Analysing and Predicting the Parking Occupancy of Micromobility Vehicles on Sidewalks

submitted in partial fulfillment for the degree of master of science

Johanna Fiebag
10974474

master information studies
data science
faculty of science
university of amsterdam

June 30, 2022

	Supervisor	Examiner
Title, Name	Dr. F.M. (Frank) Nack	Dr. S. (Sara) Mahdavi Hezavehi
Affiliation	UvA, FNWI, IvI	UvA, FNWI, IvI
Email	fnack@uva.nl	s.mahdavihezavehi@uva.nl

Universiteit van Amsterdam

Amsterdam Data Science

Gemeente Amsterdam

Analysing and Predicting the Parking Occupancy of Micromobility Vehicles on Sidewalks

Johanna Fiebag
University of Amsterdam
johanna.fiebag@gmail.com

ABSTRACT

Micromobility vehicles, such as bicycles, are often left on the sidewalk, where they limit the space of the already narrow pedestrian zone. A better understanding of micromobility parking and the possibility to predict the demand is needed to improve the management of these facilities and ultimately to prevent the obstruction of public space. Previous research was mainly focused on the parking of other vehicles, such as cars, introducing a lack of research and data related to micromobility parking.

Therefore, this research aimed to use historical counts of the number of parked micromobility vehicles along with neighborhood characteristics to analyze and predict the parking occupancy on the sidewalk. To achieve this goal, both supervised and unsupervised machine learning techniques were applied. Tree-based ensemble models proved to be suitable for predicting parking occupancy. In terms of predictive features, historical observations were the most influential predictor. The inclusion of the cluster results and neighborhood variables such as land use and the presence of points of interest further improved the predictions. Furthermore, clustering has made it possible to summarize multivariate information and to identify areas of similar characteristics.

KEYWORDS

Micromobility vehicles, parking occupancy, city accessibility, machine learning, mobility barriers, geographic information system

1 INTRODUCTION

Amsterdam is known as one of the most bicycle-friendly cities in the world. It is estimated that there are about 900,000 bicycles, compared to the population of 800,000 inhabitants [22]. The bicycle takes residents and visitors to any destination quickly and easily, whether it is going to work or university, taking the kids to school, running errands or just exploring the city. Cycling is environmentally friendly, contributes to a healthy lifestyle and can save time and money compared to other means of transport. By occupying relatively little space, the bicycle also contributes to an accessible and attractive city [21, 22].

However, there is also a downside to the frequent use of the bicycle. Due to the enormous number of cyclists, the public space is becoming increasingly crowded. Parking the bicycles, in particular, is becoming a serious challenge. In more and more parts of the city, the demand for parking spaces seems to exceed the supply [21, 22]. Due to the lack of available parking facilities and people's negligence, bicycles are often left on the sidewalk, chained to lamp posts or against bridge railings. This limits the space on the already narrow sidewalks and puts a lot of pressure on the pedestrian area. In contrast to the traffic space for other means of transport, the sidewalk fulfills several functions [23, 24]. Primarily, it is allocated as a walking space for pedestrians. In addition, the sidewalk provides space for terraces and shop displays, as well as for functional objects and street furniture. Incorrectly parked vehicles can block and limit the free passage space on the sidewalk. Although they are only temporary obstacles, they can pose a hazard to people with mobility problems and physical limitations, such as someone with a wheelchair, a walker or a pram [23, 24].

Micromobility vehicles include small-scale vehicles, such as bicycles, scooters, and skateboards, that are often used for short-distance trips [5]. A better understanding of micromobility parking and the possibility of predicting the demand is needed to improve the management of these facilities and ultimately to prevent the obstruction of public space by informally parked vehicles. With these insights, city officials can make informed decisions about implementing parking measures and, for example, optimize the placement of existing parking facilities or add more facilities in certain areas. Furthermore, parking forecasts can be integrated into route planning applications, providing pedestrians with optimized routes, so that they can avoid temporary obstacles or at least be aware of them.

While the parking behavior of other means of transport such as cars and micromobility sharing systems has received quite a lot of attention in the literature, only a limited number of studies examined the accessibility of the public space with regard to parked micromobility vehicles. The focus of these mostly observational and cross-sectional studies seems restricted to specific locations such as public transport stations [1, 27] or work and educational locations [30, 31]. Other, more generic locations in the city, such as shopping streets, or the primary parking location, i.e. residential areas, are hardly mentioned. In addition, most studies focused on historical observations and temporal information, not incorporating other external sources, such as geospatial attributes.

In the context of the above, the aim of this research was to use historical data on the number of micromobility vehicles parked on sidewalks, along with neighborhood characteristics, to analyze parking behavior and create models that can predict the number of vehicles parked on the sidewalk at a given time per sidewalk segment. The results can be used to analyze the accessibility of the city and identify potential bottlenecks.

In order to achieve the goal, the research is structured around the main research question:

To what extent can clustering and regression modeling be used for sparse spatio-temporal data on the parking behavior of micromobility vehicles to assess the accessibility of sidewalks?

To answer this research question, the supporting sub-questions are formulated:

(1) To what extent can clustering techniques be used to analyze the spatial variations in the parking occupancy of micromobility vehicles with regard to neighborhood characteristics?
(2) To what extent can regression modeling be used to predict the parking occupancy of micromobility vehicles on sidewalks?
(3) What is the impact of the neighborhood characteristics and cluster results on the performance of the regression models?

This paper consists of eight sections. In section 2, the work related to this study is discussed. Section 3 provides a general overview of the approach. The datasets used and the preprocessing steps are described in section 4. Sections 5 and 6 describe the aim, methods and results of clustering and regression modeling respectively. In section 7 the outcomes of this study are discussed based on the state of the art and the limitations are considered. Finally, the conclusion is drawn in section 8.

2 RELATED WORK

The following section presents the results of the previous literature. The section is divided into three parts. The first part takes a closer look at the application of cluster analysis to geospatial data, while the other two parts focus on the types of models and features used to predict parking occupancy.

2.1 Cluster analysis of spatial data

Cluster analysis has proven to be a valuable technique in spatial analysis and spatial data mining. The rationale of spatial clustering is that a set of spatial objects are grouped into meaningful groups so that the difference between objects in the same group is minimized [12]. Those clusters can be used to identify and locate areas with similar characteristics and to discover spatial patterns or hotspots. Those techniques have previously been applied in various domains, such as landscape ecology, customer segmentation or traffic.

An example is the research of Selvi and Caglar [25], which aimed to compare multivariate mapping of traffic accidents produced by k-means, k-medoids and hierarchical clustering. The results have shown that using clustering methods, spatial objects with similar properties can be identified. K-medoids delivered the best results in terms of cluster separation. In another study Li et al. [16] used k-means clustering to identify high-risk areas for the risk assessment of water pollution sources. Similarly, Xu et al. [29] verified the suitability of k-means clustering for spatial analysis to evaluate urban flood risk in the Haikou region of China.

