AI-Based Mode of Transportation and Destination Classification and Prediction in Origin-Destination Surveys

Ali Ahmadi

A Thesis
in
The Concordia Institute
for
Information Systems Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science
(Quality Systems Engineering) at
Concordia University
Montréal, Québec, Canada

April 2022

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Ali Ahmadi**
Entitled: **AI-Based Mode of Transportation and Destination Classification and Prediction in Origin-Destination Surveys**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science
(Quality System Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Zachary Patterson _____ Chair

Dr. Walter Lucia _____ Examiner

Dr. Ursula Eicker _____ Thesis Supervisor

Approved _____

Dr. Mohammad Mannan, Graduate Program Director

April 2022 _____

Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science
(Gina Cody School of Engineering and Computer Science)

Abstract

On AI-based mode of transportation and destination classification and prediction in origin-destination surveys

Ali Ahmadi

Travel patterns and mode choice depend on individual socio-economic attributes that need better understanding. As a result, deciding which features to investigate is a challenge in data analysis.

This study investigates people's activities and trips to explore the correlation between individual and household socio-economic attributes, neighbourhood socio-economic level and land use, and the choice of mode of transportation to access destinations in the city of Montreal. The study found that the land-use characteristics of Montreal and the shapes of its residents' travel patterns impact the design and implementation of public transportation projects throughout the census agglomeration of Montreal. These transportation infrastructure influences people's commuting behaviour patterns. How to predict these patterns using historical data and existing master plans is a major goal of this work. Machine learning and deep learning algorithms were used to predict trip destination and mode of transportation.

Numerous factors influence a person's travel pattern, including their age, residence location, and purpose of the trip. The most critical attributes were detected based on feature extraction methods and correlations between features were analyzed using a correlation heat map. This allowed to determine the most significant features to predict the trip's destination and mode of transportation.

Three most recent versions (2008-2013-2018) [1–3] of the Montreal Origin-Destination (OD) data were used. Furthermore, a comparison between the accuracy of several well-known algorithms, such as decision trees, random forests, SVMs, and feed-forward neural networks, was conducted. Comparing different results yielded from different algorithms shows that neural networks outperform all the other algorithms in terms of accuracy in predicting both modes of transportation and destination (78 percent in mode choice and 68.7 percent in destination). Therefore, it was used to predict the future trip pattern of the year 2023.

Moreover, this study proposes a Bayesian network to forecast the entire trip patterns for Montreal in 2023. This network is used to create a scaled-down version

of OD2023. For this purpose, both OD and census data were used for the past 15 years. Different characteristics of trip patterns in each year were plotted. The Bayesian network captured and modelled how the trips changed over time.

This study provides a baseline for developing an application to extract critical statistical information about trip patterns on a neighbourhood scale in Montreal.

Finally, the foundations for an application to extract critical statistical information about various trip patterns in various Montreal neighbourhoods were created. This section combined various datasets from different years, including Census, land use [4–8], and OD survey data. This application displays the extracted data in various plots and tables.

This research is meant to serve as a summary of previous studies as well as a reference for future research.

Acknowledgements

Prof. Ursula Eicker, my supervisor, deserves nothing but the highest praise. In April 2019, I began my Master of Engineering studies in the Department of Information Systems Engineering at Concordia University, and in March 2020, Prof. Eicker welcomed me to join her research team as a machine learning researcher, which marked a pivotal point in my career. My supervisor believed in me, encouraged me never to lose hope in the future, and never let me give up in the face of adversity during the pandemic. Because of her unwavering support, patience, encouragement, and wise counsel, I can now pursue my ambitions and work hard to make this world a better place. I will be eternally grateful to her for allowing me to work under her supervision.

I want to express my gratitude to Mr. Guillermo Gutierrez Morote in particular. At Concordia University, he was like a brother to me. Guillermo assisted me in my research and professional work while working in two different areas. He had dedicated a significant amount of time to me to teach me how to be an impactful person in life and continue improving my skills. Guillermo is smart, kind, and trustworthy, and I hope to have the opportunity to work with him again.

It is essential to express my gratitude to Concordia University for allowing me to attend and support international students during the pandemic. Concordia University's response to the pandemic was incredible, and I will never forget their kindness and compassion.

Last but not least, I would like to acknowledge my family as the most precious gift I have ever received because of their unwavering love, respect, and confidence. I cannot express how grateful I am to have had wise advice and a sympathetic ear. You are always keen to support me.

My family and friends have always encouraged me to reach my goals and realize my visions, which has given me the courage to explore new opportunities. I am truly fortunate to have you in my life. Thank you for everything you have done for me over the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction and Related Work

Almost everyone in a community relies on transportation in some way. This industry can interact with other industries in a variety of ways. Transportation's social, cultural, political, and economic impact on various aspects of a country's life is a vital component of the country's progress and development [10]. Congestion on the roads, unsustainable land use, depletion of non-renewable energy resources, and crashes resulting in the loss of both life and property are all repercussions of the transportation system, which plays an integral part in societal and economic transition. Some of the essential services provided by transportation are the creation and changing of society. Since its discovery, humans have been aware of it and have been seeking to develop it to fit the needs of the modern world. Among these efforts and accomplishments, academic studies are the most significant and long-term. In academic institutions, a wide range of opinions and hypotheses are examined as part of original, leading-edge scientific endeavors carried out at the cutting edge of technology.

Advances in technology have an impact on all sorts of labor, including manual labor [11]. Technological progress has a profound effect on the work done in all sectors of academic research. People's interest in learning develops as technology advances, and they begin to consider a more varied range of opinions and ideas. Technological improvements have played a crucial part in making these feats feasible during many centuries (and failures). In retrospect, waves of transportation modernization pushed traditional modes of transportation out of the way, allowing for gradual social and economic shifts. A considerable number of these services are well-established in the contemporary era's more industrialized nations. It is proven that these services exist

and are firmly embedded in people's minds, as well as the institutions and activities that are associated with them.

Millions of Canadians utilize public transportation to commute around town and to work every day. The existing public transportation system, on the other hand, abandons many individuals because it fails to convey them to their destinations [12]. Lack of access to public transportation, when combined with other socioeconomic limitations such as a lack of a car, can contribute to a state of transport poverty, prohibiting people from getting to important locations such as places of employment on time [13,14]. This thesis aims to provide exact accessibility indicators to examine the degree to which these characteristics can have a tremendous impact on the mode of transportation Canadians choose to get to their destination. These measurements are used to estimate the extent to which Canadians' socioeconomic status and personal characteristics affect their capacity to travel, as well as how many of them are at risk of slipping into the category of "transport poverty." [15] This data can be used to impact legislation aimed at increasing transit use, decreasing accessibility inequities, and removing transportation-related barriers to actively participating. It also helps urban developers and planners observe how their decisions impact commuter behavior in society.

Many Canadians in cities rely on public transportation to get around and conduct other everyday duties. In Canada's rapidly expanding cities, low-income individuals frequently have no choice but to use public transportation [11]. A wide range of social and economic disadvantages can lead to a shortage of public transportation, exacerbating transit poverty (e.g., poor health, inability to purchase a car, etc. [16]). Unemployed individuals may be unable to find work or engage in labor. In addition to spending billions on public transit, the Canadian government emphasizes social justice, diversity, and inclusion. However, the total number of people who are unable to use public transit due to a variety of deficiencies remains unknown. Low-wage workers frequently have little choice but to use public transportation. Many sections of the country rely significantly on public transportation. However, many places remain underserved because public transportation does not provide adequate access to job possibilities [17].

In the past, researchers have conducted several data studies on many forms of datasets that contain the most relevant information about the social status of the citizens and their trip patterns. A survey of these studies has been gathered in [18] and [19]. Machine Learning algorithms are the most widely used algorithms in these studies. [20] and [21] have used neural networks to forecast a model for the Origin-Destination matrix. This matrix is a model to simulate the trip frequency of zones, with element od_{ij} meaning the average number of travels from zone i zone j in one

day. The Origin-Destination dataset used in this study is an extension of this matrix.

Other studies have applied neural networks to other forms of datasets. [22] and [23] have applied graph convolutional to model passenger demands in different zones. [24], [25] and [26] have used neural networks to estimate the trip duration of different modes of transportation at different times of the day. Of these, [25] has used a special kind of recurrent neural network model called Long Short-Term Memory (LSTM) to enrich its basic artificial neural network.

Due to their high levels of interpretability, decision trees have been two of the most attractive algorithms to use in these real-life scenarios. [27] have used the decision tree algorithm on GPS point data to estimate trip purposes. It also uses GIS land use and individual characteristics data to increase the robustness of the model. These datasets (relating to the city of Montreal) have also been used in this study.

In another study, [28], the authors have used the random forest algorithm, a random ensemble of decision trees, to model the mode of transportation used in a trip. The random forest algorithm in [28] has given a chance to the authors to investigate the effect of different characteristics in the mode choice of a trip.

Authors of [29] have extended the tools developed in [28] by generalizing it to estimate the purpose of the trip. The interpretability of random forest here has given the authors a chance to investigate the influence of data selection from different seasons for training and test sets. They have also employed Aslan & Zech's test (AZ-test) [30] to increase the model's efficiency.

Some other studies have applied another famous machine learning algorithm, called the Support Vector Machines (SVMs). [31] has used the algorithm alongside neural networks to predict the mode of transportation. [32] has used the same algorithm on historical traffic data to predict the travel time. [32] has also applied Weighted Moving Average on the SVM algorithm and demonstrated their proposed algorithm's superiority over vanilla neural networks.

However, neural networks are the dominant machine learning algorithms, especially in predicting and modeling complex trip traits, such as traffic congestion and trip destination. In this regard, [26] has applied an artificial neural network to limited data from freeways and intersections to model the traffic in rush hours. [33] has used massive data from intelligent transport systems (e.g., mobile devices) to predict a vehicle's next location on the fly.

1.2 Contribution

This study has combined some of the most enriched and newest versions of datasets gathered from the citizens of Montreal. The prominent data used in this study is

the OD dataset, details of which are explained in the next chapter. Alongside the Census and Land use dataset, this study aims to analyze some of the basic patterns in trip patterns of the citizens of Montreal.

More specifically, we used many state-of-the-art machine learning algorithms, such as decision trees, random forests, and neural networks, to predict the trip destination and mode of transportation of the trips. We observed the superiority of the neural network models to other algorithms.

Later in the study, we used a statistical tool, Bayesian belief nets, and the results from the trained neural network to generate a minimal version of OD2023. To increase the robustness of the generation, we used the census dataset to evaluate the shifts in trip traits.

1.3 Thesis Objective and Motivation

This work pursues two main objectives:

- It seeks to study the correlation between the individual and household socioeconomic attributes on the choice of mode of transportation and trip destination. In other words, the first part of the study aims to predict these two features using other attributes.

- It uses historic Origin-Destination survey data and also the result from the previous objective to develop an algorithm to predict the next version of the OD data.

  Realizing how the current situation affects the choice of mode of transportation and how we can shift more to public and active transportation due to climate change and urban congestion problems is essential. It is also vital to investigate a neighborhood's socioeconomic status to achieve more equitable and accessible transit systems. Since available transportation data is sparse, and methods are needed to interpolate and extrapolate data, this work and obtaining its objectives is necessary.

1.4 Thesis Overview

This thesis is organized as follows:

- Chapter 2 presents an extensive elaboration on the datasets used in the study. The OD, census and land-use datasets have been carefully analyzed in this chapter.

- In chapter 3, technical methodologies have been explained. In this chapter, we presented the machine algorithms used in the study.
- In chapter 4, We have presented the results of the methodologies used in the study. The results have been discussed in this chapter, and the efficiency of each of the models has been observed.
- Chapter 5 introduces some interpretations of the OD data combined with the land-use data. Using bar charts and graphs, we have presented some interpretations of the data and how they have changed recently.
- Finally, in chapter 6, we conclude our work, highlight some challenges and suggest future works.

Chapter 2

Data Explanation

2.1 Montreal Origin-Destination Survey

Origin-Destination (OD) surveys are used to ascertain traffic patterns in a particular area over a specified period. The data collected can be combined with other information that will assist transportation planners in determining an area's transportation needs and developing appropriate transportation solutions. This data enables stakeholders to comprehend travel patterns and characteristics; track trends; contribute to the development of travel demand models, forecasting and plan for areawide transportation infrastructure needs and resources; and track progress toward implementing transportation policies.

Numerous survey techniques can be used to collect data in origin-destination surveys. While older methods of collecting OD data are well-studied, they are constrained in determining the scope and description of data that can be gathered. The most critical aspect of a successful survey is the participation rate, determined using origin-destination survey methods. As a result, obtaining the reliable assistance of as many study participants as possible is critical to the success of a survey.

Newer methods generate more detailed, richer data that enables planners to simulate trip routes from start to finish. Specific data collection techniques require considerable labour, such as roadside interviews and phone surveys. While roadside interviews generate a high response rate, they frequently disrupt traffic. Telephone surveys do not cause traffic disruptions but may have low response rates. For origin-destination studies, mail and web surveys are also used.

Since 1970, the Origin-Destination Survey has been conducted every five years in Montreal. It is one of the most vital transportation studies in Quebec, covering

Table 2.1: Evolution of OD Survey Coverage

	1970	1974	1978	1982	1987	1993	1998	2003	2008	2013	2018
Territory (km^2)	1200	2300	2300	3300	4700	5400	5400	5500	8200	9840	9840
Population (in thousands)	2484	2835	2954	2896	2930	3278	3499	3613	3940	4288	4447
Sample (% of households)	3.79	4.78	5.31	6.98	4.68	4.65	4.64	4.70	4.10	4.35	3.89

an increasingly significant area that includes the entire metropolitan area, the major cities of Montreal, Laval, and Longueuil, and the north and south crowns. It is a trustworthy and detailed source of information about how people get around in the Greater Montréal area on foot, by bicycle, bus, metro, train, and car. Additionally, the survey improves public transportation and road system planning and enhances Greater Montréal's urban development plans.

The Autorité régionale de transport métropolitain conducted the Origin-Destination survey in collaboration with the Ministry of Transportation of Quebec, exo, the Réseau de transport de Longueuil, the Société de transport de Montréal, the Société de transport de Laval, and the Montréal Metropolitan Community.

The Montreal OD surveys are household surveys conducted in the greater Montreal area. They are directed at all residents of occupied private households. They hope to gain a better understanding of these individuals' travel habits. These are observational surveys that provide a statistical snapshot of the various characteristics of human movement. These surveys record the origin, destination, motivation for travel, departure time, and modes of transport used for each documented trip. Furthermore, additional sociodemographic variables are collected. This survey is primarily conducted via telephone interviews and aims to create a comprehensive picture of all trips made by region residents, regardless of the mode of transport. For the first time, the 2018 OD survey included a web-based questionnaire.

The OD survey targets the entire population residing in private dwellings in each municipality included in the survey area. No travel information is collected on children under the age of four. The survey results pertaining to trips taken on business days between September 5 and December 20.

The 2018 OD survey covers the territory of the ARTM and the CMM (Communauté Métropolitaine de Montréal) and sectors for which a significant proportion of citizens commute to this territory. This territory encompasses 9,840 square kilome-

tres and is home to 158 municipalities, the same number as in 2013.

The 2018 OD survey questionnaire includes three crucial subsets of data:

- *The household information* includes the household's location, the number of members, the number of available vehicles for household members, and the household's income category.
- *Individuals' information* includes their age, sex, possession of a driver's license, possession of a monthly transportation ticket, primary occupation, and usual place of work or study for workers and students.
- *The trip information* includes the reason for the trip, the origin, the destination, the time of departure, the modes of transport used and, where applicable, the public transportation lines, modal transfer points, type of parking, carpooling, and the use of highways and bridges for trips to or from the island of Montreal.

For the purposes of sampling in OD 2018, the territory is divided into 113 (113 in OD 2013, 108 in OD 2008, and 79 in OD 2003) geographic strata defined by groupings of census tracts (CTs) or census subdivisions (CSDs) from Statistics Canada's 2016 census (2011 census for OD 2013, 2006 census for OD 2008, and 2001 census for OD 2003). The average sampling rate is 3.89 percent (4.35 percent in 2013, 4.1 percent in 2008, and 4.7 percent in 2003), although it varies slightly by geographic strata.

2.1.1 Weighting and Imputation

Two weighting factors are calculated: those estimated using household data and those estimated using population data. The geographic class division used in the weighting process corresponds to the survey's municipal sample sectors.

- *Factors applying to households*: The sample is calibrated at the margins to allow reconstruction of the distribution of private households, with the distinction of household size modulated by the distribution of people by age group; each household is classified according to its size (1 person, two people, three people, and four people and over) and the age of its members segmented by age group without regard to gender (0–14 years, 15–24 years, 25–39 years, 40–64 years, and 65 years and over). Finally, households with the same category and weighting stratum inherit the same weighting factor.
- *Factors that apply to people*: The weighting factor for individuals is based on the household factor, which has been adjusted to reconstruct the distribution

of the total reference population into gender-diverse cohorts. This procedure utilizes a unique weighting factor for each household member. The trips are calculated based on the individual's weighting factor.

2.1.2 Imputation Procedures

To calculate the weighting factors, the variables sex, age, and groupage have their missing values imputed using probabilistic models based on all available information from similar households. Additionally, a probabilistic model was used to determine the destination of indefinite travel (municipal sector only) or uncodifiable destinations.

2.1.3 Statistical Precision

Precision measures the difference between an estimate obtained from the sample and the result obtained from a complete census. This discrepancy is attributable to two errors: error sampling and observation error. Sampling error generally decreases with increasing sample size. It was evaluated by using the notion of the confidence interval. In the context of an OD survey, the interval confidence level is very variable insofar as it depends on the nature of the estimates made. Based on the assumption of a normal distribution, it is generally recognized to be that significant; the results must come from a sample comprising a minimum of 30 observations. Observation errors occur during the execution of the survey. They depend on the quality of the sampling frame, procedures for collecting information, incorrect answers, refusals of response or data processing.

2.1.4 Interpretation of the Results Tables

The results of the OD survey are presented in the form of tables that describe various characteristics of households and people residing in each of the sectors, geographic areas of the metropolitan area, and the travel characteristics that come from - or are heading towards - these sectors. There are three distinct parts in these tables:

- Sociodemographic characteristics of households and people in a sector.
- Characteristics of the movements of residents of a sector.
- Characteristics of trips produced and attracted by a sector.

Table 2.2: Trip Purpose Distribution of Residents of Laval

Trip Purpose	Outgoing Trips	Incoming Trips	External Trips
Job	22.9%	8.4%	27.1%
Education	12.8%	9.0%	3.9%
Entertainment	8.0%	5.7%	21.3%
Shopping	9.2%	8.7%	12.7%
Returning Home	33.8%	56.3%	-%
Other	13.2%	11.9%	34.8%
TOTAL (Number)	**698200**	**696200**	**21400**

Table 2.2 to 2.6 present some of the information about the trips by residents of Laval and about the general trips produced and attracted by this sector. Figure 2.1 shows the approximate location of this sector inside the Montreal island.

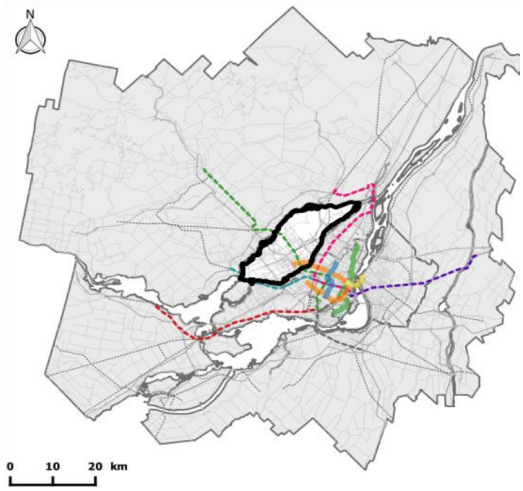

Figure 2.1: Laval's location inside Montreal [3]

The records of this data are available for every five years since 1993. We have used the recent three data records for this study, namely 2008, 2013, and 2018. Because of their irrelevance and inconsistencies with the newer data, records older than 2008 add noise and inaccuracies to the final results.

One sample of the data is depicted in figure 2.2. According to OD2013, this was trip taken by a 37-year-old woman at 4 P.M.

Table 2.3: Purpose Distribution of Trips To and From Laval

Trip Purpose	Outgoing Trips	Incoming Trips
Job	19.9%	15.3%
Education	10.9%	8.7%
Entertainment	7.5%	7.1%
Shopping	8.6%	9.2%
Returning Home	40.6%	47.5%
Other	12.5%	12.2%
TOTAL (Number)	**827100**	**827000**

Table 2.4: Mode of Transportation Distribution of Trips To and From Laval (Return Home Trips Excluded)

Mode of Transportation	Outgoing Trips	Incoming Trips
Motorized Vehicles	467200 (95.1%)	410900 (94.6%)
- Automobile	390300 (79.4%)	360100 (82.9%)
. Driver	*81.3%*	*81.7%*
. Passenger	*18.7%*	*18.3%*
- Public Transportation	66300 (13.5%)	30900 (7.1%)
. Subway	*56.4%*	*24.0%*
. STM (Bus)	*8.6%*	*6.2%*
. STL, RTL (Bus)	*66.2%*	*84.0%*
. Bimodal	*20.6%*	*10.1%*
- Other Motorized	24700 (5.0%)	23300 (5.4%)
Non-Motorized	23500 (4.8%)	22900 (5.3%)
Other	1100 (0.2%)	700 (0.2%)
TOTAL (Number)	**491500**	**434400**

Table 2.5: Mode of Transportation Distribution of Trips To and From Laval (A.M. Rush Hour Trips Only - Return Home Trips Excluded)

Mode of Transportation	Outgoing Trips	Incoming Trips
Motorized Vehicles	94.5%	93.3%
. Automobile	*71.7%*	*75.5%*
. Public Transportation	*19.4%*	*8.8%*
. Bimodal	*4.6%*	*0.5%*
. Other Motorized	*8.1%*	*9.6%*
Non-Motorized	5.3%	6.5%
Other	0.3%	0.3%
TOTAL (Number)	**235900**	**186200**

Table 2.6: Time Group Distribution of Trips To and From Laval (Trips with Motorized Vehicles Only)

Mode of Transportation	Outgoing Trips	Incoming Trips
A.M. Rush Hour	29.5%	23.2%
Day	25.7%	25.2%
P.M. Rush Hour	32.8%	38.1%
Evening	10.9%	12.2%
Night	1.1%	1.3%
TOTAL (Number)	**780000**	**780100**

Table 2.7: Trip purpose based on *motivation* key

1	Commute to Work	8	Health
2	Business Meeting	9	Drive Someone Back
3	Passing the Road	10	Looking for Someone
4	School	11	Returning Home
5	Shopping	12	Other
6	Entertainment	13	Refusal to Answer
7	Visiting a Friend		

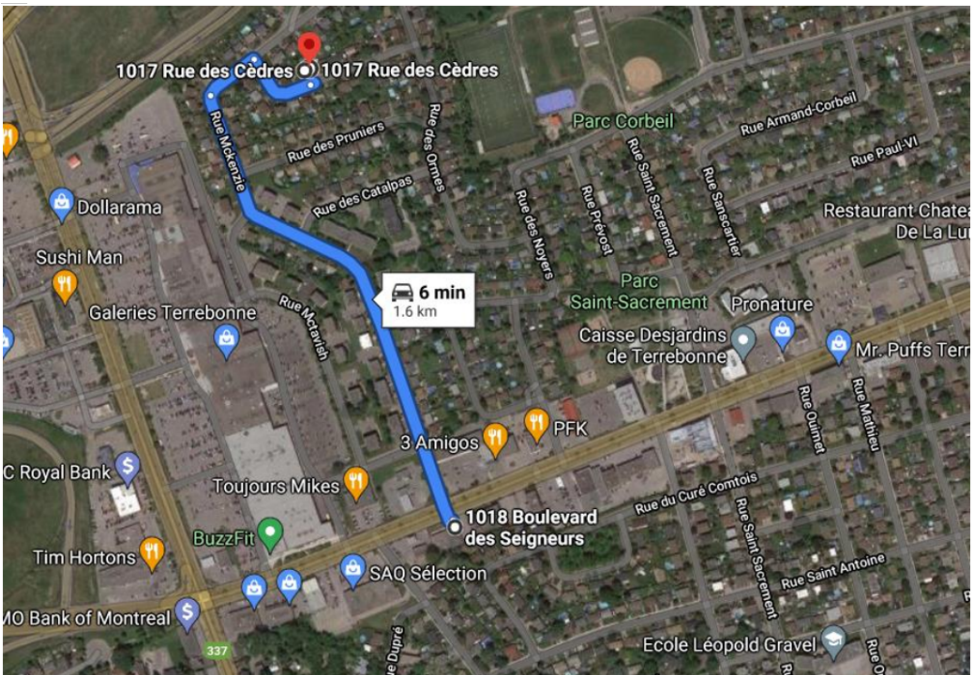

Figure 2.2: One random sample of OD2013

Table 2.8: Mode of transportation based on *Mode of transportation* key

1	Self-Drive
2	Taxi
3	Bus
4	Subway
5	Train
6	Motorcycle
7	Bike
8	Walk

2.2 Census Data

Every five years, the Census Program produces a statistical image of the country. It contains the population census as well as the agriculture census.

The population of Canada is counted in census returns, which are kept by the Government of Canada. An invaluable source of genealogy data can be found in these documents. As early as 1851, the names of all residents, their birthplaces and ages were recorded in most census records.

There is a hierarchical structure to the data collected by Statistics Canada, which breaks the country down into distinct sections.

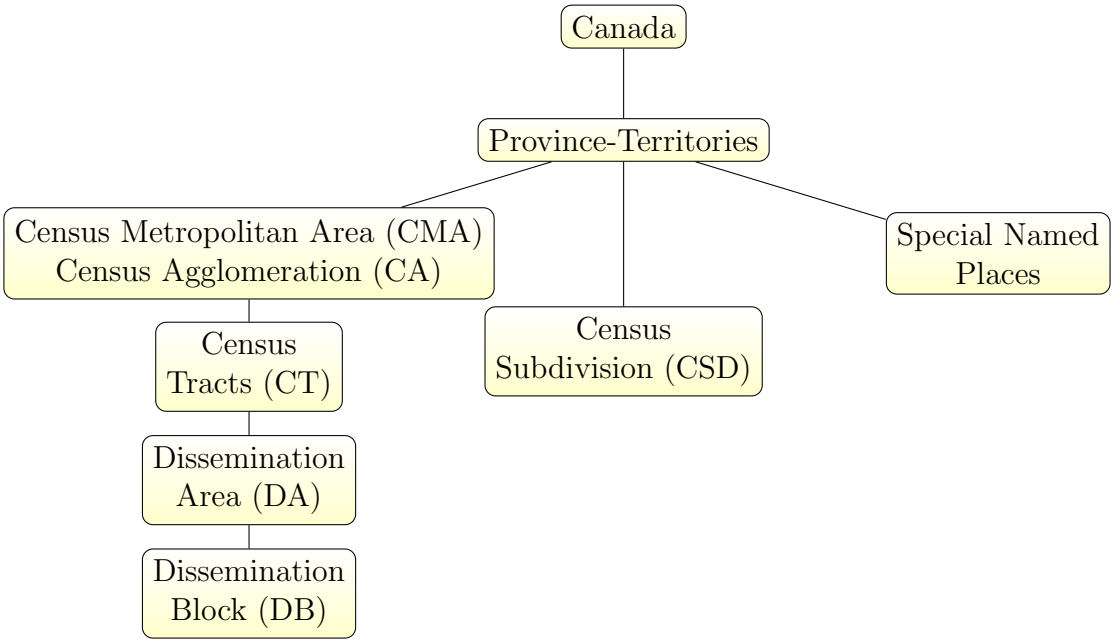

Figure 2.3: Hierarchy of census geographic entities

- Census Metropolitan Area (CMA) or Census Agglomeration (CA): A census metropolitan area must have a total population of at least 100,000 people, with at least 50,000 people residing in the core region. A census agglomeration must have a core population of at least 10,000 people to be considered.

  A census metropolitan area (CMA) or census agglomeration (CA) is made up of one or more adjacent municipalities that are centered on a population center (known as the core). A CMA must have a total population of at least 100,000

people, with at least 50,000 people living in the core. A CA must have a population of at least 10,000 people. Other nearby municipalities must have high commuting flows to be included in the CMA or CA.

CAs can be retired if the population of their core falls below 10,000. However, once a CMA is established, it remains so even if its total population falls below 100,000 or the inhabitants of its core falls below 50,000. The term "fringe" refers to small population centers with a population of less than 10,000 people. Rural areas include all areas within the CMA or CA that are not population centers.

Census tracts are created when the population of a CA reaches 50,000. Even if the population of the core drops below 50,000, census tracts in CA are preserved. Census tracts are used to divide all CMAs.

- Census SubDivision (CSD): CSD stands for "census subdivision," which refers to municipalities or areas that are considered municipal variants for statistical purposes. According to official designations introduced by federal or provincial authorities, there are 54 types of census subdivisions (CSDs).

- Census Tracts (CT): There are approximately 2,500 to 8,000 people in each census tract, which are generally stable geographic areas. Cities with a population of at least 50,000 in the previous census are home to these census metropolitan areas and census agglomerations Census tracts are first defined by a committee of local experts (such as planners, health and social workers, and educators) working with Statistics Canada. There are no census tracts that can be dissolved even if the population of an entire census metropolitan area or census agglomeration falls below 50,000. Using a set of rules, census tracts are defined. Priority is given to the following initial delineation rules.

    1. The boundaries of census tracts (CT) must be defined by physical features that can be easily recognized. Even if physical attributes are not nearby or do not exist, utility easements, boundary lines, and municipal limits can serve as CT boundaries.
    2. Ideally, the CT's population should be in the range of 2,500-8,000 people, with an average of 4,000. Central business districts, important commercial and industrial regions as well as peripheral areas may have populations that fall outside of this range.
    3. There should be as little variation in socioeconomic status and social work conditions as possible in the CT at its creation.

4. It's important to keep the CT's form as small as possible.
5. CT limits respect census metropolitan area, census agglomeration as well as province boundaries. However, CT boundaries may not usually follow census subdivision (municipality) boundaries.

- Dissemination Area (DA): A dissemination area (DA) is a small, fairly stable geographic unit formed of one or more neighboring dissemination blocks with an overall population of 400 to 700 inhabitants based on surveys from the last Census of Population Program. It is the smallest standard geographical region for which all census statistics are distributed. DAs cover all the area of Canada.

- Dissemination Block (DB): A dissemination block (DB) is a region surrounded on all sides by roadways and/or boundaries of conventional geographic areas. The dissemination block is the smallest geographical region for which inhabitants and housing counts are distributed. Dissemination blocks covering all the country of Canada.

