CHALMERS
UNIVERSITY OF TECHNOLOGY

Exploring Usage Patterns and Determinants of Shared E-scooters Using Data-driven Methods

A Study of Gothenburg, Sweden

Master's thesis in Infrastructure and Environmental Enginering

HANNES LILLIEBLAD

Department of Architecture and Civil Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY

Master's thesis ACEX30
Gothenburg, Sweden 2023

MASTER'S THESIS ACEX30

Exploring Usage Patterns and Determinants of Shared E-scooters Using Data-driven Methods

A Study of Gothenburg, Sweden

HANNES LILLIEBLAD

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Urban Mobility Systems Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Exploring Usage Patterns and Determinants of Shared E-scooters Using Data-driven Methods
A Study of Gothenburg, Sweden

HANNES LILLIEBLAD

Supervisor: Assistant Professor Kun Gao, Department of Architecture and Civil Engineering
Examiner: Associate Professor Jorge Gil, Department of Architecture and Civil Engineering

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Urban Mobility Systems Group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: An image of the spatial demand in Gothenburg

Department of Architecture and Civil Engineering
Gothenburg, Sweden 2023

Exploring Usage Patterns and Determinants of Shared E-scooters Using Data-driven Methods
A Study of Gothenburg, Sweden
Hannes Lillieblad
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

Understanding the usage patterns of shared e-scooter is crucial for decision-makers to implement effective policies and for e-scooter operators to optimize their services. This study aims to investigate the spatial and temporal usage patterns of e-scooters in Gothenburg, Sweden, and to analyze the impact of various factors, such as socio-economic and built environment factors, on e-scooter demand. Analysis of the data is done by employing data-driven methods and machine learning models, including hierarchical clustering, XGBoost, and random forest. The analysis consisted of two major parts: the usage pattern analysis, which primarily used e-scooter transaction data to analyze demand and trip patterns, and the influencing factors analysis, which analyzed several factors to determine their influence on e-scooter demand. The findings reveal distinct demand patterns across different areas and time slots in Gothenburg, with the highest average trip duration occurring during nighttime on weekends. The analysis considers 35 variables, categorized into temporal and spatial factors. Bus stops emerge as the most significant spatial determinant, followed by public buildings, commercial buildings, residential roads, other roads, and health POIs. Overall, built environment factors seem to impact e-scooter demand substantially more than socio-economic factors. Among the temporal determinants, it is the month that has the most significant influence, followed by minimum temperature, average precipitation, and weekday. A partial dependency analysis further explains the relationship between each determinant and e-scooter demand. The findings of this study provide valuable insights for e-scooter operators and policymakers to better understand and cater to the needs of e-scooter users in urban environments.

Keywords: E-scooter demand patterns; micromobility; machine learning; random forest; key determinants.

Acknowledgements

First, I would like to thank my examiner, Jorge Gil, supervisor Kun Gao, and PhD student Omkar Parishwad for their support and valuable contributions throughout the creation of this thesis. I would also like to express my gratitude towards Lukas Sandman for his aid and support during the thesis's programming and machine learning phases. Lastly, I would like to thank the entire Department of Architecture and Civil Engineering at Chalmers University of Technology and all who made this report possible.

Contents

List of Figures

List of Tables

1
Introduction

Living and traveling in a modern medium to large city often presents daily challenges. As urban areas continue to develop and expand, space for large vehicles such as cars and trucks becomes limited, making it difficult for them to compete in the urban transportation system. This necessitates the exploration of alternative modes of transport to replace cars. The city of Gothenburg, Sweden, is no exception to this trend. With the city's growth and increasing density, the city government implemented a traffic strategy in 2014 that set up goals to be achieved by 2035. This strategy aims to expand and improve public transport throughout the city, making cycling, walking, and public transport more viable and attractive options for residents and visitors alike [Göteborgs Stad Trafikkontoret, 2014]. To incentivize the usage of alternative transportation modes in the city of Gothenburg, in collaboration with NextBike, installed several bike-sharing stations [Göteborgs Stad, ndb]. These stations offer GPS-tracked bikes with seven gears but lack electric power assistance, which could deter usage in a hilly city like Gothenburg.

In recent years, e-scooters have emerged as a popular mode of transportation worldwide, including Gothenburg[Badia and Jenelius, 2021]. These electric-powered vehicles, available in large parts of the inner city, can be unlocked through each company's mobile app, driven to the desired location, and locked again. Further, they are driven with electric motors, which provide comfort and ease of navigation despite large differences in altitude. This mode of transportation has provided freedom for consumers who are now not bound to a set location, such as bus stops and bike-sharing docks or schedules [Gao et al., 2023b]. This, in turn, is changing urban mobility and has, in the long term, the potential to reduce congestion and pollution [Mehzabin Tuli et al., 2021, Gao et al., 2023a]. However, there is a need to understand the usage patterns, demand factors, and intermodality of e-scooters and other inner-city transport modes. Studies have been made concerning this subject. However, each region and city is different, and this study aims to fill this research gap in the Scandinavian region and Gothenburg specifically. This will be done by analyzing e-scooter transaction data, demographic information, and spatial data to understand usage patterns and what factors influence e-scooter demand in Gothenburg.

1.1 Aim

The thesis aims to assess and quantify the factors influencing shared e-scooter usage in Gothenburg using machine learning. Transaction data from e-scooter companies, spatial data, land use, demographic data, and other relevant data have been analyzed to answer the following research questions:

1. **What are the temporal and spatial usage patterns of e-scooter users in Gothenburg?**

2. **What factors influence the demand for e-scooters in Gothenburg, and to what extent do these factors influence the demand?**

2

Literature review

2.1 General E-scooter Information

The electrically powered stand-up scooters or e-scooters became widely available worldwide and in Gothenburg in the late 2010s and have since become an integral part of many cities' urban mobility system [Mehzabin Tuli et al., 2021]. The fleets of e-scooters in these cities are owned by private companies responsible for the distribution and rental of the vehicles [Badia and Jenelius, 2021]. The vehicles are distributed within a service area, encompassing an entire city or a smaller area. The e-scooters are equipped with a GPS tracking system to track them and ensure customers don't drive out of the service area. If this happens, the vehicle is turned off remotely and has to be returned to the service area by the user to continue the ride. The GPS also helps the customer find a company's vehicles using the phone application provided by the company, which also serves as the service interface for the customer. The vehicles can be unlocked via the phone application, driven anywhere within the permissible area, and locked again via the application. The app then calculates the fare of the ride, which often includes a start cost and an amount for every minute driven by the customer. Some companies also offer a monthly ride pass, allowing users to use the e-scooters freely with no initial cost or minute fare [Badia and Jenelius, 2021]. However, there is a limit to the amount of minutes the user can use the service during this period. The costs for some e-scooter services can be seen in Table 2.1.

There are rules the user and e-scooter companies have to abide by to use the e-scooters, which differ from city to city. In Gothenburg, e-scooters can only be driven on the road or a bicycle path, and they cannot be parked on bicycle paths or in places where they obstruct other road users. Additionally, the maximum speed of the vehicles is regulated in some parts of the city [Göteborgs Stad, nda].

Table 2.1: E-scooter ride fares

Company	**Initial cost [SEK]**	**Minute cost [SEK]**	**Monthly pass [SEK]**
VOI	10	2.50	299 [300 min limit]
Bolt	5	3.50	219 [300 min limit]
Tier	12	2.99	199 [150 min limit]
Ryde	0	2.50	299 [300 min limit]

2.2 E-scooter Studies

Since the aim of the thesis is to investigate the usage patterns and influencing factors of e-scooter usage, it is helpful to incorporate similar studies that have been made on the subject. The reports were found using Scopus and Google Scholar, where keywords such as "e-scooter", "temporal analysis", "spatial analysis", and "influencing factors" were used to find appropriate sources. These studies can influence what patterns to cover, analyze the feasibility of the results, and what factors can be included in the study.
Several studies have been conducted on the usage patterns and influencing factors of dockless e-scooters. The studies cover many cities spread throughout America, Europe, and Asia. In Section 2.2.1, where spatio-temporal usage patterns will be covered, studies using data from Berlin, Germany [Heumann et al., 2021], Nashville, Tennessee [Shah et al., 2023] and a comparative study analyzing Austin, Texas and Minneapolis, Minnesota [Bai and Jiao, 2020] will be reviewed. In the Chapter 2.2.2, where studies regarding the influencing factors of e-scooter usage are reviewed, the study using data from Berlin [Heumann et al., 2021] and the comparative study using data from Austin and Minneapolis [Bai and Jiao, 2020] as well as a study using data from Los Angeles [Yang et al., 2022] will be covered.

2.2.1 Spatio-temporal Usage Patterns

2.2.1.1 Berlin

In Berlin, the weekday demand has a lesser spike during hours 8 and 9 with a drop afterward [Heumann et al., 2021]. The demand then rises throughout the day with a maximum at 18 o'clock. Meanwhile, the morning spikes are absent on weekends, and more trips are made during the day, with a maximum at 15. After this hour a similar amount of trips are made as during the weekdays, however, after midnight the amount of trips is double that on the weekdays. Further, looking at the trip distances in Berlin, they stay fairly even through the day, with a slight drop at 4 and 5 during weekdays and at 7 during weekends.

2.2.1.2 Nashville, TN

In Nashville, Tennessee, a similar usage pattern can be seen as in Berlin. The temporal travel patterns during weekdays and weekends in Nashville are similar to those in Berlin during weekends. There is no noticeable peak during the weekday morning commute, which was present in the Berlin weekday data. In Nashville a steady rise in demand from 6 in the morning where there is a minimum until 13-14 where a maximum in demand can be seen. This pattern is present throughout the week. However, the maximum demand is greater during the weekends. This pattern in Nashville indicates that e-scooters are not used to commute to work in Nashville but are used more for recreational activities. The Nashville study also included a yearly demand analysis. This displays an increase in demand during April, May, and June, later decreasing through July. The article also included a comprehensive spatio-temporal analysis, which used a more sophisticated study that involved machine learning. The procedure involves a principal component analysis

(PCA) followed by k-mean clustering to group similar ride patterns based on their usage type. The PCA allows the user to reduce the amount of variables in a dataset by combining them into a smaller amount of variables while still retaining the information of the larger amount of variables [Jaadi, nd]. This is useful when simplifying the dataset for visualization and making the following machine learning processes faster and more efficient. However, some accuracy is lost when doing this analysis. The following process, the k-mean clustering, is an algorithm that groups data points based on similarities within the data [Shah et al., 2023].

Five distinct usage clusters were distinguished: morning work/school trips, daytime short errand trips, social trips, nighttime entertainment district trips, and utilitarian trips, and were based on temporal and spatial similarities. As such, one can discern how the built environment influences travel patterns differently. For instance, morning work/school trips were mostly from areas with low employment and parking density to business district areas and contributed to about 7% of the total trips. Daytime short errands were within areas with a high population density and a medium-high entropy (mixed land use). These trips are the most common, with 30% of the total trips taken. The social trips were mainly during the weekend evenings and took place within downtown areas with high entropy values and were about 26% of the total trips. Nighttime entertainment district trips were mainly made during weekends at night, with both the origin and end of the trips near bars and clubs. These trips make up about 16% of the total trips. Lastly, the utilitarian trips were mainly made during the daytime on weekdays, and the land use of the origin and destination were often different in type. These trips make up the remaining 22% of the trips made.

2.2.1.3 Austin, TX and Minneapolis, MN

Within the comparative study of Austin, Texas, and Minneapolis, Minnesota, the patterns differ quite substantially [Bai and Jiao, 2020]. Minneapolis shows a very even ridership throughout the week, with only a slight increase from Tuesdays to Wednesdays, which persisted throughout the rest of the week. Further, most trips in Minneapolis occur between the afternoon and evening between 12 and 00 and almost none during the morning between 6 and 12 o'clock. Meanwhile, in Austin, the ridership rises substantially through the week, with a clear maximum on Saturdays. This increase is almost entirely influenced by the rise in afternoon trips, which comprise the most significant portion of trips in the city. There is also a more extensive section of morning ridership compared to Minneapolis, with almost none at night. However, this might be due to local regulations restricting trips between midnight and early morning.