Furthermore, cluster analysis has been applied in traffic-related applications, e.g. in traffic forecasting or for detecting anomalous traffic patterns. Studies by Sfyridis and Agnolucci [26] and Gecchele et al. [11] proposed a method for estimating annual average daily traffic (AADT) in England and Italy, respectively, combining clustering for the grouping of roads with regression modeling. While Sfyridis and Agnolucci [26] used the k-prototypes algorithm in combination with a comprehensive set of numerical and categorical variables, Gecchele et al. [11] focused on temporal patterns using different types of clustering algorithms. The findings of both studies indicate that data clustering can contribute significantly to the reduction of prediction errors.

2.2 Models used for parking demand prediction

Several state-of-the-art algorithms have been used in the literature to predict parking occupancy. Three methods are distinguished: model-based, parametric and non-parametric techniques [4, 18, 28].

With model-based techniques, the underlying parking behavior is explicitly modeled based on theoretical assumptions. An example is the study by Xiao et al. [28], which presents a Markov queueing model that describes the occupancy change of a car parking facility.

Parametric and non-parametric techniques mainly use large amounts of data to train models and generate predictions [4, 18, 28]. Parametric approaches rely on statistical models, such as exponential smoothing, while non-parametric models use techniques from the field of artificial intelligence, such as decision trees.

Several studies have proven the success of parametric models [9, 32]. For instance, Yu et al. [32] suggested an autoregressive integrated moving average (ARIMA) model to predict the available parking spaces in a parking lot of a mall in China. Similarly, Fokker et al. [9] compared several forecasting models, including the seasonal ARIMA and exponential smoothing models.

A large number of studies have employed machine learning techniques, including more complex models such as artificial neural networks (ANN) [3, 4] or ensembling algorithms such as random forest (RF) [4, 8, 27]. Dias et al. [8] compared two approaches, a RF and an ARIMA model, to predict occupancy trends for bicycle sharing stations in Barcelona. Here the RF has proven to be a better solution than the ARIMA model in terms of precision. Comparable results were delivered by a study of Balmer et al. [4] in which they predicted the on-street car parking occupancy using a RF and an ANN. The RF produced more accurate predictions than the ANN.

2.3 Features used for parking demand prediction

Existing research used different attributes to predict vehicle use and parking occupancy. This includes historical data, temporal information, weather conditions and geospatial attributes.

Historical parking occupancy seems to be the most crucial feature, as it has been utilized by all related studies to predict parking occupancy [2–4, 8, 9, 17]. This data is essential to time-series models such as ARIMA. Opposed to these models, most machine learning models cannot deal with consecutive time-series data. Here the data needs to be transformed, and features that represent and summarize past observations must be extracted.

Previous research found that the inclusion of temporal information such as time of the day, day of the week or holidays has a great influence on parking occupancy [2, 8, 9, 17]. Depending on the location of a parking space, the main time of usage will change. Li et al. [17] and Arjona et al. [2] demonstrated that the time of the day/week and calendar effects, such as holidays, need to be taken into consideration in parking occupancy patterns. In addition, artistic and sports events close to the parking facilities have proven to be an essential factor for parking forecasts [9].

Furthermore, the results of several studies pointed out the importance of weather features for predicting bicycle usage. For example, Dias et al. [8] observed the impact of extreme temperatures and rainfalls on the usage of bicycles. Similarly, Badii et al. [3] concluded that the inclusion of weather features could produce significantly better predictions of car parking demand.

Few studies have proven that geospatial information contributes positively to the accuracy of predictions. Yan and Zheng [31] successfully implemented models that linked bicycle parking demand of Shanghai's central business to the land use, expressed by the number of employment positions and (destination) attraction. Similarly, Balmer et al. [4] included land use (residential, office and industrial) and point of interest features in their forecasting models, which resulted in a performance increase of 25% compared to the baseline, which only used historical occupancy and time as an input for the same models. Other geospatial features suggested to be used as inputs for the models were the wealth of a neighborhood and the proximity to places such as public transportation, schools, and businesses [8, 9].

3 APPROACH

During this project, different methods were applied to achieve the stated goals. Figure 1 provides a schematic representation of the experiment pipeline. Before the datasets were merged, they were pre-processed. Then a cluster analysis was done with the goal to identify spatial variations in the parking occupancy of micromobility vehicles with regard to the neighborhood characteristics. Finally, several machine learning models were trained to predict the parking occupancy on the sidewalks using the historical parking occupancy, neighborhood characteristics and cluster results. The models were trained with different feature combinations to assess the influence of these features.

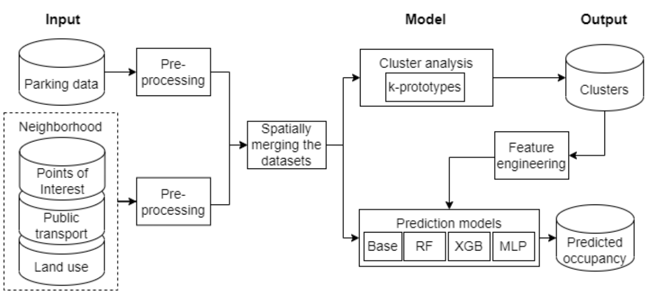

Figure 1: Schematic overview of the approach

3.1 Software

During this project, Python 3.7 was used in combination with a number of widely used data science and machine learning libraries. The pre-processing and exploratory data analysis were performed using NumPy (v1.22), pandas (v1.4.2) and GeoPandas (v0.11.0). For the clustering, the kmodes (v0.12.1) library was used. Various scikit-learn (v1.1) tools were applied during the training and evaluation of the regression models. Finally, to visualize the results of this research, matplotlib (v3.5.2), folium (v0.12.1) and seaborn (v0.11.2) were used. All code written during this research has been made available in a GitHub repository [1].

[1]Link to Github: https://github.com/Amsterdam-Internships/Micromobility-Parking

4 DATA DESCRIPTION

The city of Amsterdam has been chosen as the study area for this project. Therefore, the proposed methods were applied to sample data from that area. Several datasets regarding the parking behaviour and neighborhood characteristics were used.

The vehicle parking data was provided by Trajan [2], a traffic and mobility research bureau. In the years 2018 to 2021, they conducted a study in which both the parking capacity and the number of micromobility vehicles parked in the public space of Amsterdam were manually counted by field workers.

Each observation in the dataset contains the parking capacity and occupancy of a certain facility for a certain location at a certain time. A location can have one or more types of parking facilities. Specifically, the following variables are available: location id, street, year, part of the day, location of the parking facility (i.e. inside or outside the official facility), type of parking facility (e.g. no facility, staple or rack), capacity of the parking facility and the parking occupancy for different types of vehicles (e.g. bicycle, cargo bike or scooter). In addition, separate GeoJson files were provided that encode the geographic information, in this case polygons, of the observed locations.