2.3 Land Use

Land use refers to managing and transforming natural or wilderness areas into constructed environments like settlements and semi-natural habitats like arable fields, pastures, and managed woodlands. Land use by mankind has a lengthy history, dating back more than 10,000 years. It has been defined as "the total of arrangements, actions, and inputs that humans undertake in a particular land type," as well as "the intents and actions through which people engage with land and ecological systems." Amongst the most critical causes of ecological systems change is land usage.

As Albert Guttenberg (1959) wrote many years ago, "'Land use' is an important term in the language of city planning." Political jurisdictions commonly undertake Land-use planning and regulation to avoid land-use conflicts. Land division and use laws and regulations carry out land use plans, including zoning restrictions. Management consultancy businesses and non-governmental agencies will typically try to sway these policies before they become law.

2.3.1 Zoning

Zoning is an urban planning strategy whereby a municipality or other government body splits land into zones, each set of laws for new development. Zones can be defined for single-use (e.g., residential, industrial), or they can mix numerous suitable

activities by use, or, in the event of form-based zoning, the density, size, and shape of authorized structures can be governed by different restrictions regardless of their use. The development regulations for each zone govern whether a particular development can be approved. Zoning can prescribe a wide range of land uses, both unconditional and conditional. It could describe the size and dimensions of blocks that could be subdivided from a larger plot of land and the shape and scale of buildings. These rules are in place to help guide the growth and development of cities. In developed countries, zoning is the most frequent urban planning strategy utilized by municipal governments.

2.3.2 Scope

Zoning's primary objective is to separate uses that are deemed incompatible. Zoning is also used in practice to prevent new construction from interfering with preexisting uses and/or to preserve a community's "identity."

Zoning may include regulations governing the types of activities that are permitted on specific lots (such as open space, residential, agricultural, commercial, or industrial), the densities at which those activities are permitted, the height of buildings, the amount of space that buildings may occupy, the location of a building on the lot (setbacks), and the proportion of the lot devoted to those activities.

2.3.3 Origins and History of Zoning

Zoning districts date back to antiquity. The ancient walled city served as a model for characterizing and governing land according to its use. Outside the walls of the city were the unpleasant functions, which were typically associated with noise and stench; this was also the location of the poorest inhabitants. Unsanitary and dangerous activities, including butchering, sewage disposal, and brick firing, took place between the walls. Civil society, religious institutions, and residences were within the walls for most of the population.

2.3.4 Types

There are numerous zoning categories, some of which concentrate on regulating building types and relationship to the street of mixed uses, referred to as form-based zoning, while others focus on separating land uses, referred to as use-based zoning, or a combination of the two. Use-based zoning systems may include single-use zones,

mixed-use zones allowing for the coexistence of compatible uses, or a mixture of both single and mixed-use zones in a single system.

- Single-use zoning: Single-use zoning is a type of zoning in which only one type of use is permitted per zone. In North America, this is referred to as Euclidean zoning. Residential, mixed-use residential-commercial, commercial, industrial, and spatial zones are frequently defined as single-use zones (e.g. power plants, sports complexes, airports, etc.). Each category can be further subdivided; for example, within the commercial category, there may be distinct zones for small-retail, large retail, office use, and accommodation, while industrial may be further subdivided into heavy manufacturing, light assembly, and depot uses. In Germany, each category has a specific noise emission limit. Single-use zoning has contributed to the particular form of many cities across the United States, Canada, where a dense urban core, frequently containing highrises, is surrounded by low-density residential suburbs with extensive gardens and leafy roadways.

  Critics believe that separating daily uses from one another increases traffic because people need cars to live normally where their essential human needs are met, then get in their cars and drive throughout the day to meet those needs. Single-use zoning and urban growth have also been criticized for making it more challenging to achieve work-family balance, as greater distances must be covered to integrate the various life domains. These issues are particularly intense in the United States due to the country's high rate of automobile ownership and inadequate or inadequate urban rail and metro systems. Another source of criticism for zoning restrictions comes from those who believe the limitations violate people's property rights. A property owner may be unable to use his land for the purpose for which she purchased it because of zoning restrictions. According to some economists, single-use zoning regulations work against economic productivity and impede development in a free market economy, as inefficient zoning restrictions prevent a given area from being used more efficiently. Even in the absence of zoning regulations, a landfill, for instance, would likely gravitate toward less expensive land and away from residential areas. Single-use zoning regulations can halt innovative development such as mixed-use buildings and even prohibit otherwise permissible activities such as yard sales.

- Mixed-use zoning: Jane Jacobs, a planner and community activist, wrote extensively about the correlation between use separation and the weakness of

urban planning in New York City. She argued for dense mixed-use developments and pedestrian-friendly streets. Compared to villages and towns, where residents are generally acquainted, and low-density outer suburbs with few commuters, cities and inner-city areas face the challenge of maintaining order among strangers. This is possible in prosperous metropolitan areas with diverse uses, generating interest and attracting visitors. Some zoning systems have created mixed-use zones to facilitate the Modern Urbanist vision of walkable communities that integrate cafés, restaurants, offices, and residential development into a single zone. These continue to employ the fundamental control systems of zoning, excluding conflicting uses such as heavy industry or sewage farms while permitting compatible uses including residential, commercial, and retail activities, allowing people to live, work, and socialize within a compact geographic area.

- Form-based zoning: Not the type of land use; however, the form that land use may take is regulated by form-based zoning. For example, form-based zoning may require low setbacks, dense population, and walkability in a dense area. Form-based codes (FBCs) are intended to respond directly to a community's physical structure to promote more pedestrian-oriented and adaptable environments.

- Conditional zoning: Conditional zoning provides municipalities with additional agility and enables them to respond to the unique characteristics of a particular land use implementation. Conditional use zoning allows for permitting uses that would be prohibited under existing zoning, such as a school or a community center. Conditional utilization permits (also known as special use permits) authorize land uses that may be appropriate only in specific locations or configured or operated in a specific manner due to their unique characteristics.

- Pattern zoning: Pattern zoning is a type of zoning in which a municipality offers licensed pre-approved design ideas with an accelerated permitting process. Pattern zoning is used to lower development barriers, increase affordable housing, alleviate administrative burdens on permit review staff, and create high-quality residential designs within a suburb or jurisdiction. Additionally, pattern zoning can be used to promote specific types of buildings, such as lacking middle accommodation and affordable modest commercial properties. In some instances, municipalities acquire design patterns and build the properties themselves, while in others, municipalities sell the patterns to private

developers.

Land-use regulation is provincial jurisdiction in Canada, stemming from the constitutional mandate over property and civil rights. This authority was granted to provinces by the British North America Acts of 1867, and the 1982 Constitution Act continued it. The zoning authority is limited to real property, including land and improvements built on it becoming part of the land (in Quebec, Immeuble). The provinces delegated authority over land use within their boundaries to municipalities and regions, allowing municipalities to develop their zoning by-laws. There are regulations for land use control in provincially unorganized areas. Provincial courts are the final arbiters of appeals and revocations.

Figure 2.4: Land use characteristics of Montreal throughout recent years

Chapter 3

Methodologies

In this chapter, we elaborate on the methods used in this study to exploit the valuable information in the datasets introduced in the previous chapter. We used basic graphs and charts to create a comprehensive illustration throughout this chapter. The results of these methods on our datasets (mainly on the OD survey) are presented in more detail in the next chapter.

3.1 Feature Correlation

Before using a machine learning model to predict the main pattern of trips in the OD-data, we initialize our analytical study by investigating simple correlations between different characteristics of the trips. This helps us understand the nuances in the trip patterns better, and it also makes the result of machine learning models more interpretable.

The main challenge in this part of the work is considering different features, as there are both categorical characteristics (e.g., mode of transportation) and numerical characteristics (e.g., traveler's age). Fortunately, there are many methods for evaluating the correlation between categorical and continuous data. Based on the type of characteristics, there are three algorithms we need to consider in assessing feature correlation.

3.1.1 Categorical-Categorical Correlation

The algorithm in this type of correlation works on two categorical characteristics, such as the correlation between mode of transportation and purpose of the trip.

We use *Cramer's phi* [34] for this type of correlation, which gives a value between 0 and 1. A value of 0 indicates no associations between the features considered, and a value of 1 corresponds to a case where two variables can accurately be determined by one another. The simple task of this algorithm is to consider each possibility of values (x_i, y_j) created by two variables of x and y. The algorithm later decides if any value of x_i impacts the value of y.

Let n be the sample size of different (x_i, y_j) pairs for $1 \leq i \leq I$ and $1 \leq j \leq J$, and let n_{ij} be the number of times the value (x_i, y_j) has been observed. Also, let $|x_i|$ and $|y_j|$ represent the number of times we have observed $(x_i, .)$ and $(., y_j)$, respectively. More specifically, $|x_i| = \sum_j n_{ij}$ and $|y_j| = \sum_i n_{ij}$

Cramer's phi is calculated using the equation below. Higher values of Cramer's phi correspond to a higher association of two variables of x and y

$$V = \sqrt{\frac{\phi^2}{\min(I-1, J-1)}} \tag{3.1}$$

Where ϕ^2 has a direct relationship with chi-squared statistics and an indirect relationship with the sample size. Meaning

$$\phi^2 = \frac{\chi^2}{n} \tag{3.2}$$

and

$$\chi = \sqrt{\sum_{i,j} \frac{\left(n_{ij} - \frac{|x_i||y_j|}{n}\right)^2}{\frac{|x_i||y_j|}{n}}} \tag{3.3}$$

3.1.2 Categorical-Numerical Correlation

The algorithm used in this part evaluates the correlation between a numerical and a categorical characteristic, such as the correlation between the number of people in the household and the purpose of the trip.

We have used the correlation ratio algorithm [35] to evaluate this type of correlation. The way these algorithms work is quite similar to Cramer's phi algorithm. We first calculate the sum of deviation within different categories inside a categorical value using this algorithm. Then, we calculate the deviation of the entire sample. The ratio of the two deviations represents our objective. The ratio can take any value between 0 and 1, from the lowest degree of correlation to the highest.

Let n be the sample size of different (x_i, y_j) where x is a continuous variable and y is a categorical variable and let n_{y_j} be the number of times that value of y has been observed in category j. Also, let x_{i_j} be a value of x that has been observed alongside the value of y_j. Two types of average can be defined here as

$$\overline{x_{y_j}} = \frac{\sum_i x_{i_j}}{n_{y_j}} \tag{3.4}$$

and

$$\overline{x} = \frac{\sum_i x_i}{n} \tag{3.5}$$

Here, the correlation ratio η can be defined as

$$\eta^2 = \frac{\sum_j n_{y_j} \left(\overline{x_{y_j}} - \overline{x}\right)^2}{\sum_{i,j} x_{i_j} - \overline{x}} \tag{3.6}$$

3.1.3 Numerical-Numerical Correlation

The correlation evaluation algorithm used here works on two numerical characteristics. For instance, the correlation between the number of people in the household and the number of automobiles in the household.

In this part, we have simply applied the Pearson's r [36], which for a sample of size n consisting of tuples (x_i, y_i) can be calculated as

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}} \tag{3.7}$$

3.2 Feature Selection and Data Preprocessing

In order to work with the OD survey data, we needed to pick the most appropriate and most representative features in the dataset. From the entire set of the attributes in this dataset, we picked the features that:

- Were accurately reported. For instance, some features such as the exact coordinate of the origin and destination had been masked previously to avoid privacy invasion.
- Had few unknown values. For instance, features such as the second mode of transportation had null values for more than 80 percent of the trips since most trips were made using only one mode.

- Had the most negligible value of correlation with other attributes, hence yielded more information. For instance, two features of the exact age and age group of the traveler ultimately represented the same information. Picking them both would only add redundancy to the result of this study.

After feature selection and before being fed to any model, preprocessing on data was necessary. Since the OD data has been revised many times, the records containing nullities were of no concern to our work. However, we ended up throwing away the records that were not in accordance with the data dictionary. For instance, some rows had a value for the traveler's age that did not fit their age group. Since they were not many records with such problems, we did not let them be a part of the primary data. The OD data contains values are not in the same range. For instance, the traveler's age can be an integer between 10 and 100, whereas gender is a binary value. As a result, we needed to normalize the data to adjust its different scale values to a standard scale. In this study, Min-Max Normalization is used. The formula for min-max scaling is given as:

$$X_{new} = \frac{X_{old} - \min(X_{old})}{\max(X_{old}) - \min(X_{old})} \tag{3.8}$$

Where X_{old} is the original set, X_{new} is the scaled set, and $\min(X_{old})$ and $\max(X_{old})$ are the minimum and maximum values in the original set, respectively.

The effect of Min-Max normalization can be analyzed in 3.1. We used a random dataset with two features f_1 and f_2, with different scales for this plot. As you can see, after using normalization, both features are drawn to the same scale, and data points are much easier to distinguish [37].

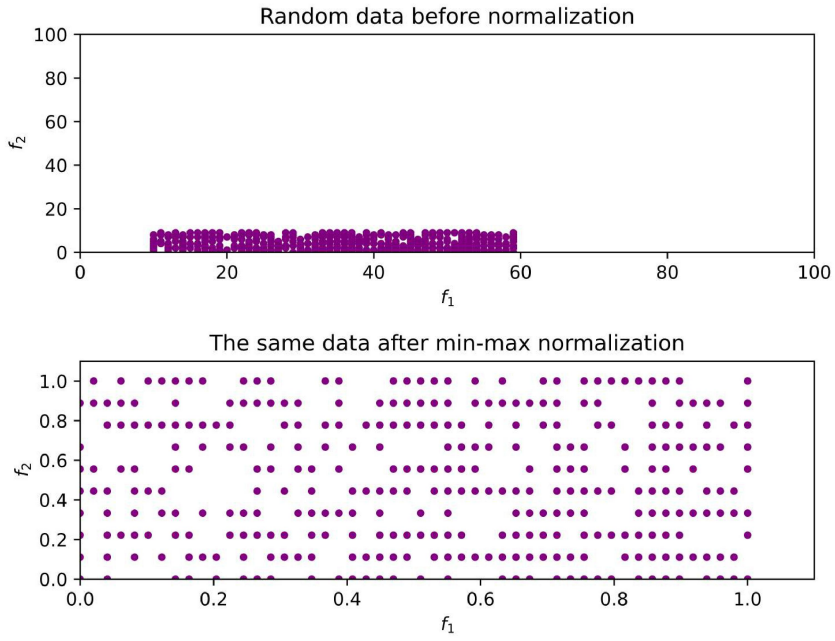

Figure 3.1: The effect of normalization on random data

3.3 Machine Learning Models

One of the main contributions of this study is to predict two of the most critical features in every trip, the mode of transportation and trip destination. Since there are a finite number of modes of transportation and places to go on a trip, predicting each of them is a classification task. In other words, the models we have used classify each trip into their proper classes, based on the modes of transportation and the destination. We have chosen some of the most famous and efficient machine learning classifiers, such as KNN, decision tree, and MLP. The OD data can be used in each model with the two main preprocessing mentioned in 3.2. In the rest of this section, we discuss each of the models we used on the data.

3.3.1 Nearest Centroid Classifier

This algorithm assigns to observation the label of the class with the nearest mean to the observation [38]. In this classifier, for an observation $\overline{x_o}$, the assigned class y_o is

$$y_o = \arg\min_{c \in C} ||\overline{\mu_c} - \overline{x_o}|| \tag{3.9}$$

where C is is the set of all classes and $\overline{\mu_c}$ is the center (mean or prototype) of class c. For all train data in class c (Q_c), the mean is calculated as

$$\overline{\mu_c} = \frac{1}{|Q_c|} \sum_{\overline{x_j} \in Q_c} \overline{x_j} \tag{3.10}$$

Since data points in OD data are compact in their feature space, assigning each class with only one prototype could have drawbacks. However, this algorithm is the simplest of classification algorithms and can yield traceable results.

Figure 3.2 illustrates the method of nearest centroid classifier for a random dataset with two features and a random test point of $[6.1, 2.7]^T$. Since this test point is closer to the mean of group one than to the mean of group two, then it is assigned to group one.

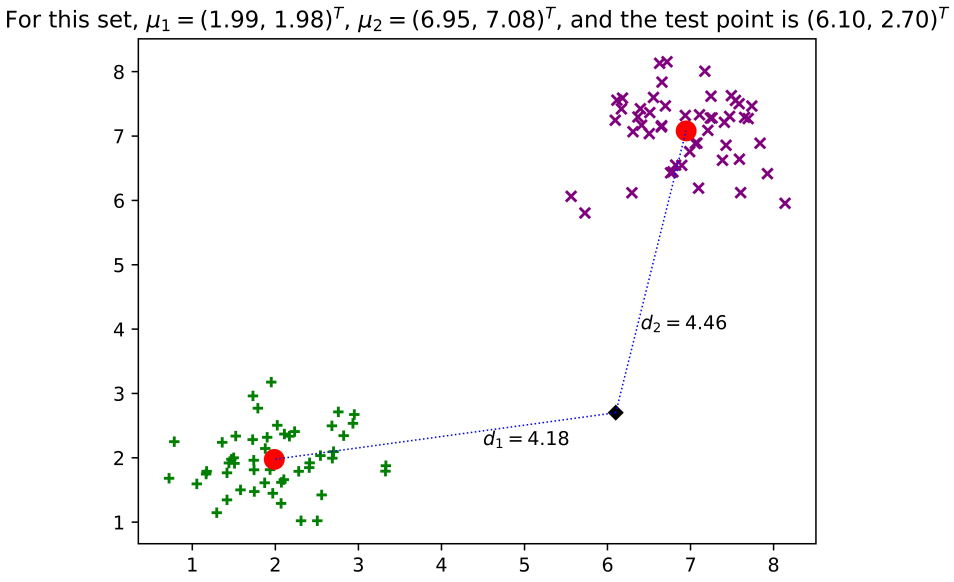

Figure 3.2: Nearest centroid algorithm on random data

3.3.2 K-Nearest Neighbor

To solve the problem of representing each class with only one prototype in the nearest centroid algorithm, we used the KNN classifier. This algorithm is a non-parametric algorithm, in which the assigned class of every test point is the class that is most prevalent among its k nearest training point neighbors [39].

In this algorithm, the value of k is chosen based on the data. Although typically a small integer, more complex datasets require higher values of k to reduce noise effects on the result. For this study, the results of choosing k from 1 to 90 have been observed separately. Figure 3.3 illustrates the method of K nearest neighbor with $k = 1$ for a random dataset with two features and a random test point of $[3.5, 8.7]^T$. Since the closest point among the nearest neighbors of different classes belongs to class 1, the test point is classified as a member of class 1.

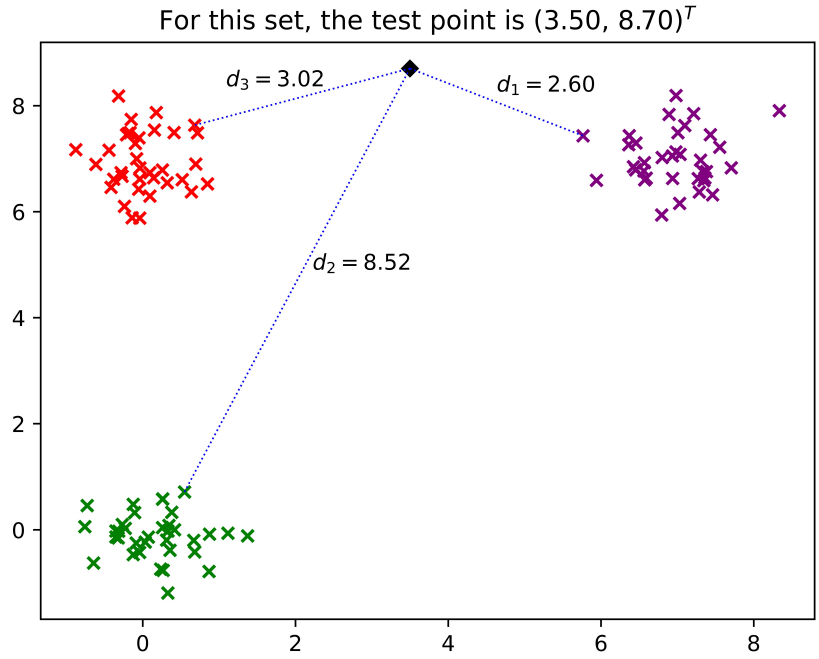

Figure 3.3: K-Nearest algorithm with $k = 1$ on random data

3.3.3 Gaussian Naive Bayes

This algorithm is a probabilistic method that assumes data features are conditionally independent of one another (hence the name, Naive) and associates a Gaussian distribution with each data feature [40].

Recall that our target was to calculate the probability of a destination, or a mode of transportation (c) given a trip record ($\overline{x}$), and $\overline{x} = [x_1, x_2, ..., x_n]^T$, where x_1 to x_n are the n separate features. In other words, we aim at calculating

$$P(c; x_1, x_2, ..., x_n) \propto P(c)p(x_1, x_2, ..., x_n; c) \tag{3.11}$$

Because of the assumption of feature independence, the probability of a set of features given the class label is equal to the product of the probabilities of each of the features given the same class label, meaning

$$p(x_1, x_2, ..., x_n; c) = \prod_{j=1}^{n} p(x_j; c) \tag{3.12}$$

and because of the Gaussian distribution assumption, each feature in a given label class has a normal distribution, meaning

$$p(x_j; c) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(x_j - \mu_c)}{2\sigma_c}\right) \tag{3.13}$$

As a result, we can write

$$P(c; x_1, x_2, ..., x_n) \propto P(c) \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(x_j - \mu_c)}{2\sigma_c}\right) \tag{3.14}$$

The objective of the algorithm is to find the best σ_c and μ_c that maximize $P(c; \overline{x})$, for any given class c and all train points belonging to that class. After learning the parameters, the selected class for test point $\overline{x_o}$ is

$$y_o = \arg\max_{c \in C} P(c; \overline{x_o}) \tag{3.15}$$

For instance, in figure 4.7, they are two classes that represent the entire 2D plane. With the assumption of Gaussian distribution, the algorithm has assigned every point in the entire 2D plane to either of the classes.

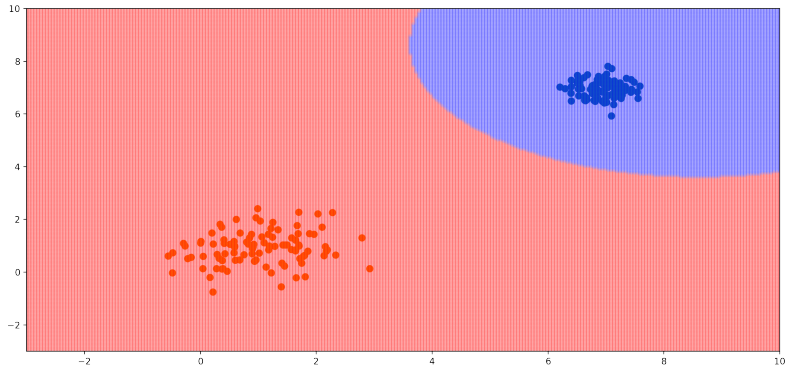

Figure 3.4: Gaussian naive bayes algorithm to classify all the points of the plane in two classes, red and blue

3.3.4 Decision Tree

The decision tree algorithm is one of the most intuitive yet effective algorithms in machine learning. By representing the entire data based on the features, this algorithm tries to find the most "useful" features and divide train data into target classes based on the values of that feature. The algorithm finds the best feature by deciding which one contributes more information to the classification task [41].

A valuable application of the decision tree algorithms is in cancer diagnosis. Imagine, based on three features of the patient's age, their weight, and the size of their tumors, a decision tree algorithm is trying to separate benign tumors from malignant ones. Working on the table 3.1, the algorithm will first decide to divide patients based on the size of their tumors because this feature seems to classify classes more coherently. More specifically, all patients with tumor size greater than four seem to have malignant tumors. Therefore, the feature of "tumor size" along with the threshold value of 4 will be the root node. For patients with a tumor size of less than 4, the algorithm will seek to find the feature and the threshold value that would yield more information. The final decision tree would be like 3.5.

Table 3.1: Information records of 19 patients

	Age	*Weight*	*Tumor Size*	*Tumor Status*
A	55	82	6.1	**Malignant**
B	52	78	4.5	**Malignant**
C	51	76	3.2	**Benign**
D	65	97	2.7	**Malignant**
E	43	71	4.7	**Malignant**
F	53	77	3.1	**Benign**
G	38	81	2.1	**Benign**
H	67	82	3.3	**Malignant**
I	29	83	3.7	**Benign**
J	53	78	5.2	**Malignant**
K	58	82	2.9	**Benign**
L	48	72	3.1	**Benign**
M	27	75	5.5	**Malignant**
N	61	95	3.6	**Malignant**
O	63	92	1.8	**Malignant**
P	53	88	2.9	**Benign**
Q	67	76	2.5	**Benign**
R	33	79	2.9	**Benign**
S	26	94	3.7	**Benign**

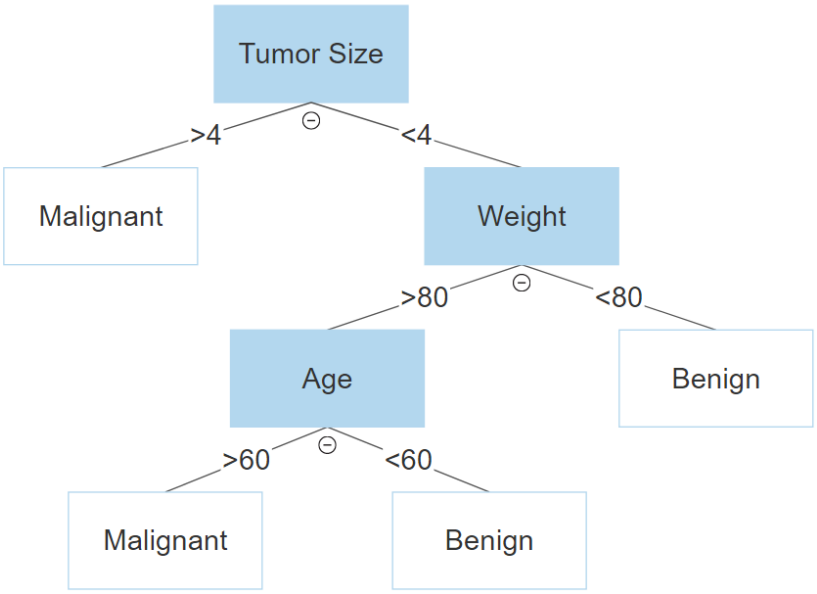

Figure 3.5: decision tree for table 3.1

3.3.5 Random Forest

As the name suggests, the random forest algorithm is an ensemble of decision trees with random configurations. In other words, the algorithm is not concerned about deciding on the best feature and the best threshold in every decision tree. Instead, it uses a random subset of features in every decision tree, and the final result is derived using a majority voting [42].

The intuition behind this algorithm is that the low correlation between decision trees yields better results than if they were to be used alone. To ensure this low correlation, decision trees have to be configured randomly, and each of them should be trained on a different subset of the main data. Although some of the trees might yield wrong results in the test phase, most of the trees would be accurate. As a result, majority voting would favor those with more accurate results.

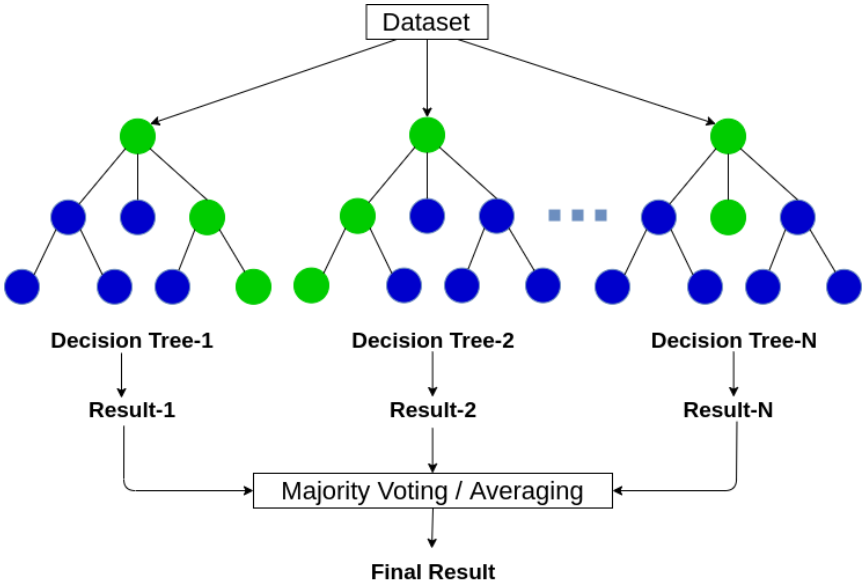

Figure 3.6: Random forest in a nutshell [9]

3.3.6 Support Vector Machine

The support vector machine (SVM) algorithm is one of the most robust algorithms in machine learning. This algorithm tries to separate data points of different classes using the "the best" boundary line. This boundary line seeks to separate different classes using the largest possible margin. The data points in all classes that push the margin towards the opposite sides are called Support Vectors. In short, the objective of SVM is to choose a hyperplane as a separator that ensure the maximum size of margin between points of each class [43].

If this hyperplane is denoted by the equation

$$\overline{w}^T\overline{x} + b = 0 \tag{3.16}$$

And if the algorithm is trying to separate two classes in figure 4.8, since data points belonging to each class falls on either side of the hyperplane, we can write

$$\begin{cases} \overline{w}^T\overline{x_i} + b > 0 & y_i = 1 \\ \overline{w}^T\overline{x_i} + b < 0 & y_i = -1 \end{cases} \tag{3.17}$$

Although not always feasible, the algorithm is trying to leave no points in the margin. As a result, equation 3.17 can be modified to form the equation

$$\begin{cases} \overline{w}^T\overline{x_i} + b \geq 1 & y_i = 1 \\ \overline{w}^T\overline{x_i} + b \leq -1 & y_i = -1 \end{cases} \tag{3.18}$$

Or equivalently

$$y_i \left(\overline{w}^T \overline{x_i} + b \right) \geq 1 \tag{3.19}$$

The main objective of the algorithm is to maximize the distance between two hyperplanes $\overline{w}^T \overline{x} + b = 1$ and $\overline{w}^T \overline{x} + b = -1$. The distance between these two hyperplanes is $\frac{2}{||\overline{w}||}$, where $||\overline{w}||$ is the L2 norm of the vector $\overline{w}$. As a result, the objective of the support vector machine algorithm is to solve the constrained optimization problem of

$$\begin{aligned} \max_{\overline{w}} \quad & \frac{2}{||\overline{w}||} \\ \text{s.t.} \quad & y_i \left(\overline{w}^T \overline{x_i} + b \right) \geq 1 \end{aligned} \tag{3.20}$$

Or equivalently

$$\begin{aligned} \min_{\overline{w}} \quad & \frac{1}{2} ||\overline{w}||^2 \\ \text{s.t.} \quad & y_i \left(\overline{w}^T \overline{x_i} + b \right) \geq 1 \end{aligned} \tag{3.21}$$

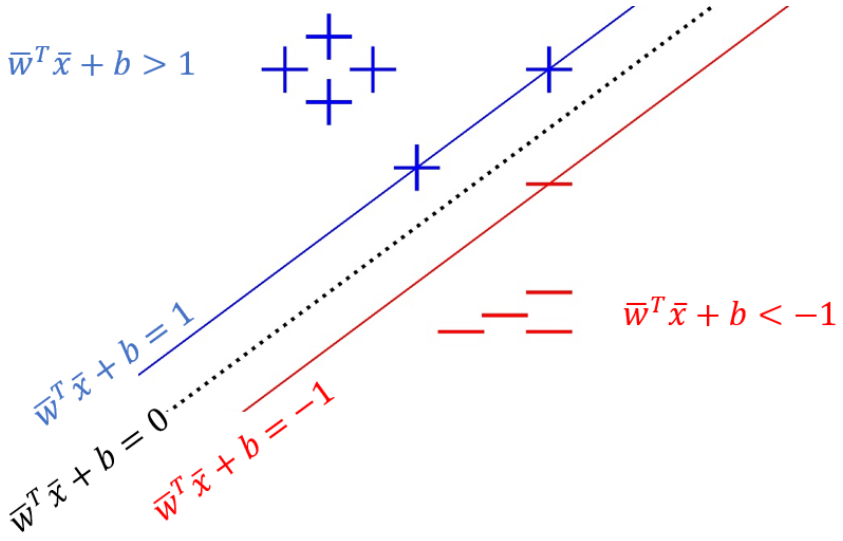

Figure 3.7: SVM algorithm in a nutshell

One of the main innovations used in SVM is the idea of the kernel trick. The intuition behind this idea is the following two points [44].

