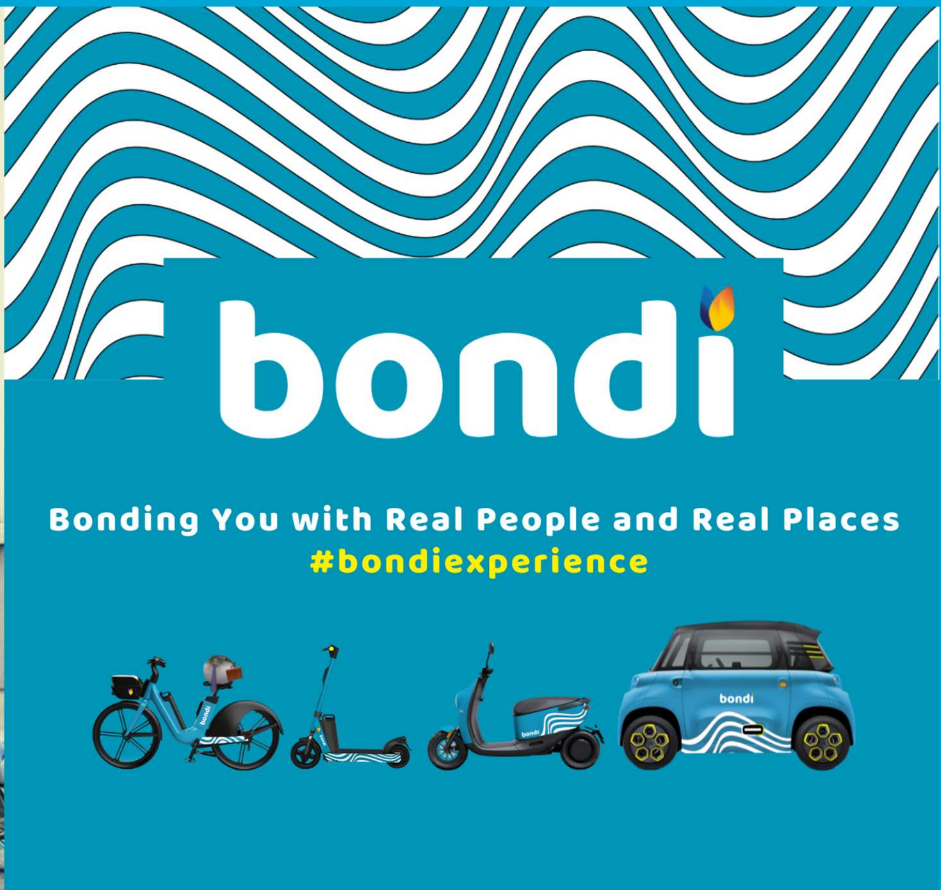
MSc Thesis in Transport, Infrastructure & Logistics

# Improving the Service of E-bike Sharing by Demand Pattern Analysis
# A Data-driven Approach

Ziru Zhang


TUDelft bondi

# Improving the Service of E-bike Sharing by Demand Pattern Analysis

## A Data-Driven Approach

by

Ziru Zhang

in partial fulfilment of the requirements for the degree of

**Master of Science**

in Transport, Infrastructure & Logistics

at the Delft University of Technology,

to be defended publicly on Monday, November 29, 2021, at 15:00.

| | |
|---|---|
| Student number: | 5061199 |
| Project duration: | May 2021 – November 2021 |
| Thesis committee: | Prof.dr.ir. S.P. Hoogendoorn, TU Delft, chair |
| | Dr. ir. N. van Oort, TU Delft, supervisor |
| | Dr. ir. P. K. Krishnakumari, TU Delft, supervisor |
| | Dr. F. Schulte, TU Delft, supervisor |
| | Chingiskhan Kazakhstan, bondi, supervisor |
| | Max Schalow, bondi, supervisor |

# SUMMARY


**Background and motivations**

Shared mobility, as sustainable transport modes, reducing the emission as well as the traffic congestions, has been increasingly popular since 2010 (European Commission, 2011). From all forms of shared mobility, shared bikes, with the additional benefits of maintaining wellbeing, have expanded rapidly around the world and e-bikes have also been gradually incorporated into these schemes with faster speed and fewer physical efforts required (Barbour et al., 2019; DeMaio, 2009). There are two main types of shared bike projects, station-based and free-floating bike sharing, where station-based ones require to pick up and return bikes at the given stations while the free-floating allows the return anywhere within the operational zone.

No matter what types of bike sharing, it is always essential to match the supply with the demand, and thereby there is elusive literature, targeting this problem in different aspects. There are mainly 5 topics related to the research on bike sharing, applicable to e-bike sharing as well: 1) studying the determinants of bike-sharing demand (Daddio, 2012); 2) statistical analysis of bike sharing projects (Noussan et al., 2019); 3) demand pattern analysis of bike sharing (C. Xu et al., 2018); 4) demand prediction of bike sharing schemes (Bao et al., 2019a); and 5) optimization of bike sharing operations (Caggiani et al., 2018). Despite a large number of existing studies, most of them targeted station-based traditional bike sharing projects and free-floating e-bike sharing was less studied. In addition to that, the widely-applied spatial analytical units for the studies targeting free-floating shared bikes are either administrative zones, such as districts or TAZ (traffic analysis zone), or equally-distributed squares, in which zones with a size of approx. 1km² are widely used. However, these units are not convenient from both the operators' and users' perspectives since 1km already exceeds the accepted distance for users to access available bikes and there are still too many choices to place bikes within a 1km² space. Thereby, a suitable unit should be introduced. Furthermore, Furthermore, existing studies mostly focus on the optimization of bike reallocation strategies but these reallocation strategies have not been evaluated to see to what extent they can improve the service in a real-life context. Even when assessing the service level of bike sharing service, only the general ridership and the ridership ratio (rides per bike per day) were considered. Therefore, a well-rounded evaluation method should be introduced to examine the effects of the operational strategies for bike sharing programmes. Additionally, there is a lack of e-bike sharing studies in general.

This research was commissioned by bondi, a start-up company providing shared mobility services in The Netherlands. Its aim is to improve the service so as to attract more rides without taking excessive reallocation actions considering the available resources. The case study in this research is then confined to bondi's e-bike sharing project in The Hague.

To fill these scientific gaps determined in the existing literature, as well as to fulfil the commissioner's goal of improving the service in order to attract more customers, the research objective has been elicited:

***To develop a data-driven approach to derive beneficial operational strategies, in order to improve the service of the e-bike sharing, by conducting data analysis, and demand pattern analysis and follow-up examination of the proposed strategies.***

**Methodology**

To achieve the research objective, this study is divided into four stages: 1) the preparation phase to study the contributing factors of demand, via literature review; 2) the data analysis to understand the correlation between the contributing factors and demand and provide insights of land use pattern, as prerequisites for the next step; 3) the demand pattern analysis, to reveal the demand pattern of bike sharing, as well as construct the corresponding operational strategies; 4) the operation phase, to adopt the proposed operational strategies as well as evaluate their effects in real life.

Thereon, the methods under each phase were developed, either consecutive or parallel to each other, organized in a systematic way and forming the data-driven approach as a whole.

For the literature review, several keywords are applied to search the relevant work, such as 'bike sharing', and 'shared bikes'. These words were combined with other words indicating the focus in different aspects, such as 'factors', 'statistical analysis', 'demand pattern', 'prediction', and 'operation'. The synonyms of these keywords are also used. In addition to the insights of applicable methods for each topic, another goal for this phase is to categorize the contributing factors of the demand for e-bike sharing projects from the existing literature.

Turning to the data analysis, the data required in this project, involving the ride data as well as the determinants, such as POIs and weather, found from the last stage, are acquired and described. Data cleaning and processing are also applied in this phase. Next to this, the relationship between the demand and the contributing factors are examined via Pearson's coefficients and multiple regression analysis, to reveal the linear correlations among data. Then, the land use pattern of the context is conducted to understand the context, facilitating the demand pattern in the later steps with an insight into the function of different spatial units. This is done by the distribution analysis of different POIs (points of interest) and the function is determined by the proportion of POIs within a unit where if there are equal or more than 50% of POIs belonging to a certain type, this unit is regarded as dominant by the corresponding function.

Followingly, in the essential stage, the demand pattern analysis is conducted. There are several steps in this stage. On top of this, a descriptive analysis is done to understand the distribution of the demand characteristics, such as the demand over different hours of a day, different days of a week, and the trip distance/duration. Second to that, the spatial analytical units are determined, based on their efficiency and effectiveness on the operation as well as the efficiency of the following algorithm (agglomerative hierarchical clustering). Apart from the administrative units, a novel unit, overlapping circles are introduced in this study, easing the efforts of operations. Based on the determined units, general demand pattern analyses are conducted to see what units attract more flows/departures/arrivals and if these units are balanced in terms of departures and arrivals all the time. Afterwards, temporal clustering is conducted and the demand pattern is studied for each of the determined temporal clusters. Based on the insights obtained from the demand pattern, rebalancing strategies are proposed. Supplementary to this, supply efficiency analysis is conducted to see how the vehicle idle time differ between different units; average travel duration and distance are also analysed on the unit level. Dependent on this, units with the least usage, indicated by longer vehicle idle time and shorter travel duration/distance are determined, and thereby the decision to adjust the operational zone is made.

Finally, these suggestions are executed in the real life, and they are evaluated to see to what extent these strategies can improve the service. Two sets of KPIs are used to assess the effects: general ridership, ridership ratio, vehicle idle time from the operators' side; net retention rate and average user expenditure from the users' side. The vehicle idle time here is computed per vehicle, instead of per unit in the last stage. Net retention rate is an indicator presenting the percentage of recurring revenue retained from existing customers over a given time period and it is set to be a month in this study. The average user expenditure expresses how much a user spends on the service on average per month. Apart from these quantitative analyses, qualitative analysis is also conducted to compare the performance of these strategies.

**Results**

After the development of methodology, these methods were applied in the case study, bondi e-bike sharing in The Hague.

Firstly, the determinants were established from the literature review, and they belong to 6 groups: 1) spatial and infrastructure factors (Hampshire, 2012); 2) weather related factors (Sabir, 2011); 3) mobility and trip characteristics (Tang et al., 2011); 4) temporal factors (Mattson & Godavarthy, 2017); 5) sociodemographic factors (Buck et al., 2013); 6) safety factors (Fishman et al., 2012). Data relevant to these factors were searched and acquired if they are available.

The commissioner of this project is bondi. It is a start-up company who provides shared-mobility services and this research focuses on its e-bike sharing project in The Hague.

The fundamental data input is provided by bondi and it consists of ride records from 19/06/2021 to 19/10/2021 where the data of the first three months were used to construct the operational strategies (corresponding to the second and the third phases in this research) and the evaluation was conducted on the whole dataset. In addition to that, data of contributing factors were obtained from the open-source database, which is composed of POI data (point of interest), weather data, public transport data, corresponding to 1), 2), 3), and 4) categories as mentioned above. The sociodemographic data is missing because of privacy concerns and there is a lack of safety data since they are usually obtained from interviews or stated preference surveys, which are out of the research scope.

After the data description, the data analysis was conducted based on these data. The correlation analyses were done by the computation of Pearson's coefficient. It is found that the number of POIs within a unit and the availability of public transport which is indicated by the number of stops and the ridership of these stops within a unit have positive impacts on the demand (the ridership). The factors under the weather groups, which are the temperature, the precipitation, and the humidity, have different effects on the ridership: the temperature is positively related to the ridership while the precipitation and the humidity affect the ridership in a negative way. The highest positive Pearson's coefficient is found between the number of the sustenance facilities and the ridership, which is 0.92 for the neighbourhood units and 0.76 for the 400m overlapping circles. The other positive values of this indicator range from 0.16 to 0.92 where the lowest coefficient is observed between the temperature and the ridership. The magnitudes of precipitation and humidity are -0.03 and -0.31 respectively. To further understand the relative importance between different factors, the multiple regression analyses were carried on: the number of the sustenance amenities, the PT (i.e., public transport) ridership, and the humidity are the most important ones under their groups.

The correlation analyses are followed by the land use pattern analysis, the distribution of POIs was visualized and the proportions of offices and recreational facilities, involving both sustenance and entertainment amenities, were computed. The proportions equal to or over 50% were identified and the corresponding functions were thereby assigned to each unit. It is observed that the central units are prone to recreation while the office-oriented units spread sparsely around the city. These insights are incorporated in the following demand pattern to see what types of units people prefer to move to/from.

Then, the demand pattern analysis was done in this case study. On top of this, a descriptive analysis was conducted. It is found there is only one obvious peak between 16:00 to 18:00, different from two peaks observed from the literature; the average trip distance is 3.65 km and the average trip duration is 18.2 minutes. Following this, the general demand pattern is analysed for two datasets between 19/06/2021 to 15/07/2021 and 30/07/2021 respectively and it figures out that the central units are the most popular and there are obvious discrepancies between the number of departures and arrivals in the central units and the beach area. The central units witness more departures while the beach area is dominant by arrivals. Thereby, the main suggestion based on these results is to reallocate the bikes from the beach area to the central units. Second to that, temporal clustering was conducted via agglomerative hierarchical clustering. It is observed that there are 5 hourly clusters, 16:00 to 16:59, 17:00 to 17:59, 18:00 to 18:59, 19:00 to 19:59, and 20:00 to 15:59, corresponding to the peak hours, the transition hours and the off-peak period separately, on the 400m overlapping circles. This clustering was also applied on the neighbourhood levels and there are 5 different clusters emerging, while the two peak hours distinguish themselves in the hourly clusters no matter what units are used. It is found from the second peak hours until the end of transition hours, users prefer to move from the office-oriented units to the recreational units and another interesting fact is that the station unit is the most popular one between 17:00 to 19:59, implying that people use e-bike sharing to serve the first and last mile to the train trips, as presented in Figure 1. Based on the insights from these two rounds of hourly clustering on different units were used to develop two sets of reallocation strategies. In addition, there is no clear daily clustering in the dataset, which is caused by that people do not have different routines in terms of the day of the week yet as this period is the first four months since the launch of this service. Parallel to the main demand pattern analyses, the vehicle idle time and average travel duration/distance were calculated for each unit. It is found that the southeast of The Hague experiences few rides or even no rides. Thereby, a suggestion, the reduction in the service area, was proposed, by leaving out these units and reallocating the bikes in those areas to the prevalent ones.

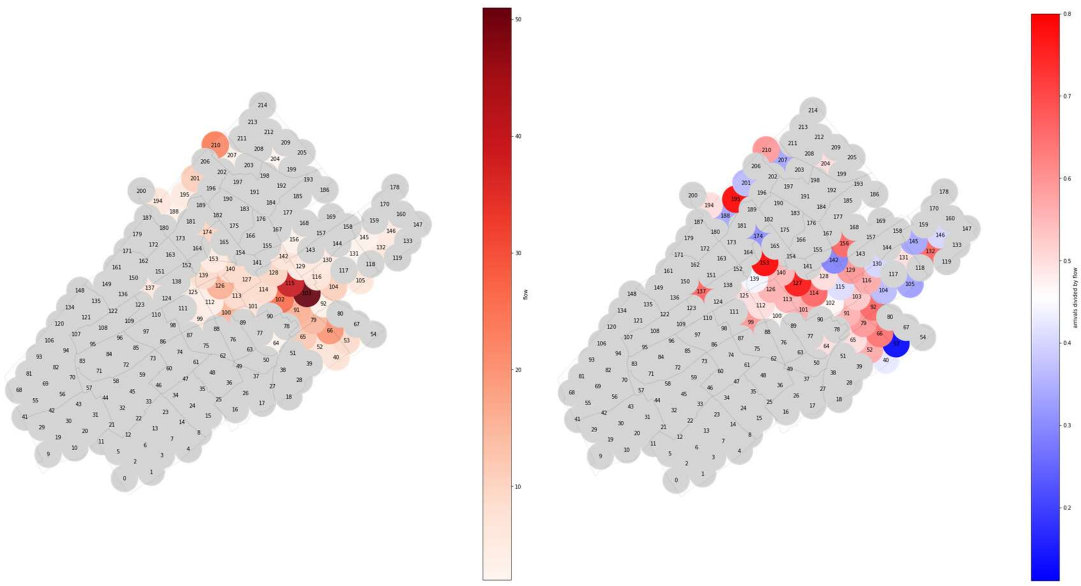


Figure 1. Heatmaps of total flows and arrival ratio of the 3rd cluster (18:00 – 18:59) in circles

Turning to the operational stage, these operational strategies were executed sequentially after the ID verification on the 17th of July until the 18th of September. Their effects were then examined, via the proposed evaluation framework and the results are presented in Table 1-3. It is found that all of them have improved the service, to different degrees though, compared to the ID verification period. The improved ridership ratio ranges from 0.52 to 0.67, compared to the reference at 0.31 for the ID verification period. The reallocation strategies based on the hourly clustering on the circle level have the most desirable effects out of all, and it has decreased the vehicle idle time per vehicle by 12 hours and 19 hours from 59 hours and 56 hours for the origin-based and the destination-based ones respectively.

Other strategies have also mitigated the vehicle idle time at around 10 hours compared to the reference case. The net retention rate is over 50% and the average user expenditure is 15 euros per user after the improvement of service, which is quite decent compared to the other existing bike schemes around the world.

Table 1. Summary of operational KPIs for different strategies

| | | | average supply | average ridership | average ridership ratio | average vehicle idle time per vehicle (h) | |
|---|---|---|---|---|---|---|---|
| | | | | | | origin-based | destination-based |
| free rides | 19/06/2021 | 21/06/2021 | 78 | 164.33 | 2.11 | 8.81 | 17.82 |
| adopting period | 22/06/2021 | 06/07/2021 | 107.06 | 105.33 | 0.98 | 47.28 | 32.14 |
| ID verification | 07/07/2021 | 16/07/2021 | 80.8 | 25.2 | 0.31 | 59.51 | 52.66 |
| first-round reallocation | 17/07/2021 | 31/07/2021 | 92.67 | 48.33 | 0.53 | 51.84 | 43.25 |
| second-round reallocation on the neighbourhood level | 01/08/2021 | 07/08/2021 | 90.29 | 54.14 | 0.59 | 57.3 | 41.08 |
| second-round reallocation on the circle level | 08/08/2021 | 08/09/2021 | 107.72 | 72.44 | 0.67 | 47.3 | 33.07 |
| without specific strategy | 09/09/2021 | 23/09/2021 | 95.93 | 61.8 | 0.64 | 50.16 | 41.08 |
| reduction in the operational area | 24/09/2021 | 18/10/2021 | 87.27 | 45.54 | 0.52 | 50.9 | 43.63 |
| overview | 19/06/2021 | 18/10/2021 | 96.14 | 63.89 | 0.66 | 40.77 | 33.05 |

Table 2. Summary of net retention rate for different months

| Time period | #users | Total revenue (€) | #Retained users | Retained users expansion (€) | #Churn users | Churn users loss (€) | NRR |
|---|---|---|---|---|---|---|---|
| 6.19 to 7.18 | 1056 | 7804.80 | | | | | |
| 7.19 to 8.18 | 425 | 3924.98 | 81 | 243.87 | 975 | 6817.41 | 15.78% |
| 8.19 to 9.18 | 448 | 6855.43 | 139 | 918.06 | 286 | 1942.66 | 86.87% |
| 9.19 to 10.19 | 333 | 4811.94 | 162 | -683.41 | 286 | 3055.01 | 52.10% |

Table 3. Summary of average user expenditure for different months

| Time period | #New users | New users revenue (€) | New user average spent (€) | Retained user average spent (€) | total user average spent (€) |
|---|---|---|---|---|---|
| 6.19 to 7.18 | | | | | |
| 7.19 to 8.18 | 344.00 | 2693.72 | 7.83 | 15.78 | 9.24 |
| 8.19 to 9.18 | 309.00 | 3955.05 | 12.80 | 20.87 | 15.30 |
| 9.19 to 10.19 | 171.00 | 1694.93 | 9.91 | 19.24 | 14.45 |

**Contributions and recommendations**

This work has both scientific and practical contributions.

The scientific contributions correspond to the gaps mentioned in the very beginning: 1) introducing an innovative spatial analytical unit, the overlapping circles to the study of bike sharing, and it is proven that the reallocation strategies derived based on this unit, are more cost-effectively; 2) adding a study in the field of e-bike sharing, providing different insights from the current studies; 3) developing a framework to evaluate the operational strategies and assessing the strategies in a real-life context, taking both operators and users into considerations.

The practical contributions are quite straightforward, meeting the requirements from the commissioner abovementioned, which is to improve the service to catch more users, so it is also beneficial for users. Another contribution in a practical way is that it provides an experience for other operators, especially small companies.

For the future development of bondi, there are also some recommendations. On one hand, reward systems are advised to be integrated in the system so as to mitigate the operational efforts as well as attract more users. Dynamic pricing schemes are a method to stimulate users to reallocate bikes spontaneously, by providing lower fees for these trips ending up at the desirable locations. It helps mitigate the efforts of reallocations by the provider. In addition, discounts are suggested to frequent users, motivating users to use this shared service more frequently, by providing a lower fee or free minutes. On the other hand, extra reallocations are advised, if possible, which allow enough supply in the residential and station areas before the regular morning peak to attract potential users, though the morning peak is not seen in the current service.

**Limitations and future work**

However, there are still some limitations in this study. The most important ones are related to data and the method respectively: 1) the explicit supply data is unavailable thereby the supply level was inferred from the ride records, and it is imprecise; 2) the evaluation method does not distinguish the effects from the operational strategies and other factors, such as the promoting campaigns or the effects of holiday seasons.

Therefore, some future work is suggested: 1) to conduct the research with better data input, including the precise supply data, and inclusion of the relationship with public transport on the defined circle units; 2) to study how to predict short-term demand of e-bike sharing project with relatively low ridership; 3) to adapt the evaluation method of the operational strategies which only focus on the effects stemming from the strategies, by providing better reference cases from the predictive models.

# ACKNOWLEDGEMENTS

This thesis marks the end of my Master's programme in Transport, Infrastructure & Logistics at TU Delft. This research targets in e-bike sharing service, one of my favourite topics in transport and I believe it is definitely the future of transport. I deeply hope this report provides new insights and can be followed by the future work.

First of all, I would like to thank Niels, who links me with bondi and teaches me a lot along during two years. Your teaching sparks our interests in public transport and shared mobility and you are always so patient, providing supports to us. Then, I would like to thank my daily supervisors, Panchamy and Frederik. Panchamy, you guide along this project and spent much time to give feedback, with countless meetings and encourage me when I am struggling, even the format of formulas. Frederik, I really want to say thank you for all your valuable inputs in this project, as you always point out the practical considerations and inform me of the user aspect.

Next, I would like to thank Max and CK, who provide me this amazing internship opportunity. I would also like to thank my other colleagues from bondi, who company me during the past 6 months. I really enjoy the family atmosphere at bondi. It is really an unforgettable experience. You trust my expertise and we work hard while keeping fun.

Third, I would like to thank my family, especially my parents, Chaoyang and Meiqiong, who raise me up and give me support in the past two decades. I could not express my gratefulness in this short page and I really miss you.

I would like to thank you my friends at TU Delft, thank you for all wonderful time together and you teach me how to do a teamwork.

Last but not the least, I would like to thank my friends, Qiao, Meiying, Yue and Yuhan, and also other friends not matter where you are. Thank you for your company along the way.

These two years and three months are extremely challenging, unexpectedly, while I learn a lot from this journey and I realize I should not rush, not matter in life or in study. Every step makes sense and now I understand what is it about. All the people aforementioned help me to believe it and inspire me to become a better person. I hope I also bring the happiness to you.

Again, I would like to express my gratefulness to all of you and to myself as well. Finally, I survived and I am really proud of my work.

Ziru Zhang

November, 2021, Delf

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACRONYMS

**ANN** Artificial Neural Network

**AR** Autoregressive Model

**ARMA** Auto-Regressive Moving Average

**BGIP** Bimodal Gaussian Inhomogeneous Poisson

**CNN** Convolutional Neural Network

**con-LSTM** Convolutional Long Short-Term Memory

**DLM** Dynamic Linear Model

**HTM** HTM Personenvervoer

**GBRT** Gradient Boosting Regression Tree

**GCNN-DDGF** Data-driven Graph Filter Model

**GEV** Generalized Extreme Value Count Model

**GMM** Gaussian Mixture Model

**GTFS** General Transit Feed Specification

**HA** Historical Average

**HDL-net** Hybrid Deep Learning Neural Network

**ISBT** Inter Station Bike Transition

**LD-BSS** Low Dimensional Model for BSS Demand Forecasting

**LSTM** Long Short-Term Memory

**MBH** Multi-Block hybrid model

**MRR** Monthly Recurring Revenue

**MV-TPR** Student-t Process Regression

**MV-GPR** Multivariate Gaussian Process Regression

**NARNN** Nonlinear Autoregressive Neural Network

**NRR** Net Retention Rate

**NS** Nederlandse Spoorwegen

**RNN** Recurrent Neural Network

**OD** Origin-Destination

**POI** Point of Interest

**PT** Public Transport

**Q** Ridership Level, Indicated by number of Ride Records

**RF** Random Forest

**RSME** Root-Mean-Square Deviation

**SGP** Similarity-Based Efficient Gaussian Process Regressor

**STGCN** Spatio-Temporal Graph Convolutional Network

**TAZ** Traffic Analysis Zone

**TL-AP** Two-level Affinity Propagation Clustering

**UBIMC** Usage Balanced Inductive Matrix Completion Model

# 1 INTRODUCTION

Shared mobility has become a trend since 2010, as a way to improve the sustainability of the transport sector and alleviate traffic congestion (European Commission, 2011). From all forms of shared mobility, shared bikes, as an active mode with an additional advantage of maintaining wellbeing, is fundamental and have been especially popular around the world (Barbour et al., 2019; DeMaio, 2009). Along with the development of the electrification in mobility sector, e-bikes emerged and have been gradually incorporated into bike-sharing schemes, with the benefits of higher travel speed and fewer physical efforts compared to conventional bikes (Fishman & Cherry, 2016).

Shared bike systems generally provide services that make bikes available for shared use on a short-term basis, distinguished from rental bike services which usually provide identical bikes for individuals for a longer term (Hu et al., 2019). It usually allows users to borrow and return bikes at different locations belonging to the same operator and in the defined service zones. This concept first came into the public gaze by a well-known community bike project in Amsterdam and there are currently more than 1,800 bike-sharing schemes including more than 9 million bikes around the world until October of 2021 ('Bicycle-Sharing System', 2021; *The Meddin Bike-Sharing World Map*, 2021). These systems are be further divided into two categories depending on their operation types: station-based shared bikes and free-floating shared bikes namely. Stations-based systems operate at the station level which requires renting and returning bikes in the given docking stations and it is only possible to pick up a bike when there are available bikes and to return bikes when there are available docks to lock the bikes in the stations. Contrarily, people are able to pick up and drop off the bikes at any suitable places under the operational zones of free-floating bike sharing systems, increasing the feasibility of bike-sharing services (Barbour et al., 2019). Thanks to this advantage, dockless shared bikes have gradually gained popularity (Z. Chen, van Lierop, et al., 2020; Fishman, 2016).

For bike sharing systems, no matter what operational types are, it is essential to understand the travel behaviour of users in order to meet potential demand, which assists in a better match between supply with demand, contributing to both operators and users (Hua, 2020; A. Li et al., 2020). There are already numerous studies related to bike sharing systems, targeting different perspectives of this topic, which ranges from influential factors of shared bikes, demand pattern of shared bikes, clustering of shared bike trips by different categories and methods, which usually feed into further predictive phase, prediction of demand in a short-term or a long-term time scope, and optimization of reallocation of shared bikes (Albuquerque et al., 2021; Eren & Uz, 2020; Fishman, 2016; Fishman et al., 2013; Galatoulas et al., 2020). It is therefore categorized into three phases: first, to identify the influential factors for the use of shared bikes; second, to describe the dataset, analyse the demand pattern as well as predict future demand, either in the short term or in the long term; third, to figure out the optimal strategies to reallocate bikes. This literature will be discussed in detail in chapter 2.

However, most of the research is related to station-based bike sharing systems. There are substantial differences between station-based shared bikes and free-floating shared bikes, and the same holds for regular bikes and e-bikes (Z. Chen, van Lierop, et al., 2020; Galatoulas et al., 2020; Gu et al., 2019). For example, free-floating shared bikes do not require users to rent and return the bike in the given station which gives a higher degree of freedom when using, while it increases the complexity when constructing the modelling since the demand is predicted per station for station-based shared bikes. Therefore, for free-floating shared bikes, spatial analytical units, such as virtual stations or traffic area zones need to be first defined as the unit of traffic so that trip generation/attractions can be modelled in these units (S. Liu et al., 2018). For e-bikes, it requires fewer physical efforts and enables a shorter travel time with the same distance compared to regular bikes. Thereby, the trip characteristics, such as travel distance, travel time varies from regular bikes. Besides, e-bikes also have an additional feature, batteries, which also affects people's choices when making a shared e-bike trip (Galatoulas et al., 2020, p.).

Furtherly, when it comes to the studies related to the operational side, there is a lack in the experiments and evaluations of different operational strategies in the real-life where current studies usually aim to determine the optimal strategies based on the dedicated while complicated mathematical models with either static or dynamic demand input. These methods are quite time-consuming and relatively unrealistic for small shared mobility operators, considering the limited human resources and uncertainty of operational action.

**To address all the gaps mentioned above, a data-driven approach is proposed, combining data analysis, demand pattern analysis as well as evaluation of proposed operational strategies, targeting in e-bike sharing projects to adopt efficacious operational strategies for small operators, in order to improve the benefits from both operators' and users' perspectives mainly by mitigating the imbalance between the demand and the supply.**

The contributions of this research are mainly three-holds, both practical and scientific ways, one is for shared service operators, one is for societal benefits and the rest is to bridge the gap in current research.

For operators, better operation strategies both improve the service quality as well as decrease the operational cost. This data-driven method provides an efficient and effective way to improve the current operation, capturing the insights by making full use of the available data. Additionally, this method is rather helpful from the operator's side, especially for small operators, preventing too many computational efforts but still in a decent way supported by several scientific methods.

For the whole society, this research assists to improve the service of shared e-bike. It is beneficial from the users' perspective which avoids the situation that people could not use shared e-bikes when they want because of the imbalance between demand and supply. Moreover, for the governments, the efficacy of shared service also improves the attractiveness of cities and the insights between shared e-bikes and public transport facilitate them to organize different transport modes in a better way. Furthermore, it promotes a more sustainable lifestyle in general.

Last but not the least, this research aims to contribute to a gap in current research as mentioned above, providing more insights into how data-driven methods, such as data analysis, and demand pattern of e-bike sharing will be applied for operational strategies with a follow-up assessment, especially for a start-up company.

This research is based on a case study of bondi e-bike sharing project in The Hague. bondi is a shared-mobility company in The Netherlands that provides sustainable and convenient shared vehicles, such as e-bikes and e-steps, to customers. This research only takes bondi's e-bike sharing in The Hague into consideration while the results are appliable for the broader operation of shared bike companies based on the proposed methods, and also provide new insights to the scientific community in the shared mobility field. A detailed description of problem definition of this study is demarcated in Chapter 3, after the literature review in Chapter 2, taking both scientific gaps as well as the interest of bondi into considerations.

## 1.1 Research Objective and Questions

Following the identified research gap and aforementioned contributions, the main research objective is formulated as follows:

***To develop a data-driven approach to derive beneficial operational strategies, in order to improve the service of the e-bike sharing, by conducting data analysis, and demand pattern analysis and follow-up examination of the proposed strategies.***

Based on the research objective, the main research question is defined as follows:

***How can data-driven methods be used for improving operations of e-bike sharing services?***

To answer the main research question, the following sub-questions are raised:

1. What are the determinants from internal trip data and external from other sources affecting the ridership, empirical analysis methods, and widely-applicable operation strategies for e-bike sharing according to literature?
2. What are the correlations of the influential factors with the ridership of dockless e-bike sharing?
3. What is the demand pattern of e-bike sharing and can it be clustered into different temporal clusters?
4. How can these insights facilitate operational strategies?
5. What key performance indicators can be applied to examine the effects of operational strategies?

6. To what extent the proposed strategies can improve the service of e-bike sharing?

## 1.2 Research Framework and Outline

According to the research objective and questions, the research is divided into 4 main stages, the preparation phase, data analysis phase, demand pattern analysis phase, and the conclusive operational phase. The structure of these stages is illustrated in Figure 1-1.

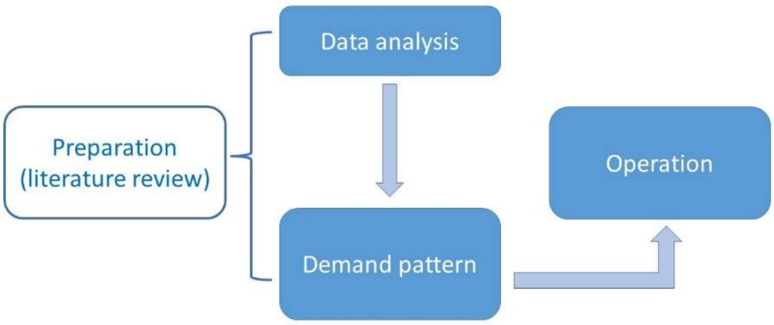



**Figure 1-1 Research structure of different stages**

The first phase, preparation, includes a literature review, the definition of the research objective and the research scope, the identification of research questions and corresponding methodologies. This phase aims to construct the framework of this study. It aims to answer the first research sub-question.

Then the second stage is the data analysis section, which includes the data description, data cleaning and data exploration, serving as a prerequisite for the demand pattern. This phase targets the 2nd research sub-question.