2.2.2 Influencing Factors

2.2.2.1 Berlin

The article that used data from Berlin mainly focused on the land use aspect and used five factors: residential, recreational, commercial, public transport, and public area [Heumann et al., 2021]. The trips were then assigned their origin and destination area type, and the total trip share for each land use type was calculated. The authors admit this is not the optimal methodology because the land use may not be associated with the trip. However, there should be some causality due to the size of the dataset. Nevertheless, residential land use is the land use type with the most departures and arrivals at around 40% each, followed by public transportation and public areas at around 20%, commercial land use at 11%, and lastly, recreational at 4%. The study also included the time of the trip in the land use analysis to investigate how the departures and arrivals in the land use types differ over time. This analysis showed that the previously mentioned percentages fluctuated fairly much throughout the day. The trips with an origin in residential areas were largest in the morning, and the arrivals for public transportation were higher in the same period. This indicates that e-scooters function as a part of the daily commute for many citizens since the same pattern is not observed during weekends. Further, the arrivals in residential areas are seen rising throughout the day, which could be the return trips made by the same commuters who left in the morning. Overall, this example indicates that the key determinants of e-scooter usage may change throughout the day and week and do not remain constant.

2.2.2.2 Austin, TX and Minneapolis, MN

The comparative study between Austin and Minneapolis used a negative binomial regression model to investigate the relationship between the influencing factors and e-scooter usage [Bai and Jiao, 2020]. This method is often used when the dependent variables' variance exceeds the mean. In this instance, the daily e-scooter demand variance is greater than the mean, which warranted this method. The influencing factors, including socio-economical and built environment types, can be seen in Table 2.2. Through the negative binomial regression model, the study found that closeness to the city center, access to public transit, and closeness to schools and other public buildings greatly correlated with high e-scooter demand in both cities. However, some differences were also present between the cities. Only in Minneapolis was a large correlation between e-scooter ridership and a young population, as well as the male population. However, the correlation was positive in both areas. Regarding the built environment, there was a large disparity when inspecting the demand in industrial areas, open spaces, parks, and transportation facilities. Austin saw a large positive correlation in areas with this environment, while Minneapolis had a negative correlation.

2.2.2.3 Los Angeles, CA

The last study to examine is the study that used data from Los Angeles, California [Yang et al., 2022]. This study uses a machine learning model called gradient boosting decision tree (GBDT), which can investigate non-linear relationships between independent and dependent variables using a series of decision trees instead of solely linear relationships. Further explanations of machine learning models and what benefits they can bring are elaborated upon in Chapter 2.3. The study also collects and uses data in the shape of census blocks, which determines the project's sample size. This aids in the data collection of data. However, the author states that a low number of these census blocks may result in an insufficient sample size, which may lower the accuracy of the results.

The study used a mixture of socio-economic and built environment variables, as seen in Table 2.2. Through the GBDT method, it was shown that the built environment had the most relative importance in comparison to the social factors. Intersection and road density stood out as the most critical factors, with 27% and 22%. The density of transit, restaurants, and employment followed these factors. The socio-economic variables such as millennial population, black population, and income had a very low impact with less than 1% relative importance each. Overall, this study highlights the importance of the built environment when estimating the demand for e-scooters in contrast to the low impact of the socio-economic factors.

Table 2.2: Observed independent variables

Variable	Berlin [Heumann et al., 2021]	Austin & Minneapolis [Bai and Jiao, 2020]	Los Angeles [Yang et al., 2022]
Socio-economic variables			
Population density		x	x
P. young population		x	x
P. male/female		x	x
P. black population			x
P. white population			x
P. High school education		x	
P. Higher education			x
Income		x	x
No. of vehicles			x
Household size			x
Built environment			
Employment density			x
Distance to city center		x	x
Transit accessibility	x	x	x
Land use diversity		x	
Land use mix		x	x
Residential	x	x	x
Commercial	x	x	x
Mixed-use		x	
Office		x	
Industrial		x	x
Institutional		x	
Open space/parks	x	x	x
Recreational	x		
Restaurants			x
Transportation	x	x	
Others		x	
Intersection density			x
Parking density			x
Bike lane density			x
Road density			x

2.3 Supervised Machine Learning

This chapter will only focus on the supervised machine learning techniques used in this study. Therefore, machine learning tools that will be used in this study such as clustering and linear regression will not be covered here. Those will, however, be introduced shortly within the methodology section.
Supervised machine learning is a computing learning tool that enables the user to discern relationships and patterns within large datasets [Awad and Khanna, 2015]. It does this through a training process where the algorithm successively gets better at recognizing the trends within the given dataset. Eventually, the model will be able to detect how each independent variable in the dataset influences the dependent variable, and if given a set of independent variables, it will be able to predict the dependent variable.
In practical terms, a supervised machine learning algorithm is given a set of images containing images of oranges and bananas (independent variables) along with the key (dependent variable) to train on. Eventually, the model will be able to recognize patterns within the images, such as colour and shape and will be able to, within a degree of certainty, predict what fruit is presented to it when given a completely new image it has not seen before.
There are different supervised machine learning algorithms with variances in their functions, which will yield varying results [Awad and Khanna, 2015]. Some of them will be listed and explained below.

Random forest creates parallel decision trees, which all go through the training process of the algorithm on different sets of data [Awad and Khanna, 2015]. These trees are called weak learners and are poor at recognizing trends and patterns within the dataset. These weak learners then combine to create a strong learner, which, through combining the knowledge of the weak learners, can make better predictions. The strong learner's ability to make correct predictions is based on averaging the predictions of the weak learners, which reduces variance in the predictions and offers good accuracy. The model also keeps track of each variable's contribution to the accuracy of the results, which tells the user the importance score for each feature.
Extreme gradient boosting, or simply XGBoost, is a form of gradient boosting decision trees (GBDT) model. GBDT, similarly to random forest, utilizes weak learners [Natekin and Knoll, 2013]. However, these trees are not trained in parallel, but in a series using a boosting method. The basis of the boosting method is to train the weak learner one at a time concerning the previous learner's mistakes, and thus, iterative improvements are made with each learner. When the training procedure is done, each learner votes on the prediction, but their votes have a weight based on each learner's performance. The XGBoost method builds on this basic GBDT formula by implementing several features, such as allowing parallel processing, which enables the algorithm to train several learners at once, which speeds up the training process and regularization, which prevents overfitting and sparsity awareness, which improves its performance where data may be missing which often occurs in real datasets.
Long short term memory or LSTM is a type of recurrent neural network or RNN often used when working with sequences of information such as languages or time-series [Kim et al., 2022]. LSTM functions using a cell state that examines an element in a data

sequence called a step. At each step, the cell state retains certain information from the previous step. What information is kept is controlled by three gates: the forget gate, which decides which information to ignore; the input gate, which determines what information should be added to the cell; and lastly, the output cell, which predicts the output by using the output gate. This method can predict e-scooter demand in an area over time using temporal data such as demand and weather. However, due to the sequential nature of LSTM, it is not suitable to predict demand using non-temporal data.

3
Methods

This section will discuss the methodology of the study. First, a literature study was conducted to get an overview of the subject to discover relevant methods and results that may be applied in this study, and answer the research questions. These theories can be seen in Section 2. Secondly, the study area was defined, and data was collected, cleaned, and processed to fit the study area and models. Lastly, the analysis was conducted. The analysis consists of two major parts: the usage pattern analysis and the influencing factors analysis.
The usage pattern analysis primarily used the e-scooter transaction data to analyze the demand and trip patterns and how they differ spatially and temporally.
In **the influencing factors anlysis**, several different factors were analyzed as to what degree they influence the demand for e-scooters in Gothenburg.
A flowchart of the methodology can be seen in Figure 3.1.

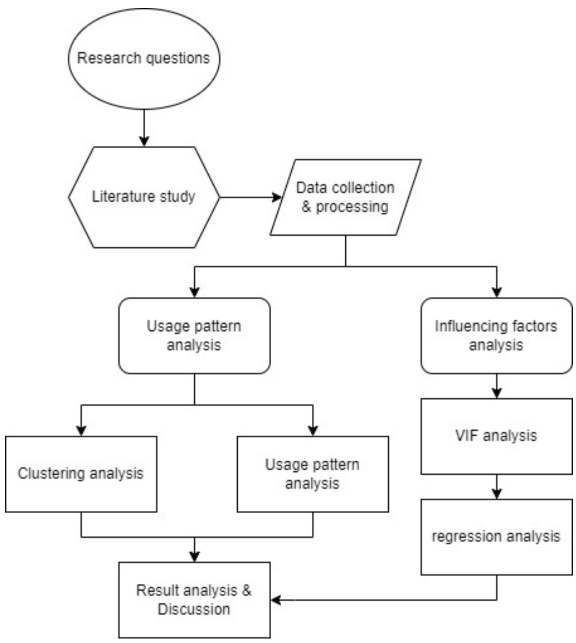


Figure 3.1: Methodology flowchart

3.1 Study Area

As stated in Section 2.1, e-scooters cannot be used in the entirety of Gothenburg and are enforced using geofencing, which causes the vehicles to shut off when exiting the

permitted area. On the mainland, the area covered stretches from Kungsladugård and Högsbohöjd in the west to Kålltorp in the east. On Hisingen Island, the area is centered around Eriksberg, Linholmen, and Backaplan.
To partition this study area into different analysis zones (AZs), the zoning system established by the city of Gothenburg municipality [Göteborgs Stad, 2022a] is used. These zones are hierarchically sorted by size in the following order, largest to smallest: Stadsområde (city zone), mellanområde (intermediate zone), primärområde (primary zone) and basområde (base zone). These zones are used for statistical data collection by the city of Gothenburg and made public through their website [Göteborgs Stad, 2022b]. To create a model with a high resolution, the smallest of the zones, the base area, is used in this study as the AZs.

3.2 Data

Several different datasets were used to achieve a holistic result for the usage pattern analysis and influencing factors analysis. These data sets are presented in Table 3.1, divided by its geographical partitioning.

Table 3.1: Data used in the study

Dataset	Source
Base zone	
Population	[Göteborgs Stad, 2022b]
No. of cars / population	[Göteborgs Stad, 2022b]
Age (proportion under 30)	[Göteborgs Stad, 2022b]
Gender	[Göteborgs Stad, 2022b]
Primary zone	
Higher education (>3 yrs)	[Göteborgs Stad, 2022b]
No. of students	[Göteborgs Stad, 2022b]
Unemployment	[Göteborgs Stad, 2022b]
Median income	[Göteborgs Stad, 2022b]
Other Data	
E-scooter transaction data	VOI
Temperature	[SMHI, nd]
Precipitation	[SMHI, nd]
Wind (max of avg)	[SMHI, nd]
Bus/tram stops	OpenStreetMap
POI data	OpenStreetMap
Roads	OpenStreetMap

The data collected for the base and primary zones is primarily statistical and demographic and is connected to each area. The primary zones comprise many base zones; therefore, in the final analysis, some AZs will have the same input since they belong to the same primary zone. Spatial and time-dependent datasets were used along with these demographic and statistical data, spatial and time-dependent datasets were used. These spatial datasets are bus stops, point-of-interest (POI) data, road data, and the e-scooter transaction data.

The bus stops, POI, and road data were collected from OpenStreetMap, a crowd-sourced geographic database that contains the geolocation of the object and the class to which it belongs. For the POI data, the class can be, e.g., hospital, cafe, or statue, and for the road, the class describes the type of road, e.g., residential, motorway, or cycleway. Temperature, precipitation, and wind speed were temporal factors to assess the effect of weather on the e-scooter demand. The weather data is gathered, and the data is given for each hour of the year.
The e-scooter transaction data serves as the basis of this report. It is gathered from the e-scooter company VOI through their API (application programming interface) and contains 754,000 trips from 284 days from 28/12/21 until 31/12/22. However, the connection to the API is sometimes lost, and thus, some days are missing. Overall, 272 days worth of data was collected out of 369 days possible. Most months have at least 20 days' worth of data. However, May, June, October, and December have 12, 15, 13, and 15 days, respectively. A summary of the data collected can be seen in Table 3.2. This dataset contains the date, time, and coordinates of when and where the trip started and ended, the distance traveled, use time, and the average speed during the trip. The destination coordinates will not be used in this study as the focus will be on what drives demand at the origin of the trip.

