Characterization of Business Establishments and Commercial Vehicle Movements Utilizing Machine Learning Techniques in Halifax, Canada

by

Niaz Mahmud

Submitted in partial fulfilment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
March 2024

Dedicated to

My Father Md Abdul Mazed

&

My Mother Arifa Nargis

Table of Contents

List of Tables

List of Figures

Abstract

This thesis develops a commercial vehicle travel demand forecasting model. First, it characterizes business establishments to investigate the agglomeration phenomenon among them. Next, it develops location choice models for five major business types (Industry, Retail, Service, Wholesale, and Transportation and Warehousing). In addition, the thesis develops commercial vehicle trip generation models, considering the agglomeration phenomenon of businesses. These models are then utilized to develop a commercial vehicle travel demand forecasting model for Halifax Regional Municipality (HRM). Furthermore, a shopping destination choice model for activity models is developed, which will be implemented within an integrated transport, land use, and energy (iTLE) modeling system to improve the behavioral representation of destination choices. The thesis extensively utilizes the business establishment dataset for HRM obtained from a third-party vendor. The uniqueness of this study lies in its novel approach to extracting agglomeration insights to develop commercial vehicle trip generation models. Furthermore, the study employs machine learning models to generate systematic choice sets for the location choice of businesses. The findings of this study offer valuable insights for commercial vehicle movements, integrated urban system modeling, business location strategies, and policymaking concerning economic development and the growth of the urban built environment.

List of Abbreviations Used

BE	Business Establishment
CBD	Central Business District
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
FTG	Freight Trip Generation
GHG	Greenhouse Gas
HPA	Halifax Port Authority
HRM	Halifax Regional Municipality
IIA	Independence of Irrelevant Alternatives
iTLE	integrated Transport, Land Use, and Energy
ML	Machine Learning
MMNL	Mixed Multinomial Logit Model
NAICS	North American Industry Classification System
QRFM	Quick Response Freight Manual
SIC	Standard Industrial Classification
TAZ	Traffic Analysis Zone

Acknowledgements

First and foremost, all praises and thanks be to Almighty Allah, the architect of the universe. I am sincerely grateful to the Almighty for blessing me with the strength and guidance to complete this thesis.

I am deeply thankful to my supervisor, Professor Dr. Ahsan Habib. His patience, mentorship, and understanding over the past two years have been invaluable. Professor Habib's support, both academically and personally, has been a cornerstone of this journey. His insightful suggestions and dedication have significantly contributed to the development of this thesis.

I am sincerely grateful to Professor Dr. Nouman Ali and Professor Dr. M. Ali Ülkü for their invaluable contributions as members of the supervisory committee. Their insightful suggestions and recommendations have significantly enhanced the quality of the thesis.

I am also grateful to Anik Vai/Sir, my undergraduate university senior, teacher, and current team member at DalTRAC. His support and mentorship have been invaluable in shaping my academic and professional path over the past two years. His guidance has been a guiding light, and I am indebted to him for his unwavering support.

Additionally, I wish to acknowledge my team members at DalTRAC for their collaboration and company. It has been a privilege to work alongside them, and I eagerly anticipate future opportunities for continued collaboration. Their diverse perspectives and expertise have enriched my research experience.

Furthermore, I am grateful to my friends and family for their unwavering support and belief in my abilities. Their encouragement has been a constant motivation, leaving a lasting impact on my academic pursuits. Their presence has made this journey all the more meaningful, and for that, I am sincerely thankful.

Lastly, I would like to express my gratitude to the funding agencies including Environment and Climate Change Canada — Climate Action and Awareness Fund, that has supported my research endeavors. Their financial assistance has enabled me to pursue this academic endeavor with dedication and focus.

Chapter 1

Introduction

1.1 Background and Motivation

Urban transportation, a cornerstone of modern cities, refers to the system of moving people and goods within a city or metropolitan area. It involves the use of different modes of transportation, including buses, subways, trams, bicycles, and pedestrian pathways. The primary aim of urban transportation is to establish effective and accessible connectivity across urban regions, facilitating the efficient movement of residents and goods. The smooth and efficient movement of goods and services is essential for a thriving economy, as it ensures that businesses can operate efficiently.

Business establishments play a crucial role in driving economic activity and generating a substantial number of personal and commercial trips during peak morning and evening hours. The sprawling business settlement patterns have led to an increase in the use of single-occupant vehicles for employees and the delivery of goods to residents. These trips, including light, medium and heavy-duty trucks, contribute to traffic congestion and emissions, including GHGs (greenhouse gases). The increased passenger vehicles are posing significant public health issues due to heightened emissions. The transportation sector accounted for approximately 25% of GHG emissions in 2019, making it the second-largest contributor in Canada (Environment Canada, 2021). In Nova Scotia, GHG emissions from the transportation sector have reached 32.3% in 2020 as shown in Figure 1.1 (Department of Environment and Climate Change, 2022). Concerns have escalated due to the increasing trend in commercial vehicle emissions observed over recent decades. Notably, freight transportation contributes around 42% of total transportation emissions (Environment Canada, 2021). Projections suggest that by 2030, emissions from freight transportation will exceed those from passenger vehicles (Ewing et al., 2020). Canada has pledged to decrease GHG emissions by 40-45 percent from 2005 levels by 2030, with the overarching objective of attaining net-zero emissions by 2050, reflecting a commitment to sustainability (Government of Canada, 2022). Consequently, freight transportation has the potential to play a crucial role in mitigating climate change. Despite extensive research on passenger travel demand, commercial vehicles have not received comparable attention in travel demand forecasting modeling (Grenzeback et al., 2000). The volume and frequency of commercial vehicle movements, as well as their emissions, are heavily influenced by the location of businesses. Therefore, a better understanding of the factors involved in selecting business locations is crucial for developing comprehensive commercial vehicle demand forecasting model.

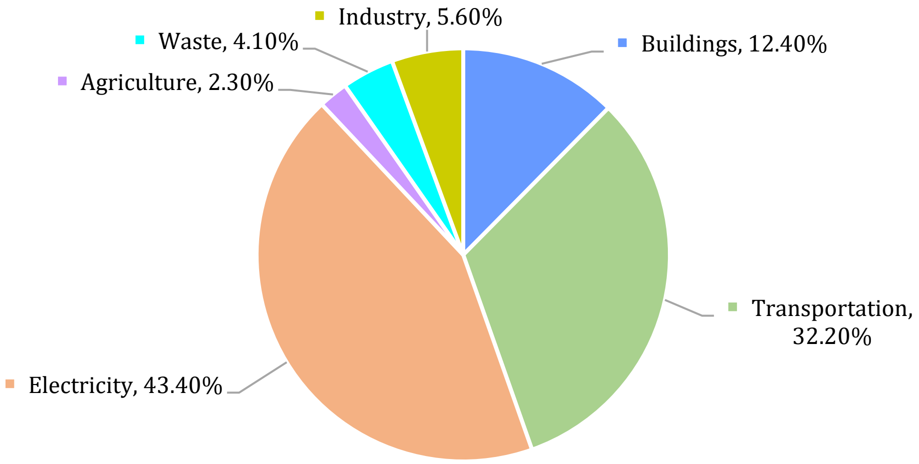

Figure 1.1 Nova Scotia's Greenhouse Gas Emissions, 2020

1.2 Research Goal

The overarching research goal is to advance the commercial vehicle travel demand forecasting modeling approaches for integrated urban systems models. This requires developing a comprehensive business location choice model to precisely model commercial vehicle movements. Additionally, it is crucial to incorporate the agglomeration phenomenon among business establishments into the commercial vehicle trip generation model to accurately discern the pattern. Furthermore, an improved shopping destination choice model is essential to enhance activity-based travel demand modeling. The aim is to capture, by predicting vehicle movements, the complex interactions and dynamics within urban areas. Ultimately, the research seeks to facilitate informed decision-making in transportation planning, and urban development, thereby advancing sustainable growth and enhancing the overall efficiency of urban systems.

1.3 Specific Objectives

The specific technical objectives of this thesis are outlined as follows:

1) To characterize business establishments and explore location choices of business establishments in the Halifax Regional Municipality (HRM).
2) To develop shopping destination choices for advancing integrated urban models.
3) To develop a commercial vehicle demand forecasting model, taking into account agglomeration of business establishments.

1.4 Relevance to Sustainable Development Goals (SDGs)

Based on the specific objectives outlined, several Sustainable Development Goals (SDGs) are pertinent. SDG 8, focusing on Decent Work and Economic Growth, aims to promote sustained, inclusive, and sustainable economic growth, along with full and productive employment opportunities for all. By examining the location choices of business establishments within the Halifax Regional Municipality, insights can be gained to create employment opportunities and foster economic growth. SDG 11, which centers on Sustainable Cities and Communities, seeks to make urban areas inclusive, safe, resilient, and sustainable. The development of shopping destination choices aligns with this goal by enhancing urban environments, promoting accessibility, reducing congestion, and ultimately improving the overall quality of life for residents. Furthermore, SDG 9, emphasizing Industry, Innovation, and Infrastructure, underscores the importance of resilient infrastructure, inclusive industrialization, and innovation. The development of a commercial vehicle demand forecasting model contributes to this goal by enhancing transportation infrastructure and logistics efficiency, thereby supporting economic development and fostering innovation. Lastly, SDG 13, addressing Climate Action, highlights the need to mitigate climate change impacts. The application of a commercial vehicle demand forecasting model can optimize transportation routes, thereby reducing emissions and contributing to efforts aimed at mitigating climate change.

1.5 Conceptual Framework

A conceptual framework is developed to achieve the objectives of this thesis (see Figure 1.2). It consists of two fundamental components of an integrated urban model: businesses and households. The framework elucidates the interactions among business establishments, households, and the transportation network, with a particular emphasis on commercial vehicle movements. It starts with the characterization of business establishments and households. The characterization of businesses involves discerning attributes such as size, type, economic sector, and agglomeration patterns, while household characterization involves understanding demographics, socioeconomic status, car ownership, and consumption patterns. After characterizing businesses, the next step is to develop location choice models for five major business establishment types (Industry, Transportation & Warehousing, Service, Wholesale, and Retail). The location choice of businesses shapes the patterns of goods movements. Moreover, retail location attracts individuals and households for shopping, resulting in significant trips. Consequently, a shopping destination choice model for households is required to predict the likelihood of choosing a particular destination. The destination choice of shopping activity for households is influenced by factors such as distance, accessibility, attractiveness, and the presence of amenities or services. The developed shopping destination model will be incorporated into an activity-based travel demand modeling framework. Central to the framework is the commercial vehicle demand forecasting model, encompassing trip generation and trip distribution. Trip generation focuses on identifying and quantifying the number of trips generated by business establishments. Factors considered include the size of the establishments, the volume of goods produced or consumed, economic activities, and industry-specific characteristics (Bastida and Holguín-Veras, 2009; Cambridge Systematics Inc., 1996; Iding et al., 2002; Lawson et al., 2012). The goal of trip generation is to estimate the total number of commercial vehicle trips originating from different zones or locations within the study area. In trip distribution, the generated commercial vehicle trips from the business establishments are distributed spatially across the transportation network. Factors influencing truck trip distribution include transportation infrastructure, access to markets, customer locations, supply chain networks, and logistical considerations (Holguín-Veras et al., 2012). The final stage, network assignment, encompasses all modes of transportation, involves assigning the trips to specific transportation networks. This step helps optimize the routing of commercial vehicles to minimize transportation costs, travel time, and environmental impacts while ensuring efficient goods movement and distribution. The outcome is a detailed understanding of how traffic flows through the transportation network, facilitating infrastructure planning and policy decision-making.

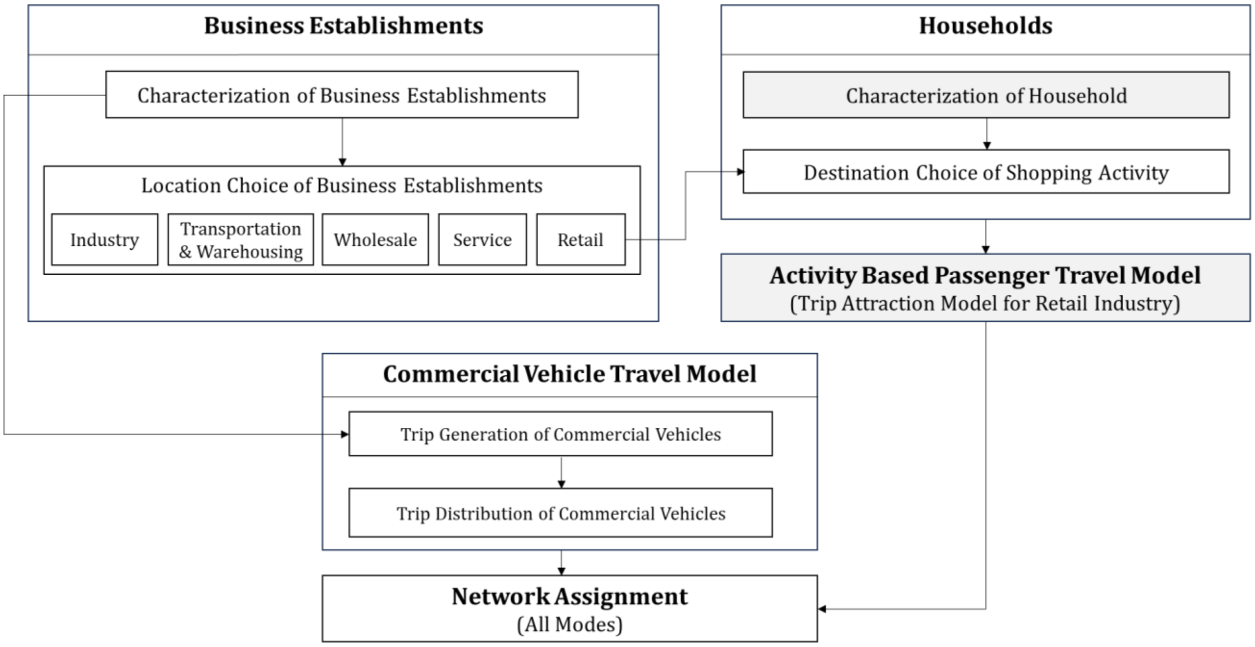
Business Establishments
Characterization of Business Establishments
Location Choice of Business Establishments
Industry
Transportation & Warehousing
Wholesale
Service
Retail
Households
Characterization of Household
Destination Choice of Shopping Activity
Activity Based Passenger Travel Model
(Trip Attraction Model for Retail Industry)
Commercial Vehicle Travel Model
Trip Generation of Commercial Vehicles
Trip Distribution of Commercial Vehicles
Network Assignment
(All Modes)

Figure 1.2 Conceptual Framework

1.6 Research Significance

The first contribution of this research lies in demonstrating an innovative two-stage framework for examining patterns of clustering and agglomeration among business establishments at the micro-level within an economic system. Additionally, the research presents comprehensive business establishment location choice models that utilize machine learning in systematic choice sets. This study also develops a systemic methodology for investigating how the concentration of business establishments affects commercial vehicle trip generation model. Moreover, it develops a comprehensive shopping destination choice model that will be integrated into an integrated transport, land use, and energy (iTLE) modeling system to enhance the representation of activity models and improve the activity-based travel demand model. The findings of this study offer valuable insights for commercial vehicle movements, business location strategies, integrated transport modeling, and policymaking concerning economic development and the growth of the urban built environment.

1.7 Organization of the Thesis

The thesis consists of six chapters and is organized as follows: Chapter 1 introduces the background, motivation, objectives of the thesis, and the conceptual framework. Chapter 2 characterizes and assesses the agglomeration of business establishments using a combination of machine learning and spatial statistics. Next, Chapter 3 presents a comprehensive location choice model for business establishments, leveraging machine learning in systematic choice sets. Chapter 4 outlines a comprehensive framework for a shopping destination choice model, combining machine learning and discrete choice modeling. Following this, Chapter 5 develops a commercial vehicle travel demand model, with a specific focus on truck trip generation while considering the agglomeration phenomenon. Finally, Chapter 6 provides conclusions, summarizes the overall contributions, and suggests potential avenues for future research.

Chapter 2

Characterization of Business Establishments[1]

2.1 Introduction

Business establishments are a critical element of urban spatial structure, shaping the distribution and function of commercial activities within a city. Urban spatial structure refers to how cities are physically organized, including the layout of buildings, roads, and green spaces, which affects how people move, interact, and live within the urban environments (Anas et al., 1998; O'sullivan, 2018). In recent decades, urban spatial patterns have experienced significant changes (Dadashpoor and Malekzadeh, 2022; Heider and Siedentop, 2020). These transformations include shifts in land use, infrastructure development, and the spatial distribution of activities within cities (Pfister et al., 2000; Coffey and Shearmur, 2002; Shearmur and Alvergne, 2003; Lee, 2007; Riguelle et al., 2007; Veneri, 2018; Zhang et al., 2019; Kwon, 2021). In particular, the spatial distribution and dynamic evolution of business establishments hold pivotal importance in urban planning. Understanding these patterns is essential for developing effective transportation policies and infrastructure solutions that support efficient, safe, and sustainable freight transportation while fostering economic growth.

[1] This chapter is adapted from:

Mahmud, N. and Habib, M. A. (2024). Characterization and Assessment of Agglomeration of Businesses Establishments: A Combination of Machine Learning and Spatial Statistics Approach. Presented at the 103rd Annual Meeting of the Transportation Research Board (TRB), Washington, D. C., January 7-11, 2024. https://annualmeeting.mytrb.org/OnlineProgram/Details/21127

The spatial evolution of urban structure, transitioning from monocentric (concentrated around one central point) to polycentric (spread across multiple centers) has attracted significant interest (Wang and Wei, 2007; Gu et al., 2009; Sun and Wei, 2014). Hu et al. (2018) explored the relationship between changing urban spatial structure and commuting behaviors in Beijing, China. In contrast, there is limited research that has investigated spatial structure from the angle of employment distribution (Liu et al., 2011; Sun et al., 2012). Although businesses are integral part of the urban structure, very few studies have explicitly focused on the characterization of businesses. Characterization of businesses refers to the process of describing and analyzing various attributes, characteristics, and spatial patterns of businesses within a particular geographic area. In summary, there is a dearth of studies focusing on assessing the spatial distribution of businesses within a medium-sized municipality, with little exploration into business characterization. This study aims to fill this gap in the existing literature by incorporating analyses of the built environment and neighborhood attributes to delineate the spatial arrangement of businesses within a medium-sized municipality.

Therefore, the objectives of this study are as follows: 1) to characterize the spatial distribution of businesses based on the built environment and neighborhood attributes, and 2) to assess the agglomeration of businesses in a medium-sized municipality. By leveraging this knowledge, cities can develop transportation policies and infrastructure solutions that support efficient, safe, and sustainable freight transportation while fostering economic growth and vitality. An unsupervised machine learning (ML) technique is utilized to characterize the business establishments considering built environment and neighborhood attributes. Following that, kernel density (a statistical method which estimates the probability density function of a random variable) estimation and factor analysis have been performed to analyze the agglomeration patterns of businesses. Finally, spatial statistics are implemented for assessing the spatial clustering of businesses. This study utilizes a comprehensive business establishment dataset to extract the exact geographic location, built environment, and neighborhood attributes.

The remainder of this chapter is structured as follows: Section two provides a synthesized literature review on the characterization of business establishments. Section three introduces the study area and relevant data. Subsequently, the framework of the study is outlined in section four. Section five elaborates on the methodology employed in the study. Following this, the subsequent section reveals the findings of this study, accompanied by a comprehensive discussion. Finally, the chapter concludes by summarizing key findings and proposing directions for future research.

2.2 Literature Review

In recent decades, urban spatial structures have undergone substantial transformations due to increased reliance on automobiles and substantial investments in road infrastructure. This shift has resulted in decentralization, as populations and job opportunities have migrated from urban cores to suburban areas. Consequently, suburban centers have emerged, diminishing the significance of traditional Central Business Districts (CBDs), and contributing to further dispersion of urban activities. The intensity of urbanization relies heavily on significantly how urban populations, economic activities, and land use aggregate.

The built environment plays a crucial role in shaping the distribution of economic activities. The influence of the built environment on the spatial distribution of urban structure is evident through the impact of the city center. It acts as a focal point around which different activities cluster, influencing the layout and density of urban development. Consequently, it indicates the central concentration of urban activities (An et al., 2019). Recent research highlights the significant influence of both primary city centers and secondary sub-centers on the spatial distribution of service, production, and manufacturing industries within urban areas (Yang et al., 2012; Li et al., 2015; Lagonigro et al., 2020). These centers serve as magnets for economic activities, attracting businesses and shaping the geographic distribution of industries. Moreover, city centers have a notable influence on population density within urban areas (Clark, 1951; Newling, 1969; Smeed, 1964). They tend to attract a higher concentration of residents due to factors such as accessibility to amenities, employment opportunities, and cultural attractions. As a result, population density often peaks in and around city centers, gradually decreasing as one moves away from these central hubs. However, recent studies have confirmed that proximity to the nearest subcenter positively influences population distribution (Huang et al., 2017). This implies that areas closer to subcenters tend to experience higher population densities, as they benefit from the accessibility and amenities offered by these secondary hubs.

The relationship between the distribution of built environment and transit accessibility has been extensively investigated. Research indicates that ring roads and rail transit systems play vital roles in facilitating people's movement, enhancing employment opportunities around subway stations, and significantly impacting the establishment of nearby businesses (Zhao et al., 2020; Tu et al., 2019; Tan et al., 2019; Zeng et al., 2020). These transportation infrastructures not only provide efficient connectivity but also act as catalysts for economic development by attracting businesses and stimulating commercial activities in their vicinity. Furthermore, studies have revealed that the existence of a ring road has a substantial influence on the spatial distribution of service, production, and manufacturing industries (Li et al., 2015). This infrastructure often acts as a critical transportation artery, facilitating goods and service movement throughout the urban area. Consequently, industries tend to cluster around ring roads due to improved accessibility and logistical advantages, shaping the geographic distribution of economic activities within the city. Mixed land uses have been shown to enhance activity levels within neighborhoods (Jacobs, 1961; Jacobs-Crisioni et al., 2014). This creates vibrant and dynamic urban environments where residents have easier access to amenities, services, and employment opportunities.

Despite extensive research on the impact of the built environment on urban development and economic activities, there is a noticeable absence of studies specifically focusing on the spatial distribution of businesses within medium-sized municipalities. Furthermore, limited exploration has been conducted into characterizing these businesses across various dimensions. Investigating both the distribution and characteristics of businesses in medium-sized municipalities could provide valuable insights into urban economic dynamics and inform policies aimed at fostering economic growth and spatial equity in these areas. The significance of this study lies in its comprehensive analysis of the spatial dynamics of business establishments within a medium-sized municipality. By examining multidimensional factors, the study aims to uncover the underlying drivers of business establishment distribution. To address multidimensionality, the study employs a machine learning-based clustering technique. In this study, a ML-based clustering technique is employed to address the multidimensionality. Additionally, kernel density and multi-distance spatial cluster analysis are utilized to analyze the spatial distribution patterns of businesses. These methods enable the validation and exploration of spatial clustering patterns, contributing to the holistic characterization of businesses within urban landscapes.

2.3 Study Area and Data

The area of interest for this research is the Halifax Regional Municipality (HRM), the capital city of Nova Scotia, Canada (Figure 2.1). This region covers an estimated area of 5,577 square kilometers. This study utilizes a large dataset of business establishments for the year 2022 obtained from a third-party vendor (Data Axle, previously known as Info Canada). Table 2.1 summarizes information about the business establishments in the study area. This reliable and comprehensive dataset contains detailed records of firms with 8-digit NAICS (North American Industry Classification System) code for HRM. The data includes valuable information about each establishment's name, type, geographic location, total number of employees, sales volume, year of establishment, credit rating, and expenses. Figure 2.2 illustrates the distribution of all business establishments throughout the study area. Most of the business establishments are located in urban and suburban areas, where the density of businesses is comparatively lower in rural areas. In total, there are twenty different types of establishments. Businesses are categorized into five broad industry types - industry, retail, service, wholesale, transportation & warehousing - based on their NAICS code to group them more effectively. Table 2.2 shows the categorization. The locations of businesses categorized by annual sales volume are depicted in Figure 2.3 through Figure 2.5.

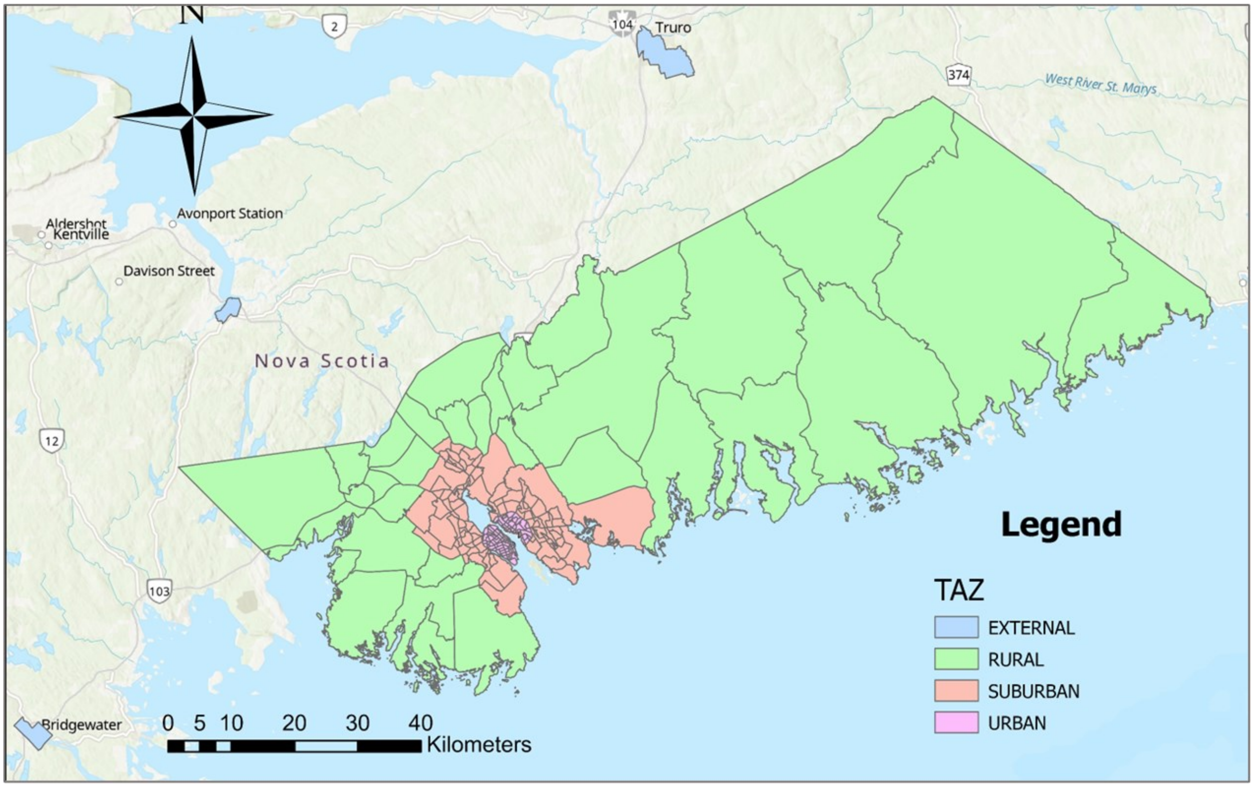

Figure 2.1 Study Area – Halifax Regional Municipality (HRM)

For geospatial analysis, the exact longitude and latitude coordinates of each business were used to geocode them in ArcGIS Pro. Additionally, the 219 traffic analysis zones (TAZ) in the Halifax Transport Network Model (Bela, 2018) were spatially joined with the establishments. This process allowed the extraction of zonal population density, zonal employment number, and zonal entropy (diversity of business establishments), essential built environment attributes. The 2021 Canadian Census provided the population density and employment numbers.

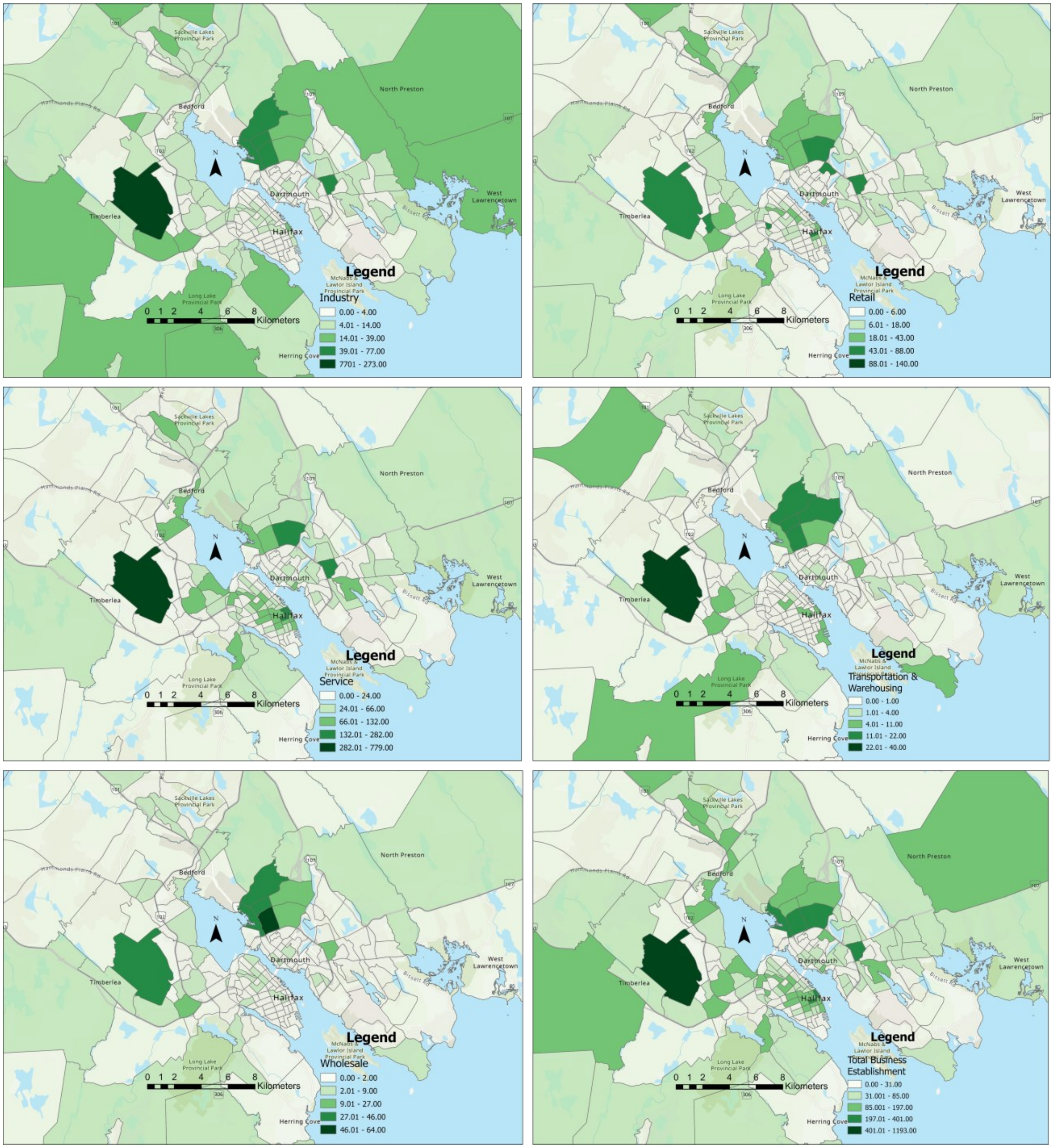

Figure 2.2 Distribution of Business Establishments in Each Traffic Analysis Zone

Sales volumes by establishment types aid in the characterization of businesses by providing essential insights into economic performance, market dynamics, and spatial distribution. Figure 2.3 displays the geographic distribution of industry and retail establishments categorized by annual sales volume. It indicates that the highest annual sales volume is observed in Bayers Lake Business Park, followed by Burnside Industrial Park.