The main purpose of the third phase is to understand the demand pattern of the e-bike sharing project, based on the case study of Bondi e-bike sharing project in The Hague. The demand pattern of e-bike sharing is analysed in this phase. Besides, descriptive analysis of data is also conducted in this stage to examine the distributions of different trip characteristics in the demand for e-bike sharing. Based on the insights gained from the demand pattern analysis, operational strategies will be determined accordingly, targeting the third and fourth research sub-questions.

Data analysis and demand pattern stages compose the crucial portions of the data-driven approach and lay a solid foundation for the proposed operational strategies.

The final stage is the extended application of the second and third stages, with the adoption of various operational strategies suggested by the demand pattern and prediction, and the following analysis of these strategies. Both qualitative and quantitative analyses are conducted to evaluate the results of these strategies. It implies how the operator(s) can use the output of demand pattern and predictive models to improve their service.

According to these phrases, the report is structured as presented in Figure 1-2. The research starts with an introduction chapter, consisting of both background and the following research objective and questions. In chapter 2, a literature review is conducted, presenting the state-of-art research methods related to the research questions, covering from prerequisite data analysis, demand pattern analysis methods to prevalent operation strategies. This chapter is also regarded as the preparation phase of the research. Based on the literature review, the research problem is deciphered in a detailed way in chapter 3, specifying the research gaps found in the literature as well as the needs from the commissioner in an explicit way. After the determination of the exact problem, the elicited methodology framework is presented, targeting to answer each of the research questions. Then, detailed methods of consecutive steps under different research stages are described in the same chapter 4. Afterwards, these methodologies are applied in chapter 5, Case Study, along with the introduction of the background of this case study and the process of applying methods. The results of the case study are then presented in chapter 6. Following this, the research itself is discussed in chapter 7 and is finally summarized in chapter 8.

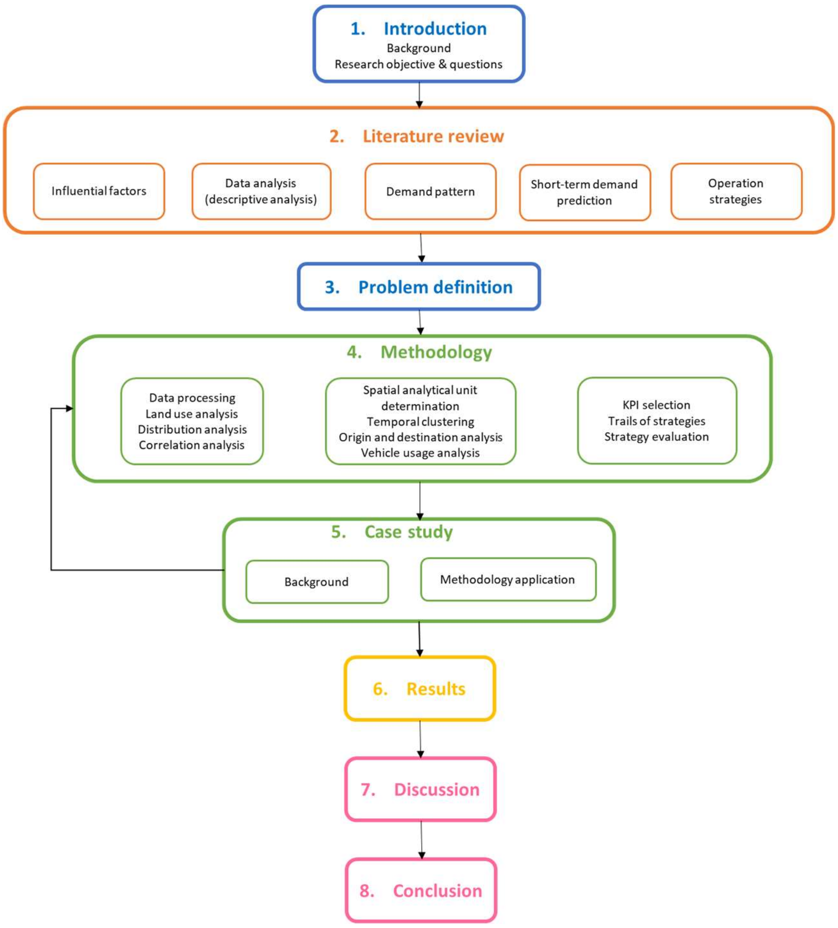



**Figure 1-2 Research outline**

# 2 LITERATURE REVIEW

This chapter describes what are the state-of-the-art methods of different aspects related to shared bike service. It involves the determinants of the demand of shared bikes, statistical analysis of these contributing factors as well as the target (i.e. demand), demand pattern analysis, prediction of shared bikes and the operational phase. This chapter plays a vital role in the whole thesis, providing extensive relevant theories and scientific findings from the literature and answering the first sub research question:

*What are the determinants from internal trip data and external from other sources affecting the ridership, empirical analysis methods, and widely-applicable operation strategies for e-bike sharing according to literature?*

To solve this question, this chapter is divided into 5 parts, tackling influential factors with respect to the use of bike sharing, data analysis of bike sharing schemes, demand pattern, prediction models and operational strategies respectively.

The search engine used here is Google Scholar, and the keywords for search are constructed using the following terms:

*"Shared mobility" OR "Bike sharing" OR "Shared bikes" OR "e-bike sharing"*

*AND*

*"impact" OR "influence" OR "factors"*

*OR*

*"demand pattern"*

*OR*

*"prediction"*

*OR*

*"operation" OR "rebalance" OR "reposition"*

"Bike sharing", as a broader term is used since e-bike sharing is a subclass under it. Most of the research methods from the bike sharing project can be applied directly for e-bike sharing projects, without particular specification though. Similar facts also hold for shared mobility projects in general, such as shared cars as the measures to do the statistical analysis, analyse the demand pattern, predict the demand and operational strategies are regarded as a synthesis to the counterparts for e-bike sharing projects.

## 2.1 Determinants of the demand for bike sharing

In this chapter, determinants of the usage of shared bikes will be discussed. Most of them are relevant to shared bikes in general while some of them are only important for e-bike sharing. Besides, there are differences between station-based shared bikes and free-floating shared bikes while only the latter one is under the research scope and therefore will be discussed in this part.

- Spatial and infrastructure factors

Spatial factors are widely observed to influence the usage of shared bikes, applicable for all types including station-based and dockless ones, regular bikes and e-bikes. These are related to socio-demographic factors as well as the built environment.

Land use pattern is closely related to the ridership as travel is a derived demand from activities and land use patterns are essential for the generation of activities. Consequently, it is observed that retail density, metro rail stations have positive effects on the usage of bike sharing projects in the Washington Metropolitan area (Daddio, 2012). Similar results are also found in other cities which population density, labour market size (i.e. employment density) are strong indicators of bike-sharing trip generation and attractions in Barcelona and Seville (Hampshire, 2012; Nair et al., 2013). Besides, education centres, commercial centres are also positively correlated to ridership levels (El-Assi & Mahmoud, 2015; X. Wang et al., 2016).

Furthermore, elevation is also a key affecting people's choice of shared bikes since a larger difference in elevation acts as a deterrent for people to use bikes due to higher efforts (Reck et al., 2021). However, the elevation differences are almost ignorable in the Dutch context.

In terms of infrastructure, the fundamental one is the availability of cycling paths which is not only essential for shared bikes but also cycling in general (Buck et al., 2013). The presence of separate bicycle paths also affects safety, which will be described later in this part. Another important factor is the availability of public transport, which is also regarded as a competitive alternative to shared bikes. Access to public transport stations is regarded as an attractive determinant when using bike sharing since bike sharing are used for the first/last mile or transfer trips (Rixey, 2013). While sometimes when there is a lack of public transport, bike sharing is the alternative to public transport and therefore the relationship between bike sharing and public transport also exerts an impact on the usage (Kong et al., 2020).

- Weather-related factors

Weather is a common factor affecting the usage of shared bikes no matter what type of shared bikes are as bikes, in general, do not provide physical shelters for users.

Weather, including precipitation, temperature, humidity, wind etc, exerts an impact on cycling in general, ranging from trip generation to travel behaviour (Sabir, 2011). It is also found that seasonal factor (month of the year) affects cycling volumes, though it is more or less related to weather factors (Tin Tin et al., 2012). It is found that colder weather, rain and wind have negative effects on cycling and people would shift to other modes during these types of weather. In the Dutch context, travellers prefer warmer weather even the temperature is higher than 25 degrees and a thermal optimum is found to be days with maximum air temperatures around 24 degrees in the Greater Rotterdam area (Böcker & Thorsson, 2014; Sabir, 2011); However, after a certain threshold, the cyclists' ridership began to decrease, with lower rate though (Miranda-Moreno & Nosal, 2011) It is also thereby observed that summer days are more welcoming for cyclists. Besides, lower humidity levels have positive effects while snow is negatively related to the ridership (El-Assi & Mahmoud, 2015).

- Mobility and trip characteristics

Trip characteristics, without doubt, are influential when making trip decisions. Trip purpose and travel time (travel distance on the other hand) are significant for choice using shared bikes. Trips with travel distances of fewer than 5 kilometres, equal to 30 minutes approximately, are more likely taken by shared bikes and e-bikes enable a longer distance (reference). Conversely, trip purpose composition differs a lot according to contexts. For example, almost half of bike-sharing trips were made for commuting in Beijing while only less than 20% were made for this purpose in Hangzhou (Tang et al., 2011).

Additionally, the system characteristics of shared bike projects also affect the usage in a substantial way. The complexity of use procedures, the service level such as available bikes, comfort of shared bikes, and also battery levels of e-bikes play unignorable roles when using shared bikes (Fishman et al., 2012; Staples, 2017). The last one, battery, is an essential feature distinguishing e-bikes from normal bikes and is a vital determinant for people to use the e-bike sharing service (Ji et al., 2014).

- Temporal factors

Temporal factors are essential for travel, and time of day, day of the week, month of the year are all involved in this range. It is found that the usage pattern of BSPs is in the nature of commute behaviours which means that the usage is usually higher during peak hours on weekdays (Hampshire, 2012; Mattson & Godavarthy, 2017).

The month of the year is already mentioned above in weather-related factors, as a correlated factor of weather, exerts an impact on the ridership.

- Sociodemographic factors

Several studies applied interviews, stated preference surveys and registration data from users to analyse the effects of sociodemographic factors on the usage. These variables consist of gender, age, income level, ownership of cars/bikes, education levels and even occupations.

Young groups, in general, are more likely to use shared bikes, especially e-bikes and the same holds for and low-income groups, proven to be consistent in several contexts such as Canada, UK and USA. However, there is an inconsistency between genders in different countries. For instance, the male took up the majority in Barclays Cycle Hire usage in London while females were dominant in Capital Bikeshare (Buck et al., 2013; Fuller et al., 2011; Ogilvie & Goodman, 2012). It is worth mentioning that ownership of private bicycles also poses a positive effect on the adoption of bike sharing (Cervero & Duncan, 2003; Fishman et al., 2013).

- Safety factors

Safety also plays a role when considering shared bikes, though there are huge variations among different contexts. For example, safety is regarded as the major barrier when making mode choice in Australia (Fishman et al., 2012). This is related to the fear of being involved in traffic accidents with other modes because of a lack of cycling path infrastructure, and therefore also overlaps with network factors (Fishman et al., 2013)

## 2.2 Data analysis of bike sharing

In general, data analysis refers to the analysis of contributing factors or the prerequisite analysis of demand patterns in the field of bike sharing. Sometimes it is also treated as a unique topic to understand the demand profiles (Noussan et al., 2019). In this review, data analysis is regarded as a separate part to reveal the demand profiles by ride attributes as well as the relationship with other factors, to serve a transition to the later demand pattern analysis, as of the defined framework as mentioned in chapter 1.2.

The main components of data analysis of shared bike projects are also dependent on their roles. As the follow-up step of studying the relationship between the ridership and the contributing factors, it serves as the quantitative proof to the impacts from factors; While as the prerequisite to the demand pattern, a descriptive analysis is usually conducted, studying the distribution of various features. However, for the first category, descriptive statistics are analysed before further analysis as well. There is no doubt that descriptive analysis is a core to understanding the data and determining the further steps to tackle with the data input.

For the descriptive analysis, some key attributes presenting the trip characteristics are under the spotlight, such as the distribution of ridership in terms of ride start hour of a day, day of a week (i.e. weekdays vs weekends), trip duration, trip distance etc (Guidon et al., 2019; Noussan et al., 2019). Additionally, the descriptive statistics of contributing factors are also analysed to describe and understand the distribution of those elements (Campbell et al., 2016; A. Li et al., 2020; Mattson & Godavarthy, 2017; Politis et al., 2020).

For the examination of effects on the ridership, regression analysis is widely applied to analyse the relationship, from which linear regression, spatial regression, ordinal regression and log-linear regression are the most prevalent (El-Assi & Mahmoud, 2015; Guidon et al., 2019; Hampshire, 2012; Miranda-Moreno & Nosal, 2011; X. Wang et al., 2016). Regression analysis tests to what extent various attributes exert an impact on the dependent variable, which is usually indicated as the number of departures (trip generation), or the number of arrivals (i.e. trip attraction). Besides, to examine the correlation for spatial features, Moran's I am also applied(Guidon et al., 2019).

Furthermore, if a stated preference survey is conducted to study the contributing factors to the use of a shared bike, a logit model is also executed to understand the decision-making process of users (Barbour et al., 2019; Politis et al., 2020). This type of researches better captures heterogeneities among users and take sociodemographic factors into account, which are usually not included in other aggregated models.

## 2.3 Demand pattern of bike sharing

The next step is to capture the demand pattern of the usage of these shared bikes. It is usually entangled with those influential factors and data analysis mentioned in the previous subchapters.

The demand pattern of shared bikes was firstly usually analysed by descriptive statistics of trip data. It is in line with the examination of the relationship between influential factors and the usage of shared bikes in descriptive analysis. it is found that the rush hours of shared bikes are concordant with commuter patterns, which is morning peak (i.e. 7.00 a.m. to 10.00 a.m.) and evening peaks which usually varies across cities (C. Xu et al., 2018). However, in this phase, the land use compositions of these origins and destinations are also analysed to see if there is a pattern in demand with the refection in trip purpose. For instance, the origin and destinations are usually located in residential and business zones, accounting for the trip purpose of commute.

Besides descriptive statistics, some clusters were also identified from the dataset by using different methods such as K-means method, agglomerative clustering and random forest classification to classify trips with similar characteristics, to gather spatial and/or temporal groups with similarities in demand or increase the predictive accuracy in later steps in this way (L. Chen et al., 2016; Hua, 2020; Jia et al., 2018; T. L. K. Liu et al., 2019). Most of these classifications are dependent on spatial and temporal features based on independent variables, where a widely-used indicator is the time of the day, while some advanced clustering was also done by processing the independent variables (Ai et al., 2019). For instance, trip types indicating the relationship with public transport were firstly inferred from revealed preference data based on spatial-temporal variables as well as factors related to public transport, and then trip types were used as rules to distinguish clusters among all trips (Kong et al., 2020). Besides, there has been an increasing number of research applying machine learning techniques to classify the pattern of usage of dockless shared bikes, which not only capture the independent variables but also the possible relationships among these factors (Zhou et al., 2020).

Demand pattern of shared bikes could be regarded as separate research, but also an essential step for demand prediction, especially for dockless shared bikes as the configuration of spatial clusters to provide the identification of virtual stations, which are followingly used as traffic zones as input of the predictive models (Hua, 2020).

## 2.4 Short-term prediction of bike sharing

After getting familiar with the demand pattern, the fundamental part, demand prediction can be done based on these pre-defined clusters. In this chapter, data sources input to the predictive model, model structures, results from models and comparison among different models will be introduced. Besides, it is worth mentioning that the main interest focuses on short-term demand prediction and therefore the following models are also under this scope.

1. Data input

The major data input, from the category of mobility and trip characteristics as mentioned earlier in this chapter, is retrieved from bike sharing projects which indicate time stamps, trajectories of trips, and sometimes occupancy of stations for station-based systems. For dockless systems, similar variables demonstrate the availability of shared bikes in the pre-defined units where traffic analysis zones and grid-based units are used dominantly. For e-bikes, the battery level could also be derived from operational data directly. Prices of shared bikes service are usually related to travel time and are derived from trip observations.

To gain data of other determinants, open-source data is derived from different sources.

For Spatial and infrastructure groups, land use factors are usually quantified by POIs and are obtained from digital maps, such as google map (Xing et al., 2020); Bicycle path is usually either treated as a binary variable (to indicate if there are available bicycle paths or not) or the coverage of bicycle paths in the aggregated zones. Thereby it is an aggregated factor in general and is mostly used for modelling mode choice compared to other alternatives (J.-R. Lin & Yang, 2011). In terms of public transport, topology characters are also attained from maps and frequency are gained from the GTFS database (Kong et al., 2020).

For weather-related factors, it is obtained from an open-source database from a national institution (Tin Tin et al., 2012); Temporal factors are also related to trip characteristics and thereby obtained from the system directly. Sociodemographic data are divided into aggregated ones, such as population and disaggregated ones, such as gender, age, etc. The former is obtained from national statistics while the latter one is largely dependent on the way of data collection. Stated preference surveys easily provide input while stated preference data is less used for predictive short-term demand of shared bikes, but more for mode choice. This information is sometimes included in the registration system of shared bikes while it is hard to collect and use because of privacy issues. For safety factors, interviews are taken to study how people feel about shared bikes but again, it is regularly used for the adoption of shared bike systems (Fishman et al., 2012).

2. Models

To predict the demand for dockless bike sharing, virtual stations or traffic area zones needed to be first defined. Predictions then are executed for these defined areas.

For the definition of virtual stations or traffic area zones, they are usually defined from the previous step as described in chapter 3. There are mainly three types of units, a grid of predefined length, traffic analysis zone and clusters, while there are also lots of methods to construct the clusters, such as K-means (Ai et al., 2019; Bao et al., 2019a; Hua, 2020; S. Liu et al., 2018; Świątek & Tomczak, 2017; S. Wang et al., 2020; C. Xu et al., 2018; M. Xu et al., 2020). Some other rules are also used to construct different models such as weather type (Sohrabi et al., 2020).

For the types of models, there are usually two main categories, non-linear models and linear models considering the characteristics of models.

For linear ones, the basic models are historical average, autoregressive model and linear regression and some models are either updated versions of these basic models or the combination of them (H. Yang et al., 2019). Andreas applied a model based on HA and AR for the shared bike system in Barcelona while Mohammed applied first/second-order polynomial models in San Francisco(Almannaa et al., 2020; Kaltenbrunner et al., 2010). Besides, generalized extreme value count model and ordinary least squares linear regression models were also used to predict the short-term demand in the United States (Sohrabi et al., 2020; H. Yang et al., 2019). They were adopted because of their simplicity and computational efficiency; However, it is worth noting that all these researches predict the demand for station-based systems instead of free-floating ones.

For the latter group, Monte-Carlo simulation and machine learning techniques have been widely applied to predict the short-term demand of shared bikes with the dominant advantage of higher prediction accuracy and the ability to capture the spatiotemporal dependency of the sharing system (L. Chen et al., 2016).

From the category of machine learning, non-linear support vector machine, non-parametric regression, convolutional neural network and recurrent neural network are widely applied in predictive models while random forest is usually used to predict the long-term demand (Bao et al., 2019a; Hua, 2020; Y. Li et al., 2015; Xiao et al., 2020; C. Xu et al., 2018; M. Xu et al., 2020).

Graph-based CNN takes spatial dependency into account while ignoring possible temporary dependencies. Similarly, long short-term memory from recurrent neural network only considers temporary dependencies. Besides, neither of these models take potential demand dependencies into considerations. However, there are several adapted models which combines the advantages of CNN and RNN to capture spatiotemporal characteristics such as conv-LSTM and it is observed that it performed better than LSTM in the context of shared bikes in Chengdu with higher prediction accuracy (Ai et al., 2019).

Besides, some probabilistic models are also used to predict the demand, such as Gaussian process and student-t regression which are also be grouped into the non-parametric regression models (Huang et al., 2019; D. Li & Zhao, 2019; J. Liu et al., 2016).

The overview of different models is seen in Table 2-1 below.

**Table 2-1 Overview of short-term predictive models applied in bike sharing systems**

| Article | Sharing Type | Location | Model | Model Type | Motivation |
|---|---|---|---|---|---|
| Kaltenbrunner et al., 2010 | SBBS | Barcelona, Spain | ARMA | linear | temporal dependency, station correlation |
| Y. Li et al., 2015 | SBBS | NYC & DC, US | GBRT; multi-similarity-based inference model | non-linear | consider meterology and station correlation; more robust, regular and easier to predict; clusters are more robust |
| L. Chen et al., 2016 | SBBS | NY & DC, US | weighted correlation network; Monte Carlo simulation | non-linear | - |
| J. Liu et al., 2016 | SBBS | NYC, US | MSWK; ISBT | non-linear | - |
| Caggiani et al., 201) | DBS | London, UK | NARNN | non-linear | predict a time series from series past values |
| (Z. Chen, Wang, et al., 2020) | SBBS | DC, US | MV-TPR; MV-GPR | non-linear | MV-TPR performs better in air quality prediction and bike rent prediction; quite mathematical |
| (C. Xu et al., 2018) | DBS | Nanjing, China | LSTM | non-linear | temporal dependency, long-term dependency |
| (Ai et al., 2019) | DBS | Chengdu, China | con-LSTM | non-linear | temporal and spatial dependencies, demand dependency |
| (S. Liu et al., 2018) | DBS | Beijing, China | LSTM; RNN | non-linear | RNN performs well in time series data and LSTM overcomes the shortcomings of gradient exploding and gradient vanishing in the backpropagation process of RNN |
| (L. Lin et al., 2018) | SBBS | NYC, US | GCNN-DDGF | non-linear | learn the hidden correlations between stations automatically |
| (Jia et al., 2018) | SBBS | NYC, US | TL-AP; GBRT | non-linear | cluster: a single station's traffic is too chaotic to predict and not neccesary to predict each station |
| (Almannaa et al., 2020) | SBBS | San Francisco Bay Area, US | DLM | linear | simplicity |
| (H. Yang et al., 2019) | SBBS | NYC, US | Ordinary Least Squares linear regression model and the Poisson regression model | linear | nested sliding window method is used to reduce the over-fitting risk; simple and efficient |
| (Bao et al., 2019b) | DBS | Shanghai, China | HDL-net | non-linear | potential to explore spatial and temporal features; combines LSTM neural network and the convolutional LSTM (Conv-LSTM) neural network in an end-to-end deep learning |
| (Z. Yang et al., 2019) | SBBS | Hangzhou, China | RF; Monte Carlo Simulation | non-linear | the interaction between users and stations; mutual inference between checkin and checkout; |
| (Wu et al., 2019) | SBBS | NYC, US | RF; GBRT; ANN | non-linear | frequently used; |
| (Huang et al., 2019) | SBBS | Bay Area & Boston, US | BGIP | non-linear | the process that people arrive at a station can be viewed as a stochastic process |
| (D. Li & Zhao, 2019) | SBBS | Chicago, US | GMM | non-linear | categorical data captures more generalized trends in bike usage predictions; GMM - flexible fitting structure and explanatory power; Markov is a simple approach to capture sequential |
| (Guido et al., 2019) | SBBS | NYC, US | LD-BSS | non-linear | limited assumptions on the model and parameters |
| (Sohrabi et al., 2020) | SBBS | DC et al., US | GEV | linear | spatial/temporal dynamics; station-specific unobserved effects; high accuracy (5% error for total demand, 75% accuracy of the station-level) |
| (Hua, 2020) | DBS | Nanjing, China | RF | non-linear | better forecasting performance, strong anti-noise ability, fast and efficient |
| (S. Wang et al., 2020) | DBS | Beijing, China | UBIMC | non-linear | effectively incorporate POIs, the spatial-temporal correlations, as well as locally balanced bike check-in/out usage constraint into a joint optimization framework |
| (M. Xu et al., 2020) | DBS | Shanghai, China | MBH | non-linear | captured spatial, temporal, and spatial-temporal properties by CNN, GRU-Net, ConvGRU-Net; Simper than LSTM; more efficient and accurate than RNN, LSTM and other ML models |
| (Y. Li & Zheng, 2020) | SBBS | NY & DC, US | SGPR | non-linear | address the data unbalance issue, and better captures the non-linearity in spatio-temporal data |
| (Xiao et al., 2020) | SBBS | Wenling, China | STGCN | non-linear | - |

3. Evaluation of models

The performance of these models is assessed in general statistics indicators. The most widely used ones are the mean absolute error, the root mean square error (i.e. RMSE) and the coefficient of determination (R square).

Additionally, for machine learning models, the dataset is divided into training data, validation and test data. Validation data is used to evaluate the performance of models and also avoid overfitting. Bayesian information criterion is also used to penalize the number of parameters, preventing overfitting problems. Besides, training time is also considered as the indicator of computation efficiency to compare the performance of models (Xiao et al., 2020; C. Xu et al., 2018).

In some studies, different machine learning models were compared. It is found that an adapted LSTM (conv-LSTM) performed better than a regular recurrent neural network (i.e. long short-term memory neural network) (Xiao et al., 2020; C. Xu et al., 2018). However, these comparisons were solely based on one context and thereby it is hard to prove that one particular type of model is better than others in general.

## 2.5 Operational strategies of bike sharing

This section reviews the literature related to the operational aspects of shared bike projects, including how the strategies are derived and how these strategies are evaluated.

Currently, there are abundant studies relevant to shared bike problems at the tactical and operational levels from the operators' side. The majority of them tackled this problem in the operational research field as optimization problems, which targets the minimization of the total cost, determined by the objectives, such as the total dispatching distance, or the total cost of travel cost, truck, loading and unloading cost with the inclusion of the benefits/loss of these reallocations (Jin & Tong, 2020; J. Liu et al., 2016). These problems are also be further divided into two types: static rebalance problems and dynamic rebalance problems. The latter category has been increasing caught attention during the past years while the previous one is still more common in this topic. The difference between these two groups is the characteristics of the input (i.e. demand) where static problems assume a static demand and the rebalance usually takes place during the night when the demand is very low while the dynamic one depends on a dynamic demand input and is undertaken when there are tangible fluctuations in demand during daytime (Caggiani et al., 2018).

The static problems are usually solved by Tabu Search, Heuristic, Branch and Cut algorithm and Mixed Integer Programming (MIP) which composes of mathematics and graph algorithms (Alvarez-Valdes et al., 2016; Chemla et al., 2013; Dell'Amico et al., 2014; Raviv et al., 2013). On the other hand, the dynamic ones predict the changing demand and take the fluctuations into account when constructing the rebalance choice, applying machine learning techniques prevalently and a detailed review of operator-based relocation strategies are be tracked in the existing literature review (Angelopoulos et al., 2018; Gavalas, 2016)

Apart from the strategies from the operators' side, the users are also counted as the main perspective when formulating the objectives in several pieces of research. It usually applies a dynamic pricing mechanism to encourage users to self-reallocate by picking up and dropping off bikes in the under-demand areas (Kloimüllner et al., 2014).

However, when comparing these methods, only the performance of the optimization algorithms are assessed, such as the efficiency of the optimization methods, demonstrated by the gap and the time consumed to solve with the defined gap (Pal & Zhang, 2017) but the real-life benefits in the service level of bike sharing projects are hardly discussed.

The lack in the evaluation of strategies' effects on the service level not only results from the disability to apply these strategies in the real-life context or the time constrains of the studies but also ascribes to the fact there is limited research in the evaluation of the service level itself. Although some studies conducted Stated Preference Survey to model the user satisfaction to understand different perspectives affecting the service level instead of examining the effects of operational strategies explicitly (Hsu et al., 2018; Mátrai & Tóth, 2016; Xin et al., 2018; Zhao et al., 2014), the existing literature in this filed of the most relevance only take daily ridership and turnover rate (ridership per vehicle per day) into account as the KPIs when deciphering the performance of these projects around the world although sometimes Multiple-criteria decision analysis (MCDM) was also applied in the appraisal, by combining different aspects of bike sharing service (Mátrai & Tóth, 2016). It is without any doubt these indicators are essential to demonstrate the level of service while they are still at a higher aggregated level. There is a need for the evaluation of the real-life effects of these proposed operational strategies.

# 3 PROBLEM DEFINITION

It is found that there are indeed some research gaps based on the literature review in the previous chapter.

Firstly, the prerequisite of all the analysis related to the demand for shared bikes is the determination of spatial analytical units. Equally distributed grids with a defined length and administrative zones have been widely applied in the existing literature because of their capability to describe and predict the demand from the researchers' perspective, while they are not friendly by the operators' and users' sides as these units do not reflect the decision-making process when people are going to use the service. For instance, a 1km grid only reflects the aggregated demand and supply in this unit, while within this unit, some locations are oversupplied while some are undersupplied and 1km is too long for a user to pick up a bike.

Secondly, the current literature in e-bike sharing mainly considers the development of these schemes, discussing the worldwide distribution of these schemes with the main focus on the supply side. Sometimes the factors affecting the adoption of e-bike sharing services are analysed based on the stated preference survey but most of them target sociodemographic factors instead of other essential ones which exert a vital effect on a single ride. Additionally, the demand pattern of e-bike sharing projects are hardly studied, let alone prediction and operational studies.

Thirdly, when it comes to the evaluation of the performance of operational strategies for shared bike projects, only ridership ratio (average rides per vehicle per day) has been applied. Doubtlessly, it is the foremost indicator for shared bikes service, indicating the success of service. However, some other indicators related to service should also be included, targeting specific aspects such as operation, and user satisfaction.

The commissioner of this research project is bondi, a start-up company in shared mobility. The prime need from their side is to increase the general ridership with the least effort considering their limited resource to manipulate excessive reallocation actions. The case study in this research is confined to the e-bike sharing service in The Hague, which launched on the 18$^{th}$ of June, 2021.

Considering the research gaps and interest from the commissioner, the research scope is then defined to preliminary data analysis, demand pattern analysis, development of feasible operational strategies and following assessment of these strategies. A Short-term demand prediction model is explored while excluded in the main report because of its limited usefulness with the fact that demand pattern already obtains the insights in general and the provider only conducts reallocation per three days instead of daily reallocation. In this way, demand patterns already provide the most valuable insights on reallocation.

# 4 METHODOLOGY

Followed by the preparation phase (i.e. the literature review, as described in chapter 2) and combined with practical considerations, the basic framework of other stages, involving theoretical methods are elicited as seen in Figure 4-1.

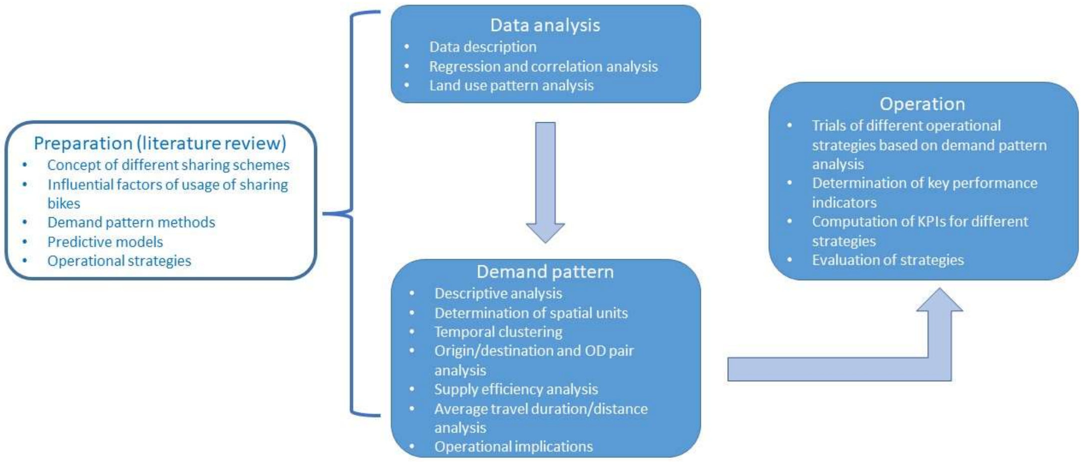



**Figure 4-1 Research framework**

In general, these steps are consecutive and complementary to each other. They are entangled under different stages and provide the next step with their output as the input.

In the data analysis section, data description, correlation analysis and land use pattern analysis are conducted. Correlation analysis is conducted after the description of data, to understand if correlation exists between the variables of interest. The results help explain the fluctuations of demand in the later investigation. Moreover, land use pattern facilitates to understand the pattern of flows between OD pairs.

This stage is followed by the demand pattern analysis, composed of descriptive analysis, determination of spatial units, temporal clustering, origin and destination analysis, supply efficiency analysis and average trip duration/distance analysis. Descriptive analysis aims to understand the statistical characteristics of essential data, such as the ridership across the hour of a day, and it is assigned to the demand pattern section since it mainly targets endogenous attributes of the demand, such as general ridership, trip duration, trip distance etc. Major operational implications and strategies are suggested based on the results from this stage. It is also worth mentioning that the selection of spatial units falls into this category as the results of descriptive analysis cast a substantial impact on it. Taking trip duration as an example, the willing walking distance to pick up an e-bike is dependent on the duration. This step also feeds back to the data analysis phase where the land use pattern analysis is investigated on the defined spatial units. After the general analysis, temporal clustering is conducted to determine the recurrent demand pattern in different hours/days, making the demand predictive and can be used to manage the fleet for different periods by the provider. Agglomerative hierarchical clustering is applied here because the clusters can be observed from the dendrogram and predefined cluster number cannot affect the outcome, ensuring the robustness of the clustering results. Based on the groups, reallocation strategies can be suggested. Furthermore, the units are investigated by vehicle idle time, average trip distance and duration. The results reveal the usage distribution over the spatial units and the operational strategies can be tailored correspondingly, with the combination of the service providers' needs.