Table 3.2: E-scooter transaction data available

Weekday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**	**Total**
Month								
Jan	4	3	3	4	3	5	5	27
Feb	4	4	4	4	4	4	4	28
Mar	3	3	3	4	3	3	3	22
Apr	4	4	4	4	5	5	4	30
May	2	2	2	2	1	1	2	12
Jun	3	3	3	2	1	1	2	15
Jul	4	3	4	4	4	4	5	28
Aug	5	5	5	4	4	4	4	31
Sep	4	4	4	5	5	4	4	30
Oct	1	1	2	3	3	2	1	13
Nov	3	3	3	3	3	3	3	21
Dec	2	2	2	3	2	2	2	15
Total	39	37	39	42	38	38	39	272

3.3 Data Processing

3.3.1 E-scooter Transaction Data

Eventual GPS malfunctions on the e-scooters and other circumstances may make some trips in the dataset outliers in some measure. Therefore, the data was cleaned before it was used in the analysis to remove these outliers and faulty data, as they might skew the results and lower the performance of the analysis. For this study, the upper and lower 2.5 percentiles of distance, trip duration, and average speed were trimmed, leaving 664,000 trips from the original 754,000. The trips' starting geolocations were then linked with the different AZs to establish how many trips originated in each AZ. This number was then divided by the number of days to calculate the average daily demand for each AZ.

3.3.2 Analysis Zones

To ensure that every analysis zone (AZ) has sufficient data to analyze and to remove any outliers where the GPS function may have malfunctioned, the AZ with less than 100 trips was removed. Afterwards, there were 406 AZ remaining. These zones are shown in Figure 3.2

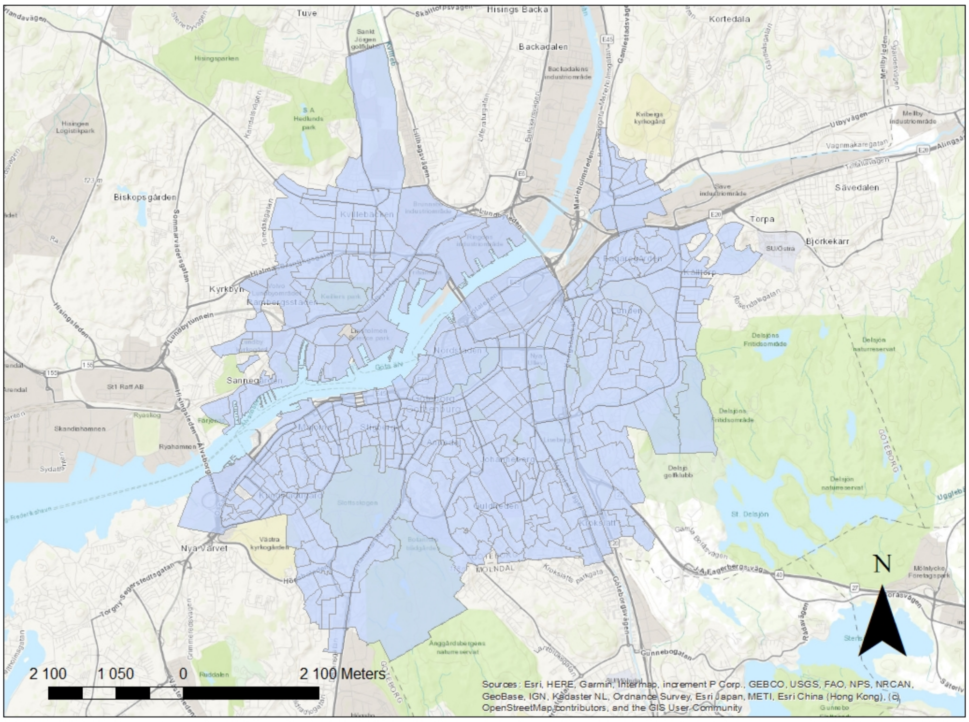


Figure 3.2: The analysis zones

3.3.3 Bus Stops

The bus stops are gathered as point of interest data, which was then, in ArcGis, given a buffer with a radius of 300 meters around them to act as an area of influence. This is a theoretical area where people are within walking distance of a bus stop. The areas were then split up along the base zone boundaries, and each base zone's total area of influence was aggregated.

3.3.4 Point of Interest Data

As stated above, the POI data have a specific class attached to each point. These classes were designated into seven major classes for easier processing and analysis. These major

classes are commercial, residential, educational, health, public, recreational, and other. What major classes each minor class belongs to is presented in Appendix A. These major classes will represent the built area in each AZ.
There are two ways to estimate the built area in each AZ [Gao et al., 2021]. One is to use the number of POIs of each class in each AZ; another is to use the term frequency-inverse document frequency (TFIDF), which is a method to weigh the importance of a word in a collection of documents. However, using the number of POIs is inaccurate when considering the built area. When collected from an online map, some POIs might be underrepresented in the dataset. This causes the ratio of that POI to be lower than in reality. TFIDF considers this eventual underrepresentation in the dataset by checking the number of each class in an AZ by the total number in the entire dataset. This is done by giving it a weight related to the class frequency in all AZ. The ratio of each class in an AZ, R_{ki}, is calculated with equations 3.1-3.4:

$$td_{ki} = \frac{N_{ki}}{\sum_{k=1}^{K} N_{ki}} \tag{3.1}$$

$$idf_k = log \frac{\sum_{i=1}^{D} N_{ki}}{\sum_{i=1}^{D} \sum_{k=1}^{K} N_{ki}} \tag{3.2}$$

$$tdidf_{ki} = td_{ki} \cdot idf_k \tag{3.3}$$

$$R_{ki} = \frac{tdidf_{ki}}{\sum_{k=1}^{K} tdidf_{ki}} \tag{3.4}$$

where:

K	The number of POI classes in AZ i
N	The number of AZs in the study area
N_{ki}	The number of POIs belonging to POI class k in AZ
td_{ki}	Term frequency of POI category k
idf_k	Weight of POI class k

3.3.5 Road-related factors

The road data is extracted as lines with a geolocation. Firstly, these lines were intersected at the borders of the AZs using ArcGIS and given the attribute of which AZ they were in. Then, similarly to the POI data, each road was attributed to one of 6 major classes depending on which minor class they belonged to. The major classes are residential, commercial, arterial, cycle, pedestrian, and other. What major classes each minor class belongs to is presented in Appendix A. The length of the road classes in each AZ was then calculated and divided by the area of the AZ to extract the road density.

3.4 Usage Pattern Analysis

As mentioned above, the usage pattern analysis is spatially and temporally conducted. For the spatial analysis, several maps were produced with the clustering method to analyze the usage pattern as it differs throughout the city. For the temporal analysis, graphs were made to show how the e-scooters were used through the year, weeks, and days.

3.4.1 Cluster Analysis

The next part of the spatial analysis was to study the usage patterns in the AZs. This was conducted with a machine-learning method called bottom-up hierarchical clustering. Clustering is a tool that seeks to group similar objects by measuring the distances between variables within the objects. In a bottom-up hierarchical clustering, all objects, or data-points, are their own clusters. The cluster is then paired with its closest pair. This process of merging clusters is then repeated until there is only one single cluster with all objects. Figure 3.3 shows a visual example of this. The clustering process can be stopped wherever on the y-axis by inserting a threshold value equaling a value on the y-axis. If this is done, the number of clusters equals the number of 'legs' at that value on the y-axis.

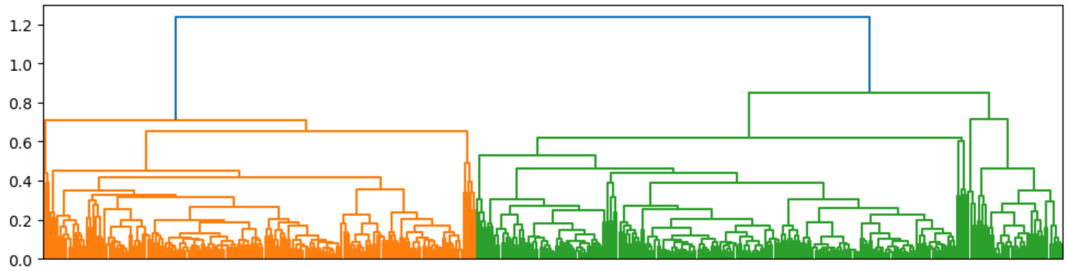


Figure 3.3: Visual example of the clustering method. The different AZs come from the x-axis, and the threshold can be seen on the y-axis.

In this application, the clustering was made for three cases: the average demand in each hour, weekday, and month in every AZ. By setting a certain threshold, an adequate number of clusters can be produced, clearly showing that different demand patterns are present throughout Gothenburg. The number of clusters in each analysis was in this case derived by trial and error. Different numbers of clusters were tested for each time resolution to find the number of clusters where the results were of quality and interpretable.

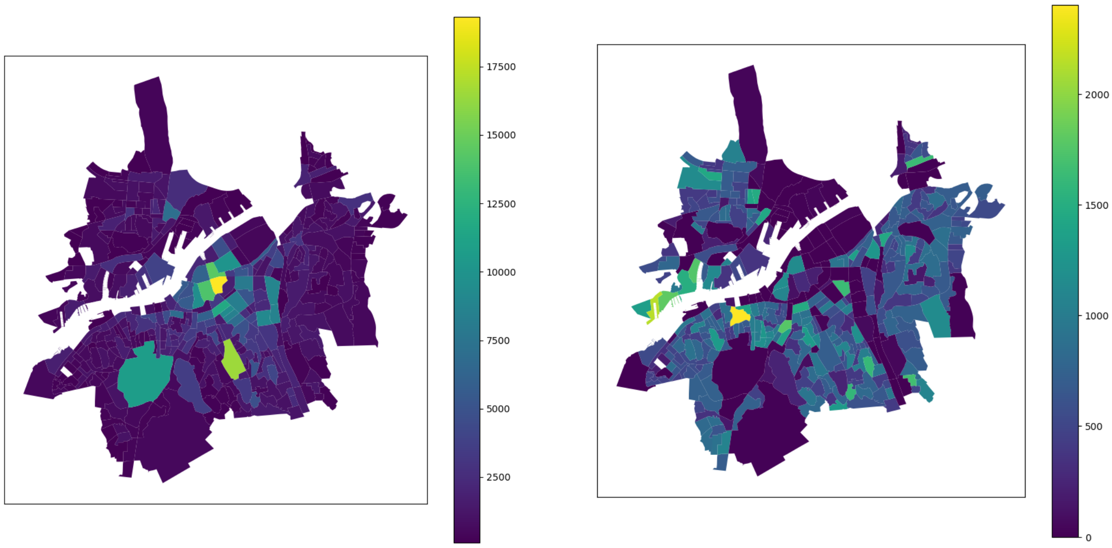
3.4.2 Temporal Analysis

Several aspects of the general usage were analyzed for the temporal analysis. This is to understand further how e-scooters are used in terms of demand, trip distance, and trip duration in different time resolutions. The analyses made can be seen in the list below:

- Average trip distance by time of day, weekday, and month.
- Average trip duration by time of day, weekday, and month.

3.5 Influencing Factors Analysis

An early assumption was made that the population in each AZ was an important factor in e-scooter demand. Therefore, a simple map was made, showing the hourly e-scooter demand by different aspects, such as the population in each area. This examines whether any simple conclusions could be made about the demand geography. The maps showing each AZ's total demand and population can be seen in Figure 3.4. It could be concluded that population alone does not induce demand; thus, more determinants were needed to analyze what influences the demand in Gothenburg.



(a) **(b)**

Figure 3.4: (a) E-scooter demand in each AZ (b) Total population within each AZ

Moving forward, the analysis included the data in Table 3.1, and a more robust analysis method was needed. The analysis was then made using supervised machine learning, a way to study the overall effect on the demand by each factor, and was conducted separately. One model is for the spatial factors, and another is for the temporal characteristics (weather, weekdays, and months). These methods were further explained in Chapter 2.3.