- Many machine learning problems (especially classifications) are easier to solve if the feature space has many independent dimensions.
- Many machine learning solutions can be defined as a function of dot products between data points.

$$\left(\sum_{i=1}^{m}\alpha_i\overline{x_i}\right)^T\overline{x}+b=0\rightarrow\sum_{i=1}^{m}\alpha_i\overline{x_i}^T\overline{x}+b=0\rightarrow\sum_{i=1}^{m}\alpha_i\langle\overline{x_i},\overline{x}\rangle=0 \tag{3.22}$$

Where α_is are the support vector coefficients, x_is are the training points, m is the number of training points, and $\langle .,. \rangle$ denotes the dot product of two vectors.

In order to make data points more linearly separable, sometimes a transformation function ϕ is used to map the data to higher dimension. As mentioned earlier, high-dimensional feature space could sometimes add relevant information that helps to solve the classification problem easier.

$$\begin{aligned} x_{\text{new}} &= \phi(x_{\text{old}}) \\ \phi &: \mathbb{R}^n \rightarrow \mathbb{R}^{n'} \\ \text{where} \quad & n' > n \end{aligned} \tag{3.23}$$

Using the transformation function, the new hyperplane equation in the SVM algorithm becomes

$$\sum_{i=1}^{m} \alpha_i \langle \phi\left(\overline{x_i}\right), \phi\left(\overline{x}\right) \rangle = 0 \tag{3.24}$$

Calculating the transformation function and the inner product of the results can be computationally expensive. Instead, as long as we can derive the inner products of the two functions, we do not need the mapping explicitly. In other words, if we define the kernel function $K(\overline{a}, \overline{b})$ as

$$K(\overline{a}, \overline{b}) = \langle \phi\left(\overline{a}\right), \phi\left(\overline{b}\right) \rangle \tag{3.25}$$

Then we can rewrite the equation 3.24 as

$$\sum_{i=1}^{m} \alpha_i K(\overline{x_i}, \overline{x}) + b = 0 \tag{3.26}$$

There are a few famous kernel functions that represent a non-linear function of the input. The most commonly used kernel function is the Gaussian kernel function, also known as the radial basis function (RBF), mostly represented as

$$K(\overline{a}, \overline{b}) = \mathcal{N}(\overline{a} - \overline{b}; 0, \sigma^2 \mathbf{I}) \tag{3.27}$$

Where $\mathcal{N}(\overline{x}; \overline{\mu}, \mathbf{\Sigma})$ is a Gaussian distribution with mean $\overline{\mu}$ and covariance matrix $\mathbf{\Sigma}$.

Other famous kernels exist such as polynomial kernel with degree d, represented as

$$K(\overline{a}, \overline{b}) = (\langle \overline{a}, \overline{b} \rangle + 1)^d \tag{3.28}$$

Figure 3.8 compares the effect of three types of kernels on the same dataset. Note how the RBF kernel is almost identical to naive Gaussian Bayes in this particular example.

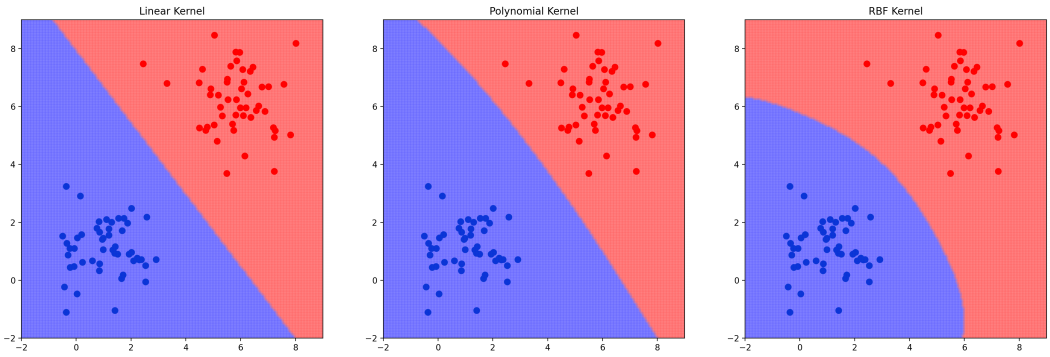

Figure 3.8: The effect of three kernels on classifying all the points of the plane in two classes, red and blue

3.3.7 Feedforward Neural Network

Feedforward neural network, also known as the multilayer perceptron (MLP), is the most widely used type of neural network for processing independent tabular data records. The purpose of a feedforward network is to develop a hypothesis function h that maps the input to the output with the least possible error rate [45]. For the process of classification (where the output can have discrete and finite values), feedforward networks can be simply modeled as

$$y_i = h(\overline{x_i}; \theta) \tag{3.29}$$

Where θ is the parameters of the neural network and $\overline{x_i}$ and y_i are its input and output, respectively. Based on an objective, by learning the values of θ, the model can find the approximate mapping between the input and the correct label associated with that output.

These networks contain no cycles or loops; Hence the information flows in only one direction, and the previous steps of the model receive no feedback from the future steps. The steps in these models are called layers. The input and output layers (that represent the input value and the output value) are mandatory in any feedforward neural network. More sophisticated models can have hidden layers as well. Each layer has an input and an output vector and represents a mapping between the input and the output. The accumulation of all the layers makes our primary hypothesis function. In general, an n-layer feedforward network can be modeled as

$$y_i = h_n\left(h_{n-1}\left(\ldots\left(h_1(x_i; \theta_1)\right)\right); \theta_{n-1}\right) \tag{3.30}$$

Where $h_j(.;.)$ and θ_j are the hypothesis function associated with the jth layer and the parameters of that layer, respectively. In simple terms, each layer maps the input vector to a higher (and sometimes lower) dimension feature space, in which some of the interesting patterns of the data can be observed and exploited. This is similar to mapping functions in SVM algorithms. However, in neural networks, the mapping function for each layer is not pre-determined and is learned throughout consecutive iterations.

The more layers a neural network has, the more complex its hypothesis function would be. This can be both a good and a bad feature of neural networks. Sometimes the pattern inside the data needs more sophisticated hypothesis functions to be exploited. On the other hand, more complex hypothesis functions lack flexibility and might not generalize. This issue will be discussed in more details further in this document.

The building blocks of each layer of the neural network are neurons. Each neuron can be thought of as a dimension in the feature space of the layer [46]. Consider figure 3.9 that represents a neural network with four layers. The network's input layer has two neurons; hence, the input layer represents a 2-dimensional feature space. ($\overline{x_i} \in \mathbb{R}^2$) Consecutive layers after the input layer represent a 7, 5, and 1-dimensional feature space.

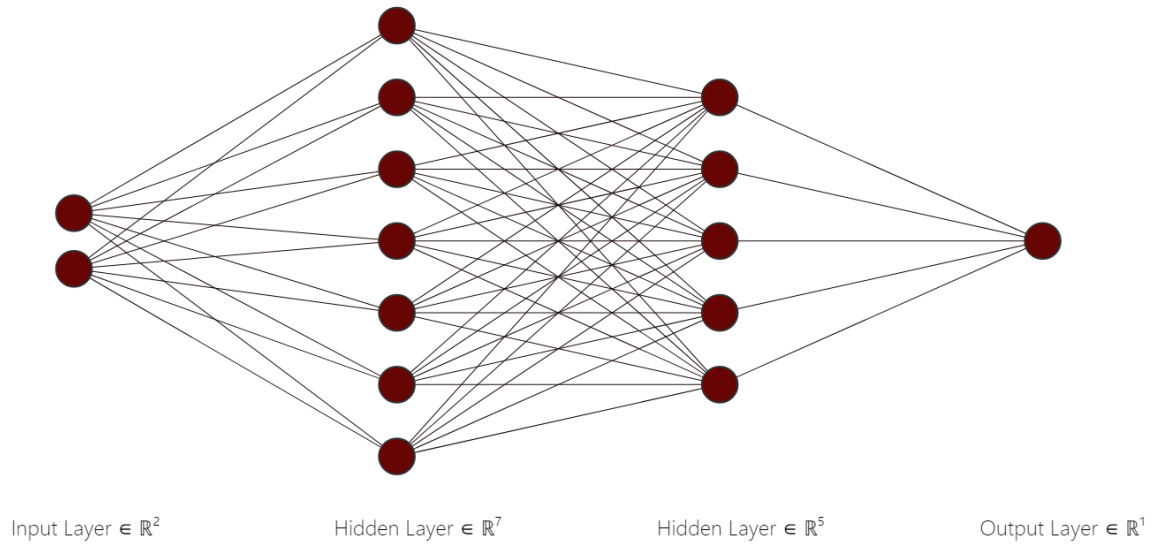

Figure 3.9: Example of an MLP with three layers

Another way of thinking of neurons is by their contribution to the hypothesis

function of the layers. The hypothesis function of layers is determined solely by their neurons. Every neuron is a functional unit that applies its parameters to the input signal and produces the input of the subsequent layers using an activation function. Below, we touch on the concepts of both the parameters of the neurons and their activation functions.

Neuron Parameters

The parameters of each neuron include a weight vector and a bias value. One responsibility of the neuron is to calculate the inner product of the weight vector and the input vector, followed by a summation with the bias value. For instance, the neuron in figure 3.10 has three inputs; therefore, it has a weight vector, $\overline{w}$ of size 3, and a bias value b. The first part of this neuron uses the following linear equation to simply calculate the mid-output of the neuron.

$$z = x_1 w_1 + x_2 w_2 + x_3 w_3 = \overline{w}^T \overline{x} + b = \langle \overline{x}, \overline{w} \rangle + b \tag{3.31}$$

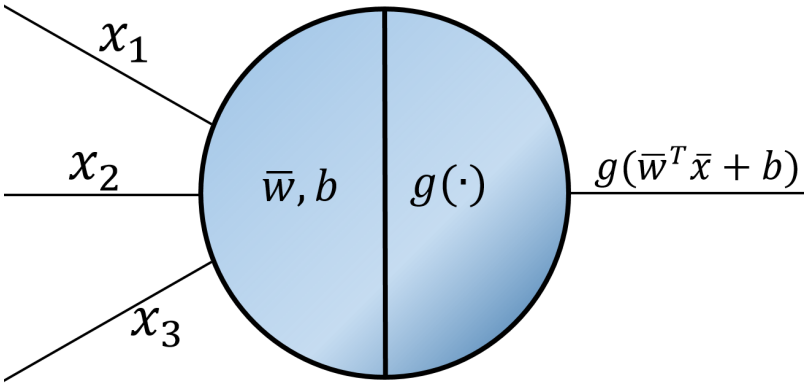

Figure 3.10: Example of a single neuron with three inputs

The final output of this neuron is calculated by applying an activation function $g(.)$ to the mid-output z.

Activation Functions

The mid-output of the neuron is too simple to have any helpful information about the input pattern. Without applying any further functions, the entire neural network model would be a linear mapping between the network's input and its output [47].

Therefore, right after the linear function of a neuron, a non-linear function is applied to the mid-output. This function decides whether the output of a neuron is strong enough to have an effect on future steps. Therefore, it is called an activation function.

There are plenty of possible activation functions that could be applied to hidden layers. Some of them are introduced hereunder.

- Rectified Linear Unit (ReLU):

  This function is the most widely used type of activation function, especially in deeper networks. It is simply defined as the positive part of its input, namely

$$g(z) = \max(0, z) \tag{3.32}$$

  The plot of this function is depicted in figure 3.11. Based on its ramp-shaped plot, it is sometimes called the ramp function as well.

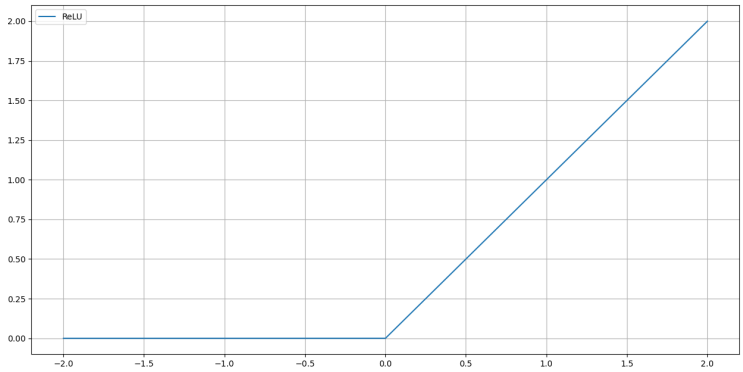

Figure 3.11: Plot of the ReLU function

- Sigmoid (Logistic):

  This function is also one of the most popular forms of activation functions. It is mainly used for binary classification. Since the output of the function ranges between 0 and 1, in binary classification, the function's output can be associated with the probability of the main output having the value 1. The sigmoid function has the equation of

$$g(z) = \frac{1}{1 + e^{-z}} \tag{3.33}$$

And its plot is depicted in figure 3.12

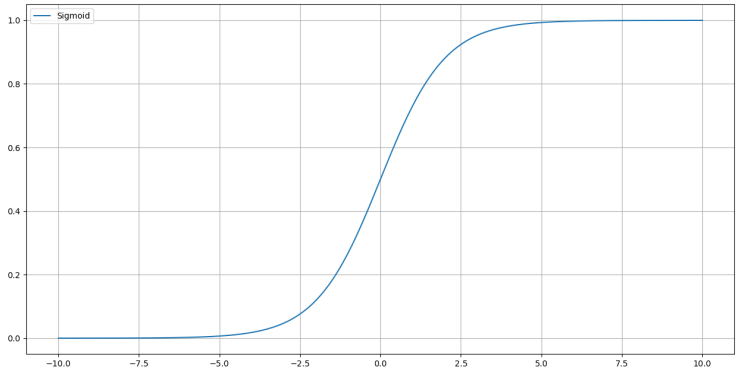

Figure 3.12: Plot of the Sigmoid function

- Tanh (Hyperbolic Tangent):

  This function is quite similar to the sigmoid activation function. Its output is in the range of -1 to 1. Hence, with proper normalization, it can be used for binary classification. It is represented with equation 3.34 and its plot is depicted in figure 3.13.

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{3.34}$$

- Leaky ReLU:

  This activation function is a variant of ReLU function. Sometimes when the learning rate is too high, ReLU function always outputs 0. (Because with higher learning rates, weight parameters become smaller and neuron mid-outputs are more likely to be negative) To avoid this problem, leaky ReLU has been proposed. It is a function that allows a positive slope when the neuron is seemingly weak. Both the equation of this function and its plot (represented in equation 3.35 and figure 3.14) are nearly identical to ReLU function.

$$g(z) = \max(z, 0.1z) \tag{3.35}$$

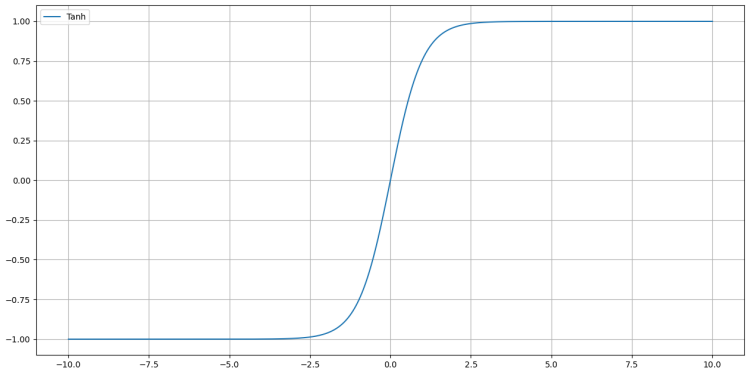

Figure 3.13: Plot of the Tanh function

We can build consecutive layers of the neural nets by putting these aforementioned building blocks (neurons and their properties). Take the single layer presented in figure 3.15 as an example. This layer has three inputs with values of and has two neurons that have parameters of $\overline{w_1} = [0.5\ 0.32\ 0.61]^T, b_1 = 0.23$ and $\overline{w_2} = [0.35\ 0.2\ 0.12]^T, b_2 = 0.1$. The neurons also use the sigmoid activation function. The output of this layer can be calculated using the following steps.

$$\overline{x} = [0.33\ 0.5\ 0.7]^T, W = \begin{bmatrix} \overline{w_1}^T \\ \overline{w_2}^T \end{bmatrix} = \begin{bmatrix} 0.5 & 0.32 & 0.61 \\ 0.35 & 0.2 & 0.12 \end{bmatrix}, \overline{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.23 \\ 0.1 \end{bmatrix}$$

$$\Rightarrow z_1 = \overline{w_1}^T x + b_1 = 0.982 \Rightarrow o_1 = \sigma(z_1) = 0.72$$

$$, \quad z_2 = \overline{w_2}^T x + b_2 = 0.399 \Rightarrow o_2 = \sigma(z_2) = 0.59$$

This example was just a simple demonstration of how forward propagation works in neural networks. However, the main question in neural networks is to decide on the value of weight vectors ($\overline{w}$) and biases (b). The purpose of the model is to assign such values to these parameters that the result of the model is the *closest* to the actual output. Many functions have been proposed to evaluate how much loss the model contains, which means that these functions quantify the difference between the expected output and the model output. Here we introduce some of the loss functions used in neural networks.

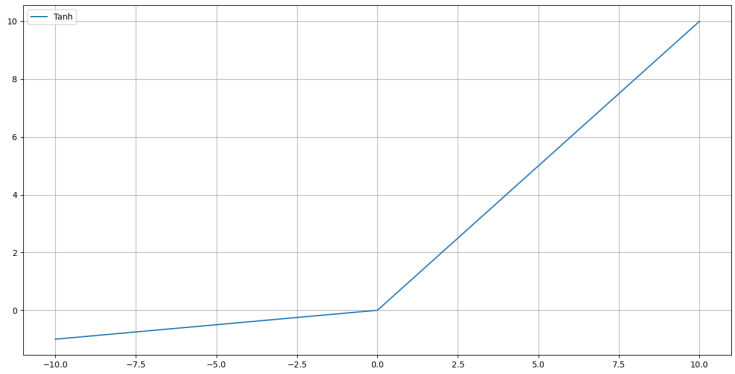

Figure 3.14: Plot of the Leaky ReLU function

Cross-Entropy Loss Function

This function is widely used in classification problems. In these problems, the last layer of the neural network outputs the probability of the input data belonging to each of the observed classes. For instance, if the number of the target classes is four, the final layer would output four values, each between one and zero, indicating the probability of the input data belonging to each of the four classes. By minimizing the cross-entropy loss function, the model tries to maximize the probability of the outputs associated with the actual class label. For one instance of the data, this function is represented with the equation

$$l = -\sum_{i=1}^{C} y_i \log \hat{y}_i \tag{3.36}$$

Squared Errors (SE)

This function is used in regression problems, where the output of the network and the expected output are absolute number values. The SE criterion calculates the squared difference between the model output and the actual output for these problems.

$$l = (\hat{y}_i - y_i)^2 \tag{3.37}$$

It is rather obvious to check that decreasing this function means the model output $(\hat{y}_i)$ is getting closer and closer to the expected output(y_i). The loss functions work only for one data point. If we want to measure the error function for all the inputs,

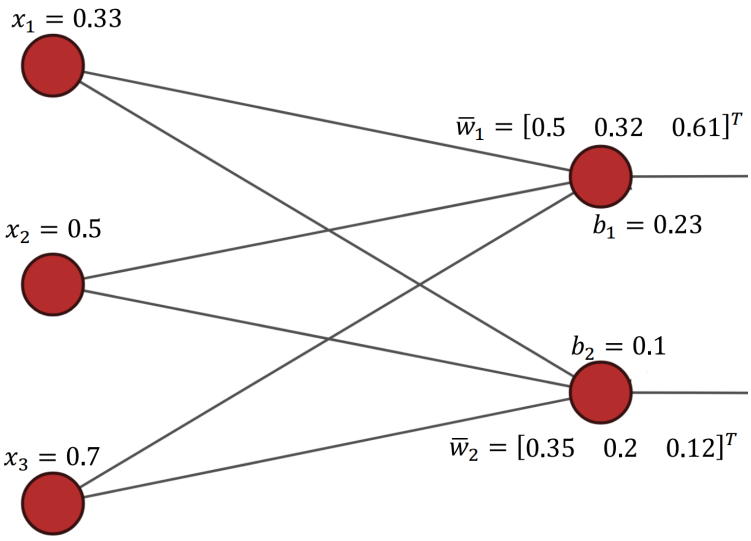

Figure 3.15: An example of a simple neural net layer

we need to take each input's average loss function values. Meaning that if for the input $\overline{x}_i$ the model encounters the loss l_i, for all the inputs, the error would be

$$J(W, b) = \sum_{i=1}^{m} l_i \tag{3.38}$$

Where m is the number of inputs, W is the set of all the weight vectors, and b is the set of all the bias values.

3.4 Data Generation Algorithm

One of the main contributions of this thesis is to use inference methods to predict the trip patterns of the year 2023. We understand that the patterns of the trips can depend on many other variations in the different characteristics of the society. However, the OD and the census datasets had the essential information one needs to know to predict a general trip pattern for the citizens of a given society. The method used in the process tries to exploit even the most minor patterns in these two datasets and capture the variation of the patterns throughout consecutive years.

The primary method used in this study is based on the Bayesian networks that are directed acyclic graphs (DAGs) that represent the dependencies between different characteristics. This chapter discusses Bayesian network concepts and defines how this algorithm is implemented on the OD and census dataset. Finally, we present the results of the algorithm.

3.4.1 Bayesian Belief Networks

As previously stated, bayesian networks (BN) are probabilistic graphical models based on directed acyclic graphs. Using a combination of graph and probability theories, they provide a compact representation of joint probability distributions. The local probabilistic models specify how these variables are combined, while the graph structure specifies statistical dependencies among the variables [48].

To model the state of an automobile, for example, the engine temperature, brake fluid pressure, tire air pressure, and so on can be used. Some of the characteristics that define the state of the automobile are connected to one another, while others are not.

A Bayesian network includes two components, A graph G and a set of distributions Θ. Each node in graph G represents a random variable, and edges represent conditional independence relationships. The set Θ of parameters specifies the probability distributions associated with each variable. Edges generally represent timing dependency (causation), so no directed cycles are permitted. A portion of a Bayesian network is shown in Figure 3.16. This section of the network is made up of a node X with variable values such as x_1, x_2,... as well as parents A and B and children C and D. The connections between nodes are directional and represent a causal relationship (e.g., A influences X or X depends on A)

The joint probability of a set of variables $x_1, ..., x_n$ is given as

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i | x_1, ...x_{i-1}) \tag{3.39}$$

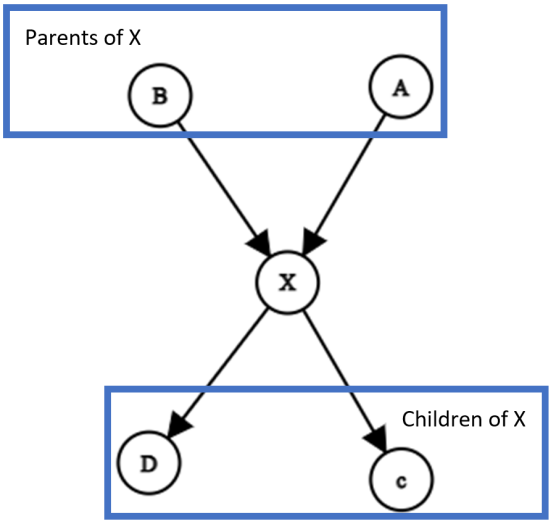

Figure 3.16: Basic example of a Bayesian network

A node x_i is in theory independent of its ancestors given its parents π, according to the conditional independence relationships encoded in the Bayesian network. Therefore

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|\pi_i) \tag{3.40}$$

Where π_i is the set of x_i's parents. We can use marginalization to answer all possible inference questions about the variables once we know the joint probability distribution encoded in the network.

Assume we have a belief net with conditional probabilities and are aware of the values or probabilities of some of the states. The Bayes rule, also known as Bayesian inference, can make informed decisions.

$$P(x_i|x_j) = \frac{P(x_j|x_i)P(x_i)}{P(x_j)} \tag{3.41}$$

The highest posterior value of the unknown variables in the net will be determined. Keep in mind that we only need to think about the connected nodes directly. The rest are conditionally self-contained.

Below are some of the examples of belief networks and their node dependencies.

In figure 3.17, nodes are sequential in terms of their relationship. This means that node B is dependent on node A, node C is dependent on B, etc. Hence, the

joint probability of all the nodes can be represented by

$$P(A, B, C, D, E) = P(A)P(B|A)P(C|B)P(D|C)P(E|D) \tag{3.42}$$

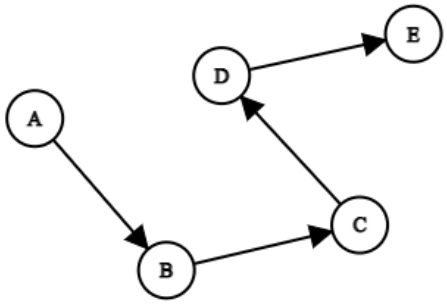

Figure 3.17: BN example 1

For figure 3.18, A and C are both root nodes. Even though all nodes are ancestors of node H, it is conditionally independent of them given their immediate parents, nodes F and G.

$$\begin{aligned} &P(A, B, C, D, E, F, G, H) \\ &\quad = P(A)P(B|A)P(C)P(D|C)P(E|B, D)P(G|E)P(F|E)P(H|F, G) \end{aligned} \tag{3.43}$$

For figure 3.19, since node A is a root node (has no ancestors), the joint probability of the entire system would be

$$P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|A)P(E|B, D) \tag{3.44}$$

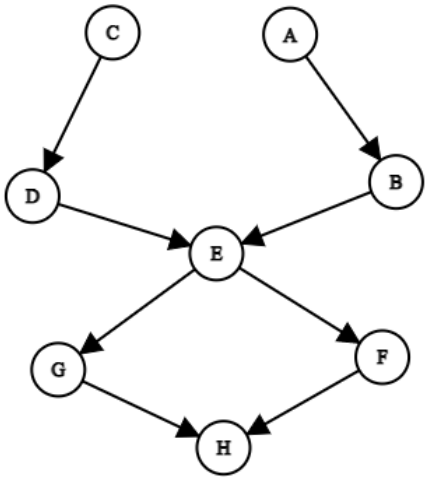

Figure 3.18: BN example 2

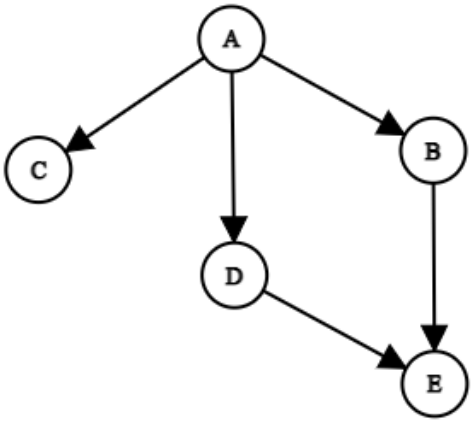

Figure 3.19: BN example 3

Chapter 4

Implementation and Data Processing

This chapter presents the detailed results of the aforementioned methodologies on our datasets. Concepts in this chapter have the same order as the previous chapter. We first depict the meaningful correlations between different trip characteristics using correlation heatmaps. In the next step, we present the configurations of each machine learning model and compare their accuracy on OD surveys. Finally, the result of the method we used to generate the independent variables of OD 2023 and then predicting the mode choice of transportation and destination is presented.

4.1 Feature Correlation

Based on the equations mentioned in 3.1, we have calculated the feature correlation between the features. The quantitative correlation is depicted in heatmaps in figures 4.1, 4.2, and 4.3 for the years 2008, 2013, and 2018, respectively. To calculate the values and also draw the correlation heatmap, we have used the `seaborn` package, a data visualization tool implemented for the `python` programming language.

As mentioned in 3.1, there are three types of correlation, and their evaluation methods differ from one another. Correlations between numerical features can have a value between -1 and 1, indicating a strong negative and a strong positive correlation, respectively. A correlation value of zero in these scenarios means no correlation between features.

On the other hand, correlations that involve categorical features cannot have negative values. In this kind of correlation, the values vary from 0 to 1. Note that this value cannot simply state the qualitative way two features are related. To solve this issue, in the next chapter and in figures to , we have presented bar charts that

present the correlation of features in a more subtle way.

The list below introduces each of the features and their possible values.

- **Categorical Features**
  - Traveler's Sex : *Male / Female*
  - Traveler's License Status: *Yes / No*
  - Time of the Trip: *0-4:59 / 5-8:59 / 9-11:59 / 12-14:59 / 15-18:59 / 19-23:59*
  - Purpose of the Trip: *Commute to Work / Business Meeting / School / Shopping / Entertainment / Visiting a Friend / Health Issues / Driving Someone Back / Looking for Someone / Returning Home*
  - Mode of Transportation: *Self-Drive / Taxi / Bus /Subway / Train / Motorcycle / Bike / Walk*
  - Trip Destination: *Each of the 751 Zones in Montreal*
- **Numerical Features**
  - Number of Vehicles in the Household : *0-14*
  - Number of People in the Household: *1-19*
  - Traveler's Age: *1-99*

Figure 4.1: Correlation heatmap for the trips, year 2008

Figure 4.2: Correlation heatmap for the trips, year 2013

Figure 4.3: Correlation heatmap for the trips, year 2018

4.2 Data Preprocessing and Machine Learning Models

The overall structure of this section is depicted in figure 4.4, and the details of each step have been described in the following subsections. The details of the algorithms of this section were primarily developed using the `numpy`, `pandas`, and `scikit-learn` packages, all three of which are among the most famous tools for data science. Most of the plots have been generated using the `matplotlib` package.