Last but not least, experiments of different strategies derived from this data-driven method are set up and conducted in the case study, and the performance is evaluated based on multiple KPIs, with the combination of qualitative analysis with practical considerations. The purpose of this phase is to understand if these derived strategies indeed impact the service positively and compare their performance to provide guidance for future operation.

## 4.1 Data analysis

### 4.1.1 Data description

This step deals with data, encompassing data acquisition, data cleaning, and data pre-processing.

- Data description

Data acquisition involves the obtainment of data of determinants and the dependent variable, ridership itself. They are introduced in sequence.

a. Ride-related data: these data are attained directly from an IT-based system linked with shared e-bikes. Each ride record has a unique ride ID, with a corresponding rider ID and a vehicle ID. Generally, it consists of pick-up timestamp, either drop-off timestamp or trip duration, pick-up location and drop-off location. Sometimes, trip distance is also included in the ride records while it depends on the charge schemes (Shaheen et al., 2010; Vogel et al., 2011).

b. Rider-related data: these data are also retrieved from the operator's system, including the information required in registration. Typical information is a user name, balance, email, phone number etc. Sometimes, ID are required because of safety and operational concerns while this verification is usually done by the third party and the operator(s) only owns the data of the status of ID verification. However, these data are difficult to use because of privacy issues. Fortunately, each rider is assigned to a unique rider ID and their series of ride activities can be analysed with privacy data contactless.

c. Vehicle-related data: this type of data is attained from the operators' system as well, composing a unique vehicle ID, either last ride ID or last ride time, and some defined characteristics by the operators' side, such as fleet category, vehicle name etc. while these data can also be inferred from the ride records, despite the fact if a vehicle is not used at all it is not be included in the ride records.

d. Spatial and infrastructure data: in general, the spatial and infrastructure features are presented as POIs (point of interest) under different categories. They are attained from open-source maps with POIs (points of interest), such as the widely-used ones, *Google map (https://www.google.com/maps), Baidu map (https://map.baidu.com/)* and *open street map (https://www.openstreetmap.org)* (C. Xu et al., 2018). The geographic database consists of POIs with the corresponding latitudes and longitudes (Faghih-Imani et al., 2017). The way to categorize POIs is based on the rules of a different database. Taking OpenStreetMap as an example, it divides amenity into 10 categories, which are sustenance, education, transportation, financial, healthcare, entertainment, art & culture, public service, facilities, waste management and others. With the given category and the geographical information, POIs are aggregated into different spatial levels and the corresponding metrics, such as the absolute amount of POIs in one unit, density of a specific type POIs in the area, proximity of the virtual station to a specific type POI are computed based on these data (Campbell et al., 2016; Faghih-Imani et al., 2017; He et al., 2019). In addition, the popularity of these amenities, which is usually indicated by the flow visiting the facility, sometimes also plays a role when computing the metric. They are acquired from the provider(s) of the facilities. In this study, the flow of interest is the popularity of public transport facilities, which is derived from the fact that shared mobility is regarded as either the alternative or complementary to other transport modes (Eren & Uz, 2020; Kong et al., 2020; Ma et al., 2020; Martens, 2004, 2007). The acquisition of flow to public transport stops is either based on the dashboard and public figures from the public transport providers or the assumptions with ridership parameters obtained from other relevant studies.

e. Weather-related data: the weather data is widely obtained from the national climate database based on different climate stations, and normally it is open-source (El-Assi & Mahmoud, 2015; Tin Tin et al., 2012). For Dutch context, these data are obtained from Koninklijk Nederlands Meteorologisch (KNMI: https://www.knmi.nl) while the data for recent days will only be released after given days. An alternative way is to use weather-related website, which contains both the weather forecast information as well as the historic weather records of the past days.

f. Sociodemographic data: these data involve private information, the obtainment of them is mainly dependent on the operator(s) whether they request these data when users register and whether the privacy policy allows the usage of these data. Generally, they cannot be used without specific requests. Another way to acquire these private data is to conduct a specified survey, stating the purpose of the survey explicitly, to collect while it is hard to match with the existing ride records (M. Chen, Wang, et al., 2020). Instead, they are used to identify how people behave towards the shared mobility considering their own context from a macro perspective, instead of targeting specific rides.

➢ Data cleaning

The data cleaning is divided into two parts, one is related to the shared service itself, including ride-related, rider-related and vehicle-related data; The other involves the data of external determinants, such as spatial and infrastructure data.

For the first category, the essential part is to clean the ride records, excluding the incomplete ride records, faulty records and outliers. The criteria are described as follows:

a. Geographical area: to filter the ride records in the zones of interest. The first step is done with the combination of vehicle-related data, figuring out if the fleet type of vehicles corresponding to the rides belongs to the fleet of related zones. The second step is to use the origin and destination information of each ride, and if they are within the considered area, the ride records are kept;

b. Trip duration: two thresholds of trip duration are set up, with a minimum of 1 minute and a maximum of 2 hours. This set of thresholds varies according to contexts, such as 2 minutes and 1 hour (Reck et al., 2021);

c. Trip distance: the trip distance is set up to be below 100 km to avoid outliers. By this rule, there are indeed some rides being filtered while the reasons behind these trips are uncertain. Similar to the second criteria, other values of this threshold are also used according to the distribution of various datasets, such as 100 meters and 3 km (S. Li et al., 2021). These thresholds are largely context-dependent and targets in the falsely ride records based on the distance;

d. Trip date: it is dependent on the period of interest and is used as a rule to filter data according to the timestamp of ride records.

For the external data, data cleaning is more straightforward. With the selected geographical area and the period of interest, the data under the research scope will be cleaned and attained.

➢ Data processing

Similarly, data processing also corresponds to the abovementioned types.

The process of the data belonging to the first type targets in ride records primarily, and it also depends on the features of raw data and the requirement of following steps. Assuming the raw data of ride records as below, which only include ride ID, rider ID, timestamp of ride start, ride duration & distance, and source & endpoint location as shown in Table 4-1:

**Table 4-1 Ride record example**

| Ride ID | Veh ID | Rider | Start timestamp | Ride duration | Ride distance | Source longitude & latitude | End longitude & latitude |
|---|---|---|---|---|---|---|---|

To obtain other important features, the following computations are made:

a. Ride end: the timestamp of the end of a ride is obtained as the sum of trip duration and ride start timestamp;

b. Day: this attribute refers to the day of the week, dependent on the ride date;

c. Ride hour: this attribute indicates the hour of the ride starting time;

d. Ride end hour: this attribute indicates the hour of the ride ending time;

e. Cost: the cost is usually dependent on the ride duration, and is calculated as the product of ride duration and the rate, which is usually expressed as how much it will cost per minute when using the service.

Besides, aggregated ridership is also obtained from the ride records as the count of all ride records within a certain period and in a given geographical unit.

To further explore the ride records with insights of vehicle and riders, the vehicle and rider data are also merged into the ride records, with the identical rider ID and vehicle ID.

When the supply data could not be obtained directly from the provider, it can be derived from the ride records based on unique vehicle ID. Supply is determined by the used vehicle with a pre-defined period. This method results in some biases, excluding the vehicles which have not been used during the defined period while it is the only way to determine the supply when the precise data is missing.

For the external independent data, the process is highly dependent on the format of the raw data and the following steps. In this research, weather data should be in hourly units considering the hourly demand pattern analysis and short-term demand prediction; POIs information is aggregated in the spatial unit, with the absolute amount and relative portion.

a. Weather data: if the raw data is in a one-hour interval, there is no need to perform an additional process. However, if the raw data is in a 3-hour interval, an assumption is needed to discretize weather data into a 1-hour interval, which is that the weather keeps constant within the 3-hour interval and therefore the same magnitudes are assigned to the consecutive 3 hours. Additionally, the weather data is also processed on the daily basis, by taking the average of weather information, such as temperature, the amount of precipitation, and humidity level of 24 hours belonging to the same day;

b. POIs data: in this research, POIs data is aggregated into 5 categories, sustenance, education, healthcare, entertainment, and office namely. They are presented as the amount of these facilities in a spatial unit. Alternatively, they are presented as the proportion out of all facilities belonging to the same unit or the proportion out of all facilities of the same specific category within the whole area;

c. Public transport data: the process of public transport data involves two data sources, the public transport stations and the ridership. Similar to the process of POIs, one indicator of public transport availability is presented as the absolute amount or relative percentage of stations in the unit. The other indicator takes ridership of stations into account, presented as the sum of all ridership of public transport stations in the unit. The later indicator considers the popularity of stations instead of only concerning the distribution of public transport facilities.

### 4.1.2 Correlation analysis

To examine the relationship between demand, which is indicated as the ridership as explained in section 4.1.1, and other factors, correlation analyses are conducted. This step is regarded as a descriptive analysis of correlation, which does not require a precise determination of different types of relationships, such as linear relationships and other non-linear relationships. The main purpose is to see if some factors have a positive or negative effect on demand instead of deciphering the relationship in a precise way. Thereby, a linear relationship is chosen as the main target of this step, and two methods, Pearson's correlation coefficient and linear regression analysis are applied.

Pearson's correlation coefficient is chosen since the factors to be examined are continuous and it is an easy way to test if there is a significant linear relationship between two variables. While, there are sometimes more than one independent variable exerting impacts on the targeted variable (which is demand in this study), and therefore, regression analysis is also conducted, as a supplement method to examine the linear relationship between one dependent variable and multiple independent variables.

- Pearson's correlation coefficient

Pearson's correlation, also known as Pearson's r, normalized the covariance between two variables and is used to reveal the linear correlation between these two variables, the formula to compute it for a sample is given as Equation 4-1:

$$r_{xQ} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Q_i-\bar{Q})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(Q_i-\bar{Q})^2}} \qquad \textbf{Equation 4-1}$$

Where $n$ is the sample size; $x_i$ and $Q_i$ are the individual sample points indexed with $i$; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the sample mean) and analogously for $\bar{Q}$; in this study, Q always indicates the ridership, aggregated in different time intervals on different spatial units and x represents different influential factors, including weather, number of POIs, and the availability of public transport.

- Multiple linear regression

For multiple determinants from the same group, linear regression is conducted. It is used to examine the strengths of different exogenous variables to the dependent variable. The method of ordinary least squares is used to determine the line which fits the data most.

The multiple linear regression is illustrated as Equation 4-2:

$$Q_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \varepsilon_i \qquad \textbf{Equation 4-2}$$

Where $Q_i$ represents the ridership level on the given spatial level in the corresponding time interval, $x_{i1}$, $x_{i2}$, … are the determinants (i.e., explanatory variables) fallen into the same category; $\beta_0$ is the constant term and $\beta_1$, $\beta_2$, … are the slope coefficients for each explanatory variable

Based on the results of multiple linear regression, the strength, as well as the significance of different variables, will be determined.

### 4.1.3 Land use pattern analysis

In the previous steps, the correlation between the demand and the POIs are examined, and it is followed by land use pattern analysis, to understand how POIs distribute over the spatial units, with the purpose to better understand the demand pattern of shared mobility service in the predefined units.

First, POIs, as described in section 4.1.1, are aggregated to the pre-determined spatial analytical units as Table 4-2 shows, in which | | represents the amount of amenities:

**Table 4-2 POIs data example**

| Unit ID | \|sustenance\| | \|education\| | \|healthcare\| | \|entertainment\| | \|office\| | \|PT\| | \|PT ridership\| |
|---|---|---|---|---|---|---|---|

Second, sustenance and entertainment are then aggregated into the same category: recreation, since pubs, restaurants, and bars fall into the category of sustenance while it is also regarded as a way of recreation, especially when identifying the travel purpose.

After the aggregation of sustenance and entertainment, the percentage of different types of facilities are calculated per unit as Equation 4-3, with the aim to determine the main function of the unit.

$$\boldsymbol{p(i,c)} = \frac{|\boldsymbol{i_c}|}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} |\boldsymbol{i_c}|}, \forall \boldsymbol{i} \in \boldsymbol{I}, \boldsymbol{c} \in \boldsymbol{C} \qquad \textbf{Equation 4-3}$$

Where $p(i,c)$ presents the proportion of a specific facility in a given location; $i_c$ is the type $i$ amenities belonging to unit $c$, and the denominator is the total amount of all facilities in this unit, regardless of their types, where the notation $|x|$ represents the absolute value of x, which is constantly used in this study; $I$ is the set containing all the types of POIs and C is the set consisting of all spatial analytical units.

This proportion indicates how different amenities are composite within a unit but does not indicate the density of the amenity at the municipality level. If the proportion of a specific type is equal to or higher than 50%, the main function of this unit is defined as this function.

## 4.2 Demand pattern analysis

The demand pattern analysis involves several stages: initially, a descriptive analysis is conducted to learn the basic features of data; secondly, the spatial analytical units need to be determined based on the insights gained from the last step; then the demand is analysed in this aggregated level and temporal clustering is conducted to see if demand pattern changes according to time; followingly, detailed demand pattern, such as the popularity of locations as origins or destinations, and different OD pairs are investigated in the determined temporal clusters; simultaneous to temporal clustering, supply efficiency, average travel distance and duration are studied in the spatial unit level to examine the variations in different units; in the final step, operational strategies are proposed based on all the insights from the abovementioned steps.

### 4.2.1 Descriptive analysis

The main purpose of descriptive analysis is to explore the data in general, studying the distribution of the ride attributes in particular.

Therefore, in this step, the distribution of different attributes, including the ride time, in terms of day of a week and ride start hour, trip duration, trip distance, which also includes the presentation of the mean and quantiles. It also helps to determine the thresholds described in chapter 4.1.1 and spatial analytical units, which will be described later on.

### 4.2.2 Determination of spatial analytical units

There are two types of spatial analytical units used in this research, one is based on the administrative units, neighbourhood, and the other, defined as an equally distributed polygon, overlapping circles, in geographical areas, which is advantageous for operations.

The administrative units are obtained directly from the division and definition from the municipality, and it is usually districts and neighbourhoods under the municipality level.

### 4.2.3 Temporal clustering

This step deals with the temporal clustering based on similarities and dissimilarities in demand of different periods. The aim of temporal clustering is to investigate if demand pattern for different periods emerges and the insights will be studies in the next step, aiding to development of operational strategies (T. L. K. Liu et al., 2019).

The whole procedures are shown in Figure 4-2: firstly, rede records are aggregated in the spatial units determined in the last step and then OD matrix is computed accordingly, at different aggregated levels as presented in Equation 4-4 and Equation 4-5; it is followed by the temporal clustering based on these OD matrices, which gathers different periods with similar features together, and are used to capture essential demand peculiarity in the next step.

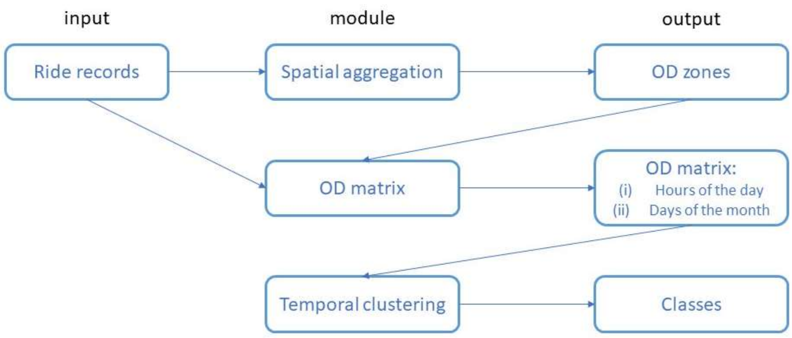



**Figure 4-2 Procedures of the demand pattern clustering**

OD matrix presents the flow between different locations with the insights on how demand (i.e., rides) are attracted and generated at zonal levels. For an OD matrix, each cell indicates the number of rides started within a given period from a particular origin to a destination on a given day.

First of all, the ride records are aggregated with the predefined spatial units as the origin and destinations, and the flow $q(x, y, t, \tau)$ then corresponds to a specific origin $x$, a destination $y$, a ride date $\tau$, and a ride hour, presented by the ride start hour, $t$.

Secondly, hourly clustering and daily clustering are considered in this work, and therefore two series of OD matrices are constructed: hourly OD matrices and daily OD matrices.

For hourly OD matrices, the flow $q$ is aggregated in the increment of 1-hour interval from 0:00 to 24:00, and it is respective to each day. Therefore, there are 24 D*k*k OD matrices in total where D specifies the number of days and $k$ is the number of zones, where each cell corresponds to the flow between the given OD pair during a specific hour for a given date $\tau$.

$$\boldsymbol{Q_t(c_o, c_d, \tau) = \sum_{x \in c_o} \sum_{y \in c_d} q(c_o, c_d, t, \tau)} \qquad \textbf{Equation 4-4}$$

$$\boldsymbol{Q_\tau(c_o, c_d, t) = \sum_{x \in c_o} \sum_{y \in c_d} q(c_o, c_d, t, \tau)} \qquad \textbf{Equation 4-5}$$

In a similar way, daily OD matrices are dependent on the aggregation on a daily basis, generating D 24*k*k OD matrices in total where each cell corresponds to the flow increment of 1-hour interval between a given OD pair ($c_o$, $c_d$) for a given date τ; $c_o$ and $c_d$ represents the origin and destination on the zonal level while x and y is the exact geolocation of the origin and destination of the ride records.

Thirdly, temporal clustering is conducted with the aggregated OD matrices as the feature vectors. Each data point consists of a corresponding OD matric. Agglomerative hierarchical clustering is chosen because there is no prerequisite of the number of clusters before conducting, and it uses a dendrogram to assist the determination of the optimal number of clusters (Rokach & Maimon, 2005).

In this research, Euclidean distance is used to compute the dissimilarity metric as shown in Equation 4-6 and ward method is applied to combining the clusters by the variance of clusters as Equation 4-7 where $\triangle\,(A,B)$ , the merging cost, is minimized during the clustering process, which is found to be the most suitable method for quantitative variables (Calinski & Harabasz, 1974).

$$\boldsymbol{d(p,q)} = \sqrt{\sum_{i=1}^{n}(\boldsymbol{q_i} - \boldsymbol{p_i})^2} \qquad \textbf{Equation 4-6}$$

Where p and q present two points in Euclidean n-space, $p_i$, $q_i$ is the Euclidean vectors and n indicates the dimension of vectors.

$$\triangle\,(\boldsymbol{A},\boldsymbol{B}) = \sum_{i\in A\cup B}\|\overrightarrow{\boldsymbol{x_i}} - \overrightarrow{\boldsymbol{m_{A\cup B}}}\|^2 - \sum_{i\in A}\|\overrightarrow{\boldsymbol{x_i}} - \overrightarrow{\boldsymbol{m_A}}\|^2 - \sum_{i\in B}\|\overrightarrow{\boldsymbol{x_i}} - \overrightarrow{\boldsymbol{m_B}}\|^2 = \frac{\boldsymbol{n_A n_B}}{\boldsymbol{n_A + n_B}}\|\overrightarrow{\boldsymbol{m_A}} - \overrightarrow{\boldsymbol{m_B}}\|^2$$

**Equation 4-7**

Where $\overrightarrow{m_j}$ is the centre of cluster $j$ and $n_j$ is the number of points in it; $\|\vec{\mathbf{x}}\|$ is the absolute value norm of $\vec{\mathbf{x}}$ on the one-dimensional vector space and $\boldsymbol{A} \cup \boldsymbol{B}$ is the union of the sets A and B.

Agglomerative hierarchical clustering, as a bottom-up algorithm, starts with the cluster number equal to the number of data points with zero merging cost since each data point is an individual cluster, and the successive converging process continues until only one cluster is left. The optimal number of clusters will then be decided based on the dendrogram.

Hourly clustering and daily clustering are conducted in this study based on the OD matrices aggregated at hourly and daily levels as described before.

### 4.2.4 Origin and destination analysis

Based on the temporal groups determined from the last step, this step explores how demand varies according to time-based on the fundamental features in origins, destinations and the magnitudes of popular OD pairs. The popularities of different locations are compared during the evolving periods.

### 4.2.5 Supply efficiency analysis

The supply efficiency analysis is conducted by examining the vehicle idle time per spatial unit. Vehicle idle time means the time when the vehicle is in place while no ride is taken in the vehicle, indicating how long the vehicle is idle between two rides (Cats et al., 2020). Commonly, the vehicle idle time only refers to the time interval between two rides while sometimes there are only 1 or even no ride records then the assignment of the location corresponding to the idle time is tricky.

To keep the consistency and include all possible idle time records, vehicle idle time is separated into two types, one corresponding to the origins of rides as $VIT^v(c_o, \tau)$, and the analogous for the destinations as $VIT^v(c_d, \tau)$.

$$\boldsymbol{VIT^v(c_o, \tau)} = \begin{cases} \boldsymbol{t}, & |\boldsymbol{R_v^\tau}| = \boldsymbol{0} \\ \boldsymbol{ts_r^v - ts_\tau}, & |\boldsymbol{R_v^\tau}| = \boldsymbol{1} \\ \begin{cases} \boldsymbol{ts_r^v - ts_\tau}, \boldsymbol{if\ r\ is\ the\ first\ ride\ in\ R_v^\tau} \\ \boldsymbol{ts_{r+1}^v - te_r^v},\ \forall \boldsymbol{r} \in \boldsymbol{R_v^\tau} \end{cases}, & |\boldsymbol{R_v^\tau}| > \boldsymbol{1} \end{cases} \qquad \textbf{Equation 4-8}$$

$$\boldsymbol{VIT^v(c_d, \tau)} = \begin{cases} \boldsymbol{t}, & |\boldsymbol{R_v^\tau}| = \boldsymbol{0} \\ \boldsymbol{te_\tau - te_r^v}, & |\boldsymbol{R_v^\tau}| = \boldsymbol{1} \\ \begin{cases} \boldsymbol{ts_{r+1}^v - te_r^v},\ \forall \boldsymbol{r} \in \boldsymbol{R_v^\tau} \\ \boldsymbol{te_\tau - te_r^v}, \boldsymbol{if\ r\ is\ the\ last\ ride\ in\ R_v^\tau} \end{cases}, & |\boldsymbol{R_v^\tau}| > \boldsymbol{1} \end{cases} \qquad \textbf{Equation 4-9}$$

$\forall v \in V$ where V is the set of all available vehicles during period τ and t is the total time duration of the period τ.

$ts$ indicates the starting time, for both the ride or the period; $te$ is the ending time for either the ride, $r$, or the period. $ts_\tau$ is the starting time for period τ and $te_\tau$ is the ending time for period τ;

$c_o$ is the original unit, and $c_d$ is the destination unit;

$ts_{r+1}^v$ is the starting time of the $(r+1)^{th}$ ride of vehicle $v$ and $te_r^v$ is the ending time of the $r^{th}$ ride of vehicle $v$; $R_v^\tau$ is the set including all rides for vehicle v, during time period τ;

If there is only 1 ride record belonging to this vehicle during the given time period, it is tricky to assign which unit this vehicle idle time belongs to. Therefore, this vehicle idle time is separated into two sub vehicle idle times, assigned to the original unit and the destination unit separately. For example, if vehicle $v$ only has a ride record during time period τ, with the attributes ride start time $ts_v$, original location (pick-up location) $c_o^v$, ride end time $te_v$ and destination (drop location) $c_d^v$. The first idle time before picking up from the starting point of this period is assigned as the vehicle idle time for the original unit, while the second idle time from terminating the ride at the destination to the ending timestamp of this period is assigned the vehicle time for the destination unit. Analogously, for more than 1 ride record, the same procedures are applied for the first and the last rides: assigning the time between the starting time of this period, $ts_\tau$ and the starting time of the first ride, $ts_r^v$; assigning the time between the end time of the last ride, $te_r^v$ and the end time of this period, $te_\tau$.

The time interval t is determined by the usage of the shared service. For example, if there is 1 ride per vehicle per day on average, the time interval could be set up as 1 day, preserving the maximal vehicle idle time of 24 hours. Similarly, if the ridership ratio is lower, such as 1 ride per vehicle per 3 days on average, the time interval should be 3 days with the maximum of vehicle idle time as 72 hours. It is worth mentioning the default time interval is 24-hours when the decision is tough to make. Additionally, the frequency of reallocation also exerts an effect on the determination of the time interval, since the main aim of this indicator is to aid the reallocation, avoiding low usage in general. Thereby, this time interval should also be compatible with the frequency of reallocation.

Based on the formula above, unit-based vehicle idle time will be calculated. To assist the decision-making of reallocation, this vehicle idle time per spatial unit per period is computed as follows:

$$\boldsymbol{AVIT_{c_o}(T_t)} = \frac{\sum_{i \in I_{c_o}^{T_t}} \boldsymbol{VIT^v(c_o,\tau)}\ \forall v \in V, \tau \in T_t}{|I_{c_o}^{T_t}|}, \forall \boldsymbol{c_o} \in \boldsymbol{C}, \boldsymbol{T_t} \in \boldsymbol{T} \qquad \textbf{Equation 4-10}$$

$$\boldsymbol{AVIT_{c_d}(T_t)} = \frac{\sum_{i \in I_{c_d}^{T_t}} \boldsymbol{VIT^v(c_d,\tau)}\ \forall v \in V, \tau \in T_t}{|I_{c_d}^{T_t}|}, \forall \boldsymbol{c_d} \in \boldsymbol{C}, \boldsymbol{T_t} \in \boldsymbol{T} \qquad \textbf{Equation 4-11}$$

$i$ indicates each vehicle idle time record, and $I_{c_o}^{T_t}$ is the set of all vehicle idle time belonging to the unit $c_o$ and during the period $T_t$ and analogously for $I_{c_d}^{T_t}$; C is the set of all spatial units; $T_t$ is the period, presenting the different operation stages across the whole period, for the ride records dataset and $T$ is the period of the whole dataset; the denominator, $|I_{c_o}^{T_t}|$ and $|I_{c_d}^{T_t}|$, is the number of all vehicle idle time belonging to unit $c_o/c_d$ during the time period $T_t$.

Based on the magnitude of $AVIT_{c_o}(T_t)$ and $AVIT_{c_d}(T_t)$ in different units, a heatmap can be visualized, describing which unit(s) a vehicle encounters shorter original-based or destination-based vehicle idle time and thereby those locations are the appealing ones to reallocate bikes.

### 4.2.6 Average travel distance and duration analysis

Similar to the supply efficiency analysis described in the last section, average travel distance and duration is computed at the spatial unit level as well. The calculations are conducted as follows:

$$\boldsymbol{ATD_c(T)} = \frac{\sum_{i \in R_c^T} \boldsymbol{travel\ distance_i}}{|R_c^T|}, \forall \boldsymbol{c} \in \boldsymbol{C} \qquad \textbf{Equation 4-12}$$

$$\boldsymbol{ATT_c(T)} = \frac{\sum_{i \in R_c^T} \boldsymbol{travel\ time_i}}{|R_c^T|}, \forall \boldsymbol{c} \in \boldsymbol{C} \qquad \textbf{Equation 4-13}$$

Where $ATD_c(T)$ indicates the average travel distance corresponding to a unit $c$ and a given period T; $R_c^T$ is the set including all the ride records originating at unit c during the time period T and $|R_c^T|$ indicates the size of the set; average travel time is computed analogously with the travel time as the object instead of travel distance.

### 4.2.7 Development of operational strategies

This step constructs a set of operational strategies dependent on all the above-mentioned analyses as demonstrated in Figure 4-3.

With the demand dynamics of different temporal clustering, an imbalance between supply and demand per spatial unit is observed, evolving according to time. Based on these insights, the major reallocation strategies will be derived for different time periods.

Moreover, vehicle idle time also indicates where vehicle experience a shorter idle time and a longer idle time, the reallocation suggestions are drawn based on these results; furtherly, small adjustments with respect to reallocation amount and location are obtained from the distribution of average travel distance and duration at the spatial unit level.

Besides the reallocation implications, some other operational suggestions are also obtained from the abovementioned analysis. The operational area will be increased or decreased based on the vehicle idle time where if there is almost no usage in the unit, it is considered to discard it considering the cost involved in the maintenance and operation of that zone. Similarly, with the adjustment in the operational area, the fleet size will be adjusted.

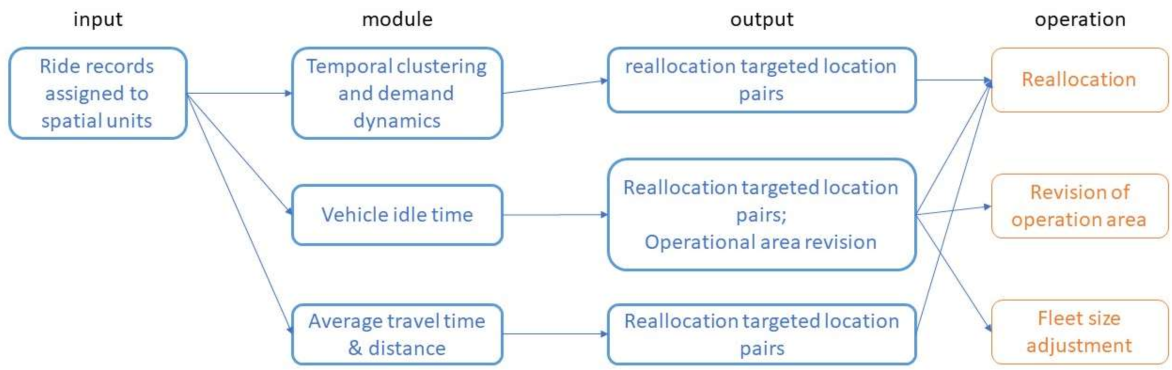



**Figure 4-3 Development of operational strategies**

## 4.3 Operation

In this stage, firstly operation strategies obtained from the previous parts are applied and they are evaluated based on a set of various KPIs (key performance indicators).

### 4.3.1 Experiment setup and execution

The experiments are set up according to different sets of operational strategies recommended in different time periods.

The duration of strategies should vary according to the context as well as the figuration of the next strategies. However, in general, the duration of different strategies should be set up as the same to be comparable if possible. If not, it should also be noted when assessing the effects of different strategies.

### 4.3.2 Key performance indicators (KPIs)

To evaluate the performance of different strategies quantitatively, different metrics, commonly known as KPIs, should be determined first.

In this research, 5 KPIs are used to examine the impacts in the sharing service after the implementation of strategies, from both operators' and users' perspectives. For the operators' interest, the daily ridership, ridership ratio, average vehicle idle time are included; The net retention rate and average user expenditure are proposed to provide insights into the growth of existing customers, by the customers' side.

- Daily ridership

This is indicated by the total ridership every single day. The sum is based on the ride start time. For example, if a ride starts at 23.:59 on 10/09/21 and ends at 00:15 on 11/09/21, it is assigned to the rides belonging to 10/09/21.

$$\boldsymbol{Q_\tau = \sum_x \sum_y \sum_t q(x, y, t, \tau)\,, \forall \tau \in T} \qquad \textbf{Equation 4-14}$$

Where $t$ is the ride start time it takes the sum of all the rides in the given day τ, as long as the origin and destinations are within the operation zones.

- Ridership ratio

This indicates the general ridership ratio per day. The supply of this day is defined as the total available fleet size in the whole operational zone.

$$\boldsymbol{r_\tau = \frac{Q_\tau}{supply_\tau}} \qquad \textbf{Equation 4-15}$$

- Average vehicle idle time

The vehicle idle time is computed in the same way as described in section 4.2.5. However, in this phase, the average vehicle-based vehicle idle time is applied instead of the unit-based ones, using the same set of vehicle idle time records, with a different aggregation way, thought. The object of this metric corresponds to the vehicle, and therefore for each vehicle, the average idle is computed based on all corresponding vehicle idle time records, regardless of the spatial units. Afterwards, the average vehicle idle time is computed by taking the average of all average idle time correspondingly to the available fleet during this time period. Similarly, there are also two types of this metric, the origin-based one and the destination-based one dependent on the category of vehicle idle time.

$$\boldsymbol{AVIT_v^{c_o}(T_t) = \frac{\sum_{i \in I_v^{c_o}} VIT_i^v(c_o,\tau)\ \forall c_o \in C, \tau \in T_t}{|I_v^{c_o}|}, \forall v \in V^{T_t}, T_t \in T} \qquad \textbf{Equation 4-16}$$

$$\boldsymbol{AVIT_v^{c_d}(T_t)} = \frac{\sum_{i \in I_v^{c_d}} VIT_i^v(c_d,\tau)\ \forall c_d \in C, \tau \in T_t}{|I_v^{c_d}|}, \forall v \in V^{T_t}, T_t \in T \qquad \textbf{Equation 4-17}$$

Where $\boldsymbol{I_v^{c_o}}$ is the set of origin-based vehicle idle records belonging to the vehicle v, and analogously applies for $\boldsymbol{I_v^d}$.

$AVITV_{c_o}(T_t)$ and $AVITV_{c_d}(T_t)$ are the average of the average origin-based/destination-based vehicle idle time per vehicle during the time period $T_t$, the denominator is the fleet size belonging to this time period within the operation zone.

$$\boldsymbol{AVITV_{c_o}(T_t)} = \frac{\sum_{v \in V^\tau} AVIT_v^{c_o}(T_t)}{|V^{T_t}|}, \forall\ T_t \in T \qquad \textbf{Equation 4-18}$$

$$\boldsymbol{AVITV_{c_d}(T_t)} = \frac{\sum_{v \in V^\tau} AVIT_v^{c_d}(T_t)}{|V^{T_t}|}, \forall\ T_t \in T \qquad \textbf{Equation 4-19}$$

- Net retention rate

Net revenue retention is a metric demonstrating the variations within the existing revenue base. It is used to describe to what extent the revenue of the existing customers grow or churn on a monthly basis (*Guide to Net Dollar Retention (NDR) - Definition, Calculation, Tips*, 2021). It indicates how much customers spent and their expenditure changes in the service over time and is a way to understand customers' satisfaction: if they are satisfied with the service, they will keep a subscription and continue spending money on it.

$$\boldsymbol{NRR} = \frac{\textit{Starting MRR}+\textit{Expansion MRR}-\textit{Contraction MRR}-\textit{Churn MRR}}{\textit{Starting MRR}} * \mathbf{100} \qquad \textbf{Equation 4-20}$$

Where MRR is the monthly recurring revenue and NRR is computed based on it.

