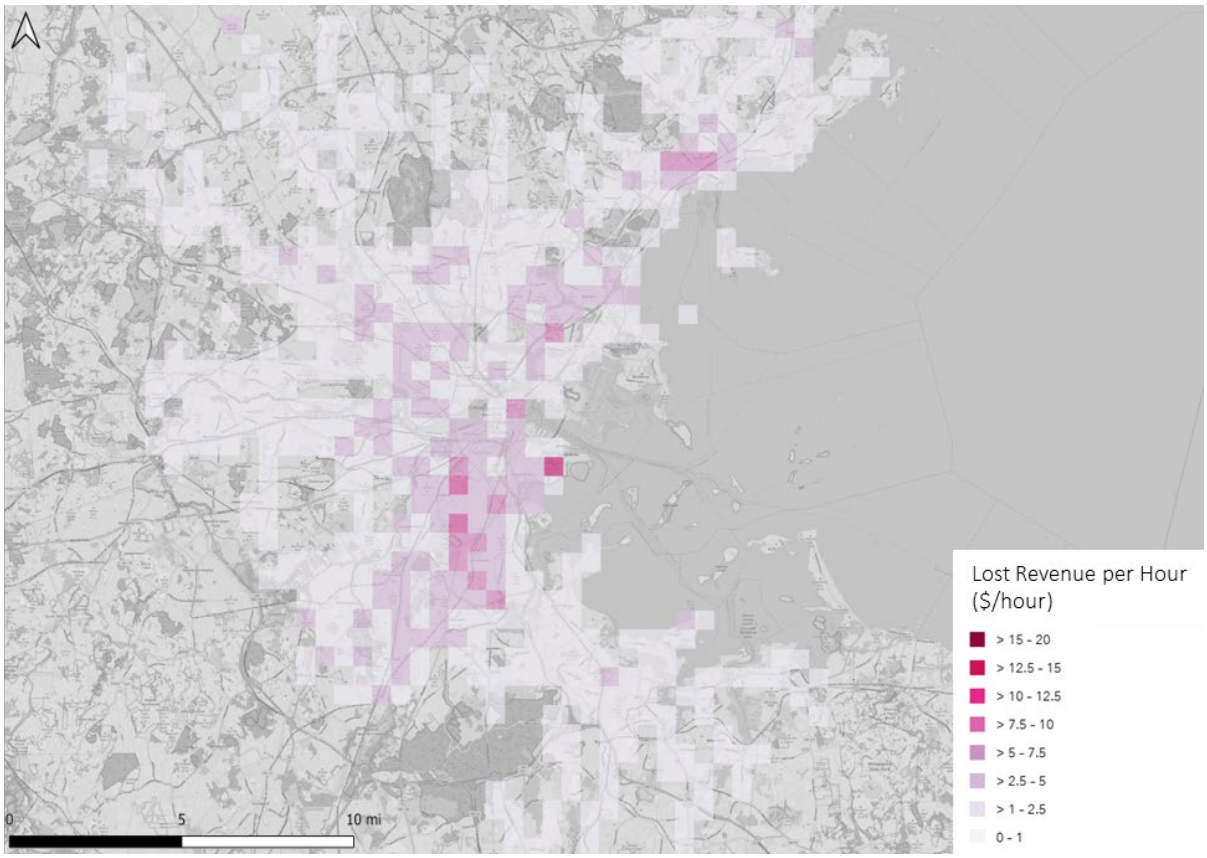

Figure E.16 Estimated Lost Revenue per Hour in AM Peak (5:30 a.m.–9:00 a.m.)