Before conducting the analyses, a variance inflation factor (VIF) analysis was done. VIF is a measure to check the presence of multicollinearity between the factors. Multicollinearity is when two or more variables in a regression model have a high correlation. For instance, if housing prices were to be predicted in a regression with the help of the total size of the house and the number of bedrooms and bathrooms, there would most likely be a high correlation between the different variables. Having correlations in a regression such as the one made makes the regression unstable and may lead to unreliable results. Therefore, variables with a VIF value higher than ten were removed from the model. The remaining and removed factors can be seen in Table 3.3.

Table 3.3: Remaining and removed factors after VIF analysis

Spatial	Temporal
Remaining	
Population	Month
Population Density	Weekday
Proportion of students	Minimum Temperature
POI commercial	Average Precipitation
POI educational	Max Precipitation
POI health	Maximum wind speed
POI other	Minimum wind speed
POI public	
POI recreational	
POI residential	
Bus stops	
Residential roads	
Arterial roads	
Commercial roads	
Cycle roads	
Pedestrian roads	
Other roads	
AZ area size	
Young residents	
Cars per population	
Removed	
Higher education	Average temperature
Average Income	Minimum precipitation
Male population	Average wind speed
Unemployment	Maximum temperature

Next, supervised machine-learning models were trained using the cross-validation procedure. The two machine learning models, random forest and XGBoost, were deemed the most appropriate for this analysis. This splits the dataset into two parts, one for training the dataset and one for validation, which is done to assess the performance and validity of the machine learning model. The model uses the influencing factors as explanatory features and the daily demand as the dependent variable. For this model, 20% of the dataset was used to validate the model.
Then, a comparison between the two different fittings protocols regression score coefficients was made to determine which model has the most accurate fitting. The regression score coefficients were R-squared (R^2), the root mean squared error (RMSE), and the explained variance (EV). The regression score coefficients were also compared to the coefficients of a standard linear regression to check the validity of using a machine learning method instead of a standard linear regression.
After this, the feature importance scores were calculated for the influencing factors. The feature importance score measures the impact each influencing factor has on the daily demand for e-scooters in Gothenburg. Charts showing the partial dependency between the factors and the demand were also created. The partial dependency is used to highlight

the impact of a single variable within the dataset. This is done by keeping all other variables static while changing the value of the chosen variable.

4
Results

This chapter is intended to showcase the results of the usage pattern analysis and the influencing factors analysis described in Chapters 3.4 and 3.5, along with observations that can be made by the results. This is to answer the questions formulated in the aim.

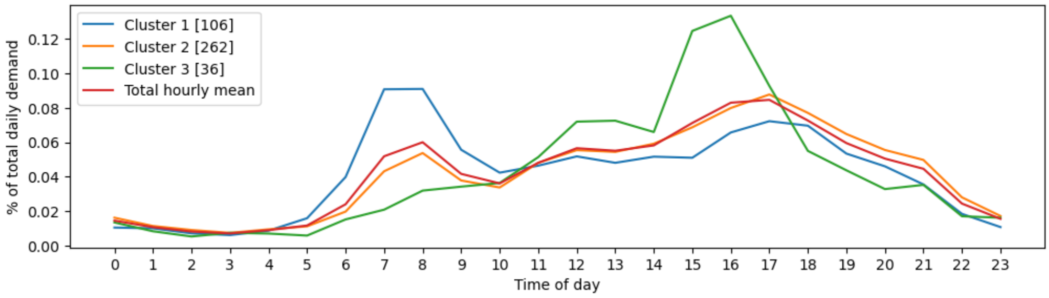
4.1 Usage Patterns

4.1.1 Clustering Results

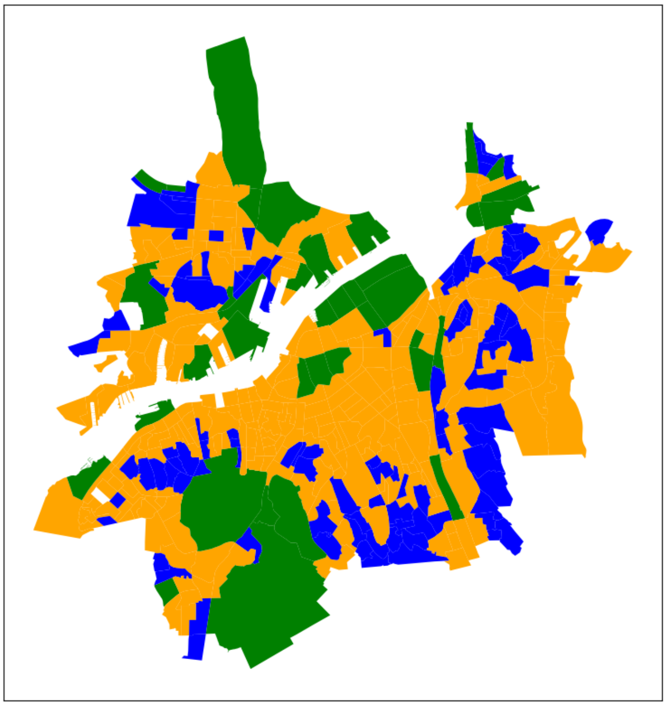
First, the result from the clustering analysis will be presented. These show the demand patterns in a specific time resolution and how they differ throughout Gothenburg. The number of clusters in each analysis was in this case derived by trial and error. Different numbers of clusters were tested for each time resolution to find the number of clusters where the results were interpretable and optimized the quality.

4.1.1.1 Hourly Demand

The first clustering analysis was made on the hourly demand and can be seen in Figure 4.1 where (a) is the clusters' average demand for each hour of the day. As can be seen, cluster 1 represents the AZs where most of the demand for e-scooters is in the morning between 6 and 9, with a slight increase between 16 and 18. The cluster is represented with the same colours in (b), and it can be seen that this usage pattern is most prominent on the outer perimeters of the city. This pattern could be explained by people using e-scooters for their morning commute to work. The second cluster pattern is the most commonly occurring in Gothenburg and represents a more stable demand throughout the days, with slight increases in the morning and evening. The third and last cluster has very low demand in the morning, then rises during mid-day, with a significant increase at 15 to 17, where it decreases afterward. This cluster is spread throughout the city but may represent areas where people work and take an e-scooter home in the evening or recreational spaces such as parks. All the clusters show a very low demand between 23 and 5 in the morning. This can be explained by the fact that most people sleep these hours and because e-scooters are shut off between these hours on the weekends to lower the risk of drunk driving.



(a)

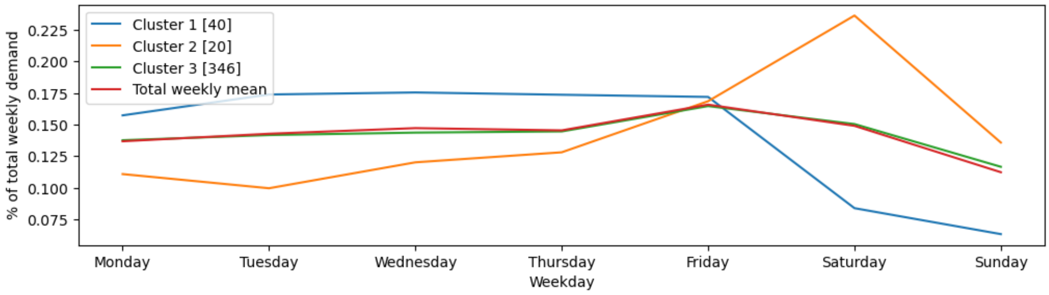


(b)

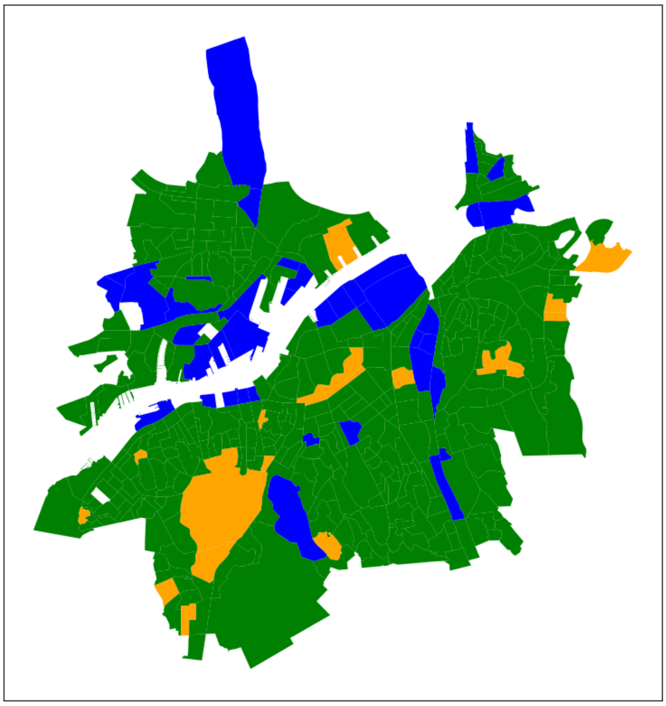
Figure 4.1: **(a)** E-scooter demand throughout the day. **(b)** Map showing what cluster each AZ belongs to.

4.1.1.2 Weekly Demand

The second cluster analysis was conducted on the demand variance of the weekdays, which can be seen in Figure 4.2. As can be seen in 4.2(a), the only difference between the clusters is how the demand differs during the weekend, where cluster 1 has a higher demand during workdays than during the weekends, cluster 2 has a higher demand during the weekends, and cluster 3 with stable demand throughout the week, which is also the most commonly occurring pattern. Cluster 1 is most likely AZ, with a higher degree of workplaces than recreational and residential land use. Cluster 2 is most likely the opposite of Cluster 1, as some AZs contain recreational areas. Coloured yellow in (b) are places such as Slottskogen, Kungsparken, and Valhalla, where many spend their weekends. The rest with a more stable demand variance most likely contain a mix of residential, commercial, and recreational activity, which induces demand no matter the weekday. However, something that can be extracted from every cluster is that demand decreases during Sundays despite what demand pattern is present the rest of the week.



(a)

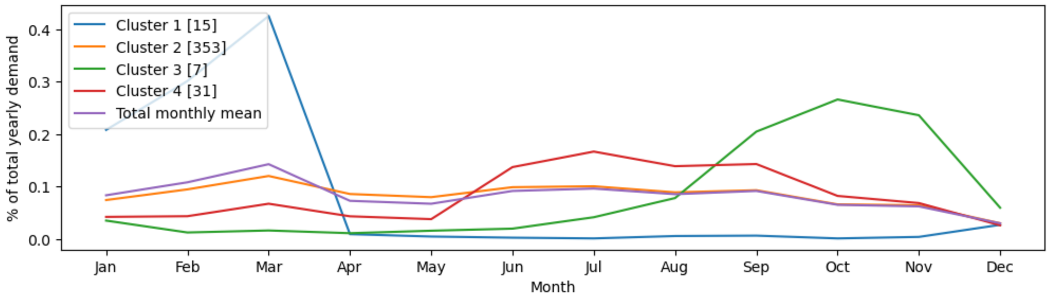


(b)

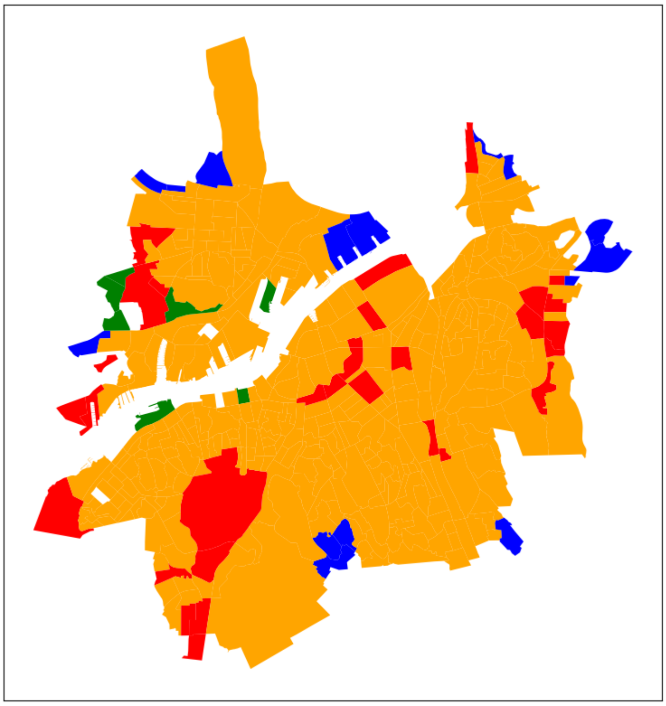
Figure 4.2: **(a)** E-scooter demand throughout the week. **(b)** Map showing what cluster each AZ belongs to.