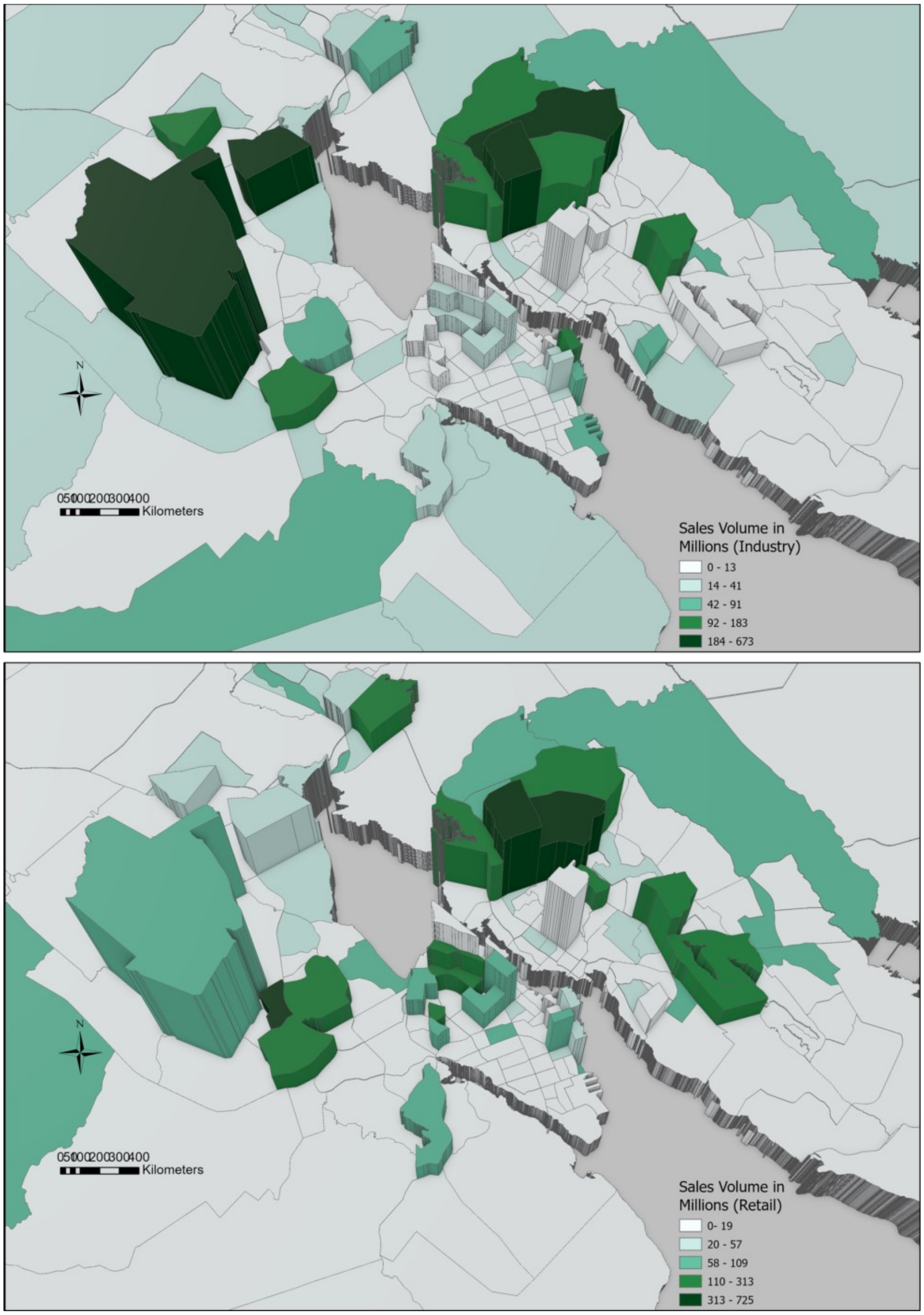

Figure 2.3 Location of Industry and Retail by Annual Sales Volume

For retail establishments, the highest sales volume is observed in Burnside Industrial Park. However, the presence of several supermarkets in the urban areas is prominent.

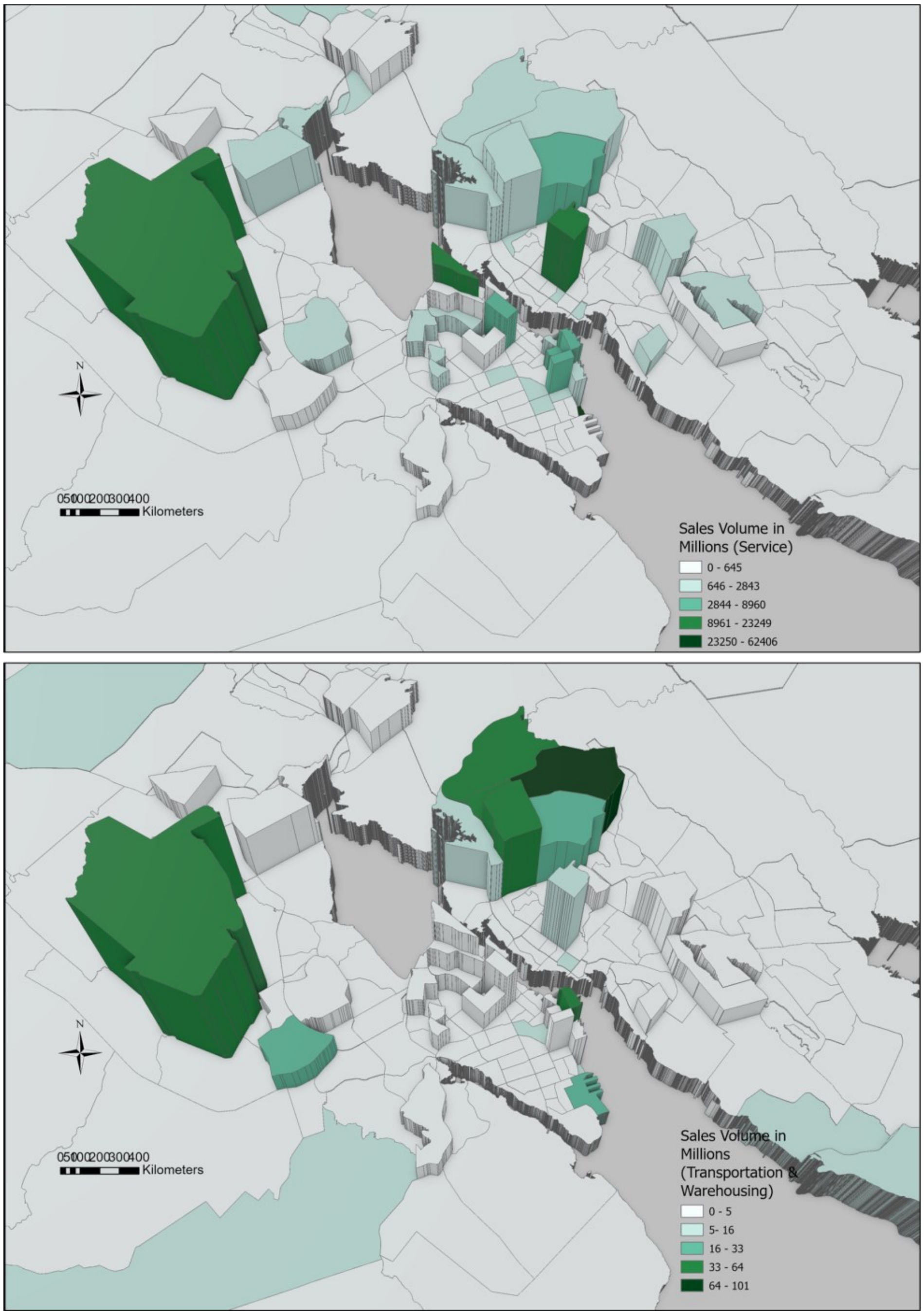

Figure 2.4 Location of Service and Transportation & Warehousing by Annual Sales Volume

The annual sales volume of transportation & warehousing is below 5 million for most of the TAZs in the urban core, suggesting a low density of transportation & warehousing activities in this area of Halifax. However, the annual sales volume of service establishments in a few TAZs of the urban core ranges between 2844 million and 8960 million (Figure 2.4).

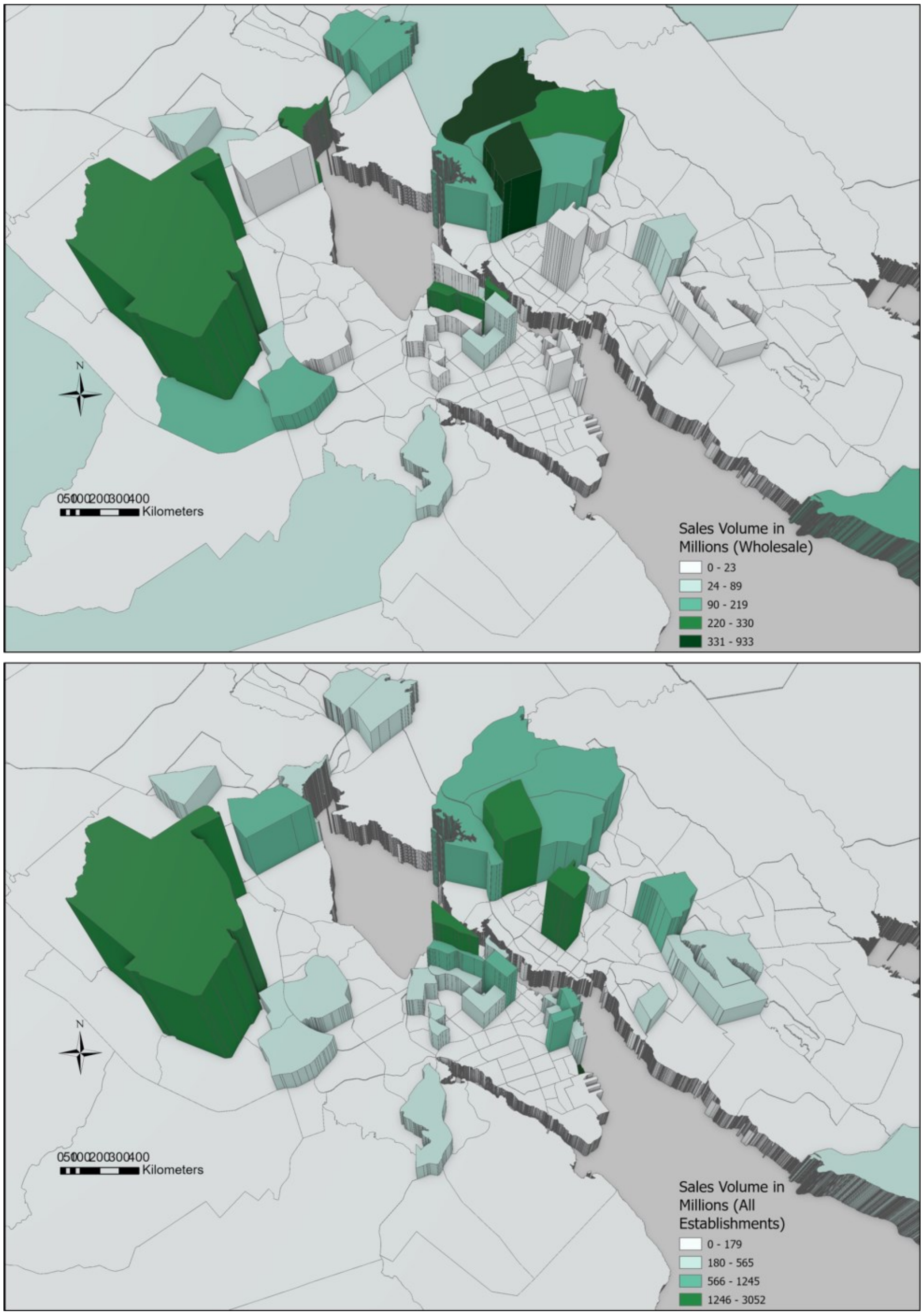

Figure 2.5 Location of Wholesale and All Establishment by Annual Sales Volume

Overall, the spatial distribution of all business establishments by annual sales volume is significant in the Bayers Lake Business Park and Burnside Industrial Park (see Figure 2.5).

Table 2.1 Descriptive Statistics of Info Canada Dataset

Establishment Types	**Count**
Accommodation And Food Services	7.48%
Administrative And Support, Waste Management and Remediation Services	4.11%
Agriculture, Forestry, Fishing, And Hunting	0.16%
Arts, Entertainment, And Recreation	2.48%
Construction	9.43%
Educational Services	3.29%
Finance & Insurance	5.50%
Health Care and Social Assistance	9.31%
Information & Cultural Industries	1.85%
Management Of Companies & Enterprises	0.04%
Manufacturing	3.32%
Mining, Quarrying, And Oil and Gas Extraction	0.21%
Other Services (Except Public Administration)	12.98%
Professional, Scientific and Technical Services	9.40%
Public Administration	2.99%
Real State and Rental and Leasing	5.11%
Retail Trade	13.84%
Transportation & Warehousing	2.56%
Utilities	0.04%
Wholesale Trade	4.66%

Table 2.2 Classification of Five Establishments (parenthesis numbers indicate 2-digit NAICS code)

Establishment Types	**Elements**
Industry	Manufacturing (31-33) Construction (23) Agriculture, Forestry, Fishing and Hunting (11) Mining and Oil and Gas Extraction (21)
Service	Information and Cultural Industries (51) Professional, Scientific and Technical Services (54) Administrative and Support, Waste Management, and Remediation Services (56) Finance and Insurance (52) Arts, Entertainment and Recreation (71) Real Estate and Rental and Leasing (53) Accommodation and Food Services (72) Management of Companies and Enterprises (55) Utilities (22) Health Care and Social Assistance (62) Public Administration (92) Educational Services (61) Other Services (except Public Administration) (81)
Retail	Retail Trade (44, 45)
Transportation and Warehousing	Transportation and Warehousing (48,49)
Wholesale	Wholesale Trade (41)

2.4 Framework of Characterization of Businesses

A framework is developed to characterize the businesses. Figure 2.6 shows the framework of the methodology. It begins with data collection and preparation, where spatial and non-spatial data on businesses, including attributes such as industry type, location, built environments and neighborhood attributes, are gathered and preprocessed to ensure consistency and compatibility across datasets. Following this, K Prototype clustering is applied to the business dataset to identify distinct clusters of businesses based on multidimensional factors. By grouping businesses with similar characteristics into clusters, considering both numerical and categorical variables, this step provides a foundational understanding of the diversity and complexity of businesses. Subsequently, kernel density estimation is conducted to visualize the spatial distribution of businesses across the study area. Generating a continuous density surface representing the intensity of business concentration within the urban landscape allows for the identification of areas of high and low business concentration. Multi-Distance Spatial Cluster Analysis using Ripley's K-function is then performed to assess the spatial clustering patterns of businesses. By calculating the K-function for different distances, the degree and scale of spatial clustering across the study area are quantified, providing insights into the spatial processes driving business agglomeration. Finally, findings from K Prototype clustering, kernel density estimation, and Ripley's K-function analysis are integrated and interpreted in the context of existing theories and literature on urban economics, spatial clustering, and business location decisions.

This study hypothesized that businesses prioritize specific built environments and neighborhood attributes in the characterization process. For example, transportation and warehousing, real estate and rental and leasing, manufacturing, construction, and public administration tend to cluster in a low population density zone. This phenomenon is likely influenced by factors such as the cost of land, accessibility to transportation routes, zoning regulations, space availability for expansion, noise and environmental concerns, and the benefits of specialization and agglomeration economies. Selected relevant attributes (population density, employment numbers, entropy) are included in the unsupervised machine learning algorithm. In the process of business establishment clustering, the selection of specific zonal attributes such as population density, employment number, and entropy can be justified based on their relevance and significance.

Population Density: A crucial factor as it indicates the concentration of potential customers within a particular area. Higher population density implies a larger market size and increased demand for goods and services, making it an essential consideration for identifying areas with a substantial customer base.

Employment Number: It provides insights into the local job market and economic activity within a particular area. Areas with a high number of employees typically indicate a robust local economy, which also suggests a stable customer base and higher purchasing power. Including zonal employment numbers in the clustering analysis helps identify regions with thriving business environments, allowing businesses to capitalize on the existing economic vitality and potential partnerships.

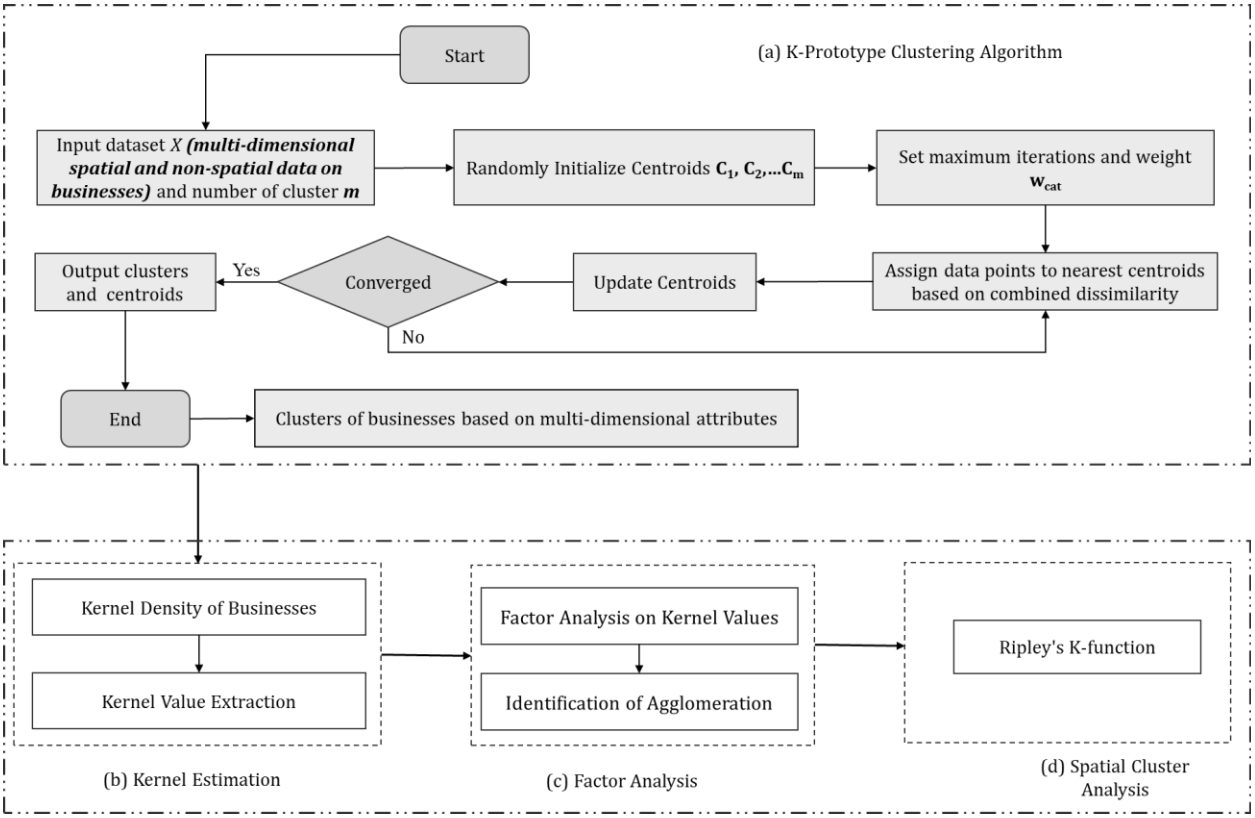
Start
(a) K-Prototype Clustering Algorithm
Input dataset X (multi-dimensional spatial and non-spatial data on businesses) and number of cluster m
Randomly Initialize Centroids $C_1, C_2, \ldots C_m$
Set maximum iterations and weight w_{cat}
Output clusters and centroids
Yes
Converged
No
Update Centroids
Assign data points to nearest centroids based on combined dissimilarity
End
Clusters of businesses based on multi-dimensional attributes
Kernel Density of Businesses
Kernel Value Extraction
Factor Analysis on Kernel Values
Identification of Agglomeration
Ripley's K-function
(b) Kernel Estimation
(c) Factor Analysis
(d) Spatial Cluster Analysis

Figure 2.6 Framework for Characterizing Business Establishment

Entropy: It plays a significant role in capturing the diversity of businesses within a given zone. By assessing the variety of establishments in an area, it becomes possible to identify regions with a vibrant and competitive commercial environment. High zonal entropy suggests diverse consumer preferences, enabling businesses to cater to a broader customer base or identify niche markets that align with specific offerings. The following formula calculates zonal entropy:

$$H(i) = -\sum_{k=1}^{K} (\Pr(k) * \log(\Pr(k))) \qquad (2.1)$$

where, *H(i)*: Zonal entropy value for zone *i*

$\Pr(k)$: Proportion of establishments in zone *i* belonging to industry type *k*

K: Total number of industry types

By considering these three attributes together, the clustering analysis has the potential to provide a comprehensive understanding of the economic landscape and the propensity of commercial vehicle movements in the region. This holistic approach aids decision-making processes related to business establishment, enabling stakeholders to identify locations based on market size, competition level, and economic conditions.

2.5 Methodology

2.5.1 K-Prototype Clustering Algorithm

K-Prototype is a machine learning-based clustering approach uniquely designed to handle numerical and categorical datasets. Derived from the k-means clustering technique, it aims to minimize the within-cluster sum of dissimilarities by using distance metrics like Euclidean distance for numerical data points and appropriate measures for categorical data (Huang, 1998). By considering the diverse characteristics of the dataset, K-Prototype effectively forms clusters that are internally cohesive and distinct from one another. This robust hybrid algorithm extends the applicability of traditional k-means clustering to real-world datasets with mixed data types, providing an invaluable tool for data exploration and pattern discovery in various domains.

Let X be the dataset with n data points and p attributes per data point(x_i), comprising both categorical and numerical variables.

Let $C = \{C_1, C_2, ..., C_k\}$ be the set of k clusters, where each cluster C_j contains a subset of data points.

The dissimilarity measure between a numerical data point x_i and a cluster centroid μ_j can be calculated using the Euclidean distance:

$$d_{num}(x_i, \mu_j) = \sqrt{\sum_{p=1}^{P}(x_{ip} - \mu_{jp})^2} \tag{2.2}$$

where x_{ip} and μ_{jp} are the p-th attributes of x_i and μ_j, respectively and P denotes the total number of attributes (both numerical and categorical).

The dissimilarity measure between a categorical data point x_i and a cluster centroid μ_j can be calculated using a distance metric appropriate for categorical data, such as the Hamming distance:

$$d_{cat}(x_i, \mu_j) = \sum_{p=1}^{P}(x_{ip} \neq \mu_{jp}) \tag{2.3}$$

where x_{ip} and μ_{jp} are the *p-th* attributes of x_i and μ_j, respectively.

The overall dissimilarity between a data point x_i and a cluster centroid μ_j can be calculated as a weighted sum of the numerical and categorical dissimilarities:

$$d(x_i, \mu_j) = (1 - \gamma) * d_{num}(x_i, \mu_j) + \gamma * d_{cat}(x_i, \mu_j) \tag{2.4}$$

where γ is a weighting factor between 0 and 1 that determines the relative importance of numerical and categorical variables.

2.5.2 Optimum Number of Clusters

The optimum number of clusters is identified based on the Bayesian Information Criterion (BIC) measure (Kim and Seo, 2014) and Silhouette Score (SS). The sudden change in BIC value indicates the optimum number of clusters. The Elbow method plots the cost (inertia) of clustering as a function of the number of clusters. Inertia measures the sum of distances between points and their assigned cluster centroids. The plot looks like an arm, and the "elbow point" is the optimal number of clusters where adding more clusters doesn't significantly decrease inertia (Marutho et al., 2018). A higher Silhouette Score is desirable as it signifies well-defined and separated clusters. A score closer to 0 or negative values may indicate the presence of overlapping or poorly separated clusters, implying that the clustering might need improvement.

2.5.3 Kernel Density Estimation

This study utilized point pattern analysis through kernel density estimations to visualize the spatial distribution of business establishments within each cluster. This study estimated kernel on a grid surface with cells measuring 1.0 km by 1.0 km. It used a radius of 0.5 km to estimate the spatial density of business establishments across the 20 types of business establishments. The kernel values, representing the expected number or intensity of firms per grid cell, effectively generate density maps for each sector. Subsequently, the analysis aimed to explore the potential existence of agglomeration of business establishments via factor analysis.

2.5.4 Factor Analysis

Factor analysis is a statistical technique employed to reduce a set of observed variables (in this context, kernel values for each establishment within a cluster) into smaller latent factors. The primary objective of factor analysis is to identify k orthogonal latent factors, wherein each factor corresponds to a set of highly correlated firm intensity variables (kernel values). In simpler terms, it endeavors to uncover underlying patterns or themes that elucidate the common variations observed in the dataset. This study theorized that if such agglomeration among the businesses is present, the resultant column vectors, portraying the kernel values for each type of establishment, would demonstrate a significant correlation, indicating the agglomeration of certain businesses. A similar methodological approach was found in the literature, where employment co-location was analyzed (Maoh and Kanaroglou, 2007; Shearmur and Coffey, 2002). Factor analysis is applied to the 20-column vectors of each establishment within the determined cluster to assess these correlations and identify potential groups of correlated factors. By implementing factor analysis to the kernel values for different industrial sectors, this study attempts to unveil patterns of agglomeration among businesses within a cluster, symbolizing the built environment and neighborhood characteristics.

2.5.5 Multi-Distance Spatial Cluster Analysis (Ripley's *K*-function)

Multi-Distance Spatial Cluster Analysis, often referred to as Ripley's *K*-function, is a method used in spatial statistics to analyze the spatial distribution pattern of a set of points within a defined region. The *K*-function measures the degree of spatial clustering or dispersion of the points relative to a random distribution. It provides information about the spatial interaction and pattern of point locations across different distances. A typical transformation of the *K*-function, often referred to as $L(d)$, is defined as:

$$L(d) = \sqrt{\frac{A \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} k_{i,j}}{\pi n(n-1)}} \tag{2.5}$$

where, d = distance; n = total number of features; A = total area of the features; $k_{i,j}$ = weight. If there is no edge correction, then the weight will be equal to one when the distance between i and j is less than d and will equate to zero otherwise.

The *K*-function can be calculated for various distances, and the resulting curve (*K(d) vs. d*) can be compared to the expected value for a random distribution. If the observed *K(d)* value is higher than expected, it indicates clustering at that distance. If the *K(d)* value is lower than expected, it suggests a regular or dispersed pattern. The *K*-function is often used in combination with its simulation envelope to test the statistical significance of clustering patterns. The simulation envelope is generated by randomly distributing points within the study region multiple times, calculating the K-function for each simulation, and then obtaining upper and lower confidence bounds.

2.6. Results and Discussions

2.6.1 Cluster Analysis

Two evaluation metrics: the Bayesian Information Criterion (BIC) value and the Silhouette Score, are used to find the optimum number of clusters of 20 types of business establishments based on the built environment and neighborhood characteristics. The BIC is a statistical criterion that helps in model selection by balancing the model's fit to the data and its complexity. The Silhouette Score measures the quality of clustering by calculating how similar an object is to its own cluster compared to others.

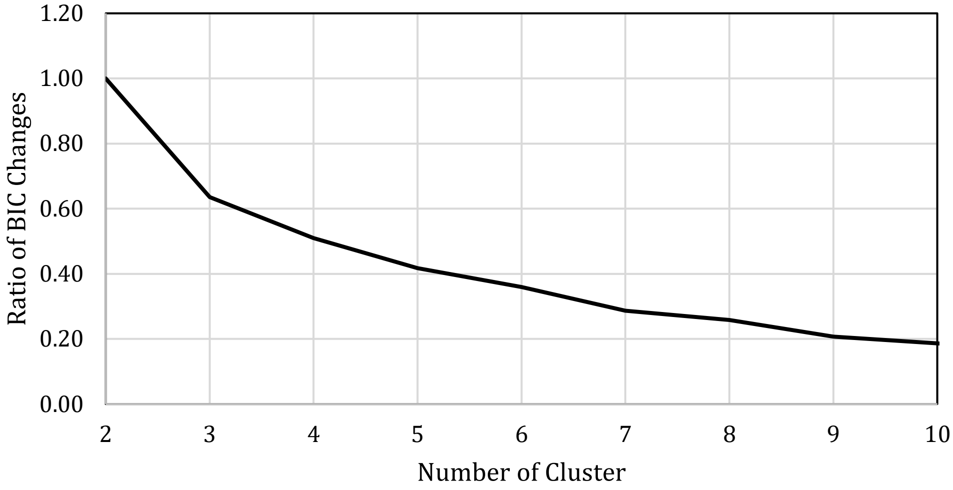

Figure 2.7 Ratio of BIC changes with Respect to Number of Clusters

Figure 2.7 shows the BIC value plotted against the number of clusters, and there is a noticeable abrupt change in the ratio of BIC value when the cluster number reaches 3. This abrupt change in the BIC value is often referred to as the "elbow point" in the plot. It suggests that the ideal number of clusters is located at the point where the BIC value or another appropriate metric shows the most significant drop (the "elbow") before reaching a plateau. In this case, the plot indicates that the BIC value sharply decreases when moving from 1 to 3 clusters but doesn't show as significant a change beyond 3 clusters. Based on this observation, the optimum number of clusters was set to 3.

However, it is essential to consider both the BIC value and the Silhouette Score to ensure the clustering solution is appropriate. The Silhouette Score helps to verify the cohesion and separation of clusters and ensures that the clustering is meaningful and well-defined. Figure 2.8 demonstrates the initial assumption of 3 clusters as the optimal number of clusters is satisfactory.

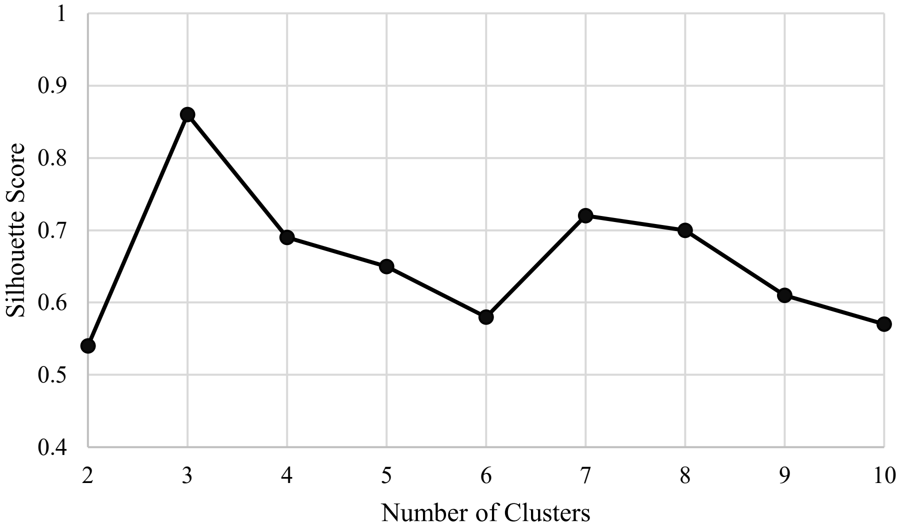

Figure 2.8 Silhouette Score with Respect to Number of Clusters

Therefore, separating the 20 types of business establishments into three distinct groups provides a meaningful and interpretable clustering solution based on the data and chosen evaluation metrics. The clustering with three clusters likely captures the underlying patterns and variations in the data effectively, making it a suitable choice for further analysis and decision-making related to these business establishments (Figure 2.9).

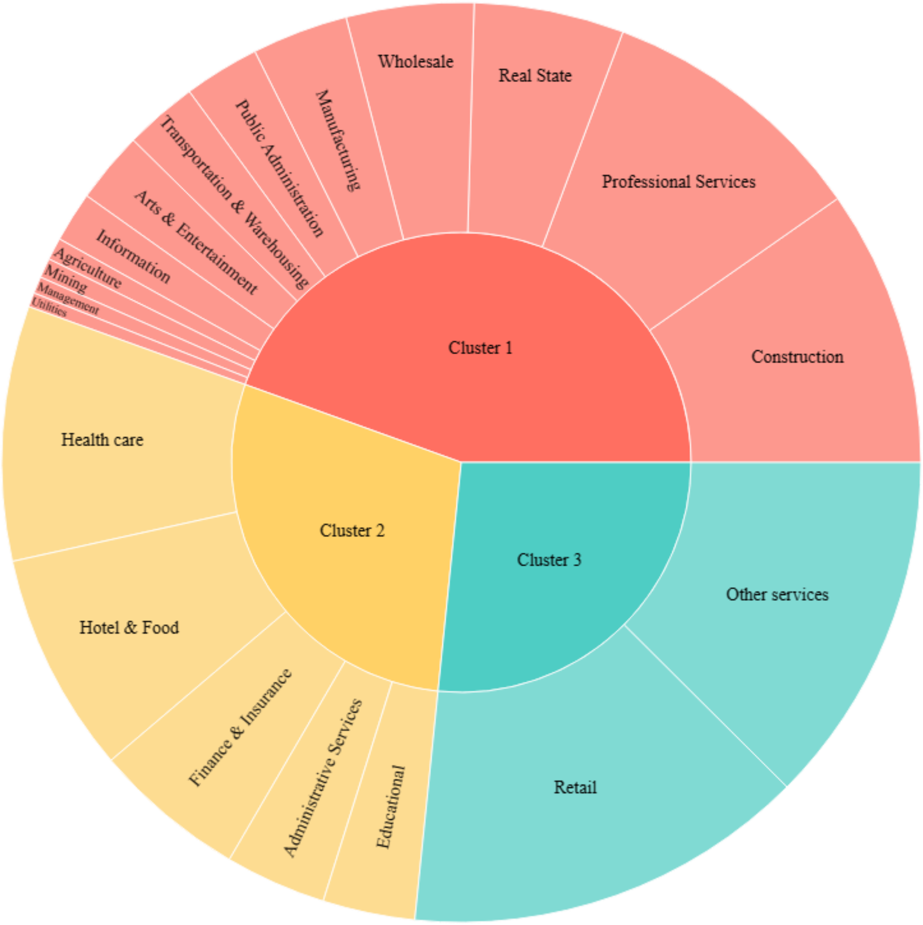

Figure 2.9 Distribution of Business Establishments within a Cluster

Cluster 1 exhibits noteworthy characteristics, namely the highest zonal entropy (Figure 2.10), indicating significant spatial heterogeneity, juxtaposed with the lowest population density. Despite its sparse population, Cluster 1 showcases the highest employment numbers and a diverse mix of business types spanning all major sectors. This combination of features suggests that Cluster 1 functions as a regional economic hub with a broad economic base and spatial diversity, offering diverse opportunities for various industries. The inclusion of resource-based and manufacturing sectors within Cluster 1 implies a focus on activities related to natural resources or production, serving a broader region. Concurrently, the presence of business types centered on innovation, such as professional services and information industries, suggests a progression beyond basic production activities towards higher value-added economic endeavors.