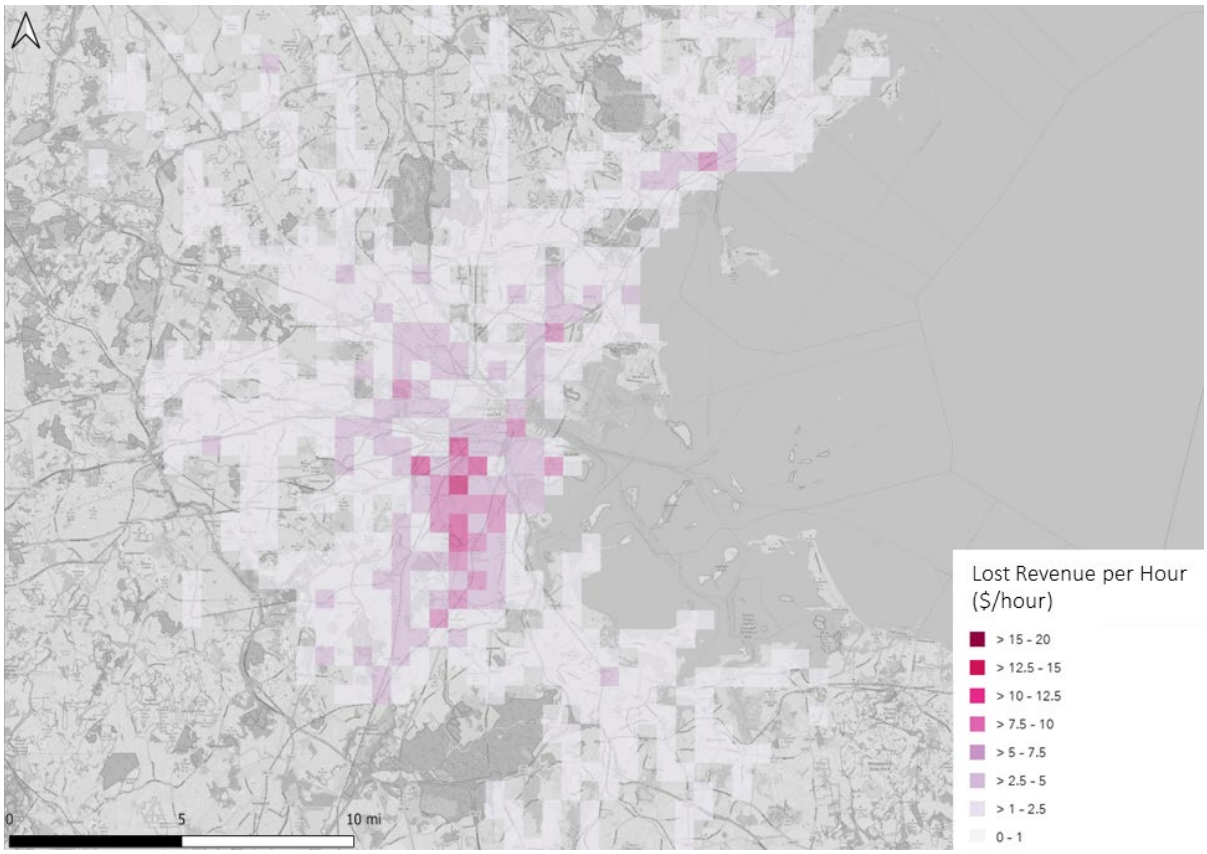

Figure E.17 Estimated Lost Revenue per Hour in Midday (9:00 a.m.–1:30 p.m.)