- Average user expenditure

Complimentary to net retention rate, average user expenditure is also computed. Three indicators are belonging to this group, total user average expenditure, new user average expenditure and retained user average expenditure.

$$\boldsymbol{average\ user\ expenditure_m} = \frac{total\ expenditure_m}{|users_m|}, \forall m \in I \qquad \textbf{Equation 4-21}$$

Where M is the set of all months and m refers to a specific month.

This class of indicators provide insights into how much users of different groups spend on the service on a monthly basis.

### 4.3.3 Evaluation of strategies

The performance of strategies is analysed both qualitatively and quantitatively.

Based on the four metrics mentioned above, strategies are compared. The base case is the time period without any particular strategies.

Qualitative analysis is also conducted since strategies are not the sole variables exerting the impacts in the service level, and therefore the effects resulting from other factors are also included in this way.

# 5 CASE STUDY

Succeeding the methodology, this chapter describes the case study in the following way. First, the case study project of bondi's shared e-bikes and the operational area of The Hague are introduced. Second, the methods outlined in chapter 4, are applied in this case study, with specific descriptions of the adoption. Nested in the application of methodologies, the time scope of data used in different steps are also demarcated.

## 5.1 Background of bondi shared e-bikes in The Hague

The mobility provider, bondi, launched the e-bike sharing service on the 19th of June 2021 in The Hague. The launch took place in the late evening of the 18th but officially started on the 19th. The fleet size of the total available e-bikes is 150 bikes while it started with 100 bikes, with a dynamic adjustment in the fleet size according to the operational situations in the following days.

The operational zone of this project also witnessed changes across the time as shown in Figure 5-1. The whole Leidschenveen-Ypenburg district is excluded from the operation all the time. Initially, the operational zone covered almost all parts of the rest 7 districts, presented in Figure 5-1 a, which are Loosduinen, Segbroek, Scheveningen, Haagse Hout, Centrum, Escamp and Laak namely, with the preclusion of some certain areas which were found to be hardly managed, such as the green space located in the outskirts of Loosduinen, Escamp, Haagse Hout and Scheveningen. Only the main residential areas of Escamp were included in the beginning, with an expansion in the later phase (where the neighbourhoods Wateringse Veld, Bouwlust en Vrederust, Moerwijk, the western part of Laakkwartier en Spoortwijk are excluded). This pioneer operational area has experienced minor revision during the launching days and was fixed before the implementation of ID verification in the system as Figure 5-1 b. Finally, the geographical operational area has been narrowed down on the 27th as shown in Figure 5-1 c, in September with the implementation of the new strategy.

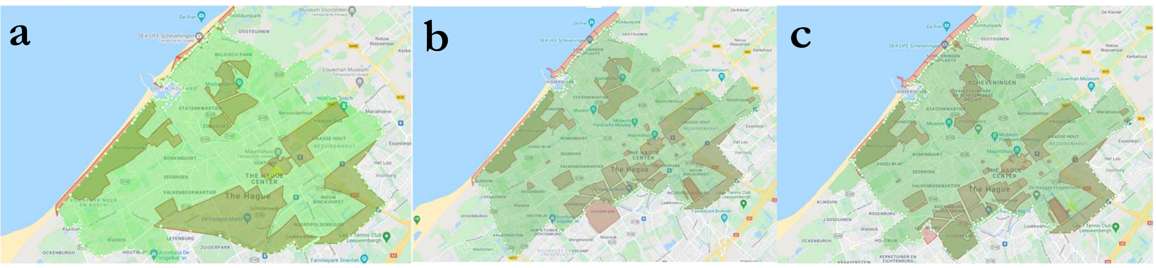



**Figure 5-1 The service area a (from 19/06 to 19/07); b (from 20/07 to 24/09); c (from 24/09 until now)**

The city of the case study, The Hague, as the third-largest city in The Netherlands with nearly 550,000 inhabitants in 2021, is composed of 8 administrative districts and 44 neighbourhoods which are named as Wijken in Dutch ('The Hague', 2021; *The Hague in Numbers*, 2021; 'Wijken en buurten in Den Haag', 2021). The neighbourhood division and spatial distribution of population are demonstrated in Figure 5-2 where it is found that the inhabitants mainly concentrate in the southern part and Laakwartier en Spoorwijk owns the largest amount of them.

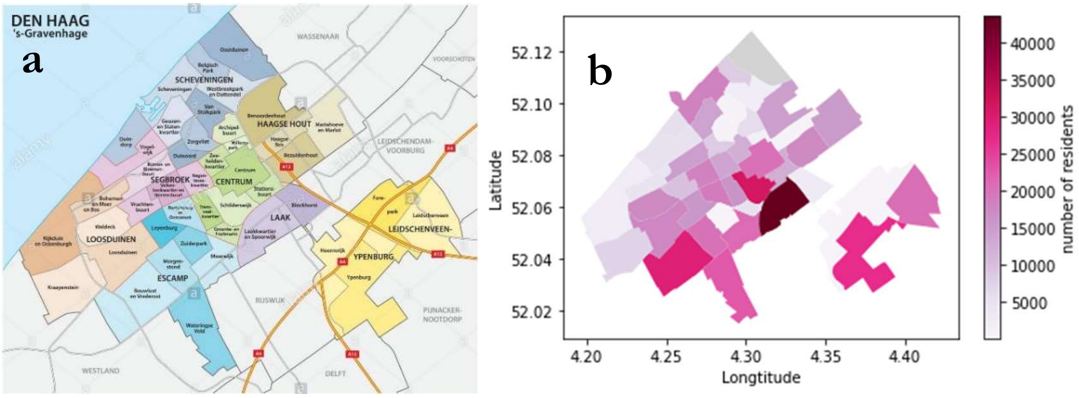



**Figure 5-2 Neighbourhood division (Rainer Lesniewski, 2021) (a); and population distribution of The Hague (b)**

In terms of transport facilities, there are more than 30 bus and tram lines, which are mainly operated by HTM and 6 train stations in The Hague as presented in Figure 5-3 and Figure 5-4. These public transport stops will be used in the following analysis of influential factors, to see if the availability of public transport expose an impact on the demand of e-bike sharing. Additionally, cycling is prevalent in The Hague, same to other Dutch cities (*Holland-Cycling.Com - Finding Your Way*, n.d., p.; Staples, 2017), which lays a solid foundation of the e-bike sharing projects.

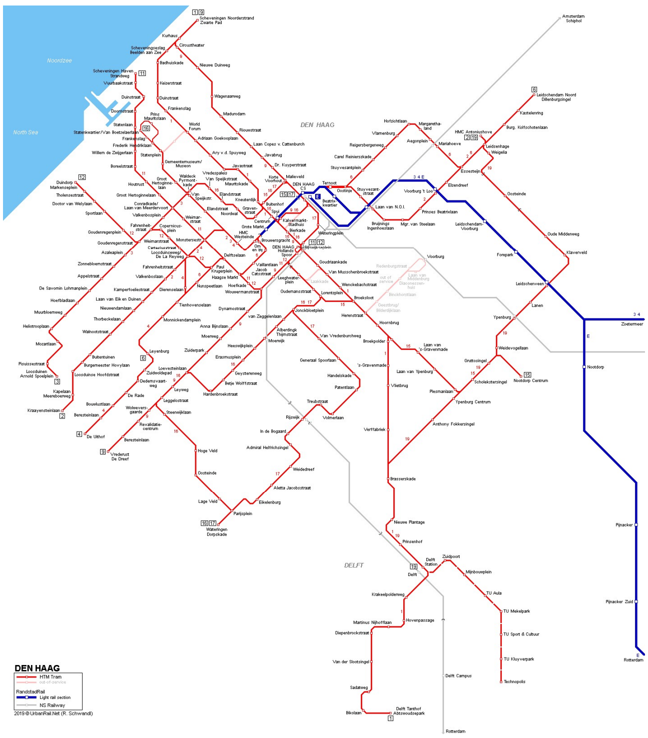


**Figure 5-3 Urban rail map of The Hague (*Den Haag Tram & RandstadRail*, 2021)**

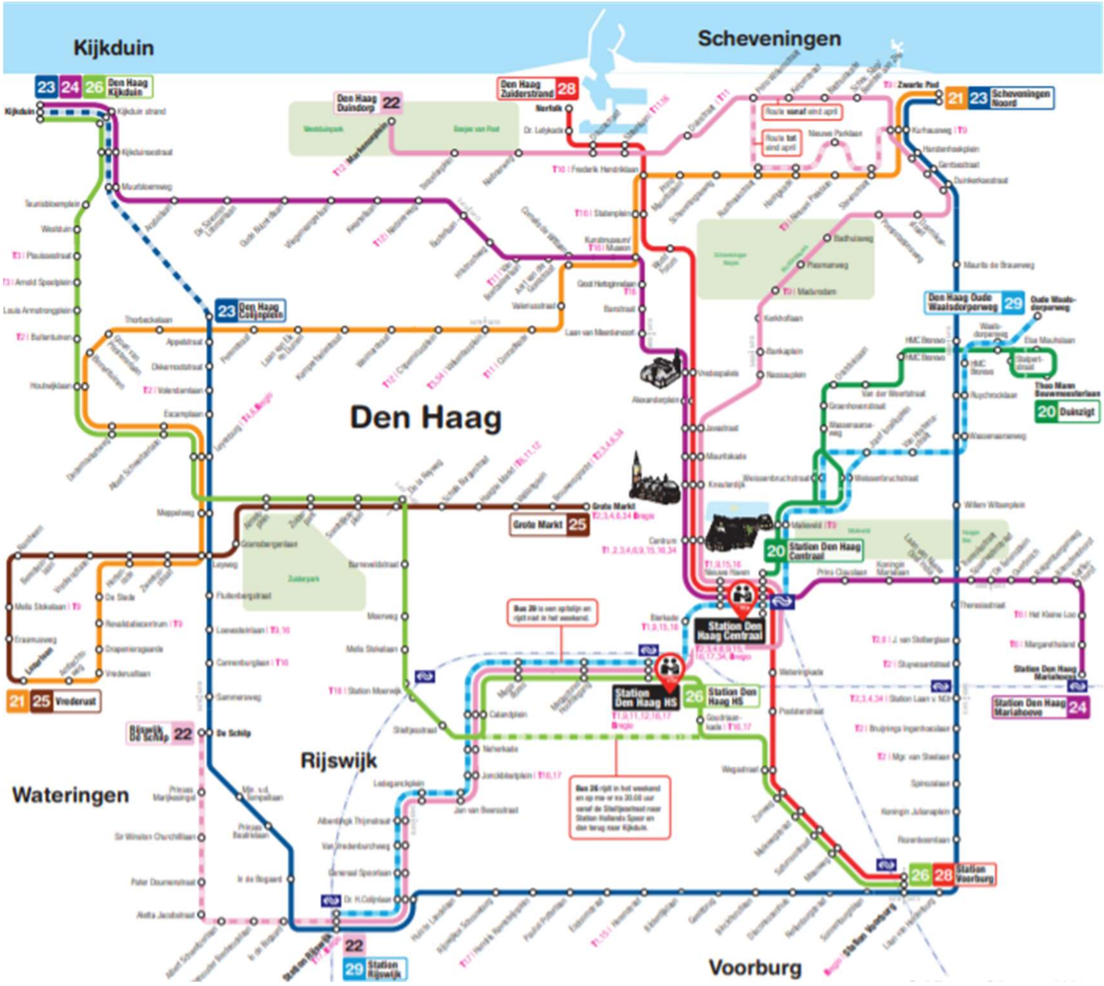



**Figure 5-4 Bus network of The Hague (source: *Line Network Map - HTM*, 2021)**

Moreover, The Hague is a tourist city by the sea with well-known beaches (*The Hague by the Sea | DenHaag.Com*, 2021). This implies a holiday season during the summer, even with the ongoing COVID-19 pandemic. Tourists, therefore, play a vital group of users of this e-bike sharing project.

## 5.2 Data analysis

This section deals with the data analysis phase in the context of bondi's e-bike sharing in The Hague, with the description of the process of the methods' applications.

### 5.2.1 Data description

This part tackles the data acquirement and the data cleaning as well as the processing of all data input.

- Ride records

The main data input is the ride records, where a unique ride ID is assigned to each ride record, and rider ID, vehicle ID, cost of each ride, ride distance and duration, ride start timestamp, source latitude and longitude (of the start point of this trip) and corresponding end latitude and longitude.

The case study focuses on the operation of bondi e-bike sharing in The Hague, and therefore this geographical scope applies to the data cleaning.

The data cleaning is done by the following criteria, using the ride records and operational zones until the 15th of July:

a. Geographical area: to filter the ride records in The Hague, firstly only rides made by vehicles belonging to "Den Haag Free Floating Fleet" are kept. It, therefore, does not contain the reallocation trips since the fleet is changed when reallocating. To avoid the mistakes that sometimes fleet type is not changed during reallocation, additional cleaning is done by filtering the rides made by bondi's staff;

b. Trip duration: two thresholds of trip duration are set up, with a minimum of 1 minute and a maximum of 2 hours. This is based on the distribution of trip duration of rides until 15/07/2021 which shows 75% of trips are under 20.74 minutes while the maximum is 501.97 minutes and the minimum is at 0.15 minutes as demonstrated in Figure 5-5 a;

c. Trip distance: the trip distance is set up to be below 100 km to avoid outliers and faulty ride records. This is based on the distribution of the trip distance of rides until 15/07/2021 which shows 75% of trips are under 4.65 km while the maximum is 276.19 km which probably results from system error and the minimum is at 0.01 km as demonstrated in Figure 5-5 b. By this rule, there are indeed some rides being filtered while the reasons behind these trips are uncertain;

d. There is an additional rule applied to the data cleaning done for ride records with the assignment of origins and destinations to the spatial analytical units (either the neighbourhood level or the circle level) as described in the methodology part: the rides with either origin or destination outside the operational zone are excluded.

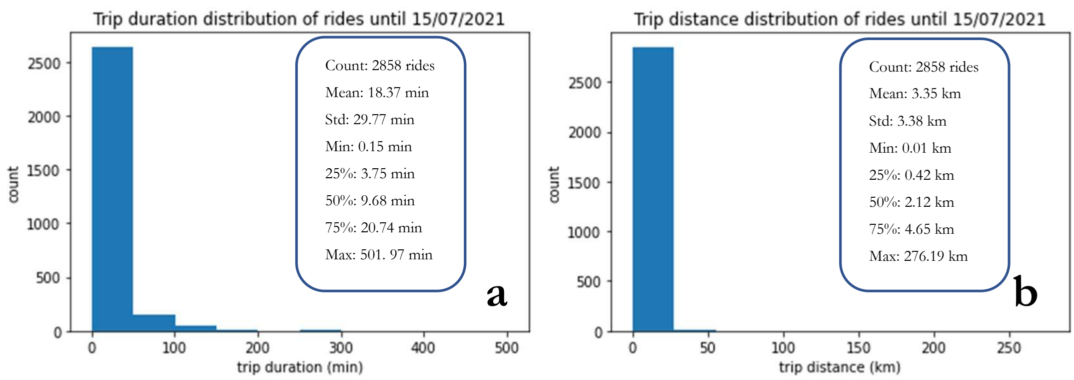



**Figure 5-5 Trip duration (a); and trip distance distribution of rides until 15/07/2021**

By these rules, there are 7858 ride records from 19/06/2021 to 19/10/2021. Various methods applied datasets with different time scopes, and the description, as well as motivation, are represented in the corresponding parts.

Followed by data cleaning, data processing of ride records was made. The raw ride data only contains partial attributes and therefore some pre-processing work needs to be done to obtain a full list of attributes related to trip characteristics.

a. Ride end: the timestamp of the end of a ride is obtained as the sum of trip duration and ride start timestamp;

b. Day: this attribute refers to a day of the week, dependent on the ride date;

c. Ride hour: this attribute indicates the hour of the ride starting time;

d. Ride end hour: this attribute indicates the hour of the ride ending time.

- Supply data

The precise supply data is missing. Therefore, the inference from ride records as mentioned in the methodology part was applied based on the obtained ride records. A 7-days interval ranging from 6 days before the targeted day to the end of the targeted day is chosen as one week is a decent and conservative time scope considering the operational actions are taken once or twice a week.

- Land use data

Land use data of The Hague are retrieved from two sources: open-source data owned by the municipality of The Hague and OpenStreetMap (*Den Haag in Cijfers - Average - Districts*, 2021; *OpenStreetMap*, 2021).

The analysis of these neighbourhoods was done by comparison among the distribution of points of interest (POIs). There are 8 categories of data related to neighbourhoods as listed below:

a. Sustenance: bar, café, fast food, food court, ice cream, pub, restaurant;

b. Education: college, library, music school, university (kindergarten, schools are excluded since the age scope of the users should be above 18 yrs.);

c. Healthcare: clinic, dentist, doctors, hospital, pharmacy;

d. Entertainment, arts & culture: arts centre, brothel, casino, cinema, community centre, conference centre, events venue, fountain, nightclub, swingerclub, theatre;

e. Office;

f. Demographical: number of residents, income level;

g. Traffic: average ownership of private cars/motors per address, public transport stops

**Table 5-1 Overview of POIs in The Hague**

| | Amount |
|---|---|
| Sustenance (a) | 1510 |
| Education (b) | 24 |
| Healthcare (c) | 227 |
| Entertainment (d) | 166 |
| Office (e) | 1224 |
| PT stops (g) | 379 |

There are 3530 POIs in The Hague under category a, b, c, d, e and public transport stops belonging to g as presented above in Table 5-1. Sustenance owns the highest amount while education facilities are the lowest one.

The data from the first 5 categories were obtained from *OpenStreetMap*, indicated by POIs with their exact location on the map; While the 6th category, including the number of residents and residential income level, was obtained from *The Hague Opendata* as aggregated data on the neighbourhood level. Similarly, the average ownership of private cars and motors per address was also aggregated data at the neighbourhood level, obtained from The Hague Opendata while the points of public transport stop, including both tram and bus stops, were obtained from HTM. By doing so, it assists the demand pattern analysis in the following sections.

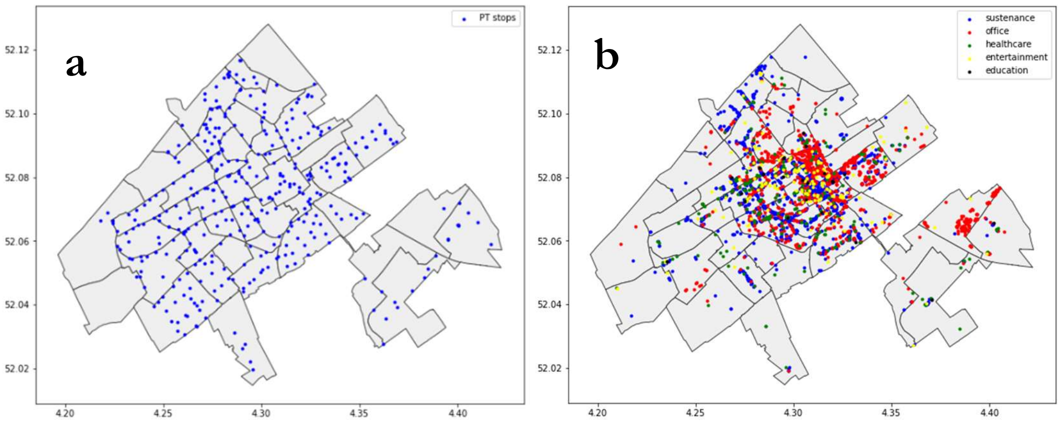



**Figure 5-6 Distribution of POIs: public transport (a) and other POIs (b)**

Tracking the overview of the distribution of POIs from all 5 categories in Figure 5-6, they spread from the central area to the outskirts on the whole. The hot locations of offices are concentrated in the northeast of the centrum, besides Bezuidenhout, which is an office-oriented neighbourhood nearby.

Additionally, the central neighbourhoods tend to be multi-functional, indicated by a balanced distribution of different amenities while outer neighbourhoods show a monofunctional tendency. For instance, the seaside neighbourhoods are mainly occupied by sustenance facilities.  For the outskirts, the amounts of POIs are usually much lower however some neighbourhoods own a high amount of a specific type of facilities. For instance, there are many offices in Forepark which is designated to accommodate workplaces.

- Sociodemographic data

Sociodemographic data on the individual level is not available in this research due to privacy issues. Only income levels and ownership of cars/motors are analysed and the results are presented in section 6.1.2.

- Weather data

There are 984 weather records in total, ranging from 19/06/2021 to 19/10/2021, at a 3-hour interval which means there are 8 records for an identical day. It was retrieved from an online resource and includes temperature, precipitation in mm and humidity as a percentage (*The Hague Historical Weather*, 2021).

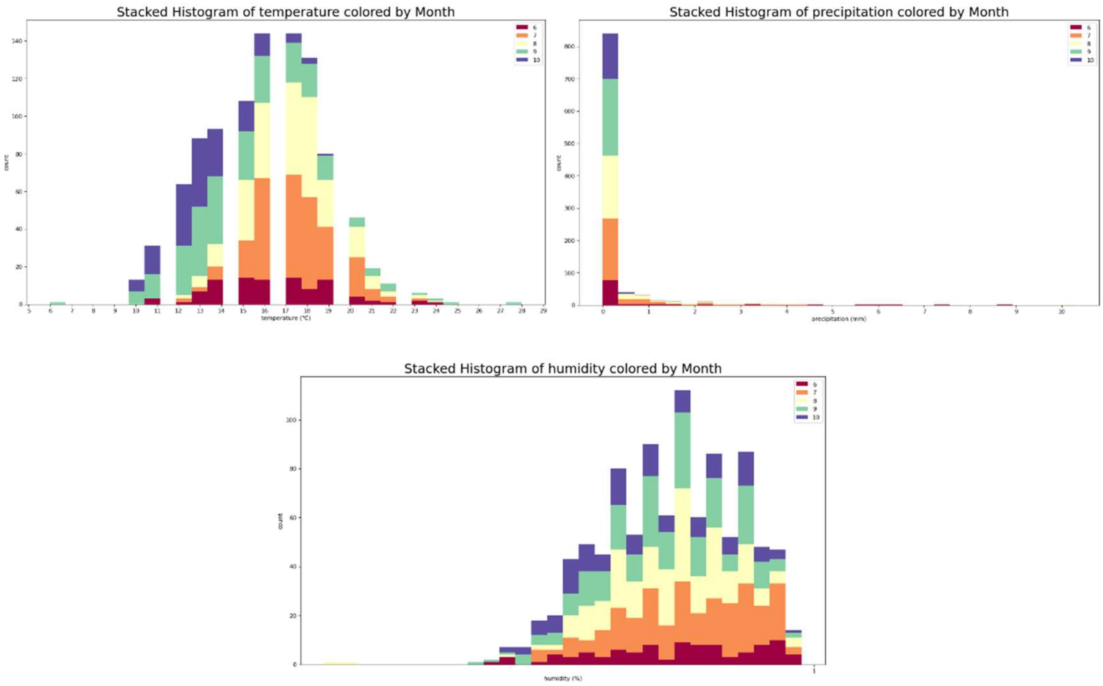



**Figure 5-7 Distribution of weather factors**

**Table 5-2 Descriptive statistics of weather factors**

| Month | | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|
| count | | 96 | 248 | 248 | 240 | 152 |
| | mean | 16.45 | 17.32 | 17.15 | 15.03 | 13.35 |
| | std | 2.67 | 1.81 | 2.03 | 3.06 | 1.85 |
| | min | 11.00 | 12.00 | 12.00 | 6.00 | 10.00 |
| | 25% | 14.75 | 16.00 | 16.00 | 13.00 | 12.00 |
| | 50% | 16.00 | 17.00 | 17.00 | 14.50 | 13.00 |
| | 75% | 18.00 | 18.00 | 18.00 | 17.00 | 14.00 |
| **temperature (℃)** | max | 24.00 | 23.00 | 24.00 | 28.00 | 19.00 |
| | mean | 0.62 | 0.40 | 0.38 | 0.01 | 0.12 |
| | std | 1.67 | 0.90 | 1.04 | 0.06 | 0.35 |
| | min | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 25% | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 50% | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 75% | 0.13 | 0.30 | 0.20 | 0.00 | 0.10 |
| **precipitation (mm)** | max | 8.60 | 6.50 | 10.30 | 0.60 | 2.30 |
| | mean | 78% | 80% | 77% | 76% | 74% |
| | std | 12% | 10% | 10% | 10% | 12% |
| | min | 48% | 55% | 22% | 45% | 50% |
| | 25% | 70% | 73% | 70% | 69% | 64% |
| | 50% | 80% | 81% | 78% | 77% | 73% |
| | 75% | 89% | 88% | 84% | 84% | 84% |
| **humidity (%)** | max | 97% | 96% | 98% | 97% | 97% |

Based on the distribution presented in Figure 5-7 and Table 5-2, it is found that July and August, as the summer season, witnessed a higher temperature in general but also accounted for a relatively large volume of precipitation. October is the coldest month with the lowest average precipitation and humidity.

- Public transport data

The geographical distribution of public transport is already presented in Figure 5-6. According to literature review in Chapter 2, the availability of public transport exerts an effect on the demand of public transport, where sometimes bike sharing is used as an alternative, a connection to public transport.

To indicate the availability of public transport, the ridership of public transport is also needed. The ridership for busline and tramline is not published and thereby an assumption was made to compute the average daily ridership in a stop unit: 631 passengers per tram stop and 368 per bus stop. This was derived from a study simulating the public transport in The Hague which provides the number of passengers per average workday of one tram line and one bus line as presented in the column of Number of passengers per average workday in Table 5-3 (van Oort & Drost, 2015). The total number of passengers per average workday is divided by the number of stops per line and the average daily ridership on the stop level is then obtained as presented in Table 5-3. Besides, the railway station ridership was retrieved from NS dashboard, in an average daily basis in 2019 as presented in Table 5-4, assigning to corresponding HTM stops (*Reizigersgedrag | NS Dashboard*, 2021).

**Table 5-3 Assumption of the ridership of PT stops**

| | Number of passengers per average workday | Number of stops | Number of passengers per stop |
|---|---|---|---|
| Tram line 15 | 12,000 | 19 | 631 |
| Bus line 15 | 7,000 | 19 | 368 |

The overview of ridership is presented in Table 5-4, and then that ridership is aggregated to stop level.

**Table 5-4 Ridership of railway stations**

| | Daily ridership |
|---|---|
| Railway station - Centraal | 104,747 |
| Railway station – Holland Spoor | 40,894 |
| Railway station – Lan van NOI | 18,824 |
| Railway station - Mariahoeven | 3,248 |
| Railway station - Moerwijk | 3,293 |
| Railway station - Ypenburg | 2,654 |

### 5.2.2 Correlation analysis

The correlation analysis is conducted by 4 datasets of various categories, supply, weather, land use data and public transport data, with the aggregated ridership data in the corresponding time interval and spatial units.

If there is only one independent attribute under the dataset, only Pearson's correlation coefficient is examined; otherwise, additional multiple regression analyses are also conducted to detect the impacts of various independent attributes.

The data used for correlation analysis are described in Table 5-5.

Table 5-5 Data description of correlation analysis

| | size | time scope | dynamics | aggregation level |
|---|---|---|---|---|
| **weather** | 336 | 19/06/2021 - 30/07/2021 | temporal | per 3 hours, the whole area |
| **supply** | 42/1638/9030 | 19/06/2021 - 30/07/2021 | temporal and spatial | daily, the whole area/neighbourhood/circle |
| **land use data** | 215/39 | 19/06/2021 - 30/07/2021 | spatial | neighbourhood/circle |
| **public transport** | 215/39 | 19/06/2021 - 30/07/2021 | spatial | neighbourhood/circle |

### 5.2.3 Land use pattern analysis

The intention of land use pattern analysis is to understand the function of different spatial units and support the following demand pattern analysis. The first unit is 44 different neighbourhoods (i.e. wijken) under 8 main districts in The Hague, which is defined by the municipality of The Hague for statistical purposes and will be used in the following demand pattern analysis('The Hague', 2021). Before August of 2021, bondi's operation was executed at the neighbourhood level. The second unit is 215 overlapping circles with a 400m radius in the operational zones of The Hague, which is chosen because of the efficiency of further operational improvement. The motivation of the 400m radius will be discussed in detail in the later section 5.3.1.

The spatial distribution of overlapping circles under 7 districts is presented in Figure 5-8. If the centroid of a circle locates within a neighbourhood, the circle is assigned to this neighbourhood based on the centroid. It is found that Centrum, Scheveningen and Escamp take charge of the majority of these circles.

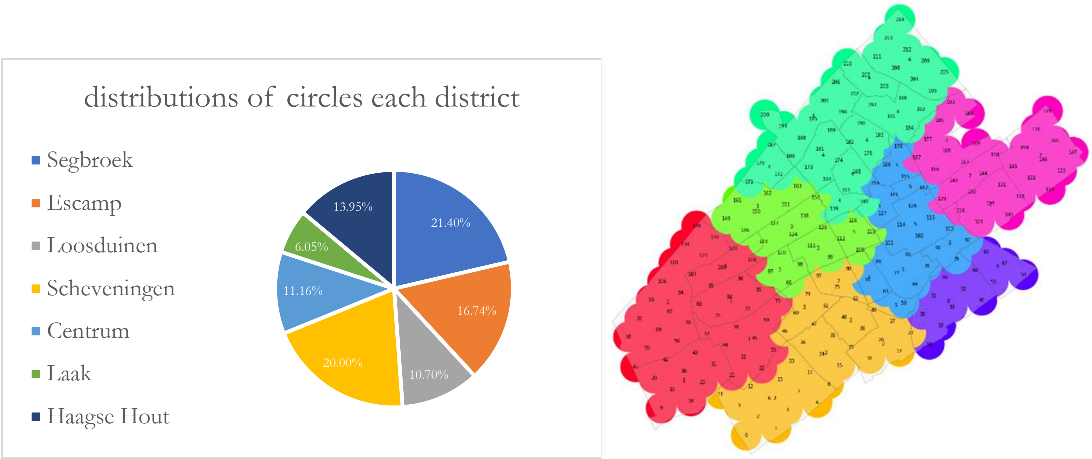



**Figure 5-8 Distribution of circles under administrative districts**

The spatial distribution of POIs is presented under section 5.2.1.

## 5.3 Demand pattern analysis

The descriptive analysis uses the ride records from 19/06/2021 to 30/07/2021; the demand pattern analyses are based on two sets of ride records, which both started from 19/06/2021 while with different ending points at 15/07/2021 and 30/07/2021 respectively. This stems from the iterative process of demand pattern analysis so as to propose strategies applied in a sequence. The first dataset is used to construct a solution aiming at the decrease in ridership after ID implementation requirements in the system; the second time scope targets strategies specific to the holiday season; the third one accounts the general strategies with a circa 3-month input when the users already got familiar and adopted their regular travel behaviour of this e-bike sharing service.

The supply efficiency analysis and average trip distance/duration use a dataset with a different time scope considering the applicability in the timeline of operational strategies. The time scope of this data input is from 19/06/2021 to 24/09/2021.

Table 5-6 illustrates the size of these datasets. The disaggregated type only uses the fleet type to filter the rides outside the operational zone, while the aggregated ones use the spatial boundaries of neighbourhoods/circles to exclude ride records with either the origin or the destination outsides. The circle units result in slightly higher data size since the boundary circles are sometimes outside The Hague with their centroids located in The Hague.

**Table 5-6 Description of size for various datasets used in demand pattern analysis**

| Time scope | disaggregated | Aggregated on the neighbourhood level | Aggregated on the circle level |
|---|---|---|---|
| 19/06/2021 to 15/07/2021 | 2323 | | |
| 19/06/2021 to 30/07/2021 | 3032 | 2978 | 2990 |
| 19/06/2021 to 24/09/2021 | | | 6735 |

### 5.3.1 Determination of spatial analytical units

Free-floating shared bikes is accompanied by rides with disaggregated pick-up and drop points. However, it is sometimes intangible to reveal patterns by these disaggregated points and is also inefficient to focus on each point from the operators' perspective. Therefore, the spatial units are crucial to aggregate data in a reasonable geographical range.

The spatial analytical unit before August 2021 was the neighbourhood, defined as wijken by the municipality of The Hague. It was also used for operational strategies, such as battery swap and vehicle reallocation strategies. The initial demand pattern analysis was therefore conducted at the neighbourhood level.

However, the size of different neighbourhoods is considerably different, with surface ranging from 14 to 337 hectares and population from 0 to 15040 residents in 2008 (Statistiek, 2017).  The variations in size increase the difficulties of operational strategies. For instance, different locations belonging to the same neighbourhood encounter different supply status where ones are oversupplied while the others are undersupplied; however, it will be evaluated as "balanced" status when rides are aggregated at the neighbourhood level. Additionally, it is possible that there are many available vehicles in the central supply point of a neighbourhood while it is too far to walk for the users in other points in the same neighbourhood, especially when the neighbourhood size is large.

Hence, smaller and equally distributed units should be applied as spatial analytical units and overlapping circles are used.