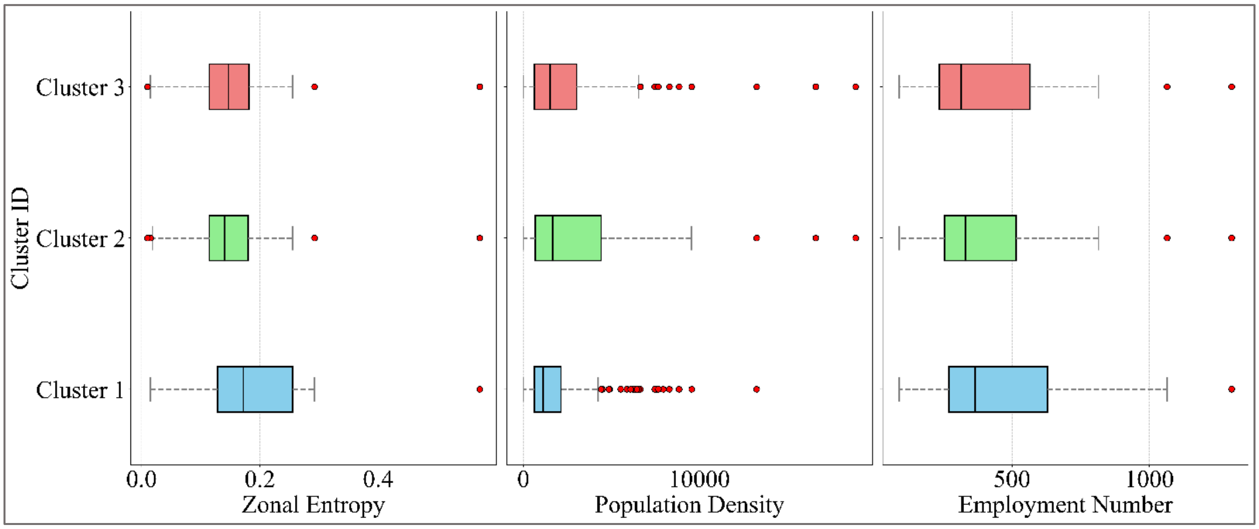

Figure 2.10 Boxplot for Each Attribute within a Cluster

The lower population density in Cluster 1 indicates the availability of ample land, which can facilitate the establishment of essential infrastructure and facilities, supporting a range of economic activities catering to a larger geographical area. This factor may also enable the development of industrial facilities, transportation infrastructure, and logistics critical for export-oriented production. The higher employment rate relative to density points towards a potential productivity advantage, likely attributable to concentrated economic activity and production chains within specific sectors. Such productivity gains may arise from the agglomeration economies and supply chain linkages within clustered industries, leading to enhanced efficiency and competitiveness. The presence of resource-based and manufacturing sectors in Cluster 1 sees it as an export base for the wider region, generating external income that subsequently circulates back into the local economy through job creation, supply chain interactions, and economic multipliers. However, the coexistence of innovation-driven industries signals a diversification process beyond traditional resource extraction and production, attracting and retaining skilled workers, thereby supporting its evolving economic base.

Cluster 2 exhibits distinctive characteristics that distinguish it from others. Notably, it is characterized by the lowest zonal entropy, signifying a spatially concentrated urban form commonly found in densely populated urban cores where economic activities are centered. The high population density in Cluster 2 indicates its classification as an urban cluster predominantly driven by the demands of a dense population. The cluster's moderate employment number concerning its population density suggests a specialization in high-productivity service industries as opposed to lower-paying tourism and retail sectors. This specialization in service-oriented businesses, trade, finance, and tourism indicates that Cluster 2 primarily serves the needs of residents and businesses within its boundaries. The observed low zonal entropy and higher land values in Cluster 2 possibly indicate limitation in spatial heterogeneity and mixed land uses. Additionally, the concentration of service and retail businesses in this cluster may result in lower wages and productivity due to the proximity-driven nature of such industries.

Cluster 3 exhibits moderate zonal entropy, population density, and employment opportunities, with a prevailing focus on retail and other service-oriented businesses. These findings suggest that Cluster 3 likely represents either a medium-sized urban cluster or a suburban area with a primary emphasis on meeting the needs of local residents through retail and service industries. This cluster can be interpreted as a regional economic hub, an urban cluster with many service industries, and a medium-sized urban/suburban cluster with a primary orientation towards serving local demands. The presence of retail and local service businesses in Cluster 3 indicates its primary role in catering to the daily requirements of its residential population. The moderate population density and employment levels further indicate a mixed land use pattern, likely comprising a combination of residential and commercial activities. The observed moderate zonal entropy values, which suggest a concentration of activity despite some mixed land uses, are consistent with the typical characteristics of medium-sized urban areas.

2.6.2 Factor Analysis on Kernel Density

The findings show a strong correlation among the business establishment types (Public Administration, Professional, Scientific and Technical Services, Information and Cultural Industries, Arts, Entertainment and Recreation, Management of Companies and Enterprises) based on their high Factor 1 loading (Table 2.3). Factor 1 likely represents a common underlying factor or characteristic that these industries share, leading them to cluster together geographically. It means that areas or regions with a high concentration of one of these industries are also likely to have a high concentration of the others. A possible interpretation of this finding could be that a specific economic or socio-cultural environment in certain regions collectively fosters the growth of these industries. For example, a city or area with a strong focus on research and technology (Professional, Scientific, and Technical Services) may also attract businesses in related fields such as Information and Cultural Industries and Arts, Entertainment, and Recreation that cater to the needs of a tech-savvy population.

The "Real Estate and Rental and Leasing" and "Utilities" industries are both closely connected to the development and management of physical infrastructure. Regions experiencing significant real estate development need expanded utility services such as electricity, water supply, and waste management, leading to the clustering of these industries.

Table 2.3 Factor Analysis on Kernel Density of 20 Types of Establishments

	Establishment Types	**Factor 1**	**Factor 2**	**Factor 3**
Cluster 1	Public Administration	**0.888**	0.229	0.163
	Professional, Scientific and Technical Services	**0.687**	0.515	0.318
	Information and Cultural Industries	**0.611**	0.332	0.242
	Arts, Entertainment and Recreation	**0.407**	0.374	0.359
	Management of Companies and Enterprises	**0.394**	0.306	0.159
	Real Estate and Rental and Leasing	0.307	**0.515**	0.241
	Utilities	0.248	**0.494**	0.306
	Wholesale Trade	0.153	0.075	**0.734**
	Manufacturing	0.192	0.166	**0.698**
	Construction	0.130	0.258	**0.657**
	Transportation and Warehousing	0.270	0.235	**0.519**
	Mining and Oil and Gas Extraction	0.202	0.184	**0.400**
	Agriculture, Forestry, Fishing and Hunting	0.017	0.132	**0.386**
Cluster 2	Administrative and Support, Waste Management, and Remediation Services	**0.734**	0.281	0.329
	Accommodation and Food Services	**0.684**	0.538	0.175
	Health Care and Social Assistance	0.158	**0.762**	0.168
	Finance and Insurance	0.384	**0.710**	0.130
	Educational Services	0.421	**0.454**	0.287
Cluster 3	Other Services (Except Public Administration)	0.463	**0.644**	0.385
	Retail Trade	0.402	**0.535**	0.287

The factor loading analysis reveals a significant correlation or agglomeration among the following industries: Wholesale Trade, Manufacturing, Construction, Transportation and Warehousing, Mining and Oil and Gas Extraction, and Agriculture, Forestry, Fishing, and Hunting. The high factor loadings indicate a tendency for these industries to cluster together in specific geographic regions. Several factors may contribute to this clustering phenomenon. Firstly, supply chain and complementary relationships exist among these industries, where entities in one sector rely on the services and products of others for smooth operations. For instance, manufacturing companies depend on the wholesale trade sector for effective distribution, and transportation and warehousing play a vital role in linking these industries. Secondly, proximity to abundant natural resources may attract businesses in mining, agriculture, and related sectors, leading to spatial concentration.

The "Administrative and Support, Waste Management, and Remediation Services" sector provides essential support services to businesses, such as waste disposal, cleaning, security, and administrative functions. Consequently, regions with a thriving business community, particularly those with a concentration of hotels, restaurants, and other hospitality establishments, would generate a higher demand for these support services, leading to the clustering effect.

The factor loadings from the conducted analysis indicate a notable agglomeration among the Health Care and Social Assistance, Finance and Insurance, and Educational Services industries. For example, a region with a strong presence of educational services may also attract professionals in the healthcare and financial sectors due to the demand for their services from students, educators, and school staff.

Based on the factor loadings provided, there appears to be a significant agglomeration or clustering among the “Other Services (Except Public Administration)” and “Retail Trade” industries. For example, areas with a high concentration of retail outlets may see an increased demand for various services to support these businesses, such as repair and maintenance, entertainment, and personal care services.

2.6.3 Multi-Distance Spatial Cluster Analysis (Ripley's K-function)

The K-functions for each establishment have been estimated using ArcGIS Pro. As illustrated by Ripley's K-function (Figure 2.11), the observed K(r) of all establishments is higher than the upper confidence envelope in the range of 25-30 km. Figure 2.11 indicates that sectors tend to be clustered within 30 km of each other and not randomly or uniformly distributed. The exact distance at which clustering occurs can provide valuable insights. For example, it may indicate the effective range of attraction or interaction between sectors.

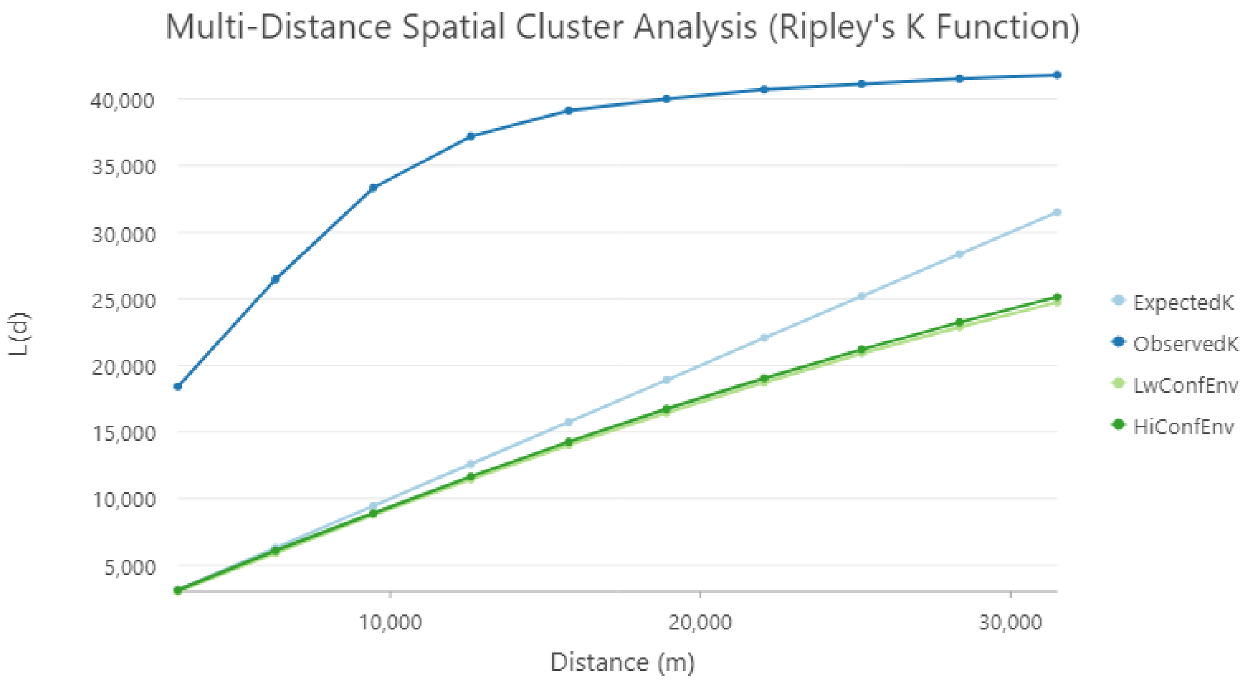

Figure 2.11 Multi-Distance Spatial Cluster Analysis (Ripley's K-function)

2.7 Conclusion

The framework presented in this study to discern characterization of business establishments incorporated multi-dimensional built environment and neighborhood attributes. The application of the unsupervised machine learning technique with the spatial statistics methods allows a data-driven approach to understand the intricate characterization of business establishments. This is deeply influenced by the surrounding built environment and neighborhood factors. This approach can help reveal underlying patterns and connections that might not be visible through traditional geospatial and statistical methods. The micro-level analysis yields rigorous insights into how certain types of businesses cluster and how the surrounding built environment and neighborhood characteristics influence their characterization. The outcomes from this study indicate that wholesale trade, manufacturing, construction, transportation and warehousing, mining and oil and gas extraction, as well as agriculture, forestry, fishing, and hunting, tend to cluster in a specific geographic location characterized by spatial heterogeneity, juxtaposed with the lowest population density and highest employment opportunities. Investigations into analogous clustering tendencies have been discerned across various sectors, encompassing public administration, professional, scientific, and technical services, information and cultural industries, arts, entertainment, and recreation, as well as the management of companies and enterprises while illustrating the phenomenon of agglomeration among these sectors.

However, this study is subject to certain limitations, as it exclusively focuses on a constrained set of the built environment and neighborhood attributes, namely population density, employment number, and zonal entropy, as the influencing factors governing business establishments' clustering and agglomeration. Although these attributes significantly influence business location decisions, they do not encompass the entirety of the factors that potentially shape the spatial distribution of businesses. The built environment constitutes a multifaceted framework containing various elements, including transportation infrastructure, proximity to suppliers and customers, accessibility to amenities, adherence to land use regulations, crime rates, and environmental conditions. These additional factors might be necessary to ensure the ability of this study to offer a comprehensive understanding of business location patterns and the fundamental mechanisms driving agglomeration phenomena. It is imperative to incorporate a broader range of built environment and neighborhood attributes alongside other pertinent factors to overcome this limitation and extend the scope of this study. Including socioeconomic data, such as income levels, education levels, and consumer behavior, might offer an in-depth insight into how business establishments interact with the local population dynamics while depicting the clustering and agglomeration events. By addressing these aspects in future research, the complex dynamics of business location decisions and agglomeration patterns could be captured through robust understanding.

Nevertheless, the proposed framework that reveals the agglomeration patterns of business establishments contribute to the existing literature, by systematically analyzing clustering and spatial distribution process. The findings from this study offer valuable insights for commercial goods movement modeling, business location strategies, commercial vehicle movement modeling, and policymaking related to economic development and urban built environment.

Chapter 3

Location Choice Model of Business Establishments[2]

3.1 Introduction

Business establishment plays a crucial role in spatial development and is a fundamental component of integrated land use and transportation models. As the primary sources of economic activities, business establishments exert a strong influence in drawing and generating a significant number of personal and commercial trips in the morning and evening peaks (Bell, 1991; Armstrong, 1972). Commercial trips involving both light and heavy-duty trucks lead to traffic congestion and significantly contribute to emissions, including greenhouse gases (GHGs). The growing concerns arise from the upward trend in commercial vehicle emissions witnessed over the past few decades (Environment and Climate Change Canada, 2022). However, the existing literature has extensively examined passenger travel demand, yet the realm of travel demand forecasting modeling has not given comparable attention to commercial vehicles (Ziomas et al., 1995; Grenzeback et al., 2000). The volume and frequency of commercial vehicle movements, along with the resulting emissions, are greatly influenced by the location of businesses. Hence, a better

[2] This chapter is adapted from:

Mahmud, N. and Habib, M. A. (2024). A Comprehensive Business Location Choice Model Leveraging Machine Learning in Systematic Choice Set. Presented at the 103rd Annual Meeting of the Transportation Research Board (TRB), Washington, D. C., January 7-11. https://annualmeeting.mytrb.org/OnlineProgram/Details/21127

Mahmud, N. and Habib, M. A. (2024). A Comprehensive Business Location Choice Model Leveraging Machine Learning in Systematic Choice Set. *Transportation Research Record. (Accepted)* understanding of the process involved in choosing business establishment locations is essential to develop a comprehensive freight demand modeling.

The location choice of business establishments is conceptualized as two distinct steps: the search process and the location choice process (Alcácer and Chung, 2014b). During the search process, businesses gather information, evaluate market conditions, and identify potential locations that meet their specific requirements (Balbontin and Hensher, 2019). This search procedure attempts to create a group of alternatives that are feasible and realistic. Following the search process, businesses carefully evaluate the shortlisted options based on predetermined criteria and preferences before making a final decision on the final location (Elgar et al., 2009). However, most existing studies on the location choice of businesses have focused primarily on the second step while neglecting the initial search process (Maoh, 2005). The location search step is imperative for reliable and precise location choice estimations and predictions, and by ignoring this step, the findings might be unable to capture the full complexity of businesses' location decisions.

The business establishment location choice model is primarily implemented utilizing the framework of multinomial logit models (MNL) (Abraham and Hunt, 1999; Khan, 2002). Most earlier studies used random sampling of alternatives to generate the choice set (De Bok and Sanders, 2005a). However, these methodologies do not fully depict the behavioral dynamics of a business establishment's location search process. The inclusion of one or more unreliable alternatives in the choice set is, therefore, quite likely to take place, which will end up resulting in biased parameter estimations for the choice model. Few studies have attempted to create a systematic choice set to address this issue (De Bok and Sanders, 2005b;

Elgar et al., 2009). However, business location choice decision is influenced by various factors, including distance to the central business district (CBD), proximity to transport infrastructure, socioeconomic attributes, and agglomeration factors (Abraham and Hunt, 1999; Maoh and Kanaroglou, 2005). Considering several characteristics to generate a reasonable choice set yields the challenge of building different types of location alternatives, which may be challenging from a methodological and computational perspective. Hence, the primary research question of this study is how to develop a more sophisticated approach to capture the multi-dimensional factors to develop a better location choice model of businesses. Hence, this study advances a novel machine learning (ML)-based clustering method to generate a plausible choice set for business location choice modelling.

The rest of this chapter is organized as follows: In section two, a literature review synthesizing establishment location choice models is presented. Section three introduces the study area and relevant data. Following this, the conceptual framework of the study is outlined in section four. Section five details the methodology employed in the study. The subsequent section unveils the findings of this study, accompanied by a thorough discussion. Finally, the paper concludes by summarizing key findings and suggesting directions for future research.

3.2 Literature Review

Integrated land use and transportation models have primarily focused on residential location choices and associated commuting patterns. However, businesses play a crucial role in generating trips involving commercial vehicle movements. Although the development of large-scale integrated models has grown over the last two decades, commercial vehicles have been underrepresented in travel demand modeling. In recent years, there has been an increase in the number of commercial vehicle movement models (Eisele et al., 2013). Most existing studies have implemented freight demand forecasting modelling utilizing two main methods: 1) commodity-based modeling and 2) trip/tour-based modeling. Commodity-based modeling focuses on analyzing the flow of goods based on their characteristics, origin, destination, and volume, providing insights into infrastructure needs and policy impacts (Fischer et al., 2000). On the other hand, trip/tour-based modeling examines individual vehicle trips and interactions with the transportation network, offering detailed insights into congestion, travel times, and operational decisions (Bela and Habib, 2019). In both commodity-based modeling and trip/tour-based modeling approaches for freight demand modelling, the location of business establishments plays a vital role.

Business location choice modeling is a multidisciplinary field that draws insights and methodologies from various academic disciplines (Kumar and Kockelman, 2008; van Wissen, 2000). The location decisions of businesses can often be categorized into three main factors: accessibility, office profile, and business profile (Balbontin and Hensher, 2021). These factors encompass various aspects that influence where a business chooses to locate its operations (Carlton, 1979; Lee, 1982). Transportation and infrastructure are critical considerations as businesses seek locations with good access to highways, airports, ports, and public transit systems (Bodenmann and Axhausen, 2012; Iseki and Jones, 2018; Weterings and Knoben, 2013). Such proximity minimizes transportation costs and facilitates the efficient movement of goods and personnel. The locational preferences of businesses are influenced by household income, population density, and proximity to the CBD (Maoh and Kanaroglou, 2009; Waddell and Shukld, 1993). High household incomes attract businesses seeking affluent consumers, densely populated areas appeal to those targeting larger customer bases, and proximity to the CBD impacts accessibility and networking opportunities for various firms. Additionally, businesses often favor locations with agglomeration benefits, where the concentration of similar or related industries fosters synergies, knowledge-sharing, and improved access to skilled labor and resources (Hansen, 1987; Gabe and Bell, 2004; De Bok and Van Oort, 2011; Rosenthal and Strange, 2004; Backman and Karlsson, 2017). In the existing literature, a limited number of studies specifically address business establishment location choice from a freight perspective focusing on the movement and transportation of goods and commodities (de Jong and Ben-Akiva, 2007; Wisetjindawat et al., 2007; Pourabdollahi et al., 2012; Roorda et al., 2010; Cavalcante and Roorda, 2013).

Business establishment location choice modeling involves considering a wide range of available alternatives for a business, which can vary from hundreds at the zonal levels to thousands at the parcel levels. The two-step business establishment location choice model that aims to mimic search process, has some difficulties in dealing with a wide range of spatial alternatives. One of the most widely adopted techniques to address the challenges associated with many alternatives is a random sampling of alternatives. Manski (1977) introduced a two-step discrete-choice modeling framework, wherein the first stage involves deriving a subset of choice alternatives from the universal choice set. This initial selection process can be executed through the application of predetermined criteria for choice set selection or through a random selection approach (Ben-Akiva and Lerman, 1985; Swait and Ben-Akiva, 1987). McFadden (1978) demonstrated that a subset including the observed choice and a random sample from the potential choices might be used as a substitute for the entire choice set. Although random sampling-based approaches offer computational advantages, it is important to be aware of the potential bias and inconsistent parameter estimates that can arise due to improper representation of alternative choice sets (Rashidi and Mohammadian, 2015). Several studies on modelling the location choice of business establishments have been conducted to address this issue, utilizing observed data to create structured choice sets. For instance, Elgar et al. (2009) searched for businesses employing two geographical anchor points: the existing location and the location of the firm's owner. The chosen set was then constructed by drawing an ellipse around these two points. Then, a multinomial logit model based on random sampling is estimated utilizing the constructed choice set. De Bok and Sanders (2005b) applied route choice modelling principles to generate systematic choice sets for each identified business relocation. A representative collection of possible business locations was constructed by developing progressive subsets of alternatives (Adler, 1993). These systematic choice sets have considered a specific location choice attribute, while multiple factors affect the location choice process of businesses. Recently, there have been attempts to utilize machine learning techniques in the residential location choice model to develop a systematic choice set. For instance, Orvin and Fatmi (2023) employed the Gaussian mixture model (GMM) for their residential location search model. However, the application of machine learning in the business location choice model remains scarce in literature.

The literature review suggests that there are limited studies which represent the behavioral process of business location choice. Therefore, the primary research question of this study revolves around developing a more refined approach to include the multi-dimensional location choice attributes in the process-oriented location choice model of businesses. The significance of this research lies in its potential to deepen the understanding of business location choice behavior, improve predictive precision, guide strategic decision-making for businesses, inform policymaking, and optimize resource allocation. In essence, bridging this research gap holds both theoretical implications and practical applications for urban freight movement. In this study, a novel machine learning (ML) approach has been introduced in the two-step location choice model of the business establishment to generate a finite number of plausible alternatives while precisely preserving the multi-dimensional location choice attributes in a single frame. An extensive business establishments dataset for 2022 is utilized in this study to develop the location choice model of businesses as part of a comprehensive commercial vehicle demand modeling for Halifax Regional Municipality (HRM).

3.2.1 Contribution of the Study

This study contributes significantly to the transportation and land use modeling and geography literature as it models the location choice of business establishments. The strength and novelty of this study lie in its proposed framework, which effectively combines machine learning techniques with conventional econometric modeling to develop a comprehensive two-stage location choice model of business establishments. One of the substantial contributions of this study is adopting an unsupervised machine learning algorithm to address the multi-dimensionality of influential factors of the location choices of businesses to address the first stage of the location decision. A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm was implemented to identify clusters of locations based on multidimensional attributes of business location choices. Following the clustering process, econometric modeling (logit/mixed logit model) has been utilized to model the location choice of businesses as part of a comprehensive freight demand modeling.

3.2.2 Potential of DBSCAN for Generating Business Location Choice Sets

DBSCAN holds significant potential for generating business location choice sets. It is a popular clustering algorithm that efficiently identifies dense regions in spatial data. When applied to business location data, it can help identify clusters of potential business sites that share similar characteristics and are located near each other. The applicability lies in its ability to handle noise and outliers in the data, which is crucial when dealing with real-world business location datasets that may contain inaccuracies or irregularities. By eliminating noise and focusing on dense regions, DBSCAN can identify meaningful clusters of potential business locations, helping decision-makers understand the distribution and spatial patterns of suitable areas for new business establishments. In addition, it does not require specifying the number of clusters beforehand, unlike some other clustering algorithms. This attribute is particularly advantageous in generating business location choice sets, as it allows for a more flexible and adaptive approach when dealing with varying market demands and changing geographical conditions. Additionally, the ability to detect irregularly shaped clusters makes it well-suited for capturing complex spatial patterns that might not be easily identifiable through other clustering methods. Consequently, it assists in creating a more accurate and realistic location choice sets, providing valuable insights for businesses to make informed decisions.

3.3. Study Area and Data

The area of interest for this research is the Halifax Regional Municipality, the capital city of Nova Scotia, Canada. This study employs a substantial dataset of business establishments from the year 2022, acquired from a third-party vendor (Data Axle, formerly known as Info Canada). The details of the study area and the dataset are discussed in Chapter 2. Additional information has been collected from 2021 Canadian Census. In this study, the location choice of a business establishment is considered at the parcel level.

Table 3.1 Descriptive Statistics of Explanatory Variables

Descriptive Statistics	**Minimum**	**Maximum**	**Mean**
Parcel Area(m^2)	10.43	80881.38	308.90
Distance to CBD(m)	0	92233.72	9223.37
Distance to Highway(m)	8.15	82777.46	4468.84
Distance to Business Park(m)	0	115952.00	9223.37
Distance to Bus Stop(m)	1.80	9223.37	3972.54
Distance to Local Street(m)	0	9223.37	51.50
Distance to Mall(m)	0	92233.72	9223.37
Population Density	1.1	18897.80	1489.98
Employment Number	90	1300	416.73
Zonal Entropy	0.01	0.57	0.14
Industry Count within 500m	0	64	2.03
Retail Count within 500m	0	130	3.14
Transportation & Warehousing Count within 500m	0	21	0.44
Wholesale Count within 500m	0	50	0.69

Table 3.1 demonstrates the descriptive statistics of the explanatory variables used in the location choice model. Because the census data is presented in an aggregate manner, it was necessary to disaggregate the information to obtain data at the individual parcel level. To achieve this, the approach employed in this study followed the method proposed by Bracken and Martin (1989), known as cross-area interpolation. This technique utilizes kernel estimation to convert a spatial distribution based on census centroid data into a continuous

density surface. Subsequently, this density surface is overlaid onto parcels, making it compatible with other Geographic Information System (GIS) data. By implementing this method, several variables were generated at the parcel level.

3.4 Conceptual Framework

A comprehensive two-stage location choice model of business establishments is developed in this study (Table 3.1). The developed business establishment location choice model operates at the parcel level, representing the utmost precision in its characterization.

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) generates clusters of parcels based on the relative importance of the multi-dimensional business location choice attributes. The generated clusters, eliminating the noises, are then utilized as a reliable selection pool to create the choice set of a business. The decision to discard outlier clusters in the selection pool is based on specific objective — to develop viable business locations within the study area of 162,623 real estate objects, encompassing both residential and commercial real estate. The use of DBSCAN helps identify and eliminate individual real estate objects considered as noise, such as those with unusual characteristics, in the context of business location choice. These outliers may not align with the typical profile of a viable business location. Consequently, these outlier clusters are excluded during the development of business location choice sets, streamlining the focus on clusters with higher potential. Next, econometric modeling is implemented to model business establishment location choice employing the generated plausible choice sets. In this study, a total of 10 parcels are selected randomly, including the one that has been pre-determined, to create the final choice set. The rationale for selecting this specific number is to minimize computation time.

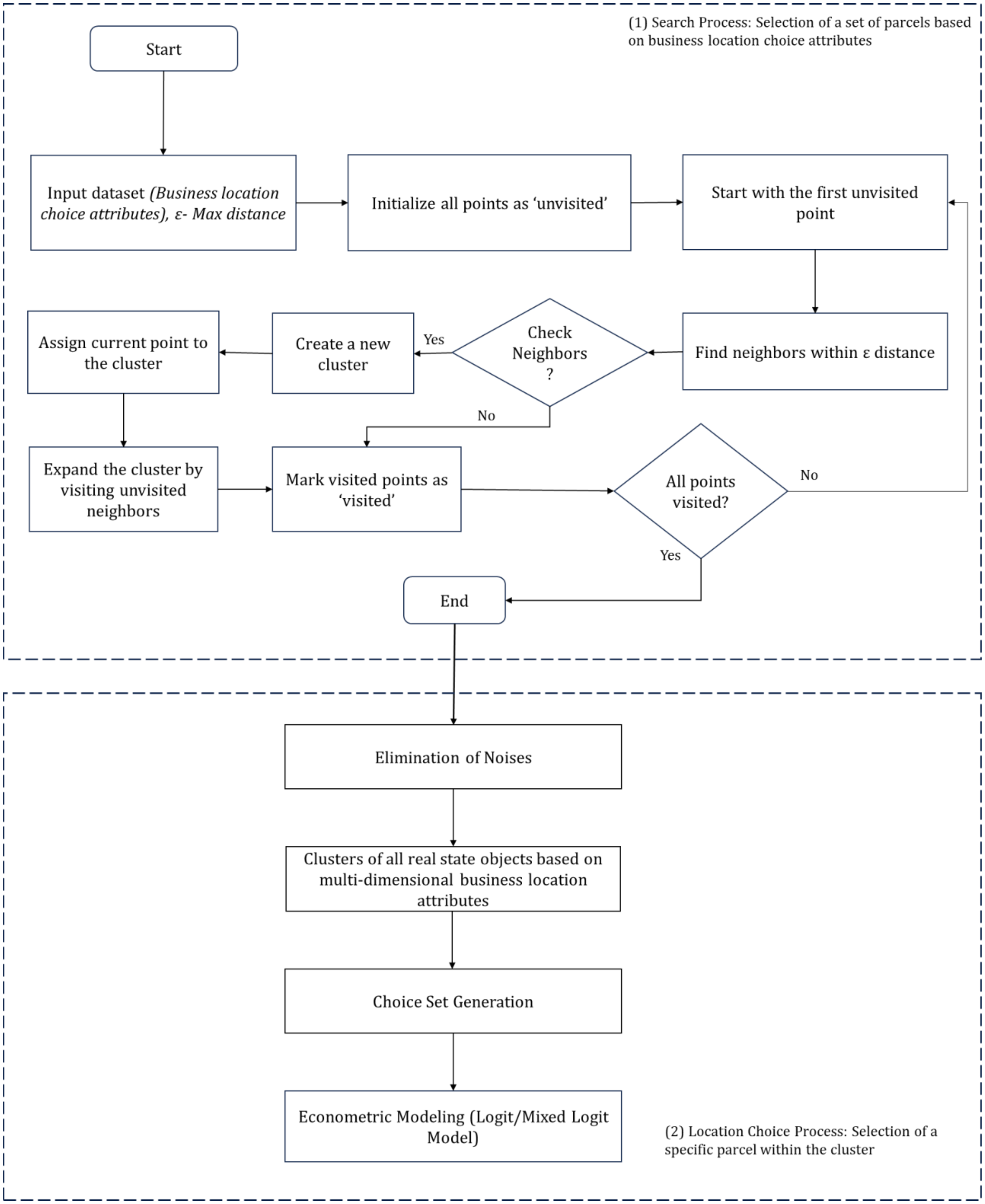
(1) Search Process: Selection of a set of parcels based on business location choice attributes
Start
Input dataset (Business location choice attributes), ε- Max distance
Initialize all points as 'unvisited'
Start with the first unvisited point
Find neighbors within ε distance
Check Neighbors ?
Yes
No
Create a new cluster
Assign current point to the cluster
Expand the cluster by visiting unvisited neighbors
Mark visited points as 'visited'
All points visited?
No
Yes
End
Elimination of Noises
Clusters of all real state objects based on multi-dimensional business location attributes
Choice Set Generation
Econometric Modeling (Logit/Mixed Logit Model)
(2) Location Choice Process: Selection of a specific parcel within the cluster

Figure 3.1 Conceptual Framework for Two-Stage Business Establishment Location Choice Model

This study hypothesized that businesses prioritize multi-dimensional location choice attributes in selecting a business location to depict the first stage of the location choice process, mimicking the search process. For instance, industries encompassing transportation and warehousing, real estate, and rental and leasing, manufacturing, construction, and public administration exhibit a propensity to cluster within regions characterized by low population density. This phenomenon is conjectured to be attributable to several influential factors, notably the cost of land, accessibility to transportation routes, zoning regulations, availability of space for expansion, considerations pertaining to noise and environmental impacts, as well as the advantages arising from specialization and agglomeration economies. The most significant multi-dimensional location choice attributes are included in the unsupervised machine learning algorithm to categorize real estate objects for each establishment. The generated cluster based on the multi-dimensional location choice attributes have the potential to provide a comprehensive understanding of the economic landscape and potential market opportunities in an area.

Utility maximization is assumed to influence businesses to mimic the second stage of the location choice model. In this stage, the decision-maker is assumed to evaluate a set of alternatives and choose the one that maximizes their utility or satisfies their preferences. This holistic approach aids decision-making processes related to business establishments, enabling stakeholders to identify locations based on market size, competition level, and economic conditions.

3.5. Methodology

3.5.1 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a widely used clustering algorithm for multi-dimensional spatial data based on density. This algorithm seeks to effectively partition data points into clusters, concurrently detecting noise points that do not exhibit strong association with any cluster. This algorithm fundamentally depends on two key parameters: the neighborhood radius denoted as ε, and the minimum number of points required to establish a dense region (*minPts*).

Let, X be a dataset, which consists of n data points $\{x_1, x_2, ..., x_n\}$.