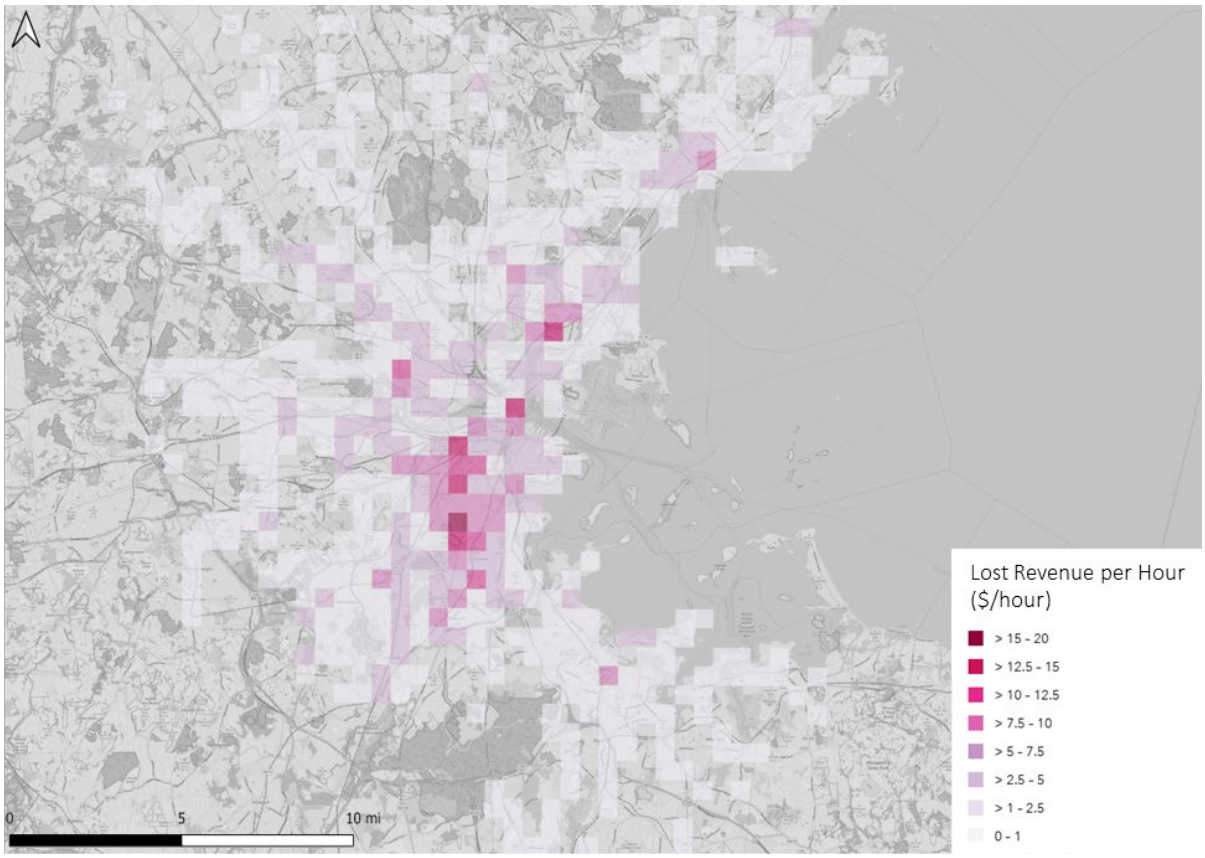

Figure E.18 Estimated Lost Revenue per Hour in Midday School (1:30 p.m.–4:00 p.m.)