4.1.1.3 Monthly Demand

The third and last cluster analysis was conducted for the yearly demand, with the average demand ratio for the months in each cluster shown. For this analysis, 4 clusters are shown with two clear outliers, 1 and 3. As seen in (b) in Figure 4.3, both are mainly situated at the edges of the city, and this could be explained by changes in the permissible riding zone set by VOI. However, changes in the geofencing have not been published by the e-scooter operator. The same conclusion could be made about cluster 4. However, some of the AZs coloured red in (b), such as Slottskogen and Kungsparken, indicate areas where demand would be higher in the summer months, which is displayed in (a). However, most AZs belong to cluster 2, which has a stable demand throughout the year with a slight decrease in the winter months. However, in both clusters 2 and 3, there is a slight decrease in demand from March to April. This is most likely due to the decrease in the total amount of e-scooters permitted in Gothenburg, which took effect on April 1st, 2022 [Göteborsposten, 2022]. Before this date, there was no maximum set by lawmakers, but an estimated maximum of 9500 e-scooters was reduced to 4000 vehicles.



(a)

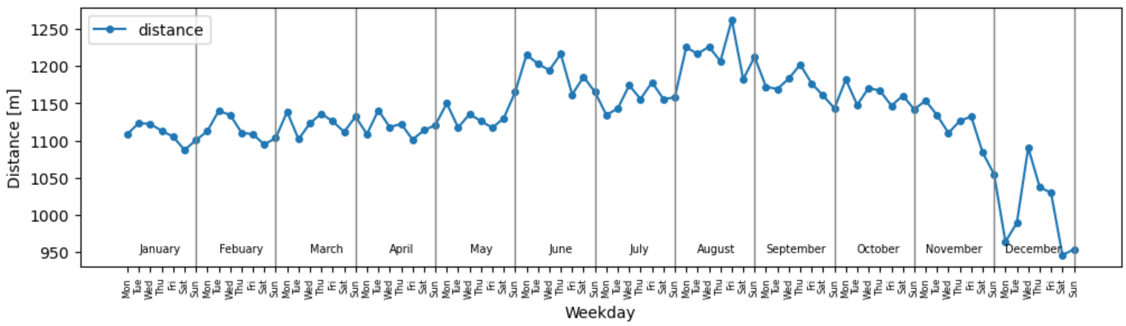


(b)

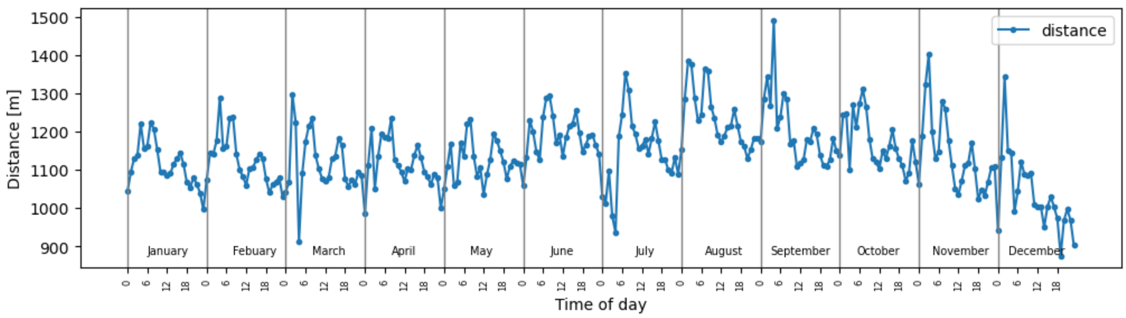
Figure 4.3: (a) E-scooter demand throughout the year. **(b)** Map showing what cluster each AZ belongs to.

4.1.2 Temporal Usage Patterns

Next, results from the temporal analysis where trip duration and trip distance were analyzed concerning hourly, weekday, and monthly resolutions. In Figure 4.4, the average distance traveled in each weekday in every month is displayed in (a), and the average distance traveled for each hour of the day in every month is displayed in (b). In (a), the average distance stays relatively even between each month, with some longer trips taking place during June and August; however, the weekday does not seem to impact travel distance significantly. In (b), there is no significant difference between the months. However, longer trips seem to be taken in the mornings around 7-8, along with a smaller, second spike in the afternoon at 17.



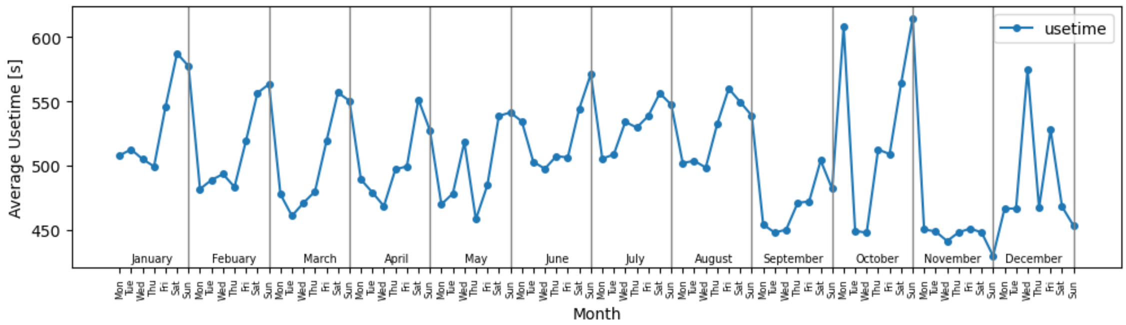
(a)



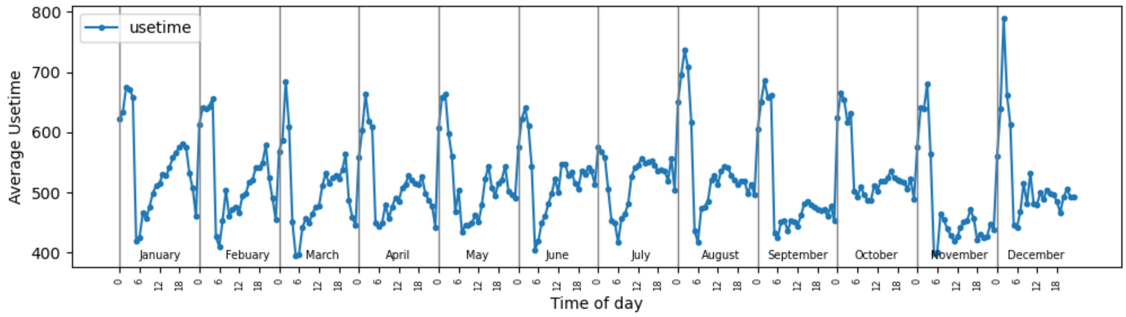
(b)

Figure 4.4: **(a)** Average trip distance for each weekday in each month. **(b)** Average trip distance for each hour of the day in each month.

In Figure 4.5, the average trip duration is shown in the same manner as the trip distance above. However, in contrast to trip distance, the trip duration shows a much more distinct pattern than the distance. The longest trips are taken on Fridays and Saturdays from midnight to about 4 in the morning. This might be due to the lack of public transport during these hours and the fact that people tend to stay up later on these days.



(a)



(b)

Figure 4.5: **(a)** Average trip duration for each weekday in each month. **(b)** Average trip duration for each hour of the day in each month.

4.2 Key Determinants

4.2.1 Spatial Variables

Observing the regression scores in Table 4.1, it shows that the random forest model performed better than XGBoost and the linear regression. Due to this superiority, only the results of the random forest model will be shown and discussed in this report.

Table 4.1: Regression Scoring for each model

	Random Forest	XGBoost	Linear regression
R-squared (R^2)	0.502	0.292	0.195
Root Mean Square Error (RMSE)	5.569	5.712	6.753
Explained Variance (EV)	0.463	0.442	0.226

In Figure 4.6, the feature importance score of the random forest can be seen for the spatial factors shown in Table 3.3. The results of the model show that the most impactful determinant of e-scooter demand in a specific area is the number of bus stops in an AZ or, more precisely, the total area of influence by bus stops described in Chapter 3.3.3. Bus stops have a feature importance score of 0.32, about 3.6 times larger than the second largest score, the public POI ratio, followed by commercial POIs, residential roads, and other roads. After these, the remaining factors scores remain relatively equal with a slight difference. The factors with the lowest importance scores are residential POIs, other POIs, and educational POIs.

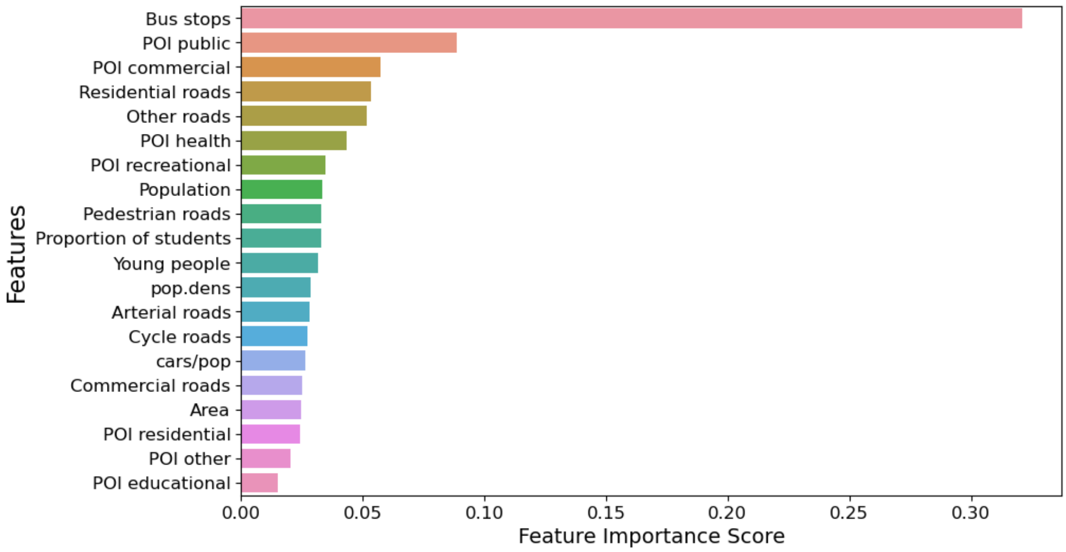


Figure 4.6: Graphic showing the impact on e-scooter demand by each spatial factor using random forest

4.2.2 Temporal Variables

Further, looking at the temporal determinant impact analysis results shown in Figure 4.7, one can see a clear result. The most impactful factors on the demand are the month of the trip and the minimum temperature of the day. The average precipitation, the weekday, max precipitation, max wind speed, and the minimum wind speed follow these.