A data point x_i is considered a core point if the number of points within its neighborhood $N(x_i)$, defined by the threshold distance ε, is greater than or equal to *minPts*. Mathematically, it is expressed as:

$$|N(x_i)| \geq minPts. \qquad 3.1)$$

The neighborhood $N(x_i)$ of a data point x_i consists of all data points x_j from the dataset X that are within a distance ε from x_i, formally represented as:

$$N(x_i) = \{x_j \in X \,/\, dist\,(x_i, x_j) \leq \varepsilon\} \qquad 3.2)$$

The reachability distance is a crucial concept in DBSCAN, which is the maximum of the distance between a core point p and another data point q (denoted as dist(p, q)) and the threshold ε. The reachability distance is represented as:

$$\text{Reachability distance } (p, q) = max\,(\varepsilon, dist\,(p, q)) \qquad (3.3)$$

This measure enables the identification of directly density-reachable points, wherein a data point q is directly density-reachable from a core point p if q is included in the neighborhood *N(p) of p, i.e.,* $q \in N(p)$.

Additionally, DBSCAN employs the concepts of density-reachability and density-connectivity. A data point q is density-reachable from a core point p if there exists a sequence of core points $\{p_1, p_2, ..., p_n\}$ such that $p_1 = p, p_n = q$, and each p_i is directly density-reachable from p_{i-1} for $i = 2, 3, ..., n$. Furthermore, a data point q is density-connected to a core point p if both p and q are density-reachable from a shared core point o.

DBSCAN identifies clusters as non-empty sets of data points that satisfy specific conditions. A cluster C contains points where each point p in C is directly density-reachable from some core point q within C, and all core points in C are density-connected. Data points that are not part of any cluster are considered noise points or outliers. To identify optimal values for ε and *minPts* in the DBSCAN algorithm, a comprehensive approach can be employed. The analysis commences with reachability plots, facilitating a visual assessment of distances between data points and aiding in the initial determination of a suitable ε. Following this, an iterative testing process involves experimenting with different ε and *minPts* values to fine-tune parameters while mitigating the risk of over-segmentation. To ensure a robust parameter selection, validation metrics like silhouette analysis and cluster stability measures can be applied, providing quantitative insights into cluster quality and consistency. Thus, DBSCAN effectively discovers clusters of arbitrary shapes and demonstrates resilience to noisy datasets, making it a valuable tool in various spatial data clustering applications.

3.5.2 Mixed Logit Model

This study has formulated distinct location choice models for each economic sector: Industry, Retail, Service, Wholesale, and Transportation & Warehousing. To comprehensively address unobserved heterogeneity within business establishments in a given sector, this study employs a mixed multinomial logit (MMNL) modeling approach. Recognizing the diverse preferences inherent in businesses, the adoption of a standard logit model assuming homogeneity proves insufficient in adequately capturing these intricacies. The utilization of the mixed logit model allows for the accommodation of variations in preferences and considers unobservable factors that impact location decisions. This sophisticated modeling approach enhances predictive accuracy, thereby enabling businesses and policymakers to make more informed decisions.

Let U_{ni} represent the random utility of an establishment *n* derived from a chosen location *i*. The random utility U_{ni} can be described according to the following equation:

$$U_{ni}=V_{ni}\ \beta+\mathcal{E}_{ni} \tag{3.4}$$

where,

V_{ni} = the deterministic part in the utility (a function of attributes related to location of businesses)

$\mathcal{E}_{ni}$ = a random error term (assumed to be identically and independently distributed (IID) across individuals and alternatives)

here, β represents the difference in preferences; $f(\beta|\theta)$ denotes the density function under the overall parameter θ. In the scenario where there is no difference in preferences (β

remains fixed), the probability L_{ni} of choosing a location from a choice set is calculated following equation:

$$L_{ni}(\beta) = \frac{e^{[V_{ni}(\beta)]}}{\sum_{j=1}^{I} e^{[V_{nj}(\beta)]}} \tag{3.5}$$

In the mixed logit model, the parameter β is randomly changed. Consequently, it becomes essential to multiply the distribution of β in order to derive the conditional selection, accounting for the existence of random preference differences. The ultimate form of the mixed logit model with random preference differences is as follows:

$$P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta \tag{3.6}$$

The form of the distribution for $f(\beta|\theta)$ is typically chosen based on practical considerations and can include options such as a normal distribution, a lognormal distribution, or a uniform distribution, among others. The expression for the random coefficient, utilizing a normal distribution as the mixing distribution, can be stated as follows:

$$\beta = \overline{\beta} + \sigma^* \eta \tag{3.7}$$

where $\bar{\beta}$ is mean of the random coefficient; σ is the standard deviation of the random coefficient; η is standard normal variable, typically referring to a random variable with a mean of 0 and a variance of 1. This involves using 200 Halton draws, a method for generating random samples from the standard normal distribution, which can be utilized to simulate the values of the random coefficients.

The log-likelihood function is utilized to estimate the model parameters, and its goodness of fit is measured using the AIC (Akaike Information Criterion) and R^2 values.

3.6 Results and Discussion

3.6.1 Choice Set Generation Based on Cluster Analysis

In the clustering process, three distinct clusters (Cluster 0, Cluster 1, and Cluster 3) have been identified. The presence of cluster label -1 indicates noise points that were not assigned to any specific cluster. To assess the quality of clustering results obtained through the application of the DBSCAN algorithm, one of the commonly used matrices is Silhouette Coefficients.

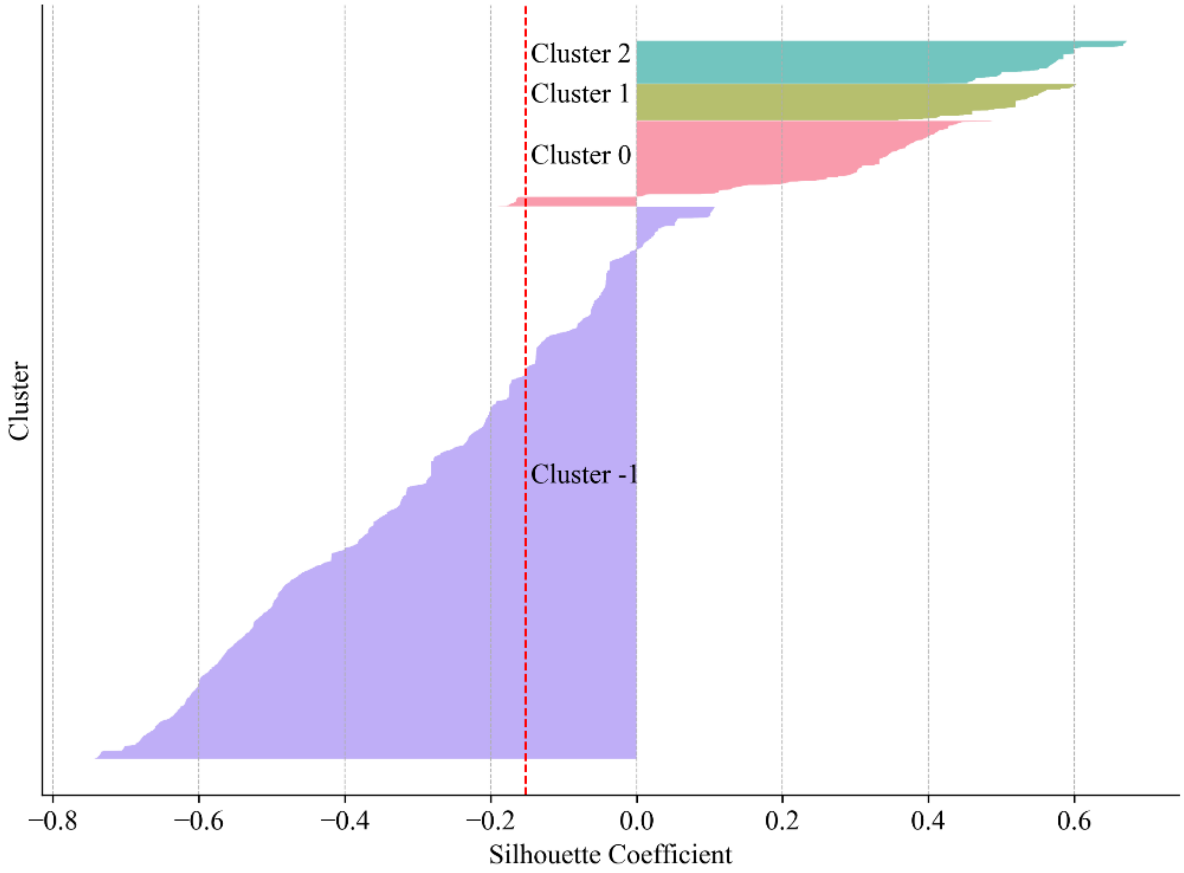

Figure 3.2 Silhouette Plot for DBSCAN

The Silhouette Coefficient values represents the similarity of each data point to its assigned cluster compared to other clusters, with a range from -1 to +1. The Silhouette Coefficient plot showcases horizontal bars, one for each cluster, with their widths indicating the number of data points within the respective clusters. Interpretation of the Silhouette Plot involves examining the Silhouette Coefficient values for each cluster. Positive values close to +1 indicate well-clustered data points appropriately assigned to clusters, while values near 0 suggest data points near decision boundaries and suboptimal clustering. Figure 3.2 demonstrates a reasonably separated clusters obtained through the application of the DBSCAN algorithm, offering a reliable and meaningful clustering outcome. The following meaningful insights are extracted from the generated cluster by utilizing density plots (Figure 3.3) and a heatmap (Figure 3.4) for all attributes within each cluster:

Cluster 0 (*Accessible but less affluent suburban areas*): It comprises locations that exhibit a relatively close proximity to key urban amenities, such as the central business district (CBD), highways, business parks, public transit, and local streets. However, these areas are notably further away from shopping malls. The population density and income levels in this cluster hover around the average, while employment levels are below average. Moreover, the diversity of employment in Cluster 0 is also below average, indicating a lower range of economic activities. These characteristics collectively suggest that Cluster 0 represents accessible but less affluent suburban areas with some clustering of economic activity, albeit with lower diversity. Businesses that might thrive in this cluster include discount retailers catering to value-conscious customers, fast food restaurants strategically located close to residential communities, and various service-based establishments serving the surrounding suburban population.

Cluster 1 (*Less accessible, lower density suburban areas*): Cluster 1 is situated at considerable distances from the central business district (CBD), highways, and business parks. However, these areas enjoy average accessibility to public transit, local streets, and shopping malls. Cluster 1 exhibits significantly lower population density, slightly below-average income, and employment levels. Moreover, the diversity of employment in this cluster is approximately average. This suggests that Cluster 1 represents less accessible, lower density suburban areas with a moderate mix of industries. Businesses likely to thrive in this cluster include big box retailers, requiring substantial land area, automotive-related ventures such as car dealerships and repair shops, and storage facilities well-suited for lower density regions.

Cluster 2 (*Accessible, higher density areas near business parks and highways*): It comprises locations that are farthest from public transit, local streets, and shopping malls but enjoy the closest proximity to highways and business parks. The distance to the central business district (CBD) is approximately average. This cluster exhibits significantly higher population density while income and employment levels are slightly below average. Moreover, the diversity of employment in Cluster 2 is below average, indicating a more limited mix of industries. These characteristics collectively suggest that Cluster 2 represents accessible, higher density areas situated near business parks and highways. Businesses that might flourish in this cluster include office buildings seeking locations with excellent transportation access and amenities, hotels catering to business travelers, restaurants targeting affluent customers willing to travel further for a premium dining experience, and specialty retailers offering unique products to an upscale demographic.

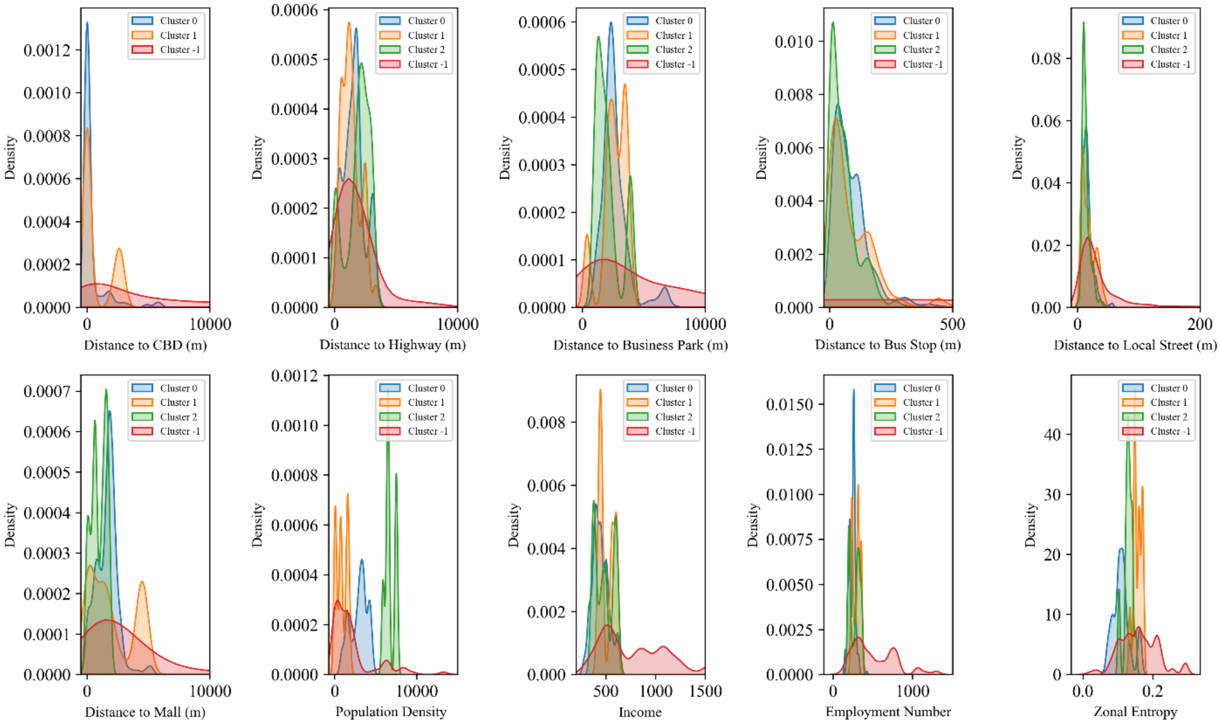

Figure 3.3 Density Plots for All Attributes within Each Cluster

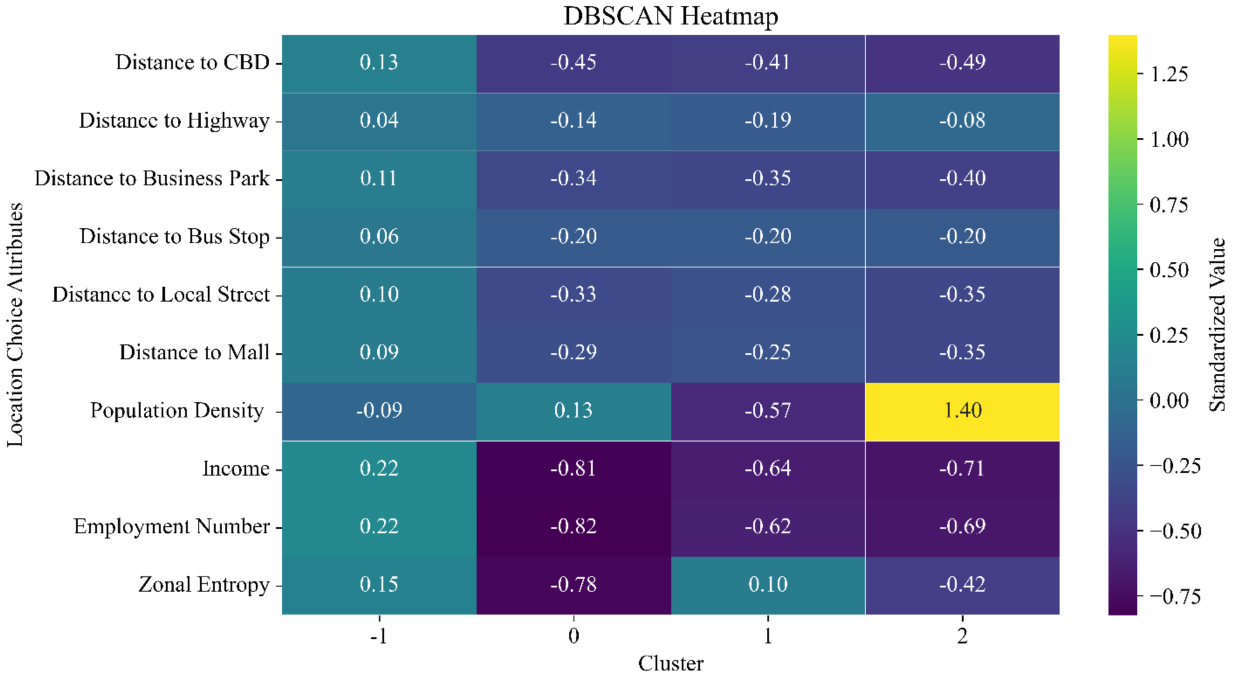

Figure 3.4 Heatmap for All Attributes within Each Cluster

The revealed underlying structure of the clusters provides essential support for data-driven decision-making. Once the noises are removed, the clusters of parcels that emerge serve as the basis for establishing a rational selection pool for a business establishment. Subsequently, 10 parcels are randomly chosen, including the one that has been pre-selected, to form the ultimate choice set. The rationale behind choosing this specific number is to minimize computation time. In addition, it reduces the likelihood of violating the independence of irrelevant alternatives (IIA) since the unobserved attributes of locations in the same neighborhood are likely to be similar (Manski, 1977). McFadden (1978) demonstrated that a subset, comprising the observed choice and a random sample from the potential choices, can serve as a substitute for the entire choice set. This study postulated that such an approach would eliminate unreliable alternatives from the selection pool.

Table 3.2 Comparison between One-Stage and Two-Stage Location Choice Model

	Industry		**Retail**		**Service**		**Wholesales**		**Transportation & Warehousing**	
Goodness of fit	One Stage Model	Two Stage Model	One Stage Model	Two Stage Model	One Stage Model	Two Stage Model	One Stage Model	Two Stage Model	One Stage Model	Two Stage Model
Restricted log likelihood	-2516.13	-2309.49	-3009.72	-2763.14	-892.53	-626.34	-1003.84	-849.65	-564.41	-439.79
Log likelihood function	-1683.56	-1409.23	-1863.23	-1550.76	-311.47	-173.15	-563.43	-411.49	-237.62	-155.39
R-squared (constants only)	0.331	0.39	0.381	0.4388	0.651	0.7235	0.439	0.5157	0.579	0.6467
R-squared (constants only) (adjusted)	0.326	0.3889	0.373	0.4382	0.641	0.7224	0.431	0.5142	0.571	0.6448
AIC	3395.12	2846.5	3746.46	3121.5	642.8	366.3	1146.36	843.62	491.24	328.8

The effectiveness of the systematic choice set was evaluated by employing a mixed multinomial logit model with randomly selected alternatives. The findings (Table 3.2) reveal that the systematic choice set model achieved better goodness of fit compared to the model with randomly chosen alternatives. Thus, confirming the validity of the initial hypothesis, which suggests that eliminating inaccurate alternatives through the systematic choice set generation improves the goodness-of-fit.

3.6.2 Results of Location Choice Model of Five Business Sectors

The estimated coefficients of the explanatory variables, accompanied by their respective *t*-statistics in parentheses are presented in Table 3.3. The coefficients represent the effect of each explanatory variable on the likelihood of a business choosing that location, holding all else constant. Positive coefficients indicate a higher likelihood and negative coefficients a lower likelihood.

Industry: The proximity to a business park within a 500-meter radius has a substantial impact on attracting medium-sized industry businesses with 10-99 employees. These firms seem to find the presence of business parks important, likely due to the access to shared resources, suppliers, and expertise that can support their growth and expansion. Business parks typically offer large, flexible spaces, which meet the requirements of growing medium-sized firms. On the other hand, being close to a Central Business District (CBD) decreases the likelihood of industry businesses choosing that location. The higher land costs near CBD make them less attractive, and the limited, inflexible spaces may not suit industrial needs. Additionally, traffic congestion near CBD leads to longer travel times for suppliers and employees, making such areas less appealing. Interestingly, despite the advantages of business parks, the overall proximity to these parks reduces the likelihood of industry firms choosing to locate there, suggesting a general preference for locations away from business parks. However, other significant factors, such as proximity to bus stops and local streets, play crucial roles in determining location choices for industry businesses. Moreover, locating in a rural area slightly increases the likelihood of industry businesses selecting that option, while proximity to other industry businesses has a positive impact, likely due to potential supply chain links and synergies. Conversely, proximity to wholesale businesses strongly decreases the likelihood of location, possibly due to competition for similar spaces. The significant standard deviations for certain variables indicate heterogeneity within the industry sector, suggesting that different factors influence location choices for various industry firms, depending on their specific needs and characteristics. These determinants might encompass establishment-specific attributes such as employment numbers, annual sales volume, and floor area. Furthermore, unobserved contextual effects and potential measurement errors may contribute to the observed dispersion, thereby underscoring the multifaceted nature of the decision-making processes within this sector.

Retail: In the context of choosing a location for retail, certain factors have been identified as influential in the decision-making process. One of the key findings is that proximity to the CBD, highways, bus stops, business parks, and existing retail outlets is associated with a higher likelihood of selection. This suggests that being closer to customers, essential infrastructure, and other businesses improves accessibility and overall desirability for retail establishments. Additionally, larger retail outlets, such as supermarkets and big box stores, tend to see higher volumes of goods movements and truck deliveries due to their extensive inventory. Consequently, having easy access to highways and major roads, which the model confirms as desirable, facilitates efficient truck deliveries and the smooth movement of goods to and from these retail locations. Another significant factor impacting the choice of retail location is land use diversity, quantified by the concept of entropy. The results indicate that higher land use diversity positively influences the likelihood of selecting a particular location. This suggests that having a diverse mix of potential customers in the vicinity can be advantageous for retail businesses.

Moreover, the suburban setting emerges as a notable aspect affecting the decision of retail location. On average, being in a suburban area increases the likelihood of selection. However, the situation is not uniform among all retailers, as indicated by the standard deviation term. Some retailers strongly prefer suburban locations, while others do not find them as appealing. The attractiveness of suburban areas to retailers may stem from various factors, including potentially lower costs, larger land parcels, and more parking options. Nevertheless, the observed variation in retailers' preferences for factors like customer base, costs, and agglomeration contributes to the heterogeneity in the desirability of suburban locations for retail businesses. Thus, understanding these diverse factors is essential for making informed decisions when choosing the most suitable retail location.

Service: The choice of location for service uses is influenced by several key factors. Larger parcel areas, proximity to the CBD, and access to public transit are associated with higher log odds of selecting a suitable location for a service business. This is because larger sites can accommodate the operational needs of service enterprises, such as storage, vehicle parking, and maneuvering areas. Being close to the CBD is advantageous for service uses as it allows them to reach corporate clients located in downtown areas, thereby expanding their customer base. Moreover, easy access to public transit is particularly beneficial for lower-income workers, as it facilitates their commute to service locations.

However, certain factors are found to have a negative impact on the log odds of choosing a location for service uses. Proximity to highways, business parks, and urban areas is associated with a decrease in the likelihood of selection. While major road access is still necessary, service businesses often prefer locations with less highway proximity to avoid noise and congestion. Standalone locations are also preferred over business parks due to the flexibility and cost-effectiveness they offer. Non-urban locations are favored as well, as they provide ample space, lower costs, and less congestion, while still being accessible, as indicated by the positive effect of CBD proximity. Higher land use diversity and the presence of surrounding industries and retailers also play a role in location selection, albeit to a lesser extent. One notable finding is the significant heterogeneity in preferences regarding highway proximity among service uses. This variation is attributed to the diverse needs and customer bases of different service businesses, leading to differing opinions on the importance of highway proximity in their choice of location.

Wholesale: Wholesalers consider various factors when choosing a location for their operations. Findings shows that proximity to highways and positions within business parks positively influence the likelihood of wholesalers selecting a specific site. Easy access to highways enables efficient transportation networks for shipping and receiving goods, while business parks offer valuable benefits such as larger spaces, well-established infrastructure, and access to other businesses that can cater to the wholesalers' needs. On the other hand, certain factors are associated with lower odds of wholesalers choosing a location. Higher population density, rural settings, and proximity to CBD tend to discourage their selection. Wholesalers typically prefer less densely populated, more suburban areas that provide ample space for storage, parking, and logistics activities. However, accessibility remains essential, and even in non-rural areas with lower density, the necessary infrastructure and accessibility are still available. Interestingly, wholesalers' preferences for proximity to CBD show significant heterogeneity. The variations in their supply chains and customer bases lead to diverse needs for CBD accessibility and location preferences in general. Overall, the findings align with wholesalers' focus on efficient goods movement.

Transportation & Warehousing: Results reveal several distinct preferences within the transportation industry. Larger sites are favored due to their ability to accommodate crucial operational needs such as vehicle staging, loading/unloading, and material storage. On the other hand, smaller transport uses prefer locations near highways, as they heavily rely on efficient road networks. Moreover, businesses within the transportation sector tend to opt for locations that are farther from the central business district (CBD), which offer more flexibility, ample space, and cost advantages while still ensuring accessibility. Another key finding is the preference for clustering with related transport businesses, indicating a desire for collaboration and synergy. However, surrounding industries are not preferred, suggesting a desire for a dedicated space within the transportation sector. Surprisingly, proximity to business parks and rural locations has only marginal effects, with public transit access having a relatively smaller impact on location preferences. Standalone locations are favored over business parks due to their advantages in terms of flexibility, space, and cost. Despite being non-urban, such locations offer large and cost-effective sites, making them appealing options for transportation businesses despite their requirement for accessibility. Site size preferences also vary significantly among different transportation uses, reflecting the diverse needs and operations within the industry.

Table 3.3 Parameter Estimation of Business Location Choice Model

Explanatory Variables	**Industry**	**Retail**	**Service**	**Wholesales**	**Transportation & Warehousing**
(Number of employee<10)*(Distance to Highway<500m)					.80(2.18)**
(Number of employee10-99)*(Distance to Business Park<500m)	.28(2.20)*				
Parcel Area(ln)	.85(18.49)***		1.39(8.78)***		1.04(4.25)***
Distance to CBD<500m	-.21(-1.98)**	.26(1.81)*	2.78 (4.63)***	-0.60(-1.82)*	-1.01(-2.85)***
Distance to Highway<500m		.49(5.29)***	-1.87(-3.69)***	.44(2.09)**	0.55
Distance to Business Park<500m	-1.26(-2.93)***	.72(1.92)*	-3.58(-2.67)***	1.24(2.09)**	-2.19(-1.96)**
Distance to Bus Stope<1000m	.28(1.84)*	.61(3.32)***	1.53(3.23)***		1.004(1.72)*
Distance to Local Street<200m	-.57(-1.94)*	-.92(-2.48)**			
Distance to Mall<500m		-.62(-3.25)***			
Population Density per square km(ln)				-.12(-1.98)**	
Employment(ln)	.23 (2.58)***				
Entropy	1.91 (1.86)*	.74(5.89)***	1.03(2.81)***	.68(2.75)***	
Rural	.26(1.67)*			-1.14(-3.75)***	.83(1.75)*
Suburban		.34(3.18)***			
Urban			-6.66(-6.91)***		
Industry Count within 500m	.29(12.59)***		.09(2.29)**	-.09(0.10345)***	-.08(-2.03)**
Retail Count within 500m		.21(18.04)***	.14(4.77)***		
Transport Count within 500m				.23(2.23)**	2.45(9.10)***
Wholesale Count within 500m	-.13(-4.10)***			.57(8.72)***	
Entropy(Stdev)	5.69(0.04)**				
Employment Number(ln) (Stdev)	.349(0.02)*				
Rural(Stdev)	.42(0.04)*				
Suburban(Stdev)		2.01(3.75)***			
Distance to Highway<500m (Stdev)			1.43(2.27)**		
Distance to CBD<500m (Stdev)				1.99(2.96)***	
Parcel Area(ln) (Stdev)					.95(2.82)***
***, **, * ==> Significance at 1%, 5%, 10% level.					
Goodness of fit					
Restricted log likelihood	-2309.49	-2763.10	-626.30	-849.65	-439.79
Log likelihood function	-1409.23	-1550.76	-173.15	-411.49	-155.39
R-squared (constants only)	0.39	0.4388	.7235	0.5157	.6467
R-squared (constants only)(adjusted)	0.3889	0.4382	.7224	0.5142	.6448
AIC	2846.5	3121.5	366.3	843	328.8

3.7 Conclusion

This study presents a novel process-oriented business location choice model that incorporates multi-dimensional built environment and neighborhood attributes. The integration of the unsupervised machine learning technique with the mixed multinomial logit (MMNL) model allows a data-driven approach to understand the intricate phenomenon of the two-stage location choice process of business establishment. This approach can help reveal underlying patterns and connections that might not be visible through traditional one step multinomial logit model. The rigorous analysis yields insights into how certain types of businesses conduct the initial search process considering the influential built environment and neighborhood characteristics in a two-step location choice model. The outcomes from this study indicate that the choice of location for service businesses is influenced by several factors, such as larger parcel areas, proximity to the CBD, and access to public transit, which are associated with higher log odds of selection. On the other hand, factors like proximity to highways, business parks, and urban areas have a negative impact on the likelihood of selection, with standalone, non-urban locations preferred for their flexibility, cost-effectiveness, and access to diverse customer bases. The transportation industry prefers larger sites for crucial operational needs and smaller transport businesses choose locations near highways. They prioritize standalone locations over business parks, prefer clustering with related transport businesses, and the volume of goods movements and truck traffic varies widely based on the scale and type of each transportation use.

However, this study is subject to certain limitations and future directions, as it exclusively focuses on a constrained set of the accessibility factors (transportation infrastructure, proximity to major roads, highways, airports, and public transportation options) as the influencing features governing business establishments' location choice. Although these attributes significantly influence business location decisions, they do not encompass the entirety of the factors that affects location choice of businesses. The inclusion of office and business profile factors in the location choice model for business establishments holds the potential to significantly enhance decision-making accuracy. By incorporating these factors, businesses can make well-informed choices about their preferred locations. The office profile factor allows businesses to identify spaces that cater to their specific spatial needs, work environment preferences, and technological requirements, resulting in increased productivity and employee satisfaction. Simultaneously, the business profile factor aids in selecting locations aligned with strategic objectives and target markets, promoting growth opportunities and networking prospects within the region. By reinforcing brand identity and considering long-term viability, businesses optimize resource allocation and cost-effectiveness. Furthermore, to enhance the comprehensiveness of location choice modeling, it is recommended to extend the proposed model to incorporate qualitative data, including interviews with business owners. In terms of practical considerations, exploring efficient data collection methods is crucial, with a focus on participant selection, interview methodologies, and the reliability of gathered information. The integration of qualitative data offers potential advantages, providing a more holistic understanding of business location choice dynamics. Moreover, incorporating qualitative insights can refine and validate the quantitative model, potentially improving its predictive accuracy and making the location choice model more relevant to real-world scenarios. The developed model has the potential to address temporal dynamics in real-world applications through the development of hybrid models (a combination of a time-series forecasting model to predict future conditions and then using DBSCAN to cluster locations based on those predicted conditions. The proposed model may encounter certain challenges in practical implementation, particularly in the areas of parameter tuning, cluster adaptability, and memory efficiency. Making strategic parameter choices and implementing optimizations are crucial for ensuring accurate modeling.

The developed two-step location choice model focusing on major economic sectors, including industry, retail, service, wholesale, and transportation, constitutes a noteworthy contribution to the existing literature. The integration of this novel and potentially insightful method provides valuable and in-depth insights for various domains, such as commercial vehicle movement modeling, urban planning, business location strategies, and policymaking concerning economic development. Most importantly, this study represents a substantial advancement in location choice modeling in developing a commercial goods movement modeling for an integrated transport, land use and energy modeling system.

Chapter 4

Shopping Destination Choice Model[3]

4.1 Background

Activity-based travel demand modeling incorporates destination choice as a crucial component, encompassing activities such as shopping, dining, or recreation. These models predict the likelihood of choosing a particular destination, considering variables such as distance, accessibility, attractiveness, and the presence of amenities or services. Shopping serves as a significant catalyst for the economy and drives considerable travel demand. With in-person shopping remaining the predominant form of retail, surpassing e-commerce by a wide margin, individuals travel to stores to fulfill their shopping needs, resulting in a substantial portion of overall trips. Shopping destination choice is pivotal in understanding the flow of people and goods. Hence, a better understanding of the process involved in shopping destination location choice is essential to enhance the predictive accuracy of activity-based travel demand modeling.

The COVID-19 pandemic and associated mitigation policies brought about significant changes in travelers' attitudes, businesses, and supply chains (Grashuis et al., 2020; Shen et al., 2022), resulting in a shift in the way individuals make daily travel destination choices.

[3] This chapter is adapted from:

Alam, Jahedul MD., Mahmud, N., and Habib, M. A. (2024). A Comprehensive Framework for Shopping Destination Choice Model: Combination of Machine Learning and Discrete Choice Modeling. Accepted for presentation at the 17th International Conference on Travel Behavior Research. July 14-18, 2024, Vienna, Austria.

Consequently, travel choices, including work and shopping preferences, have become increasingly complex, particularly affecting how people select their shopping destinations. Although the COVID-19 pandemic is no longer a major threat, it has led to lasting changes in travel and shopping habits, impacting businesses and retail with potential long-term effects (Statistics Canada, 2023). Therefore, a modeling framework of destination choice is of paramount importance to capture the post-COVID changes in shopping destination choices.

In travel demand modeling, destination choice modeling plays a pivotal role by examining potential locations through factors like socio-economic attributes, distance, destination characteristics, and zone-specific features. The destination choice modeling approach comprises two integral steps: the destination search process, which generates a viable set of choices, and the modeling process, facilitating a deeper understanding of individual shopping destination choice behavior. However, the presence of a large number of alternatives in the choice set is a challenge to spatial choice modeling (Habib and Miller, 2009) and is likely to increase computational complexity (Bouzouina et al., 2021; Lee et al., 2010). Traditional destination choice models use zones, such as traffic analysis zones (TAZ), as potential alternative locations that anchor all shopping trips into a single zonal centroid. Bradley et al. (2010) introduced SACSIM, a regional forecasting model using DaySim, an activity-based econometric model. It associates activities with specific parcel locations, allowing detailed land use and urban design data to be incorporated, which can extend to include business unit characteristics for a better understanding of shopping travel. Sivakumar and Bhat (2007) developed a conceptual and econometric framework for non-work activity location choice that incorporates spatial cognition, heterogeneous preference behavior, and spatial interaction. Suel and Polak (2018) emphasized the importance of including online shopping in destination choice modeling but noted challenges in finding data for discrete choice models.