The vehicle parking dataset contains parking counts from 2018, 2019, 2020 and 2021, with one count per location per year. While the counts in 2018, 2019 and 2021 took place in the afternoon, in 2021 the count was done in the evening. The number of locations observed varies from year to year. Most counts were carried out in 2018, with 32,510 locations. Both 2019 and 2021, contain approximately 15,600 locations, while 2020 has only about 13,000 locations.

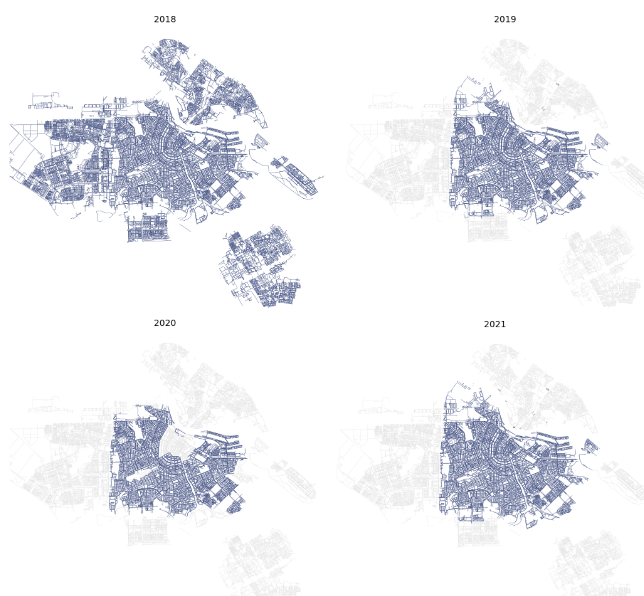

Figure 2: Locations in the vehicle parking count dataset observed in the years 2018, 2019, 2020, 2021

[2]Trajan. Bicycle study. Available from: https://www.trajan.nl/werkterreinen/fietsonderzoek.html

Table 1: Description of the neighborhood datasets used in this project

Dataset	Description	Length
Points of interest (function)	The location of different types of points of interest with non-residential functions (e.g. horeca, office, retail)	32,995
Points of interest (horeca)	The location of different types of points of interest with horeca functions (e.g. cafe, restaurant, coffeeshop)	4,171
Land use	The location and surface area for three land use categories (i.e. residential, services and work)	17,963
Public transport	The location and modality type of public transport stations.	231
Shopping area	The location and surface area of different types of shopping areas (e.g. core shopping area or neighborhood center)	132

Looking at the geographical distribution of the observations (see Figure 2), it appears that 2018 is the only year where counts were made outside the city ring (A10), including the districts Nieuw-West, Zuidoost and Noord. In 2020 observations are missing for part of the canal belt and on the outskirts of the city. This means that overall, only one observation is available for more than 50% of all locations, while 38% of the locations have been counted in all four years.

In addition, the dataset for this project was enhanced with several open-source data sources[3] containing neighborhood characteristics that potentially influence the parking occupancy. Table 1 provides a brief description of these additional datasets. The datasets contain different geometric features that represent the observed location (see Figure 3). While locations in some datasets (e.g. points of interest) are indicated by simple points, others have more complex shapes such as (multi-)polygons (e.g. parking data). Furthermore, some datasets contain information about the sidewalks (e.g. parking data), while others contain information about the surrounding buildings (e.g. land use) or streets (e.g. public transport).

4.1 Data preparation

Before the data was analyzed and used to build the prediction models, it was pre-processed. This included converting the data into a proper format, dealing with outliers, missing values, and extracting valuable features. The raw parking data was in the long format, which means that there were multiple occupancy counts for each location for the different parking facility types. In order to prepare the data for the analysis, the data was reshaped from the long to the wide format, aggregating the parking occupancy per location.

4.1.1 Missing values. In the first step, all datasets were inspected for missing values. In the neighborhood datasets, no observations with missing values were found. For the parking data, each location was required to contain a valid geometric feature to merge the parking data with the neighborhood features. 150 locations without this information were therefore removed. Furthermore, 79 locations in the 2019 data were found to have no parking counts at all. Those data points were completely removed because, without any parking data, they could not provide valuable information for the analysis As the deletion only affected about 0.5% of the locations, this had no negative impact on the quality of the dataset.

4.1.2 Data merging. In order to analyze the relation between the neighborbhood characteristics and the parking behavior, the datasets were merged, using the geospatial information of each location. Depending on the geometric object, the datasets were merged using different techniques. Data covering areas, such as buildings (e.g. land use), has been merged with the parking data based on the intersection of their geometric objects. Intersection means that the boundary or interior of the object intersects with that of the other in some way. If several observations have been assigned to one location, it has been decided to keep only the observation with the largest overlap. In the event that no intersecting observation was found, the closest observation was matched to the location in the parking data. Datasets, such as the points of interest and public transport, which were represented by points off the sidewalk, were merged with the parking data based on the distance between their geometric objects. The maximum distance within which the nearest geometry can be retrieved is set at 100 m. This distance has been chosen based on [21], which states that 100 m is the maximum distance people are willing to walk from the parking facility to their final destination. Some results contained multiple output records for a single input record. For this reason, a pivot table was created in which the neighborhood information was aggregated. The resulting attributes represent the number of locations per point of interest within 100 m of a sidewalk segment.

Figure 3: Geometric features of the datasets

4.1.3 Feature extraction. The next step involved extracting and creating relevant features to enhance and summarize the available information. First, categorical attributes were one-hot encoded so that numerical algorithms accept them. Any type of vehicle parked incorrectly on the sidewalk can create a barrier to pedestrians. For that reason, a feature was added to the dataset that summarized the total parking occupancy of all vehicle types for each location. The surface area of the sidewalk segments can differ per location. In order to be able to compare the values of different locations, attributes were created with the parking occupancy in relation

[3] City of Amsterdam. Open Maps Data Amsterdam. Available from: https://maps.amsterdam.nl/open_geodata/

to the surface area of a location. For the points of interest (POI) features, the frequency of occurrence of each function category was calculated by dividing the number of points of a category by the total number of points at a location.

Since ordinary machine learning algorithms are unable to deal with data in a time series format, autoregressive features were added that represent historically observed values: lag features that shift a variable by a defined time period and rolling window features that summarize a number of historical values. The choice of the number of lags and the size of the window was limited because only a maximum of four historical values were available in the data. For the window feature, the values of all previous years were averaged for each year. In addition, lag features have been created using the parking values of the past one to three years.