Figure 4.4: preprocessing and machine learning steps in a nutshell

4.2.1 Data Preprocessing

A min-max normalization was applied to the dataset before using the machine learning models. The effect of this normalization can be seen in figure 4.5. For this figure, we have picked 200 random samples of the OD dataset and plotted the data points based on their mode of transportation (which can have eight different values) and trip destination (which can have 751 different values that is the number of census tract zones inside the Montreal CMA.)

4.2.2 Machine Learning Models

Figures 4.6 to 4.9 portray how each machine learning algorithm approaches the classification task. For the purpose of these representations, we have selected 200 OD samples with modes of transportation of Personal Vehicle, Bus, and Subway. The models' function in these figures was to classify the entire 2D plane into its representative mode of transportation. Note that for this part, we have applied the PCA feature reduction algorithm on the entire dataset to decrease the number of practical features to two principal components (PC1 and PC2) in order to make it possible to visualize the result in a 2 dimensional plot. Whereas in the primary model, we have used all selected features.

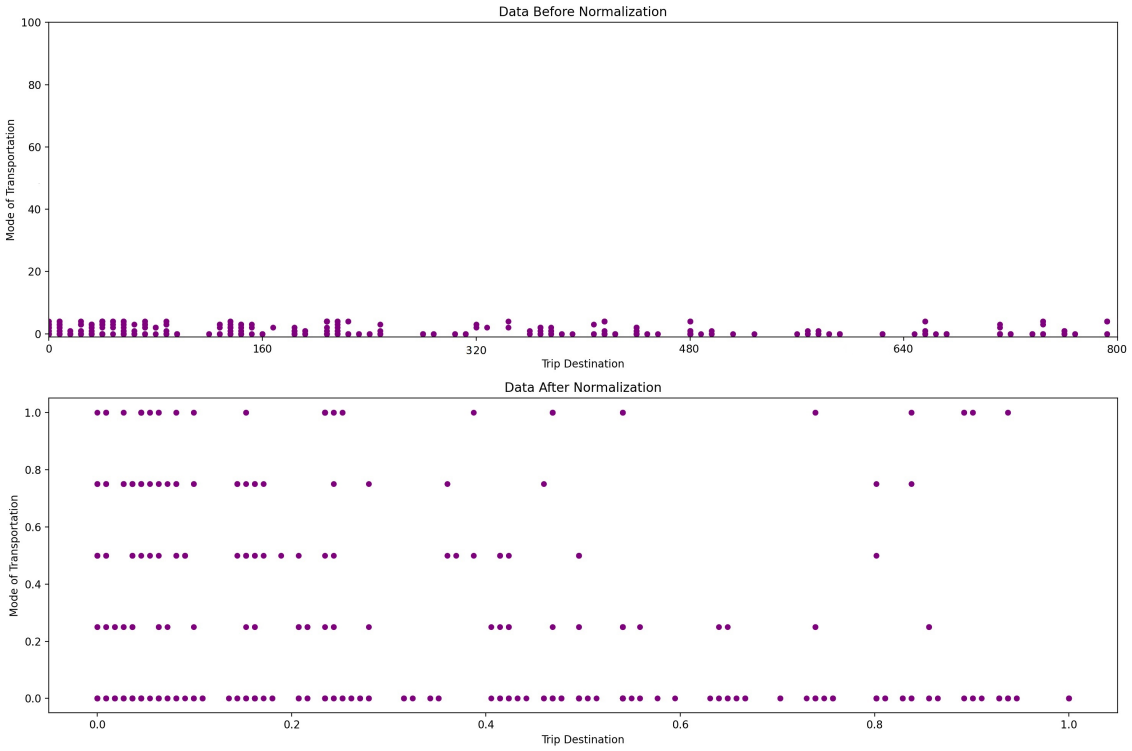

Figure 4.5: The effect of normalization on OD data

After training and testing seven mentioned machine learning models to predict the mode of transportation and trip destination, table 4.1 and table 4.2 compare the results of these algorithms in terms of their accuracy. For all the models, we used OD2008 and OD2013 as the training data, and OD2018 as the test data. Note that we have compared the results of top-1 and top-5 accuracy for each model for the trip destination, whereas, for the mode of transportation, top-1 and top-2 accuracies have been compared. This is simply because the number of possible modes of transportation is much less than the possible number of destinations (8 to 751)

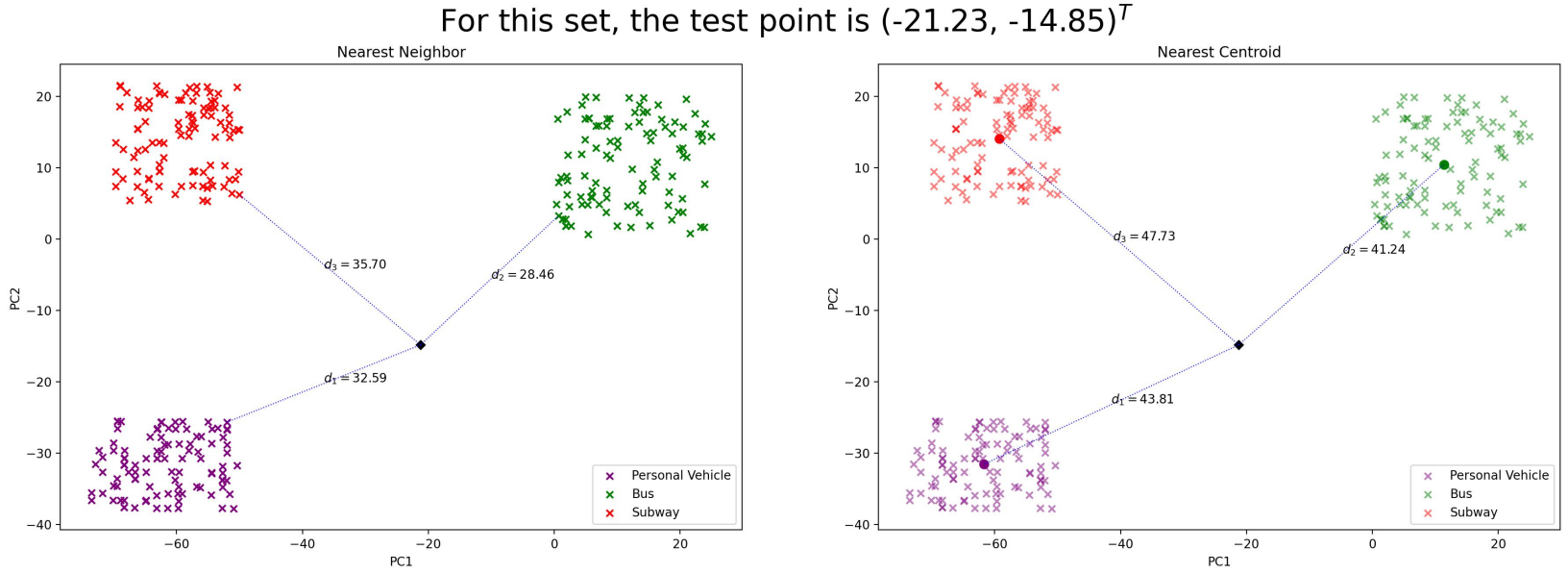

Figure 4.6: Nearest Centroid and Nearest Neighbor algorithm on OD data

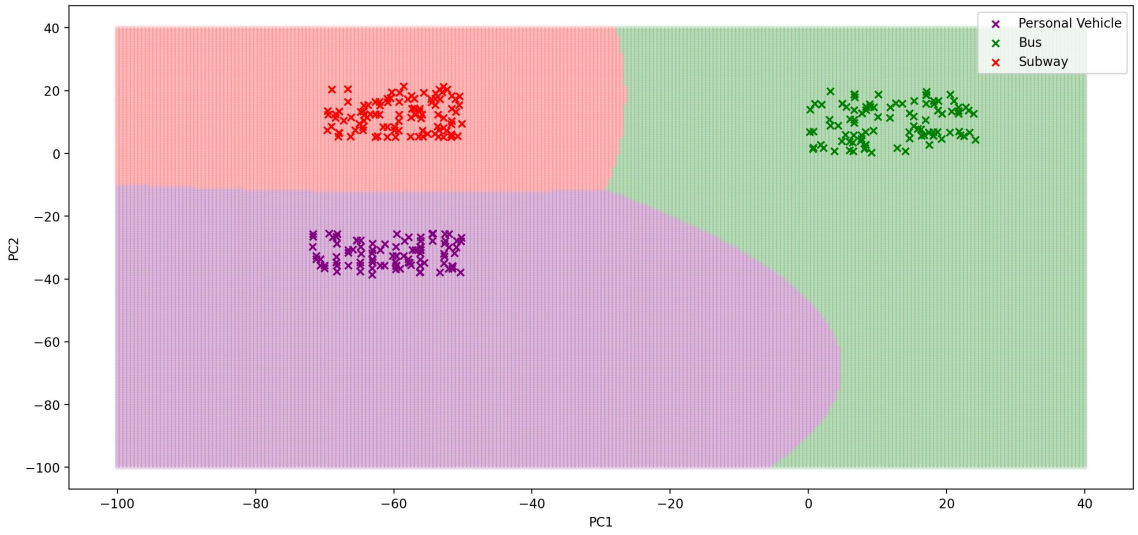

Figure 4.7: Gaussian naive Bayes algorithm on OD data

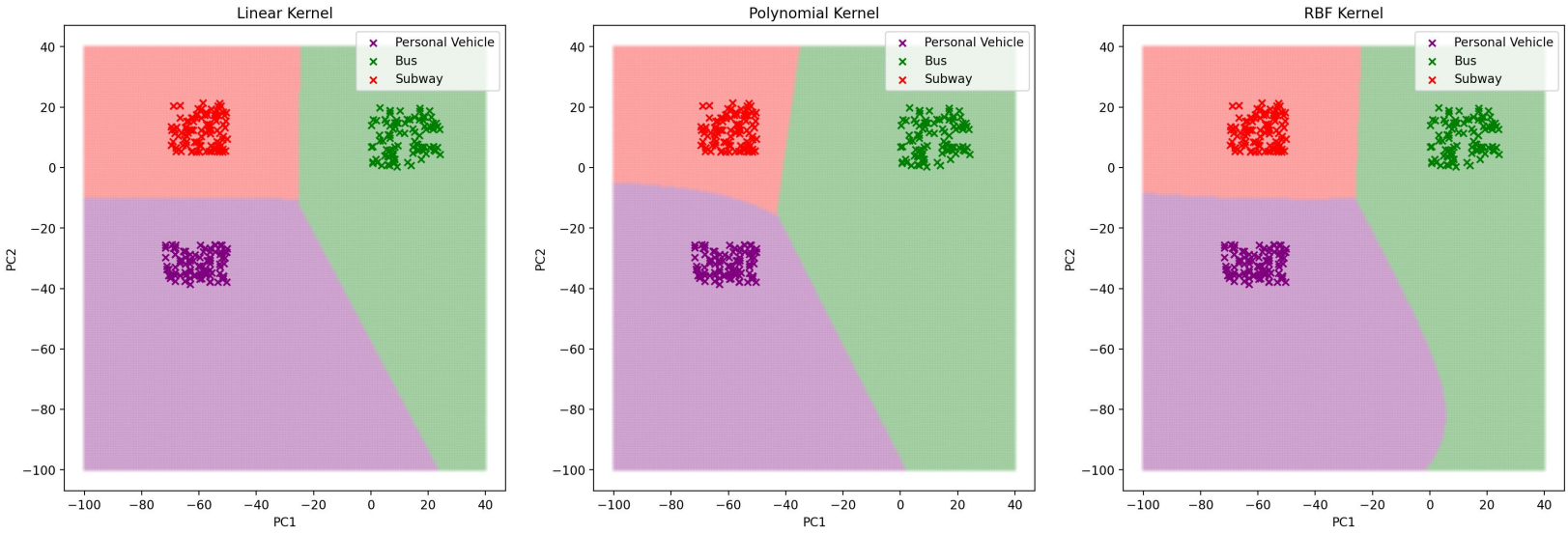

Figure 4.8: Support Vector Machine algorithm on OD data

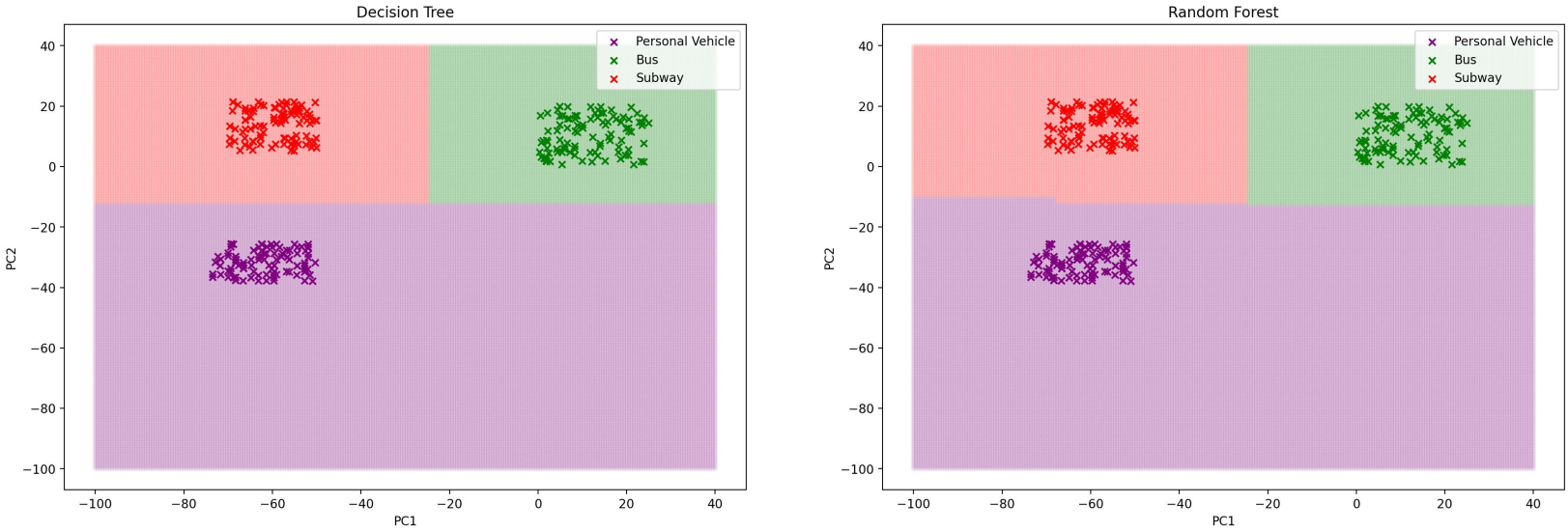

Figure 4.9: Decision Tree and Random Forest algorithm on OD data

4.2.3 Feedforward Neural Network

As you can see in tables 4.1 and 4.2, neural network outperformed all of the other models in terms of their accuracy. Hence we extended our study on this particular model.

- The input of the models is in $\mathbb{R}^9$, meaning it has nine independent features. (Remember that the number of the features selected was eleven, and since we are trying to predict two of them, the input is left with nine features). The output is in $\mathbb{R}^8$ for the mode of transportation and $\mathbb{R}^{751}$ for the trip destination. (Number of classes in each target)

- We start adding more layers to the network in the tuning phase by using test and trial to deepen the network. The model's accuracy was highest after adding seven layers. Adding further layers caused the model to overfit to the train data and perform poorly on the test data. Hence, the final model has seven layers. The number of neurons in each hidden layer has also been decided upon trial and error and chosen based on the highest accuracy.

- For hidden layers of the network, we used the standard ReLU function. This function adds non-linearity to the model and enables it to exploit complex relationships between the input and the output. However, for the final layer, we used the Sigmoid function. This function outputs values between zero and one, which can be perceived as the probability of the input belonging to a particular output class. Therefore, it is used for classification tasks, such as ours.A general model of our neural network used is presented in figure 4.10.

This model, with the aforementioned settings, yielded a much higher accuracy than other settings; hence we reconfigured its combination of the training data many times and captured the difference in each configuration. We ran our neural network once with the train set and the test of the same year (one for each year separately), and then we trained it on OD2008 and OD2013 and tested it on OD2018. Since the correlation of the features of the trips had almost stayed the same throughout the years, we expected that regardless of using different years as the train and the test set, the model would yield high accuracy with higher volumes of training data. This fact is precisely what we observed, which has also been presented in tables 4.3 and 4.4. The idea that a model trained on previous years can be used in the future with a slight drop in accuracy shows the generality of our model.

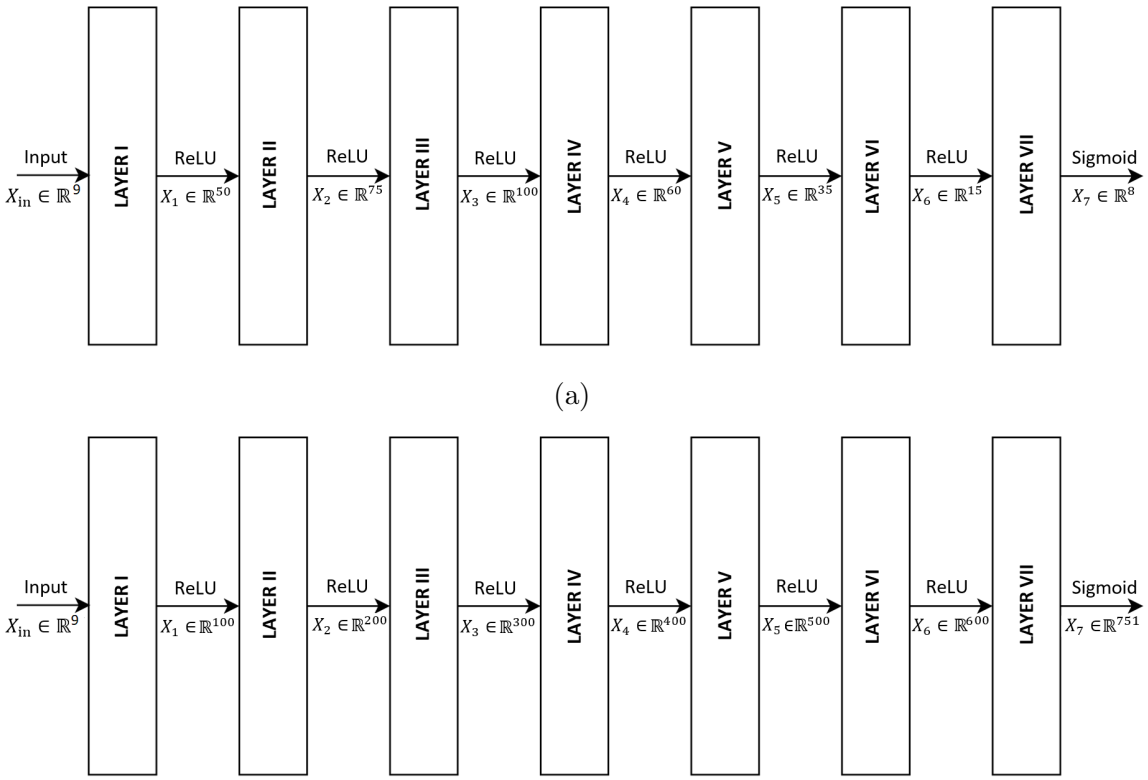

Figure 4.10: The neural networks used to predict the mode of transportation(a) and the trip destination(b)

Table 4.1: Accuracy comparison of machine learning models for destination prediction

Algorithm		**Top-1 Accuracy**	**Top-5 Accuracy**
Nearest Centroid		5.83%	19.63%
KNN	1-NN	26.80%	34.23%
	2-NN	24.32 %	33.12%
	89-NN	27.02%	34.34%
	90-NN	27.28 %	35.53%
Naïve Gaussian		13.21%	41.22%
Decision Tree		60.42%	63.27%
Random Forest		60.81%	75.28%
SVM	linear kernel	14.06%	17.21%
	RBF kernel	16.71%	23.11%
	polynomial kernel	21.96%	24.72%
Neural Network (MLP)		68.71%	85.43%

Table 4.2: Accuracy comparison of machine learning models for mode of transportation prediction

Algorithm		**Top-1 Accuracy**	**Top-2 Accuracy**
Nearest Centroid		48.59%	62.23%
KNN	1-NN	55.38%	61.38%
	2-NN	59.28 %	69.83%
	89-NN	62.96%	71.44%
	90-NN	62.85 %	71.58%
Naïve Gaussian		68.51%	82.68%
Decision Tree		65.40%	74.04%
Random Forest		72.81%	88.51%
SVM	linear kernel	70.34%	75.31%
	RBF kernel	71.68%	78.03%
	polynomial kernel	71.84%	79.23%
Neural Network (MLP)		78.09%	92.81%

Table 4.3: Accuracy comparison for different combination of the train and test data configurations of the neural network for trip destination prediction

Train Set	**Test Set**	**Top-1 Accuracy**	**Top-5 Accuracy**
OD2008 (80% sample)	OD2008 (20% sample)	67.16%	83.70%
	OD2018	66.96%	83.51%
OD2013 (80% sample)	OD2013 (20% sample)	67.92%	83.17%
	OD2018	67.31%	82.77%
OD2018 (80% sample)	OD2018 (20% sample)	68.17%	84.67%
OD2008, OD2013	OD2018	68.71%	85.43%

Table 4.4: Accuracy comparison for different combination of the train and test data configurations of the neural network for mode of transportation prediction

Train Set	**Test Set**	**Top-1 Accuracy**	**Top-2 Accuracy**
OD2008 (80% sample)	OD2008 (20% sample)	73.51%	90.71%
	OD2018	73.17%	90.02%
OD2013 (80% sample)	OD2013 (20% sample)	77.11%	91.07%
	OD2018	75.39%	91.67%
OD2018 (80% sample)	OD2018 (20% sample)	77.63%	91.72%
OD2008, OD2013	OD2018	78.09%	92.81%

4.3 OD's Independent Variable Generation

In this section we elaborate on the methods of generating independent features for the next version of the OD data. The dependent features were the mode of transportation and the trip destination. We have already trained an MLP model based on historical data to predict these two features by using neural network. Now we have to generate other nine independent features for future in order to predict the mode choice of transportation and destination for the year 2023.

Since OD data is collected every five years, it just makes sense to call the next version of the OD as OD2023. (The latest version is OD2018) Therefore, we label our generated OD as OD2023.

4.3.1 Deriving the Probability Distribution Using the Bayesian Belief Network

Each of the nine independent characteristics in the OD could have a finite number of discrete values. For instance, a traveler's sex is a binary value, and the number of people in their household is an integer between one and thirteen. Hence, the number of parameters required to establish a row inside the OD data is limited.

This study used Bayesian belief networks to calculate the probability of each feature having a specific value. In other words, Bayesian networks were used to derive the probability distribution of the data.

As discussed in 3.4, the Bayesian network needs to define the "parent" feature of each characteristic. In other words, the diagram of parent-child correlation between independent features should be determined before any further calculations. We used the correlation heatmap of the characteristics from previous OD records to derive this structure. We assigned a parent-child relationship to pairs of characteristics with the highest correlations throughout the last three years (2008, 13, and 18). Figure 4.11 represents the Bayesian network that we used for our dataset.

To elaborate on how the network works, consider three features, trip origin (denoted by α), the purpose of the trip (β), and the number of people in the household (γ). Based on figure 4.11, the joint probabilities of these features (based on the figure 4.11) can be represented as

$$P(\alpha, \beta, \gamma) = P(\alpha)P(\beta|\alpha)P(\gamma|\beta) \tag{4.1}$$

Using the distribution $P(\alpha, \beta, \gamma)$ we can generate the sample data with three features of α, β, and γ. However, for calculating the joint probability, we still need $P(\alpha)$, $P(\beta|\alpha)$, and $P(\gamma|\beta)$. To derive these basic probabilities, we simply used the OD

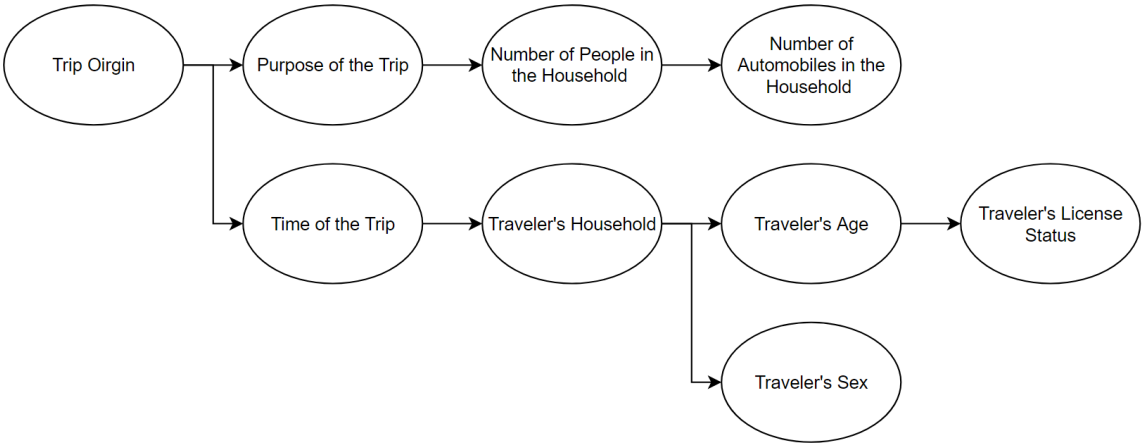

Figure 4.11: Bayesian parent-child correlation diagram used to derive the probability distribution of the independent features in OD data

data. Since each feature only had one parent, we only had to derive the distribution of one feature, or at most a pair of features each time. For instance if α could have values of $\{\alpha_1, ..., \alpha_n\}$, the probability of $\alpha = \alpha_i$ is calculated as

$$P(\alpha = \alpha_i) = \frac{N(\alpha = \alpha_i)}{N} \tag{4.2}$$

where N is the size of the dataset, and $N(\alpha = \alpha_i)$ is the number of times that the α feature has the value of α_i. The conditional probability $P(\beta = \beta_j | \alpha = \alpha_i)$ can be computed using the Bayes theorem.

$$P(\beta = \beta_j | \alpha = \alpha_i) = \frac{P(\alpha = \alpha_i, \beta = \beta_j)}{P(\beta = \beta_j)} \tag{4.3}$$

4.3.2 Generating the Upcoming Data Distribution Using Linear Fitting

Using the discussed Bayesian network in the previous section, we derived the necessary distribution that makes the overall distribution of the entire OD data in each year. (from 2003 to 20018) To generate the probability distribution of the next version of the OD data, we used linear fitting. For instance, to calculate $P_{23}(\alpha)$ we fitted a linear model on $P(\alpha)$ from previous OD records. ($P_{03}(\alpha)$, $P_{08}(\alpha)$, $P_{13}(\alpha)$, and $P_{18}(\alpha)$). However, there an important issue to note here:

The probability distribution of the traveler's age and sex based on their household zone has been derived using a linear fit on the census dataset. Since this dataset targeted the residents in each zone and their personal traits, it was a more reliable source of information for the two features mentioned above.

4.4 OD2023 Generation and Testing

Finally, we fed the nine features generated by a Bayesian network for 2023 (features in figure 4.11) into our trained MLP model to predict the mode of transportation and trip destination. This MLP was trained on OD2008, Od2013 and OD2018. By concatenating these two features with the independent ones, the entire new version of the OD is complete. An overview of the whole process of this generation is depicted in figure 4.12.

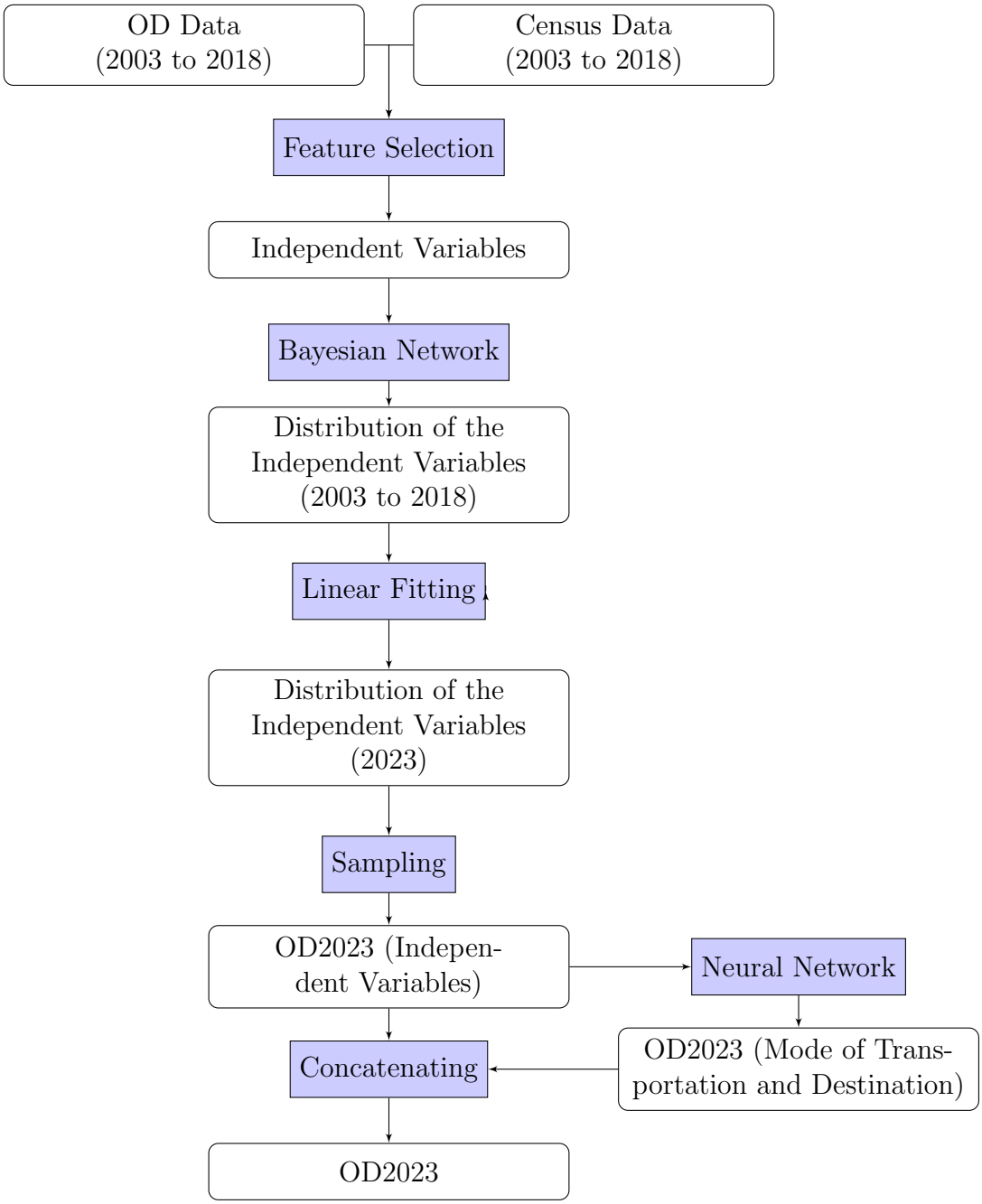

Figure 4.12: Overview of the workflow for predicting OD2023 Mode choice of transportation and destination.