The study on shopping trips has explored models for generating shopping destination choices, primarily using the concept of gravity modeling (Haynes and Fotheringham, 1985). Gravity models assume that the likelihood of choosing a shopping destination decreases with distance and increases with shopping center size (González-Benito, 2005; Huff, 1963; Simmonds and Feldman, 2011). Hassan et al. (2019) implemented a two-stage destination choice modeling framework, initially employing a rule-based fuzzy logic model to create a latent choice set, followed by its integration into a discrete choice model for destination selection behavior estimation. Phan et al. (2022) used matrix factorization with Bayesian personalized ranking to create personalized destination lists from a user-zone-visited frequency matrix derived from a travel survey. Kikuchi et al. (2001) developed a destination choice model using a coordinates-based methodology to resolve the issues of zone systems in destination choice modeling. They employed Monte Carlo Markov Chain for efficient choice probability evaluation and behavior simulation with a large number of alternatives. Jonnalagadda et al. (2001) employed a tour-based microsimulation to model destination and mode choices in San Francisco as part of a travel demand forecasting model. Their destination choice models utilized multinomial logit with 40 randomly chosen traffic zones as potential destinations. However, decision-makers undergo an essential selection process often overlooked by most models, which narrows down alternative choices (Habib and Miller, 2009).

The literature indicates a scarcity of studies representing the behavioral process of shopping destination location choice. Therefore, the primary research question of this study is centered on improving estimation and prediction accuracy in the two-step shopping destination choice model by incorporating multi-dimensional attributes of the behavioral process. This study introduces a novel machine learning approach to the two-step shopping destination choice model, creating a finite set of viable alternatives while accurately preserving the behavioral process within a single framework. The contribution of this study lies in the application of two machine learning algorithms (Random Forest and K-Prototype) to accommodate multi-dimensional factors in developing plausible choice sets that capture individual preferences. Following the choice set generation, econometric modeling (mixed logit model) has been utilized to model the shopping destination choice.

4.2. Methodology

This study develops a two-stage shopping destination choice modeling framework that involves destination location choice set generation and shopping destination choice modeling. Figure 4.1 presents a sequential modelling framework for shopping trip destination choices at the business establishment level. It implements the Random Forest algorithm to identify relevant features for both shopping destination location choice set generations and discrete choice modeling. The location choice set development process involves a clustering of shopping destinations that utilizes features derived from RF modeling. Individuals within clusters are identified and alternative locations are randomly chosen based on individuals' cluster memberships. Choice sets are then utilized in mixed logit modeling for predicting shopping destinations.

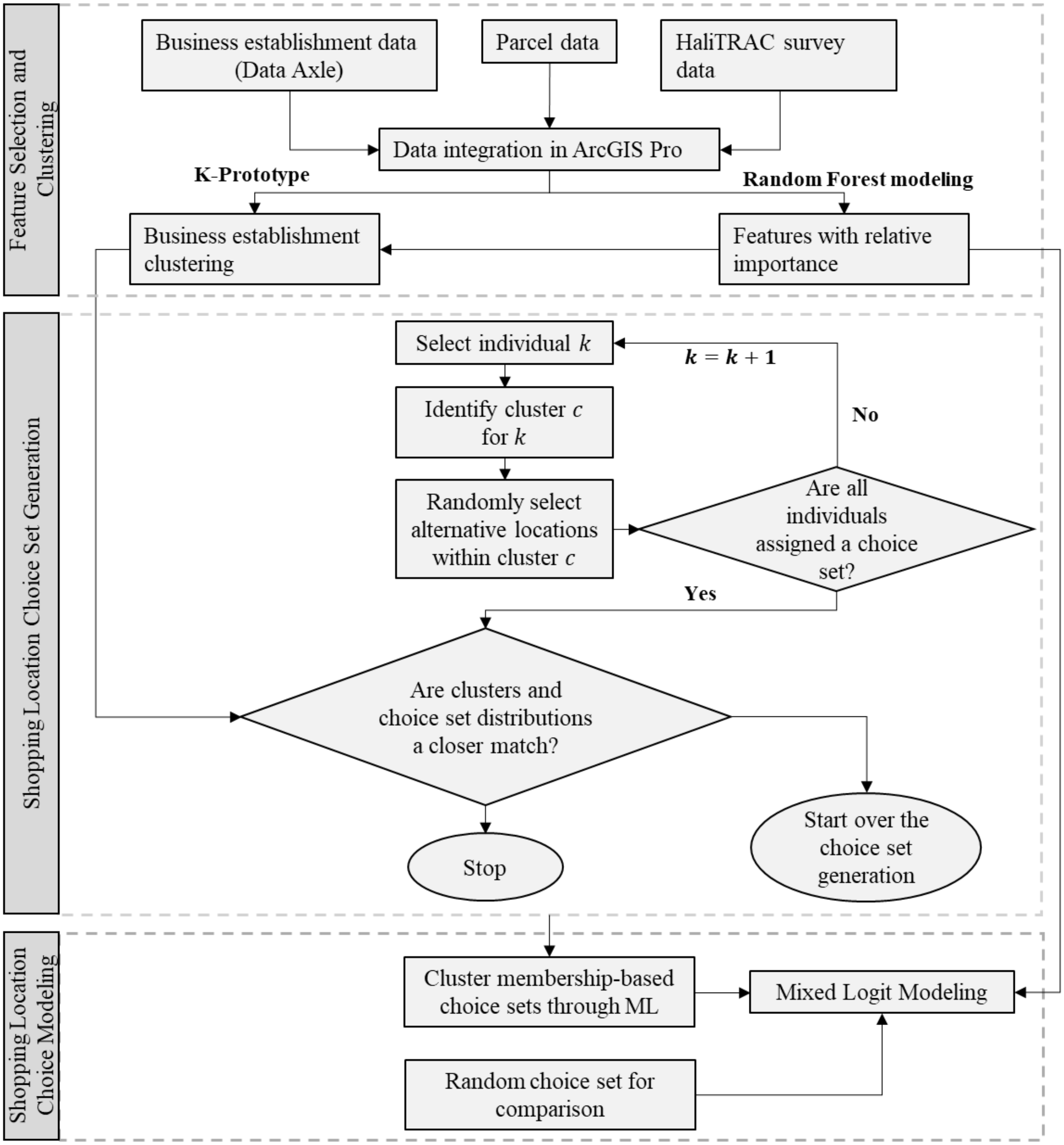
Feature Selection and Clustering
Business establishment data (Data Axle)
Parcel data
HaliTRAC survey data
Data integration in ArcGIS Pro
K-Prototype
Random Forest modeling
Business establishment clustering
Features with relative importance
Shopping Location Choice Set Generation
Select individual k
$k = k + 1$
Identify cluster c for k
No
Randomly select alternative locations within cluster c
Are all individuals assigned a choice set?
Yes
Are clusters and choice set distributions a closer match?
Stop
Start over the choice set generation
Shopping Location Choice Modeling
Cluster membership-based choice sets through ML
Mixed Logit Modeling
Random choice set for comparison

Figure 4.1 Conceptual Framework of Shopping Trip Destination Choice Modeling

4.2.1 Halifax Travel Activity (HaliTRAC) survey data

This study utilizes data from the 2022 Halifax Travel Activity (HaliTRAC) survey. The HaliTRAC survey reached out to 50,500 households in HRM through three stages. In the first phase, 4,000 postcards were sent to households, inviting them to take part in the online survey. The second phase employed random digit dialing to text around 32,000 households, offering them the opportunity to participate online or via telephone. For the third phase, 14,500 households were selected randomly through landline sampling, using both address and home phone numbers. This group had the option to complete the survey online, via mail, or by phone. Additionally, a fourth phase targeted HRM residents through social media, engaging approximately 135,000 Meta users.

The 2022 HaliTRAC survey gathered data in three main categories: household, individual, and trip information, each with varying dataset sizes. It was distributed to households, resulting in 3,731 responses. Each member of the households was requested to provide responses, leading to a total of 5,095 individuals contributing to the survey, averaging 1.37 per household. The survey achieved an overall completion rate of 13.6%. It gathered information on a 24-hour travel activity log for each individual. The survey also covered vehicle details, residential location, ownership, and demographic information of household members, along with trip specifics like locations, times, modes, and purposes. This survey offers valuable insights into travel patterns within HRM. Typically, households in the area own an average of 1.57 vehicles. Residents tend to make an average of 2.8 trips daily, with 77.6% of these trips conducted by car. The average distance covered by HRM residents is 25.6km, with 27.8km of that distance traveled by car. Additionally, 64.9% of trips in the region are undertaken by individuals traveling alone. Table 4.1 presents the descriptive statistics of all explanatory variables retrieved from the HaliTRAC survey data utilized for modelling purposes.

The summary statistics of various characteristics in the HaliTRAC data are compared with the 2021 census data across three levels: household attributes (such as dwelling type, tenure, household size, and income); individual attributes (age, gender, employment status, and occupation); and trip attributes (such as mode choice and commuting duration). For instance, the average household size reported in the HaliTRAC survey is 2.46, whereas the average household size according to the 2021 Census is 2.20. In most instances, the distribution of these attributes closely matches that of the census data, with variations ranging from 0.01% to 6%. As a result, the sample is deemed representative of the population data for HRM.

4.2.2 Business Establishment Data

An extensive dataset of business establishments for 2022 (obtained from Data Axle) helps identify 2,702 shopping business locations in Halifax Regional Municipality (HRM). It includes 30,851 complete firm records with 8-digit NAICS codes. The dataset describes establishments, their locations, business types, employee sizes, and sales volumes. This study focuses on 2,702 shopping trips, categorized into "Routine shopping (S1)" and "Special item shopping (S2)" types, accounting for 42.79% and 57.21% of all locations, respectively. S1 has 48 types, S2 has 49 types, all geocoded and matched with survey data for analysis.

Table 4.1 Descriptive Statistics of Model Explanatory Variables

Variable Name	Description	Mean/ Proportion	Standard Deviation
Business establishment characteristics			
Sales volume	Total annual volume sold in 1000	6.61	3.83
Employee size	Total employee number		
Business type	Dummy, if it is Routine shopping (e.g., retails, superstore, convenience store) = 1, 0 otherwise	95%	n/a
Individual Characteristics			
Household annual income			
< $15,000	Dummy, if the household annual income is less than $15,000 = 1, 0 otherwise	1.70%	n/a
$15,000 - $24,999	Dummy, if the household annual income is in between $15,000 - $24,999 = 1, 0 otherwise	3.27%	n/a
$25,000 - $34,999	Dummy, if the household annual income is in between $25,000 - $34,999 = 1, 0 otherwise	4.39%	n/a
$35,000 - $49,999	Dummy, if the household annual income is in between $35,000 - $49,999 = 1, 0 otherwise	9.84%	n/a
$50,000 - $74,999	Dummy, if the household annual income is in between $50,000 - $74,999 = 1, 0 otherwise	18.08%	n/a
$75,000 - $99,999	Dummy, if the household annual income is in between $75,000 - $99,999 = 1, 0 otherwise	17.72%	n/a
$100,000 - $149,999	Dummy, if the household annual income is in between $100,000 - $149,999 = 1, 0 otherwise	22.63%	n/a
$150,000 - $199,999	Dummy, if the household annual income is in between $150,000 - $199,999 = 1, 0 otherwise	13.33%	n/a
> $200,000	Dummy, if the household annual income is more than $200,000 = 1, 0 otherwise	9.04%	n/a
Travel and Network Characteristics			
Travel time	Travel time to shopping destination from origin location	11.53	7.69
Distance to central business district (CBD)	Distance between each business location to CBD, m	2537.17	3845.51
Distance to nearest business park	Distance between each business location to nearest business park, m	3792.26	3778.85
Distance to nearest mall	Distance between each business location to nearest mall, m	1752.73	1814.03
Distance to nearest highway	Distance between each business location to nearest highway, m	1023.73	673.23
Distance to nearest local road	Distance between each business location to nearest local road, m	26.37	15.91
Distance to nearest bus stop	Distance between each business location to nearest bus stop, m	145.23	363.33
Built Environment Characteristics			
Land use index	Measurements of spatial land use intensity		n/a
Retail concentration	Number of shopping businesses within 300 m of the alternative destinations	17.72	17.17

4.2.3 Random Forest-Based Feature Selection for Shopping Destination Choice Sets

This study employs Random Forest (RF) Modeling to discern factors affecting individuals' shopping destination choices. Figure 4.2 illustrates the process within RF modeling.

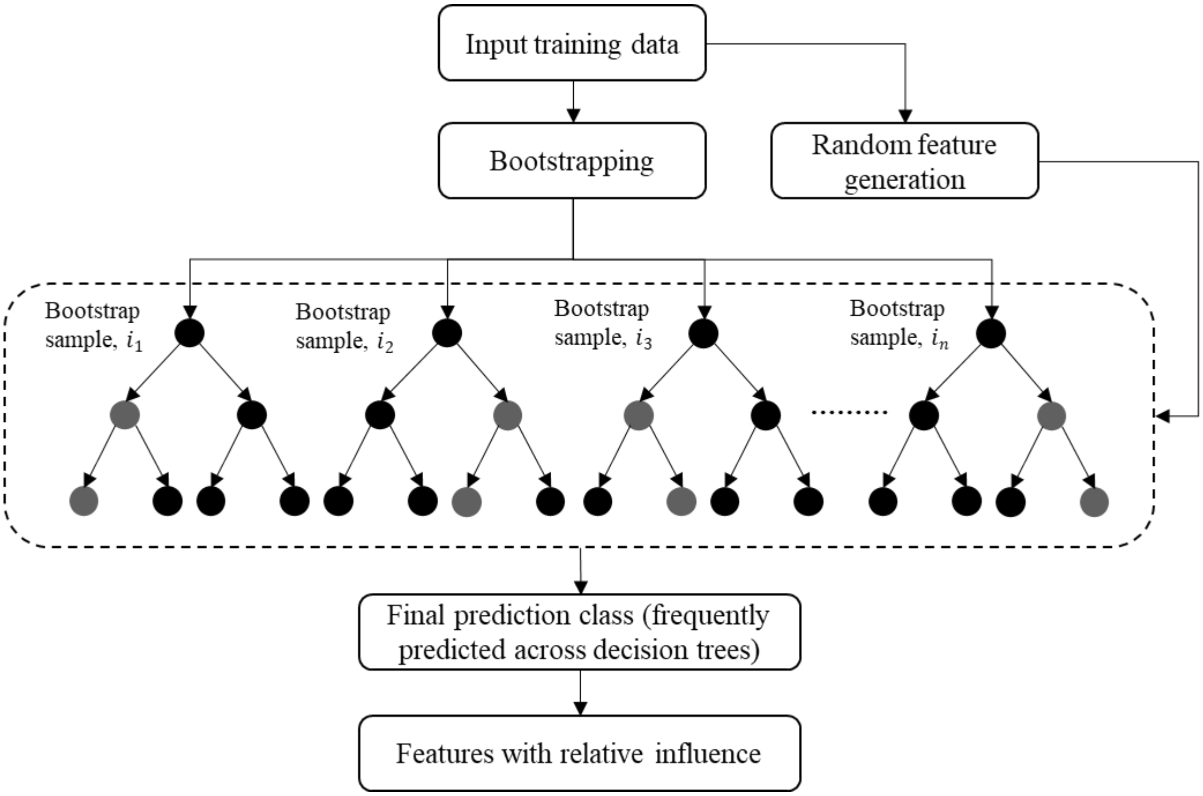

Figure 4.2 Random Forest Modeling Process

Constructing a decision tree involves creating a subset of features randomly for training the tree with bootstrap sample and the tree grows until it reaches the maximum depth. The importance of each attribute is measured using the Gini impurity index as expressed in equation 1. The Gini value is used for calculation of splitting mother node for a particular variable given that the Gini values are less for the descendent nodes. The average decrease of Gini values for a variable in all decision trees across all forests is the measure of the importance of that feature. Suppose a candidate variable x_i with a possible number of instances as $C_1, C_2, \dots \dots . C_m$, then Gini impurity index $G(x_i)$ can be expressed as follows.

$$G(x_i) = \sum_{C=1}^{m} p(x_i = C_m)(1 - p(x_i = C_m)) \tag{4.1}$$

where $p(x_i = C_m)$ is the probability of instance.

4.2.4 Destination Choice Set Generation: K-Prototype Algorithm

The K-prototype handles mixed numerical and categorical data by minimizing within-cluster dissimilarity, computed from distances between numerical data points and dissimilarities between categorical data points.

Let, X be the dataset with n data points and p attributes per data point x_i. Any data point can have either categorical or numerical variables. Suppose $C = \{c_2, c_3, \dots \dots . c_k\}$ be the set of k clusters, where each cluster C_j contains a subset of data points. The dissimilarity measure between a numerical data point x_i and a cluster centroid μ_j can be calculated using the following Euclidean distance in equation 4.2.

$$d_{num}(x_i, \mu_j) = \sqrt{\sum_{p=1}^{P}(x_{ip} - \mu_{jp})^2} \tag{4.2}$$

where x_{ip} and μ_{jp} are the p-th attributes of x_i and μ_j, respectively and P denotes the total number of attributes (both numerical and categorical).

The dissimilarity measure between a categorical data point x_i and a cluster centroid μ_j can be calculated using a distance metric appropriate for categorical data, such as the Hamming distance:

$$d_{cat}(x_i, \mu_j) = \sum_{p=1}^{P}(x_{ip} \neq \mu_{jp}) \tag{4.3}$$

where x_{ip} and μ_{jp} are the *p-th* attributes of x_i and μ_j, respectively.

The overall dissimilarity between a data point x_i and a cluster centroid μ_j can be calculated as a weighted sum of the numerical and categorical dissimilarities according to equation 4.4.

$$d\left(x_i, \mu_j\right) = (1 - \gamma) * d_{num}\left(x_i, \mu_j\right) + \gamma * d_{cat}\left(x_i, \mu_j\right) \tag{4.4}$$

where γ is a weighting factor between 0 and 1 that determines the relative importance of numerical and categorical variables.

Clusters obtained from the K-Prototype modeling are used to generate choice set for the corresponding individuals and then the choice sets are entered into the mixed logit modeling of destination choice.

4.2.5 Shopping Destination Choice Modeling: A Mixed Logit Modelling (MXL) Approach

This study employs a random utility-based discrete modeling approach, specifically the mixed logit (MXL) modeling technique to investigate individuals' shopping destination choice behavior. The study adopts MXL modeling to capture the potential heterogeneity of individual preferences in choosing shopping locations while considering varying factors, including travel time, and proximity of the destinations. Let V_{pr} represent the utility of an individual p derived from a chosen shopping destination r. The utility V_{pr} can be described according to equation 4.5.

$$V_{pr= \beta_{pr}X_{pr}+\varepsilon_{pr}} \tag{4.5}$$

Here, X_{pr} is the vector of observed explanatory attributes such as economical and regional attributes of business establishments, travel, and network attributes, socioeconomic and built environment characteristics, and ε_{pr} is the random error term, which is assumed to be identically and independently distributed (IID). Then the probability P_{pr} of choosing a destination from a choice set is calculated following the equation 4.6.

$$P_{pr} = \frac{\exp\left[\beta_{pr}X_{pr}+\varepsilon_{pr}\right]}{\sum_{1}^{r}\exp\left[\beta_{pr}X_{pr}+\varepsilon_{pr}\right]} \qquad (4.6)$$

This study follows the maximum log likelihood technique to estimate the model parameters and there is no analytical solution to this function due to its closed form. Hence the function is maximized using Quasi-Monte Carlo simulation with 200 Halton draws. The simulated log likelihood function (SimLL) used in this study is expressed below.

$$SimLL = \sum_{t=1}^{T} \ln \frac{1}{D} \sum_{d=1}^{D} P_{pr}(\beta_{pr}{}^{d}) \qquad (4.7)$$

where, T is the total alternatives and D is the total number of draws.

The model is estimated in NLOGIT6.0 platform and the goodness of fit of the model is evaluated in terms of log likelihood function, AIC and R^2 values.

4.3 Results and Discussion

4.3.1 Important Features for Constructing Shopping Destination Choice Sets

The feature selection identified nineteen factors through Random Forest modeling. Figure 4.3 shows the factors with their relative importance concerning the selection of shopping destination locations.

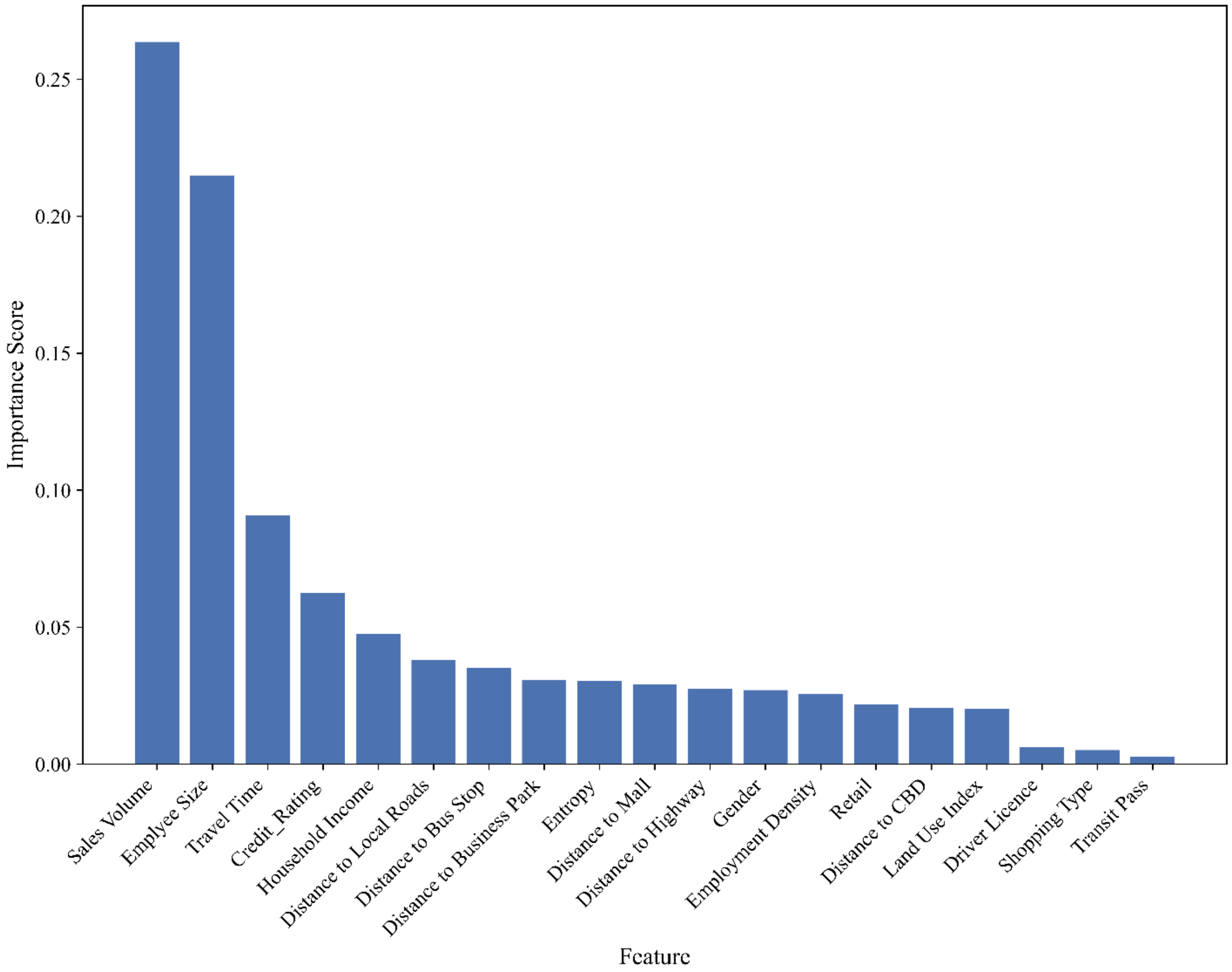

Figure 4.3 Feature Selection through Random Forest

RF modeling analyzed economic variable and travel cost, proximity, socioeconomic, and built environment factors. This yields the first 18 factors with importance weight ranging from 0.02 – 0.25 except the last two factors regarding the ownership of driver license and transit pass, which are found to show almost zero influence on the location choices; hence discarded from the analysis. Results show that sales volume (0.25), employee size (0.175), and time to reach the business locations are the top factors that attract individuals to shop at different business stores. Travel time (0.075) captures spatial constraints to choosing shopping locations and demonstrates a significant influence on shopping destination choices.

Individual characteristics such as household income are critical for choosing shopping locations. Proximity and built environment factors, including distance to central business districts and land use also influence shopping location choices. In the choice set generation process, this study uses the top economical (i.e., sales volume, employee size) and travel factors (i.e., travel time) to understand the agglomeration of business locations and spatial constraints to choosing them. Proximity, individual and built environment variables are likely to influence shopping location choices to the same degree and used for mixed logit modeling of location choices. This study also includes sales volume and travel times within location choice modeling to capture individual responses to the economic aspects of business establishments and their varying degree of accessibility. The mixed logit model presented in the following section identifies the combination of all or a subset of factors (identified through RF process) that captures the combined effects of economical, proximity, built environment, and heterogenous preferences of individuals on shopping location choices.

4.3.2. Shopping Destination Choice Set Generation

This study incorporates RF-identified top economical (i.e., sales volume, employee size) and travel factors (i.e., travel time) in the K-Prototype clustering algorithm to construct shopping destination choice sets. Three clusters are determined through the Elbow method (Thorndike, 1953). Figure 4.4 shows the changes in the BIC ratio with respect to number of clusters, which varies with the number of clusters and identifies the optimum number of clusters. The point on the line demonstrating a steeper change between preceding and the following section of the line indicates the optimum number of clusters as prescribed by the

Elbow method. In this study, cluster #3 represents the steep changes. Figure 4.5 explains the clusters concerning numerical variables utilized in the clustering process.

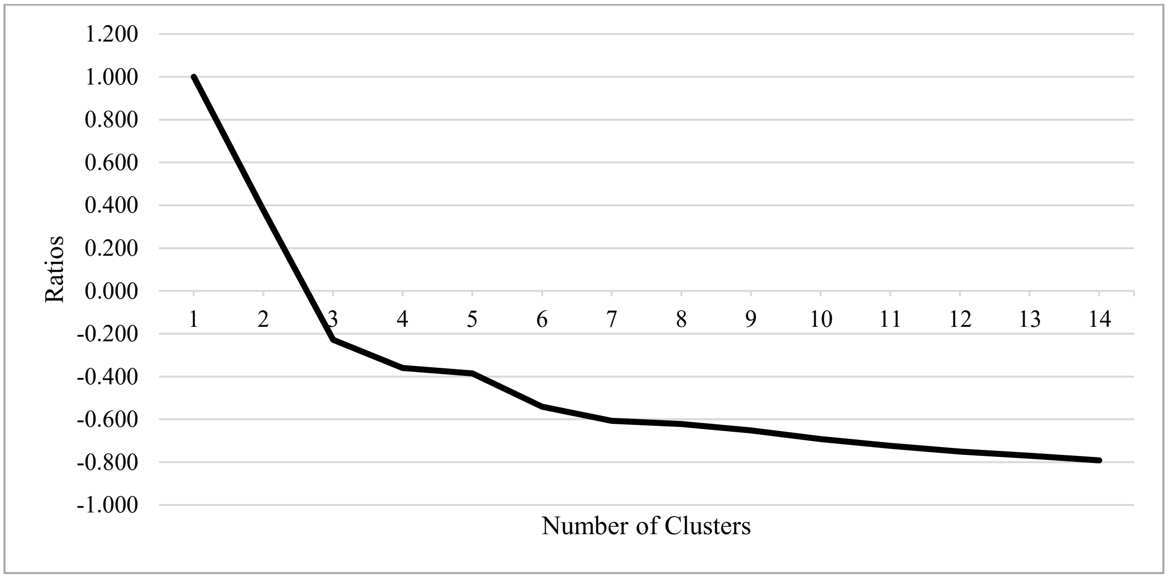

Figure 4. 4 Change of Auto-Clustering BIC with respect to numbers of clusters

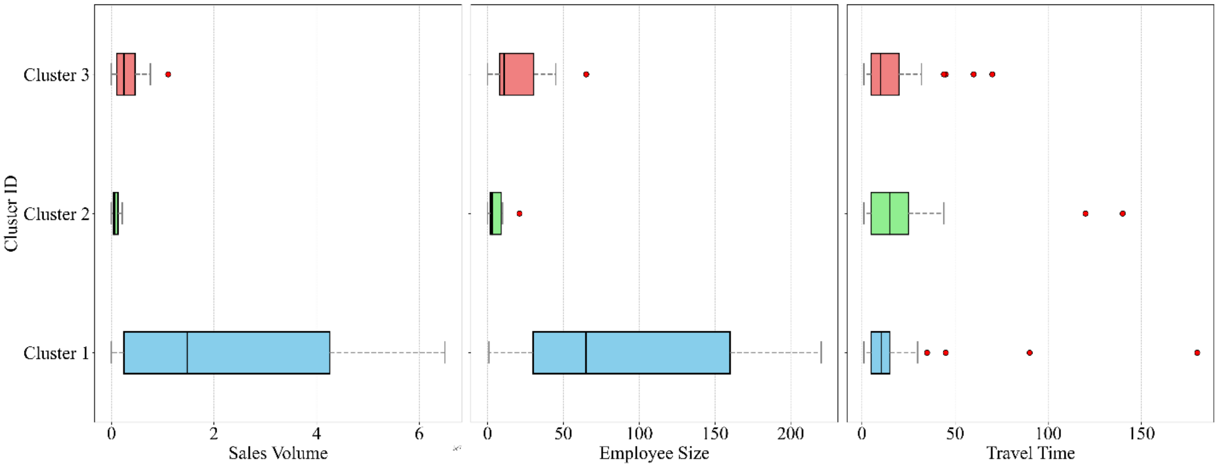

Figure 4.5 Analysis of Clusters with Respect to Sales and Employee Size and Travel Time

Clustering results show that cluster #2 and cluster #3 encompass small to medium size businesses in terms of sales volume and employee size. For example, these may include convenience and liquor stores for routine shopping, and stores for personal special item shopping. Conversely, cluster #1 represents the businesses with higher sales and employee size, such as retail and superstores for routine shopping, and home centers, and appliance stores for special item shopping. The travel time variable indicates that all types of businesses are well distributed over the Halifax Regional Municipality (HRM), and average travel time is slightly higher for cluster #2 in which there are several small businesses located in suburban area. The Clustering process also identifies individuals' membership to clusters which helps in choosing alternative locations accounting for individual preferences. Results show that 33%, 26%, and 41% individuals belong to cluster# 1, 2, and 3, respectively.

This study randomly selects alternative shopping destinations for each individual based on their membership of the clusters. This process enables the representation of individual preferences in choosing alternative destination locations. Choice sets are generated considering business establishments for two shopping types: routine shopping (S1) and special item shopping (S2). For example, retail and superstores are places of S1 category while home center, and appliances are the locations of S2 type. This study determines the distributions of shopping types resulting from the chosen business establishment locations in the choice sets. Then the distribution is compared to that of the clusters to ensure the representativeness of the generated choice sets. Figure 4.6 shows the distributions of shopping types in both clusters and choice sets obtained through the K-Prototype and cluster membership-based location selection processes.

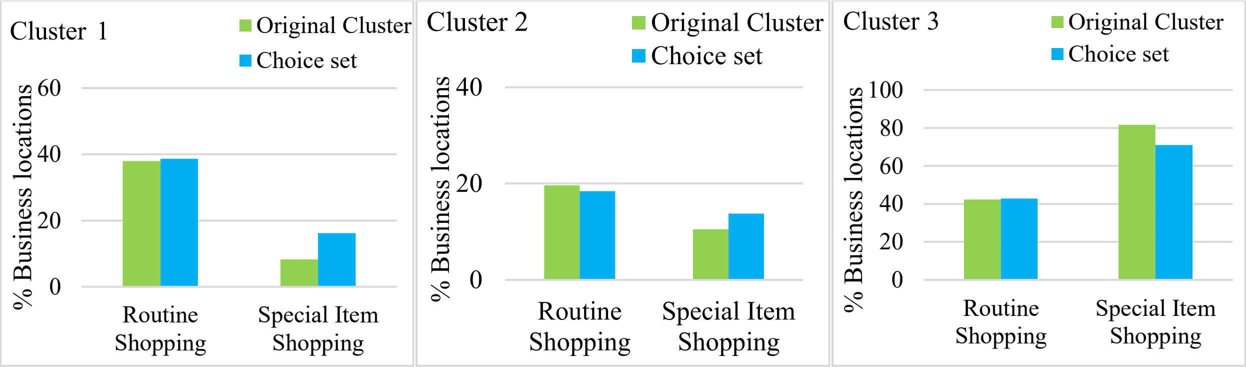

Figure 4.6 Verification of Choice Sets in Relation to the Clusters Created Through K-Prototype

Results show that the deviation between two distributions varies ranging from 0.5 and 10%, for routine and special item shopping yielding an acceptable choice sets for mixed logit modelling. The MXL model developed is found statistically significant demonstrated by different measures of model goodness fit presented in Table 4.2. The model shows a Log likelihood function value of -502.41 and an R^2 value of 0.26. This study also demonstrates the effectiveness of the ML-based location selection process in destination choice modeling (Model #1 in Table 1) with respect to models that utilize choice sets generated randomly. Five additional random choice sets are generated and used to develop five corresponding MXL models (Model #2 – Model #6). Results show that MXL models developed using randomly generated choice sets provide poor model fitness values (R^2: 0.00016 – 0.00074). R^2 for Model #2-Model #6 are significantly lower as the sample size is relatively small (320 observations) and random selection of alternatives adds difficulties to establish a strong connection between dependent and independent variables.