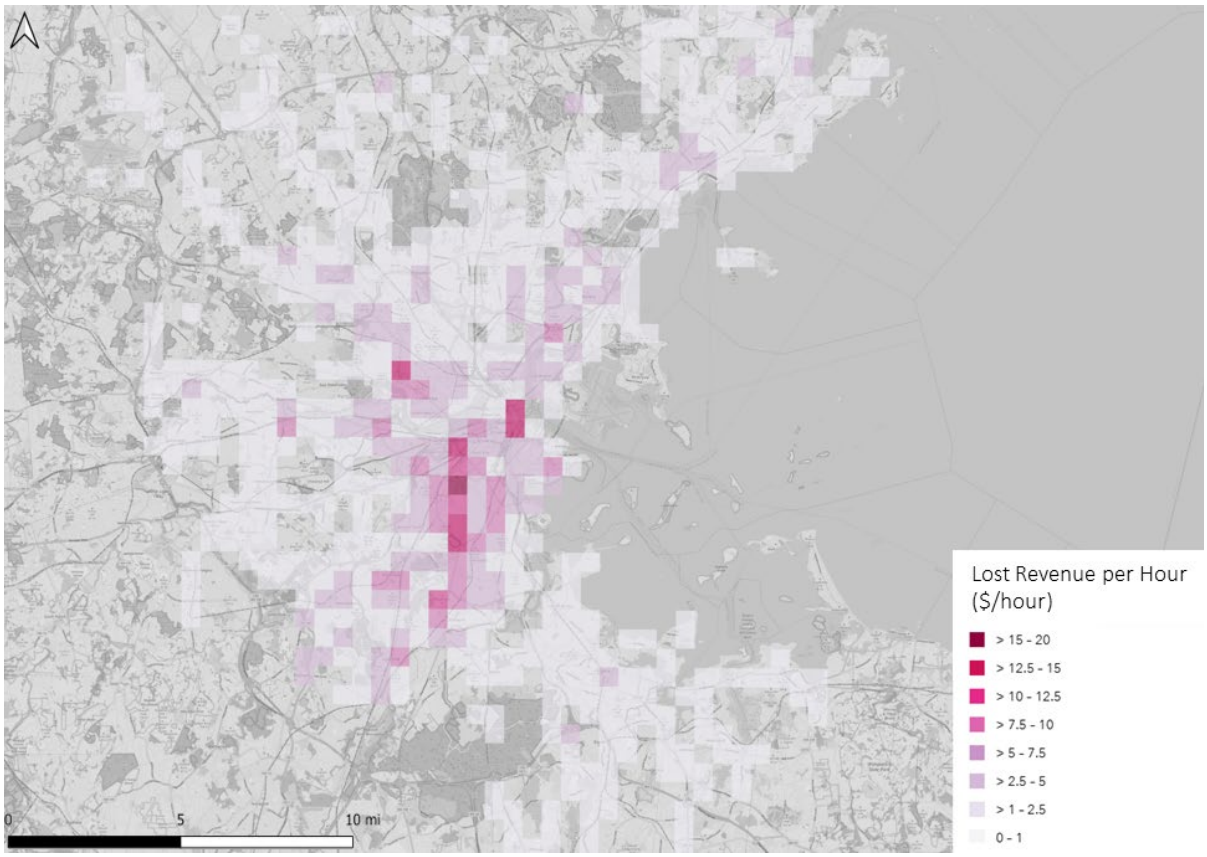

Figure E.19 Estimated Lost Revenue per Hour in PM Peak (4:00 p.m.–6:30 p.m.)

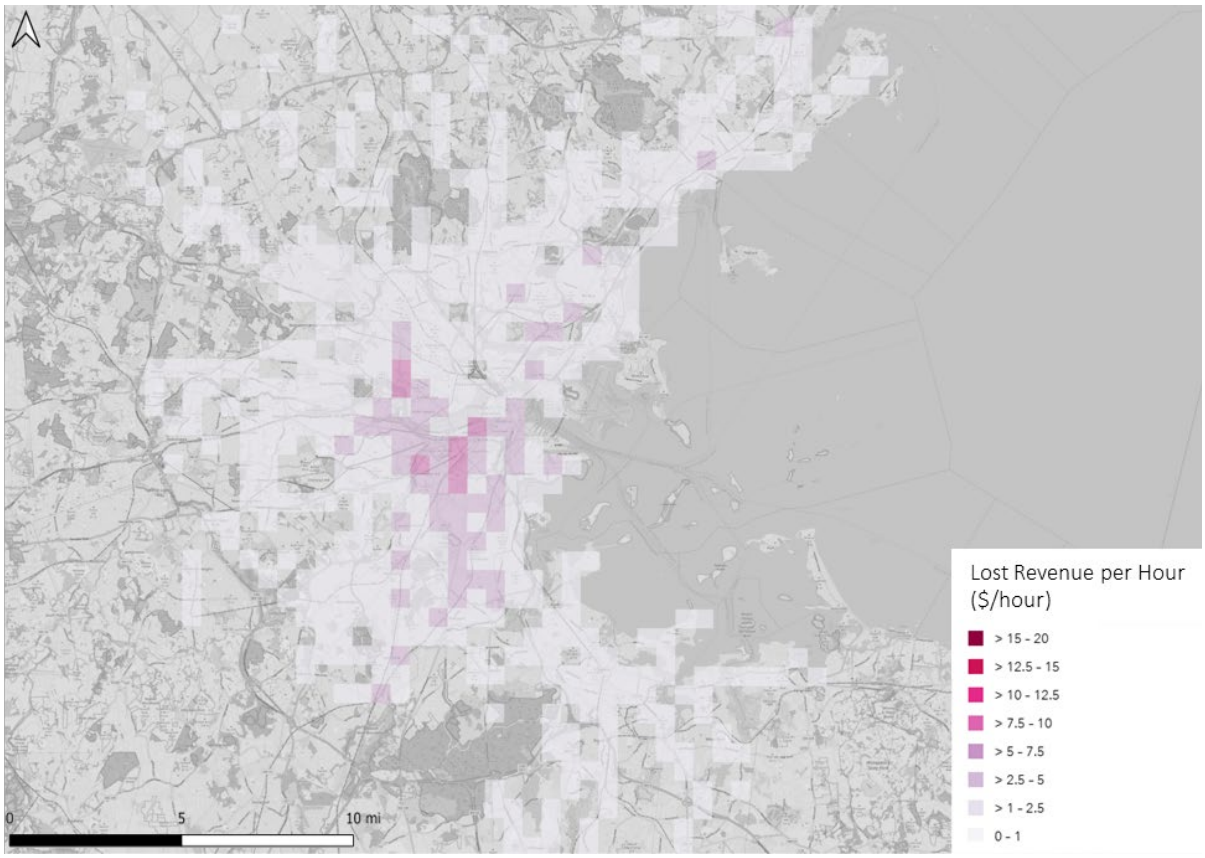

Figure E.20 Estimated Lost Revenue per Hour in Evening (6:30 p.m.–11:59 p.m.)