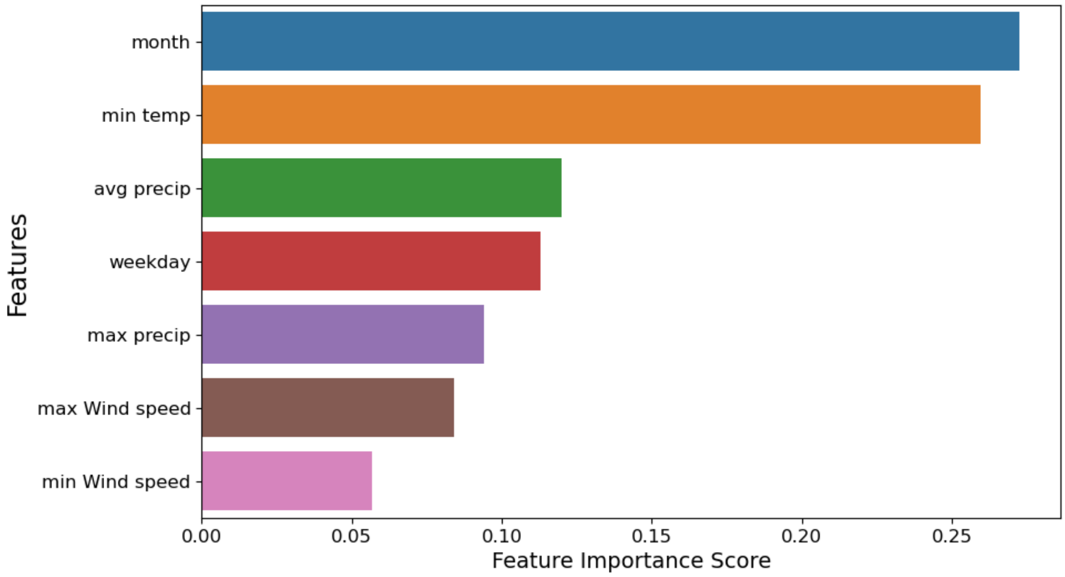


Figure 4.7: Graphic showing the impact on e-scooter demand by each temporal factor using random forest

4.2.3 Partial Dependency

However, what is not shown in Figure 4.6 and 4.7 is how the factors impact the demand for e-scooters. In Figure 4.8, some of the partial dependence charts are shown. These show the relation between the factor and the demand for e-scooters. Figures of the remaining impact charts can be found in Appendix B. Figure 4.8 (a) shows the impact of bus stops. It clearly shows that the demand for e-scooters increases with the prevalence of bus stops in the AZs while having a low effect otherwise. In (b), the partial dependence of residential buildings can be seen. More residential AZs have a lower demand for e-scooters than AZs with a higher degree of commercial activity, as shown in (c). The last chart (d) shows the young population's impact on the demand. From the ratio of 0 to 0.2, it shows some impact on the demand. However, this might be due to a lack of data. After a ratio of 0.3 in an AZ, it clearly shows that demand rises sharply with the amount of young people.

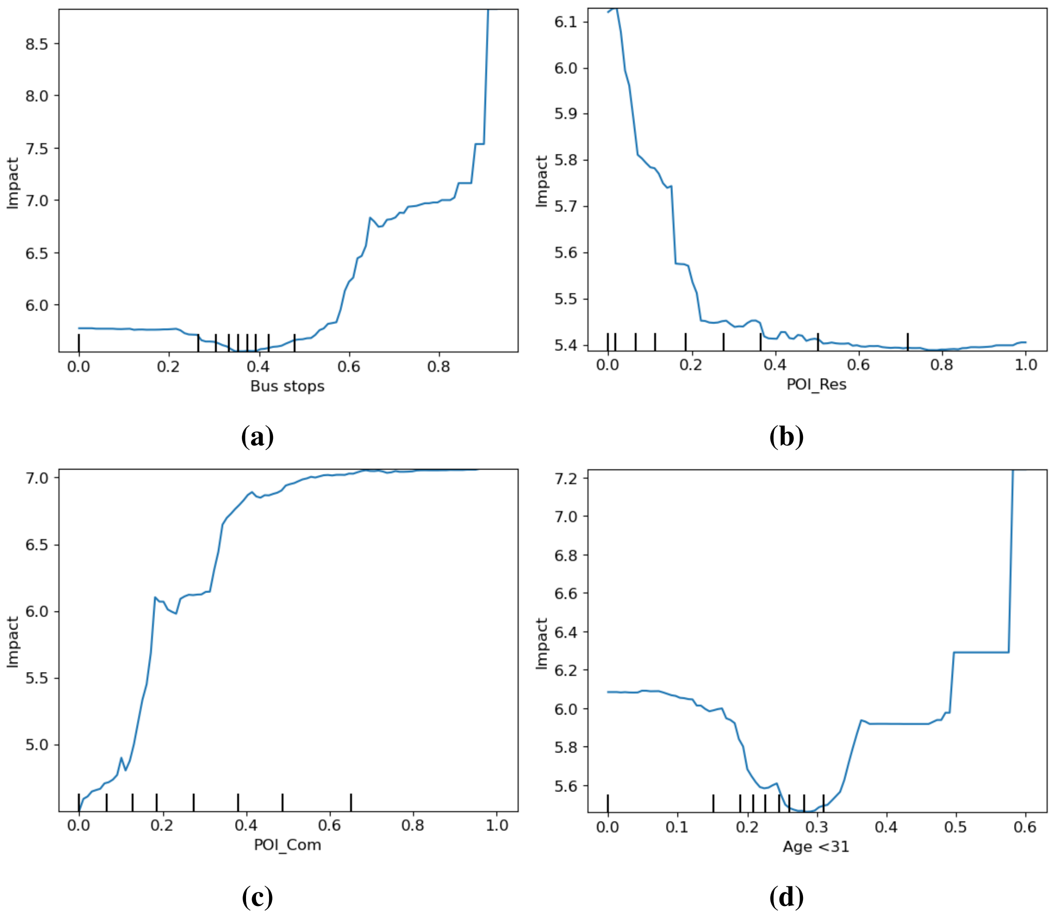


Figure 4.8: Graphs showing the pure impact of **(a)** Bus stops **(b)** Residential POIs **(c)** Commercial POIs **(d)** People aged 30 and younger

5
Discussion

This chapter will discuss the methodology and results of the report. This is done to examine flaws and investigate implementations of the results while discussing possible further investigations.

5.1 Methods & Data

This study solely studied the origins of each e-scooter trip to study the demand. However, only regarding origins answer a part of the usage pattern and structure. Implementing the destination of e-scooter trips would allow one to analyze the flow of e-scooters within Gothenburg accurately. For instance, regarding Figure 4.1, it can be seen that different hourly travel patterns exist in different parts of the city. A more educated travel analysis could be made if a similar clustering model were implemented concerning the destination of the trips. If the clusters where afternoon departures dominated showed a majority of arrivals in the morning, an assumption could be made that there are mainly daily commuters in that area. Employing this broader scope in the analysis of the influencing factors would also be of interest. It would enable the analysis of where people are traveling as well. Generally, the analysis would be larger in scope, covering the entire trip from origin to destination.

Observing Table 2.2, it can be seen that the primary zone data has a lower resolution than the base zone data. Since the analysis is made with the base zones as AZs, some variables will be identical for the AZs within a particular primary zone. This will affect the result of the machine learning analysis regarding these variables. In this case, the results for these variables will have a lower accuracy since the machine learning algorithm won't be able to capture patterns adequately. Therefore, an optimal solution would be to train the algorithm on a model with a lower resolution, with the primary zones as AZs, and adequately capture these intricacies.

5.2 Missing Data and Rule Changes

As stated in Chapter 3.2, only about 71% of the days within the period of the e-scooter data capture were gathered. The periods not captured were often in larger spans, which resulted in some months having less data available for the analysis, especially May, June, October, and December. In extension, this resulted in these months having very little data on certain days of the week. Observing Table 3.2, it can be seen that May and June only have one day's worth of data for Fridays and Saturdays, and October has one day's worth of data for Mondays, Tuesdays, and Sundays. The effect of this lack of data is that outliers in the data have a larger effect on the overall result in both the analysis of influencing factors and usage patterns. Its effect on the influencing factor is hard to quantify due to the nature of the analysis. However, it can have a major effect on the clustering analysis seen in Chapter 4.1.1 and especially the temporal usage patterns seen in Chapter 4.1.2. Looking at (a) in Figure 4.5, A very erratic pattern can be observed in May, October, and December, which does not conform to the patterns in other months. This is most likely due to the lack of data in these months, and one ought to be careful when making conclusions based on the data from these months. This pattern could be explained by holidays or other events, but these claims should be substantiated with other data.

Through the duration of the data capture from 28/12/21 until 31/12/22, there were two changes to laws affecting e-scooters in the city of Gothenburg [Göteborsposten, 2022] [Trafiken.nu, 2022]. The first one, enacted on April 1st, reduced the number of e-scooters allowed within Gothenburg. The second one, enacted September 1st, made driving or parking e-scooters on sidewalks or walkways illegal. The effects of the first rule change can be seen in Figure 4.3 with an apparent decrease in overall demand after March. It also affected the influencing factors analysis, which can be seen in the partial dependence charts, which show the relationship between demand and the month in Appendix B, which shows that the demand reduces after March. Therefore, this partial dependency graph is likely not entirely accurate, and the results shown in Figures 4.7 and 4.7 might not be completely accurate. However, the correlation shown still holds statistical integrity despite this.

5.3 Key Determinants Interpretations

Looking at the results in Chapter 4.2, it is clear that bus stops have by far the most significant feature importance score, and by observing the partial dependency graphs, this relationship is positive, meaning more bus stops in an area lead to higher e-scooter demand. However, it is essential to note that these variables do not exist in a vacuum. The underlying factors as to why there might be more bus stops in an area might be because this area has high commercial activity or is home to a larger transportation hub such as a train station. These variables would most likely see an increased demand without the high number of bus stops. Therefore, the relative importance of bus stops might also be somewhat inflated by these other factors. Despite this, the influence of bus stops should not be neglected. When looking at the results from a larger perspective, it can be seen that the built environment has the largest impact on e-scooter demand compared to the socio-economic determinants. This finding aligns with other research papers [Yang et al., 2022]. The socio-economic variable with the most influence was population, the 8th most important factor in the random forest model.

The partial dependency can also be used to find interesting patterns between demand and factors. Often, threshold effects can be observed once a specific value is seen. For instance, inspecting the non-linear relationship between the demand and population in Appendix B, such an effect is present. There is little to no demand in areas with a population of less than 500 people, but once the population rises above 500, there is a sharp increase in demand. Similar patterns can be seen for other determinants such as minimum temperature, number of students, bicycle roads, and other POIs. The opposite effect, where the demand suddenly decreases, can be seen in wind speed, weekdays, and precipitation. E-scooter companies can use these effects to determine where vehicles are placed or what areas to expand into.

5.4 Implementations

It has been shown that travel patterns and motivations can differ significantly between cities [Badia and Jenelius, 2021] [Bai and Jiao, 2020]. This study showed how e-scooters are used in one city, Gothenburg, and as such, the results should mainly be applied to Gothenburg. Therefore, implementations in the form of policies or practical changes to e-scooter-sharing businesses should be made with caution in cities other than Gothenburg. However, the study can contribute to making policy decisions to encourage e-scooter usage or reduce it, depending on the space and time. Figure 4.6 showed that bus stops are heavily linked with e-scooter usage. This implies that many use the e-scooter as a part of their daily commute to places not close to public transportation options or simply as a last-mile solution. Suggestions such as these should be backed by qualitative research reports to uncover the travel habits of the public so that the decision-makers can make a full implementation. However, research such as this report could be used as a foundation to suggest changes to public transport routes in areas where there seems to be a lack of commuting options for the last part of the trip. It can also be of value to see where to install special parking areas for e-scooters to incentivize proper parking of e-scooters. Further, it can be seen in Figure 4.5 that long trips are frequent on weekends at night when there is a risk of drunk driving. To decrease this risk, decision-makers could implement and encourage other transport modes in areas where this behavior is most frequent.

Another implementation of the study is the ability to create a model based on the influencing factors analysis to track and predict future e-scooter demand based on the built environment factors, social factors, and temporal factors used in this study. For instance, if an e-scooter company were to expand its geofencing area, the company and government could predict the demand and, by extension, predict the number of e-scooters needed for the extended area. Further implementations could also be made to predict the spatio-temporal usage patterns. The results shown in the clustering analysis in Chapter 4.1.1 could, together with the temporal determinant analysis results shown in Figure 4.7, predict the demand for each AZ in an hourly and weekly solution while including weather forecast data. This would allow e-scooter companies to make short-term decisions about where to allocate e-scooters and estimate what and how many trips are needed within each AZ to maximize efficiency within the system.
In conclusion, the study enables the development of more effective e-scooter-sharing systems in Gothenburg and, to some extent, other cities. This includes integrating e-scooters with different modes of transportation and developing pricing and marketing strategies that consider the target population's specific usage patterns. Doing this will enable commuters and those who only use e-scooters occasionally to make conscious decisions with more transport mode options, thus opening the possibility of reducing carbon emissions, congestion, and noise.