There are two reasons why 400-metres is chosen as the radius of overlapping circles: on one hand, 400 metres approximately correspond to 5-minute walking based on the Dutch average walking speed at 4.5 km/h (Waterstaat, 2019). 5-minute is widely used as the threshold of catchment areas and therefore this concept is also applied here ('Basics', 2011; Sarker et al., 2019); on the other hand, a sensitivity analysis was conducted to ensure the robustness of 400 meters in terms of revealing the demand pattern in hourly clustering.

To examine the robustness of the unit, different radiuses, as presented in Table 5-7, are applied to determine if circles in different sizes exert an effect on the outcome of temporal clustering in the sequent analysis. It is found that the radius of 200 metres generates 860 units in The Hague, which turns out to be computational-inefficiency when conducting agglomerative clustering (which leads to a crack due to memory error even with a 24G RAM).

**Table 5-7 Data size of overlapping circles with different radius**

| | Number of units in The Hague |
|---|---|
| 200m | 860 |
| 300m | 385 |
| 400m | 215 |
| 500m | 137 |
| 600m | 96 |

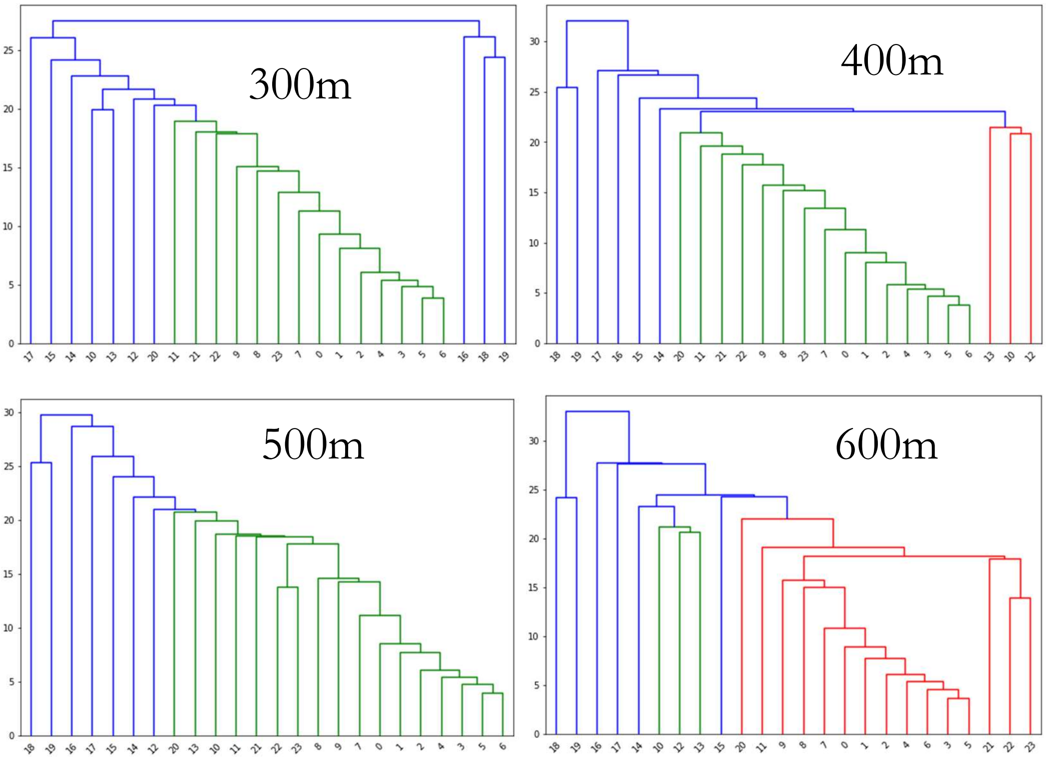



**Figure 5-9 Dendrograms of hourly clustering in overlapping circle units with different radius**

It is observed that with a predefined cluster number as 5, all units (i.e. 300m, 400m, 500m and 600m) present the same clustering results as indicated in Figure 5-9: 16, 17, 18, 19 and other hours. These results prove the robustness of 400m since the variations in radius do not expose a significant impact on the clustering results when the given cluster number is 5. Considering the catchment area, 400 meters is then chosen as the radius of overlapping circles for further analysis.

### 5.3.2 Temporal clustering

The temporal clustering was conducted for both spatial analytical units, and the datasets are introduced as Table 5-6.

### 5.3.3 Supply efficiency analysis

Supply efficiency analysis was analysed by computing the average vehicle idle records per unit as described in section 4.2.5.

Considering the fact that the operational strategies are taken once or twice per week, 3-days is chosen as the time interval for computation. This means the vehicle idle time is 72 hours at maximum.

Two metrics targets in origins and destinations were computed and the size of valid circle units and vehicle idle records is presented in Table 5-8. The valid circle units refer to the units with at least one existing vehicle idle record. If there are no vehicle idle records in the unit, the default maximal vehicle idle time, 72 hours are assigned to these units for visualization convenience.

**Table 5-8 Data description of the size of vehicle idle time**

| | Valid circle units | Vehicle idle records |
|---|---|---|
| Origin-based vehicle idle time | 159 | 9839 |
| Destination-based vehicle idle time | 157 | 9800 |

### 5.3.4 Average travel distance and duration

Average travel distance and duration was computed on a unit basis according to the method mentioned in section 4.2.6. The number of valid units as origins containing ride records is 139 during this period. The units without any ride records are assigned with a default value, 0, for the convenience of visualization in section 6.2.5.

# 6 RESULTS

After depicting the case study, this chapter presents the results from the methods application introduced in chapters 5.2 and 5.3. First, the results of data analysis are reported and visualized, including the distribution of the main trip characteristics, as well as the examination of correlation analysis and spatial description (i.e. land use pattern analysis). Then, the outcomes of demand pattern analysis are illustrated and discussed, followed by the derivation of recommended operational strategies. Finally, these operational strategies are implemented and the following impacts are evaluated and inspected.

## 6.1 Data analysis

The data preparation was already described in chapter 5.2. In this section, the results of the correlation analysis are presented.

### 6.1.1 Correlation analysis

A summary table of Pearson's correlation coefficient for all abovementioned variables is listed in Table 6-1. It can be identified weather has a lower correlation level out of all 4 types of variables and land use pattern has the highest linear correlation with the ridership level. Unexpectedly, the precipitation is observed to be of low correlation with the ridership.

**Table 6-1 Summary of Pearson's correlation coefficient**

| Variable/Spatial level | | the whole area | neighbourhood | circle |
|---|---|---|---|---|
| supply | | 0.43 | 0.18 | 0.49 |
| weather | temperature | 0.16 | - | - |
| | precipitation | -0.03 | - | - |
| | humidity | -0.31 | - | - |
| land use | #sustenance | - | 0.92 | 0.76 |
| | #office | - | 0.9 | 0.61 |
| | #entertainment | - | 0.92 | 0.74 |
| | #healthcare | - | 0.61 | 0.73 |
| | #education | - | 0.74 | 0.33 |
| public transport | #stops | - | 0.34 | 0.26 |
| | ridership | - | 0.66 | 0.29 |

To have deeper understandings of how these variables affect the ridership levels, several multiple regression analyses were conducted on weather, land use and public transport categories.

For the regression of weather category demonstrated in Table 6-2, ride hour is also included as an additional attribute as it is the ridership level is also highly correlated with the ride hour. Only humidity under the weather category is observed to be significant at 95% confidence level to this regression and it exerts a strong negative effect on the ridership. Temperature is found to be insignificant under the regression form, which is associated with the gentle variations among the dataset.

**Table 6-2 Regression results of weather factors**

| R-squared | 0.178 | |
|---|---|---|
| | Coefficient | Standard error |
| Constant | 26.35** | 8.13 |
| Temperature (°C) | -0.25 | 0.29 |
| Precipitation (mm) | 0.40 | 0.49 |
| Humidity (%) | -23.97** | 6.13 |
| Hour_start | 0.49** | 0.086 |

**: significant at 95% confident interval; *: significant at 90% confident interval

For land use data, two regression analyses were conducted, on the level of the neighbourhood and 400m overlapping circle respectively, as seen in Table 6-3 and Table 6-4. The neighbourhood level is observed to have a higher model fitness, indicated by R-squared. All variables, except for the amount of healthcare and education, are found to be significant at a 95% confidence interval in both units. It is found that sustenance facilities have positive effects at a similar magnitude of offices, and entertainment amenities are proven to exert the highest positive impacts in both cases. Surprisingly, educational facilities are found to influence ridership in a negative way.

**Table 6-3 Regression results of POIs on the neighbourhood level**

| R-squared | 0.899 | |
|---|---|---|
| | Coefficient | Standard error |
| Constant | 41.45** | 20.43 |
| \|sustenance\| | 1.80** | 0.80 |
| \|office\| | 2.13** | 0.83 |
| \|healthcare\| | -3.64 | 4.17 |
| \|entertainment\| | 12.51** | 4.98 |
| \|education\| | -38.33 | 25.06 |

**: significant at 95% confident interval; *: significant at 90% confident interval

**Table 6-4 Regression results of POIs on the circle level**

| R-squared | 0.693 | |
|---|---|---|
| | Coefficient | Standard error |
| Constant | 4.85* | 2.60 |
| \|sustenance\| | 1.89** | 0.56 |
| \|office\| | 1.09** | 0.30 |
| \|healthcare\| | -2.50 | 4.59 |
| \|entertainment\| | 11.52** | 1.47 |
| \|education\| | -16.83** | 6.83 |

**: significant at 95% confident interval; *: significant at 90% confident interval

When it comes to public transport, the number of public transport stops is only significant on the neighbourhood level at a 90% confident interval while public transport ridership is significant at a 95% confident interval in both units, inspected by Table 6-5 and Table 6-6. From the result, it can be observed that the number of stops has a more substantial effect on the ridership while it is generated from the difference in the range of these two attributes. Taking the regression on the neighbourhood level as an example, the maximum number of public transport stops is 25 in the dataset while the maximum PT ridership is 24293. After multiplying the coefficient, it contributes to 266 and 972 rides separately, confirming that ridership certainly has a higher positive effect on the ridership.

**Table 6-5 Regression results of public transport on the neighbourhood level**

| R-squared | 0.486 | |
|---|---|---|
| | Coefficient | Standard error |
| Constant | -27.90 | 57.01 |
| \|PT\| | 10.62* | 5.47 |
| PT ridership | 0.04** | 0.01 |

**: significant at 95% confident interval; *: significant at 90% confident interval

**Table 6-6 Regression results of public transport on the circle level**

| R-squared | 0.111 | |
|---|---|---|
| | Coefficient | Standard error |
| Constant | 9.67 | 5.56 |
| \|PT\| | 7.82** | 2.58 |
| PT ridership | 0.00** | 0.00 |

**: significant at 95% confident interval; *: significant at 90% confident interval

### 6.1.2 Land use pattern analysis

The results of land use pattern analysis based on two types of spatial analytical units are illustrated in this subsection. The neighbourhood level includes the analyses on the income level, residential statistics, ownership of car/motors while the counterpart of 400m overlapping circles are absent because of a lack of data.

- In the neighbourhood unit

Comparing the spatial distribution of average disposal income per resident and number of residents per neighbourhood in Figure 6-1, it can be found that the northern neighbourhoods are richer and are with fewer residents while the southern counterparts are denser with lower income levels.

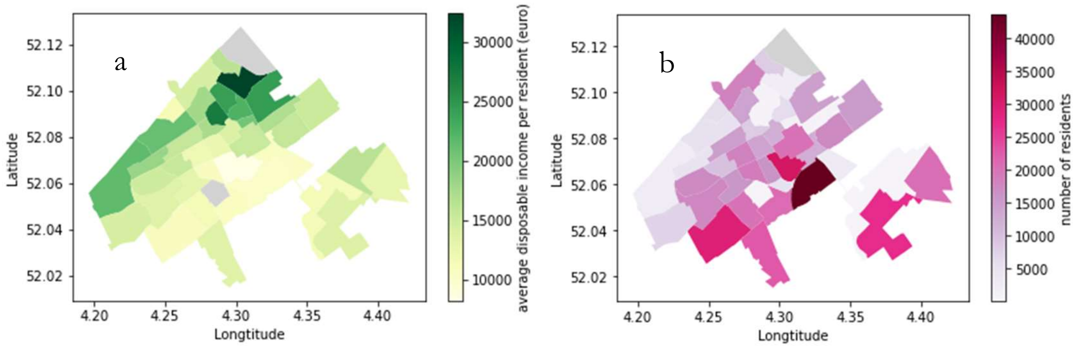



**Figure 6-1 Heatmaps of income level (a); and population (b)**

Turning to the ownership of private cars and motors as shown in Figure 6-2, the outskirt neighbourhoods tend to own private cars, and a similar observation is seen in terms of the ownership of motors. However, motors are less common in The Hague, compared to private cars. Unfortunately, the ownership of bicycles in the neighbourhood unit is unavailable.

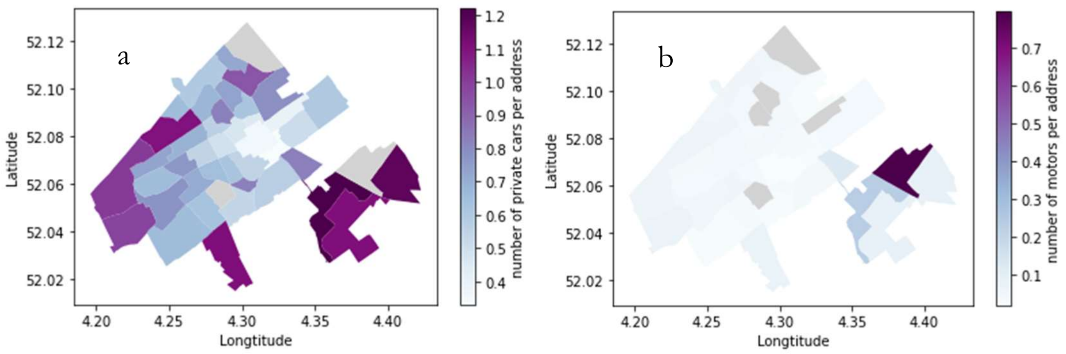



**Figure 6-2 Heatmaps of the ownership of private cars (a); and motors (b)**

For public transport facilities, which are tram and bus services provided by HTM in The Hague, the distribution of stops is relatively equal, except that there is a clear lack in the outer greenspace-oriented zones, such as Kijkduin en Ockenburgh, Kraayenstein and Oostduinen, and the district of Leidschenveen-Ypenburg.

Furtherly, the proportion of different amenities were analysed as described in 4.1.3. For the sustenance facilities, the north area of Scheveningen area next to the beach, and the district of Escamp, are engaged with a high proportion of sustenance amenities out of 5 main categories of POIs. While the office distributions vary somewhat sparsely, where Forepark, HaagseBos, Zorgvliet, Van Stolkpark, and Groente-en Fruitmarkt stand out among these neighbourhoods. The percentages of healthcare, entertainment and education amenities are rather low compared to sustenance and office. However, it is found that Ypenburg is occupied with a fairly high portion of healthcare and education amenities compared to other areas. Also, healthy services are prevalent in Kraayenstein, Waldeck and Leyenburg. Generally, healthy amenities are prone to locate in residential areas instead of the central areas. Interestingly, Kraayenstein also owns a high percentage of entertainment facilities and Vogelwijk is the most entertaining area when it comes to the percentage of entertainment facilities out of all facilities. Mariahovev en Marlot is also distinguished, with lots of sports facilities. For educational facilities, the absolute value is quite low since only higher education is considered.

N.B.: these figures only indicate the function at neighbourhood levels instead of the absolute amount of a particular type of amenities in this area. Taking Duindrop as an example, it presents a high share of office at 66.7% while there are only 4 offices in this area. The high proportion is triggered by a lack of other facilities but it is still defined as an office-oriented area since office facilities largely outweigh other facilities (i.e., sustenance, entertainment, healthcare and education) in this area.

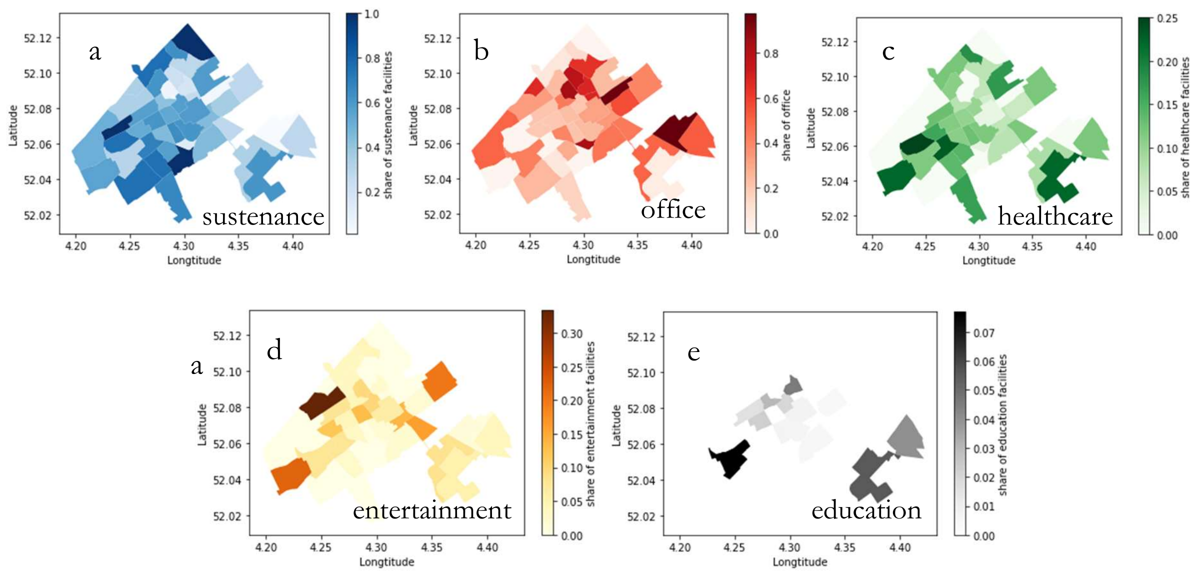



**Figure 6-3 Heatmaps of POIs on the neighbourhood levels**

It is seen in Figure 6-3 that neighbourhoods in the centrum are prone to be relatively multifunctional but prevalent with recreational amenities, while other neighbourhoods are usually dominant with (a) certain types(s) of amenities.

According to the definition dependent on the share of different facilities, as described in chapter 4.1.3. Sustenance amenities, such as bars and restaurants, and entertainment facilities, such as theatres, sports facilities, are all classified as an indicator of recreation. If the proportion of office or recreation is higher than 50%, this unit is then defined as an office-oriented or recreational-oriented functional unit. They are visualized in Figure 6-4.

In general, recreational zones are relatively concentrated in Centrum, the beach area, and the outer residential zones are also prevalent by recreational facilities, mainly the sustenance ones. The office-dominant neighbourhoods mainly spread between Centrum and Scheveningen districts, as well as Haagse Hout.

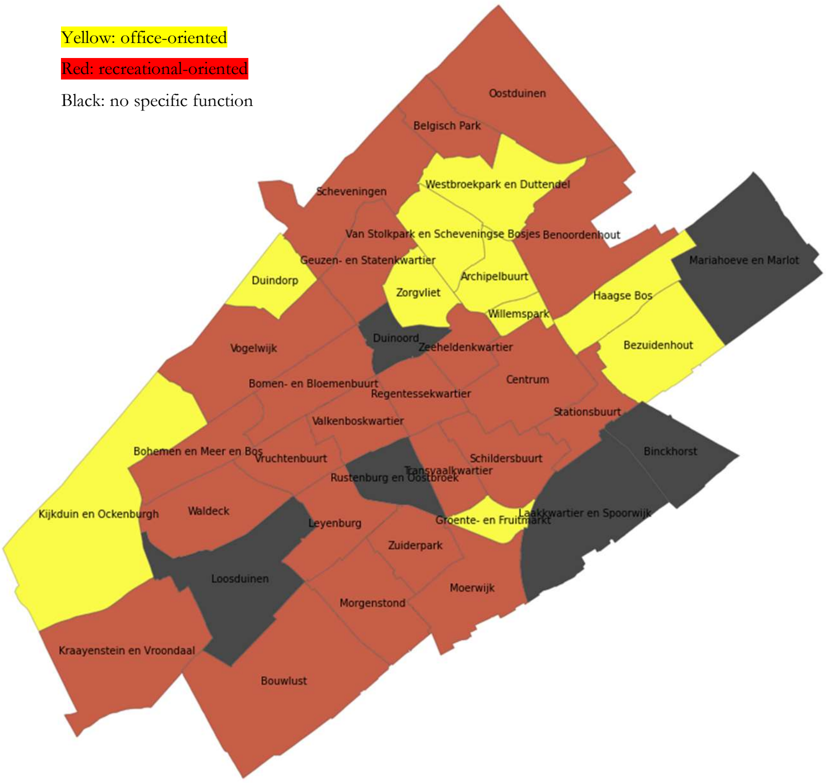



**Figure 6-4 Neighbourhoods with functions**

- In the 400m overlapping circle unit

The same analyses were conducted on the overlapping circle level as demonstrated in Figure 6-5. It presents similar patterns as the neighbourhood one while the units are smaller and thereby the absolute amounts of facilities are lower.

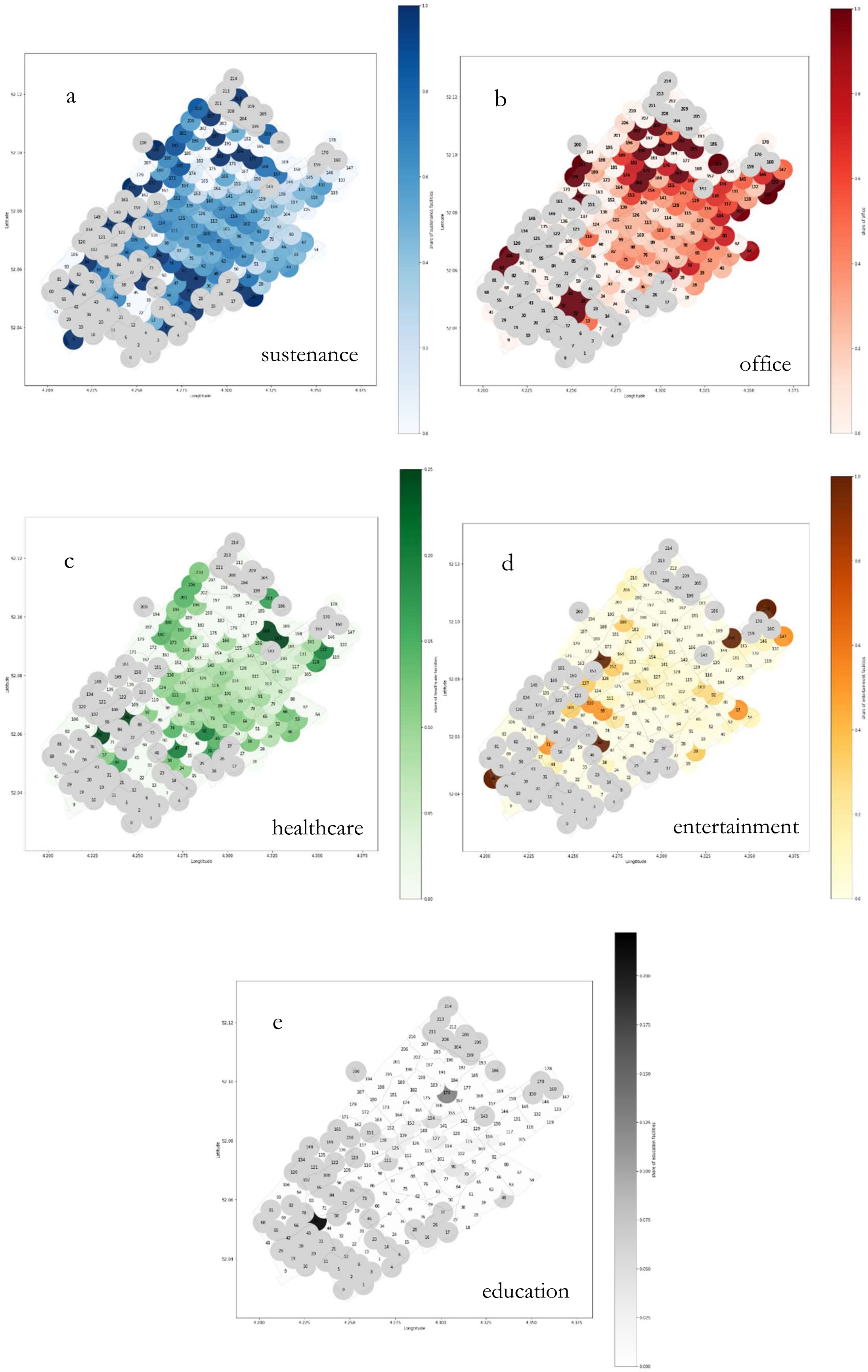



**Figure 6-5 Heatmaps of POIs on the circle levels**

Furthermore, the main function of different units should be defined in order to analyse the demand pattern, viz., the office and recreational zones.

The definition of recreational-oriented and office-oriented zones are the same as before. If the proportion of office or recreation is higher than 50%, this unit is then defined as either an office-oriented or recreational-oriented functional unit as Figure 6-6 shows. The results are somewhat different from its counterpart on the neighbourhood level, especially for the outer areas. Different locations within the same neighbourhood also demonstrate heterogeneity in functions.

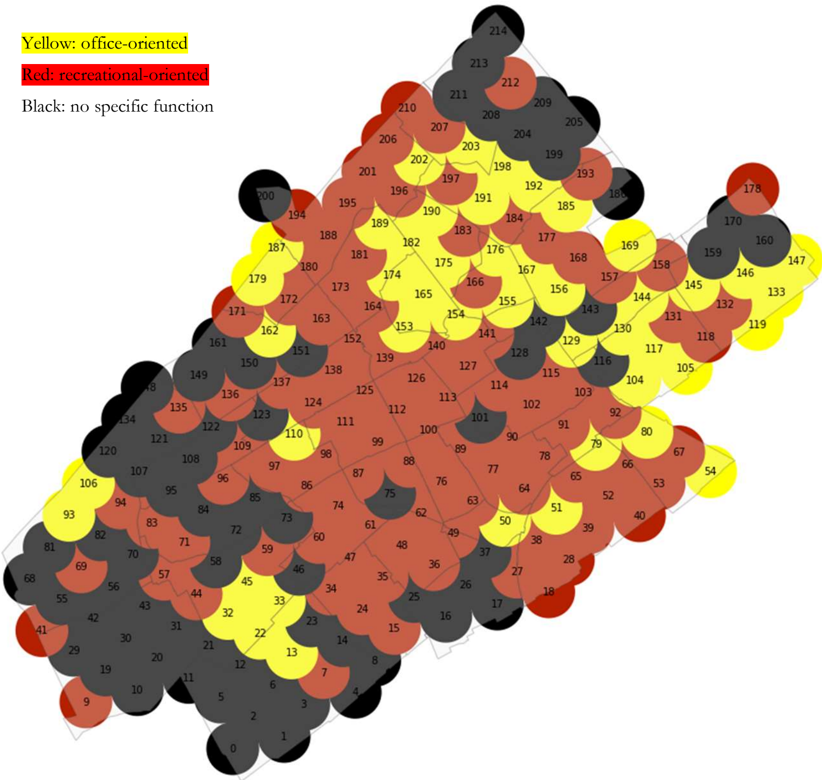



**Figure 6-6 Circles with functions**

It is observed that the recreational areas are mainly concentrated in the central and southwestern part of The Hague, especially in Centrum and along the beach area, still. However, Laak is also engaged with many recreational-oriented areas. For workplace function, it plots relatively sparsely. There are quite some office-oriented areas in Scheveningen, Centrum and Haagse Hout.

## 6.2 Demand pattern analysis

The results of demand pattern analysis are demonstrated in this section, including the output of descriptive analysis, general demand pattern analysis as the first-round demand pattern analysis using the dataset from 19/06/2021 to 15/07/2021 and the second-round demand pattern analyses in two levels with temporal clustering involved using the second dataset from 19/06/2021 to 30/07/2021. After that, supply efficiency analysis and average travel distance/duration analysis are illustrated with the data input until 24/09/2021. It is followed by a summary concluding the operational strategies derived from these outcomes.

### 6.2.1 Descriptive analysis

This section presents the descriptive statistics of the essential data input, ride records, with a time scope between 19/06/2021 to 30/07/2021.

➢ Ridership

Examined the ridership in Figure 6-7, for the first 4 days, the ridership was exceptionally high because the service is free of charge during this time period. Afterwards, the ridership continually declined, with several fluctuations though. The ID verification, installed on the 17th of July, which aims to avoid misbehaved rides, has exerted a negative effect on the ridership, decreased the ridership to circa 30, which also arises from the bad weather condition. Followingly, reallocation strategies obtained from the data-driven methods have been implemented, which will be discussed in detail in later section 6.2.6, and they contribute to the rebounce of ridership to 60 rides per day approximately.

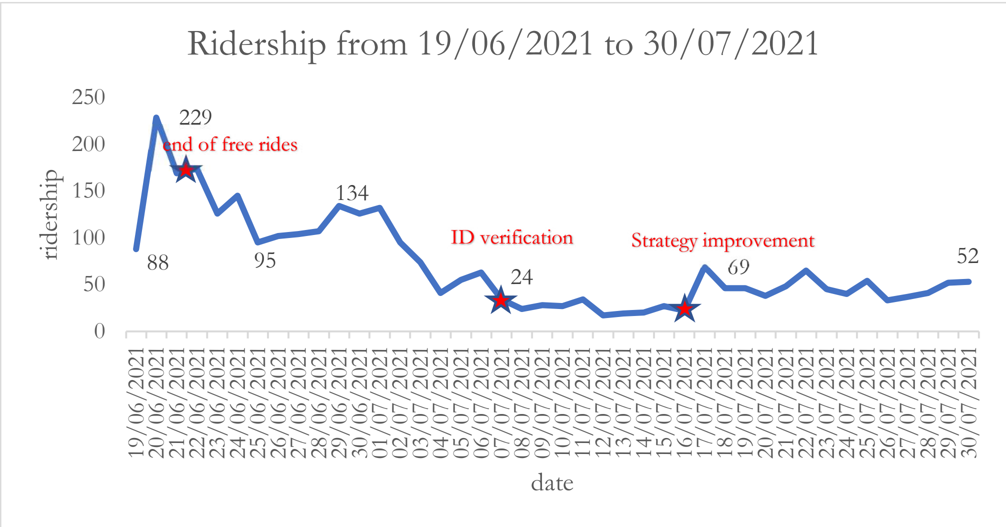



**Figure 6-7 Ridership overview**

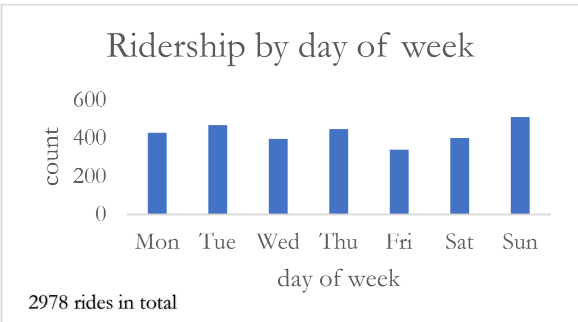



**Figure 6-8 Ridership by day of week**

For ridership in terms of day of the week (see Figure 6-8), there is no obvious difference between the ridership per day, indicating that the ride day does not insert a noticeable effect on the ridership while its impacts on the demand pattern will be described in section 6.2.3.

Additionally, the demand does not either present the typical morning peak pattern as seen Figure 6-9. The ridership witnesses a growth from the morning in general and there is an evening peak at around 6 p.m. After the peak, the ridership starts to decline. A detailed analysis is conducted by temporal clustering considering the spatiotemporal pattern of the demand in the chapter of 6.2.3.

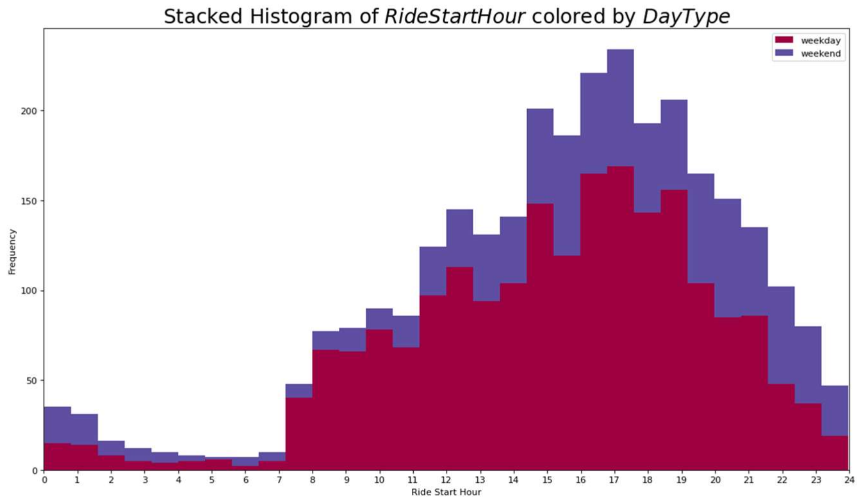



**Figure 6-9 Distribution of ride start hour overtime of a day**

➢ Trip duration vs trip distance

The histogram of ride distance shown in Figure 6-10, illustrates that the majority of trips are shorter than 10 km and the peak is between 0 to 2 km. The average travel distance is 3.65 km and the median is 2.67 km.