4.1.4 Outliers. Finally, the data was inspected for possible outliers. Almost all numerical characteristics related to parking capacity and occupancy were positively skewed, including an excessive amount of zeros. The maximum value for the number of vehicles parked outside the parking facility was 3000 per 100 m^2. Assuming that a standard bicycle occupies approximately 1.2 m^2 of area [22], any value above 85 vehicles per 100 m^2 was assumed to be an outlier and therefore removed.

4.1.5 Final dataset. After pre-processing, the dataset contained 43,906 observations. The target variable was identified as the number of vehicles parked outside the official parking facilities on the sidewalk per 100 m^2. Concerning the parking data, the dataset contained the capacity and four variables aggregating the historical parking behavior. In addition, there were 16 features regarding the points of interest with non-residential functions and 12 features regarding the points of interest with horeca functions. Lastly, the dataset included three features describing the land use, one binary variable for the shopping area and two features for the public transport stations.

5 CLUSTERING SPATIAL DATA

Previous studies have shown the success of different clustering techniques for geospatial data and the traffic domain [11, 25, 26, 29]. Therefore, in the first part of this project, cluster analysis was applied with the goal of grouping similar sidewalk segments based on their neighborhood characteristics and parking occupancy. These clusters were used to analyze spatial variations in the parking behavior and to identify problematic locations along with their neighborhood properties.

5.1 K-prototypes

In this study, the partition algorithm k-prototypes is applied based on its success in related work [26]. Partition clustering algorithms start with an initial partitioning and then iteratively relocate objects between clusters based on some criteria until an optimal (local) partition is reached. K-means and k-medoids are the two simplest and best-known partition algorithms. A disadvantage of those methods is that they only accept one type of data, numerical or categorical, respectively. To overcome this problem, Huang [14] proposed k-prototypes, which integrates distance metrics for numerical data and dissimilarity measures for categorical data to enable clustering objects of mixed data types.

5.2 Experimental setup

5.2.1 Validation. Like other partition algorithms, k-prototypes requires specifying k, the number of clusters to be generated. To find the most optimal value for k, several data-driven approaches are available [15]. Using categorical variables, k-prototypes does not allow calculation of the distance-based metrics such as the Silhouette coefficient. Therefore, it was decided to implement the Elbow method.

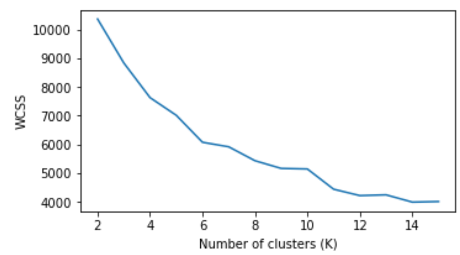

Figure 4: Selecting k number of clusters using the Elbow method

For the Elbow method, the within-cluster sum-of-squares (WCSS), or the sum of squared distances of all the samples and their cluster centroid, was used. This value was calculated and plotted for 2 to 15 numbers of clusters. The optimal number of clusters is represented by the point where the slope tends to decrease. Based on Figure 4, it was concluded that the optimal number of clusters for this study was six.

5.2.2 Normalisation. The numerical attributes in the data have different scales. With distance-based algorithms such as k-prototypes, large-scale features are given a higher weight, leading to dominance in the clustering. To avoid this, standardizing the numerical features is recommended. Since several variables, such as land use and frequency of POI occurrence, were already in the range of 0 to 1, it was decided to apply min-max normalization to comprise the rest of the features to the same scale.

5.3 Results

The following paragraphs present the results of the cluster analysis. A qualitative analysis was carried out to study the differences among clusters and to investigate potential insights that the clusters can provide about the relation between the parking behavior and neighborhood characteristics. This analysis involved visualizing the geospatial distribution of the resulting clusters by projecting them onto a map of Amsterdam (see Figure 5) and using descriptive statistics for the most relevant characteristics (see Table 2). Specifically, the mean for numerical and the frequency for categorical variables was calculated per cluster, and non-parametric tests were performed to test whether the difference between clusters was significant.

In cluster 0 (pink), the most common venues are offices (0.85). The land use in this cluster is mainly residential (0.8) and a smaller

Figure 5: Cluster results plotted on the map of Amsterdam

part is work (0.13). Figure 5 shows that the included locations are spread over the city, with accumulations in the canal belt in the Centrum and close to the Vondelpark in Zuid. The parking pressure in this cluster is high (occupancy of 7.3 vehicles per 100 m^2).

The second-largest cluster with 3,373 observations is cluster 1 (orange). In this cluster, there is a relatively high frequency of education (0.11) and high land use of services (0.78). Locations in this cluster are again spread throughout the entire city. Some interesting locations in this cluster are the University of Amsterdam at Science Park and the Vrije Universiteit in Zuid. The parking pressure on these sidewalks is medium (occupancy of 5.6 vehicles per 100 m^2).

Cluster 2 (red), the smallest cluster (n = 1,276), contains a high frequency of horeca venues (0.6), especially cafes and restaurants, and a medium frequency of retail (0.22). In addition, the proportion of shopping streets (0.53) and tram stops (0.27) in this cluster is higher than in other clusters. The observations are mainly located in the city center and throughout the city along the city streets in the districts Centrum, Pijp (Zuid) and Baarsjes (West). The parking pressure in this cluster is the highest (occupancy of 8.6 vehicles per 100 m^2).

With about 21,300 observations, cluster 3 (light blue) is the most common cluster in Amsterdam. The map shows that this cluster covers large parts of the city, except the city center. Sidewalk segments in this cluster are mainly characterized by a very high land use of housing (0.97) and a low to medium parking pressure (occupancy of 4.03 vehicles per 100 m^2).

The locations on the outer parts of the city, in neighborhoods like Sloterdijk, Zuidas, Westerpark, Westpoort and Nieuw West, mainly belong to cluster 4 (green). This cluster is characterized by a very high land use of work (0.91). There are also quite a few offices and companies, with a frequency of 0.27 and 0.12 respectively. With

Table 2: Mean values of the main features of the cluster results

	Cluster					
	0	1	2	3	4	5
Location count	1,474	3,373	1,276	21,370	2,875	1,907
Offices freq.	0.85	0.02	0.07	0	0.27	0.03
Retail freq.	0.04	0.02	0.22	0	0.03	0.88
Horeca freq.	0.02	0.01	0.6	0	0.01	0.05
Companies freq.	0.01	0.01	0.01	0.01	0.12	0
Education freq.	0.01	0.11	0.01	0.01	0	0.01
Residential land use	0.8	0.13	0.57	0.97	0.03	0.64
Work land use	0.13	0.09	0.05	0.01	0.91	0.03
Services land use	0.07	0.78	0.38	0.02	0.06	0.32
Shopping area	0.14	0.15	0.53	0.05	0.1	0.55
Tram station	0.17	0.14	0.27	0.08	0.08	0.29
Occupancy in per 100m^2	3.43	2.83	3.76	1.95	1.78	3.63
Occupancy out per 100m^2	3.89	2.78	4.85	2.08	1.78	3.81
Occupancy per 100m^2	7.32	5.6	8.61	4.03	3.56	7.44
Capacity per 100m^2	4.53	4.6	4.75	2.25	2.66	4.56

a total occupancy of 3.56 vehicles per 100 m^2, this cluster's parking pressure is the lowest.