6
Conclusion

This study analyzed e-scooter ridership data from the year 2022 to explore travel patterns and the key determinants that influence e-scooter trips with the following research questions as a guide:

- How are e-scooter users travelling, and what temporal and spatial user patterns are present in Gothenburg?
- What factors are influencing the demand for e-scooters in Gothenburg? How much do these indicators influence the demand?

Regarding the first research question, many patterns with different time resolutions were explored. For the daily demand, there is a slight spike in the morning, and later in the evening, there is a maximum demand at hour 17. However, this pattern differs in some areas in Gothenburg, with the two spikes being exaggerated. In the weekly resolution, the demand remains steady, with a slight drop in the weekend. Some areas do not conform to this pattern, with either an increase or a sharper drop in the weekend. However, the demand pattern on a yearly resolution does not have a clear pattern. This is due to rule changes on the 1st of April, which lowered the number of e-scooters in Gothenburg to 4000 [Göteborsposten, 2022]. Because of this rule change, there is a drop-off in trips from March to April, which makes it difficult to extrapolate a pattern with substantial certainty. Nonetheless, there is a slight increase during June and July, which decreases into the autumn and winter. This increase in demand during the summer is more prominent in some areas, indicating that these areas have seasonal dependencies and are where people like to spend their summers. These areas include parks like Slottskogen and Kungsparken and recreational areas like Heden and Valhalla.
Further, regarding the temporal patterns, the longest distances travelled are during weekday morning hours. In contrast, the highest average use time is during nighttime on weekends. These opposing patterns can most likely be attributed to the longest distances to work and the most efficient travel path being travelled. The longest use times are during hours when people might travel to and from weekend entertainment in the city centres and bars.

Moving onto the second research question, the clear outlier with the largest feature importance was bus stops, followed by public buildings, commercial buildings, residential roads, other roads, and health POIs according to the random forest model, which had the superior regression score compared to the XGBoost model. Overall, the built environment had a larger impact on e-scooter demand than the socio-economic factors in an area.

However, it is essential to remember that these variables do not exist in isolation. Bus stops are often present in areas with high commercial activity or population.

Lastly, the implementations of this study are multifaceted, but the city-specific nature of travel patterns and motivations cautions against direct application to other cities. However, it can help influence policy decisions by optimising public transportation routes, predicting future e-scooter demand, and leveraging spatio-temporal usage patterns to improve the efficiency of e-scooter-sharing systems and reduce the environmental impact.

Bibliography

[Awad and Khanna, 2015] Awad, M. and Khanna, R. (2015). *Efficient Learning Machines Theories, Concepts, and Applications for Engineers and System Designers*. Apress Berkeley, CA. doi: https://doi.org/10.1007/978-1-4302-5990-9.

[Badia and Jenelius, 2021] Badia, H. and Jenelius, E. (2021). *Shared e-scooter micromobility: A review of travel behavior, sustainability, infrastructure use, safety and policy measures* . doi:10.13140/RG.2.2.19225.95841.

[Bai and Jiao, 2020] Bai, S. and Jiao, J. (2020). Dockless e-scooter usage patterns and urban built environments: A comparison study of austin, tx, and minneapolis, mn. *Travel Behaviour and Society*, 20.

[Gao et al., 2023a] Gao, K., Li, A., Liu, Y., Gil, J., and Bie, Y. (2023a). Unraveling the mode substitution of dockless bike-sharing systems and its determinants: A trip level data-driven interpretation. *Sustainable Cities and Society*, 98:104820.

[Gao et al., 2023b] Gao, K., Yang, Y., Gil, J., and Qu, X. (2023b). Data-driven interpretation on interactive and nonlinear effects of the correlated built environment on shared mobility. *Journal of Transport Geography*, 110:103604.

[Gao et al., 2021] Gao, K., Yang, Y., Li, A., and Qu, X. (2021). Spatial heterogeneity in distance decay of using bike sharing: An empirical large-scale analysis in shanghai. *Transportation Research Part D: Transport and Environment*, 94. doi: http://dx.doi.org/10.1016/j.trd.2021.102814.

[Göteborgs Stad, 2022a] Göteborgs Stad (2022a). Kartor - områdesindelning. https://goteborg.se/wps/portal/enhetssida/statistik-och-analys/geografi/kartor.

[Göteborgs Stad, 2022b] Göteborgs Stad (2022b). Statistikdatabas göteborgs stad. http://statistikdatabas.goteborg.se/pxweb/sv/.

[Göteborgs Stad, nda] Göteborgs Stad (n.da). Elsparkcyklar. https://goteborg.se/wps/portal/start/trafik-och-resor/trafik-och-gator/cykling-och-cykelvagar/elsparkcyklar.

[Göteborgs Stad, ndb] Göteborgs Stad (n.db). Lånecyklar (styr ställ). https://goteborg.se/wps/portal/start/gator-vagar-och-torg/cykling-och-cykelvagar/lanecyklar-styr-o-stall.

[Göteborgs Stad Trafikkontoret, 2014] Göteborgs Stad Trafikkontoret (2014). Trafikstrategi för en nära storstad. https://goteborg.se/wps/wcm/connect/32f1301c-7e10-4f6d-a0fa-ee4f1c2f3f3a/Trafikstrategi_Slutversion_swe_web_140402.pdf?MOD=AJPERES.

[Göteborsposten, 2022] Göteborsposten (2022). Klart: Antalet elsparkcyklar ska minskas drastiskt. https://www.gp.se/nyheter/g%C3%B6teborg/klart-antalet-elsparkcyklar-ska-minskas-drastiskt-1.67102386.

[Heumann et al., 2021] Heumann, M., Kraschewski, T., Brauner, T., Tilch, L., and Breitner, M. H. (2021). A spatiotemporal study and location-specific trip pattern categorization of shared e-scooter usage. *Sustainability*, 13. doi: https://doi.org/10.3390/su132212527.

[Jaadi, nd] Jaadi, Z. (n.d). A step-by-step explanation of principal component analysis (pca). `https://builtin.com/data-science/step-step-explanation-principal-component-analysis`.

[Kim et al., 2022] Kim, S., Choo, S., Lee, G. L., and Kim, S. (2022). Predicting demand for shared e-scooter using community structure and deep learning method. *Sustainability*, 14. https://doi.org/10.3390/su14052564.

[Mehzabin Tuli et al., 2021] Mehzabin Tuli, F., Mitra, S., and Crews, M. (2021). Factors influencing the usage of shared e-scooters in chicago. *Transportation Research Part A: Policy and Practice*, 154:164–185. ISSN 0965-8564.

[Natekin and Knoll, 2013] Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7. doi: https://doi.org/10.3389/fnbot.2013.00021.

[Shah et al., 2023] Shah, N. R., Guo, J., Han, L. D., and Cherry, C. R. (2023). Why do people take e-scooter trips? insights on temporal and spatial usage patterns of detailed trip data. *Transportation Research Part A*, 173.

[SMHI, nd] SMHI (n.d). Ladda ner meteorologiska observationer. `https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer/#param=airtemperatureInstant,stations=core/`.

[Trafiken.nu, 2022] Trafiken.nu (2022). Ny lag för elsparkcyklar. `https://trafiken.nu/goteborg/arkiv/2022/september/ny-lag-for-elsparkcyklar/`.

[Yang et al., 2022] Yang, H., Zheng, R., Li, X., Huo, J., Yang, L., and Zhu, T. (2022). Nonlinear and threshold effects of the built environment on e-scooter sharing ridership. *Journal of Transport Geography*, 104. doi: https://doi.org/10.1016/j.jtrangeo.2022.103453.

A

Appendix A

A.1 Classification for the OSM attributes

A.1.1 Land use POIs

Commercial

Hotel, hostel, restaurant, market place, bicycle shop, fast food, cafe, bar, supermarket, convenience, department store, pub, outdoor shop, bookshop, clothes, optician, car rental, nightclub, bakery, laundry, hairdresser, beverages, florist, travel agent, furniture shop, stationery, recycling clothes, bicycle rental, beauty shop, doityourself, video shop, sports shop, mobile phone shop, shoe shop, jeweller, toy shop, car dealership, gift shop, greengrocer, butcher, computer shop, vending parking, vending machine, car wash, guesthouse, retail, kiosk, brewery,industrial, office,commercial,warehouse.

Residential

Bungalow, allotment house, shed, semidetached house, detached, dormitory, hut, house, residential, apartments, shelter.

Educational

University, school, college, kindergarten.

Health

Hospital, clinic, dentist, pharmacy, doctors, chemist, veterinary

Public

Community centre, bank, atm, post office, museum, memorial, toilet, library, embassy, police, post box, fire station, town hall, courthouse, train station, civic, government, public.

Recreational

Garden centre, park, theatre, tourist info, cinema, stadium, playground, sports centre, attraction, picnic site, fountain, viewpoint, artwork archaeological, arts centre, drinking water, car sharing, theme park, bench, pitch, sports hall, church, religious, cathedral, historic, stadium, ruins, synagogue, chapel, pavilion, mosque.

Other

Recycling paper, waste basket, tower, recycling, recycling glass, wayside cross, comms tower, observation tower, general, newsagent, water well, camera surveillance, other.

A.1.2 OSM Roads classification

Residential

Residential, living street.

Commercial

Service, tertiary, tertiary link.

Arterial

Primary, secondary, motorway, trunk, primary link, trunk link, secondary link, motorway link.

Cycleway

Cycleway, bridleway, track.

Pedestrian

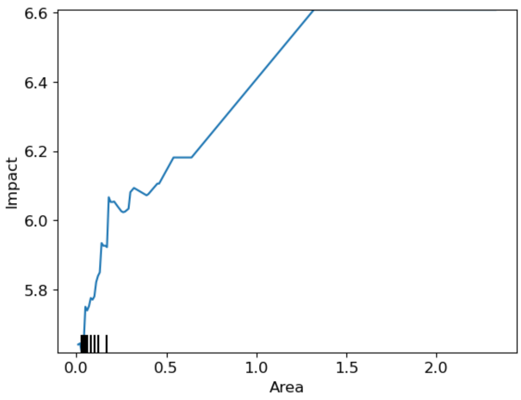
Pedestrian, footway, path, steps, track grade2, track grade4, track grade1, track grade3, track grade5.