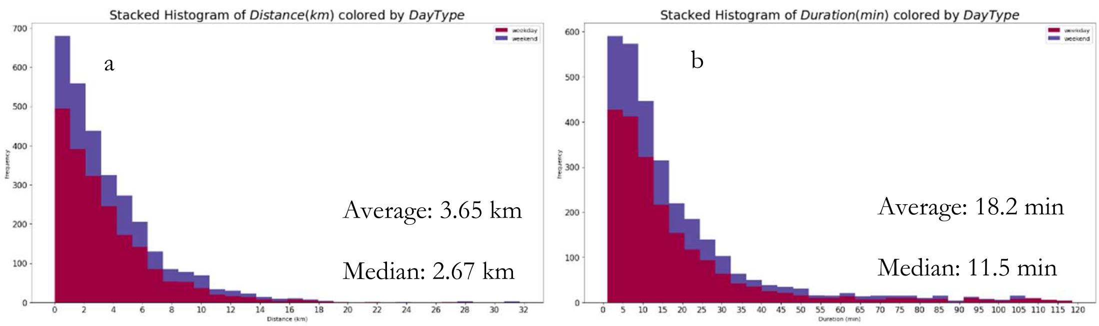



**Figure 6-10 Distribution of trip distance in km (a); and trip duration in minute (b)**

Meanwhile, the inspection of trip duration shown in Figure 6-10 indicates that most rides are shorter than 50 mins and the peak is between 0 to 10 minutes. The average travel duration is 18.2 minutes while the median is 11.5 min, next to the peak range.

### 6.2.2 General demand pattern analysis

General demand pattern analysis firstly deals with the pattern by pick-up and drop-off points on a disaggregated level. Then, the same procedures are applied for ride records, while aggregated on the neighbourhood and 400m overlapping circles respectively.

➢ Disaggregated analysis

Figure 6-11 presents the disaggregated spatial plots of all ride records in this period in The Hague. It is observed that the central zones are the most prevalent areas no matter in terms of pick up and drop points. While it is also worth noting that the demand is also dependent on the supply as introduced in the previous part 6.1.1. Besides, there are also some zones occurring to be attractive, such as the zones along the beach, which provide entertainment activities. The neighbourhood is used as the spatial unit in the next section.

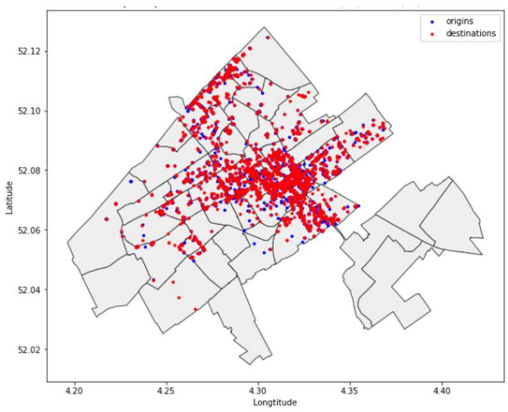



**Figure 6-11 Spatial plots of 2978 rides between 19/06/21 to 30/07/21 in The Hague**

➢ First-round demand pattern analysis on the neighbourhood level (19/06/2021 – 15/07/2021)

The first-round demand pattern analysis was conducted at the neighbourhood level. The results are presented in Figure 6-12. It is observed that Centrum is always the hottest spot in terms of departures and arrivals and the central areas are favoured compared to other areas. Besides, the beach area also catches attention, standing out as the heat spots.

Comparing the difference between arrivals and departures per unit, it can be found that the arrivals outweigh departures in the majority of neighbourhoods, despite Centrum and Bezuidenhout. However, Belgisch Park and Oostduinen situated along with the beach witness overwhelming arrivals, observed from Figure 6-13.

Thereby, it is recommended to rebalance the bikes from the beach areas to Centrum and Bezuidenhout. Another operational strategy is to place bikes in batches when rebalancing the bikes, making the bikes more noticeable and attractive instead of placing them separately in the initial launching phase.

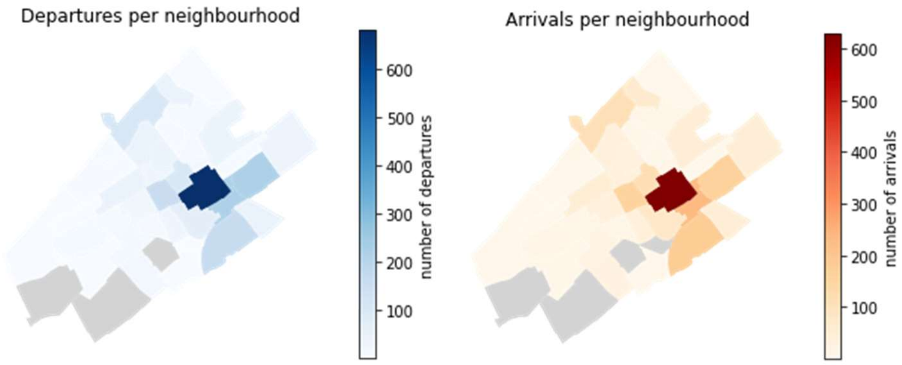



**Figure 6-12 Heatmaps of departures and arrivals of the first-round**

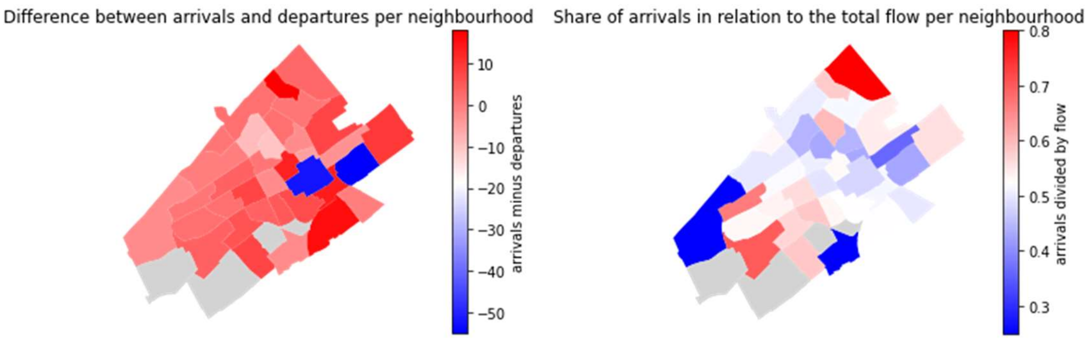



**Figure 6-13 Heatmaps of difference between departures and arrivals of the first-round**

- Second-round demand pattern analysis on the neighbourhood level (19/06/2021 – 30/07/2021)

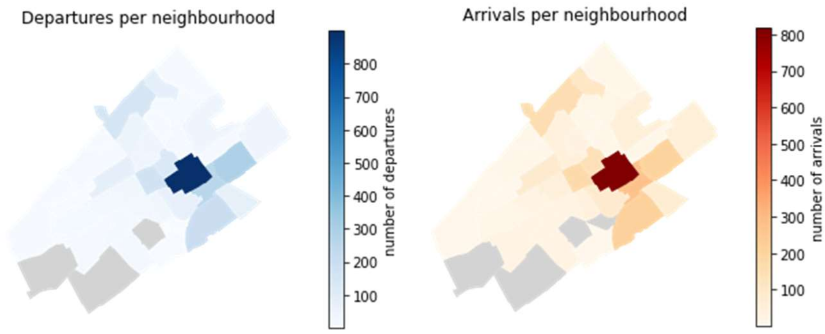



**Figure 6-14 Heatmaps of departures and arrivals of the second-round in neighbourhoods**

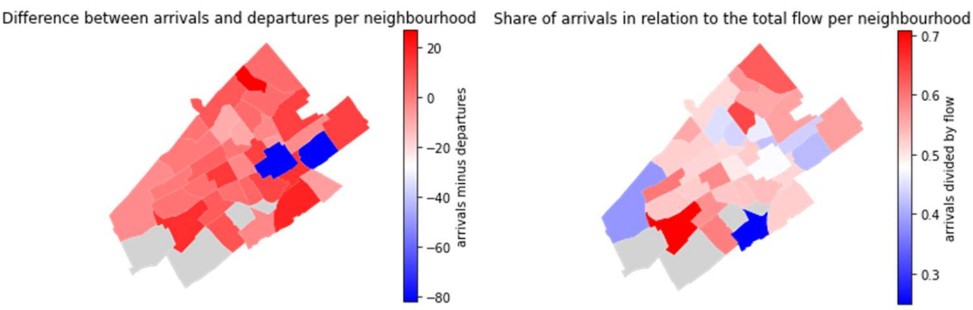



**Figure 6-15 Heatmaps of difference between departures and arrivals of the second-round in neighbourhoods**

It is observed from Figure 6-14 that the centrum is the most popular hotspot, without doubt, followed by Laakkwartier en Spoorwijk, Bezuidenhout, Scheveningen and Belgisch Park neighbourhoods in terms of the absolute flow (i.e., the sum of departures and arrivals). Considering the gap between arrivals and departures in Figure 6-15, the centrum and Bezuidenhout areas are the only two neighbourhoods where the departures exceed arrivals and Van stolkpark is the most balanced area. While there is a tendency that people prefer to end their trips in Belgisch Park and Laakkwartier en Spporwijk instead of embarking on their trips there. This fact implies that the reallocation in these two areas is lack. When it comes to the arrival ratio, it is found that the arrivals outweigh the departures to a large extend in Loosduinen and Van stolkpark while the departures highly overshadow the arrivals in Moerwijk and Kijkduin en Ockenburgh with low absolute amounts of trips, though. Contrarily, the arrival ratios in the centrum are usually around 50% in the central areas despite their striking difference between arrivals and departures. While their flow is quite high and therefore the gap does not contribute to an obvious portion in ratio in these areas.

Tracking the overview chord diagram in Figure 6-16 which presents how bikes flow among various neighbourhoods, it is found that the foremost trips start and end up in the same area. There are still some outstanding OD pairs, however. Taking trips originating at the centrum area as the example, Stationsbuurt with mixed functions, Bezuidenhout and Laakkwartier en Spoortwijk where workplaces are the dominant, are the popular destinations from the centrum. It is highly likely that these trips are commute trips. Considered another heat spot, Scheveningen, it is observed that the centrum and Belgisch Park (which is also a beach area) are also popular destinations for trips starting from Scheveningen.

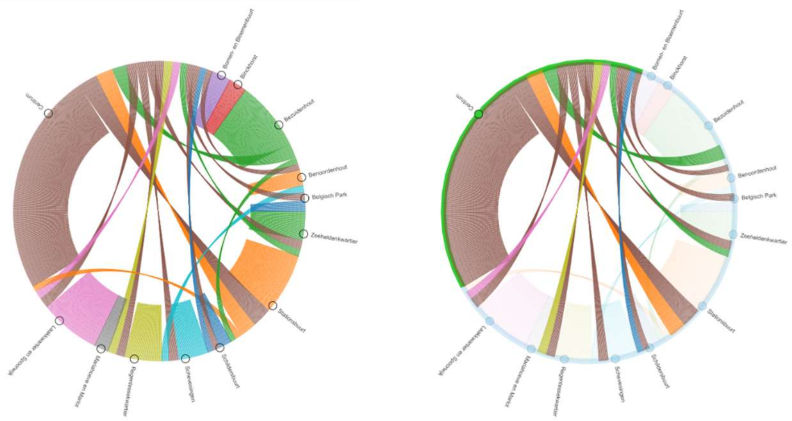


**Figure 6-16 Chord diagram of the second-round in neighbourhoods**

➢ Second-round demand pattern analysis on the 400m-circle level (19/06/2021 – 30/07/2021)

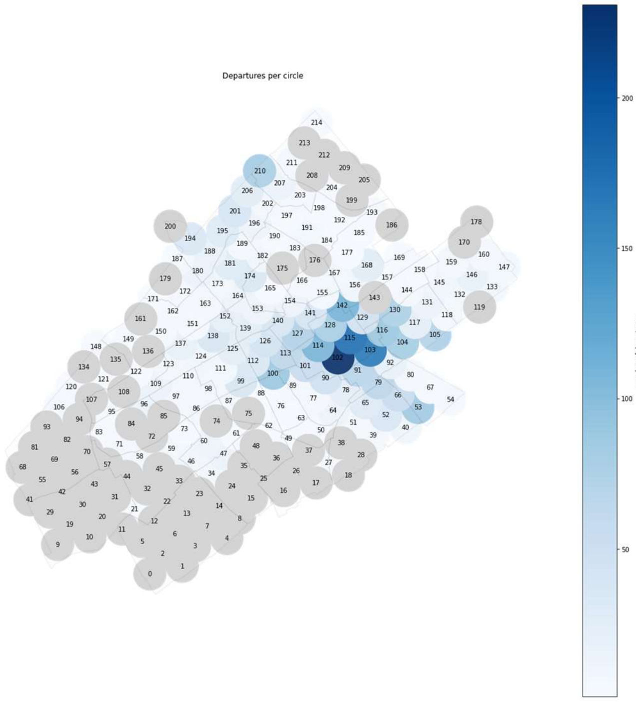



**Figure 6-17 Heatmap of departures of the second-round in circles**

Generally, the hotspots are concentrated in the central areas and some areas nearby, such as the circles located in Bezuidenhout (i.e., 104, 105, 116 and 130 which is between Haagse Bos and Bezuidenhout) and the area 100 situated among Regentessekwartier, Transvaalkwartier and Schilderswijk, seen in Figure 6-17. The beach area is also welcomed, with a sub hotspot, zone 210, near De Pier. However, the southeaster areas are the least dense, with almost no departures in those circles.

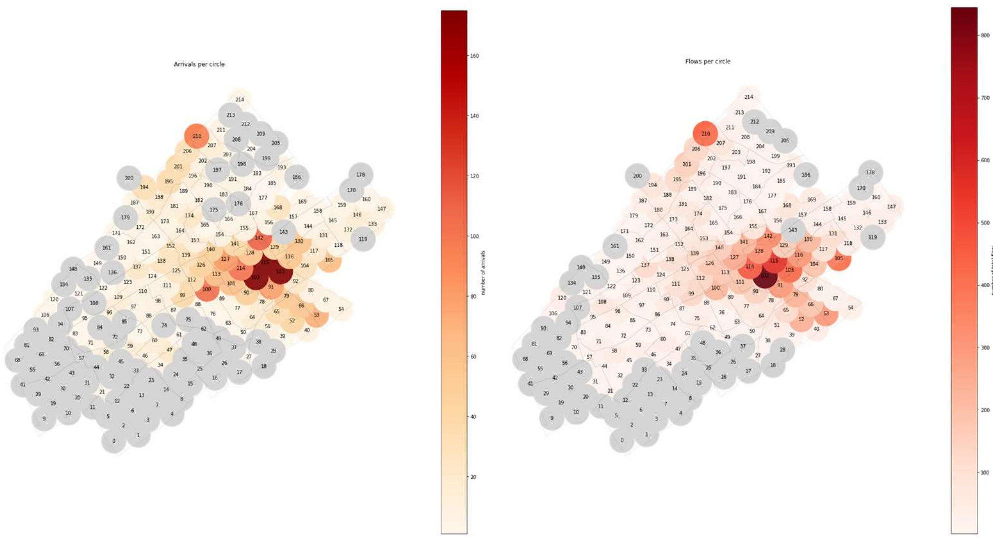



**Figure 6-18 Heatmaps of arrivals and total flows of the second-round in circles**

A similar pattern is observed from the distribution of arrivals as well as the total flow in Figure 6-18.

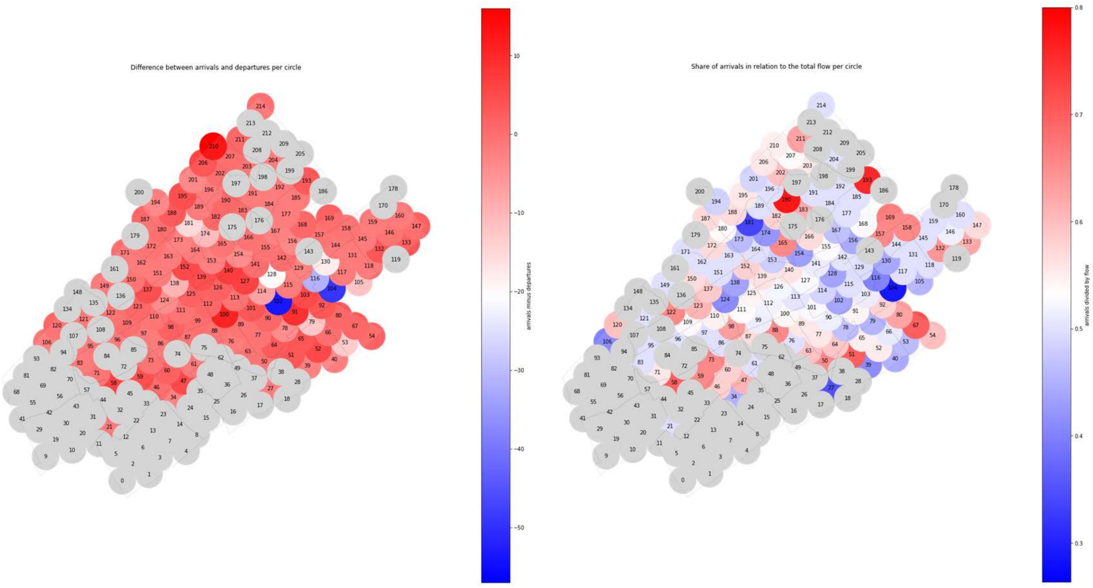



**Figure 6-19 Heatmaps of difference between departures and arrivals of the second-round in circles**

Displayed in Figure 6-19, the majority of hotspots mentioned above are dominant with departures, despite the popular zones near the beach, presenting a slightly arrival-oriented tendency.

In general, the trips spread from the centre to the outer areas.

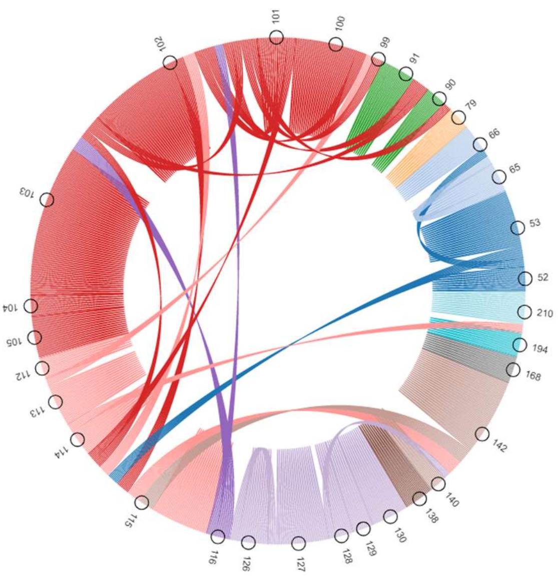



**Figure 6-20 Chord diagram of the second-round in circles**

From the chord diagram in Figure 6-20, the most popular OD-pairs are still quite concentrated in the central areas, especially in Centrum, Stationsbuurt and Laak district, where most of them are recreational-oriented zones because of the dominance of sustenance amenities. However, some units near the beach also distinguish, such as 210, nearing the pier.

### 6.2.3 Temporal clustering

Clustering analyses are conducted on both neighbourhood and overlapping circles levels. The detail analysis of the neighbourhood units can be seen in Appendix A. In this part, only the results on the overlapping circles are presented.

- Hourly clustering

Given 5 clusters, a decent number for clustering, hour 16, 17, 18, 19 and other periods emerge to be 5 separate classes by agglomerative hierarchical clustering with ward method by Euclidean distance similarity as the dendrogram indicates in Figure 6-21.

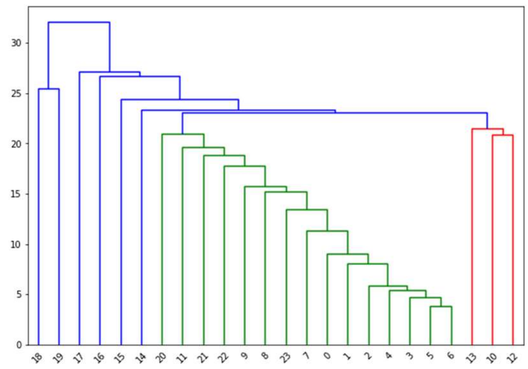



**Figure 6-21 Dendrogram of hourly clustering on the circle level**

1. The first peak hour: 16

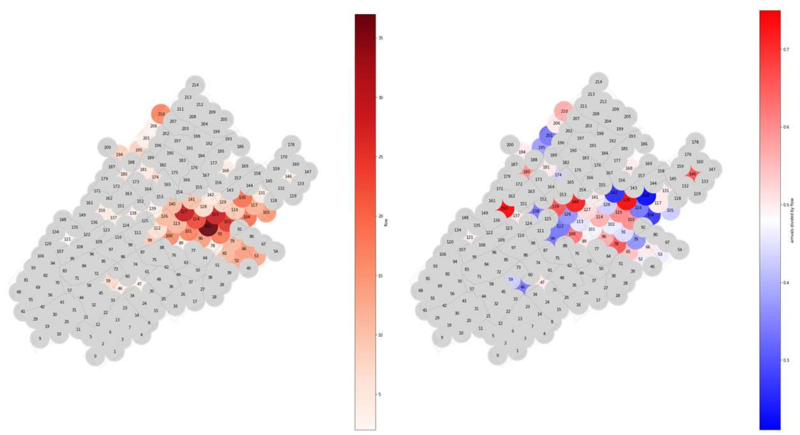


**Figure 6-22 Heatmaps of total flows and arrival ratio of the 1st cluster in circles**

16 is the first peak hour. For this period, the rides gather in Centrum and the beachside area as seen in Figure 6-22. Besides these two areas, the rides are relatively sparse in the outer units at a low magnitude. The most prevailing spot is 102, followed by 127 and 115 which belong to Centrum and are recreational-oriented areas. Worth noting, the other units of attention are generally recreational-oriented zones despite the units located in Bezuidenhout, which is 130, 117, 105 and 104 where the main function is office. Taking the function into account, these four office-oriented units present a departure-dominant flow in

the first peak hour. As for the central area, the arrival/departure pattern is rather heterogeneous and vary per unit.

Tracking the chord diagram in Figure 6-23, the main components of rides flow among the central units. Taking the essential unit, 102 as an example, the majority of rides originating at this unit also ends at this area, while 127 and 90 situate at Zeeheldenkwartier and the border of Centrum and Schildersbuurt are also preferred destinations, consistent with the average travel distance in the descriptive analysis part. Taking another heat spot, 210 near De Pier, as the representative of the beach area, the destination composition is relatively distributed, from 130 in Bezuidenhout which is far away from the beach, to 174 located in Zorgvilet with a reasonable distance to the beach and 196, the unit nearby in the other side of the beach, near the harbour.

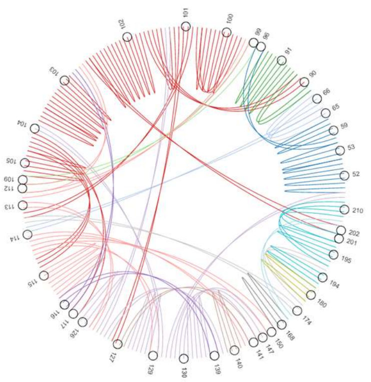


**Figure 6-23 Chord diagram of the 1st cluster in circles**

2. The second peak hour: 17

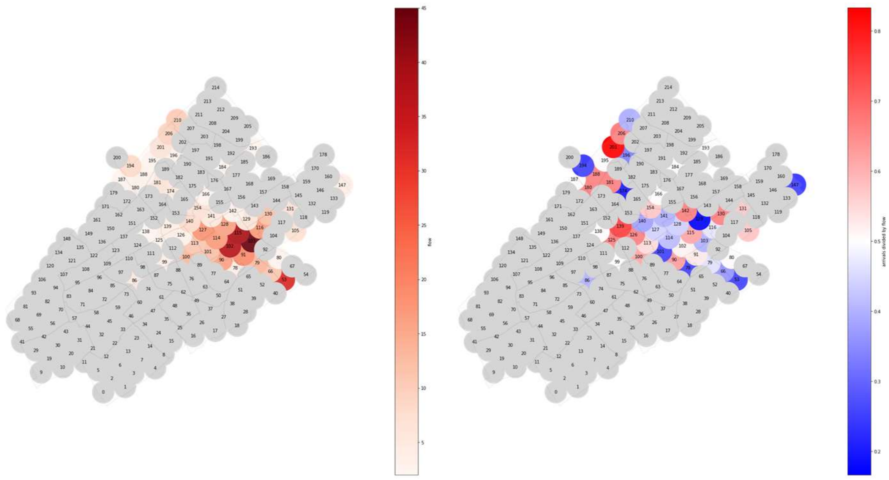


**Figure 6-24 Heatmaps of total flows and arrival ratio of the 2nd cluster in circles**

In terms of the distribution of flow seen in Figure 6-24, this period demonstrates a similar pattern as the first peak hour. The central areas are still the hottest. However, the most attractive unit shifts to the unit 103 which nears the train station (i.e., Den Haag HS) and this unit also reveals a slight departure tendency, and the alike phenomenon appears in unit 129 which is close to Den Haag Centraal station, with an even higher departure ratio. This suggests that during this hour, travellers prefer to pick up the bikes near the train station, which is conceivable to serve the last mile of their trips, compensating the train trips. The central units are generally balanced between departures and arrivals, with slightly more departures.

Additionally, the beach units, in general, attract more trips than the last hour, mainly as the destinations rather than trip origins, while the residential areas are less appealing.

Considering the chord diagram in Figure 6-25, the most frequently occurring unit is 103, located between Centrum and Stationbuurt near the HS station, and the destinations from this unit are quite sparse, covering the other units in the central area as well as the beach area. Surprisingly, the unit, 53, located between Laakkwartier en Spoorwijk and Binckhorst, is also found to be a popular origin, whose destinations are mainly the units in Centrum district.

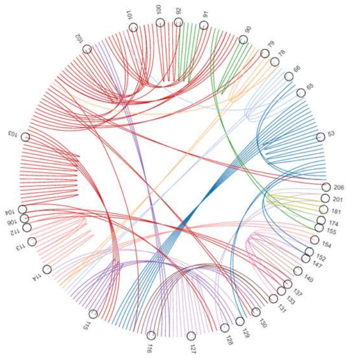


**Figure 6-25 Chord diagram of the 2nd cluster in circles**

3. The first transition hour: 18

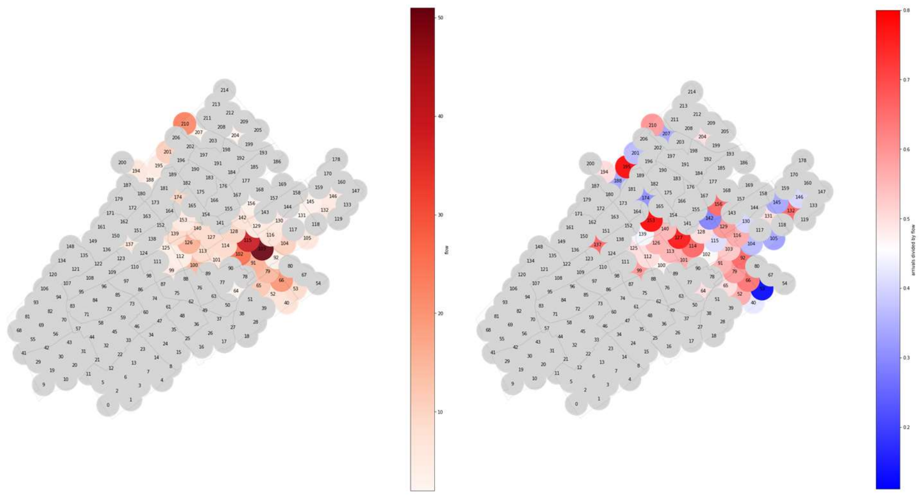


**Figure 6-26 Heatmaps of total flows and arrival ratio of the 3rd cluster in circles**

In this first transition hour, the units in Bezuidenhout and Mariahoeve en Marlot only catch attention with a low flow, though, observed from Figure 6-26. Some of them (131 and 132) are recreational-oriented and the rest of them (105, 104, 130, 145 and 146) are office-oriented units. Office-oriented units present a predilection of more departures while the recreational ones tend to attract arrivals instead.

In this period, the central area is quite balanced, with a marginal dominance of arrivals, especially for the north-western zones, between Regentessekwartier and Centrum. The unit, 53, is still overwhelmed by departures but the general flow in this unit is less significant compared to the previous hour.

In terms of the chord diagram in Figure 6-27, unit 103 is the most prevalent origin/destination, with an almost balanced status in the magnitude of departures and arrivals. It is recognised that the trips in this unit is either the first-mile trip or the last mile trip for train trips.

Additionally, unit 174, an office-oriented unit situated between Geuzen en Statenkwartier and Zorgvliet, is also grown to be important in this area, as a popular origin with the majority of trips ending at the beach area. The beach unit, 210 near the pier, is also a welcomed destination of trips generated in Haagse Hout and Centrum.

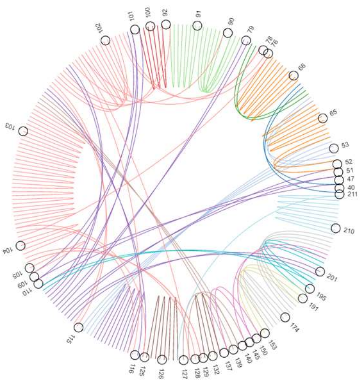


**Figure 6-27 Chord diagram of the 3rd cluster in circles**

4. The second transition hour: 19

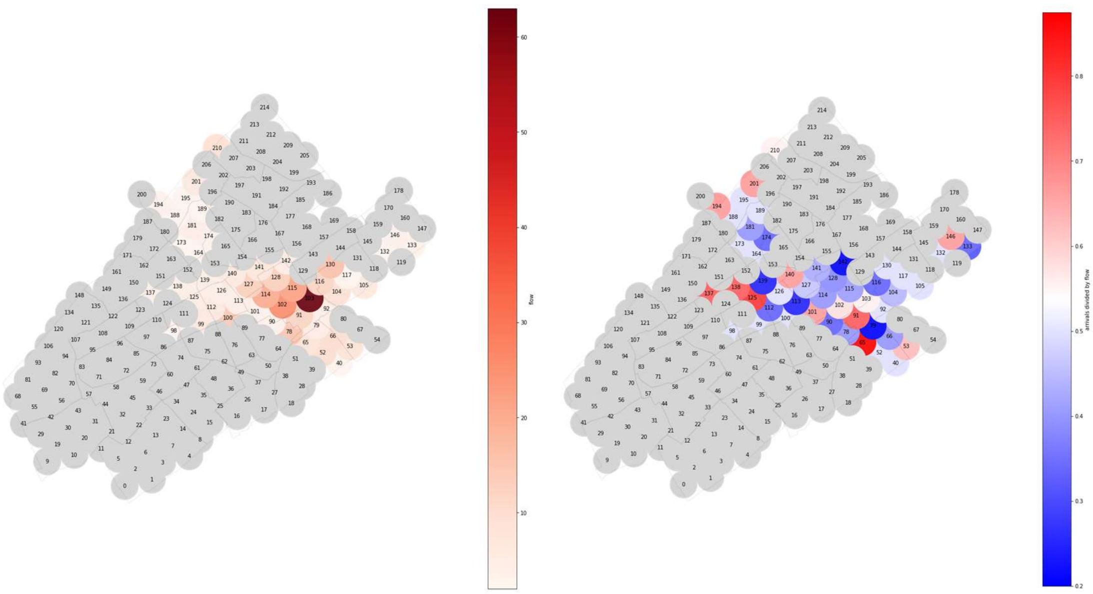


**Figure 6-28 Heatmaps of total flows and arrival ratio of the 4th cluster in circles**

The unit near the train station, 103, is still the most popular unit in this time period, with almost equal departures and arrivals, echoing the deduction that these trips serve the first/last mile supplementary train trips as presented in Figure 6-28. Rides are relatively distributed to other units while the central areas are still prevalent as all the time. In this time period, the central units are generally dominated with more departures while the arrivals are towards the outer units. The same implication is found in the chord diagram in Figure 6-29.

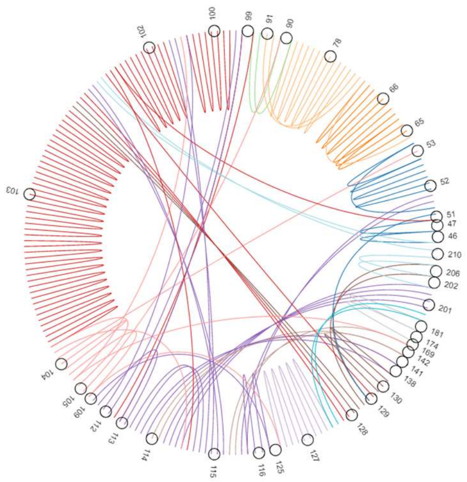



**Figure 6-29 Chord diagram of the 4th cluster in circles**

5. Others: 20:00-15:59

The off-peak period ranges from 20:00 to 16:00, covering the longest period of a day and therefore the ride distribution covers the widest area consequently, especially the residential zones in the outer units, as Figure 6-30 shows. Invariably, the central areas are still quite essential, followed by units next to Centrum district which is located in Serbroek and Laak district.