The last cluster, 5 (dark blue), has a very high frequency of retail (0.88). As in cluster 2, the share of shopping streets and tram stops is also higher here than in other clusters (0.55 and 0.29 respectively). Looking at Figure 5, it can be seen that the observations in this cluster are located on the main city streets, which connect different neighborhoods, such as Centrum, Baarsjes and Oud Zuid. The parking pressure is comparably high as in cluster 0.

Finally, the non-parametric Kruskal-Wallis test was used to determine whether the median of included variables in the six clusters differed significantly. For all features displayed in Table 2, the p-value was smaller than the significance level (a = 0.05). It was therefore concluded that at least one cluster for each variable significantly differs from the other clusters.

Overall, clusters with higher parking occupancy are mainly characterized by high frequencies of retail and horeca and a certain amount of residential buildings. The combination of these neighborhood characteristics attracts two types of parking behaviour: short-term parking for shopping and recreational purposes and long-term parking for residents. Especially for a short visit to a shop, people tend to park their vehicles outside a rack on the sidewalk. Moreover, the parking space for the residents of these central locations is scarce, so vehicles are often parked in the public space. Low parking is associated with a high land use of work and companies, mainly in the outer neighborhoods. This can be explained by a much lower density of points of interest in these areas, which attracts fewer people and therefore fewer parked vehicles. Companies and offices might also offer more private parking facilities.

6 PREDICTING PARKING OCCUPANCY

In the second part of this study, different regression models were applied to predict parking occupancy. In addition to historical observations, neighborhood characteristics and cluster results were used as input features in the models, and their influence on the predictions of parking occupancy was investigated.

6.1 Regression models

Based on the previous literature, several parametric and non-parametric approaches have demonstrated to provide accurate predictions regarding parking occupancy. Most time-series forecasting models, such as ARIMA, require a minimum of 50 observations for accurate estimation [13, 19]. Since the dataset in this study contained a maximum of four observations per location, it was not possible to use these methods. Therefore, it was decided to focus on non-parametric machine learning methods in this study.

A number of state-of-the-art machine learning algorithms were selected, which have proven to perform well for parking occupancy forecasting: random forest (RF), XGBoost and multilayer perceptron (MLP). Consequently, a baseline was defined based on the average parking behavior of a location over the past three years. For reasonable predictive power, the trained models should yield better results than the baseline.

6.1.1 Random forest. Random forest is an ensemble method that consists of a collection of individual decision trees [6]. The model is formed by generating a large number of decision trees (DT), each using a different bootstrap sample of the features in the data. For the final prediction, the RF takes the average output of all individual DTs. An advantage of random forests is that the method is able to model complex non-linear processes and is well suited for high-dimensional data sets with various data types.

6.1.2 XGBoost. XGBoost, short for Extreme Gradient Boosting, is another ensemble tree method that implements the gradient boosting framework [7]. Opposed to a RF, XGBoost sequentially creates new DTs that predict the residuals of its predecessor. These are then added together to make the final prediction. To minimize the loss when adding new DTs, the algorithm uses gradient descent. Similarly to RF, XGBoost can efficiently handle large, sparse datasets with mixed data and missing values.

6.1.3 Multilayer perceptron. The last model, a multi-layer perceptron, is a feed-forward artificial neural network (ANN) that consists of an input layer, an output layer and an arbitrary number of hidden layers [10]. During training, the model implements back-propagation to find suitable parameters and minimize the error. An advantage of the multi-layer perceptron is that it works well with large datasets and can model highly nonlinear functions.

6.2 Features included in the models

Several studies have shown that parking occupancy predictions can be improved when different data sources are added to the historical parking data [2–4, 8, 9, 17]. The data in this study does not contain a clear indication of time, and, in addition, only a maximum of four historical observations are available. It was therefore not possible to use temporal information or time-bound data such as the weather or events in the prediction models. However, the large number of locations covered in the data provided an excellent opportunity to investigate the influence of geospatial factors on the forecasts.

Furthermore, based on the cluster results, new features were constructed. The labels of the clusters were directly used as a categorical feature. In addition, the mean historical parking occupancy was calculated per cluster.

To investigate the influence of the features, the models were trained using different combinations of variables. In total, five models were trained, using only the historical counts, the historical counts together with the neighborhood and/or cluster characteristics and only the neighborhood and cluster characteristics.

6.3 Experimental setup

6.3.1 Hyperparameter optimization. In order to increase the performance of the regression models, hyperparameter optimization was applied to find the most optimal combination of parameters. For each model, different sets of hyperparameters were specified and evaluated using random search. Random search was chosen over grid search because it is computationally lighter and faster. During the search, 500 different combinations were randomly searched within a given search space. The search spaces and the selected parameters can be found in the GitHub repository [4].

6.3.2 Feature selection. During feature selection, a subset of the most relevant features was created to improve the models' performance and reduce noise and computational costs. Tree-based ensemble methods, such as RF or XGBoost, offer the possibility to give the impurity-based importance of the included explanatory variables. These insights were exploited to investigate the contribution of the attributes to the performance of the models and to select the most relevant subset of features.

6.3.3 Data splitting. The data was split into three subsets to train, validate, and test the models. Because the data in this research contains a temporal component, it was required to maintain the chronological order during the split to avoid look-ahead bias. Due to the scarcity of the data related to its temporal component, only one split was possible, using 2019 for training (n = 15,462), 2020 for validation (n = 12,880) and 2021 for testing the models (n = 15,564).

6.3.4 Validation. Metrics commonly used in research to evaluate the performance of the models are the mean average error (MAE), the mean squared error (MSE) and the root mean squared error (RMSE). MAE represents the average of the absolute residuals in the dataset, while the MSE is calculated by taking the average of the squared residuals. As the name implies, the RMSE is the square root of the MSE. The MAE and RMSE were preferred because they have the same units as the dependent variable, making them easier to interpret. Furthermore, the coefficient of determination, also known as R squared (R^2), was used for the validation. R^2 indicates how well the predictive variables in the model explain the variability in the response variable.

[4] Link to Github: https://github.com/Amsterdam-Internships/Micromobility-Parking

6.4 Results

The following section describes the results of the regression models used to predict the parking occupancy of micromobility vehicles on the sidewalk.