4.3.3 Results of Shopping Destination Choice Model

The shopping destination choice model (Model #1) in this study tested all factors suggested by RF modeling. Factors broadly include business establishment and socioeconomic characteristics, travel, and network attributes, and built environment. Table 4.2 presents the results of shopping destination location choice model.

Table 4. 2 Comparison and Parameter Estimation of MXL Models

a) Comparison of MXL models: ML-based location selection and random choice sets

Modelling approach	Choice set#	Model#	Log likelihood function	AIC	R^2
ML-based location selection and destination choice modeling	Choice set #1 (ML-based)	Model #1	-502.41	1020.8	0.26
Random location selection and destination choice modeling	Random choice set #2	Model #2	-663.36	1344.7	0.00016
	Random choice set #3	Model #3	-662.99	1344	0.00071
	Random choice set #4	Model #4	-663.17	1344.4	0.00044
	Random choice set #5	Model #5	-662.98	1344	0.00074
	Random choice set #6	Model #6	-663.09	1344.2	0.00056

b) Results of the ML-based MXL model (Model #1) for shopping location choice

Variables	Coefficients	t-stat
Business Establishment and Socioeconomic Characteristics		
*Sales volume * household annual income ≥ $74,999*	-0.1017**	-2.25
*Routine shopping * CBD*	-0.00011*	-1.92
Travel and Network Attributes		
^{R}*Travel time (continuous)*	-0.3150***	-6.63
CBD household annual income ≤ $34,999*	-0.00011	-0.28
Distance to the nearest business park (continuous)	0.00079	1.37
Built Environment		
Land use index (continuous)	1.3499**	2.29
^{R}*Retail concentration within 300 m of the alternative destinations (continuous)*	0.0330	1.50
Standard Deviation		
^{R}*Travel time to shopping locations (continuous)*	0.1182***	4.07
^{R}*Retail concentration within 300 m of the alternative destinations (continuous)*	0.2457***	4.25

^{R}Random parameters, ***, **, * refer to the significance at 1%, 5%, and 10% level.

Model results suggest that land use (coefficient = 1.3499, *t*-stat = 2.29) significantly influences shopping destination location choices indicating that people want to visit shopping units located in a mixed land use area. The reason is that the mixed land use area encompasses a wide range of shopping stores and provides an opportunity of performing multi- purpose activities and shopping (i.e., routine shopping and special item shopping). This study analyzes weekday shopping; hence, traffic congestion is a critical factor in shopping destination location choices. Results suggest that the longer the travel time the less likely people will visit a business location for shopping (coefficient = -0.3150, *t*-stat = -6.63) on weekdays. Similarly, people are unlikely to travel long distances for routine shopping (e.g., grocery) (-0.00011, *t*-stat = -1.92). People also like to shop at their desired stores located in proximity to the nearest business park (coefficient = 0.00079 and *t*-stat = 1.37. The random parameter "Retail concentration within 300 m of the alternative destinations (Standard deviation = 0.2457, *t*-stat = 4.25)" and "Travel time to destinations (Standard deviation = 0.1182, *t*-stat = 4.07)" are found significant at 99% level. This indicates the heterogeneity in shopping destination choices across populations. For example, people may willingly shop at a store in an area with a high concentration of retail stores within 300m of their shopping destinations, while there are people who may not intend to visit those locations. This can be referenced with another finding which postulates that people with a high household income (≥ $74,999) are likely to shop at the store with lower sales volume. The reason can be argued that high income people reside in suburban areas, except the south-end of Halifax and prefer to meet their routine shopping needs by the local retail stores due to longer distances to other shopping locations. CBD and surrounding low to medium income area in the Halifax Peninsula are densely populated and use the nearby multiple retail stores yielding high sales volume for these shopping locations. High income groups also showcase an affordability to shop in convenience and superstore which are expensive for low/medium income population. Additionally, there is heterogeneity in the perception of travel time to shopping locations in a sense that individual may have preference for a specific shopping store/center regardless of its geographic locations or time required to arrive.

4.4 Conclusion

This study advances the modeling of shopping destination choices by tackling existing challenges in the processes of generating location choice sets and modeling location choices. It combines machine learning and discrete choice modeling to address these gaps, as evidenced by the goodness-of-fit of the developed model. The novelty lies in the efficiency of the method in handling multi-dimensional aspects of parcel-level location searching and unobserved heterogeneity in shopping behavior across individuals.

The model is tested in the Halifax region to understand how people select different shopping establishments for routine and special item shopping. Location choice set generation identifies three clusters of business establishments based on sales volume, employee size, and travel time, informed by the feature selection process within the Random Forest modeling framework. Cluster #3 comprises shopping establishments (41%), including superstores, retail, and convenience stores, with large sales volumes and employee sizes. Clusters #1 and #2 consist of small to medium-sized business establishments, accounting for 58% of total shopping establishments. Although travel time is similar for clusters #1 and #3, it is slightly higher for cluster #2, where stores for special food and items are located at a greater distance. Generated choice sets closely match the distributions of the clusters, with a small deviation of 0.5 – 10%. Results from the mixed logit model reveal heterogeneity in shopping destination choice behavior across all individuals. Variables such as travel time, land use, and proximity to central business districts and business parks are likely to influence individuals’ attitudes towards shopping location choices. Five additional mixed logit models using random choice sets show that cluster-based choice sets provide behavioral insights that enhance prediction.

However, the study has limitations. The model designed to predict shopping destination choices primarily focuses on key factors such as business establishment and built environment characteristics, travel, and network attributes. Yet, due to the utilization of a logit model estimated through random sampling of alternative business locations, the ability to estimate alternative specific constants is restricted. Moreover, while the study sheds light on several important factors influencing consumer behavior, it overlooks crucial aspects such as pricing dynamics, product quality considerations, and the evolving landscape of online shopping trends. Additionally, the socio-demographic attributes of consumers, which can significantly impact their shopping preferences and habits, are not fully accounted for. Addressing these limitations could significantly enhance the robustness and applicability of the model in predicting shopping destinations. The integration of pricing dynamics, product quality, and online shopping trends could provide a more comprehensive understanding of consumer behavior in the retail landscape. Furthermore, exploring a wider range of consumer behaviors and preferences, including socio-demographic characteristics, would provide a more holistic understanding of the intricacies involved in selecting shopping destinations, ultimately contributing to more effective decision-making processes in retail planning and development.

Nevertheless, this study fills a gap in destination choice modeling and provides a framework that incorporates multi-dimensional features in searching suitable locations and modeling shopping destination choices. The model analyzes parcel-level shopping behavior, offering specificity about individual shopping preferences and locations. Results from the model can guide decision-makers in land use planning and the design of demand and location-efficient firm development strategies.

Chapter 5

Development of Commercial Vehicle Demand Forecasting Model[4]

5.1 Introduction

Commercial vehicles play a vital role in urban transportation, facilitating the movement of goods and services within cities. From delivery trucks to freight carriers, these vehicles are the backbone of urban logistics, ensuring that essential items reach businesses and consumers efficiently. This role is particularly important in a port city like Halifax, located on Canada's Atlantic coast. As the capital of Nova Scotia, Halifax boasts two container terminals and an intermodal terminal within its downtown area, generating significant truck traffic, especially during peak hours. The Port of Halifax stands as one of the nation's largest commercial ports, serving as the Atlantic gateway to Canada and benefiting from its ice-free harbor, modern infrastructure, and top-notch security (Bonney, 2013).

However, commercial vehicles also have a substantial impact on traffic networks (Singh and Santhakumar, 2022). Their presence can lead to increased congestion, especially during peak hours and in urban areas where traffic volumes are already high. This congestion can result in delays, longer travel times, and reduced overall efficiency of the transportation system. Despite rail transportation playing a role in freight movement, trucks remain the primary

[4] This chapter is adapted from:

Mahmud, N., Arunakirinathan, V., and Habib, M. A. (2024). Contextual Modification of Quick Response Freight Manual Trip Generation Models: A Machine Learning Approach. Accepted for presentation at the 2024 World Symposium on Transport and Land Use Research, Bogota, Columbia, June 17-20.

mode for transporting commercial goods, significantly impacting local traffic congestion and environmental quality. Moreover, the size and weight of commercial vehicles can put additional strain on road infrastructure, leading to increased maintenance costs and the need for more frequent repairs (Lightstone et al., 2021). This can further disrupt traffic flow and inconvenience other road users, exacerbating operational challenges, especially during peak hours when commercial vehicles mingle with commuter traffic.

Additionally, commercial vehicles are significant contributors to greenhouse gas (GHG) emissions due to their reliance on fossil fuels, primarily diesel and gasoline. These emissions come from the combustion of fuel in the vehicle's engine and include carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), which are major contributors to climate change (Environment and Climate Change Canada, 2022). Addressing the environmental impact of commercial vehicles is crucial for mitigating climate change and ensuring sustainable transportation systems in cities like Halifax.

However, most of the existing integrated urban model tends to overlook the incorporation of commercial vehicles, primarily due to the scarcity of freight data. This oversight can hinder the accuracy and comprehensiveness of these models, as commercial vehicles play a crucial role in urban transportation systems. The lack of reliable freight data results primarily from the high costs and time required for conducting freight surveys. This constraint has impeded the incorporation of a commercial vehicle demand forecasting model into the integrated transportation, land use, and emissions model. Thus, the development of such a model is crucial for Nova Scotia, considering the unique features of this region.

Therefore, this chapter introduces a commercial vehicle demand forecasting model, with a particular focus on the trip generation model taking into account the agglomeration effects of business establishments within Halifax's transportation network. This commercial vehicle demand forecasting model is validated using traffic count data collected from Halifax Regional Municipality (HRM).

5.2 Literature Review

The traditional commercial vehicle travel demand forecasting model comprises commercial trip generation, distribution, and traffic assignment components. In contrast to the mode choice model utilized in passenger travel demand forecasting, the commercial vehicle travel demand forecasting model bypasses the mode choice aspect. Instead, it directly estimates the origin-destination trip table by employing trip generation rates and trip distribution models at the Traffic Analysis Zone (TAZ) level. By employing distinct trip rates for various types of trucks, the need for a mode split step is automatically obviated (Federal Highway Administration, 2007).

Freight trip generation (FTG) refers to the number of commercial vehicle trips generated or attracted by a business establishment (Bastida and Holguín-Veras, 2009; Holguín-Veras et al., 2013). It comprises two main elements: freight trip attraction (FTA) and freight trip production (FTP). FTG depends on land use, economic activity, and company attributes such as employment and business area (Brogan, 1980; Cambridge Systematics Inc., 1996; Iding et al., 2002; Lawson et al., 2012). To facilitate the integration of commercial vehicle travel demand forecasting models for practitioners, several valuable compilations of freight trip generation have been developed. For instance, the ITE’s Trip Generation model estimates truck traffic as a fixed ratio of overall traffic (Trip Generation, 2008). The Quick Response Freight Manual (QRFM) offers commercial vehicles trip generation models to ensure the inclusion of commercial vehicles demand in planning models (Cambridge Systematics Inc., 1996). This approach provides a more comprehensive evaluation of transportation needs, thereby supporting sustained growth.

The accuracy of commercial vehicle travel demand forecasting model relies significantly on freight trip generation models. To address the scarcity of freight data, primarily resulting from the high costs and time required for conducting freight surveys, researchers have taken initiatives. One such effort involves the spatial transferability (the ability of a model to be applied effectively across different spatial contexts or locations) of FTG models, although their applicability to new regions lacks reliable estimation (Balla and Sahu, 2023; Venkadavarahan and Marisamynathan, 2023). Effective transferability requires data analysis from diverse sources and the use of sophisticated analytical methodologies. Holguín-Veras et al. (2013) investigated the transferability of a freight trip generation model across different contexts. They utilized zonal employment data per industry segment and Standard Industrial Classification (SIC) codes to estimate freight trip generation for various vehicle categories, drawing from the QRFM. To improve the model's transferability, they devised a synthetic correction procedure. This procedure aimed to broaden the transferability of fixed freight trip generation rates, enabling their utilization across industries where freight trip generation does not necessarily correlate directly with business size. However, their transferability approach does not consider the agglomeration effects of businesses.

The clustering of economic activities, known as agglomeration, has a notable impact on truck trips. This influence arises from heightened demand for goods and services within concentrated areas, necessitating increased freight transportation. Additionally, the close proximity of suppliers and customers in these regions leads to more efficient truck trips due to shorter distances. Furthermore, agglomeration encourages businesses to strategically optimize their supply chains, selecting locations that minimize transportation costs. Overall, agglomeration shapes truck trips by driving demand, fostering proximity-based efficiencies, and guiding strategic supply chain decisions to reduce overall transportation expenses. Hence, the primary research question of this study is how to develop a systematic method for the transferability of FTG models that can accommodate the effects of agglomeration of businesses. This development aims to improve the estimation and prediction accuracy of transferability, which is essential for the development of a comprehensive commercial vehicle demand forecasting model.

The research question mentioned above inspires this study to introduce a machine learning (ML)-based clustering method to reflect the effects of agglomeration of business to determine the commercial vehicles trip generation models. Therefore, this study aims to develop a systematic approach to investigate the process of estimating commercial vehicle trip generation models. By addressing this objective, this study aims to develop a commercial vehicle travel demand forecasting model in the context of the Halifax Regional Municipality (HRM).

This study makes a significant contribution to the transportation literature by addressing one of the critical issues in commercial vehicles travel demand forecasting model. It enhances the transferability of commercial vehicle trip generation models. Specifically, it examines various business establishments and refines commercial vehicles trip generation models to account for the agglomeration phenomenon within businesses. The strength and uniqueness of this study lie in its proposed framework, which effectively integrates ML techniques to capture the agglomeration phenomena of business establishments. A noteworthy contribution is the adoption of an unsupervised ML algorithm to address the multi-dimensionality of influential factors in the clustering of businesses related to agglomeration. This study utilized a K-Prototype clustering algorithm to identify clusters of businesses based on multidimensional built environments and neighborhood characteristics. Following the clustering process, the commercial vehicle trip generation models are adjusted to incorporate the agglomeration effects of businesses, thereby improving the accuracy of transferability.

The remainder of this chapter is organized as follows: In section three, the modeling approach is presented. Section four introduces the study area and relevant data. Following that, section five outlines the calculation of truck trip generation. Section six provides validation of the commercial vehicle demand forecasting. The subsequent section reveals the findings of the traffic assignment, along with a thorough discussion. Finally, the chapter concludes by summarizing key findings and suggesting directions for future research.

5.3 Modelling Approach

To develop the commercial vehicle demand forecasting model, this study has employed a trip-based modeling approach, which consists of three main components: commercial vehicle trip generation, commercial vehicle trip distribution, and traffic assignment (Figure 5.1). The developed commercial vehicle demand forecasting model has been integrated with the Activity-Based Travel Demand model of the Halifax Regional Municipality. It offers daily traffic volume categorized by vehicle type, including light, medium, and heavy trucks.

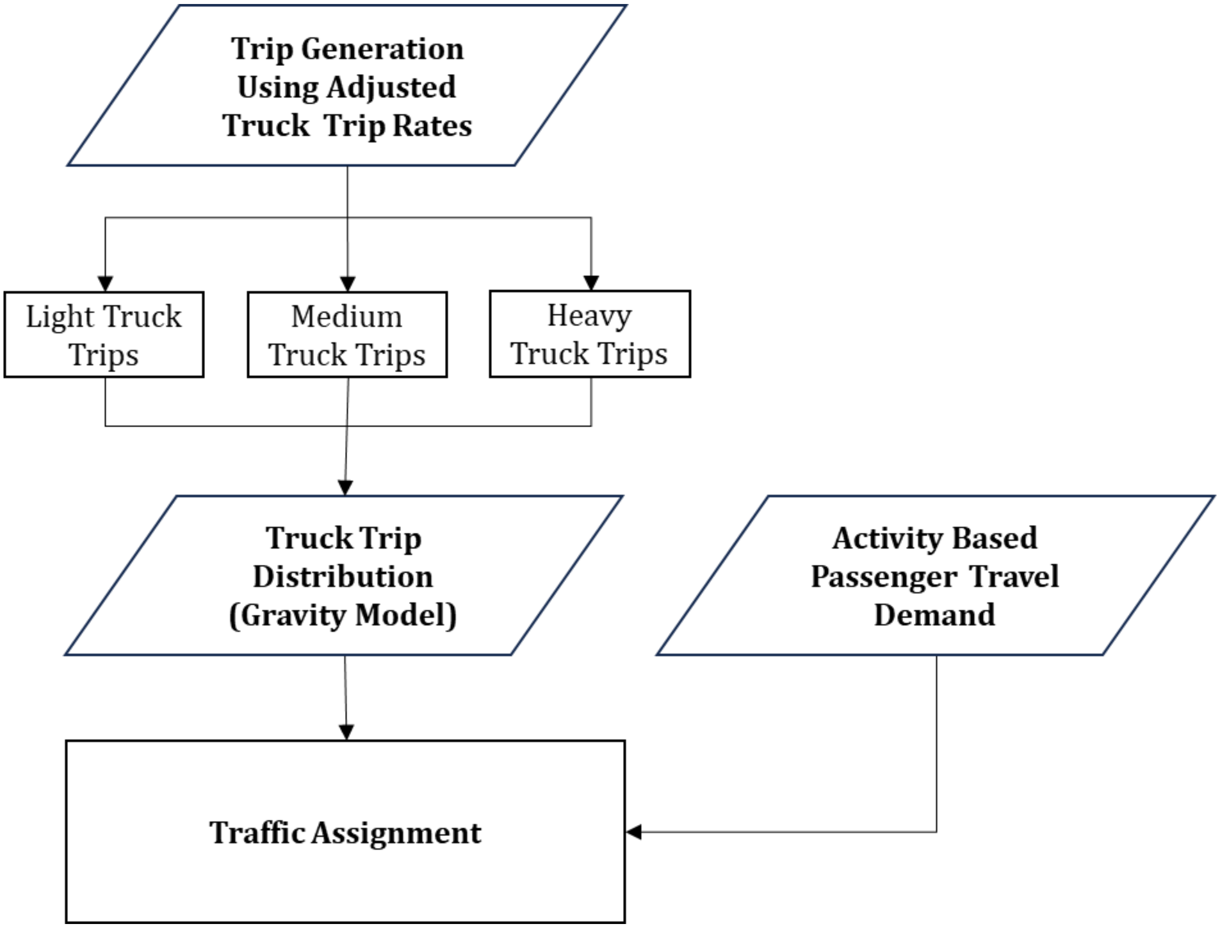
Trip Generation Using Adjusted Truck Trip Rates
Light Truck Trips
Medium Truck Trips
Heavy Truck Trips
Truck Trip Distribution (Gravity Model)
Activity Based Passenger Travel Demand
Traffic Assignment

Figure 5.1 Commercial Vehicle Demand Forecasting Modeling Framework

Truck Trip Generation

A framework based on machine learning is developed to determine the truck trip rates taking into account the agglomeration effects of businesses, as illustrated in Figure 5.2. The process initiates by collecting data, encompassing variables like zonal entropy, employment, population density, and truck trips for each type of business establishments from QRFM. Subsequently, the K-Prototype clustering algorithm is employed to categorize the data based on both categorical and numerical features. To refine this clustering process, the optimal number of clusters is determined through techniques like Bayesian Information Criterion (BIC) and Silhouette Score (SS). Each data point is then assigned to its respective cluster. Following this, mean values for zonal entropy, employment, and population density are computed within each cluster, forming the foundation for subsequent adjustment factor calculations. Reference values, such as overall means across all clusters, are estimated.

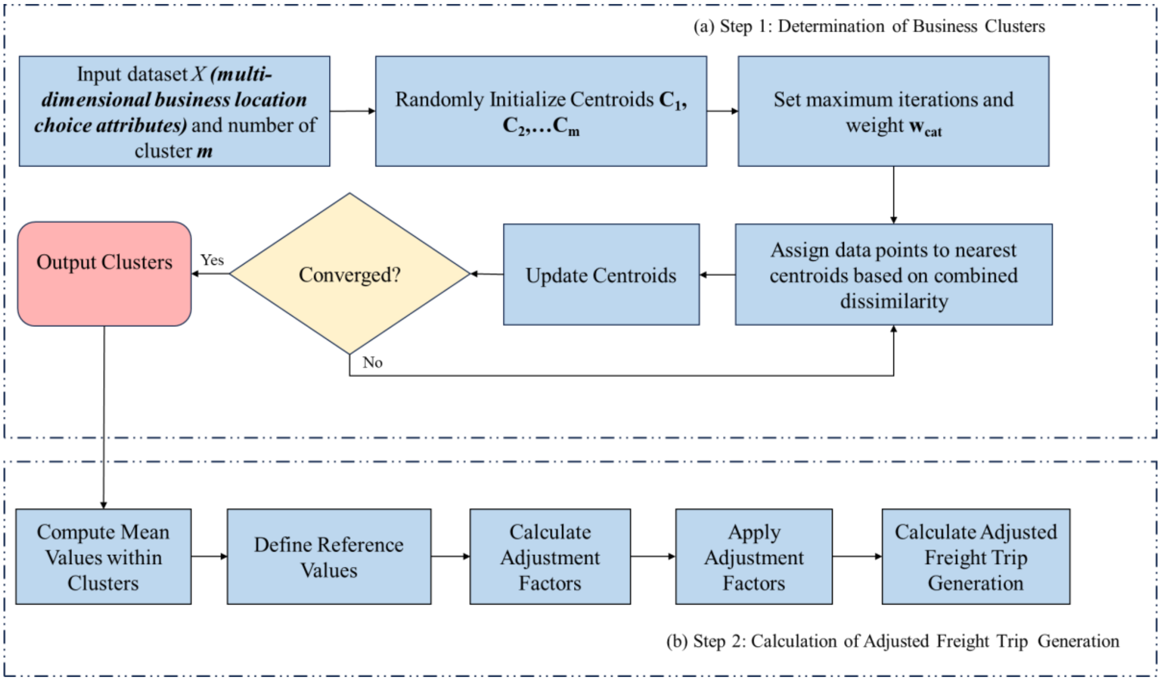
(a) Step 1: Determination of Business Clusters
Input dataset X (multi-dimensional business location choice attributes) and number of cluster m
Randomly Initialize Centroids C_1, C_2,…C_m
Set maximum iterations and weight w_{cat}
Assign data points to nearest centroids based on combined dissimilarity
Update Centroids
Converged?
Yes
No
Output Clusters
Compute Mean Values within Clusters
Define Reference Values
Calculate Adjustment Factors
Apply Adjustment Factors
Calculate Adjusted Freight Trip Generation
(b) Step 2: Calculation of Adjusted Freight Trip Generation

Figure 5.2 Conceptual Framework for Adjustments ff QRFM Truck Trip Rate

Adjustment factors are then calculated by evaluating the ratio of the mean values to the reference values for each variable. These adjustment factors capture how specific cluster characteristics deviate from the overall averages. Finally, the adjustment factors are applied to the truck trip rates for each business type within each cluster, resulting in more refined estimates that account for the unique characteristics identified through clustering. The first step of the framework is implemented in Chapter 2.

Truck Trip Distribution

After collecting production and attraction values for each type of truck, the next step involves constructing an origin-destination (O-D) matrix tailored to the specific characteristics of each truck type. This matrix serves as a fundamental tool for understanding the flow of goods and services between different locations. The gravity model, a widely used method in transportation planning, forms the basis for creating these O-D matrices. This model operates on the principle that the flow of trips between two locations is directly proportional to the product of their production and attraction values, while inversely proportional to the distance between them.

Traffic Assignment

After preparing the origin-destination (O-D) matrices for commercial vehicles, a multiclass traffic assignment is conducted using the user equilibrium assignment principle within the developed transport network model. Passenger demand is derived from the activity-based passenger travel demand model for HRM. The user equilibrium multiclass traffic assignment principle is solved using a standard method, aiming to minimize overall travel time across congested road links. This principle ensures that congested links discourage additional travelers from using them, given the availability of alternative routes. Therefore, the technique iteratively computes both link flow and cost to establish equilibrium conditions within the network.

5.4 Study Area and Data

This study is focused on Halifax Regional Municipality (HRM), Nova Scotia, Canada. Halifax plays a pivotal role as a port linking shipping companies to over 150 countries. Oversight of the primary container ports, the Fairview Cove Terminal (70 acres) and the South End Terminal (74.5 acres), is under the purview of the Halifax Port Authority (HPA). Major truck routes and CN Rail lines link these ports.

Almost all roads allow access to private vehicles, but restrictions are placed on truck traffic. Some road sections entirely prohibit trucks, while others impose nighttime bans (from 9 pm to 7 am). A few major connector roads permit unrestricted truck access at all times. The study area encompasses 219 Traffic Analysis Zones (TAZs), with 93 representing urban areas, 93 representing suburban areas, and 33 representing rural areas. Halifax's downtown core functions as a hub for commercial, residential, and office spaces. Surrounding this core are suburban areas, mainly residential with some industrial and commercial zones. As one moves away from the Halifax peninsula, rural areas become more prevalent. The transport network model includes three external TAZs located in Truro, Windsor, and Bridgewater for long-haul truck movements.

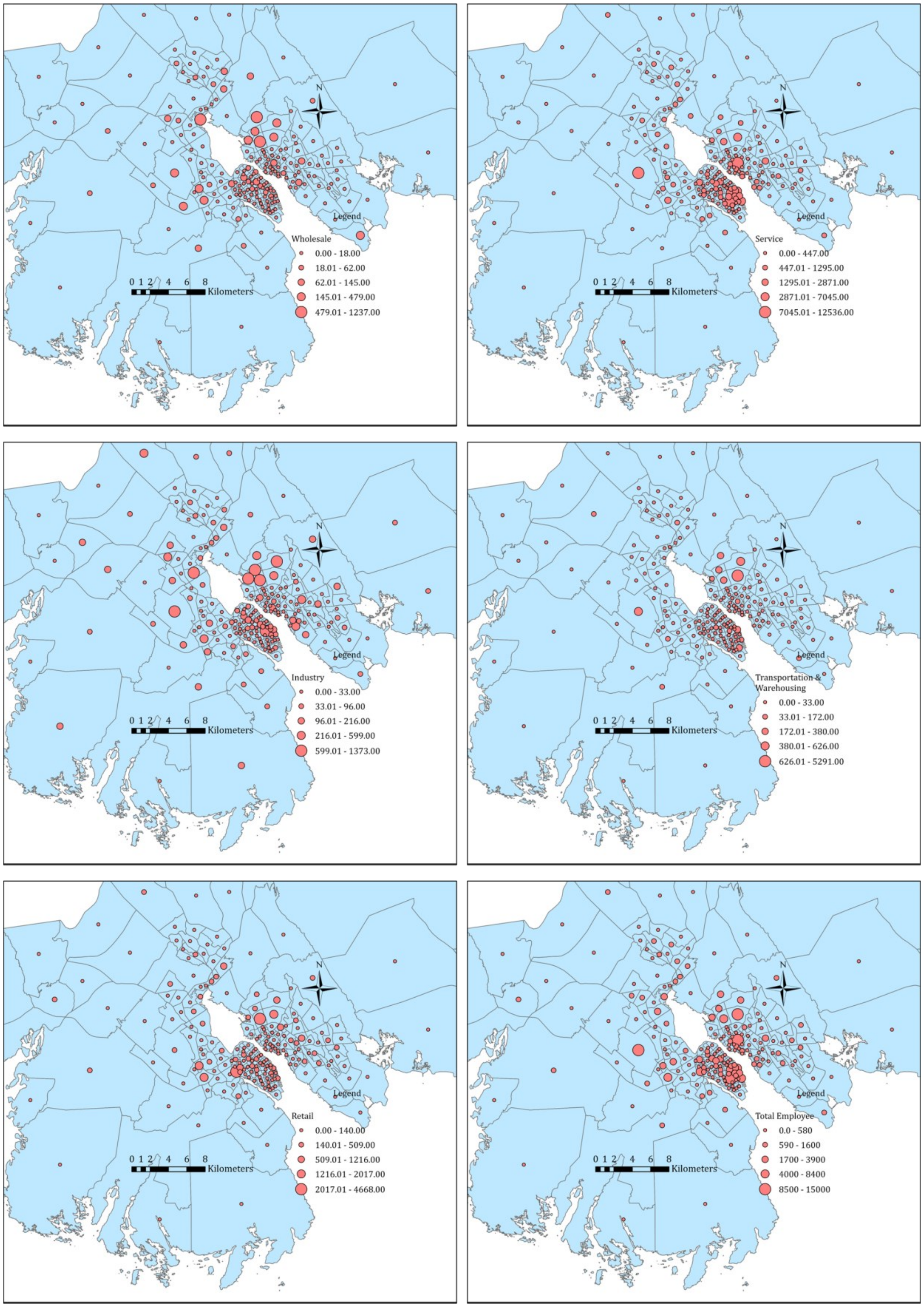

Figure 5.3 Employee Distribution of Various Business Establishments in (HRM)

To construct the commercial vehicle demand model, this study utilizes an extensive dataset of business establishments, obtained from Data Axle. Details of the dataset and study area are described in Chapter 2. Figure 5.3 displays the employee distribution among various business establishments in HRM.

5.4.1 Summary of Quick Response Freight Manual (QRFM)

To address the absence of freight in transportation models, the Quick Response Freight Manual (QRFM) was developed (Cambridge Systematics Inc., 1996). Due to the limited number of studies incorporating freight into the traffic demand model, the manual seeks to offer comprehensive information about the freight transportation system and the variables that impact the freight demand.

Table 5.1 QRFM Truck Trip Generation Rate (parenthesis numbers indicate 2-digit NAICS code)

Generation Variable (Number of Employments/Households)	**Light Trucks (Four-Tire Trucks)**	**Medium Trucks (Single Unit Trucks 6+ Tires)**	**Heavy Trucks (Combination Trucks)**
Agriculture, Mining, and Construction (11, 21, 23)	1.11	0.289	0.174
Utilities, Manufacturing, Transportation/ Communications/ and Wholesale (22, 42, 31, 32, 33, 48, 49)	0.938	0.242	0.104
Retail (44)	0.888	0.253	0.065
Office and Services (44, 51-56, 61, 62, 71, 72, 81, 92)	0.437	0.068	0.009
Households	0.251	0.099	0.038

The QRFM comprises a set of Freight Trip Generation (FTG) models that offer planners straightforward procedures and transferable parameters, which can be used to develop commercial vehicle demand forecasting model. This comprehensive approach enhances the presence of freight considerations in transportation planning. Freight Trip Generation (FTG) models in the QRFM utilize zonal employment data categorized by industry segments, employing the Standard Industrial Classification (SIC) as the classification system. Table 5.1 shows the FTG rates for various vehicle classes destined for a traffic analysis zone.

5.5 Estimation of Truck Trip Generation

In order to determine the truck trip generation rates for various business types, this chapter has relied on the findings from Chapter 2. The procedure is described in detail in the conceptual framework (Figure 5.2).

A sample calculation illustrating the adjustment of Truck Trip Generation Rates is provided with a specific focus on Cluster 3. This cluster demonstrates distinct mean values for zonal entropy, employment, and population density, recorded at 0.157502, 408.2290122, and 2424.992406, respectively. To establish a benchmark, overarching reference values representing the overall mean values across all clusters are considered: 0.191381 for zonal entropy, 423.8329 for employment, and 2476.495 for population density.

$$\text{Zonal Entropy Adjustment Factor} = \frac{\text{Mean (Zonal Entropy)}}{Reference_{Zonal\ Entropy}} = \frac{0.157502}{0.191381} = 0.8227$$

$$\text{Employment Adjustment Factor} = \frac{\text{Mean (Employment)}}{Reference_{Employment}} = \frac{408.2290122}{423.8329} = 0.9631$$

$$\text{Population Density Adjustment Factor} = \frac{\text{Mean (Population Density)}}{Reference_{Population\ Density}} = \frac{2424.992406}{2476.495} = 0.9792$$

The computation of adjustment factors involves assessing the ratios between the mean values and their corresponding reference values. this yields the zonal entropy adjustment factor at 0.8227, the employment adjustment factor at 0.9631, and the population density adjustment factor at 0.9792. following this, the application of these adjustment factors to the QRFM freight trip rate for retail trade in cluster 3, which was initially established at 0.888 trips per employee, yields an adjusted rate of 0.8227 * 0.9631 * 0.9792 = 0.78 for four-tire trucks. This adjusted rate provides insight into how cluster-specific characteristics influence the freight trip generation rate for retail trade within Cluster 3. Table 5.2 shows the adjustment factor and adjusted trip rate for other establishment types.