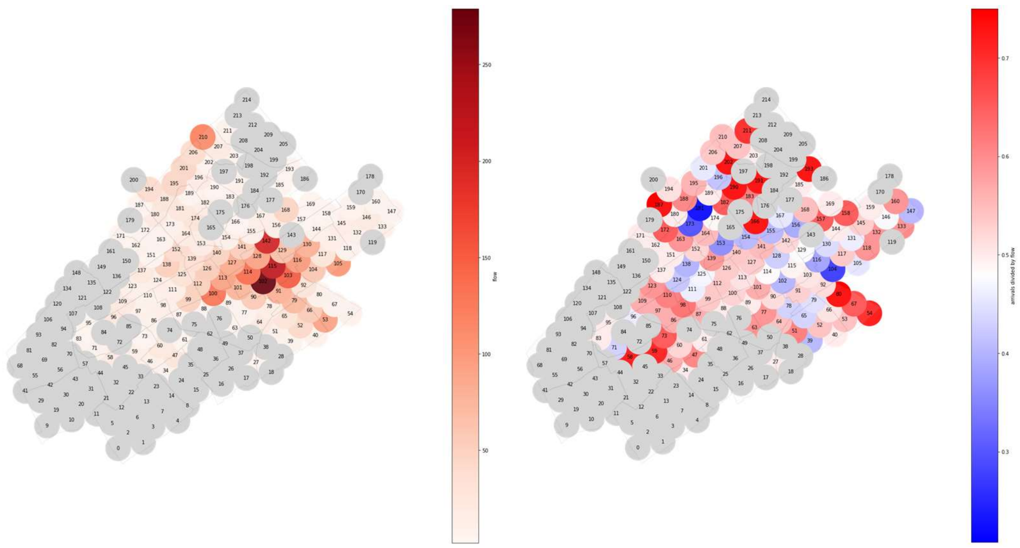


**Figure 6-30 Heatmaps of total flows and arrival ratio of the 5th cluster in circles**

Accordantly, the outer residential units are either balanced with similar departure and arrival ratios, or prone to more arrivals in this time period.

Taking the chord diagram (shown in Figure 6-31) into account, the central units, such as 142, 102, and 115 are prevalent as both origins and destinations. Apparently, it is observed that almost or more than a half of the trips originate and terminate at the same unit from the top 50 occurrent OD pair from this diagram, which means the trip is relatively shorter in this period.

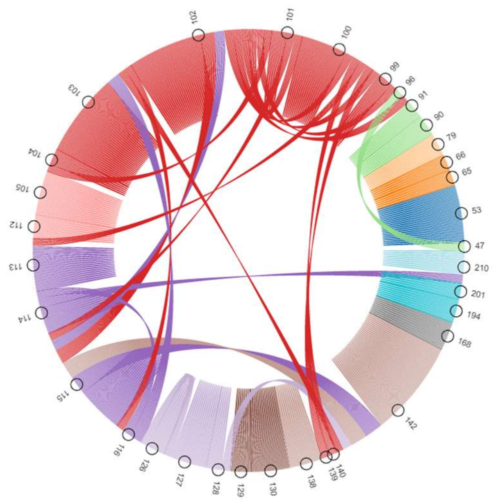



**Figure 6-31 Chord diagram of the 5th cluster in circles**

6. Descriptive comparisons among different hourly clusters on the circle level

**Table 6-7 Statistics of average travel distance and duration for different clusters**

| | Average travel distance (km) | Average travel duration (min) |
|---|---|---|
| 16 | 3.38 | 18.92 |
| 17 | 3.96 | 19.07 |
| 18 | 3.89 | 22.18 |
| 19 | 3.41 | 18.94 |
| Others | 3.64 | 17.28 |
| Total | 3.65 | 18.13 |

Comparing the average travel distance in kilometres and the average travel duration in minute of 5 different hourly clusters presented in Table 6-7, it is obvious that only the off-peak period (the other period) presents a shorter average travel duration than the global average travel duration. The travel distance in this period is also slightly lower than the overall average one. Conversely, the hour, 17 and 18 both presents longer trip distance and duration. Hour 16 and 19, reveal a different pattern, with a shorter average travel distance while a longer travel duration, implying a lower ride speed in these two hours.

➢ Daily cluster

The cluster analysis in the circle level does not reveal apparent daily clusters, consistent with the counterpart on the neighbourhood level in Appendix A, as presented in Figure 6-32.

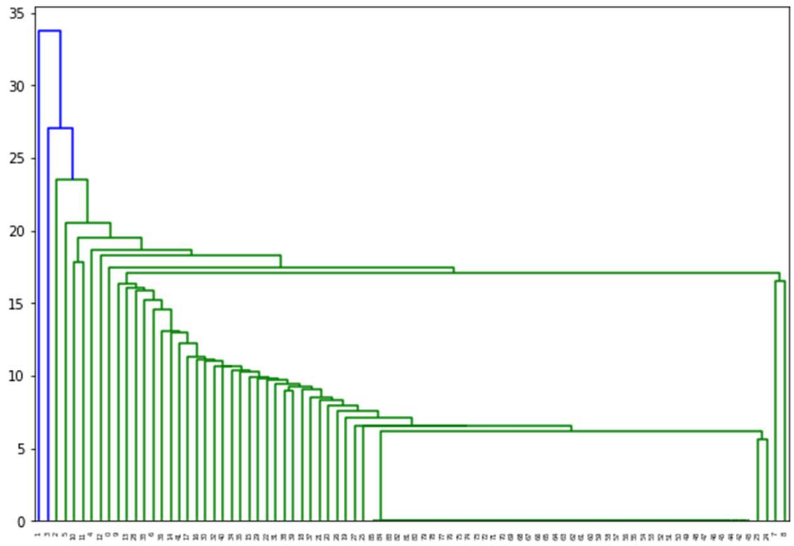



**Figure 6-32 Dendrogram of daily clusters on the circle level**

From the dendrogram shown in Figure 6-32, the results reveal uninterpretable daily clusters, at least there are no clear clusters in terms of day of the week. Instead, the days merge into different clusters in their stage of the operational period. For example, given 5 clusters, the second day cluster into a class, and the same hold for the other clusters in general. This implies that different days of a week are not dominant role in the daily clustering of the demand pattern within these two months. Instead, the evidence points out the likelihood that the adopting process of this sharing service (i.e., the habitual formation process) exerts stronger impacts on users' behaviour.

### 6.2.4 Supply efficiency analysis (until 22/09/21)

The overview of average vehicle idle time per unit is presented in Table 6-3. The origin-based metric has a slightly higher average vehicle idle time in general. N.B.: the average vehicle idle time presented in the summary table is based on the units with at least one vehicle idle time record, and therefore the units where there are no rides at all are not counted.

**Table 6-8 Statistics overview of supply efficiency analysis**

| | Valid circle units | Average vehicle idle time (h) |
|---|---|---|
| Origin-based vehicle idle time | 159 | 23.19 |
| Destination-based vehicle idle time | 157 | 22.51 |

➢ Origin-based vehicle idle time

Figure 6-33 presents the distribution of origin-based average vehicle idle time, where one data point corresponds to one unit. It is found that most units experience an average vehicle idle time between 20 to 40 hours. Figure 6-33 b includes all units, including the units without rides and thereby assigned with a default idle time, 72 hours.

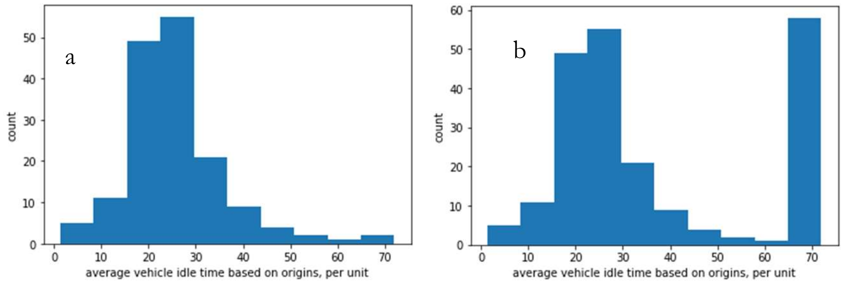



**Figure 6-33 Distribution of origin-based average vehicle idle time per unit (a); and including default values (b)**

It is observed from Figure 6-34 that the outer units usually experience a longer vehicle idle time, especially those located in the southwest.

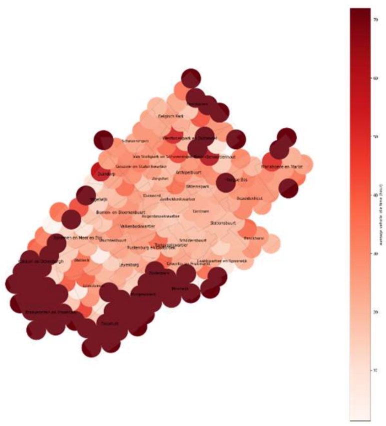


**Figure 6-34 Heatmap of origin-based average vehicle idle time per unit**

➢ Destination-based vehicle idle time

Similar findings are observed from the destination-based metrics, where the southwest part obviously lacks uages, as presented in Figure 6-35 and Figure 6-36.

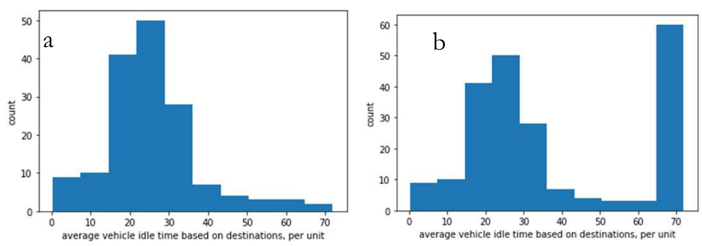



**Figure 6-35 Distribution of destination-based average vehicle idle time per unit (a); and including default values (b)**

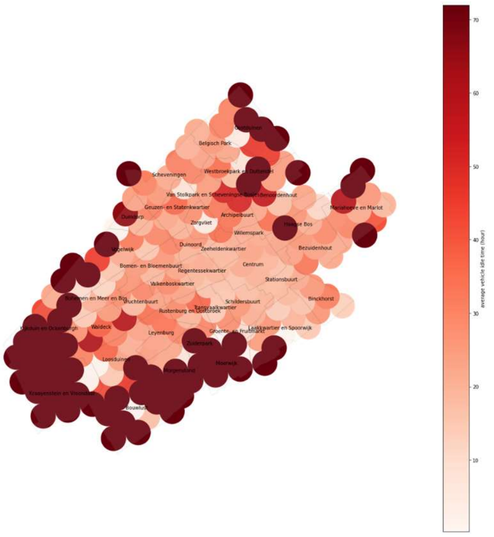


Figure 6-36 Heatmap of destination-based average vehicle idle time per unit

### 6.2.5 Average travel distance and duration analysis (until 24/09/21)

Average travel distance and duration indicates that there is a lack of use in the southwest part as presented in Figure 6-38 and Figure 6-40, aligned with the observations in section 6.2.4. Moreover, the units located between the outskirts and central areas, which are usually those office-oriented or residential units, witness a longer trip duration as well as trip distance.

Table 6-9 Overview of average travel distance and duration analysis

| | |
|---|---|
| #valid circle units | 139 |
| Average trip distance (km) | 3.22 |
| Average trip duration (min) | 16.48 |

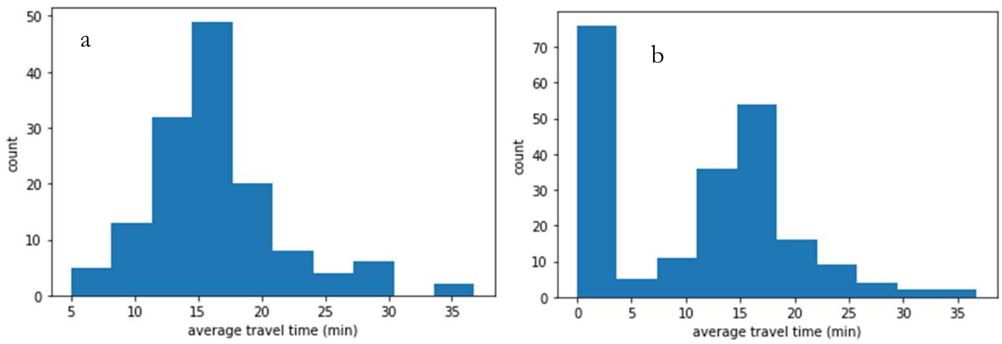



Figure 6-37 Distribution of average travel time and includes default values

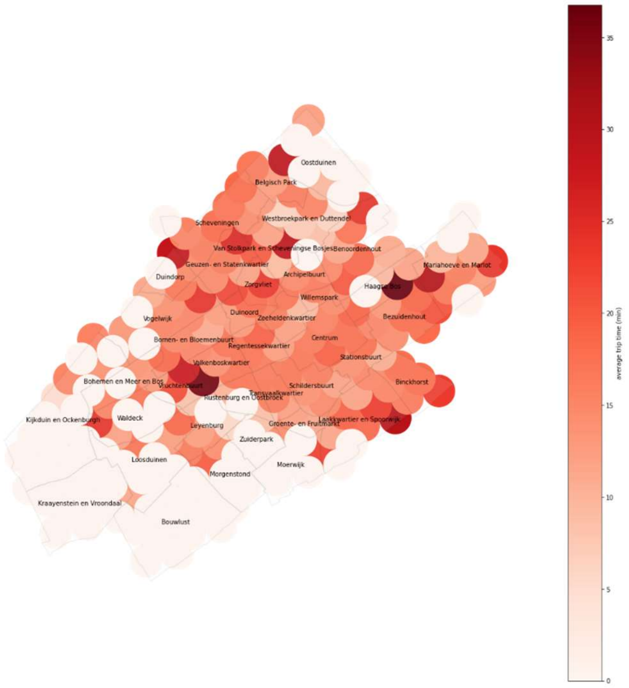



**Figure 6-38 Heatmap of average travel time per unit**

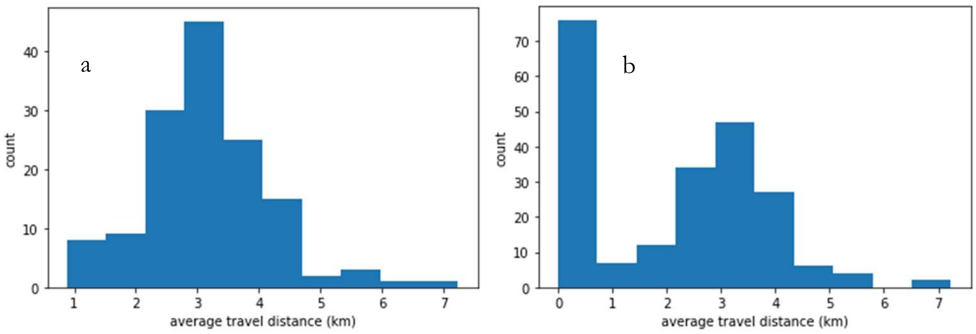



**Figure 6-39 Distribution of average travel distance and includes default values**

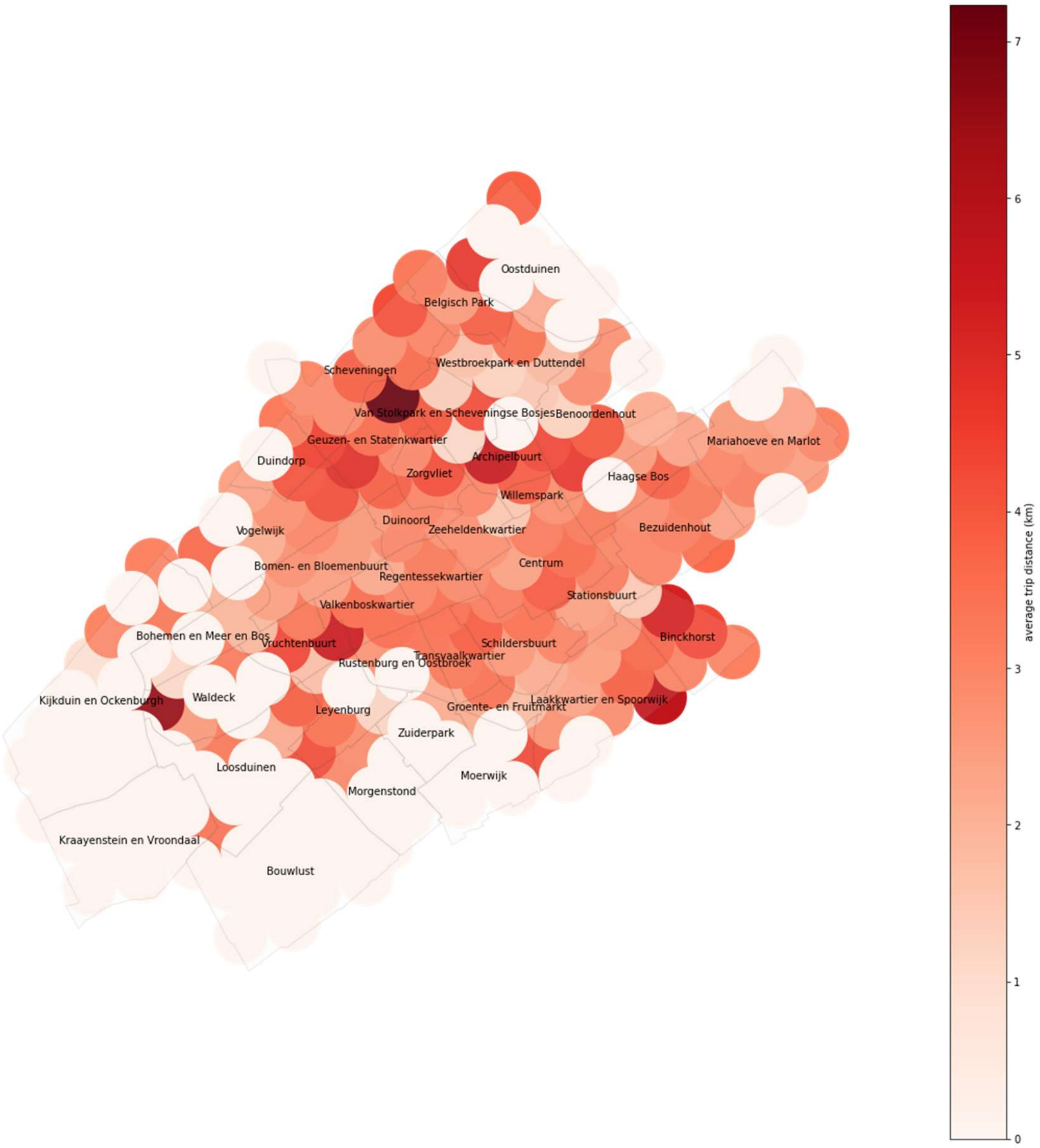



**Figure 6-40 Heatmap of average travel distance per unit**

### 6.2.6 Proposal of operational strategies

There are two sets of operational strategies based on the results from demand pattern analysis: The first set comes from the results of temporal clustering and the second is based on the output of supply efficiency as well as travel distance/duration analysis.

Considering flow, the imbalance between arrivals and departures, and reallocation efforts indicated by the distance of reallocation, some suggestions towards reallocation based on different units and in different time periods are made:

- Implications on the neighbourhood level

1. During the midday period, it is found that there are more departures in Bezuidenhout while more arrivals in Stationsbuurt nearby. Therefore, it is recommended to reallocated more bikes from the area near HS and Cnetraal station in Stationsbuurt to Amalia van Solmsstraat and Schenkka,m de in Bezuidenhout;
2. During the first peak hour, it is also recommended to assign more bikes to Bezuidenhout for the same hotspots to meet the potential demand, but the bikes should be drawn from Brickhorst which presents an arrival-tendency in this period;
3. During the second peak hour, it is recommended to reallocate bikes from S.R. Haagse Hout or Kampen in Mariahoeven with an arrival-tendency to Stationsbuurt;
4. During the transition hours, it is suggested to put more bikes in Bezuidenhout and Centrum from Laakwartier;
5. During the sleeping hours, it is advised to reallocate more bikes to Regentessekwartier and Centrum from Benoordenhout.

**Table 6-10 Summary of reallocation strategies based on hourly clusters in neighbourhoods**

| | **time period** | **from** | | **to** | |
|---|---|---|---|---|---|
| **midday** | **10:00-15:59** | Stationsbuurt | near HS or Centraal | Bezuidenhout | Amalia van Solmsstraat and Schenkkade |
| **1st peak** | **16:00-16:59** | Brickhorst | | Bezuidenhout | Amalia van Solmsstraat and Schenkkade |
| **2nd peak** | **17:00-17:59** | Mariahoven | S.R. Haagse Hout or Kampen | Stationsbuurt | near HS or Centraal |
| **transition** | **18:00-19:59** | Laakkwartier | | Bezuidenhout and Centrum | Amalia van Solmsstraat and Schenkkade |
| **sleeping** | **20:00-9:59** | Benoordenhout | | Regentessekwartier and Centrum | Loosduinseweg |

➢ Implications on the 400m circle level

1. For the heat spots, are mainly located in the central areas and they are rather balanced in general. However, in hour 16, unit 115 experiences overwhelming arrivals than departures, and thereby it is recommended to reallocate bikes in 115 to 103 to prepare the next busy periods before 103 become the hottest spot, serving more potential rides;

2. During the off-peak hours, there are normally more arrivals occurring in the pier unit (210) while more departures in the harbour unit (201). Therefore, it is advised to shift some bikes from 210 to 201`;
3. For the peak hours, 16 and 17 and the first transition hour 18, unit 53 located in Laak is an unignorable area, which is recommended to prepare more bikes. The bikes should be drawn from less busy areas with outweighed arrivals, such as the unit 78 in hour 16.

**Table 6-11 Summary of reallocation strategies based on hourly clusters in circles**

| | **time period** | **from** | | **to** | |
|---|---|---|---|---|---|
| **1st peak** | **16:00-16:59** | Centrum | 115 | Station | 103 |
| **2nd peak** | **17:00-17:59** | Schilderswijk in Centrum | 78 | Laak | 53 |
| **1st transition** | **18:00-18:59** | | | Laak | 53 |
| **2nd transition** | **19:00-19:59** | | | | |
| **Off-peak** | **20:00-15:59** | beach | 210 | beach | 201 |

➢ Comparison of reallocation strategies based on different units

The reallocation implications found from these two different units are quite diverse. However, it is almost identical that during the peak hours and the transition hours, Stationsbuurt is extremely popular and there is a need to reallocate more bikes to capture more potential rides. For the remaining ones, the suggestions in circles are more precise and therefore indicates different recommendation from the neighbourhood level.

These two subsets of rebalancing strategies derived from the temporal clustering were applied in different time periods, as described in the previous part of this chapter.

Subsequently, the second set of operational strategies was proposed according to results of supply efficiency and average travel distance/duration analysis as follows: since the southwest part is clearly lack of use, it is reasonable to shrink the operational areas to exclude these areas, so as to improve the service efficiency and prioritize the popular areas with more bikes. Addition to that, it mitigates the efforts of reallocation as well as battery swaps.

➢ Summary of operational strategies

Based on the results of demand pattern analysis, there are in total 5 recommended operational strategies as presented in Table 6-12. Three out of five focus on the reallocation strategies as the visualization in Figure 6-41 with clear starting time and ending time. They are consecutive from 17/07/2021 to

08/09/2021. Additionally, bikes are always placed in batches since the beginning and since 24/09/2021, the operational areas have been decreased dependent on the recommendation from supply efficiency analysis and average travel distance/duration analysis.

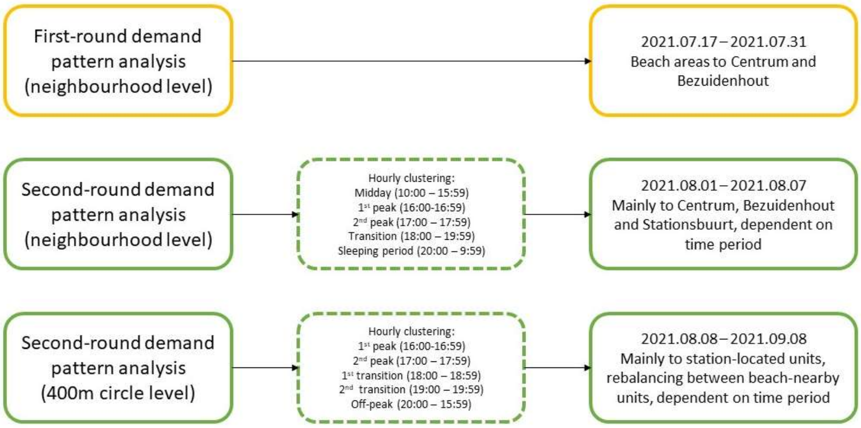



**Figure 6-41 Development of rebalancing strategies**

**Table 6-12 Summary of Operational strategies**

| | Start time | End time |
|---|---|---|
| Place bikes in batches | 2021.07.17 | Until now |
| First-round reallocation between Beach and Centrum/Bezuidenhout | 2021.07.17 | 2021.07.31 |
| Second-round reallocation on the neighbourhood level (hourly-based) | 2021.08.01 | 2021.08.07 |
| Second-round reallocation on the 400m circle level (hourly-based) | 2021.08.08 | 2021.09.08 |
| Reduction in the operational zones | 2021.09.24 | Until now |

## 6.3 Operation

The priority of bondi's operational strategy is the battery swap all the time, ensuring the usability of shared e-bikes. For reallocation, there was only 1 run of reallocation of e-bikes per week on average before the implements of the suggested strategies.

### 6.3.1 Introduction of operational strategies

Because of the limited accuracy of the short-term demand prediction model, all of the operational strategies were derived from the demand pattern analysis as described in 6.2.6. There are in total 5 operational strategies with various time scopes where 3 out of 5 are reallocation strategies, consecutive in the timeline from 17/07/2021 to 08/09/2021, seen in Table 6-12 in the last part.

The details of the last two-round reallocation are presented in Table 6-10 and Table 6-11 in the previous section, and the exact action taken per reallocation is dependent on the execution time of the action, which is usually between 10:00 to 16:00 during daytime.

### 6.3.2 Computation of KPIs of operational strategies

The KPIs from the operational sides are presented in Table 6-13. It is obvious that the period of the implementation of ID verification performs the worst out of all the periods. During the periods with rebalancing, all KPIs are improved to different degrees. Among three reallocation periods, it is clear that the reallocation strategies based on hourly clustering have stronger positive effects no matter on ridership or the mitigation of vehicle idle time. Besides, the reallocation on the circle level has the foremost advantages on the improvement of the service level.

**Table 6-13 Summary of Operational KPIs**

| | | | average supply (\|vehicles\|) | average ridership (\|rides\|) | average ridership ratio | average vehicle idle time per vehicle (h) | |
|---|---|---|---|---|---|---|---|
| | | | | | | origin-based | destination-based |
| free rides | 19/06/2021 | 21/06/2021 | 78 | 164.33 | 2.11 | 8.81 | 17.82 |
| adopting period | 22/06/2021 | 06/07/2021 | 107.06 | 105.33 | 0.98 | 47.28 | 32.14 |
| ID verification | 07/07/2021 | 16/07/2021 | 80.8 | 25.2 | 0.31 | 59.51 | 52.66 |
| first-round reallocation | 17/07/2021 | 31/07/2021 | 92.67 | 48.33 | 0.53 | 51.84 | 43.25 |
| second-round reallocation on the neighbourhood level | 01/08/2021 | 07/08/2021 | 90.29 | 54.14 | 0.59 | 57.3 | 41.08 |
| second-round reallocation on the circle level | 08/08/2021 | 08/09/2021 | 107.72 | 72.44 | 0.67 | 47.3 | 33.07 |
| without specific strategy | 09/09/2021 | 23/09/2021 | 95.93 | 61.8 | 0.64 | 50.16 | 41.08 |
| reduction in the operational area | 24/09/2021 | 18/10/2021 | 87.27 | 45.54 | 0.52 | 50.9 | 43.63 |
| overview | 19/06/2021 | 18/10/2021 | 96.14 | 63.89 | 0.66 | 40.77 | 33.05 |

From the users' perspective on a monthly basis, the figures related to net retention rate are illustrated in Table 6-14. The month from 19$^{th}$ of August to 18$^{th}$ of September has the highest NRR at 86.87%. However, NRRs of all three months are below 100%. Taking the average user expenditure into account, it is easily observed that retained users always have higher average expenditure than the new users, as seen in Table 6-15.

**Table 6-14 Summary of NRR**

| Time period | \|users\| | Total revenue (€) | \|Retained users\| | Retained users expansion (€) | \|Churn users\| | Churn users loss (€) | NRR |
|---|---|---|---|---|---|---|---|
| 6.19 to 7.18 | 1056 | 7804.80 | | | | | |
| 7.19 to 8.18 | 425 | 3924.98 | 81 | 243.87 | 975 | 6817.41 | 15.78% |
| 8.19 to 9.18 | 448 | 6855.43 | 139 | 918.06 | 286 | 1942.66 | 86.87% |
| 9.19 to 10.19 | 333 | 4811.94 | 162 | -683.41 | 286 | 3055.01 | 52.10% |

**Table 6-15 Summary of average user expenditure**

| Time period | \|New users\| | New users revenue (€) | New user average spent (€) | Retained user average spent (€) | total user average spent (€) |
|---|---|---|---|---|---|
| 6.19 to 7.18 | | | | | |
| 7.19 to 8.18 | 344.00 | 2693.72 | 7.83 | 15.78 | 9.24 |
| 8.19 to 9.18 | 309.00 | 3955.05 | 12.80 | 20.87 | 15.30 |
| 9.19 to 10.19 | 171.00 | 1694.93 | 9.91 | 19.24 | 14.45 |

### 6.3.3 Evaluation of the operational strategies

Based on the results shown in 6.3.2, it is found that the reallocation accordingly to the hourly clustering demand pattern on the circle level has the most remarkable advantages, contributing to the highest ridership and the lowest vehicle idle time by the operator's side.

By the users' side, the time interval of KPIs is exactly 1-month, different from the KPIs in the operators' aspect. Thereby, it is hard to compare these two sets of KPIs. The time scope of the three reallocation strategies is partially overlapped with the second month (from 19$^{th}$ of July to 18$^{th}$ of August) while the third month also accounts for more than half days of the execution period of the third reallocation strategy. The third month performs the best considering both NRR and average user expenditure.

Nevertheless, NRRs of all three months are below 100%. It indicates that the customer royalty is not fully established. It is quite common for a new service/product and especially for a start-up company. The operational strategies are proven to be beneficial to improve the service, indicated by an increase in NRRs and average user expenditure, though with substantial fluctuations.

The fourth month (9.19 to 10.19), experience a decline in NRR and total revenue, while the average expenditure of new users and all users do not witness a considerable decrease, even with the shrink in the operational areas.

The last strategy, reduction in the operational area, has a lower positive effect while it decreases the efforts of battery swaps and reallocation to a large degree, not reflected in the KPIs used in this evaluation, though.

It is found that rebalancing has better effects while it requires more effort in operation. However, reduction in the service area, in the other way, relieves the efforts of operation without harm to the service level. Thereby, they are complementary to each other and executed at the same periods.

### 6.3.4 Discussion of the evaluation results

Based on the assessment of these operational strategies, it is proved that they do improve the service level. However, there are still some limitations underlying the evaluation.

On one hand, the comparisons are done between the reference case of the ID verification and the other time periods. There are many external factors affecting people's willingness to use the service though, such as weather, summer holidays, and some other unknown variables.

On the other hand, it is also worth mentioning that there is another free-rides week from 06/08/2021 to 13/08/2021, which has also exerted positive effects on the ridership. Offline promotion campaigns from the marketing department in bondi also attract users to a substantial extent. Thereby, the improvements in the service do not originate in these abovementioned strategies completely.

# 7 DISCUSSION

This chapter presents a discussion of the whole research, targeting the data, methods and the results. First of all, results from the previous chapter 6 are reflected, comparing with the parrel work. This is followed by the description of the research contributions, in both scientific and practical ways. Then, the limitations in terms of data and methods are suggested, with recommendations of future work.

## 7.1 Results reflection and synthesis

In this section, results from different phases, data analysis, demand pattern analysis and operation namely, are presented with emphasis, interpreted and compared with parallel work.

- **The Data analysis**

**Table 7-1 Results of Pearson's coefficients of different factors with the demand**

| Spatial level | | the whole area | neighbourhood | circle |
|---|---|---|---|---|
| supply | | 0.43 | 0.18 | 0.49 |
| weather | temperature (℃) | 0.16 | - | - |
| | Precipitation (mm) | -0.03 | - | - |
| | Humidity (%) | -0.31 | - | - |
| land use | \|sustenance\| | - | 0.92 | 0.76 |
| | \|office\| | - | 0.90 | 0.61 |
| | \|entertainment\| | - | 0.92 | 0.74 |
| | \|healthcare\| | - | 0.61 | 0.73 |
| | \|education\| | - | 0.74 | 0.33 |
| public transport | \|stops\| | - | 0.34 | 0.26 |
| | \|ridership\| | - | 0.66 | 0.29 |

**Relationships between the contributing factors and the demand**

For the data analysis stage, supply, weather, POIs and public transport availability, as contributing factors, were examined as determinants for the ridership using a linear regression. The results, represented by Pearson's coefficient, can be seen in Table 7-1 and multiple regression with the ridership. It is found that despite the weather, all other three factors have a positive effect on the ridership, to a different degree, ranging from 0.18 to 0.92. It means users prefer to start their rides in an area with more POIs (more services), more (attractive) PT stops and more available bikes. The positivity brought by the supply is quite straightforward since people only use the bike when there are available bikes and the operators tend to place the bikes in the hotspots of the city, denser than other locations. This finding was also be observed from the previous work though most of the work study the relationship between the supply and the demand for station-based bike sharing schemes in which the number of stations or the density of docked stations results in a higher ridership (Zhao et al., 2014).