6.4.1 Model comparison. Table 3 shows the performance statistics, in terms of MAE, RMSE and R^2, for each model and feature set combination using the test set. When comparing all models, it can be seen that they have a relatively similar performance. In general, the tree-based ensemble models XGBoost and RF outperformed the MLP regressor for all three performance metrics. The MAE values of the RF and XGBoost were almost the same for all feature combinations, but with regard to RMSE and R^2, the RF outperformed XGBoost. In addition, the RF delivered better predictions using only the neighborhood features and cluster results. Looking at the RMSE and R^2, all models outperformed the baseline, while the difference for the MAE was not much distinctive. The performance of the fitted MLP models was very similar for nearly all feature combinations. Furthermore, the MAE of the XGBoost and MLP model, only including the neighborhood and cluster features, was worse than that of the baseline.

Based on the performance metrics, the random forest regressor using historical, neighborhood and cluster features was identified as the best model. The RF could explain 58% of variation in the predicted parking occupancy, and, on average, delivered a 2.20 difference between the predicted and observed parking occupancy in terms of vehicles parked per 100 m^2 sidewalk. The higher RMSE value of 4.58 indicates that there was some variation in the magnitude of the residuals. Overall, the residuals were reasonably well distributed, with a mean error of 0.001, implying that the model has not greatly over- or underestimated parking occupancy.

Table 3: Error metrics on the test set for each combination of feature set and model

Model	Features			MAE	RMSE	R^2
	Hist.	Neigh.	Clus.			
Baseline	X			2.28	7.23	-0.05
XGBoost	X			2.28	4.8	0.54
	X		X	2.25	4.75	0.55
	X	X		2.23	4.73	0.55
	X	X	X	2.21	4.7	0.56
		X	X	2.39	5.18	0.47
Random forest	X			2.28	4.71	0.55
	X		X	2.24	4.71	0.56
	X	X		2.23	4.63	0.57
	X	X	X	**2.20**	**4.58**	**0.58**
		X	X	**2.28**	**4.71**	**0.56**
Multilayer perceptron	X			2.28	4.76	0.54
	X		X	2.26	4.86	0.53
	X	X		2.30	4.85	0.53
	X	X	X	2.29	4.85	0.53
		X	X	3.33	5.99	0.28

Figure 6 provides more insight into the performance of the random forest models for both the train and the test set. It can be seen that the MAE is almost twice as high for the test set when the historical observations are included in the model. This indicates that the models are probably overfitting. When the exact historical data is excluded, the overfitting is reduced. This can be seen from the MAE values of the rightmost model, which are much closer. Thus, by including the exact historical parking count as a predictor, the model seems unable to generalize on the test data anymore. This overfitting problem was also found for the other regression models and using the other error statistics.

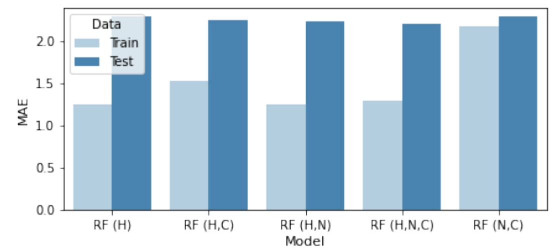

Figure 6: MAE of the random forest models on the train and test set using historical (H), neighborhood (N) and/or cluster (C) features

6.4.2 Spatial prediction variation. An assessment of the spatial variability of the prediction performance was carried out by comparing the model's performance across all sidewalk segments. Figure 7 on the next page shows the number of parked vehicles on the sidewalk predicted by the RF regressor together with the actual values and the residuals for each sidewalk segment.

Overall, the two maps with the predicted and actual parking occupancy show clear similarities in the distribution of areas with high and low parking occupancy, with higher values in the center and lower values in the outer neighborhoods. However, especially in the canal belt, the model underestimated the parking counts. This is also reflected by the red sidewalk segments in the residuals map. For residential areas outside the center, the opposite is the case. Here the predictions are slightly higher than the actual values. Furthermore, it is noticeable that in areas with higher parking counts, the magnitude of the residuals increases to a certain extent, which can be seen by more saturated areas in the residuals map.

6.4.3 Feature importances. Finally, the importance of the different features was investigated. Looking at Table 3, the performance of the tree-based models was slightly improved by adding neighborhood and cluster variables (4% for the RF). This indicates that historical features were the most predictive factor for parking occupancy. In fact, for the MLP, the inclusion of additional features led to a decrease in the model's performance. Excluding the historical parking counts resulted in poorer performance for both XGBoost and MLP. However, the RF achieved almost the same performance as using only the historical observations as a predictor.

As mentioned before, tree-based models such as the RF can provide feature importances. Figure 8 illustrates the importance of the variables included in the RF model using all features. The historical parking occupancy, both the mean and lag feature, were identified as the most important predictors. Their importance scores are three times higher than the values of the other input features.

Figure 7: Predictions of the random forest regressor (left), actual values (middle) and residuals (right)

Figure 9 provides the feature importances of the RF model including the neighborhood and cluster features. Here the most crucial feature is the surface area of the sidewalk. Regarding the neighborhood and cluster characteristics, the lists of the most important characteristics are almost identical for both models. All three land use features, residential, work and services, have proven to be of some importance for the prediction of parking occupancy. A number of points of interest features, namely offices, horeca and retail are also among the most influential features. Other important neighborhood features are the tram stations and shopping streets. Finally, the cluster variables, i.e. the cluster category itself and the parking occupancy averaged per cluster, have contributed valuable information to the models.

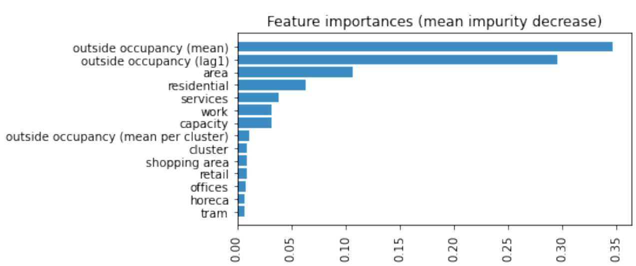

Figure 8: Feature importances of the random forest using historical, neighborhood and cluster features

The importance scores of the neighborhood and cluster characteristics are higher in the model that included only these characteristics. When including the exact historical parking count, it seems that the model gave much more value to this attribute, limiting the influence of other features. This is consistent with the findings from Table 3. Here, adding the neighborhood and cluster features only resulted in a 4% improvement of the MAE, while the model with only these features performed just as well as the model with only the exact historical counts. This again supports the assumption that the models with the exact historical parking counts were overfitting.

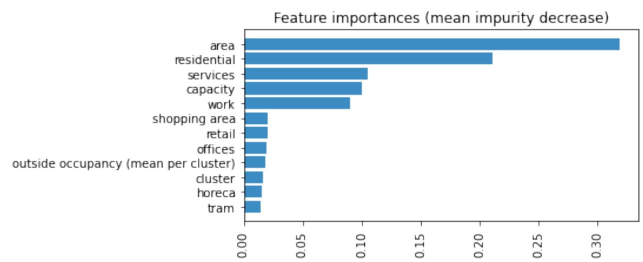

Figure 9: Feature importances of the random forest using neighborhood and cluster features

7 DISCUSSION

This section discusses the observed results, links them to previous research, and attempts to answer the proposed research questions.