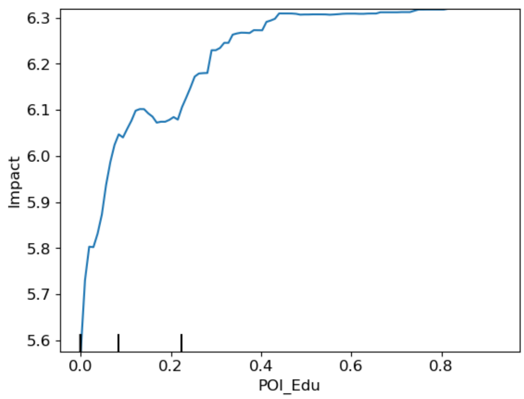
Other

Unclassified

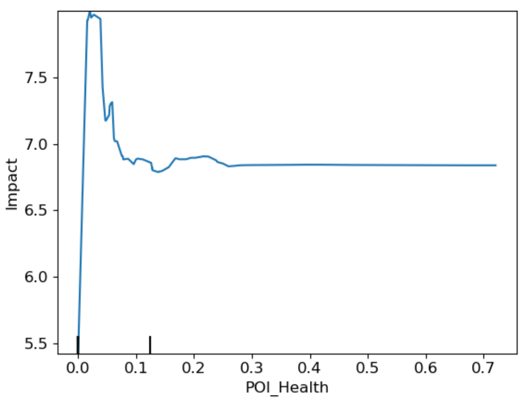
B

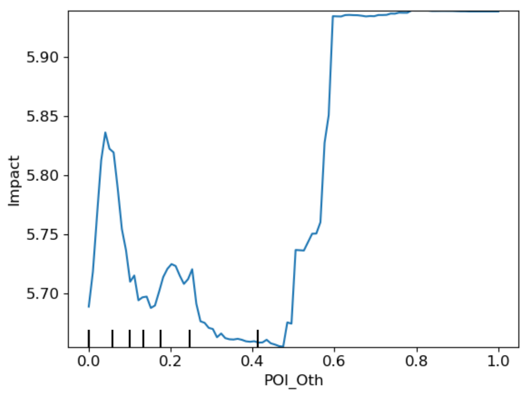
Appendix B: Partial dependency plots



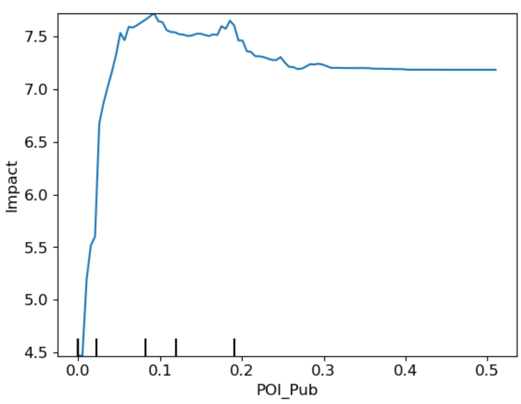


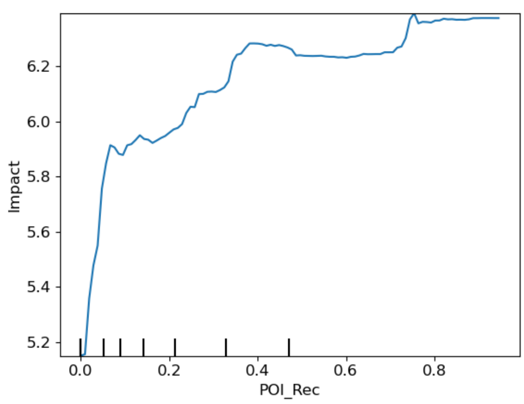
(a) **(b)**





(c) **(d)**





(e) **(f)**

Figure B.1: Graphs showing the pure impact of **(a)** Area **(b)** Educational POI **(c)** Health POIs **(d)** Other POIs **(e)** Public POIs **(f)** Recreational POIs

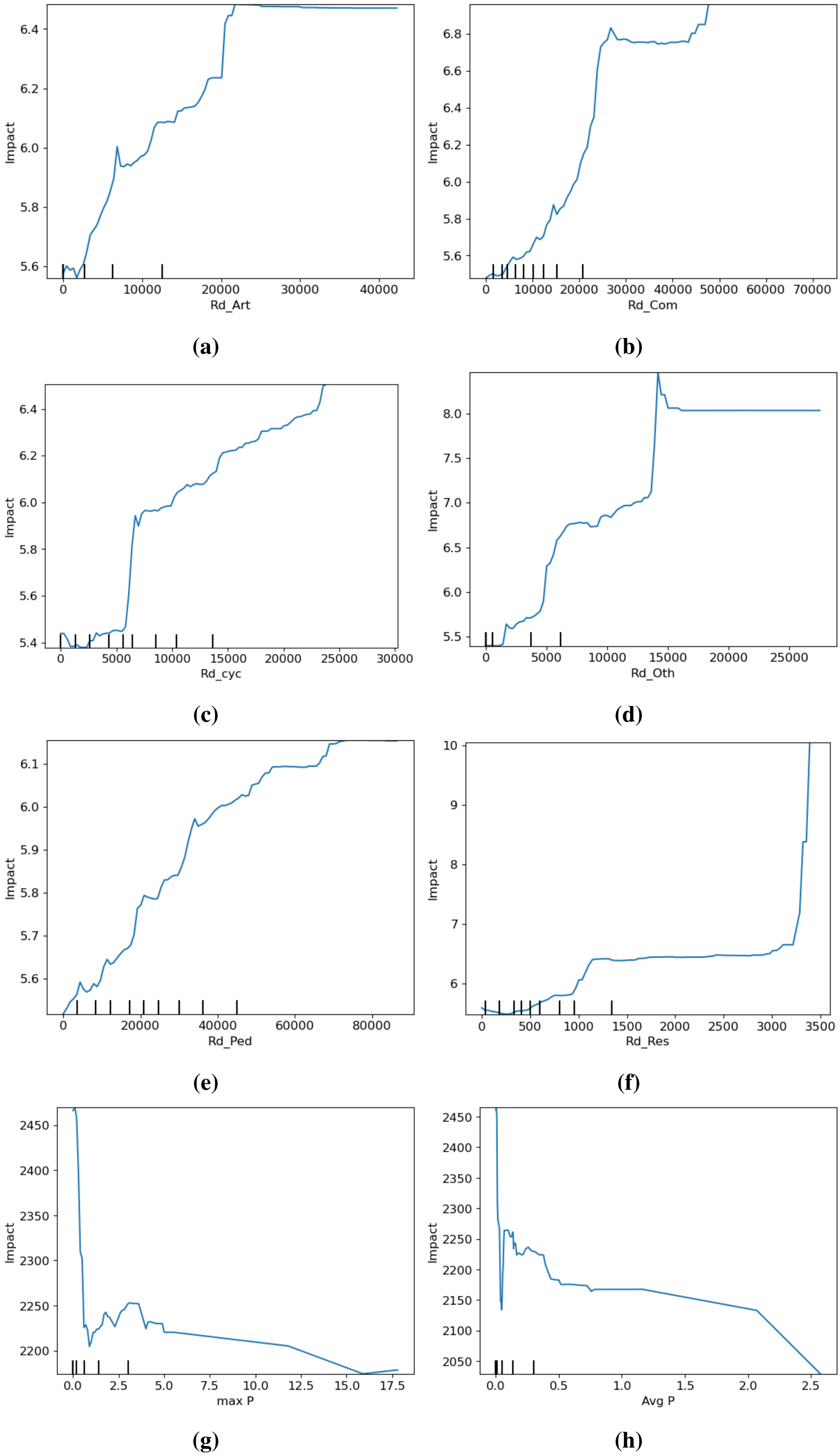
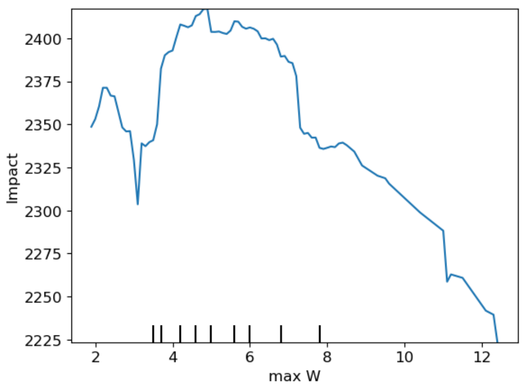
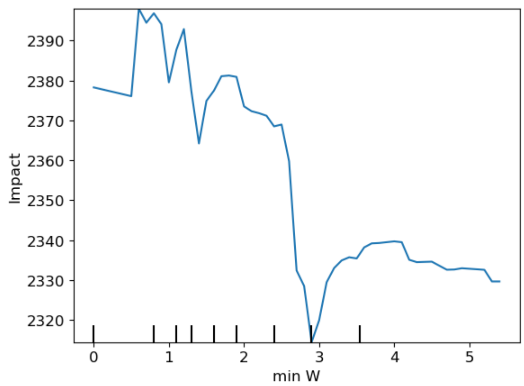


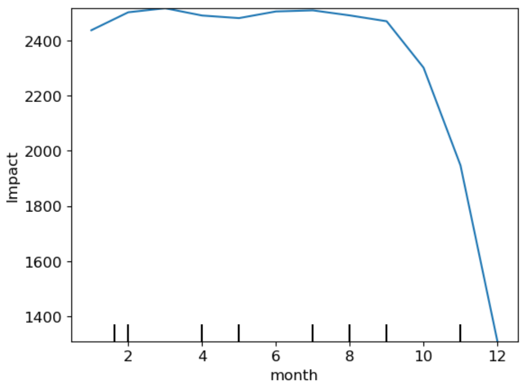
Figure B.2: Graphs showing the pure impact of **(a)** Arterial roads **(b)** Commercial roads **(c)** Cycle roads **(d)** Other roads **(e)** Pedestrian roads **(f)** Residential roads **(g)** Maximum precipitation **(h)** Average precipitation



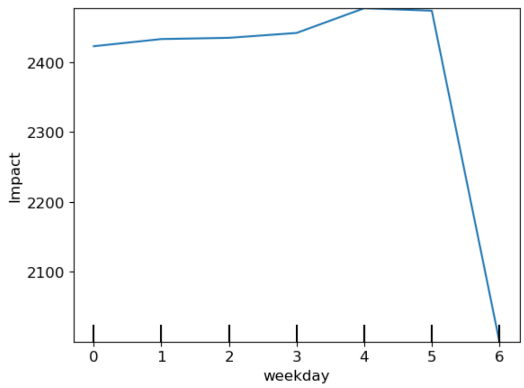
(a)



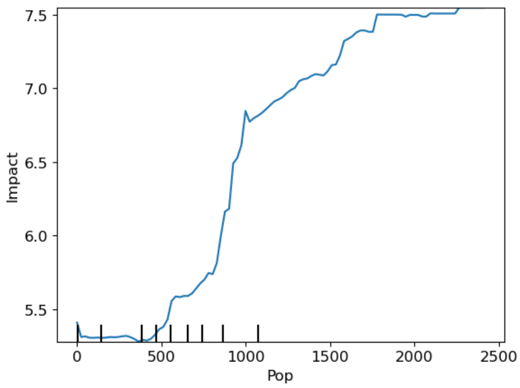
(b)



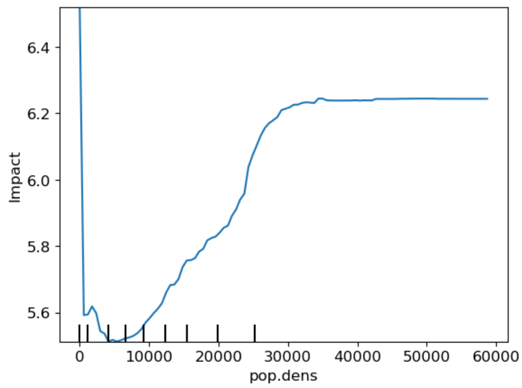
(c)



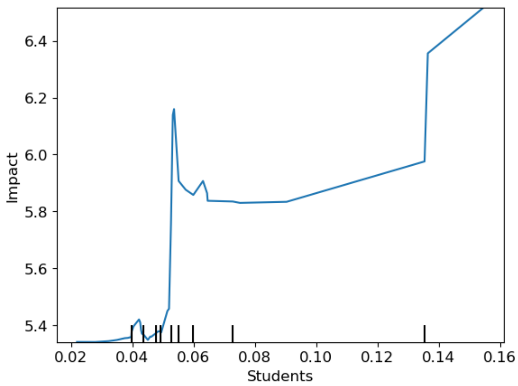
(d)



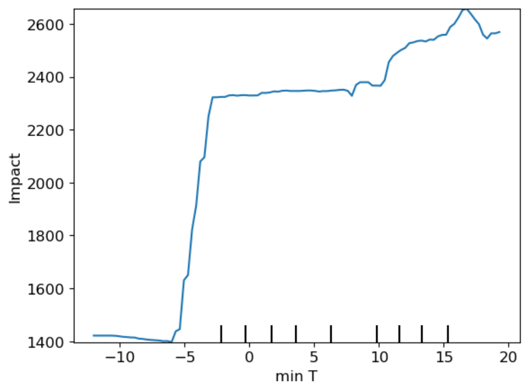
(e)



(f)



(g)



(h)

Figure B.3: Graphs showing the pure impact of **(a)** Maximum wind **(b)** Minimum wind **(c)** Month **(d)** Weekday **(e)** Population **(f)** Population density **(g)** Students **(h)** Students