Testing this model was quite challenging. Since we did not have any ground truth data to compare our results, we first used the same methodology to generate the OD2018 and compared our results to the actual OD2018. (Keep in mind that for generating OD2018, we only used OD2003, OD2008, and OD2013, whereas, for OD2023, we used OD2018 as on top of these datasets.) Figures 4.13 to 4.16 compare some of the feature histograms of the actual OD2018 and our created ones. As you can see, our model has performed well while trying to generate a miniature version of OD2018.

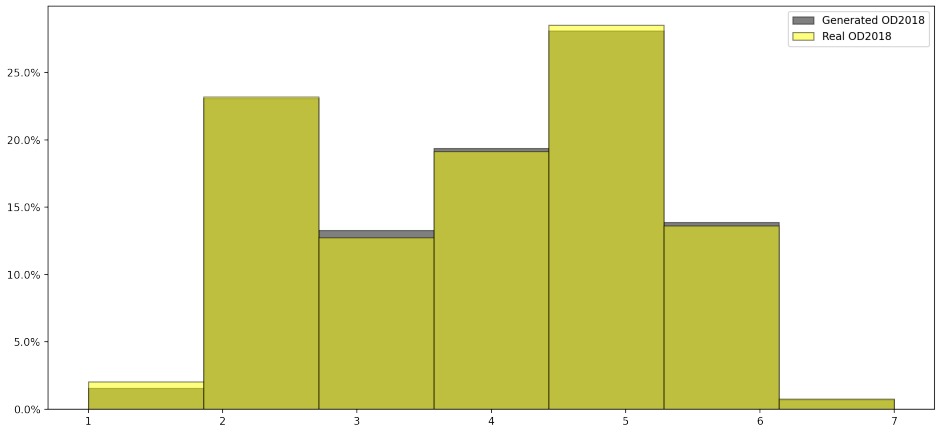

Figure 4.13: Comparison of the time group of the trip for the real OD2018 and generated OD2018

In the final step of the data generation, we depicted the correlation heatmap of the features in the generated data (OD2023) in figure 4.17. Also, figures to represent how features are correlated with one another in the generated OD2023, compared to previous OD records.

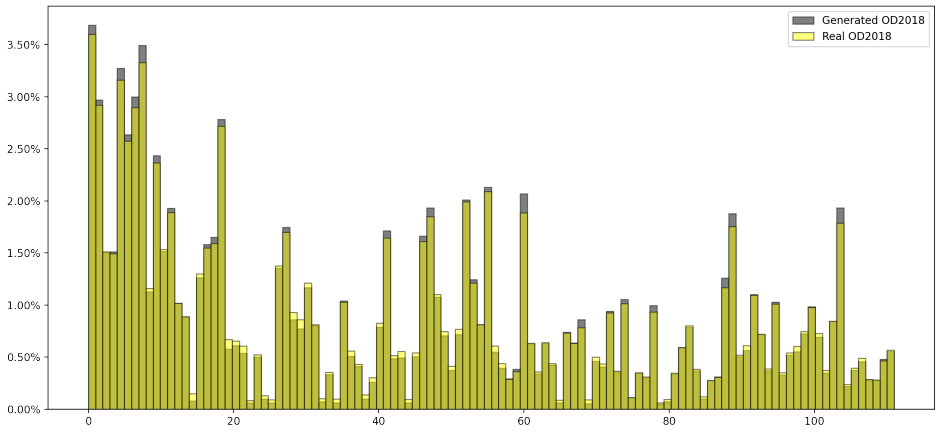

Figure 4.14: Comparison of the origin of the trip for the real OD2018 and generated OD2018

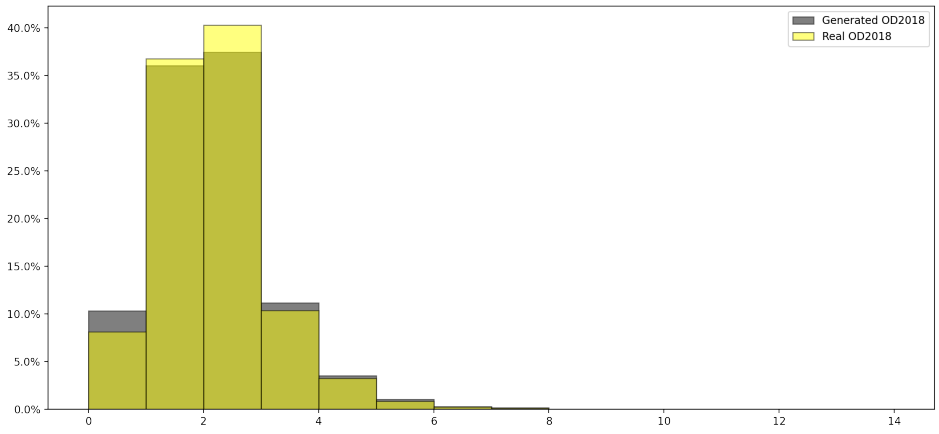

Figure 4.15: Comparison of the number of automobiles in household for the real OD2018 and generated OD2018

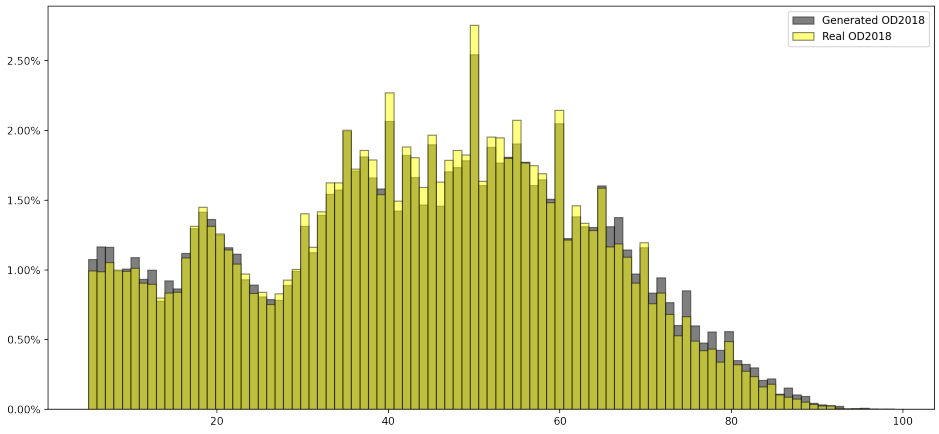

Figure 4.16: Comparison of the traveler's age for the real OD2018 and generated OD2018

Figure 4.17: Correlation heatmap for the trips, year 2023 (Generated)

Chapter 5

Data Interpretation

In this chapter, we consider some of the trends in the trips patterns of the citizens of Montreal and how they changed throughout the last couple of years. More specifically, we divide this chapter into two major sections.

In the first section of the chapter, using the landuse dataset, we chose four zones (census tracts) of Montreal, a residential zone, a commercial zone, an industrial zone and a green-space zone. From 2012 to 2020, we can see how the land use characteristics of each zone have changed over time. Based on the information available, we can see a link between land use characteristics and travel patterns that each zone produces or attracts.

In the second section of this chapter, we take a broad look at all of the trips of the citizens of Montreal. We aim to study **how** the different characteristics of the trips are correlated with one another. The figures represented in this chapter make it easier to understand the values of the heatmaps represented in the previous chapter.

5.1 Example Zones

We chose four zones that were some of the trip's hotspots for this section. A commercial zone, an industrial zone, a residential zone, and a green space area are the four zones. The usage of each of these zones was characterized by the area of the buildings inside of them, indicated by the landuse dataset. Figures 5.1 and 5.3, show how the utility of the buildings of the industrial and commercial zones have changed throughout the years 2014, 2016, 2018, and 2020. Figures 5.2 and 5.4 does the same job on a geographical map. For the other two zones, please refer to Appendix.

5.1.1 Commercial (With Code 0.0326)

Because of its shopping centers, this area is classified as a commercial zone. However, the area of industrial buildings in the zone is also significant. In addition, as shown in figure 5.1, some commercial buildings have been demolished and turned into vacant land in recent years.

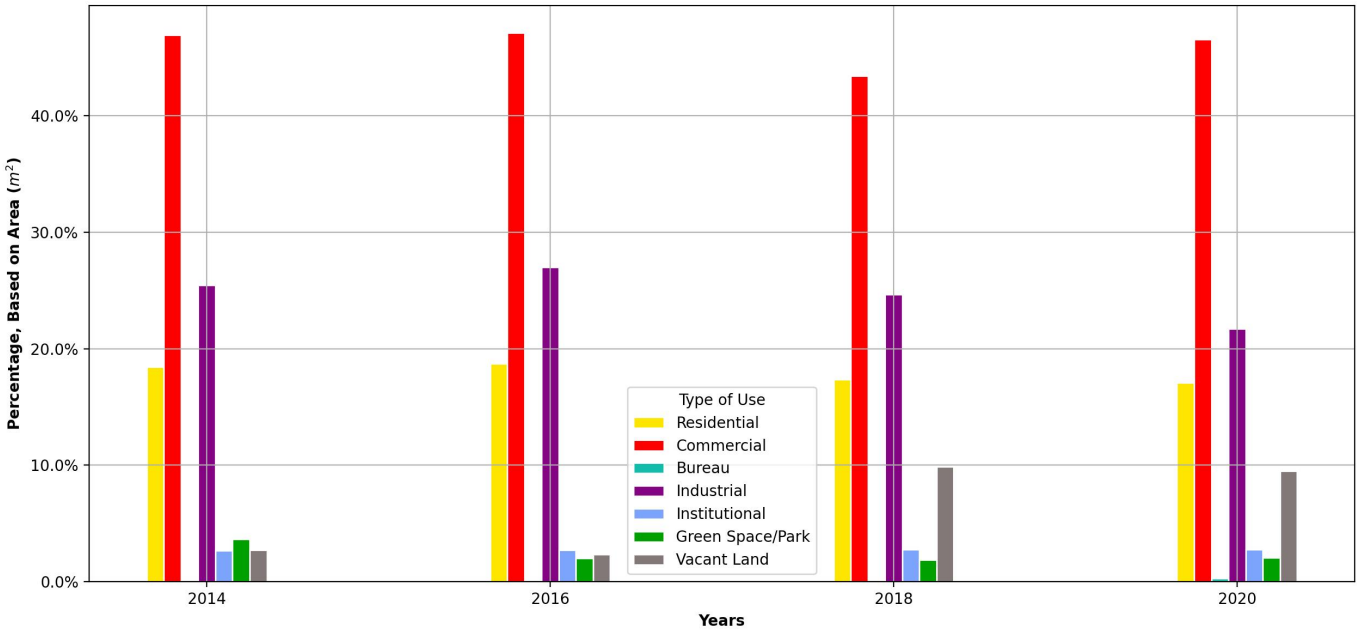

Figure 5.1: Area Distribution of Different Types of Places in the Commercial Zone (Code 0.0326)

Figure 5.2: Land Use characteristics of a Commercial Zone (Code 0.0326) Throughout Recent Years

5.1.2 Industrial (With Code 0.0642)

This area is currently an industrial zone transitioning to a commercial one. As shown in figure 5.3, the volume of commercial buildings is growing over time. These last two zones demonstrate a trend in which industrial and commercial zones are mixed to create more enriched surroundings.

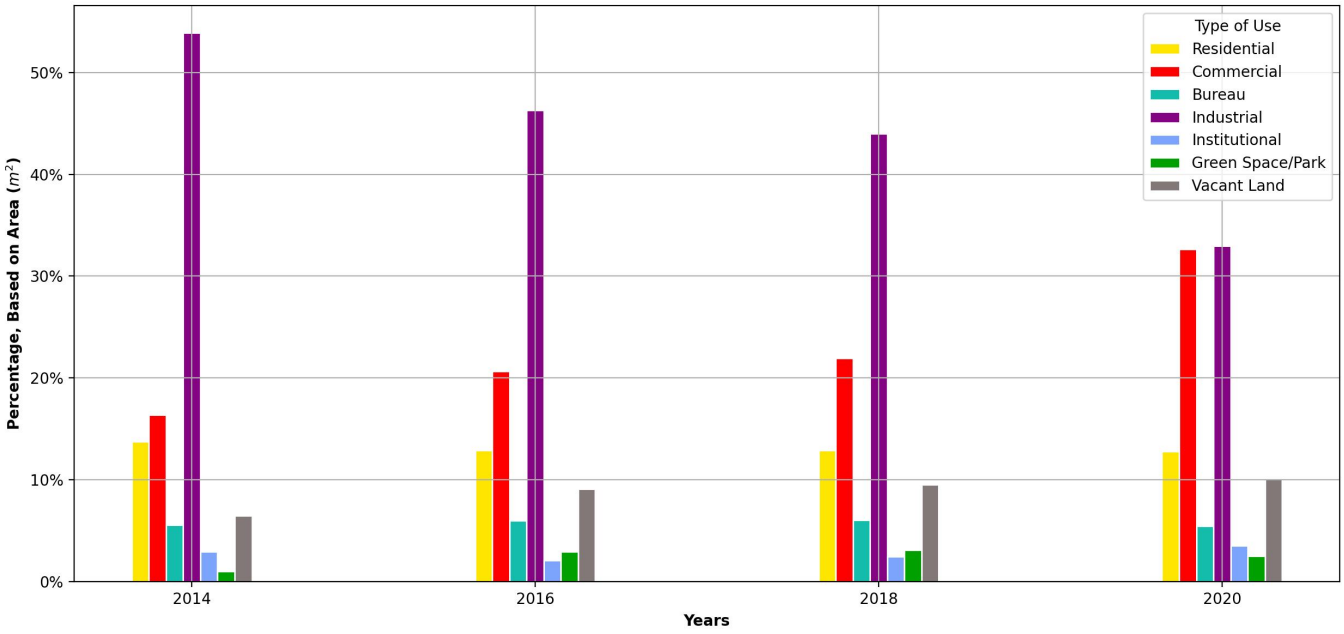

Figure 5.3: Area Distribution of Different Types of Places in an Industrial Zone (Code 0.0642)

Figure 5.4: Land Use characteristics of an Industrial Zone (Code 0.0642) Throughout Recent Years

5.1.3 Attracted Trips

This subsection will look at the motivation, mode of transportation, and approximate time of trips attracted by each of these zones.

Commercial Area

According to the commercial characteristics of this zone, the majority of the trips attracted to this area are either going to work or shopping, which accounts for nearly 60% of all trips attracted to this area.

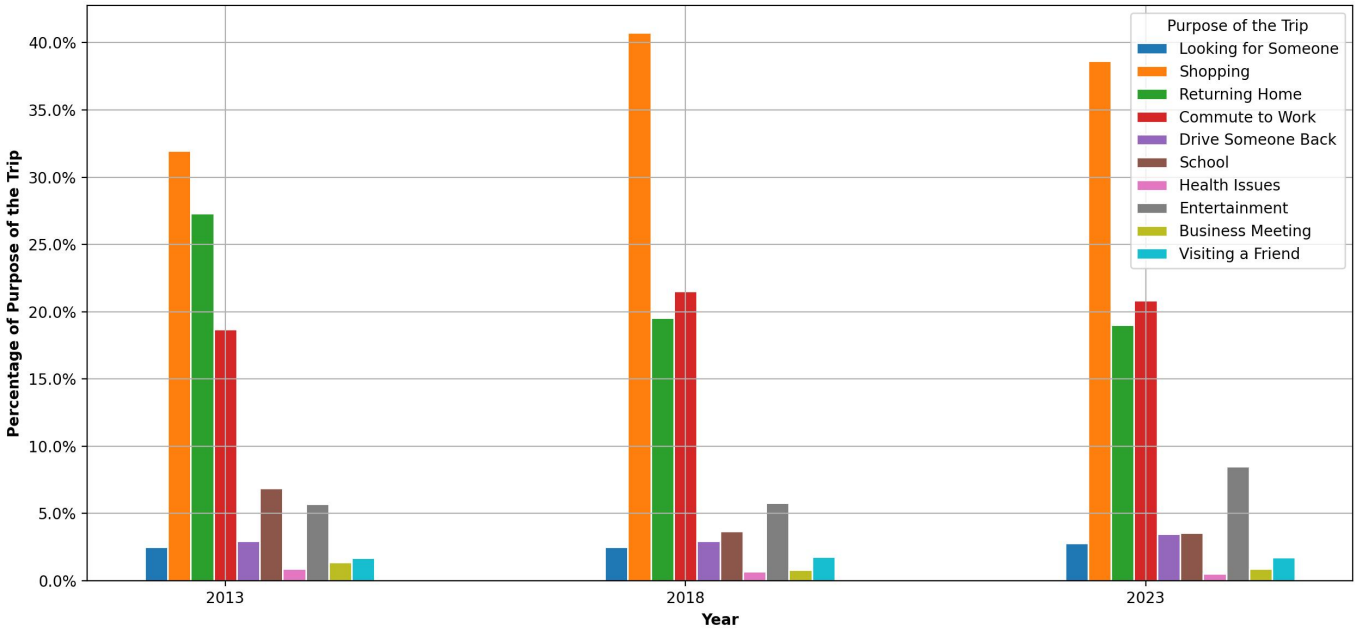

Figure 5.5: Purpose Distribution of the Trips Attracted by a Commercial Zone

According to projections, the personal car is the primary mode of transportation used to reach this zone, which has been increasing year after year and is expected to reach more than 65 percent by 2023. Another point worth mentioning is that this area is almost adjacent to one of the main subway stations on the Montreal green line, but as we can see, only 5% of the trips drawn by this are by subway. As a result, we can conclude that the majority of commuters believe that accessing

this zone is much more accessible by personal vehicle than by public transportation. The regional municipality should invest more in cycle paths or re-evaluate their walkability standards in the mentioned region and invest more in optimizing their bus network in the area to attract more commuters to increase the use of green transportation methods.

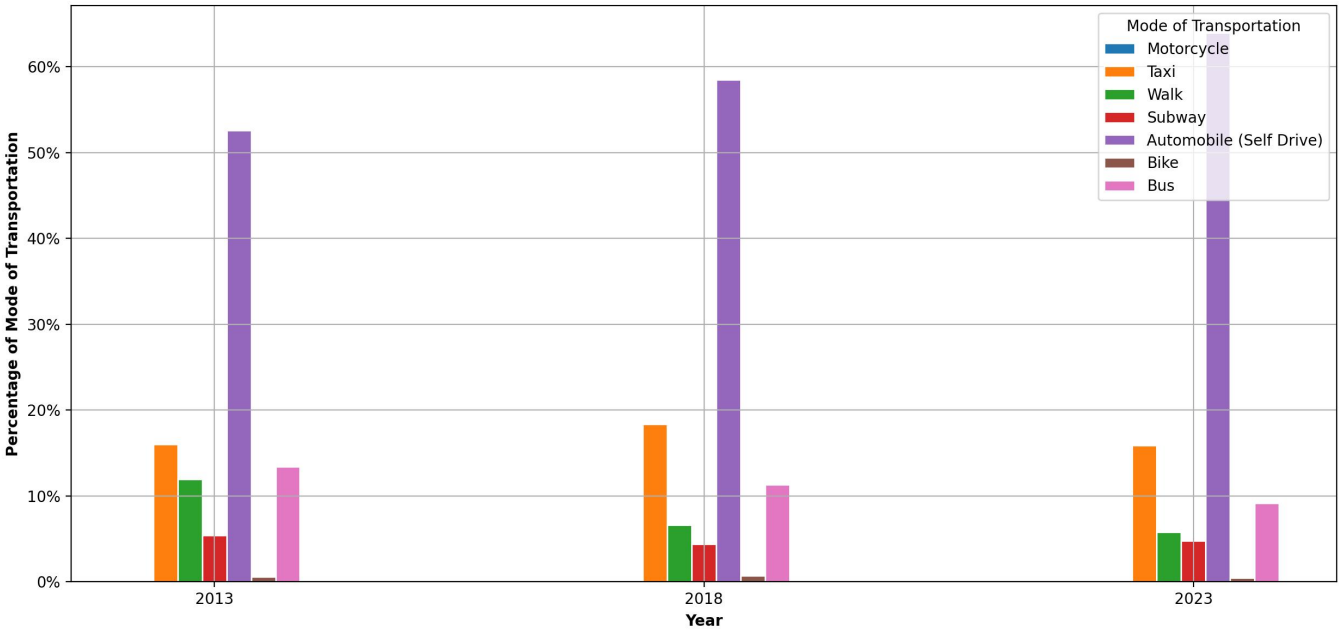

Figure 5.6: Mode of Transportation Distribution of the Trips Attracted by a Commercial Zone

The approximate time distribution of the commercial zone-attracted trips shows that the majority of the trips occurred between 5-18:59. Based on the purpose of the trips discussed previously, shopping was the most popular, so we can see that the number of trips attracted by the zone is around 10% in total during the closure or before the opening of shops.

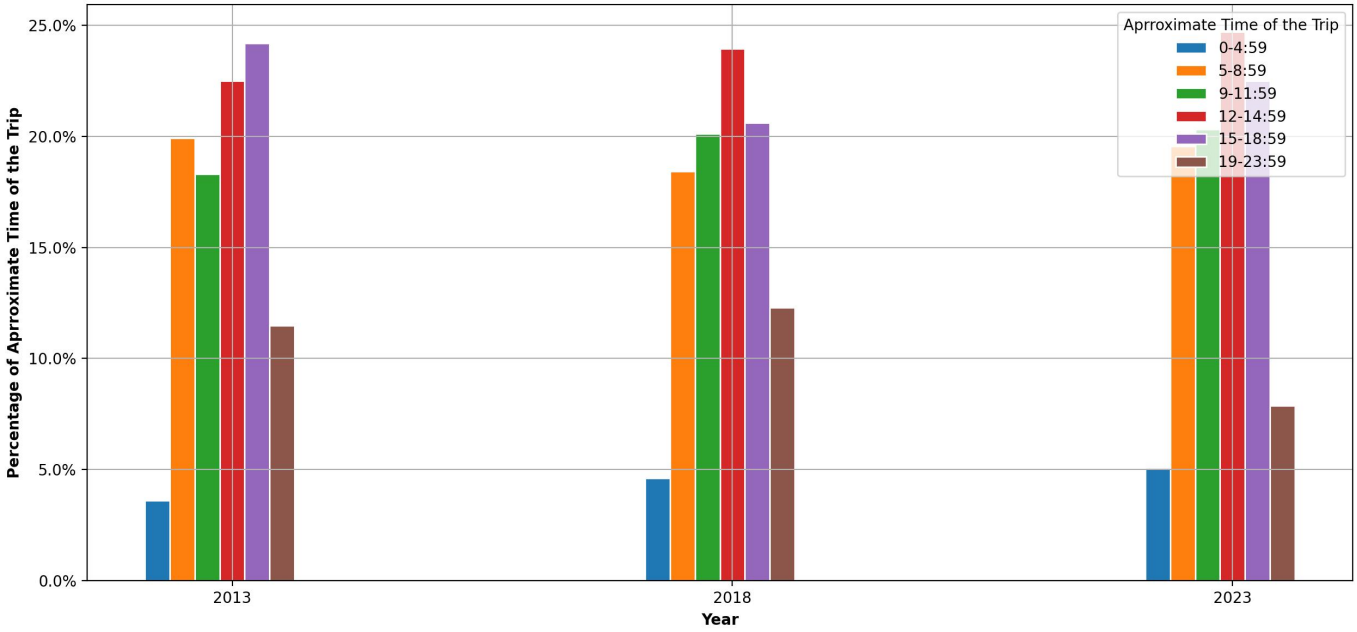

Figure 5.7: Approximate Time Distribution of the Trips Attracted by a Commercial Zone

Industrial Area

This zone, as previously stated, is primarily a mix of industrial and commercial characteristics. In this case, shopping and commuting to work account for more than 68 percent of all trips attracted by this zone.

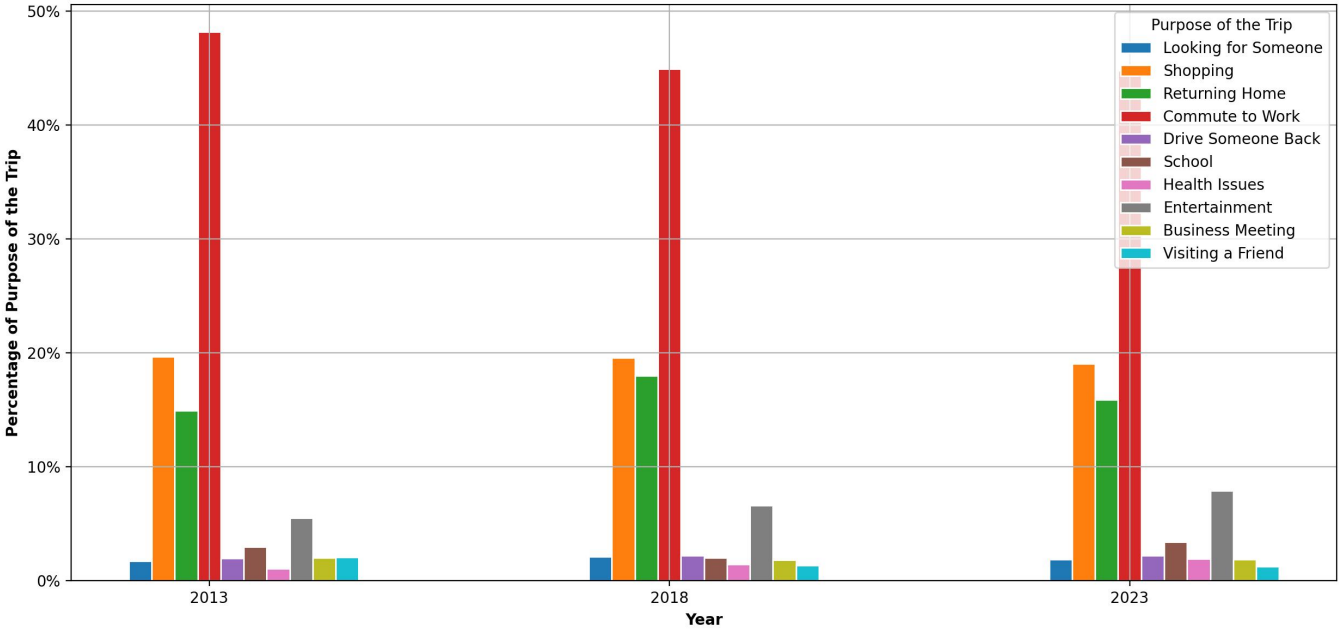

Figure 5.8: Purpose Distribution of the Trips Attracted by an Industrial Zone

As the commercial zone, the personal vehicle is the primary mode of transportation for people commuting to this industrial zone, accounting for nearly 80% of all commutes. Unfortunately, public transportation, cycling, and walking account for less than 10% of the total mode of transportation. This analysis can assist us in identifying areas where public transportation is insufficient in each zone.

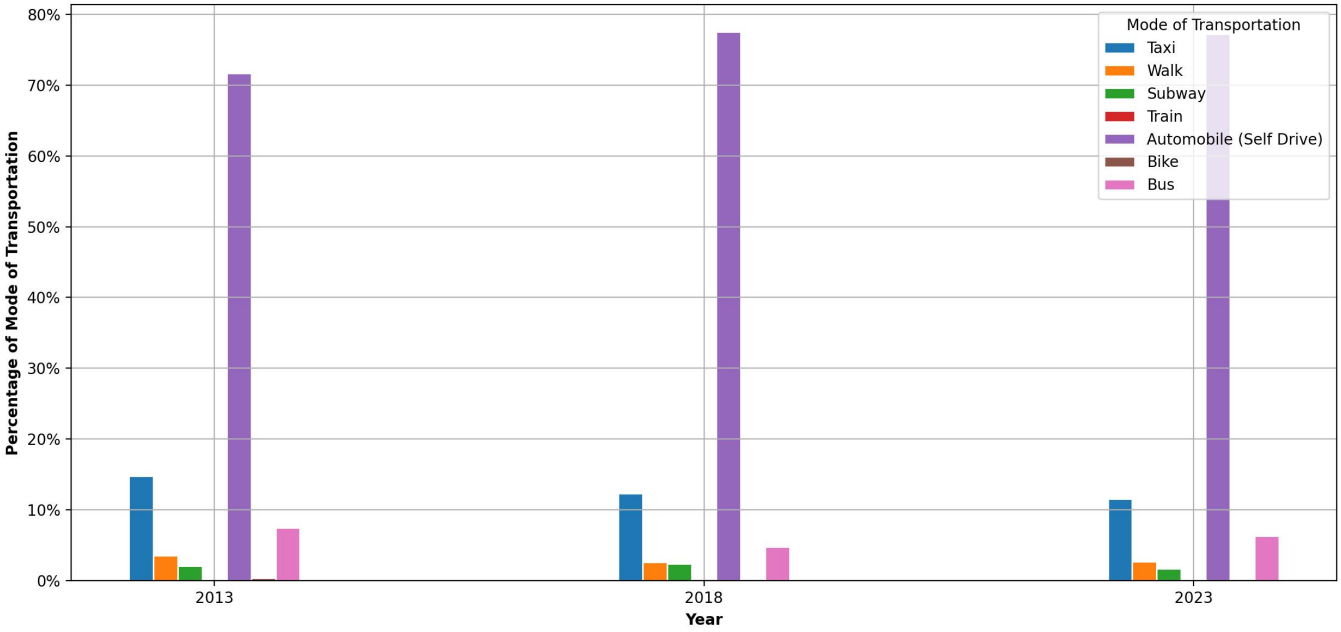

Figure 5.9: Mode of Transportation Distribution of the Trips Attracted by an Industrial Zone

We identified that commuting to work has the highest frequency percentage of trips attracted to this zone based on the motivation of the trips, so the approximate time distribution of the trips shows that these trips mostly happen between 5-8:59 when employees want to get to their working location within this zone.