7.1 Research questions

7.1.1 To what extent can clustering techniques be used to analyze the spatial variations in the parking occupancy of micromobility vehicles with regard to neighborhood characteristics? The cluster analysis aimed to divide sidewalk segments into groups based on their properties, maximizing both the similarity of features within clusters and the dissimilarity between clusters. Clustering, specifically k-prototypes, has made it possible to summarize and visualize multivariate information by dividing sidewalk segments into meaningful clusters. Based on those clusters, spatial patterns and areas with comparable properties could be identified during the exploratory phase of this research. In addition, the clusters helped to understand the relationships between certain neighborhood characteristics and the parking behavior of micromobility vehicles. With the help of domain knowledge, these insights could be interpreted and explained. The findings of the cluster analysis can directly be applied for urban planning to improve the management of parking facilities, for example by adding extra parking racks in busier areas. However, the clusters could also be further exploited by various systems such as forecasting models. In addition, it may be interesting to apply this approach to other traffic-related applications, such as car parking or pedestrian crowds.

Similar techniques have been widely used in many fields as mentioned earlier [11, 25, 26, 29], but used scarcely in the analysis of locations and their characteristics in relation to micromobility parking. Our findings are in line with the conclusions of previous research in that they also found that clustering has added value for identifying locations with similar properties.

7.1.2 To what extent can regression modeling be used to predict the parking occupancy of micromobility vehicles? In this study, three machine learning methods, namely random forest, XGBoost and multilayer perceptron, were used to predict the parking occupancy of micromobility vehicles. Both tree-based models slightly outperformed the MLP for all feature combinations. This might be explained by their relative simplicity and their robustness to overfitting. It is also possible that random search has not found the most optimal combination of hyperparameters. The difference between the performance of the tree-based models was insignificant, with the RF slightly outperforming the XGBoost. Based on those results, it is suggested to further investigate the applicability of tree-based models for parking occupancy prediction in future research.

In previous research, there is no consensus about which type of model is more appropriate for predicting parking occupancy. Both, tree-based ensemble and ANN models, have proven to deliver promising results [3, 4, 8, 27]. In addition, several studies have used parametric approaches such as ARIMA to make reliable predictions. [9, 32].

Overall, the findings of the regression modeling gave a general idea of how machine learning models can be used in predicting parking occupancy of micromobility vehicles. However, due to the small number of historical observations, the possibilities with regard to the design of the forecasting models were very limited. Considering that the models only provide one-year ahead forecasts, the added value of these forecasts for managing micromobility parking is rather small. Applications, such as planning accessible routes, may require shorter-term forecasts.

7.1.3 What is the impact of the neighborhood characteristics and cluster results on the performance of the regression models? To investigate the importance of neighborhood information and cluster results, the regression models in this study were trained using different sets of features. In addition, importance scores of the tree-based models were used to confirm conclusions about the influence of those features.

By including both geospatial and cluster features in the tree-based models, the performance of the models improved compared to using historical data alone. Although the improvement was only about 4%, the characteristics have proven to be of some value in predicting parking occupancy. This could be confirmed by the comparable performance delivered by the model only using the historical parking counts and the model using the neighborhood and cluster features. In general, the historical observations of a location were found to be an essential factor when predicting parking occupancy. However, including the exact historical counts led to overfitting of the models. Removing this feature and only using the historical data summarized in the cluster results has decreased overfitting.

The same conclusions were drawn when comparing the feature importance scores of the RF models. For the model including all features, the historical parking occupancy scored highest, while neighborhood and cluster variables were less than half as important. The importance scores of the features in the model without the exact historical counts were generally higher. The most promising characteristics were the land use variables and points of interest such as retail and horeca. Those findings are consistent with the results of the clustering part, where the same variables showed a high discriminating value between the clusters.

The results of this study are largely in line with previous research, where adding additional information also led to a performance increase [4, 9, 31]. Those studies identified similar variables, such as land use of offices [4] and public transport stations [9], as important features. However, the performance improvement in other studies was significantly greater than in this study. This may be explained by the fact that the exact historical parking counts have made the models susceptible to overfitting. Furthermore, a number of the related studies were focused on the parking behavior of another type of vehicle, such as cars, in which these factors may play a greater role.

7.2 Limitations

One major limitation of this study was the scarcity of the data related to its temporal component. Only a small amount of historical data was available for each location, with a maximum of four observations. As stated before, parking of (micromobility) vehicles is a spatial-temporal issue, indicating that parking behavior changes over time depending on the geospatial environment. Previous research has shown that time and time-related features, such as weather and holidays, have a great value in predicting parking occupancy [2, 3, 8, 9, 17]. With the dataset provided in this study, it was not possible to exploit and analyze the influence of these key factors. Consequently, this has limited the possibilities with regard to the design and outcomes of this study.

To address this limitation and provide reliable results, more historical data is needed. This data must cover at least the different times of the day (morning, afternoon and evening), as well as the different days of the week (weekday vs weekend) and months in a year. Based on this information, different time-related factors can be included in the analysis. In addition, more historical data offers the possibility to use other forecasting models, such as ARIMA, that require a minimum of 50 observations [13, 19].

Given the large number of locations in this dataset, it seems unfeasible to perform the counts for the entire city. To tackle this problem, it is possible to count only certain locations that represent a group of similar locations in terms of their geospatial properties. For this, use can be made of the results obtained during the clustering analysis. Another option for gathering the necessary data is to use photos of sidewalk segments. Previous research has demonstrated that those photos can be used to capture parking data, including arrival and departure times and parking durations [20].

8 CONCLUSION

Parking micromobility vehicles is becoming an increasing problem in urban areas such as Amsterdam. A better understanding of micromobility parking and the possibility to predict the demand is needed to improve the management of these facilities and ultimately prevent obstruction of the public space by informally parked vehicles. Previous research mainly focused on the parking of other means of transportation, such as cars or vehicle sharing systems, introducing a lack of research and data related to micromobility parking. In addition, most studies used historical data and calendar effects only, ignoring the impact of other features, such as neighborhood characteristics.

Therefore, this research aimed to use historical data on parked micromobility vehicles to analyze and predict the parking occupancy on the sidewalk. The parking data was enhanced with different neighborhood properties and the findings of the cluster analysis. Results showed that tree-based ensemble models such as random forest and XGBoost are suitable for predicting parking occupancy. In terms of predictive features, historical observations have been found to be the most influential predictor. However, using the exact historical counts led to overfitting. Including cluster results and geospatial variables, such as land use and the presence of points of interest, has further improved the predictions. Moreover, clustering allowed summarizing multivariate information and identifying spatial patterns in parking occupancy related to certain neighborhood characteristics.