Table 5.2 Adjusted Trip Rate for Different Establishment Types

Establishment Types	**Adjustment Factor**	**Adjusted Trip Rate**		
		Four-Tire Trucks	Single Unit Trucks (6+ Tires)	Combination Trucks
Public Administration	0.97	0.42	0.07	0.01
Professional, Scientific and Technical Services	0.97	0.42	0.07	0.01
Information and Cultural Industries	0.97	0.42	0.07	0.01
Arts, Entertainment and Recreation	0.97	0.42	0.07	0.01
Management of Companies and Enterprises	0.97	0.42	0.07	0.01
Real Estate and Rental and Leasing	0.97	0.42	0.07	0.01
Utilities	0.97	0.91	0.23	0.10
Wholesale Trade	0.97	0.91	0.23	0.10
Manufacturing	0.97	0.91	0.23	0.10
Construction	0.97	1.08	0.28	0.17
Transportation and Warehousing	0.97	0.91	0.23	0.10
Mining and Oil and Gas Extraction	0.97	1.08	0.28	0.17
Agriculture, Forestry, Fishing and Hunting	0.97	1.08	0.28	0.17
Administrative and Support, Waste Management, and Remediation Services	1.05	0.46	0.07	0.01
Accommodation and Food Services	1.05	0.46	0.07	0.01
Health Care and Social Assistance	1.05	0.46	0.07	0.01
Finance and Insurance	1.05	0.46	0.07	0.01
Educational Services	1.05	0.46	0.07	0.01
Other Services (Except Public Administration)	0.78	0.34	0.05	0.01
Retail Trade	0.78	0.69	0.20	0.05

Adjustments of Special Generators

The ports, consisting of two container terminals and an intermodal terminal, serve as significant drivers of commercial vehicle activity in Halifax. A study conducted by Marinova

Consulting Ltd. provided data on the average daily number of standard containers transported between the intermodal terminal and ports. The Cargo Statistics of Halifax Port Authority (HPA) for 2011 disclosed the cargo volume handled by the port of Halifax. This cargo volume is converted into the number of trucks using guidelines from the quick response freight manual. These truck trips are then allocated within the container terminals based on their cargo throughput capacity. Truck trips originating from or destined for these special generators are combined with previously calculated trip generation values for the corresponding Traffic Analysis Zones (TAZs).

5.6 Validation of the Commercial Vehicle Demand Forecasting Model

To assess the effectiveness of the developed commercial vehicle trip generation models, this study conducts multiclass traffic assignment within the Regional Transport Network Model developed for Halifax. Figure 5.4 and Figure 5.5 illustrate the distribution of zonal truck trip production. The results highlight the significant role played by two ports, the intermodal terminal, Burnside, and Bayer's Lake as major truck trip generators. Moreover, the Halifax peninsula and the core area of Dartmouth contribute significantly to trip generation. Within HRM, 17% of port truck trips are connected to Burnside, while 43% of total truck trips originate from or are destined outside HRM.

Furthermore, the high volume of passenger cars in suburban areas suggests commuting traffic towards the downtown core, alongside truck traffic. Rural areas exhibit a lower population compared to urban and suburban areas, resulting in a lower number of passenger car trips. For long-haul trips, the majority of production and attraction occur at the seaport, container terminals, intermodal terminal, and airport. This spatial pattern is consistent for both the production and attraction of long-haul trips.

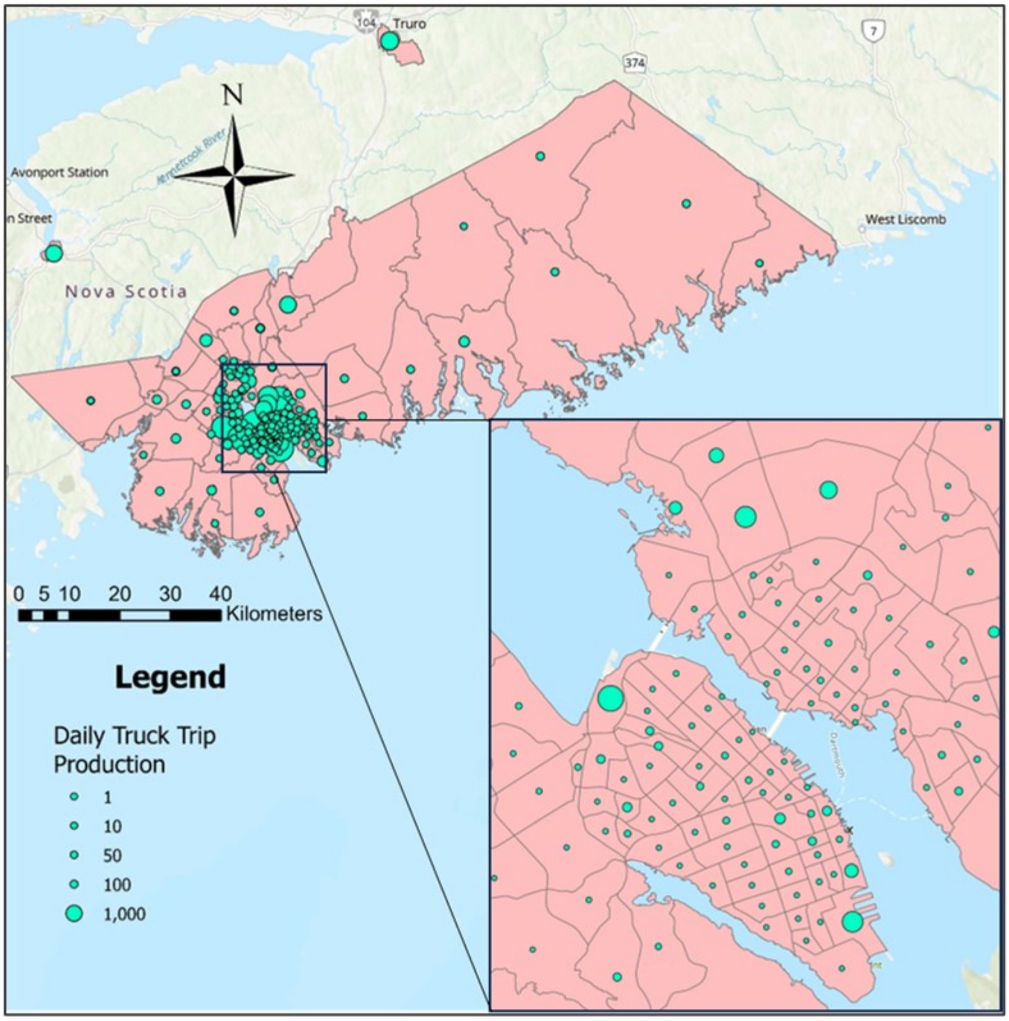

Figure 5.4 Distribution of Daily Truck Trip Production at TAZs Using QRFM Truck Trip Rates

The use of multiclass traffic assignment provides the advantage of incorporating multiple modes into a single traffic assignment, allowing for a comprehensive understanding of overall traffic congestion. The user equilibrium multiclass traffic assignment principle, a standard method, is applied to minimize the overall travel time for all available vehicles in the network, making it the preferred procedure for any transport network model. During the traffic assignment, input matrices for various vehicle classes, such as delivery trucks, long-haul trucks, and passenger cars, are collectively assigned. In this study, a daily multiclass traffic assignment is carried out using the daily passenger car volumes of HRM. To validate truck flow in the simulation model, a traffic volume-based approach is used. The comparison involves assessing the deviation between simulated and observed truck counts, with the evaluation being performed based on R^2 values.

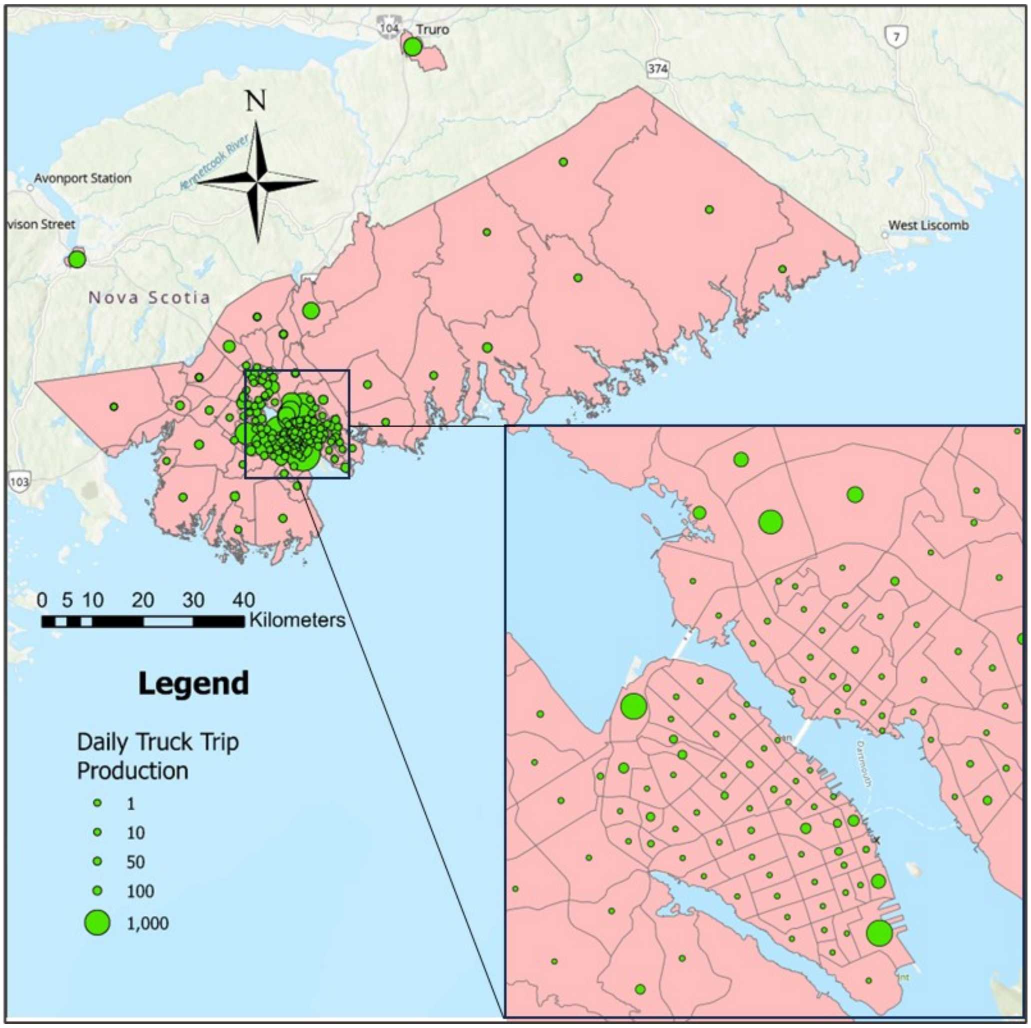

Figure 5.5 Distribution of Daily Truck Trip Production at TAZs Using Adjusted QRFM Truck Trip Rate

The R^2 value for the Quick Response Freight Manual (QRFM) stands at 0.5172 (Figure 5.6). A high R^2 value, nearing 1.0, signifies a strong correlation between observed field traffic counts and simulated data, indicating accurate replication of observed traffic count information by the model. Conversely, an R² value close to 0 suggests a weak correlation between field and simulated traffic count data. Thus, the R² value of 0.5172 in this case indicates a moderate correlation between the observed and simulated data.

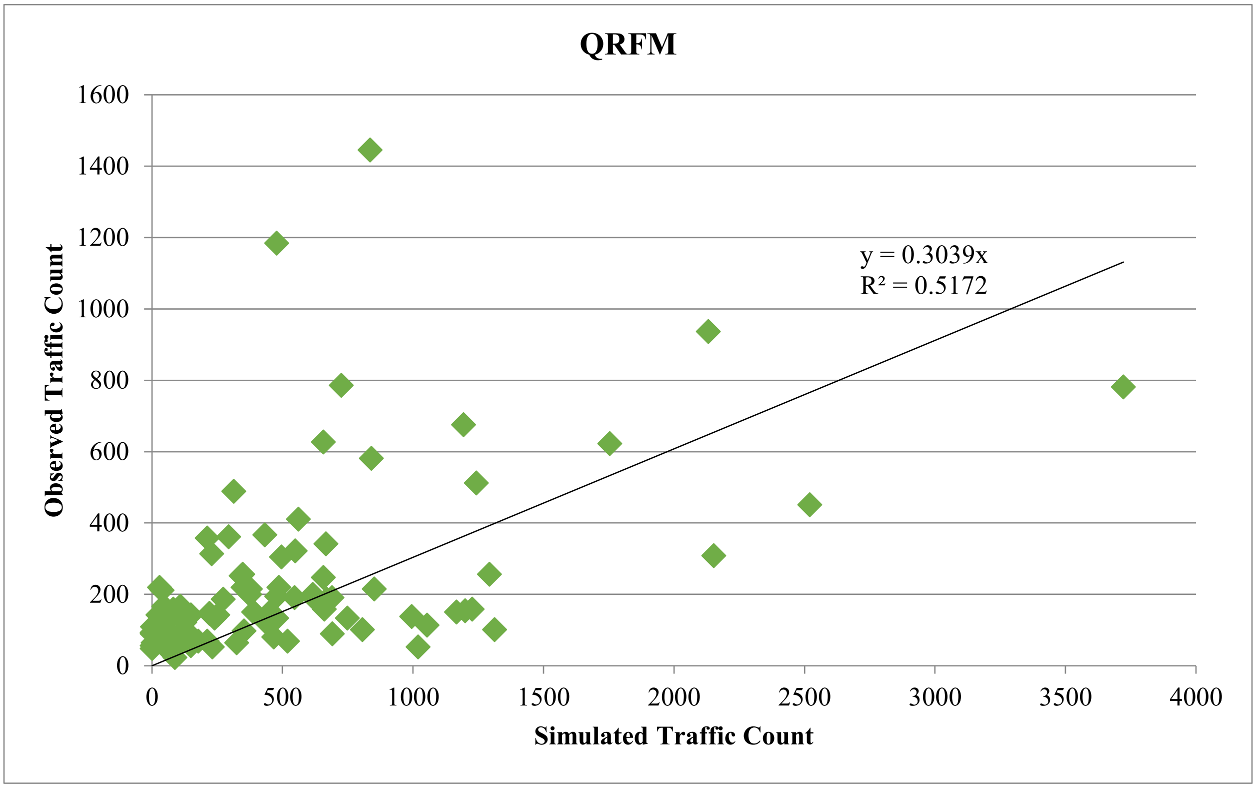

Figure 5.6 Comparison of Observed and Simulated Traffic Count Using QRFM Truck Trip Rate

On the other hand, the R² value for the adjusted QRFM is calculated as 0.8772 (Figure 5.7). Here, an R² value of 0.8772 is notably high, indicating that the statistical model utilized possesses considerable explanatory power. Approximately 87.72% of the variability in the observed data can be explained by the model. This suggests a robust correlation between the simulated values and the actual observed values, demonstrating that the model effectively captures and explains the patterns and trends present in the data. Consequently, it emerges as a dependable tool for analysis or prediction within the context of the study or analysis being conducted. Therefore, the outcomes of the validation tests illustrate that the modified freight trip generation model surpasses the original freight trip generation model.

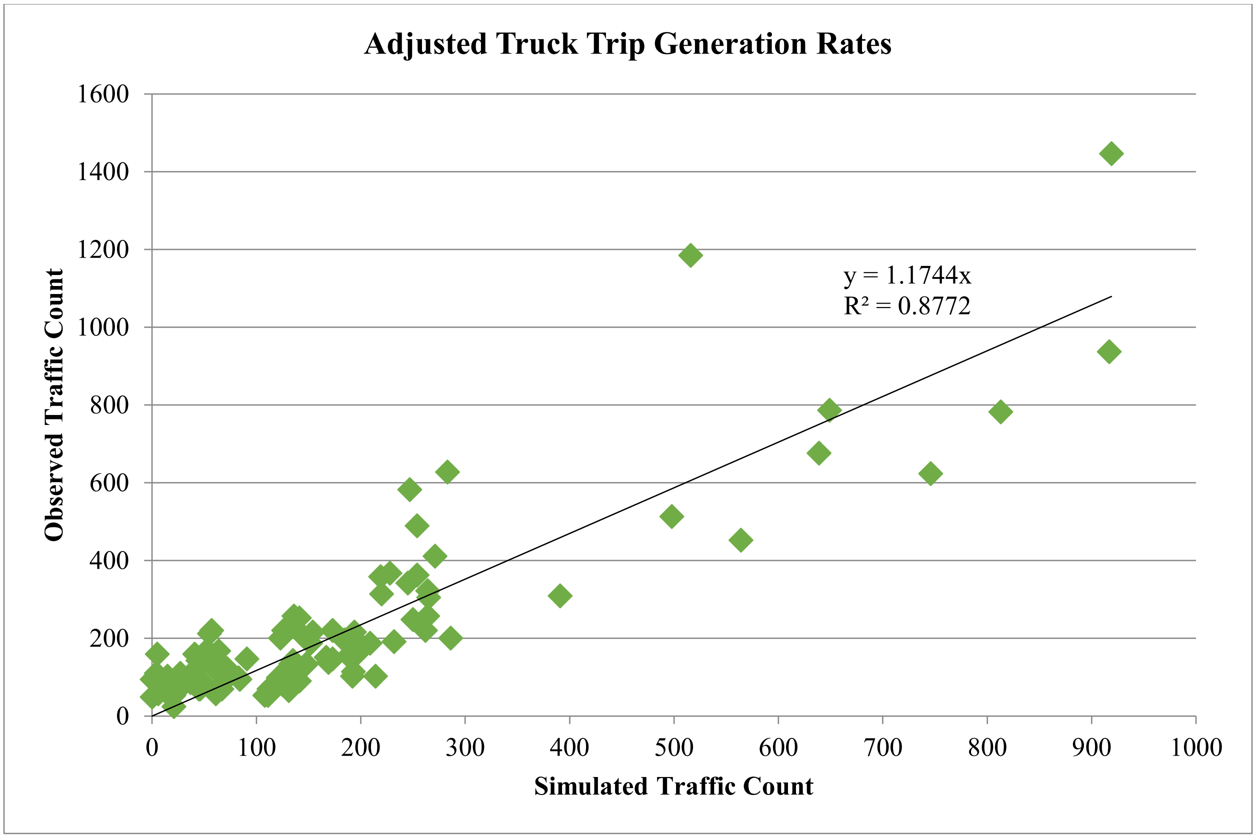

Figure 5.7 Comparison of Observed and Simulated Traffic Count Considering Adjusted Truck Trip Generation Rates

5.7 Traffic Assignment

This study examines the volume of traffic on links resulting from a multiclass traffic assignment in the Halifax transport network model. Figure 5.8 shows the total daily link volume for all vehicles after the multiclass traffic assignment in the Halifax transport network model. The results indicate that the highest vehicle flow occurs in the urban core area, particularly across two bridges connecting the twin cities of Halifax and Dartmouth. About 50% of the total traffic flows through these bridges, 27% on highways, 15% on major arterials, and 4% on other collector roads.

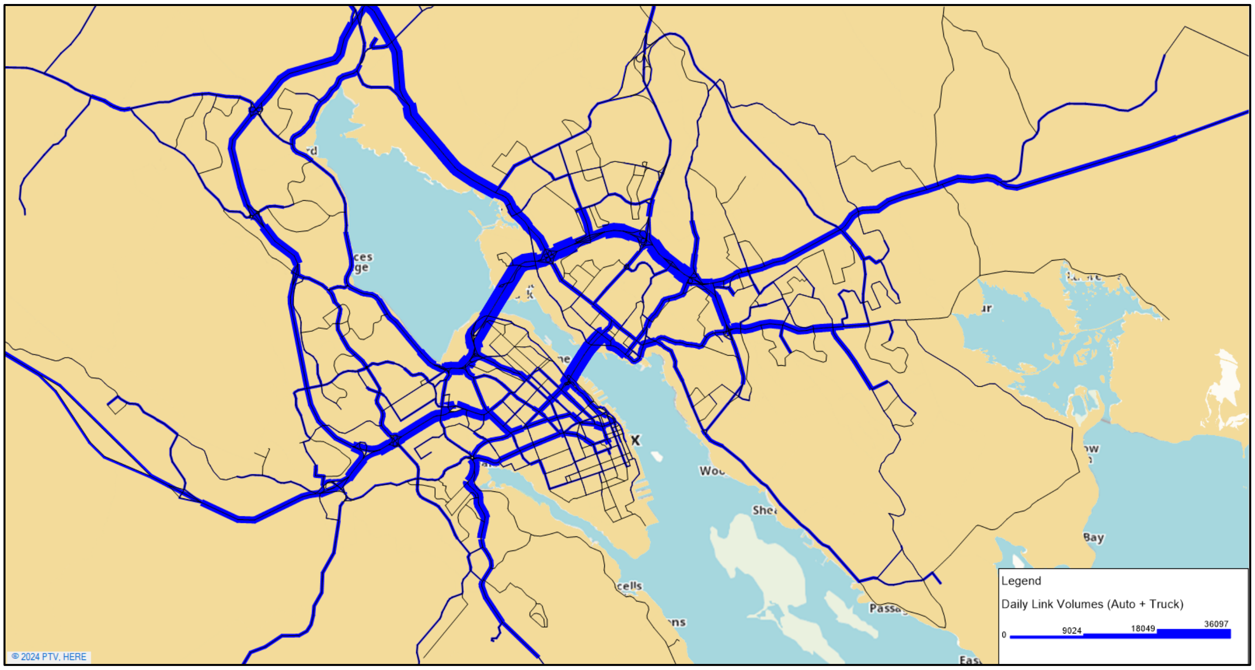

Figure 5.8 Daily Total Link Volume of All Vehicles

Figure 5.9 presents the total daily link volume specifically for trucks. Thicker lines in the diagrams represent higher implied traffic volume, while thinner lines indicate lower volume. A significant portion of heavy truck traffic pass through highways, primarily heading outwards towards the Canadian hinterlands in Quebec, Ontario, and New Brunswick, particularly via Highway 102. According to the assignment findings, approximately 65% of heavy truck movements pass through three external zones: Truro (linked to Highway 102), Windsor (linked to Highway 101), and Bridgewater (linked to Highway 103), accounting for 22% and 13% respectively. Additionally, trucks serve destinations such as Cape Breton, Prince Edward Island, other parts of Canada, and international locations through Truro, while the southwestern part of Nova Scotia is served by Windsor and Bridgewater. Only a small fraction, 12.5%, of heavy truck flow is observed within the urban core network, with the majority, 60.4%, occurring in suburban areas, which aligns with the concentration of industrial and commercial zones in those areas. The port and airport act as significant generators of long-haul truck traffic, influencing truck flows on nearby links. These results shed light on the impact of heavy truck movements on traffic congestion within the network (Figure 5.10).

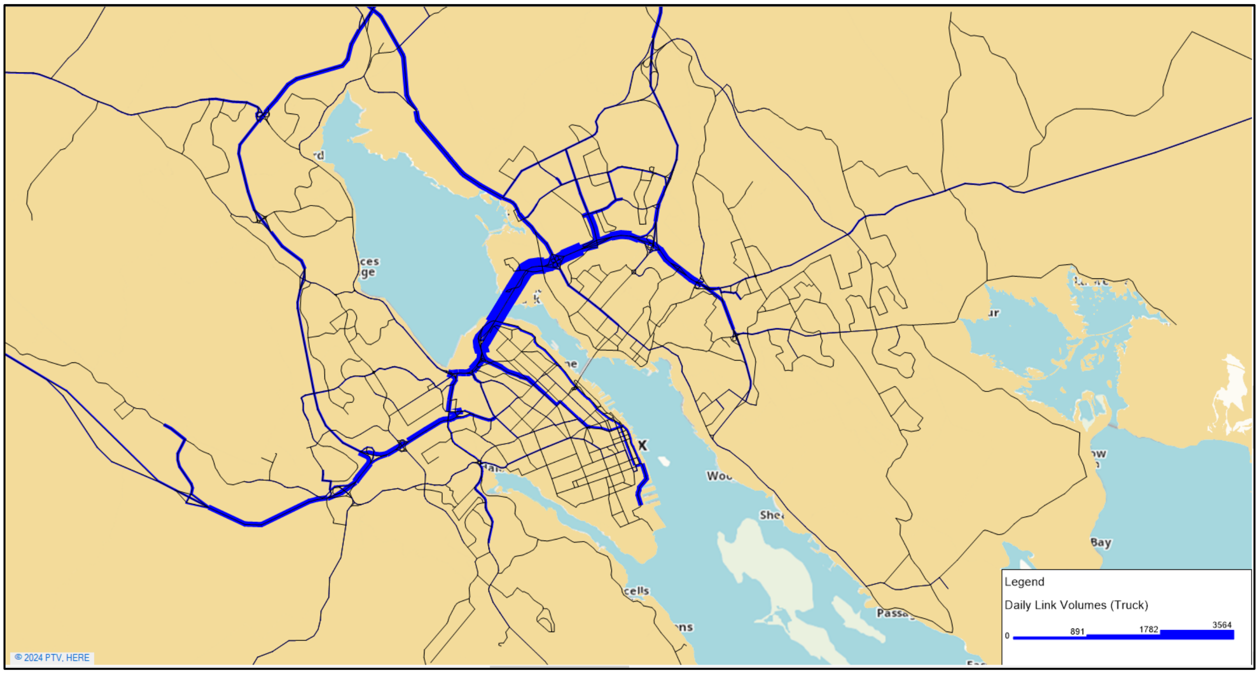

Figure 5.9 Daily Total Link Volume of All Trucks

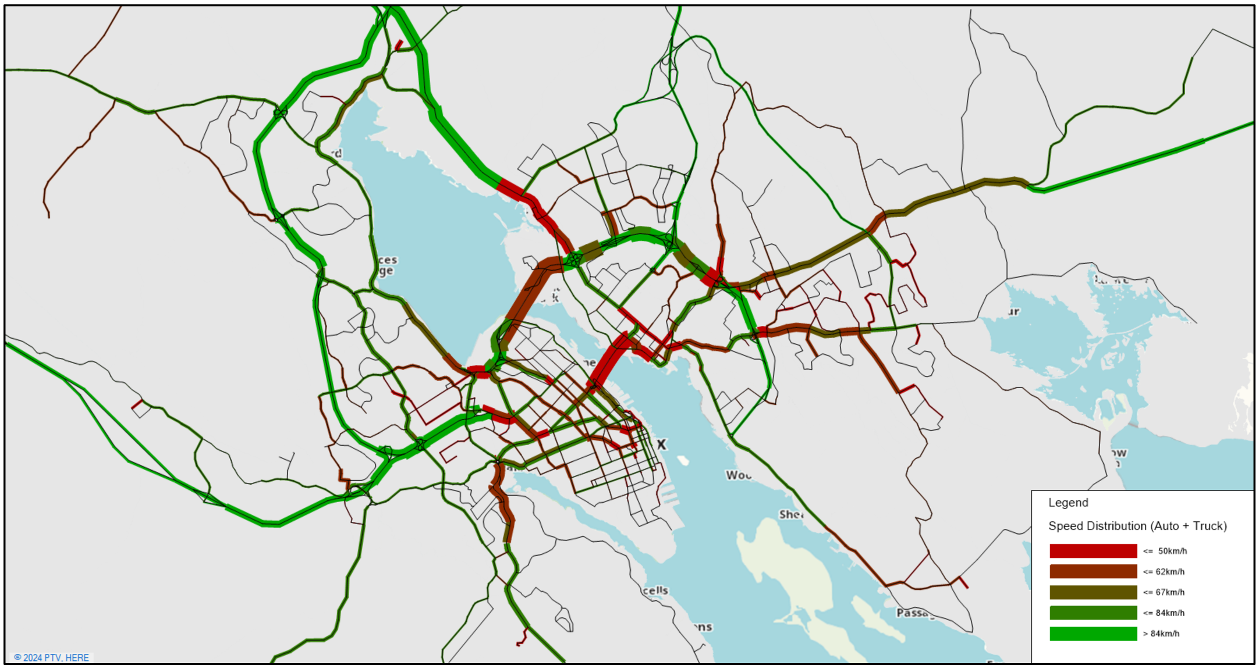

Figure 5.10 Speed Distribution of All Vehicles

5.8 Conclusion

This chapter introduces an enhanced commercial vehicle travel demand forecasting model tailored for the Halifax Regional Municipality (HRM). A significant enhancement lies in the trip generation aspect. Bela (2018) used trip generation relied on truck rates per employee across various industry categories, a common method in trip-based commercial vehicle travel demand forecasting models. However, this rate-based approach often resulted in significant over- or under-predictions of zonal trip generation, contingent upon the distribution of industry categories. Therefore, this study has opted for updated truck trip generation models that consider the clustering of business establishments, addressing this drawback. Moreover, this study has successfully addressed the challenges associated with the transferability of truck trip generation models and proposed a framework to enhance their accuracy within the context of developing a commercial vehicle travel demand forecasting model. The key innovation lies in the incorporation of the agglomeration phenomenon, where businesses cluster together, influencing dynamics of commercial vehicles. The validation results indicated that the adjusted truck trip generation model, accounting for business agglomeration, outperformed the conventional model. The high R^2 value of 0.8772 affirmed the strong explanatory power of the modified model, demonstrating its effectiveness in capturing and explaining patterns and trends in the data. This validation underscores the reliability and practical applicability of the proposed framework.

However, this study is subject to certain limitations, as it exclusively focuses on a constrained set of the built environment and neighborhood attributes, namely population density, employment number, and zonal entropy, as the influencing factors governing business establishments' clustering and agglomeration. Although these attributes significantly influence business location decisions, they do not encompass the entirety of the factors that potentially shape the spatial distribution of businesses. The built environment constitutes a multifaceted framework containing various elements, including transportation infrastructure, proximity to suppliers and customers, accessibility to amenities, adherence to land use regulations, crime rates, and environmental conditions. These additional factors might be necessary to ensure the ability of this study to offer a comprehensive understanding of business location patterns and the fundamental mechanisms driving agglomeration phenomena. It is imperative to incorporate a broader range of built environment and neighborhood attributes alongside other pertinent factors to overcome this limitation and extend the scope of this study. Including socioeconomic data, such as income levels, education levels, and consumer behavior, might offer an in-depth insight into how business establishments interact with the local population dynamics while depicting the clustering and agglomeration events. By addressing these aspects in future research, the complex dynamics of business location decisions and agglomeration patterns could be captured through robust understanding. Furthermore, this study validated the model results using traffic count data. There is an opportunity to directly verify the freight trip generation rates through an establishment-based survey.

Nevertheless, the proposed framework that reveals the freight trip generation models contribute to the existing literature, by systematically adjusting the rates utilizing agglomeration insights. The findings from this study offer valuable insights for commercial goods movement modeling, business location strategies, commercial vehicle movement modeling, and policymaking related to economic development and urban built environment.

Chapter 6

Conclusion

6.1 Summary of the Research

The importance of developing a commercial vehicle demand forecasting model is emphasized at the start of this thesis. The movement of commercial vehicles has a significant contribution to the overall GHG emissions from the transportation sector. Unlike the passenger travel demand forecasting model, the commercial vehicle movement model has received less attention in achieving the net-zero emissions goal by 2050. Since business establishments generate a substantial number of commercial trips, it is required to characterize business establishments and develop location choice models for businesses to develop a commercial vehicle demand forecasting model for Halifax Regional Municipality. Moreover, retail locations attract consumers for shopping, resulting in a substantial number of personal vehicle trips, which necessitates the development of a shopping destination choice model to enhance the activity-based passenger travel demand forecasting model. To address the challenges outlined in Chapter 1, this thesis introduces a conceptual framework that considers the characterization of business establishments and households, explores the agglomeration and location choice of businesses, develops a passenger trip attraction model for retail businesses, and estimates truck trip generation utilizing insights from the agglomeration among businesses.

Chapter 2 characterizes business establishments based on built environment and neighborhood attributes. The aim of this chapter is to investigate the patterns and agglomeration among business establishments within an economic system. This study employs an extensive business establishments dataset of Halifax Regional Municipality for the calendar year 2022 and Canadian Census 2021 information to extract built environment and neighborhood attributes. The findings obtained from this comprehensive study suggest that wholesale trade, manufacturing, construction, transportation, and warehousing demonstrate a proclivity to concentrate in a distinct geographical area characterized by spatial heterogeneity, juxtaposed with low population density and abundant employment opportunities. The findings of this study could provide valuable insights for commercial vehicle movements, integrated urban system modelling, planning, business location strategies, and policymaking concerning economic development and the growth of urban built environment.

Chapter 3 presents a comprehensive two-stage location choice framework for business establishments as part of a goods movements modeling. This study aims to formulate a systematic methodology for investigating the location choice of business establishments within Halifax Regional Municipality (HRM). The findings obtained from this comprehensive study suggest that wholesalers prioritize proximity to highways and positions within business parks for their operations while avoiding higher population density and CBD proximity. Transportation businesses seek larger sites and locations near highways, favoring clustering with related transport companies and valuing accessibility and cost-effectiveness over proximity to business parks or rural settings. The findings of this study could provide valuable insights for commercial vehicle and goods movement modeling, business location strategies, and policymaking concerning sustainable urban development.

Chapter 4 introduces a two-stage modeling framework for parcel-level shopping destination choice, taking into account multi-dimensional attributes and the diversity in shopping location preferences. The study unfolds in two main phases: (i) devising a location-search process involving feature selection and choice set generation, and (ii) constructing an econometric model to delve into individual shopping location preferences while considering unobserved heterogeneity. Results from the MMNL model indicate that as travel time and distance from the central business district increase, people are less inclined to visit stores for routine shopping (e.g., groceries). The analysis of random parameters reveals that while a high concentration of retail outlets around a desired shopping location may attract some individuals, others may still choose not to shop there. Similarly, individuals may be willing to travel longer distances to stores for special items shopping. The models developed in this study will be integrated into an integrated transport, land use, and energy (iTLE) modeling system to enhance the representation of destination choices and improve the activity-based travel demand model.

Chapter 5 provides a commercial vehicle movements model with a particular focus on commercial vehicle trip generation taking into account the agglomeration phenomenon. This study aims to assess the transferability of freight trip generation (FTG) models and proposes a framework to enhance their transferability within the context of developing a comprehensive freight demand forecasting model. The assessment involves a thorough examination of various methods for predicting commercial vehicle trip generation, drawing specific insights from the Quick Response Freight Manual (QRFM). Additionally, the study develops a systematic methodology to investigate how the concentration of business establishments, known as agglomeration, affects the transferability of freight trip generation models. The primary goal is to improve both the transferability and accuracy of these models. The study utilizes a rich business establishments dataset for the year 2022 to achieve its objectives. In addition, traffic count data from the Halifax Regional Municipality (HRM) is used to validate the developed model. The findings suggest that making systematic adjustments to account for agglomeration among businesses improves the transferability of freight trip generation models. These results have significant implications for commercial vehicle and goods movements modeling, logistics management, and policymaking related to sustainable urban development.