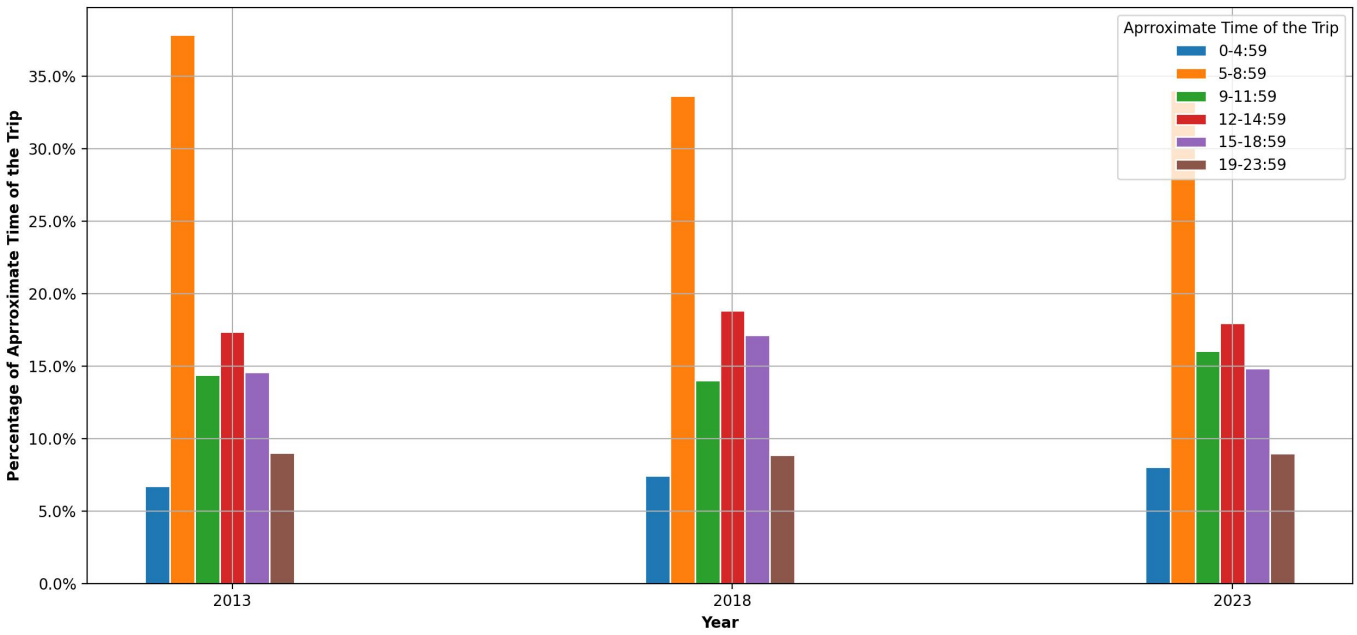

Figure 5.10: Approximate Time Distribution of the Trips Attracted by an Industrial Zone

5.1.4 Produced Trips

This section will look at the motivation, mode of transportation, and approximate time of trips produced in each of these zones.

Commercial Area

The primary purpose of this zone's trips is to return home. We know this is a commercial zone based on the analysis of the land use characteristic, so going back home as the primary purpose with more than 65 percent of all frequencies is an accurate analysis.

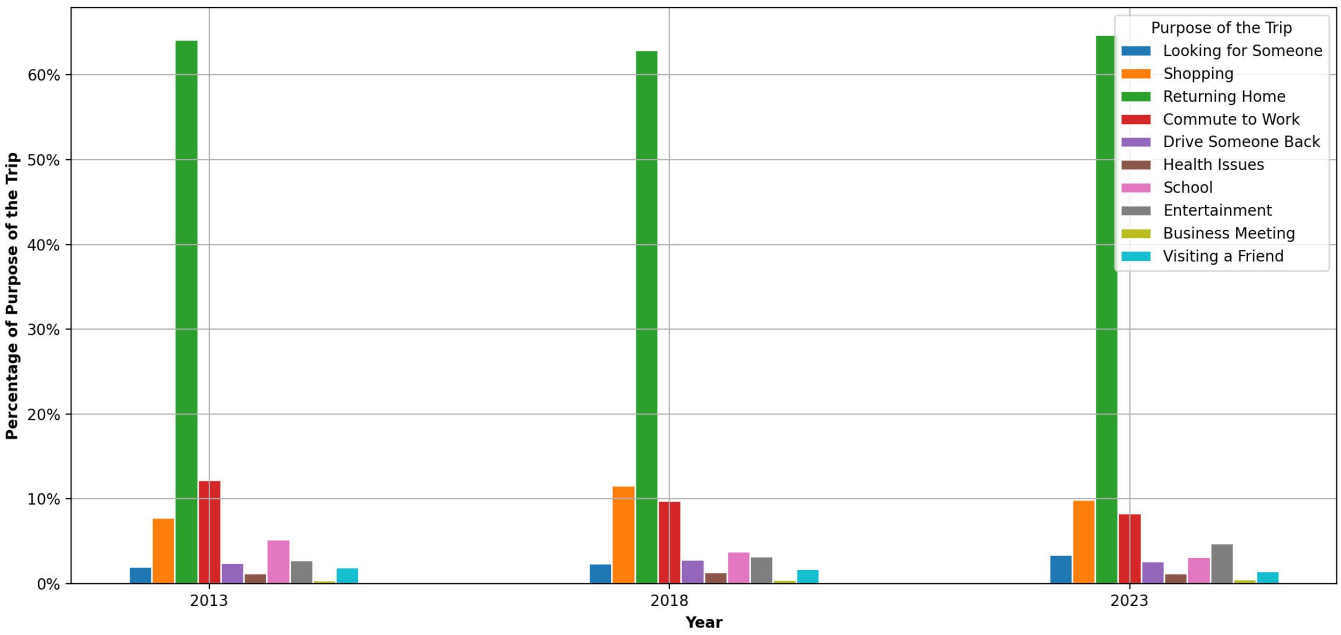

Figure 5.11: Purpose Distribution of the Trips Produced by a Commercial Zone

The primary mode of transportation for the produced trips is still the personal vehicle, which is understandable given that commuters who used their car to get to this zone are now using it to return home. Also, this trend is growing continuously through 2023.

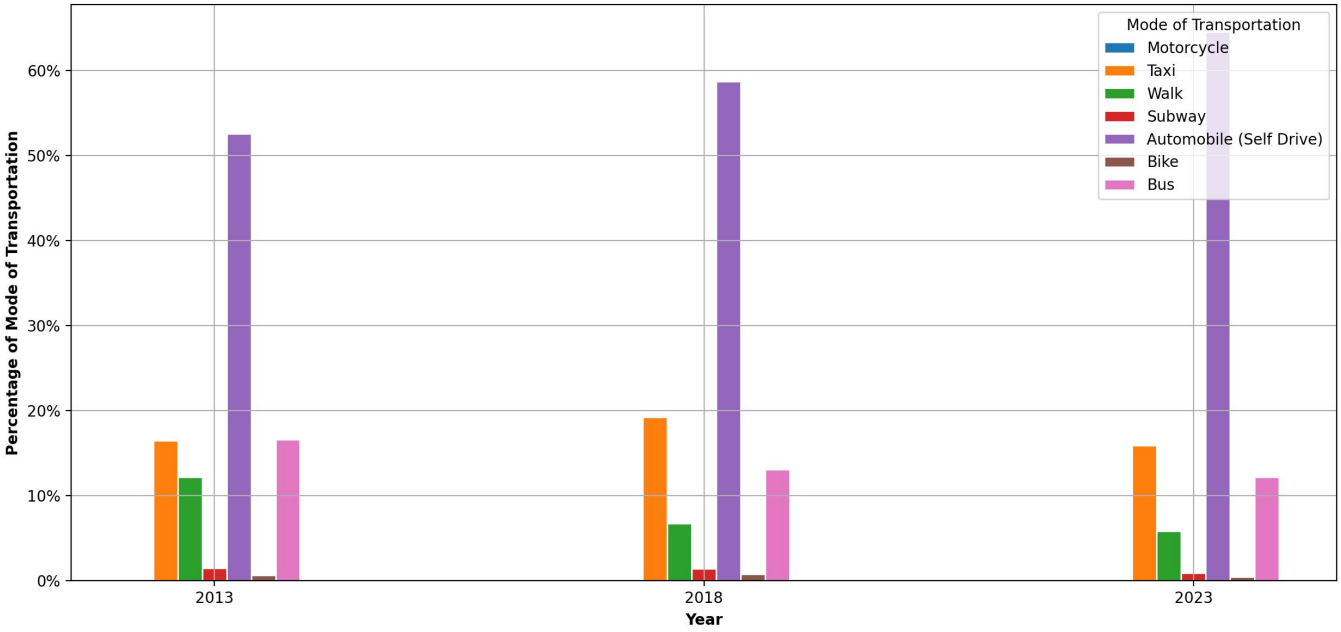

Figure 5.12: Mode of Transportation Distribution of the Trips Produced by a Commercial Zone

The approximate time of the trips this zone produced is the next feature after analyzing the purpose and mode of transportation. Based on the majority of the trips' purpose, which was to return home, we can deduce that they began between 12 and 18:59, when either people were returning from shopping or employees were returning home after working hours.

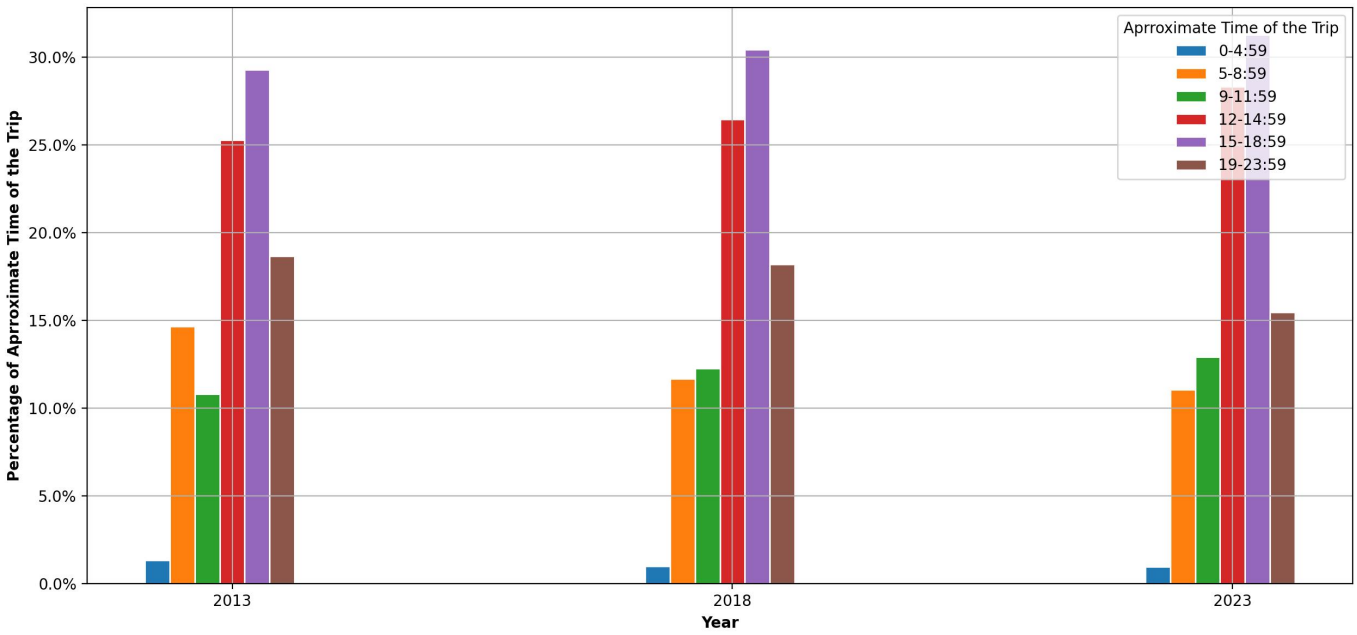

Figure 5.13: Approximate Time Distribution of the Trips Produced by a Commercial Zone

Industrial Area

Similar to the commercial zone we investigated previously, we can see that nearly 70% of all trips produced by this industrial zone are for the purpose of returning home.

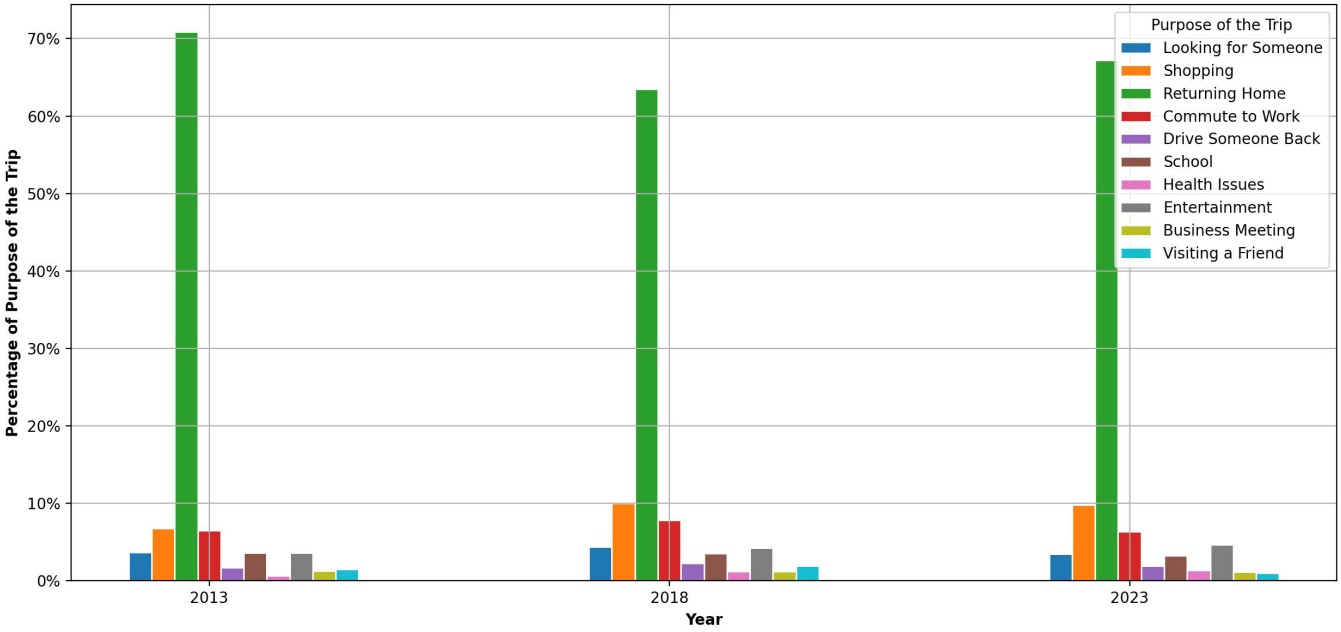

Figure 5.14: Purpose Distribution of the Trips Produced by an Industrial Zone

We also discussed the lack of adequate access to public transportation and other modes of transportation other than personal vehicles in this zone, so, like the attracted trip, the majority of the produced trips in this zone are done by personal vehicles, accounting for more than 70% of the total.

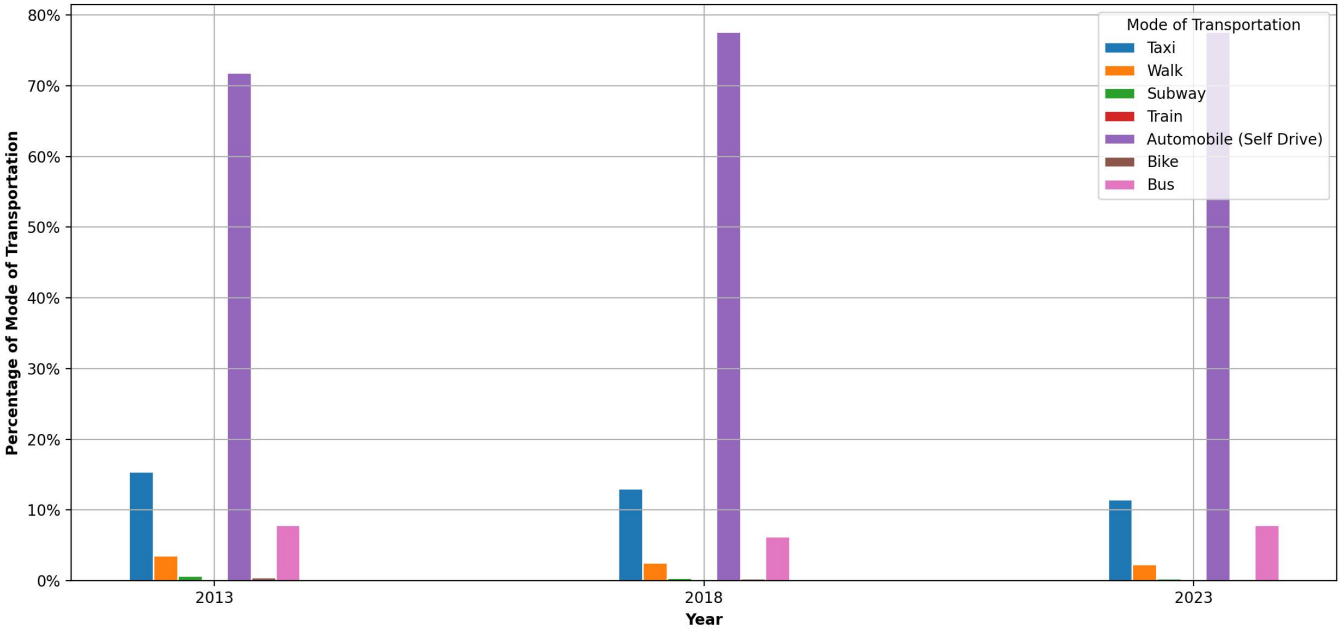

Figure 5.15: Mode of Transportation Distribution of the Trips Produced by an Industrial Zone

The primary purpose of the trips that this zone produced was to get back home. As a result, it's understandable that nearly half of these trips begin between 15 and 18:59, when the majority of businesses and shops in the area begin to close.

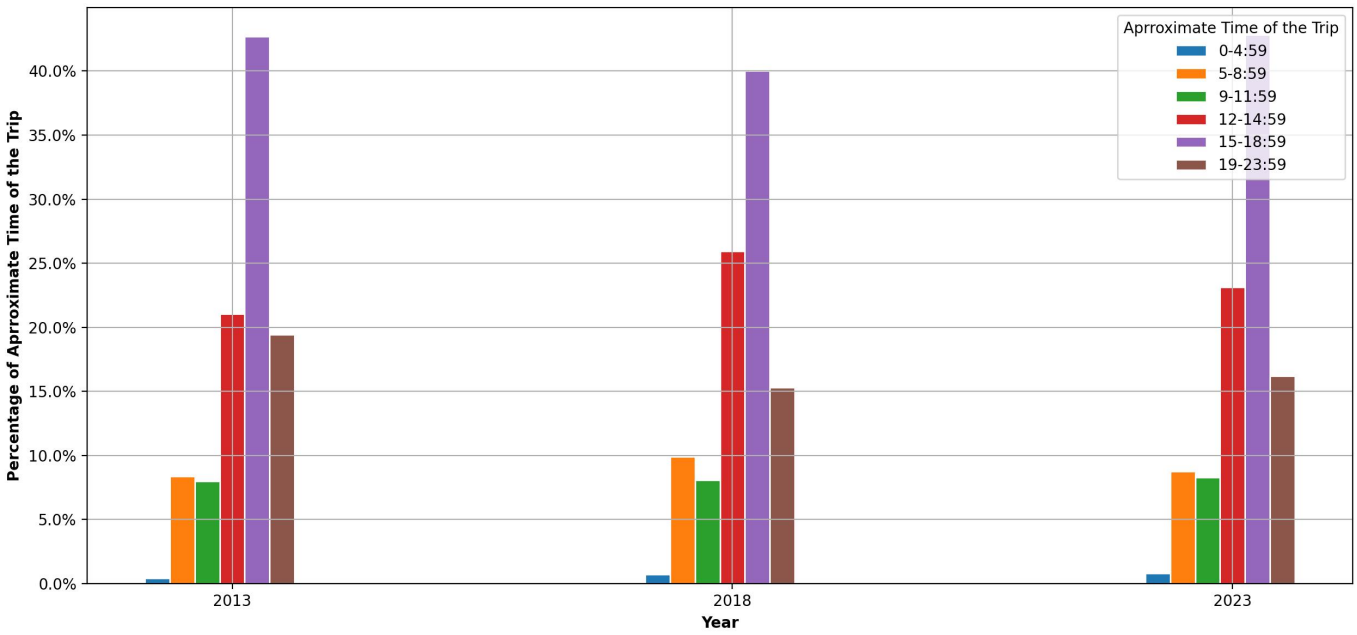

Figure 5.16: Approximate Time Distribution of the Trips Produced by an Industrial Zone

5.2 Trip Characteristics Correlation

Heatmaps were used in the previous chapter to show which trip characteristics had significant correlations. However, the value assigned to a categorical feature in a heatmap cannot be used to show how features are related. It simply shows whether or not there is a strong link between them. This section shows how some of these characteristics are related to one another and how that relationship has evolved.

5.2.1 Number of Automobiles in the Household Based on the Traveler's License Status

Most travelers have one or two vehicles in their household, as shown in figure 5.17, and this pattern has remained consistent for the past 15 years. People without a driver's license, on the other hand, are more likely to have no cars, whereas those

with a driver's license are more likely to have more cars. With each passing year, the number of people in their household with more than three or four cars rises, especially among those who have a driver's license.

5.2.2 Number of People in the Household Based on the Traveler's License Status

People without a driver's license living in more crowded families, as shown in Figure 5.18. This could be due to the fact that the majority of people with driver's licenses are adults who live alone. People without driver's licenses, on the other hand, are typically children from a more prominent family. However, the gap between the two groups has narrowed in recent years, indicating a decrease in the number of people living in a family or household.

5.2.3 Traveler's Age Based on Their License Status

People who have a driver's license are much older than those who do not, as shown in Figure 5.19. There is also a large percentage of people over the age of forty who do not have a driver's license. Figure 5.19, on the other hand, shows that the number of older adults without driver's licenses is steadily decreasing.

5.2.4 Number of People in the Household Based on the Purpose of the Trip

People who travel for school tend to come from more prominent families than those who travel for other reasons, as shown in Figure 5.20. This is also true if you are looking for someone or driving someone back. This is also due to the fact that children typically come from larger families than adults.

5.2.5 Traveler's Age Based on the Purpose of Their Trip

Except for going to school, as shown in Figure 5.21, different trip purposes have had the same proportion of travelers of various ages. Children between the ages of seven and eighteen, on the other hand, have been taking trips to school. This simple pattern has remained consistent in recent years.

5.2.6 Traveler's License Status Based on the Purpose of Their Trip

This pattern was seen in previous correlations. Children do not have driver's licenses, and they are the ones who make the majority of school trips. Therefore, trips to go to school have been taken by those who mostly do not have a license. Other trip purposes are dominated by travelers with licenses. In addition, as shown in figure 5.22, an increasing number of people have obtained driver's licenses over time.

5.2.7 Approximate Time of the Trip Based on Its Purpose

Figure 5.23 shows that trips to school, driving someone, and going to work have all occurred in the early morning. The purpose of the trips taken in the afternoon and noon, on the other hand, has been to return home. Trips for other reasons have rarely occurred in the early morning or even before noon. The majority of trips have taken place either early in the morning or late in the afternoon (also known as rush hour).

Figure 5.17: Number of Automobiles in the Household Based on the Traveler's License Status

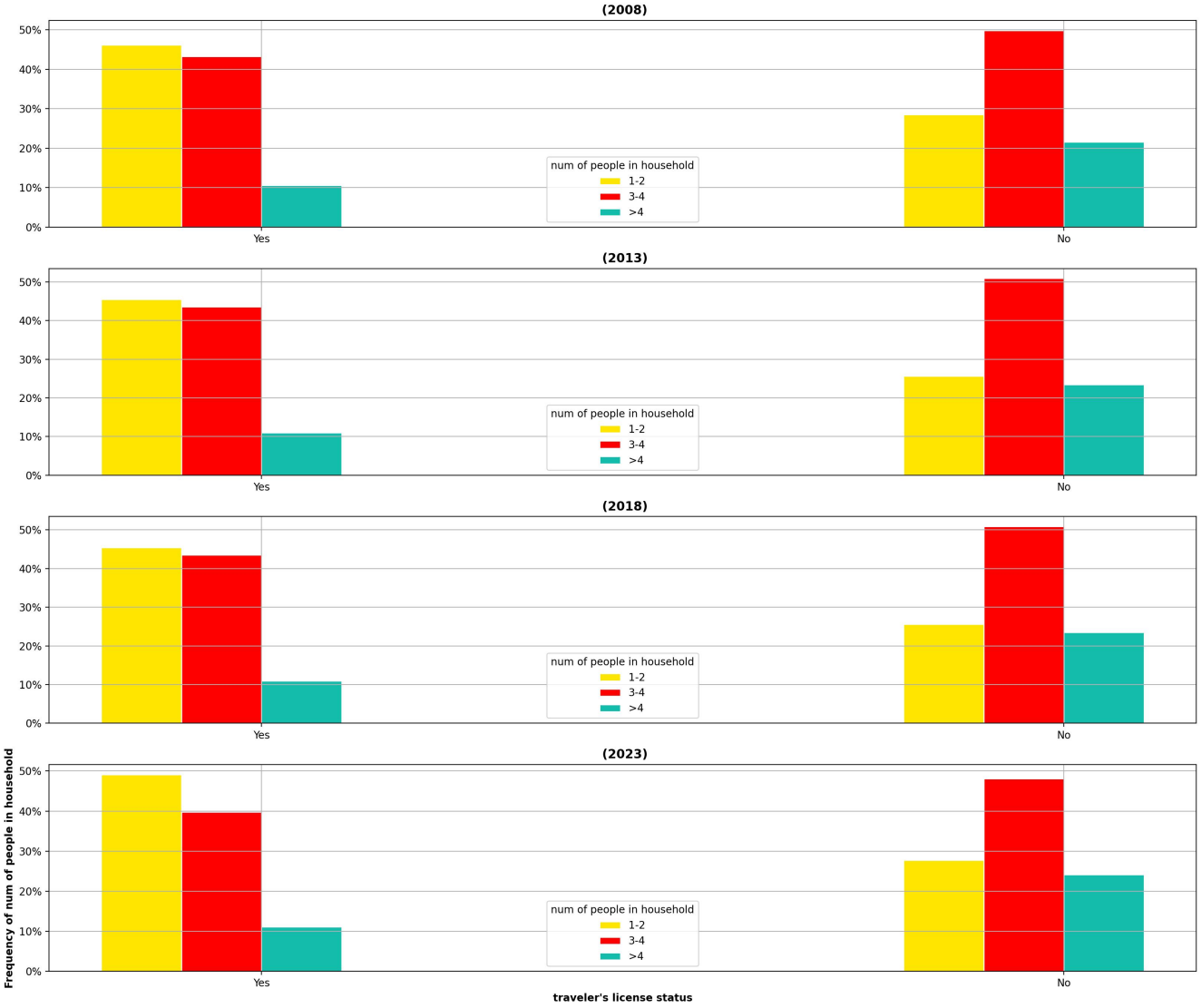

Figure 5.18: Number of People in the Household Based on the Traveler's License Status