Despite some limitations, the present study has demonstrated the potential of predictive modeling for micromobility parking and the importance of considering geospatial features. Insights from this research can be used by city officials to tackle bottlenecks in the city and thus increase the accessibility of the public space.

Future research into predicting parking occupancy of micromobility vehicles should focus on establishing a data collection framework tailored to the goals of the desired application. With the correct data, features such as calendar effects or weather conditions can be included in the models. Those models may be able to deliver more reliable predictions.

REFERENCES

[1] David Arbis, Taha Hossein Rashidi, Vinayak V Dixit, and Upali Vandebona. 2016. Analysis and planning of bicycle parking for public transport stations. *International journal of sustainable transportation* 10, 6 (2016), 495–504.

[2] Jamie Arjona, MªPaz Linares, Josep Casanovas-Garcia, and Juan José Vázquez. 2020. Improving parking availability information using deep learning techniques. *Transportation Research Procedia* 47 (2020), 385–392.

[3] Claudio Badii, Paolo Nesi, and Irene Paoli. 2018. Predicting available parking slots on critical and regular services by exploiting a range of open data. *IEEE Access* 6 (2018), 44059–44071.

[4] Michael Balmer, Robert Weibel, and Haosheng Huang. 2021. Value of incorporating geospatial information into the prediction of on-street parking occupancy–A case study. *Geo-spatial Information Science* 24, 3 (2021), 438–457.

[5] Cynthia Bennett, Emily Ackerman, Bonnie Fan, Jeffrey Bigham, Patrick Carrington, and Sarah Fox. 2021. Accessibility and The Crowded Sidewalk: Micromobility's Impact on Public Space. In *Designing Interactive Systems Conference 2021*. 365–380.

[6] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[7] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[8] Gabriel Martins Dias, Boris Bellalta, and Simon Oechsner. 2015. Predicting occupancy trends in Barcelona's bicycle service stations using open data. In *2015 sai intelligent systems conference (intellisys)*. IEEE, 439–445.

[9] Elisabeth S Fokker, Thomas Koch, Marco van Leeuwen, and Elenna R Dugundji. 2022. Short-term forecasting of off-street parking occupancy. *Transportation Research Record* 2676, 1 (2022), 637–654.

[10] Matt W Gardner and SR Dorling. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment* 32, 14-15 (1998), 2627–2636.

[11] Gregorio Gecchele, Riccardo Rossi, Massimiliano Gastaldi, and Armando Caprini. 2011. Data mining methods for traffic monitoring data analysis: A case study. *Procedia-Social and Behavioral Sciences* 20 (2011), 455–464.

[12] Jiawei Han, Jae-Gil Lee, and Micheline Kamber. 2009. An overview of clustering methods in geographic data analysis. *Geographic data mining and knowledge discovery* 2 (2009), 149–170.

[13] Martin Hecht and Steffen Zitzmann. 2021. Sample size recommendations for continuous-time models: Compensating shorter time series with larger numbers of persons and vice versa. *Structural Equation Modeling: A Multidisciplinary Journal* 28, 2 (2021), 229–236.

[14] Zhexue Huang. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery* 2, 3 (1998), 283–304.

[15] Leonard Kaufman and Peter J Rousseeuw. 1990. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics.

[16] Chunhui Li, Lian Sun, Junxiang Jia, Yanpeng Cai, and Xuan Wang. 2016. Risk assessment of water pollution sources based on an integrated k-means clustering and set pair analysis method in the region of Shiyan, China. *Science of the total environment* 557 (2016), 307–316.

[17] Jiachang Li, Jiming Li, and Haitao Zhang. 2018. Deep learning based parking prediction on cloud platform. In *2018 4th International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, 132–137.

[18] Zhenyu Mei, Wei Zhang, Lihui Zhang, and Dianhai Wang. 2020. Real-time multistep prediction of public parking spaces based on Fourier transform–least squares support vector regression. *Journal of Intelligent Transportation Systems* 24, 1 (2020), 68–80.

[19] Errol E Meidinger. 1980. *Applied time series analysis for the social sciences*. Sage Publications.

[20] David A Moskovitz and Nikki Wheeler. 2011. Bicycle parking analysis with time series photography. *Transportation research record* 2247, 1 (2011), 64–71.

[21] City of Amsterdam. 2015. Kader Fietsparkeren. Available from: https://assets.amsterdam.nl/publish/pages/865242/kader_fietsparkeren.pdf.

[22] City of Amsterdam. 2017. Meerjarenplan Fiets 2017-2022: Voor fietsers en een gezonde en bereikbare stad. Available from: https://assets.amsterdam.nl/publish/pages/865242/meerjarenplan_fiets_2017-2022.pdf.

[23] City of Amsterdam. 2020. Beleidskader ruimte voor de voetganger. Available from: https://assets.amsterdam.nl/publish/pages/865282/inspraakversie_beleidskaderruimtevoorvoetganger-toegankelijk.pdf.

[24] City of Amsterdam. 2020. Voortgangsrapportage Toegankelijkheid 2020: Iedereen doet mee! Available from: https://assets.amsterdam.nl/publish/pages/887472/voortgangsrapportage_toegankelijkheid_2020.pdf.

[25] Huseyin Zahit Selvi and Burak Caglar. 2018. Using cluster analysis methods for multivariate mapping of traffic accidents. *Open Geosciences* 10, 1 (2018), 772–781.

[26] Alexandros Sfyridis and Paolo Agnolucci. 2020. Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling. *Journal of Transport Geography* 83 (2020), 102658.

[27] Casper Antoon Slenders. 2018. Predicting bicycle facility usage through check-in and check-out data.

[28] Jun Xiao, Yingyan Lou, and Joshua Frisby. 2018. How likely am I to find parking?–A practical model-based framework for predicting parking availability. *Transportation Research Part B: Methodological* 112 (2018), 19–39.

[29] Hongshi Xu, Chao Ma, Jijian Lian, Kui Xu, and Evance Chaima. 2018. Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China. *Journal of hydrology* 563 (2018), 975–986.

[30] Junbin Xu, Zhiyong Zhang, and Jian Rong. 2012. The forecasting model of bicycle parking demand on campus teaching and office district. *Procedia-Social and Behavioral Sciences* 43 (2012), 550–557.

[31] Kefei Yan and Jilong Zheng. 1994. Study of bicycle parking in central business district of Shanghai. *Transportation Research Record* 1441 (1994), 27.

[32] Fengquan Yu, Jianhua Guo, Xiaobo Zhu, and Guogang Shi. 2015. Real time prediction of unoccupied parking space using time series model. In *2015 International conference on transportation information and safety (ICTIS)*. IEEE, 370–374.