The thesis supports Quadruple Bottom Line (QBL) sustainability (see, Ülkü & Engau, 2021) through various means. Firstly, it offers valuable insights into the clustering patterns of businesses, facilitating economic growth by identifying areas conducive to business concentration and fostering social cohesion through the creation of employment opportunities and preservation of local cultural identities. Secondly, by analyzing business location preferences within Halifax Regional Municipality (HRM), it informs urban planning decisions that align with economic goals. Additionally, the thesis enhances economic viability by providing a shopping destination choice model that informs retail development strategies according to consumer preferences. Lastly, it contributes to environmental sustainability by developing a commercial vehicle demand forecasting model aimed at reducing congestion and emissions.

6.2 Research Contributions

This thesis contributes to the literature on land use and transportation modeling in several areas. It provides insights into agglomeration and location choices among businesses, individual shopping destination preferences, and commercial vehicle trip generation. Additionally, it explores the use of machine learning algorithms in modeling commercial vehicle movements. The thesis fills several research gaps by developing models specifically for the Halifax Regional Municipality (HRM) in Nova Scotia, Canada. A brief description of these research contributions is outlined below.

- This study presents a framework for assessing the characterization of business establishments, combining an unsupervised machine learning technique with spatial statistics methods. This approach facilitates a data-driven approach to comprehending the complex phenomenon of business establishment clustering and agglomeration based on their surrounding built environment and neighborhood characteristics. This approach possesses the potential to unveil latent patterns and associations that may remain obscured when utilizing conventional geospatial and statistical techniques. The micro-level analysis offers robust insights into the clustering tendencies of specific business types and the factors influencing their agglomeration.
- This study presents a novel approach that leverages machine learning techniques to generate a systematic choice set, thereby improving the representation of realistic and reasonable location alternatives. Info Canada Business Establishments dataset 2022 is employed to achieve the aim of this study. Combining an unsupervised machine learning technique with the mixed multinomial logit (MMNL) model facilitates a data-driven approach to enhance the precision and robustness of business establishment location choice models. This approach possesses the potential to unveil latent patterns and heterogeneity among potential choice

alternatives that may remain obscured when utilizing a conventional multinomial logit model (MNL). This thorough analysis offers robust insights into the factors influencing the location choice of business establishments.

- The study pioneers a novel approach that combines machine learning (ML) with random utility-based discrete choice modeling, specifically utilizing the mixed multinomial logit (MMNL) model for parcel-level shopping destination choice.
- The study introduces a new approach that utilizes machine learning techniques to group different businesses based on multiple characteristics, effectively illustrating the agglomeration phenomenon. Insights from these business clusters are then applied to enhance the accuracy of predictions made by truck trip generation models.

6.3 Limitations

This thesis offers significant contributions across various domains of urban commercial vehicle travel demand forecasting modeling. However, the study also has some limitations, which are outlined as follows:

- The agglomeration of businesses was examined by taking into account zonal entropy, population density, and employment numbers during clustering using k-prototype algorithms. Other factors influencing agglomeration, such as distance to the central business district (CBD), proximity to highways, and distance to business parks, could also be included in the clustering algorithm. In instances where there is a high number of dimensions (the curse of dimensionality), CLIQUE (Clustering in QUEst) could be employed, as it has the capability to handle multi-dimensional attributes.

- This study focused on a limited range of accessibility factors, such as transportation infrastructure, proximity to major roads, and highways, which shape the decisions of business establishments regarding their location. While these attributes indeed play a crucial role in business location decisions, they do not represent the full spectrum of factors influencing such choices. Integrating office and business profile factors into the location choice model for business establishments has the potential to greatly improve decision-making accuracy. By considering these additional factors, businesses can make more informed decisions about their preferred locations.
- The shopping destinations choice model overlooks crucial aspects of consumer behavior, such as pricing, quality, and online shopping. Additionally, it fails to account for the socio-demographic characteristics of consumers. Integrating these factors could enhance the comprehensiveness of the model for selecting shopping destinations. This thesis primarily focuses on attributes like travel time and land use index, indicating the need for further research to incorporate a broader range of consumer behaviors.
- This study assesses the accuracy of the truck trip generation model by comparing it to traffic count data, as business establishment/zonal level truck trip generation data is unavailable. Once this data becomes accessible, the model can be further tested against it, offering additional insights into commercial vehicle travel demand forecasting.

6.4 Recommendations for Future Research

This study has offered valuable insights into characterizing business establishments, and developing a commercial vehicle demand forecasting model, which outputs the link-based daily traffic volumes by modes, travel times and average speed. The resulting data can be utilized to estimate greenhouse gas (GHG) emissions. Immediate future work could involve the estimation of GHG emissions. There are several methods for estimating emissions using the output of the commercial vehicle demand forecasting model. One approach is to utilize the Motor Vehicle Emission Simulator (MOVES) platform (US Environmental Protection Agency, 2015). Additionally, another method involves employing the novel equation proposed by Ülkü et al. (2012) for CO_2 emissions, which considers both weight and volume efficiencies of vehicles.

To further advance the field, there is a potential to extend these findings by developing an agent-based firmography and commercial vehicle movement model. These models could be integrated with the existing Integrated Transportation, Land Use, and Energy Modelling System (iTLE) for the Halifax Regional Municipality. One notable gap in iTLE is its lack of consideration of commercial vehicle movements. Since commercial vehicles significantly contribute to overall GHG emissions, addressing this gap is vital for accurately understanding and mitigating environmental impacts. Therefore, future research endeavors may prioritize incorporating GHG emissions considerations into the iTLE framework, ensuring a more comprehensive understanding of the environmental implications of transportation and land use decisions.

References

Abraham, J. E., and Hunt, J. D. (1999). Firm Location in the MEPLAN Model of Sacramento. *Transportation Research Record: Journal of the Transportation Research Board*, *1685*(1), 187–198. https://doi.org/10.3141/1685-24

Adler, J. L. (1993). Route choice: Wayfinding in transport networks. *Transportation Research Part A: Policy and Practice*, *27*(4), 338–339. https://doi.org/10.1016/0965-8564(93)90007-8

Alam, Jahedul MD., Mahmud, N., and Habib, M. A. (2024). A Comprehensive Framework for Shopping Destination Choice Model: Combination of Machine Learning and Discrete Choice Modeling. *17th International Conference on Travel Behavior Research*.

Alcácer, J., and Chung, W. C. (2014). Location strategies for agglomeration economies. *Strategic Management Journal*, *35*(12), 1749–1761. https://doi.org/10.1002/smj.2186

An, D.D., X. Tong, K. Liu, and E. H. W. Chan. (2019). Understanding the Impact of Built Environment on Metro Ridership Using Open Source in Shanghai. *Cities* 93: 177–187. doi:10.1016/j.cities.2019.05.013.

Anas, A., R. Arnott, and K. A. Small. (1998). Urban Spatial Structure. *Journal of Economic Literature*, *36*(3), 1426–1464.

Armstrong, R. B. (1972). *The Office Industry: Patterns of Growth and Location*. MIT Press, Cambridge, Mass.

Backman, M., and Karlsson, C. (2017). Location of New Firms: Influence of Commuting Behaviour. *Growth and Change*, *48*(4), 682–699. https://doi.org/10.1111/grow.12200

Balbontin, C., and Hensher, D. A. (2019). Firm-specific and location-specific drivers of business location and relocation decisions. *Transport Reviews*, *39*(5), 569–588. https://doi.org/10.1080/01441647.2018.1559254

Balbontin, C., and Hensher, D. A. (2021). Understanding business location decision making for transport planning: An investigation of the role of process rules in identifying influences on firm location. *Journal of Transport Geography*, *91*, 102955. https://doi.org/10.1016/j.jtrangeo.2021.102955

Balla, B. S., and Sahu, P. K. (2023). Assessing regional transferability and updating of freight generation models to reduce sample size requirements in national freight data collection program. *Transportation Research Part A: Policy and Practice*, *175*, 103780. https://doi.org/10.1016/j.tra.2023.103780

Bastida, C., and Holguín-Veras, J. (2009). Freight generation models: Comparative analysis of regression models and multiple classification analysis. *Transportation Research Record*, *2097*, 51–61. https://doi.org/10.3141/2097-07

Bela, P. L. (2018). *A Framework of Multiclass Travel Demand Forecasting and Emission Modelling, Incorporating Commercial Vehicle Movement for the Port City of Halifax, Canada*. Dalhousie University.

Bela, P. L., and M. A. Habib. (2019). Development of an Urban Transport Network and Emission Model for the Port City of Halifax, Canada. *TAC-ITS Canada Joint Conference*.

Bell, D. A. (1991). Travel impacts arising from office relocation from city to suburbs. In *Transportation* (Vol. 18).

Ben-Akiva, M. E., and S. R. Lerman. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Billings, S. B., and Johnson, E. B. (2016). Agglomeration within an urban area. *Journal of Urban Economics*, *91*, 13–25. https://doi.org/10.1016/j.jue.2015.11.002

Bodenmann, B. R., and Axhausen, K. W. (2012). Destination choice for relocating firms: A discrete choice model for the St. Gallen region, Switzerland. *Papers in Regional Science*, *91*(2), 319–341. https://doi.org/10.1111/j.1435-5957.2011.00389.x

Bonney, J. (2013). *Canada's Big 4 Container Ports Put Focus on Infrastructure*. https://www.joc.com/article/canadas-big-4-container-ports-put-focus infrastructure_20130902.html

Bouzouina, L., Baraklianos, I., Bonnel, P., and Aissaoui, H. (2021). Renters vs owners: The impact of accessibility on residential location choice. Evidence from Lyon urban area, France (1999–2013). *Transport Policy*, *109*, 72–84. https://doi.org/10.1016/j.tranpol.2021.05.022

Bracken, I., and Martin, D. (1989). The Generation of Spatial Population Distributions from Census Centroid Data. *Environment and Planning A: Economy and Space*, *21*(4), 537–543. https://doi.org/10.1068/a210537

Bradley, M., Bowman, J. L., and Griesenbeck, B. (2010). SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. *Journal of Choice Modelling*, *3*(1), 5–31. https://doi.org/10.1016/S1755-5345(13)70027-7

Brogan, J. D. (1980). *Improving Truck Trip-Generation Techniques Through Trip-End Stratification*. https://onlinepubs.trb.org/Onlinepubs/trr/1980/771/771-001.pdf

Cambridge Systematics Inc. (1996). *Quick Response Freight Manual*.

Carlton, D. W. (1979). *Why New Firms Locate Where They Do: An Econometric Model*. Cambridge, MA.

Cavalcante, R. A., and Roorda, M. J. (2013). Freight Market Interactions Simulation (FREMIS): An Agent-based Modeling Framework. *Procedia Computer Science*, *19*, 867–873. https://doi.org/10.1016/j.procs.2013.06.116

Clark, C. (1951). Urban Population Densities. *Journal of the Royal Statistical Society. Series A (General)* 114 (4): 490–496. doi:10.2307/2981088.

Coffey, W. J., and Shearmur, R. G. (2002). Agglomeration and dispersion of high-order service employment in the Montreal metropolitan region, 1981-96. *Urban Studies*, 39 (3), 359–378. https://doi.org/10.1080/00420980220112739

Cuthbert, A. L., and Anderson, W. P. (2002). Using Spatial Statistics to Examine the Pattern of Urban Land Development in Halifax–Dartmouth. *The Professional Geographer*, *54*(4), 521–532. https://doi.org/10.1111/0033-0124.00347

Dadashpoor, H., and Malekzadeh, N. (2022). Evolving spatial structure of metropolitan areas at a global scale: A context-sensitive review. *GeoJournal*, 87(5), 4335–4362. https://doi.org/10.1007/s10708-021-10435-0

De Bok, M., and Sanders, F. (2005a). Firm Relocation and Accessibility of Locations. *Transportation Research Record: Journal of the Transportation Research Board*, *1902*(1), 35–43. https://doi.org/10.1177/0361198105190200105

De Bok, M., and Sanders, F. (2005b). Firm Relocation and Accessibility of Locations. *Transportation Research Record: Journal of the Transportation Research Board*, *1902*(1), 35–43. https://doi.org/10.1177/0361198105190200105

De Bok, M., and Van Oort, F. (2011). Agglomeration economies, accessibility and the spatial choice behavior of relocating firms. *Journal of Transport and Land Use*, *4*(1), 5. https://doi.org/10.5198/jtlu.v4i1.144

de Jong, G., and Ben-Akiva, M. (2007). A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B: Methodological*, *41*(9), 950–965. https://doi.org/10.1016/j.trb.2007.05.002

Department of Environment and Climate Change. (2022). *Nova Scotia's Climate Change Plan for Clean Growth Our Future*. https://climatechange.novascotia.ca/

Eisele, W. L., Schrank, D. L., Bittner, J., and Larson, G. (2013). Incorporating Urban-Area Truck Freight Value into the Urban Mobility Report. *Transportation Research Record:*

Journal of the Transportation Research Board, *2378*(1), 54–64. https://doi.org/10.3141/2378-06

Elgar, I., Farooq, B., and Miller, E. J. (2009). Modeling Location Decisions of Office Firms. *Transportation Research Record: Journal of the Transportation Research Board*, *2133*(1), 56–63. https://doi.org/10.3141/2133-06

Environment and Climate Change Canada. (2022). *Greenhouse Gas Emissions Canadian Environmental Sustainability Indicators*. https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html

Environment Canada. (2014). *Canada's Emissions Trends*. https://www.canada.ca/en/environment-climate-change/services/climate-change/publications/emission-trends-2014.html

Environment Canada. (2021). *National inventory report 1990-2019: greenhouse gas sources and sinks in Canada: Part 1*. https://publications.gc.ca/site/eng/9.506002/publication.html

Ewing, M., Kim, C., Lee, J., and Smith, C. (2020). *The next frontier for climate action*. https://www.pembina.org/pub/next-frontier-climate-action

Federal Highway Administration. (2007). *Quick Response Freight Manual II*. https://rosap.ntl.bts.gov/view/dot/67831

Fischer, M., Ang-Olson, J., and La, A. (2000). External Urban Truck Trips Based on Commodity Flows: A Model. *Transportation Research Record: Journal of the Transportation Research Board*, *1707*(1), 73–80. https://doi.org/10.3141/1707-09

Gabe, T. M., and Bell, K. P. (2004). Trade-offs between Local Taxes and Government Spending as Determinants of Business Location. *Journal of Regional Science*, *44*(1), 21–41. https://doi.org/10.1111/j.1085-9489.2004.00326.x

González-Benito, Ó. (2005). Spatial competitive interaction of retail store formats: modeling proposal and empirical results. *Journal of Business Research*, *58*(4), 457–466. https://doi.org/10.1016/j.jbusres.2003.09.001

Government of Canada. (2022). *Net-Zero emissions by 2050*. https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050.html

Grashuis, J., Skevas, T., and Segovia, M. S. (2020). Grocery Shopping Preferences during the COVID-19 Pandemic. *Sustainability*, *12*(13), 5369. https://doi.org/10.3390/su12135369

Grenzeback, L., Reilly, W. R., Roberts, P. 0, Stowers, J. R., and Grenzeback, L. R. (2000). Washing-ton, D.C. In *J. R. Stowers*. Sydec, Inc.

Gu, Y., Zheng, S., and Cao, Y. (2009). The identification of employment centers in Beijing. *Urban Development Studies*, 16(9), 118–124 (In Chinese).

Habib, M. A., and Miller, E. J. (2009). Reference-Dependent Residential Location Choice Model within a Relocation Context. *Transportation Research Record: Journal of the Transportation Research Board*, *2133*(1), 92–99. https://doi.org/10.3141/2133-10

Hansen, E. R. (1987). Industrial location choice in São Paulo, Brazil. *Regional Science and Urban Economics*, *17*(1), 89–108. https://doi.org/10.1016/0166-0462(87)90070-6

Hassan, M. N., Najmi, A., and Rashidi, T. H. (2019). A two-stage recreational destination choice study incorporating fuzzy logic in discrete choice modelling. *Transportation Research Part F: Traffic Psychology and Behaviour*, *67*, 123–141. https://doi.org/10.1016/j.trf.2019.10.015

Haynes, K. E., and Fotheringham, A. S. (1985). *Gravity and Spatial Interaction Models*. (SAGE series in Scientific Geography; Vol. 2). Sage.

Holguín-Veras, J., Jaller, M., Sánchez-Díaz, I., Campbell, S., and Lawson, C. T. (2013). Freight Generation and Freight Trip Generation Models. In *Modelling Freight Transport* (pp. 43–63). Elsevier Inc. https://doi.org/10.1016/B978-0-12-410400-6.00003-3

Holguín-Veras, J., Jaller, M., Sanchez-Diaz, I., Wojtowicz, J., Campbell, S., Levinson, H., Lawson, C., Powers, E. L., and Tavasszy, L. (2012). *Freight Trip Generation and Land Use*. Transportation Research Board. https://doi.org/10.17226/23437

Holguín-Veras, J., Sánchez-Díaz, I., Lawson, C. T., Jaller, M., Campbell, S., Levinson, H. S., and Shin, H.-S. (2013). Transferability of Freight Trip Generation Models. *Transportation Research Record: Journal of the Transportation Research Board*, *2379*(1), 1–8. https://doi.org/10.3141/2379-01

Hu, L., Sun, T., and Wang, L. (2018). Evolving urban spatial structure and commuting patterns: A case study of Beijing, China. *Transportation Research Part D: Transport and Environment*, 59, 11–22. https://doi.org/10.1016/j.trd.2017.12.007

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, *12*, 283–304.

Huang, D.Q., Z. Liu, X.S. Zhao, and P.J. Zhao. (2017). Emerging Polycentric Megacity in China: An Examination of Employment Subcenters and Their Influence on Population Distribution in Beijing. *Cities* 69: 36–45. Doi:10.1016/j.cities.2017.05.013.

Huff, D. L. (1963). A Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*, *39*(1), 81. https://doi.org/10.2307/3144521

Iding, M. H. E., Meester, W. J., and Tavasszy, L. (2002). Freight trip generation by firms. *42nd Congress of the European Regional Science Association: "From Industry to Advanced Services - Perspectives of European Metropolitan Regions."*

Institute of Transportation Engineers. (2008). *Trip Generation (8th ed.)*

Iseki, H., and Jones, R. P. (2018). Analysis of firm location and relocation in relation to Maryland and Washington, DC metro rail stations. *Research in Transportation Economics*, *67*, 29–43. https://doi.org/10.1016/j.retrec.2016.11.003

Jacobs, J. (1961). *The Death and Life of Great American Cities.* New York, NY: Random House.

Jacobs-Crisioni, C., P. Rietveld, E. Koomen, and E. Tranos. (2014). Evaluating the Impact of Land-use Density and Mix on Spatiotemporal Urban Activity Patterns: An Exploratory Study Using Mobile Phone Data. *Environment & Planning A* 46 (11): 2769–2785. Doi:10.1068/a130309p.

Jonnalagadda, N., Freedman, J., Davidson, W. A., and Hunt, J. D. (2001). Development of Microsimulation Activity-Based Model for San Francisco: Destination and Mode Choice Models. *Transportation Research Record: Journal of the Transportation Research Board*, *1777*(1), 25–35. https://doi.org/10.3141/1777-03

Khan, A. S. (2002). *A System for Microsimulating Business Establishments: Analysis, Design and Results.* Department of Civil Engineering. http://archives.ucalgary.ca

Kim, D., and Seo, B. (2014). Assessment of the number of components in Gaussian mixture models in the presence of multiple local maximizers. *Journal of Multivariate Analysis*, *125*, 100–120. https://doi.org/10.1016/j.jmva.2013.11.018

Kwon, K. (2021). Polycentricity and the role of government-led development: Employment decentralization and concentration in the Seoul metropolitan area, 2000–2015. *Cities*, 111, Article 103107. https://doi.org/10.1016/j. cities.2021.103107

Kumar, S., and Kockelman, K. M. (2008). Tracking Size, Location, and Interactions of Businesses. *Transportation Research Record: Journal of the Transportation Research Board*, *2077*(1), 113–121. https://doi.org/10.3141/2077-15

Lagonigro, R., J. C. Martori, and P. Apparicio. (2020). Understanding Airbnb Spatial Distribution in a SouthernEuropean City: The Case of Barcelona. *Applied Geography* 115: 102136. doi:10.1016/j.apgeog.2019.102136.

Lawson, C. T., Holguín-Veras, J., Sánchez-Díaz, I., Jaller, M., Campbell, S., and Powers, E. L. (2012). Estimated Generation of Freight Trips Based on Land Use. *Transportation Research Record: Journal of the Transportation Research Board*, *2269*(1), 65–72. https://doi.org/10.3141/2269-08

Lee, B. (2007). "Edge" or "edgeless" cities? Urban spatial structure in U.S. metropolitan areas, 1980 to 2000. *Journal of Regional Science*, 47(3), 479–515. https://doi.org/ 10.1111/j.1467-9787.2007.00517.x

Lee, B. H. Y., Waddell, P., Wang, L., and Pendyala, R. M. (2010). Reexamining the Influence of Work and Nonwork Accessibility on Residential Location Choices with a Microanalytic Framework. *Environment and Planning A: Economy and Space*, *42*(4), 913–930. https://doi.org/10.1068/a4291

Lee, K. S. (1982). A model of intraurban employment location: An application to Bogota, Colombia. *Journal of Urban Economics*, *12*(3), 263–279. https://doi.org/10.1016/0094-1190(82)90018-3

Li, J.M., W.Z. Zhang, H.X. Chen, and J.H. Yu. (2015). The Spatial Distribution of Industries in Transitional China: A Study of Beijing. *Habitat International* 49: 33–44. Doi:10.1016/j.habitatint.2015.05.004.

Li, X. M., J. L. Zhu, and Y. Wang. (2015). Spatial Differences of Residential Quarter Floor Area Ratio: A Case Study of Dalian. *Progress in Geography* 34 (6): 687–695. (In Chinese). doi10.18306/dlkxjz.2015.06.004.

Lightstone, A., Belony, T., and Cappuccilli, J. (2021). *Understanding Goods Movement in Canada: Trends and Best Practices*. Transportation Association of Canada (TAC).

Liu, X., Sun, T., and Li, G. (2011). Research on the spatial structure of employment distribution in Beijing. *Geographical Research*, 30(7), 1262–1270 (In Chinese).

Mahmud, N., and Habib, M. A. (2024). Characterization and Assessment of Agglomeration of Businesses Establishments: A Combination of Machine Learning and Spatial Statistics Approach. *103rd Annual Meeting of the Transportation Research Board (TRB)*.

Mahmud, N., and Habib, M. A. (2024). A Comprehensive Business Location Choice Model Leveraging Machine Learning in Systematic Choice Set. *103rd Annual Meeting of the Transportation Research Board (TRB)*.

Mahmud, N., Arunakirinathan, V., and Habib, M. A. (2024). Contextual Modification of Quick Response Freight Manual Trip Generation Models: A Machine Learning Approach. *2024 World Symposium on Transport and Land Use Research*.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3), 229–254. https://doi.org/10.1007/BF00133443

Maoh, H., and P. Kanaroglou. (2007). *Modeling the Failure of Small and Medium Size Business Establishments in Urban Areas: An Application to Hamilton, Ontario*.

Maoh, H. F. (2005). *Modeling Firm Demography in Urban Areas with an Application to Hamilton, Ontario: Towards an Agent-Based Microsimulation Model*.

Maoh, H. F., and Kanaroglou, P. S. (2005). *Agent-Based Firmographic Models: A Simulation Framework for the City of Hamilton*.

Maoh, H., and Kanaroglou, P. (2007). Geographic clustering of firms and urban form: a multivariate analysis. *Journal of Geographical Systems*, *9*(1), 29–52. https://doi.org/10.1007/s10109-006-0029-6

Maoh, H., and Kanaroglou, P. (2009). Intra-metropolitan Location of Business Establishments. *Transportation Research Record: Journal of the Transportation Research Board*, *2133*(1), 33–45. https://doi.org/10.3141/2133-04

Marutho, D., Hendra Handaka, S., and Wijaya, E. (2018). The Determination of Cluster Number at k-mean using Elbow Method and Purity Evaluation on Headline News. *2018 International Seminar on Application for Technology of Information and Communication*.

McFadden, D. (1978). Modelling the Choice of Residential Location. *Transportation Research Record*, 672, 72-77.

Newling, B.E. (1969). The Spatial Variation of Urban Population Densities. *Geographical Review* 59: 242–252. Doi:10.2307/213456.

O'sullivan, A. (2018). *Urban economics*. McGrawHill.

Orvin, M. M., and Fatmi, M. R. (2023). A residential location search model based on the reasons for moving out. *Transportation Letters*, 1–15. https://doi.org/10.1080/19427867.2023.2222990

Pfister, N., Freestone, R., and Murphy, P. (2000). Polycentricity or dispersion?: Changes in center employment in metropolitan Sydney, 1981 to 1996. *Urban Geography*, 21(5), 428–442. https://doi.org/10.2747/0272-3638.21.5.428

Phan, D. T., Vu, H. L., and Miller, E. J. (2022). A new approach to improve destination choice by ranking personal preferences. *Transportation Research Part C: Emerging Technologies*, *143*, 103817. https://doi.org/10.1016/j.trc.2022.103817

Pourabdollahi, Z., Mohammadian, A. (Kouros), and Kawamura, K. (2012). A behavioral freight transportation modeling system. *Proceedings of the 14th Annual International Conference on Electronic Commerce*, 196–203. https://doi.org/10.1145/2346536.2346574

Rashidi, T. H., and Mohammadian, A. (Kouros). (2015). Behavioral Housing Search Choice Set Formation. *International Regional Science Review*, *38*(2), 151–170. https://doi.org/10.1177/0160017612461356

Riguelle, F., Thomas, I., and Verhetsel, A. (2007). Measuring urban polycentrism: A European case study and its implications. *Journal of Economic Geography*, 7(2), 193–215.

Roorda, M. J., Cavalcante, R., McCabe, S., and Kwan, H. (2010). A conceptual framework for agent-based modelling of logistics services. *Transportation Research Part E: Logistics and Transportation Review*, *46*(1), 18–31. https://doi.org/10.1016/j.tre.2009.06.002

Rosenthal, S. S., and Strange, W. C. (2004). *Chapter 49 Evidence on the nature and sources of agglomeration economies* (pp. 2119–2171). https://doi.org/10.1016/S1574-0080(04)80006-3

Shaver, J. M., and Flyer, F. (2000). Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States. *Strategic Management Journal*, *21*(12), 1175–1193.

Shearmur, R. G., and Coffey, W. J. (2002). Urban Employment Subcenters and Sectoral Clustering in Montreal: Complementary Approaches to the Study of Urban Form. *Urban Geography*, *23*(2), 103–130. https://doi.org/10.2747/0272-3638.23.2.103

Shearmur, R., and Alvergne, C. (2003). Regional planning policy and the location of employment in the Ile-De-France: Does policy matter? *Urban Affairs Review*, 39(1), 3–31. https://doi.org/10.1177/1078087403253557

Shen, H., Namdarpour, F., and Lin, J. (2022). Investigation of online grocery shopping and delivery preference before, during, and after COVID-19. *Transportation Research Interdisciplinary Perspectives*, *14*, 100580. https://doi.org/10.1016/j.trip.2022.100580

Simmonds, D., and Feldman, O. (2011). Alternative approaches to spatial modelling. *Research in Transportation Economics*, *31*(1), 2–11. https://doi.org/10.1016/j.retrec.2010.11.002

Singh, S., and Santhakumar, S. M. (2022). Empirical analysis of impact of multi-class commercial vehicles on multi-lane highway traffic characteristics under mixed traffic conditions. *International Journal of Transportation Science and Technology*, *11*(3), 545–562. https://doi.org/10.1016/j.ijtst.2021.07.005

Sivakumar, A., and Bhat, C. R. (2007). Comprehensive, Unified Framework for Analyzing Spatial Location Choice. *Transportation Research Record: Journal of the Transportation Research Board*, *2003*(1), 103–111. https://doi.org/10.3141/2003-13

Smeed, R.J. (1964). The Traffic Problem in Towns. *Town Planning Review* 35 (2): 133. doi:10.3828/tpr.35.2.n482755235577655.

Statistics Canada. (2023). *Retail e-commerce and COVID-19: How online sales evolved as in-person shopping resumed. https://www150.statcan.gc.ca/n1/pub/11-621-m/11-621-m2023002-eng.htm*

Sun, B., and Wei, X. (2014). Spatial distribution and structure evolution of employment and population in Shanghai Metropolitan Area. *Acta Geographica Sinica*, 69(6), 747–758 (In Chinese)

Sun, T., Wang, L., and Li, G. (2012). Distributions of population and employment and evolution of spatial structures in the Beijing Metropolitan Area. *Acta Geographica Sinica*, 67(6), 829–840 (In Chinese).

Suel, E., and Polak, J. W. (2018). Incorporating online shopping into travel demand modelling: challenges, progress, and opportunities. *Transport Reviews*, *38*(5), 576–601. https://doi.org/10.1080/01441647.2017.1381864

Swait, J., and Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, *21*(2), 91–102. https://doi.org/10.1016/0191-2615(87)90009-9

Tan, R.H., Q.S. He, K.H. Zhou, and P. Xie. (2019). The Effect of New Metro Stations on Local Land Use and Housing Prices: The Case of Wuhan, China. *Journal of Transport Geography* 79: 102488. doi:10.1016/j. jtrangeo.2019.102488.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263

Tu, J.J., S.Q. Tang, Q. Zhang, Y. Wu, and Y.C. Luo. (2019). Spatial Heterogeneity of the Effects of Mountainous City Pattern on Catering Industry Location. *Acta Geographica Sinica* 74 (6): 1163–1177. (In Chinese). Doi10.11821/dlxb201906007.

Ülkü, M.A. (2012). Dare to care: shipment consolidation reduces not only costs, but also environmental damage. *International Journal of Production Economics* 139, 438–446. https://doi.org/10.1016/j.ijpe.2011.09.015.

Ülkü, M.A., and Engau, A. (2021). Sustainable supply chain analytics. In Industry, Innovation and Infrastructure; Leal Filho, W., Azul, A.M., Brandli, L., Lange Salvia, A., Wall, T., Eds.; *Encyclopedia of the UN Sustainable Development Goals*; Springer: Cham, Switzerland, pp. 1123–1134.

US Environmental Protection Agency. (2011). *Smart growth. https://www.epa.gov/smartgrowth*

US Environmental Protection Agency. (2015). *Emission Adjustments for Temperature, Humidity, Air Conditioning, and Inspection and Maintenance for On-road Vehicles in MOVES2014*. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100NOEM.txt van Wissen, L. (2000). A micro-simulation model of firms: Applications of concepts of the demography of the firm. *Papers in Regional Science*, *79*(2), 111–134. https://doi.org/10.1007/s101100050039

Venkadavarahan, M., and Marisamynathan, S. (2023). Development of freight trip generation model using observed and unobserved information of supply chain characteristics for a sustainable urban transformation. *Journal of Cleaner Production*, *421*, 138500. https://doi.org/10.1016/j.jclepro.2023.138500

Veneri, P. (2018). Urban spatial structure in OECD cities: Is urban population decentralising or clustering? *Papers in Regional Science*, 97(4), 1355–1374. https://doi.org/10.1111/pirs.12300

Waddell, P., and Shukld, V. (1993). Manufacturing location in a polycentric urban area: a study in the composition and attractiveness of employment subcenters. *Urban Geography*, *14*(3), 277–296. https://doi.org/10.2747/0272-3638.14.3.277

Wang, F., and Zhou, Y. (1999). Modelling urban population densities in Beijing 1982-90: Suburbanisation and its causes. *Urban Studies*, 36(2), 271–287.

Weterings, A., and Knoben, J. (2013). Footloose: An analysis of the drivers of firm relocations over different distances. *Papers in Regional Science*, *92*(4), 791–809. https://doi.org/10.1111/j.1435-5957.2012.00440.x

Wisetjindawat, W., Sano, K., Matsumoto, S., and Raothanachonkun, P. (2007). Microsimulation Model for Modeling Freight Agents Interactions in Urban Freight Movement. *Transportation Research Board 86th Annual Meeting*.

Yamamoto, T., Kitamura, R., and Kishizawa, K. (2001). Sampling Alternatives from Colossal Choice Set: Application of Markov Chain Monte Carlo Algorithm. *Transportation Research Record: Journal of the Transportation Research Board*, *1752*(1), 53–61. https://doi.org/10.3141/1752-08

Yang, Z.S., R. Sliuzas, J.M. Cai, and H.F.L. Ottens. (2012). Exploring Spatial Evolution of Economic Clusters: A Case Study of Beijing. *International Journal of Applied Earth Observations & Geoinformation* 19: 252–265. doi:10.1016/j.jag.2012.05.017.

Yao, L., and Y. Hu. (2020). The Impact of Urban Transit on Nearby Startup Firms: Evidence from Hangzhou, China. *Habitat International* 99: 102155. doi:10.1016/j.habitatint.2020.102155.

Zhang, T., Sun, B., Li, W., Dan, B., and Wang, C. (2019). Polycentricity or dispersal? The spatial transformation of metropolitan Shanghai. *Cities*, 95, Article 102352. https://doi.org/10.1016/j.cities.2019.05.021

Zhao, J.P., M.H. Lu, and H.C. Liu. (2020). Characteristics and Influencing Factors of Spatial Distribution of Headquarters of Listed Firms in Beijing. *Economic Geography* 40 (1): 12–20. (In Chinese). Doi: 10.15957/j.cnki.jjdl.2020.01.002.

Ziomas, I., Suppan, P., Papayannis, A., Melas, D., Fabian, P., Tzoumaka, P., RappenglÜCk, B., Balis, D., and Zerefos, C. (1995). A contribution to the study of photochemical smog in the Greater Athens area. *Contributions to Atmospheric Physics*.
https://doi.org/refwid:16736