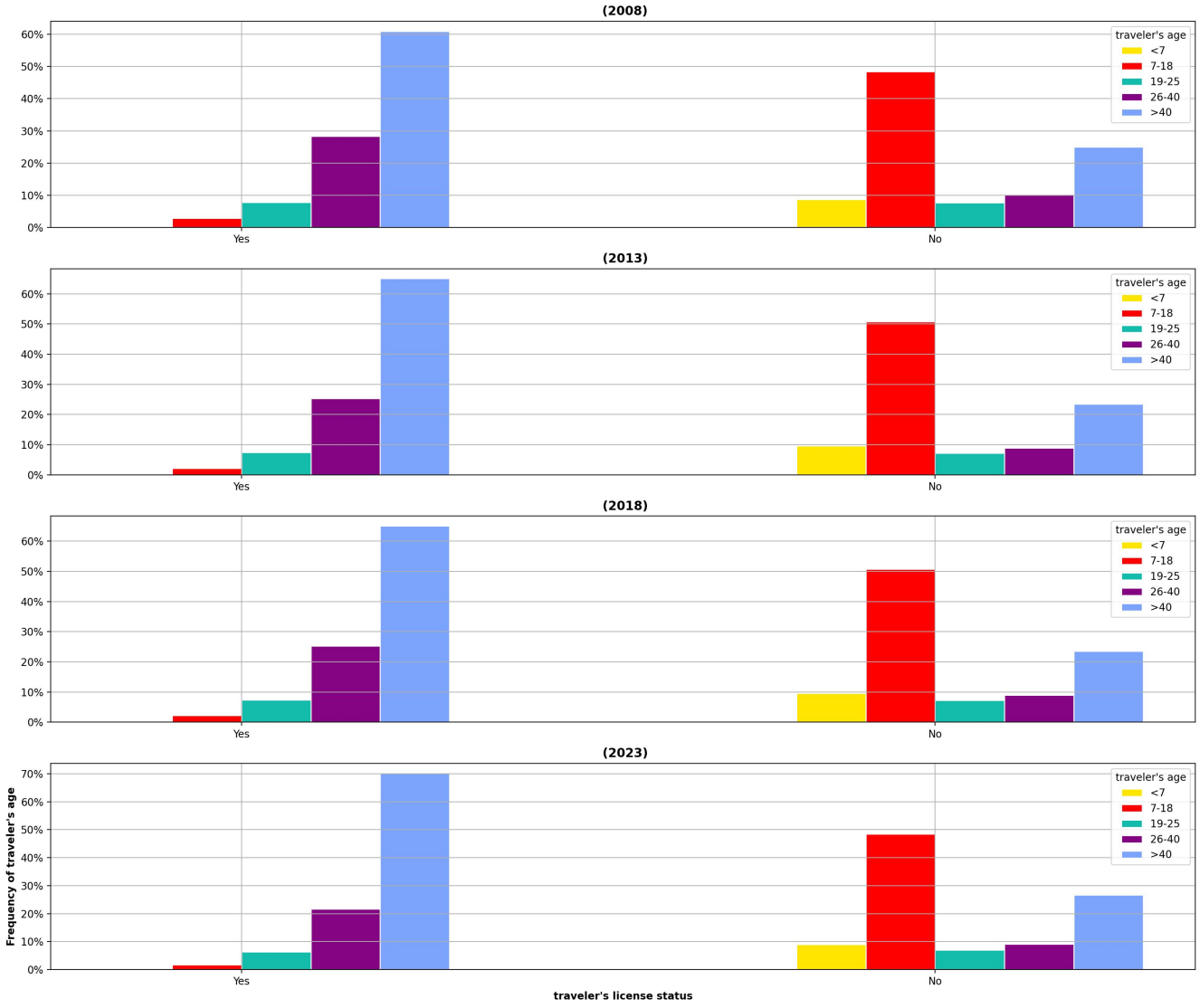

Figure 5.19: Traveler's Age Based on Their License Status

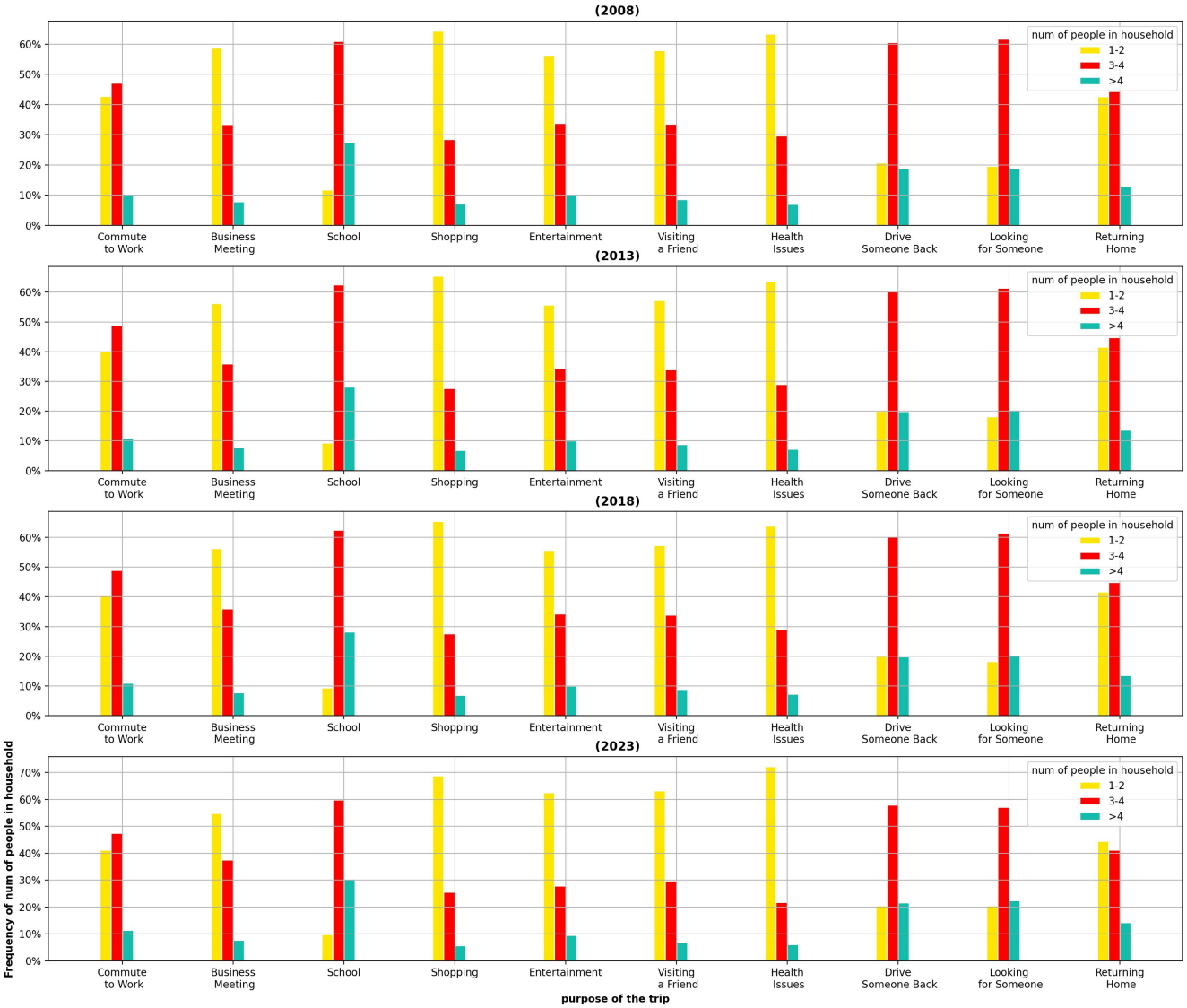

Figure 5.20: Number of People in the Household Based on the Purpose of the Trip

Figure 5.21: Traveler's Age Based on the Purpose of Their Trip

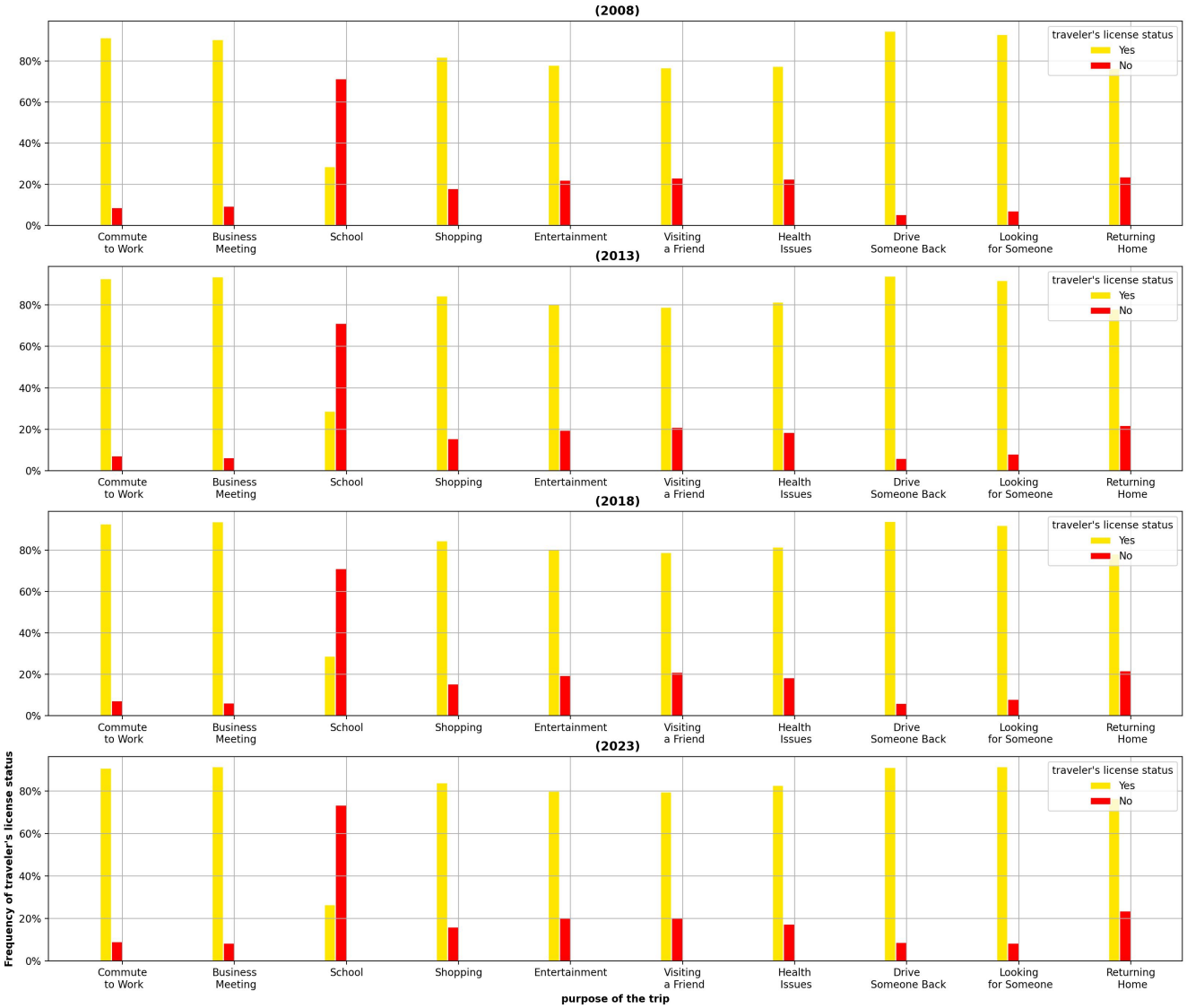

Figure 5.22: Traveler's License Status Based on the Purpose of Their Trip

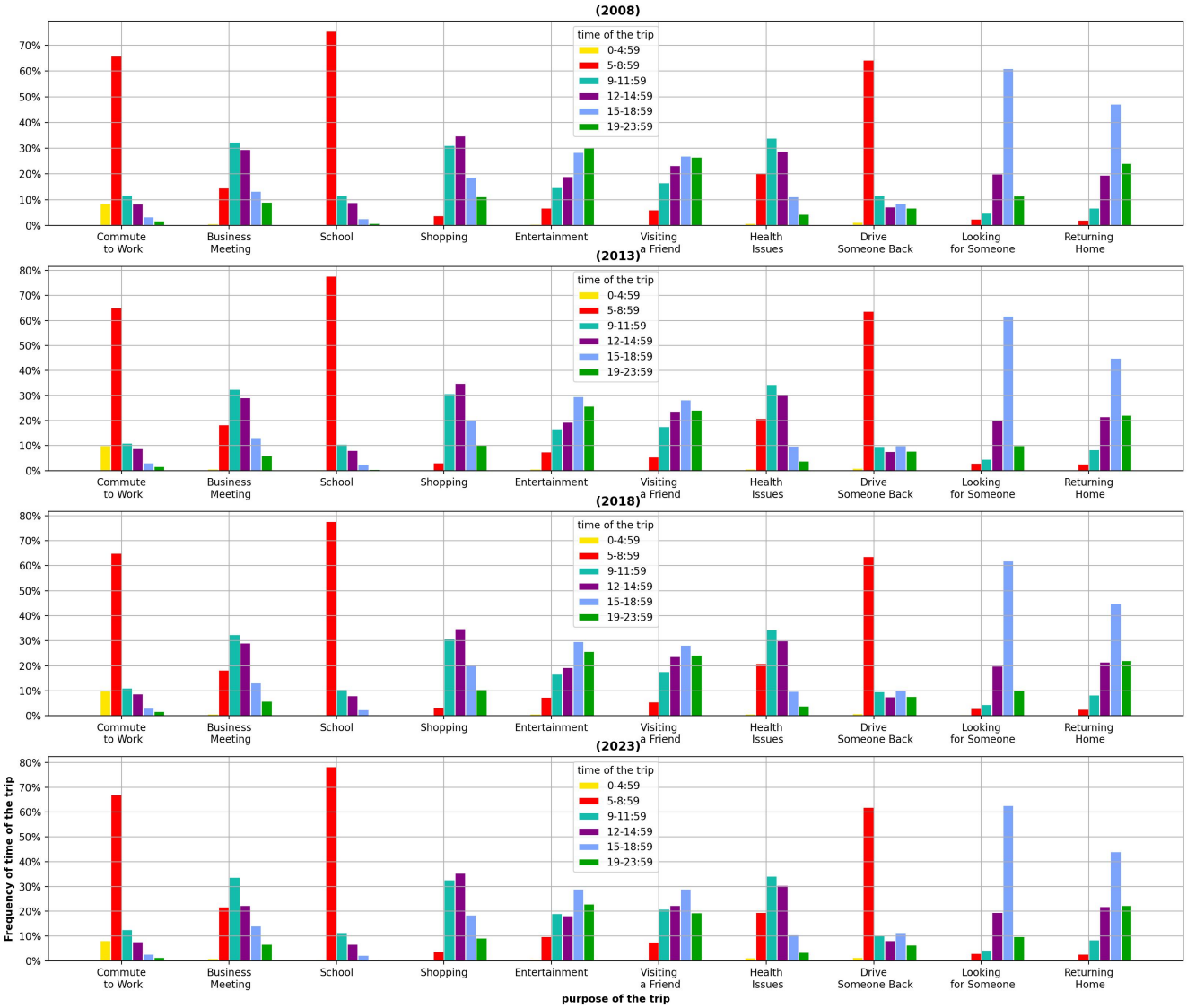

Figure 5.23: Approximate Time of the Trip Based on Its Purpose

Chapter 6

Conclusion

With the expanding concept of smart cities, doing cutting-edge research in intra-city journeys is more critical than ever. The available data in this sector is comprehensive enough to be a reliable study source. The goal of this particular study was to look into the travel habits of Montreal residents. Some of the most recent versions of the Origin-Destination, Census, and Land Use data were used for this purpose.

The OD survey included detailed information regarding the residents' travels. It includes information about the traveler as well as the journey itself. We employed existing methodologies to examine the dependencies between these qualities in the first step of this research. To graphically portray these dependencies, we presented our findings in the form of heatmaps and charts.

We attempted to present a model in the second phase of this research that could accurately predict the "trip destination" and "mode of transportation" given other trip characteristics. We chose the most relevant and clean features in this step, such as the traveler's age, the trip's origin, and so on. We then used min-max normalization to normalize the data before feeding it to various machine learning models. We compared the models' accuracy and found that a multi-layer neural network with ReLU and softmax activation functions produced the best results. The neural network was able to equally consider all the relevant characteristics to finalize its decision while predicting the mode of transportation. As a result, it is a reliable model to use when working with sequential trip records. Hybrid models are thought to produce even higher accuracies for time-series data. These types of data will be studied in the future.

The study's next step was the census data and the OD survey. In this step, we attempted to break the entire probability distribution of the trip characteristics into their individual prior and posterior probabilities using a well-known parametric

model called Bayesian networks. From 2003 to 2018, we used the correlation map of the pairs of characteristics to create the network. The task of generating trip patterns for 2023 was reduced to a linear regression task using this network. In other words, we used a Bayesian network to generate a miniature version of the OD2023, fitting a linear model of the pattern change in the census and OD data, calculating the predicted prior and posterior probabilities, and fitting a linear model of the pattern change in the census and OD data.

To address the issue of the aggregated level of study, we recommend doing future studies in the individual trip pattern, using the obtained OD 2023 to build a model for an activity-based transportation model. To build this model, we will need new data collection methods that are not reliant on five-year surveys because, for example, situations like the Covid-19 pandemic and its impact on transportation are not captured in these surveys. We need more reliable and accurate methods for collecting real-time travel data, particularly in public transportation systems that already have adequate infrastructure, such as smart cards. By implementing this study, we hope to demonstrate the importance of AI and data processing for the analysis of next-generation smart cities. To improve and extend the data sets, we must modernize outdated data collection methods by investing in automated and centralized data collection methods and IoT across the city and country.

The model can help urban and transportation experts predict future OD surveys based on a neighborhood's historical and socioeconomic attributes. It can produce a higher resolution prediction by adding more accurate time-series data. If sufficient disaggregated data is provided, the model can also be used as a starting point for creating a model that predicts the entire trip chain for a person over a given time based on their socioeconomic characteristics. The predicted result can be fed into a simulation engine to create an accurate scenario of a city's occupants' transportation behavior for the future.

This thesis has been accepted as a primary presenter in the biennial National Travel Monitoring Exposition and Conference (NatMEC) 2022. The goal of the conference is to increase the efficiency and effectiveness of multimodal traffic monitoring programs covering motorized, bicyclist, and pedestrian movements to enhance data-driven decisions in areas of performance management, planning and design, asset management, safety, and program administration.

Appendix A

Appendix

This appendix presents the same information in Chapter 5 for two other zones. One zone is a residential zone with code 0.0002, and the other is a green area with zone .0229. Characteristics of these two zones have been relatively consistent throughout the years.

A.1 Residential (With Code 0.0002)

This zone is almost purely a residential zone in the center of Montreal. Figure A.1 shows the utility of the buildings in this zone. As you can see, more than 70 percent of the area of the buildings in the zone is covered by residential buildings.

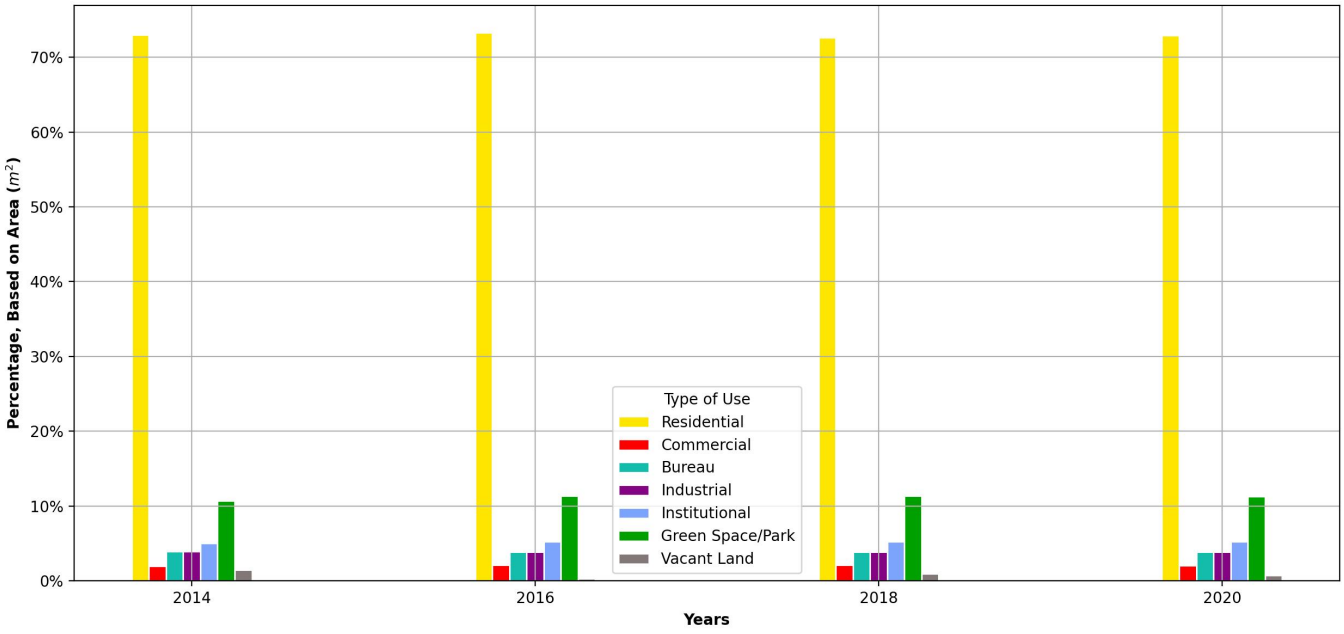

Figure A.1: Area Distribution of Different Types of Places in a Residential Zone (Code 0.0002)

Figure A.2: Land Use characteristics of a Residential Zone (Code 0.0002) Throughout Recent Years

A.2 Green Space (With Code 0.0229)

Based on figure A.3, this zone is almost entirely a green space area. However, some institutions around the region help manage the parks.

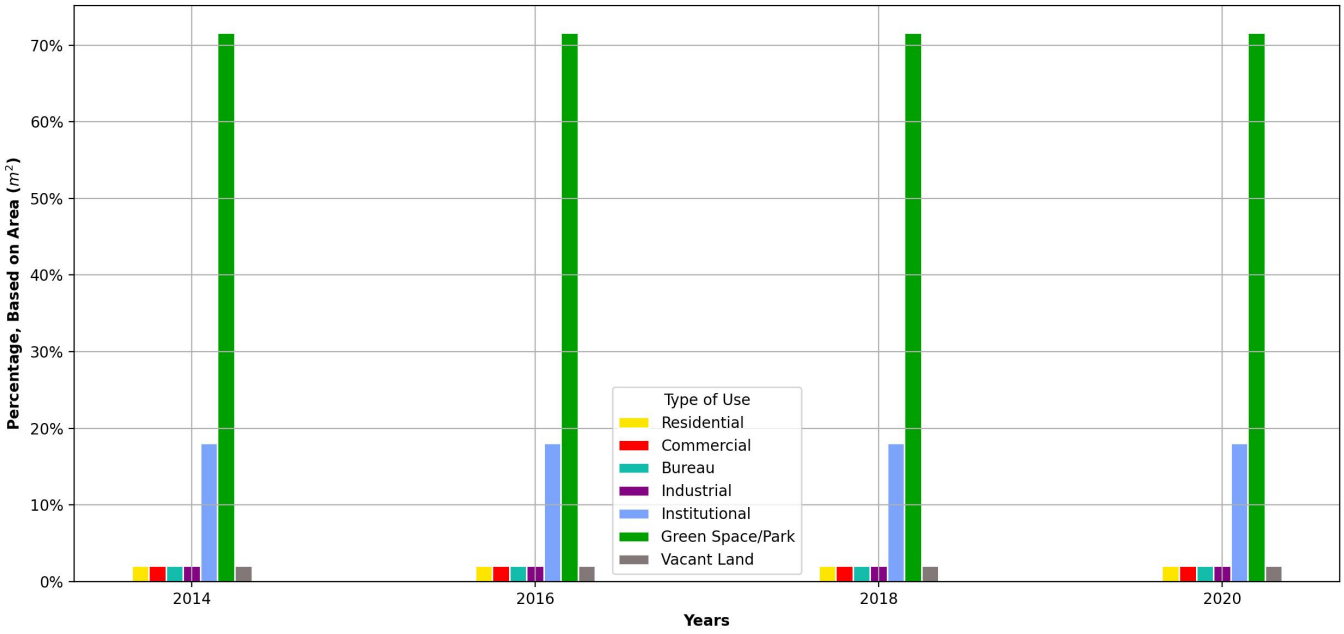

Figure A.3: Area Distribution of Different Types of Places in a Green Space Zone (Code 0.0229)

Figure A.4: Land Use characteristics of a Green Area (Code 0.0229) Throughout Recent Years

A.3 Trips Attracted by The Zones

Residential Area

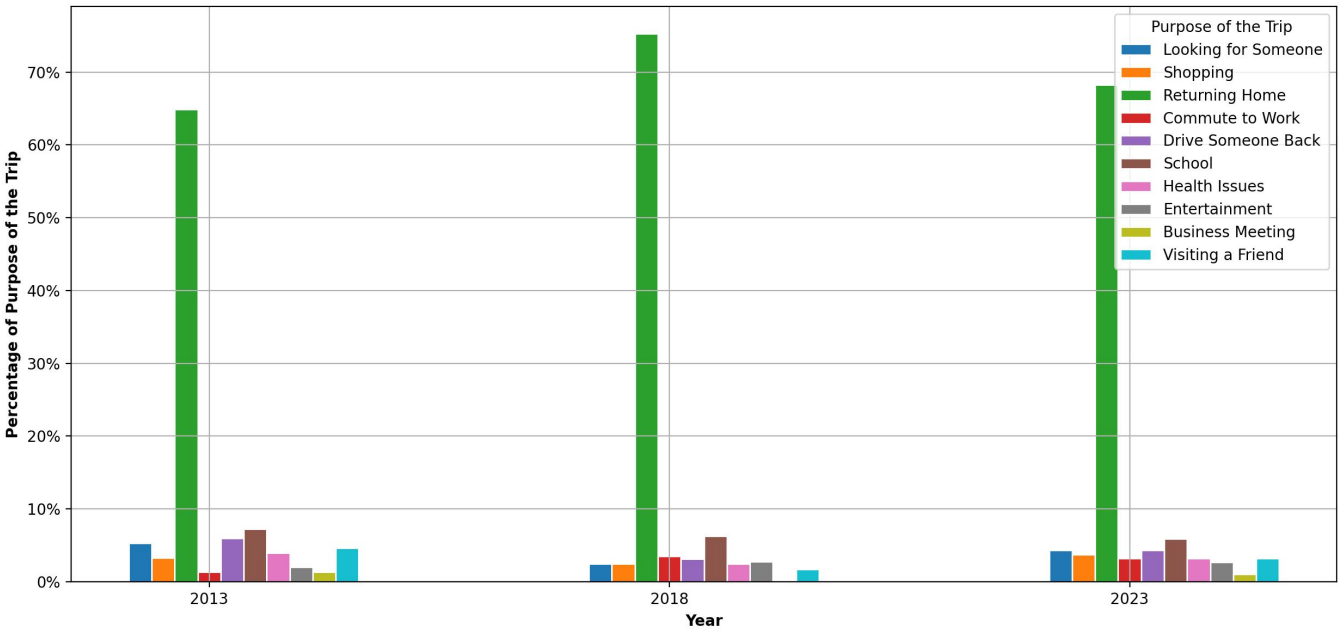

Figure A.5: Purpose Distribution of the Trips Attracted by a Residential Zone

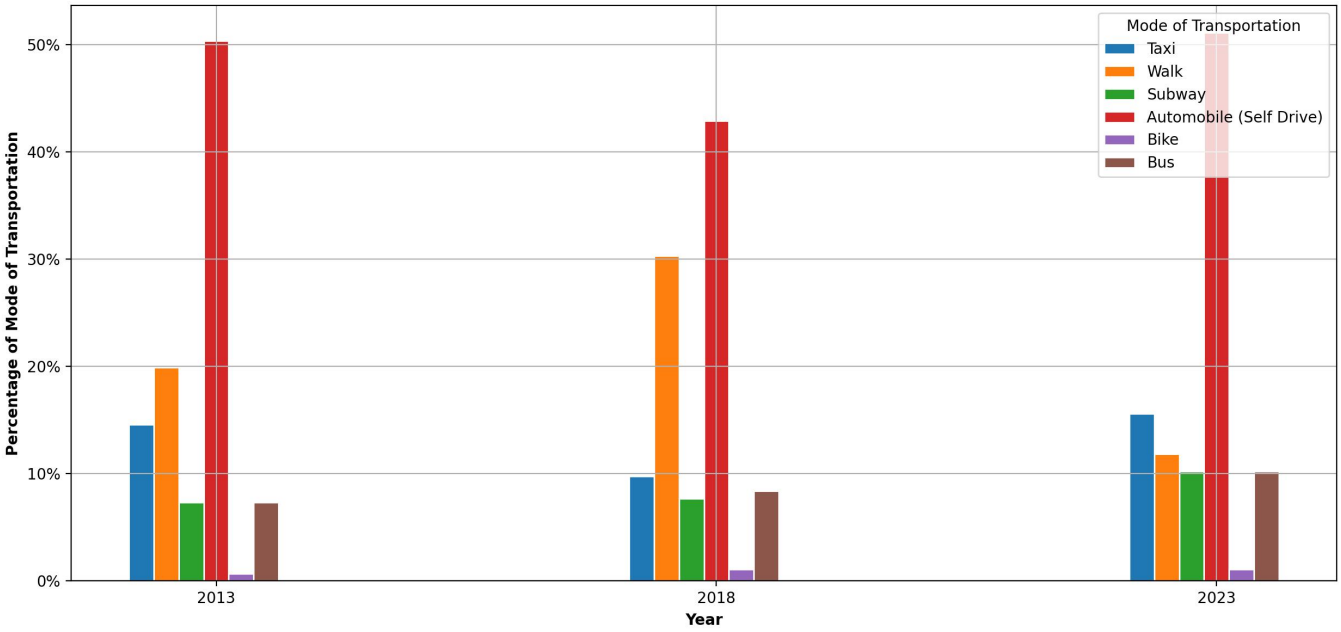

Figure A.6: Mode of Transportation Distribution of the Trips Attracted by a Residential Zone

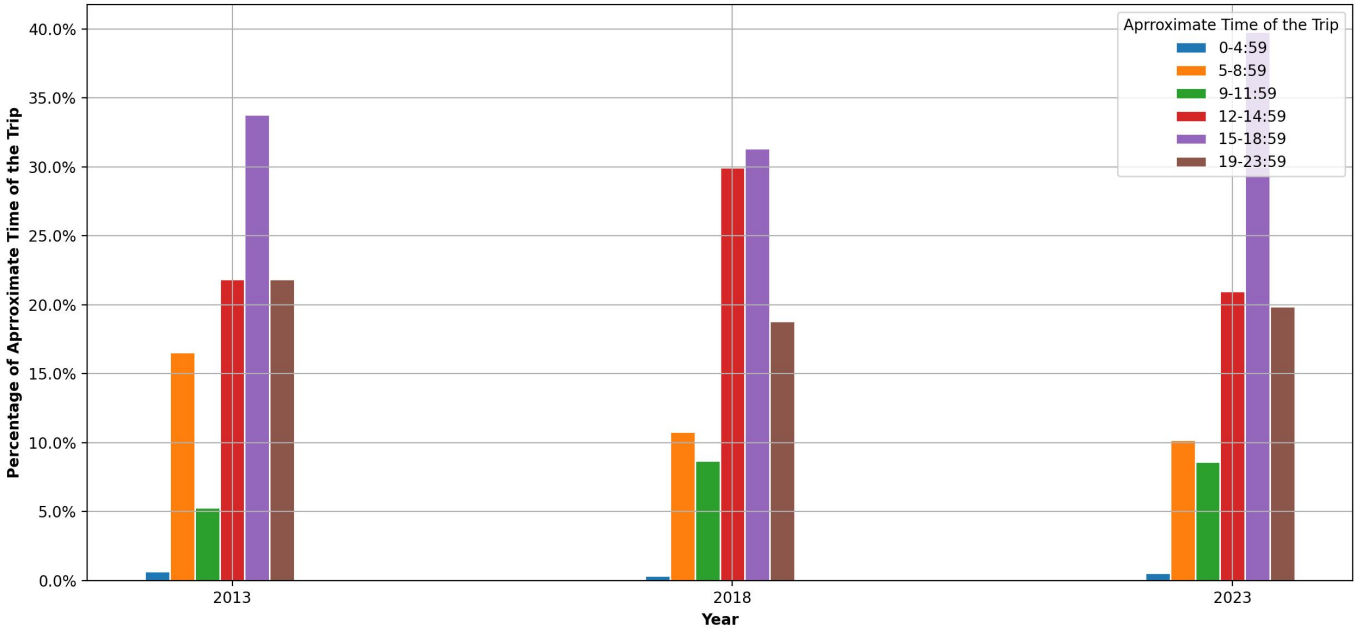

Figure A.7: Approximate Time Distribution of the Trips Attracted by a Residential Zone

Green Space Area

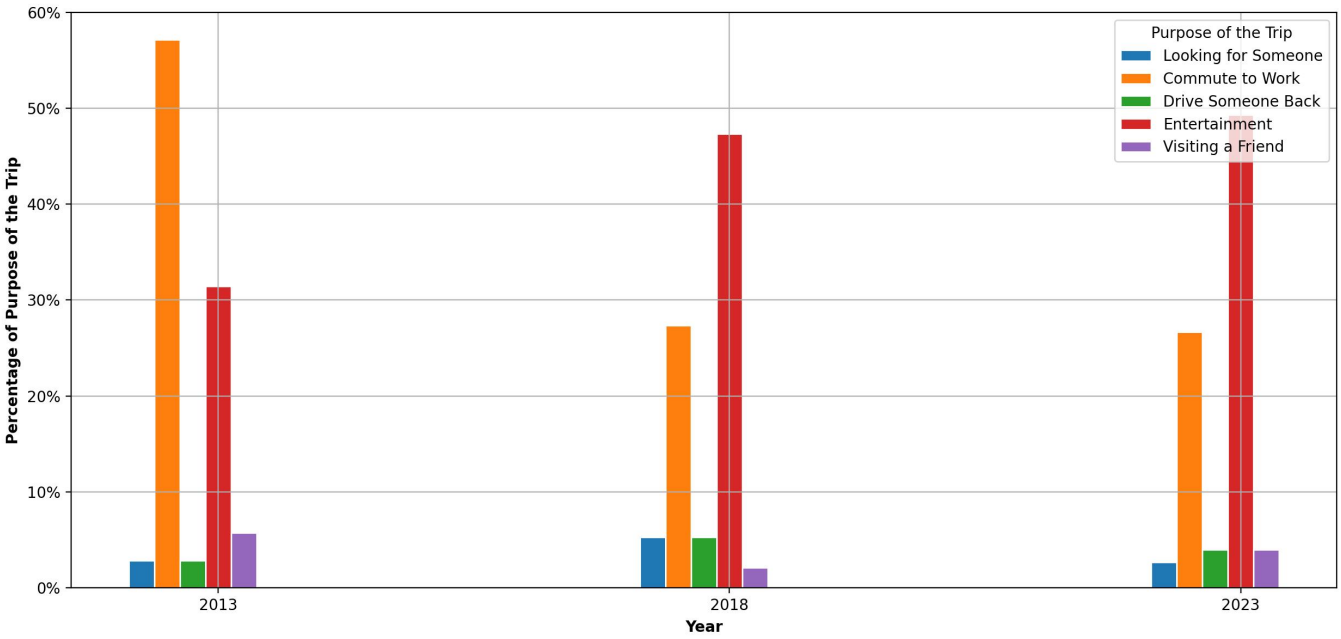

Figure A.8: Purpose Distribution of the Trips Attracted by a Green Space Zone

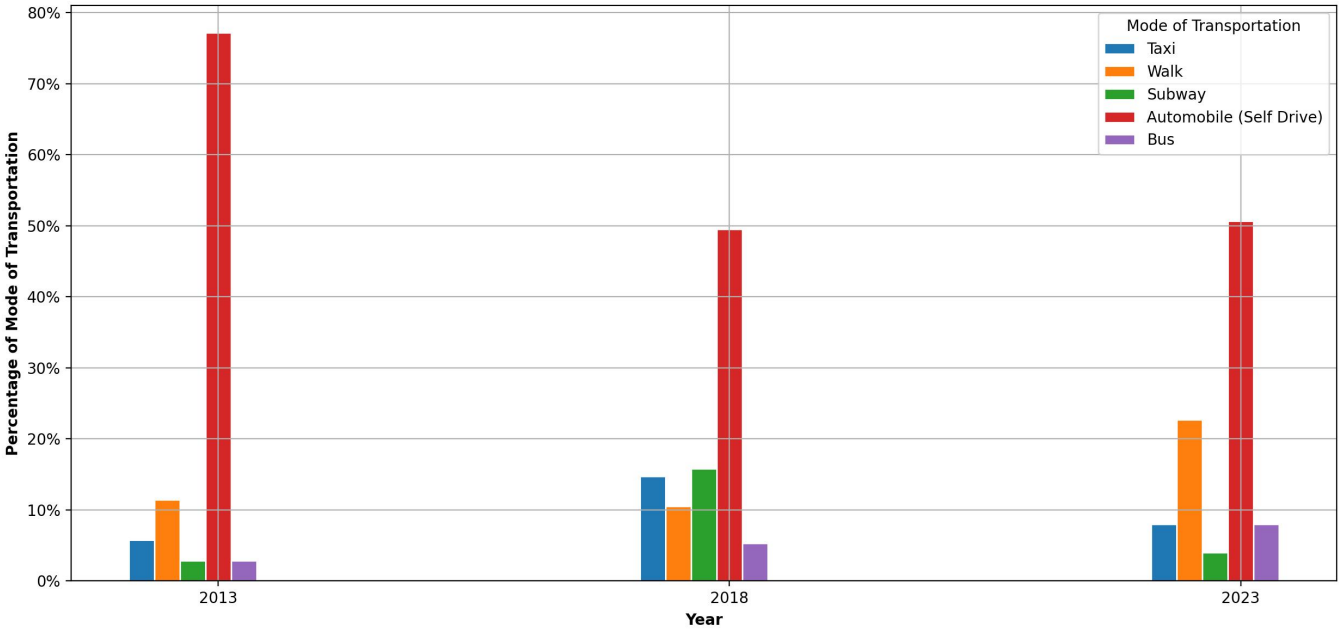

Figure A.9: Mode of Transportation Distribution of the Trips Attracted by a Green Space Zone

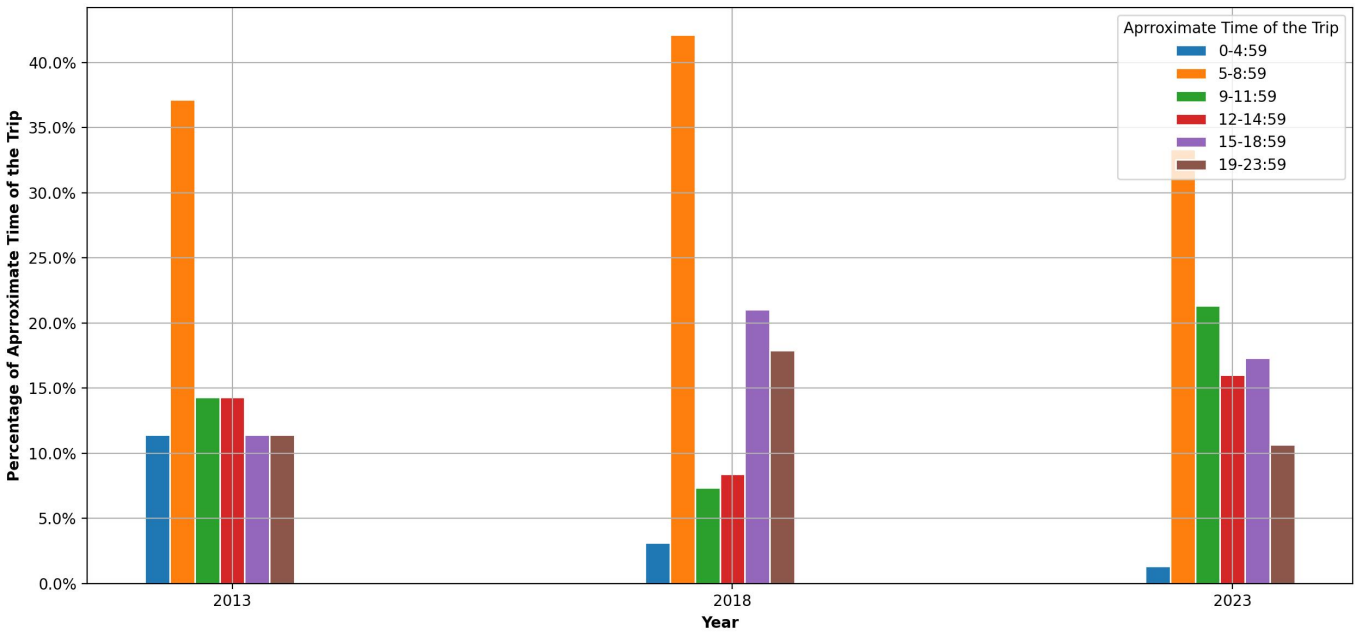

Figure A.10: Approximate Time Distribution of the Trips Attracted by a Green Space Zone

A.4 Trips Produced by The Zones

Residential Area

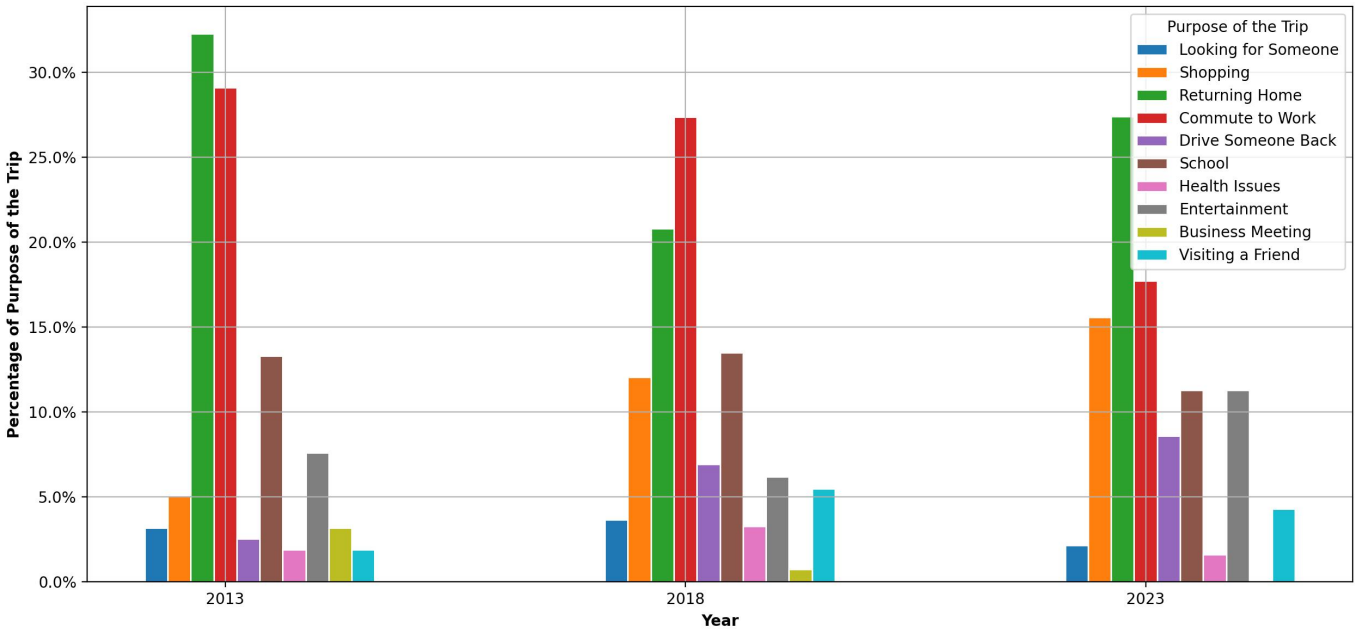

Figure A.11: Purpose Distribution of the Trips Produced by a Residential Zone

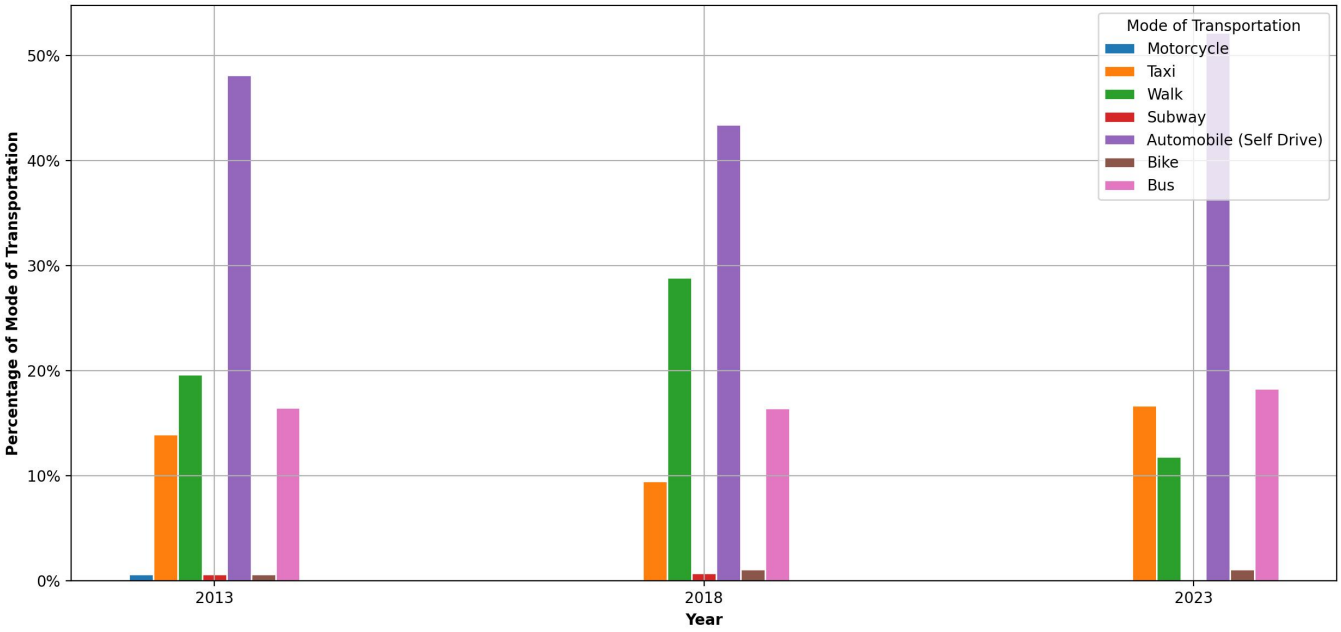

Figure A.12: Mode of Transportation Distribution of the Trips Produced by a Residential Zone

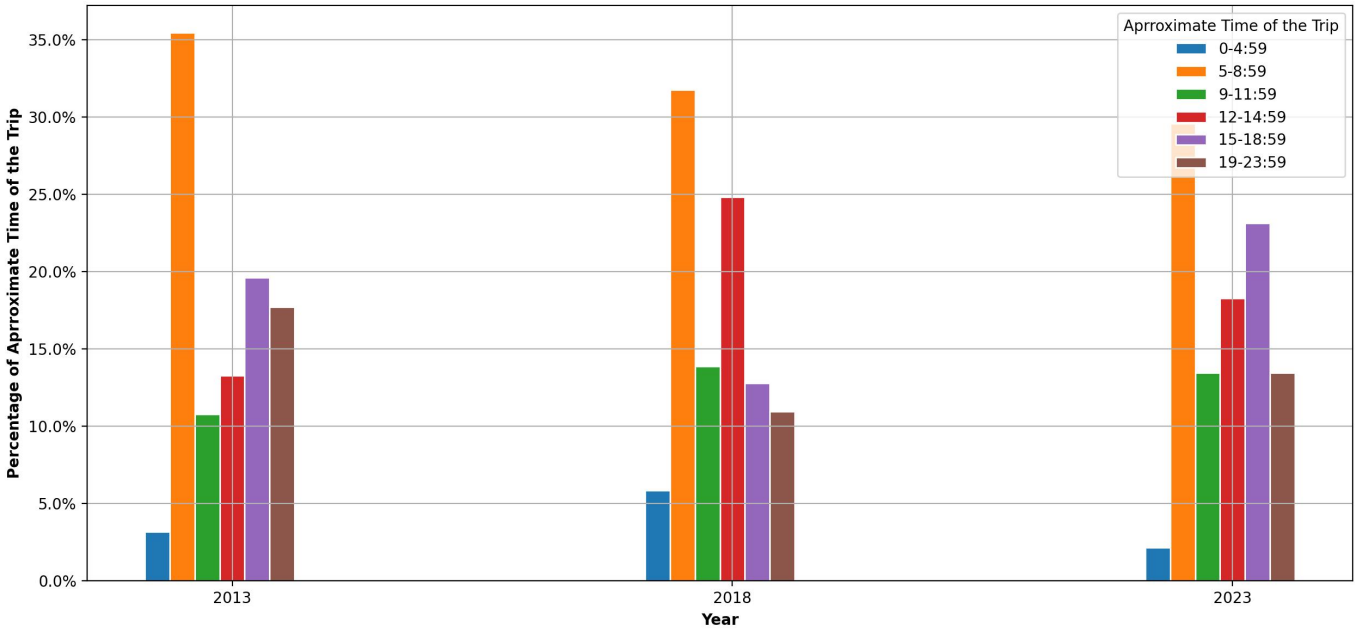

Figure A.13: Approximate Time Distribution of the Trips Produced by a Residential Zone

Green Space Area

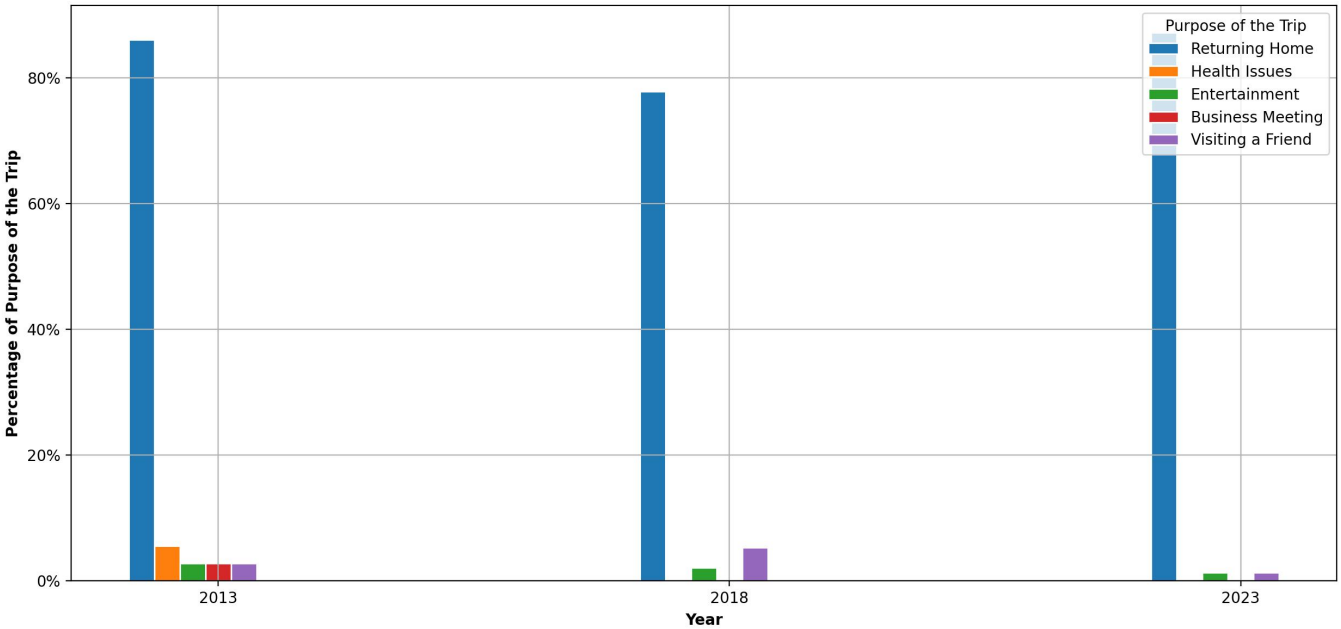

Figure A.14: Purpose Distribution of the Trips Produced by a Green Space Zone

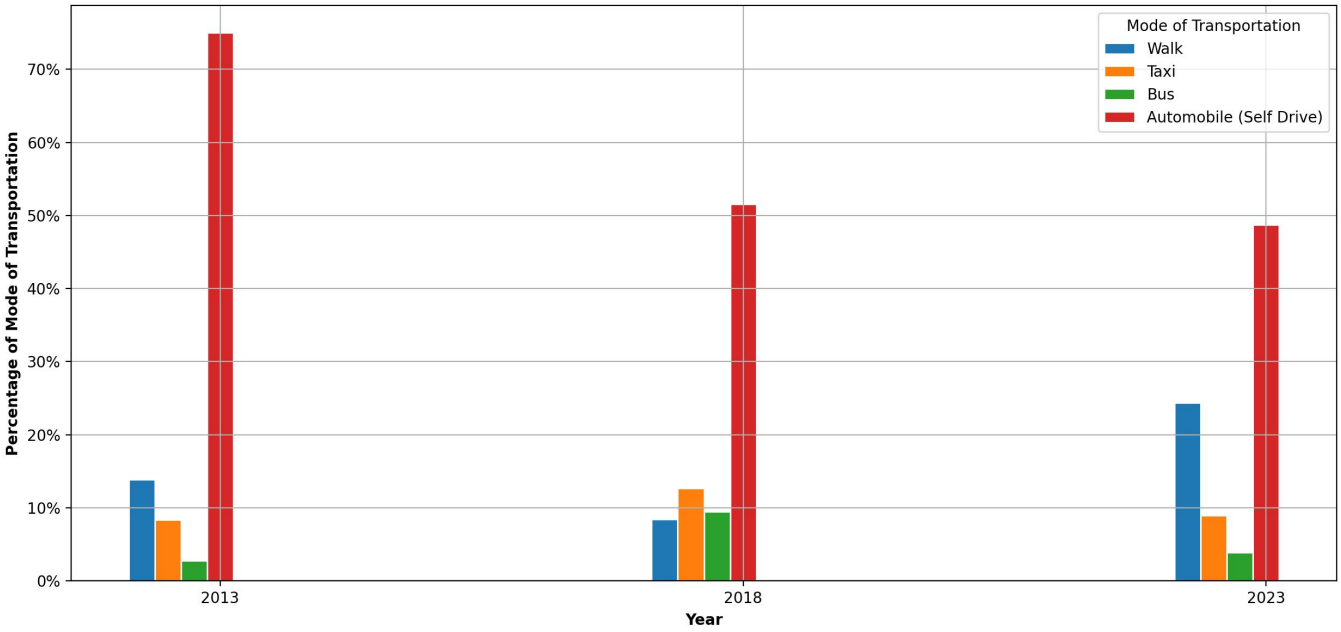

Figure A.15: Mode of Transportation Distribution of the Trips Produced by a Green Space Zone

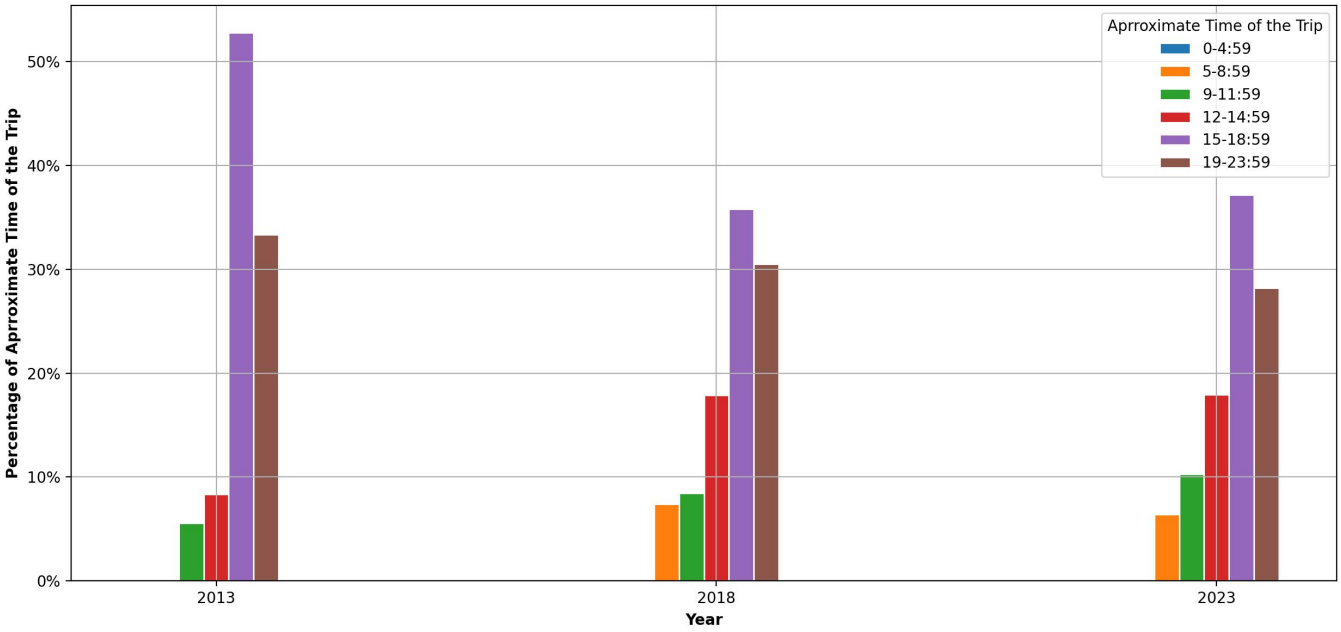

Figure A.16: Approximate Time Distribution of the Trips Produced by a Green Space Zone

List of References

[1] L'Agence Metropilitaine de Transport (AMT). 2008 origin-destination survey for the montreal region, 2008.

[2] L'Agence Metropilitaine de Transport (AMT). 2013 origin-destination survey for the montreal region, 2013.

[3] L'Agence Metropilitaine de Transport (AMT). 2018 origin-destination survey for the montreal region, 2018.

[4] Communauté métropolitaine de Montréal (CMM). Données numériques de la cartographie d'utilisation du sol, 2012.

[5] Communauté métropolitaine de Montréal (CMM). Données numériques de la cartographie d'utilisation du sol, 2014.

[6] Communauté métropolitaine de Montréal (CMM). Données numériques de la cartographie d'utilisation du sol, 2016.

[7] Communauté métropolitaine de Montréal (CMM). Données numériques de la cartographie d'utilisation du sol, 2018.

[8] Communauté métropolitaine de Montréal (CMM). Données numériques de la cartographie d'utilisation du sol, 2020.

[9] Random Forests Understanding. `https://ai-pool.com/a/s/random-forests-understanding`. Accessed: 2022-02-24.

[10] Faruk Cirit. Sürdürülebilir kentiçi ulaşım politikaları ve toplu taşıma sistemlerinin karşılaştırılması. *T.C. Kalkinma Bakanligi*, (2891), 2014.

[11] Josefina Ades, Philippe Apparicio, and Anne-Marie Séguin. Are new patterns of low-income distribution emerging in canadian metropolitan areas? *Canadian Geographer*, 56:339–361, 2012.

[12] Sébastien Breau. Rising inequality in canada: A regional perspective. *Applied Geography*, 61:58–69, 2015. Spatial Inequality.

[13] Scott W. Allard and Sheldon Danziger. Proximity and opportunity: How residence and race affect the employment of welfare recipients. *Housing Policy Debate*, 13(4):675–700, 2002.

[14] Olof Aslund, John Osth, and Yves Zenou. How Important is Access to Jobs? Old Question - Improved Answer. CReAM Discussion Paper Series 0925, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London, October 2009.

[15] Neil Bania, Laura Leete, and Claudia Coulton. Job access, employment and earnings: Outcomes for welfare leavers in a us urban labour market. *Urban Studies*, 45(11):2179–2202, 2008.

[16] Evelyn Blumenberg. Immigrants and transport barriers to employment: The case of southeast asian welfare recipients in california. *Transport Policy*, 15:33–42, 01 2008.

[17] Geneviève Boisjoly and Ahmed El-Geneidy. Daily fluctuations in transit and job availability: A comparative assessment of time-sensitive accessibility measures. *Journal of Transport Geography*, 52:73–81, 04 2016.

[18] Junping Zhang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12:1624–1639, 12 2011.

[19] Hoang Nguyen, Minh Kieu, and Tao Wen. Deep learning methods in transportation domain: A review. *IET Intelligent Transport Systems*, 12, 07 2018.

[20] Gusri Yaldi, Michael Taylor, and Wen Yue. Forecasting origin-destination matrices by using neural network approach: A comparison of testing performance between back propagation, variable learning rate and levenberg-marquardt algorithms. *ATRF 2011 - 34th Australasian Transport Research Forum*, 01 2014.

[21] Jilin Hu, Bin Yang, Chenjuan Guo, Christian S. Jensen, and Hui Xiong. Stochastic origin-destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pages 1417–1428, 2020.

[22] Yuandong Wang, Hongzhi Yin, Hongxu Chen, Tianyu Wo, Jie Xu, and Kai Zheng. Origin-destination matrix prediction via graph convolution: A new perspective of passenger demand modeling. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 1227–1235, New York, NY, USA, 2019. Association for Computing Machinery.

[23] Deborah Shmueli, Ilan Salomon, and Daniel Shefer. Neural network analysis of travel behavior: Evaluating tools for prediction. *Transportation Research Part C: Emerging Technologies*, 4(3):151–166, 1996.

[24] Ayobami E. Adewale and Amnir Hadachi. Neural networks model for travel time prediction based on odtravel time matrix. *CoRR*, abs/2004.04030, 2020.

[25] Pranjali Deshmukh. Travel time prediction using neural networks: A literature review. pages 1–5, 08 2018.

[26] Yang Hai, T. Akiyama, and T. Sasaki. Estimation of time-varying origin-destination flows from traffic counts: A neural network approach. *Mathematical and Computer Modelling*, 27(9):323–334, 1998.

[27] Zhongwei Deng and Minhe Ji. Deriving rules for trip purpose identification from gps travel survey data and land use data: A machine learning approach. 2010.

[28] Long Cheng, Xuewu Chen, Jonas De Vos, Xinjun Lai, and Frank Witlox. Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society*, 14:1–10, 2019.

[29] Lei Gong, Ryo Kanamori, and Toshiyuki Yamamoto. Data selection in machine learning for identifying trip purposes and travel modes from longitudinal gps data collection lasting for seasons. *Travel Behaviour and Society*, 11:131–140, 2018.

[30] B. Aslan and Guenter Zech. New test for the multivariate two-sample problem based on the concept of minimum energy. *Journal of Statistical Computation and Simulation*, 75:109 – 119, 2003.

[31] Hichem Omrani. Predicting travel mode of individuals by machine learning. *Transportation Research Procedia*, 10:840–849, 2015. 18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands.

[32] Subrina Akter, Tanjil Huda, Lutfun Nahar, and Shamima Akter. Travel time prediction using support vector machine(svm) and weighted moving average(wma). *Electrical Engineering Research*, 01 2020.

[33] Tomáš Mikluščák, Michal Gregor, and Aleš Janota. Using neural networks for route and destination prediction in intelligent transport systems. In Jerzy Mikulski, editor, *Telematics in the Transport Environment*, pages 380–387, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[34] Michael Kearney. *Cramér's V*. 12 2017.

[35] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. The correlation ratio as a new similarity measure for multimodal image registration. volume 1496, pages 1115–1124, 10 1998.

[36] Wilhelm Kirch, editor. *Pearson's Correlation Coefficient*, pages 1090–1091. Springer Netherlands, Dordrecht, 2008.

[37] J. Sola and Joaquin Sevilla. Importance of input data normalization for the application of neural networks to complex industrial problems. *Nuclear Science, IEEE Transactions on*, 44:1464 – 1468, 07 1997.

[38] Sumet Mehta, Xiangjun Shen, Jiangping Gou, and Dejiao Niu. A new nearest centroid neighbor classifier based on k local means using harmonic mean distance. *Information*, 9(9), 2018.

[39] Kashvi Taunk, Sanjukta De, Srishti Verma, and Aleena Swetapadma. A brief review of nearest neighbor algorithm for learning and classification. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 1255–1260, 2019.

[40] Ali Haghpanah Jahromi and Mohammad Taheri. A non-parametric mixture of gaussian naive bayes classifiers based on local independent features. In *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, pages 209–212, 2017.

[41] Bahzad Jijo and Adnan Mohsin Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2:20–28, 01 2021.

[42] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[43] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 144–152, New York, NY, USA, 1992. Association for Computing Machinery.

[44] Nello Cristianini and Elisa Ricci. *Support Vector Machines*, pages 928–932. Springer US, Boston, MA, 2008.

[45] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[46] Marius-Constantin Popescu, Valentina Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8, 07 2009.

[47] Abhishek Panigrahi, Abhishek Shetty, and Navin Goyal. Effect of activation functions on the training of overparametrized neural nets, 2020.

[48] Marco Antônio Pinheiro de Cristo, Pável Pereira Calado, Maria de Lourdes da Silveira, Ilmério Silva, Richard Muntz, and Berthier Ribeiro-Neto. Bayesian belief networks for ir. *International Journal of Approximate Reasoning*, 34(2):163–179, 2003. Soft Computing Applications to Intelligent Information Retrieval on the Internet.