**POIs (points of interest):** the positive effects from POIs are due to the fact that travel is a derived demand and it is closely related to the activities. An area with more POIs represents a potential of more activities and thereby attracts more population, resulting in more rides. From all types of POIs, entertainment, office and sustenance amenities have a higher magnitude for the generation of rides, which is also reflected in the temporal pattern in the later phase. This result is in line with the findings from the literature where POIs, in general, contribute to a higher bike usage, by slightly different proofs, such as a mixture of POIs leading to higher turnover rate (Faghih-Imani et al., 2017; A. Li et al., 2020). Out of all categories of POIs, education and healthcare facilities are observed to impact the usage undesirably, while these negative coefficients are caused by the relatively lower values of these attributes themselves, as indicated by Table 5-1 and Figure 5-6.

**The availability of public transport:** similarly, the availability of PT also facilitates a higher usage as in line with the previous work (A. Li et al., 2020). However, in this study, the indicator of the availability of PT is presented by the ridership. The number of POIs acts as a confounder in this case and as a consequence, the attractiveness of the unit because of a higher number of POIs also contributes a higher ridership.

**The weather-related factors:** as for weather factors, it is found that warmer temperatures lead to higher ridership level while precipitation and humidity decrease the ridership, indicated by Pearson's coefficient. When it comes to the multiple regression analysis including all three indicators representing the weather category, as indicated in Table 6-2, cannot be identified having a significant impact on the ridership, same as the case study in Park City (He et al., 2019). The temperature growth, surprisingly, contributes to a lower ridership in the regression result, different from the findings in Pearson's coefficient and the parallel research where an increase in temperature leads to a higher ridership (El-Assi & Mahmoud, 2015; He et al., 2019; Miranda-Moreno & Nosal, 2011; Tin Tin et al., 2012). However, this is caused by the inconsiderable variations rooted in the temperature data within this 3-months period. Humidity is found to have a negative effect both in Pearson's coefficient and the multiple regression, and this result is in harmony with several previous work (El-Assi & Mahmoud, 2015; Miranda-Moreno & Nosal, 2011).

**Land use pattern**

Additional to correlation analysis, the function analysis based on the distribution of POIs were also conducted and the results on the circle level are shown in Figure 7-1. The central areas are dominated by recreational functions while the office-oriented units distribute relatively sparse. These functions assisted to analyse the following demand, to examine whether flow has a tendency between zones with different functions during a certain period.

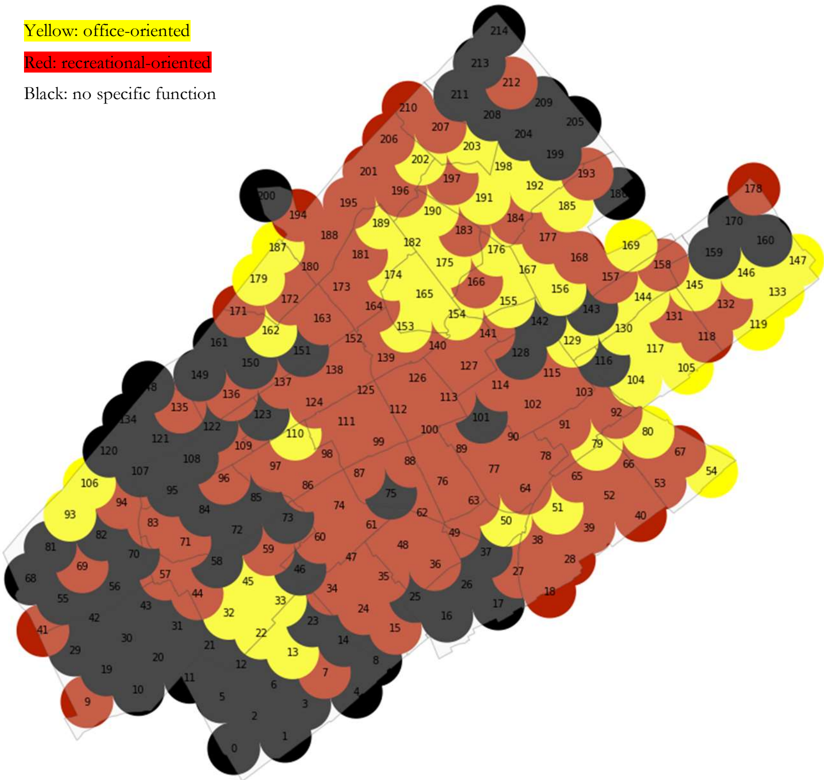



**Figure 7-1 Functions of units on the circle level**

**Future improvements in the data analysis:** to improve the data analysis stage in this study, some processing of data should be conducted. First, data is aggregated in a 3-hours interval for the examination of weather factors with the ridership, while daily aggregation is helpful to provide more insights, allowing the comparisons of the significance of these factors as well as the coefficients. Besides, only the current weather is considered in this case while the historic weather, such as the weather condition in the past 3-hours also impacts the ridership since some of people plan their trip in advance. Additional to the weather data, other spatial processing towards POIs, such as the mixture of POIs as mentioned previously to see if the diversity of POIs affects the ridership, is also the direction of future work.

- **Demand pattern**

**Demand pattern analyses and hourly clustering analyses**

Transitioning to the results from the demand pattern analysis, the first insight concerns the ridership across the hour of a day. Different from the previous studies, there is only one obvious peak period, starting at around 16:00 and ending at around 19:00 in the case study, as presented in Figure 6-14, while most of the other schemes usually witness two peaks, morning peak and evening peak respectively (S. Li et al., 2021; Miranda-Moreno & Nosal, 2011; Tin Tin et al., 2012; Xing et al., 2020). The peak period in this study corresponds to the evening peak in other studies, though it is earlier than the regular evening peak, which usually starts at 17:00 (Ma et al., 2019). The general demand pattern analysis illustrates a widely-observed phenomenon that the central areas are always prevalent by more rides than the outskirts (S. Li et al., 2021; C. Xu et al., 2018). The hourly clustering determined by agglomerative clustering indicates 5 clusters, among which two clusters correspond to the two hours in the peak, the other two represents the two hours after the peak, and the remaining periods cluster together as indicated in Table 7-2. It is found that during the peak and transition clusters, people tend to leave from the office-oriented zones to the recreational-oriented zones, as indicated in Figure 7-2 as a representative, which is parallel with the previous studies (Xing et al., 2020).

**Table 7-2 Summary of hourly clusters on the 400m overlapping circle level**

| Hourly cluster | Time period |
|---|---|
| 1st peak hour | 16:00 to 16:59 |
| 2nd peak hour | 17:00 to 17:59 |
| 1st transition hour | 18:00 to 18:59 |
| 2nd transition hour | 19:00 to 19:59 |
| Off-peak periods | 20:00 to 15:59 |

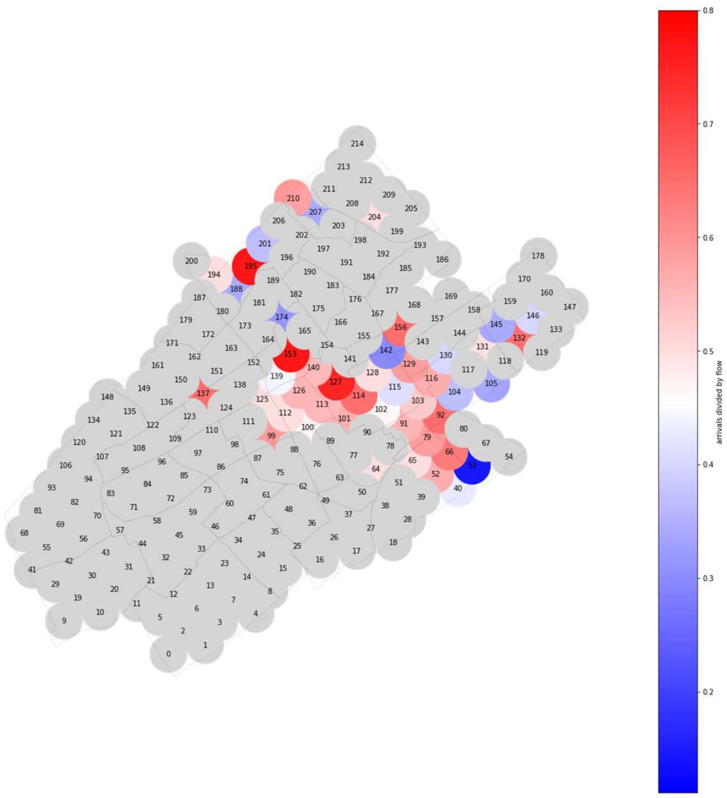



**Figure 7-2 heatmap of departure ratio at the 1st transition hour**

However, the inflow and outflow of residential zones are not discussed explicitly in this study because of a lack of data on the overlapping circle level and thereby the inference of commute trips is not well-supported. Apart from that, starting from the second peak hour at 17:00, the station-nearby unit experiences increasing popularity, firstly mainly for trip generation but gradually attracting both departures and arrivals in the following hours until 20:00, implying the fact that users tend to use this e-bike sharing to serve the first and last mile of train trips, in accordance with the previous studies (S. Li et al., 2021; Xing et al., 2020).

**Supply efficiency and average travel distance/duration analyses**

These are followed by the analysis of supply efficiency analysis and travel distance/duration per unit. It is also found that the e-bikes witness less usage in the outskirts while some units located in the central areas and the office-oriented units nearby experience the most efficient usage as indicated in Figure 7-3 and Figure 7-4. There is no existing literature, to the best of the author's knowledge, in the bike sharing field performing similar analysis and thereby the results are hard to compare. The results from the general and hourly demand pattern analyses also come to the similar results in this part.

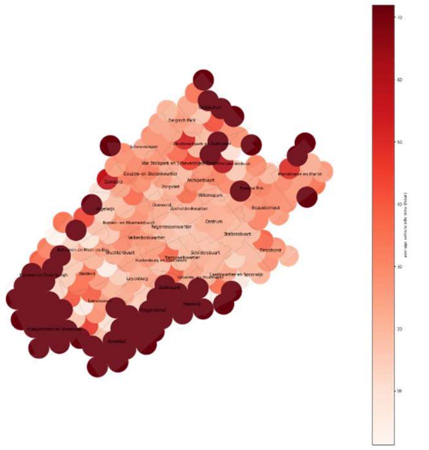


**Figure 7-3 Heatmap of origin-based average vehicle idle time per unit**

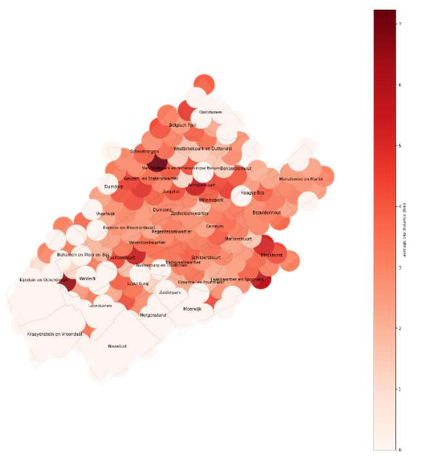


**Figure 7-4 Heatmap of average travel distance per unit**

**Development of the operational strategies**

When it comes to the development of operational strategies based on these results obtained from the data-driven analysis, there are mainly two types of operational strategies, which are rebalancing and reduction in the service zone. The reallocation suggests the reallocation between different units to balance the supply with the demand while the adjustment in the operational zone aims to prioritize the areas of the most interest, leaving out the areas with the least use or no use.

- **The operational phase**

The last phase tackles the execution and evaluation of the derived operational strategies.

**Ridership ratio:** the foremost KPI used in the evaluation is the ridership ratio per bike per day. The adopting periods, including free ride days, experience a high ridership ratio at over 2 in the beginning and gradually shrinks to around 1 ride per bike per day. The period of ID verification without any operational strategies, witnesses the lowest ratio, at 0.31 and this indicator bounces to between 0.5 to 0.7 with the application of different operational strategies as presented in Table 7-3. Though it is observed that operational strategies indeed facilitate a healthier ridership ratio, this e-bike sharing service still presents a lower ratio, compared with other schemes as shown in Figure 7-5 and the full details can be seen in Table C-1 in Appendix C.

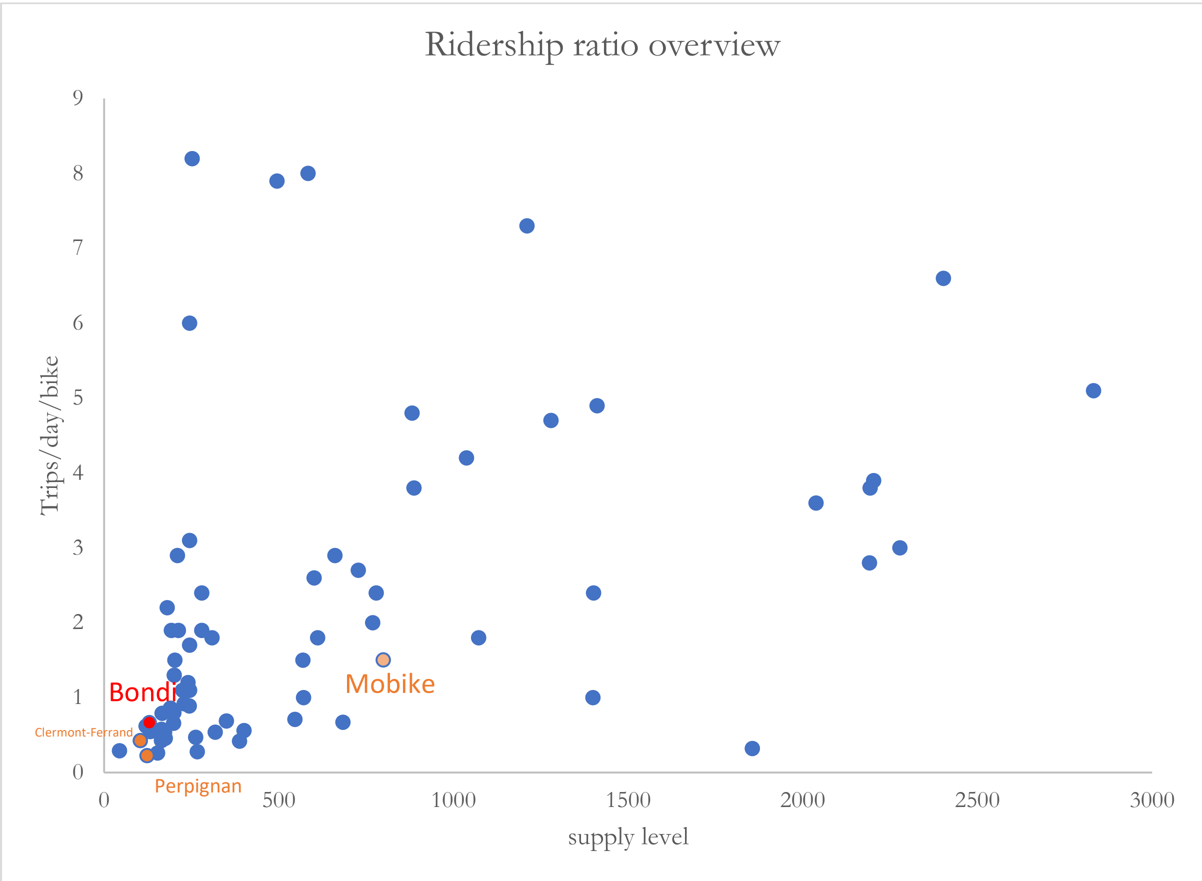
One of the reasons is the high ownership of bicycles in The Netherlands and thereby the usage of bike sharing is much lower in this case study while the table does not target the projects in The Netherlands. Also, the supply level in this study is under 150 bikes across the whole period, even considering the possible underestimations rooted in the supply inference, which is quite low compared to other schemes listed in Table C-1 and there are only four projects, Alacant, Boulder, Clermont-Ferrand and Perpignan, having a similar supply level at around 130. The first two schemes present a similar ridership ratio as this case study and the latter two have a lower ridership ratio at around 0.4 and 0.2 respectively. Considering a similar context in Delft, the Mobike project, with a much higher supply between 600 to 1000, also only saw the ridership ratio at around 1.5 in 2018 (Boor, 2019). All these figures prove the decency of the results in this case study as well as the efficacy of operational strategies.

**Vehicle idle time:** turning to the second KPI, the vehicle idle time. The operational strategies help decrease the vehicle idle time, no matter origin-based or destination-based, by around 10 hours, with fluctuations across different strategies, compared to 59 hours and 52 hours per e-bike during the ID verification period. This indicator, along with the ridership and ridership ratio, suggest that the reallocation based on hourly clustering on the overlapping circles has the best performance on the improvement of the service.

**Table 7-3 Summary of Operational KPIs**

| | | | average supply (\|vehicles\|) | average ridership (\|rides\|) | average ridership ratio | average vehicle idle time per vehicle (h) | |
|---|---|---|---|---|---|---|---|
| | | | | | | origin-based | destination-based |
| free rides | 19/06/2021 | 21/06/2021 | 78 | 164.33 | 2.11 | 8.81 | 17.82 |
| adopting period | 22/06/2021 | 06/07/2021 | 107.06 | 105.33 | 0.98 | 47.28 | 32.14 |
| ID verification | 07/07/2021 | 16/07/2021 | 80.8 | 25.2 | 0.31 | 59.51 | 52.66 |
| first-round reallocation | 17/07/2021 | 31/07/2021 | 92.67 | 48.33 | 0.53 | 51.84 | 43.25 |
| second-round reallocation on the neighbourhood level | 01/08/2021 | 07/08/2021 | 90.29 | 54.14 | 0.59 | 57.3 | 41.08 |
| second-round reallocation on the circle level | 08/08/2021 | 08/09/2021 | 107.72 | 72.44 | 0.67 | 47.3 | 33.07 |
| without specific strategy | 09/09/2021 | 23/09/2021 | 95.93 | 61.8 | 0.64 | 50.16 | 41.08 |
| reduction in the operational area | 24/09/2021 | 18/10/2021 | 87.27 | 45.54 | 0.52 | 50.9 | 43.63 |
| overview | 19/06/2021 | 18/10/2021 | 96.14 | 63.89 | 0.66 | 40.77 | 33.05 |

**Figure 7-5 Ridership ratio overview of bike sharing schemes around the world**



**Net retention rate and average user expenditure:** to the users' side, the net retention rate is around 50% to 80% after the implementation of the suggested strategies. Searching for similar metrics targeting bike sharing services, it is found that the retention rate is also below 100%, ranging from 20% to 70% Even the world's biggest bike sharing operator, Lime only claims having a retention rate of 60% in the US (Shaheen et al., 2014; SmartCitiesWorld, 2017). Besides, this rate for bondi is quite acceptable considering the fact that this is only the first 4-months of this service and the operator is a start-up company. The last discussed KIP, avg. user expenditure, shows an increase when comparing the second month with the third and fourth, demonstrating an increase from 9 to 15 € per user. Moreover, it is found that retained users have a tendency to spend more money per trip compared to new users, and the months with the operational strategies indeed witness a higher average user expenditure.

To summarize the operational phase, the different implemented strategies boost the service of e-bike sharing service in this case study, regardless of the limitations laying on the data and methods themselves, which will be mentioned in the following sections 7.3.

- **Implications**

The results of this study are composed of three stages, data analysis, demand pattern analysis and the operational part. The implications are elicited accordingly below.

**Correlation between weather and the attraction of units:** the results from the data analysis suggest that the demand for e-bike sharing is dependent on several determinants, such as supply, weather, and POIs. Areas with more POIs and public transport indeed attract more rides, by the generation of more activities. This implies that the operator should prioritize these central areas with diverse activities. However, some of them have a stronger relationship with the weather, especially those areas having typical outdoor activities. For instance, if there is heavy precipitation, e.g., at the beach, this results in less demand, even though with several points of sustenance and entertainment, are not an ideal location to reallocate the bikes. Contrarily, the areas having a high density of POIs generating indoor activities, are more resilient to the weather changes.

**Attraction of office-oriented units during the evening peak:** second to that, it is observed that there is only one early evening peak in the e-bike sharing service. It gives a hint that people prefer to use shared e-bike when they do not rush to work; They are more willing to use shared e-bikes to leave work to recreational areas or back home. This statement is also supported by the findings from the hourly clustering where the most popular OD pairs are always between the office-oriented zones/station-unit and the recreational zones. It is thereby recommended that the service provider should pay attention to reallocate more bikes in the office-oriented zones before the peak.

**Combination of different operational strategies:** the derived reallocation strategies, follow the suggestions mentioned above and it is found that they indeed improve the ridership. Additionally, the results of the operation infer that even though the rebalancing actions have the most significant effects on the ridership, it increases the efforts of staff necessary to relocate the bikes, unavoidably. Operational strategies apart from the reallocation need to be taken into account as well when dealing with the operation, such as the reduction of the service area. It has several benefits: first, it makes better use of available vehicles, by reallocating them to the areas with shorter vehicle idle time; second, the physical efforts are deduced to a large extent by a narrower operational area; third, the revised operational area are well equipped with more supply. However, it is still worth noting this strategy disappoint some users if their origins/destinations are not within the service area anymore and therefore a considerate analysis needs to be conducted beforehand, as done in chapters 6.2.4 and 6.2.5.

**Limited use of the short-term demand predictive model:** another implication from this research is that if the ride records are limited and the ridership level is relatively low, the predictive model is possibly not an ideal tool to help with the operation, especially for short-term operational strategies.

## 7.2 Research contributions

This part describes the contributions of this study, from both scientific and practical perspectives.

This research aims to bridge the scientific gaps as mentioned in chapter 3, and the detailed contributions are as follows:

**Spatial analytical units:** first, it applies two spatial analytical units, administrative neighbourhoods and 400m equally-distributed overlapping circles in the demand pattern analysis, with additional temporal clustering. This study compares the results from these two different units and also derives different reallocation strategies accordingly. It is found that the latter one (400m circles), which is still an innovative unit, have a more desirable result, both in the execution itself, easing the reallocation efforts by more precise locations, and in the positive effects on the service**.** The current studies usually only tackle one selected spatial analytical level and the comparison between different units are unavailable while this study provides such insight.

**Study on free-floating e-bike sharing service:** second, this study itself targets free-floating e-bike sharing services, which is less discussed in the current literature. It adds some knowledge to this field and can be used as a reference for future work. It also synthesizes the findings with the counterparts of traditional bike sharing projects in chapters 7.1, revealing the similarities and dissimilarities of the current studies in various aspects.

**Evaluation**: third, which is also the most important part, it contributes to the evaluation of operational strategies. As mentioned before, the existing research does not assess the effects of different operational strategies in a real-life setting while this research bridges the gap by providing a well-rounded evaluation, considering both the operator and the users and both quantitatively and qualitatively. Additionally, it introduces several novel KPIs to evaluate the e-bike sharing service, including vehicle idle time, net retention rate as well as average user expenditure. It provides a guideline for fellows to evaluate the operational strategies of e-bike sharing services, in a systematic way.

**Improvement of the service and example for other provides:** switching to the practical side, the foremost benefit, as stated in the title of this research, is to improve the service and it indeed facilitates a better service by the data-driven approach as presented in this study. The improvements were achieved cost-effectively, without costing too many resources while reaching quite desirable results. It is a successful case that can be followed by other small operators. Apart from the service providers, the improved service level also facilitates to the users, and the whole society, by providing a convenient transport service. It also benefits the city and contributes to the future of transport.

## 7.3 Limitations and future work

This section describes the limitations rooted in this study, from both the data-related and method-related perspectives, followed by the potential of future work, specific to those mentioned limitations.

### 7.3.1 Data-related limitations and future solutions

**Ride records:** the discussion firstly comes to the essential components of data, the ride records. The amount of ride records is quite limited, lower than 10,000 in total within a 4-month operation. This, unavoidably led to many undesirable results of the applied methods as follows: first, the demand pattern between weekdays and weekends are unclear and the similar finding occurred in the results of daily clustering, with the emergency of clusters which are hard to interpret; Second, the inaccuracy of LSTM short-term predictive model presented in Appendix B is mainly caused by the limited size of data size; Apart from this, the low level of ridership also increases the complexity of prediction. For instance, the highest ridership in unit 105 within a 3-hour interval is only 4 and the most frequency readerships are 0 and 1, which makes the dependent variable look random to some extent.

**Supply data:** moreover, the availability of some data sources is also quite constrained. Undoubtedly, the fleet size is an important indicator, no matter when studying the correlations or assessing the impacts of those applied strategies. However, in this study, the dynamic fleet size is not directly available and instead, it was thereby inferred from the ride records. This leads to the fact that if a bike is not used, then it is not counted into the fleet. The limited size of ride records even worsened this situation.

**Availability of other data:** other data of interest also confront the situation of unavailability. The weather data is in a 3-hour interval while in reality travel demand is affected by the weather, especially the precipitation explicitly and shorter time interval, such as 1-hour interval is more ideal for the research, no matter in the aspect of correlation check or as independent variables in the predictive model. Furthermore, the income, population and ownership of private cars and motors are only available on the neighbourhood level and therefore are not connected to the circle level.

**Time scope of data:** for further studies, a longer time scope is desirable and help address the abovementioned limitations to a large degree. Second, if the hourly ridership is still low, the ridership aggregation on the longer time interval should be considered and applied in the predictive model, which is facilitative to the accuracy. It is concluded that the time interval used in the short-term demand predictive model should be determined and adjusted according to the distribution of the dependent variable.

### 7.3.2 Method-related limitations and future solutions

The limitations rooted in the methodology is explained in this section, along with the proposals on how these disadvantages will be overcome in the future.

**Supply inference:** first of all, the way to infer the supply data, as discussed in the previous part, is not perfect. This method does not take the unused vehicles into account, which underestimates the fleet size. Additionally, it used a 7-days interval to determine the fleet in the case study while some vehicles are withdrawal within this period, leading to an overestimation of supply. This resultes in the impreciseness of the supply data, which undermined the further analyses, including the correlation analysis between the supply and the demand, vehicle idle time computation as well as the predictive model. These limitations can be alleviated by a larger dataset and eliminated with the availability of separate vehicle data with time stamps.

**Correlation analysis:** secondly, the correlation analysis was only conducted between the main variable of interest, which is the ridership, as the dependent variable, and dependent variable(s) belonging to the same category of determinants. Besides, only the linear correlation was examined by Pearson's coefficient and multiple regression analysis while nonlinear correlations also exist. Apart from this, the variables, except the weather data, are all linked to spatial units, and thereby spatial autocorrelation should also be tested. A well-rounded correlation analysis, including non-linear correlation, the correlation between the demand and a combination of different exogenous determinants, as well as spatial autocorrelations, is recommended to overcome these downsides.

**Spatial analytical units:** the third limitation is connected to the spatial analytical units. Two spatial analytical units, the neighbourhoods and the overlapping circles were determined without considering the spatial autocorrelation. As a result, the demand pattern per unit is not able to distinguish itself from other units, contributing to the ambiguity of results of daily clustering. Additional spatial clustering, before temporal clustering, is an option to address this issue. However, it will increase the operational efforts, which undermines the initiative of the unit of the overlapping circles.

**Vehicle idle time:** the next comes to the computation vehicle idle time. As suggested in the first statement, the results are partially biased because of the disadvantages rooted in the supply data used in this study and it will be eliminated with a better supply data source, either explicit time-series vehicle data or a larger set of ride records. Moreover, a time interval was set to compute the vehicle idle time and 3-days is applied in this study. It means the longest vehicle idle time is 72 hours by default. However, there are some bikes that are not well reallocated experiencing a vehicle idle time longer than 72 hours in reality and this method thereby underestimates this metric. Besides, if a vehicle was already withdrawn in reality while was accounted into supply due to the method used to infer the fleet size, a longer vehicle idle time will be assigned to this bike, even it is not under operation during this period in reality. A proper time interval, compatible with the regular operational interval mitigates this advantage. In this study, 3-days are already the most situation time interval while the operational interval is not fixed, which results in possible drawbacks.

**Evaluation of the operational strategies:** this is followed by the reflections of the evaluation phase. In general, the method applied in this study assessed the effects on the service as a whole, instead of targeting on the influences exerted by the operational strategies solely. Therefore, impacts are also yielded by other attributes, such as the weather, other determinants as mentioned in chapter 2.1, as well as other strategies taken by the provider (i.e., marketing campaigns). They were entangled with each other and thereby it is hard to perform abductive reasoning. It is difficult to overcome in the evaluation, and thereby the causality is not strong as thought. Furthermore, there are even more external factors exerting the effects on people's travel behaviour, especially during this ever-changing period under the effects of the ongoing pandemic, COVID-19. The baseline in the case study is the situations during the ID implementation without any deliberative strategies in July while other suitable reference cases are solutions of this issue. For instance, results of a predictive model with controlled hypothetical attribute levels are better references to their counterparts in reality. Another limitation in the evaluation method is related to one of the KPIs, Net Retention Rate. It was computed on a monthly basis while the execution of different operational strategies was not exactly one month and thereby their effects are not well reflected in this KPI. Similarly, the other KPI from the customers' side, user average expenditure, was also calculated on a monthly basis while the time interval is flexible and easy to adjust accordingly. On the contrary, the problems rooted in NRR is hard to deal with since it is not realistic to evaluate the recurring revenue in a shorter period while these impacts are also evaluated by other KPIs to some degree.

### 7.3.3 Future work

Further research directions are suggested in this section.

**Overlapping circles and relationship to public transport:** first of all, it is suggested to apply overlapping circles as the spatial analytical units when studying bike sharing services for future work, aggregated with more data on this level, such as population, ownership of other transport modes, if available. Additionally, the relationship to public transport, which are modal complementation, modal substitution and modal integration as studied in the relevant work, is recommended to include in the data analysis, demand pattern analysis, and even predictive models. Operational strategies should also be constructed by taking this aspect into account, tailoring delicate strategies for different user groups with different relationships to public transport.

**Predictive models for low-ridership cases:** secondly, it is also interesting to investigate what predictive model(s) can forecast the short-term demand of a relatively small bike sharing project. In this study, only weather and temporal factors are included and thereby the results are not accurate as expected. Other independent factors should be fed into the model to see if they actually improve the performance of the prediction. With accurate predictive models, more operational strategies will be constructed, tailored in those areas of interest even though they are not the hotspots in the current stage.

**Evaluation of the operational strategies:** last but not the least, it is recommended to continue with the exploration in the evaluation of operational strategies in the real-life, and better baselines for evaluation should be introduced in future work, distinguishing the effects from the operational strategies and other factors. Besides, it is better to make the timeline of all KPIs in the same steps. Another suggestion is to apply the operational strategies on a monthly basis, instead of different time scopes, caused by time limits, in this study.

# 8 CONCLUSION

The research gaps stated in chapter 3 are: 1) developing a spatial analytical unit of bike sharing service to improve the efficiency and efficacy of the operation; 2) studying the contributing factors and demand pattern of e-bike sharing service; 3) proposing a framework to evaluate the performance of operational strategies in the real-life context, incorporating additional KPIs besides ridership ratio. To fill these gaps as well as satisfy the need from the commissioner, which is to increase the ridership via the improvement of the service in a cost-effective way, the objective of this research was formulated to develop a data-driven approach to derive beneficial operational strategies so as to improve the service of e-bike sharing, with the follow-up evaluations to see if the developed operational strategies indeed improve the service of e-bike sharing.

The research was carried out with the scope of bondi's e-bike sharing in The Hague in The Netherlands. The main data input related to the operation of e-bike sharing is provided by bondi, combined with other data obtained from the open-source database. The time scope of data is four months, ranging from 19/06/2021, the service launch day, until 19/06/2021, where the data of the first three months were used to construct operational strategies via a data-driven method, including the data analysis and demand pattern analysis and the whole dataset was applied to evaluate the results of the proposed strategies.

This research is consist of four stages: 1) the preparation phase done by literature review; 2) the data analysis including data description, correlation analysis and land use pattern analysis; 3) the demand pattern analysis involving descriptive analysis, general demand pattern analysis and corresponding clustering analysis, supply efficiency analysis, average travel distance and duration analysis, and the following development of operational strategies; 4) the operational phase including the setup and execution of the proposed strategies, and the following evaluation.

The methodology was proposed in a general way, allowing the application for other work, and then these methods were applied in the case study, targeting bondi's e-bike sharing project in The Hague. The results were then presented and discussed.

This study bridged the research gaps and fulfilled the needs of bondi as mentioned at the very beginning of this chapter. Overlapping circles were proposed and applied as the spatial analytical units of the e-bike sharing project, and the results were compared with the counterparts acquired from the traditional units, the administrative zones neighbourhoods, and it turned out the proposed operational strategies on the circle level had a better performance, examined by the following evaluation framework. Second to that, the influential factors and demand pattern of e-bike sharing service was also studied, in the context of bondi in The Hague. The most significant and innovative part of this research is the development of the evaluation framework to examine the performance of operational strategies in a real-life context, targeting the third research gap.

This chapter starts with the presentation of the main findings, followed by the answers to the research questions, and then the contributions are summarised, ending with the recommendations of future work.

## 8.1 Main findings

This section presents the main findings of four stages, the preparation phase, the data analysis, the demand pattern analysis and the operation phase.

**Determinants of the demand:** first of all, the preparation phase discovered the existing methods widely applied in the current studies related to bike sharing projects and figured out the influential factors of e-bike sharing demand. Currently, the research in this field focused on 4 main topics, contributing factors of bike sharing demand, exploration of bike sharing projects, demand prediction, and optimization of the operation. The factors exerting effects on the demand are grouped into 6 types, which are spatial and infrastructure factors, weather-related factors, temporal factors, mobility and trip characteristics, sociodemographic factors and safety factors.

Following the determination of contributing factors of e-bike sharing demand, the research was carried out based on a case study of bondi e-bike sharing service in The Hague.

**Correlation analyses and land use pattern:** a data analysis was conducted to understand the relationship between the abovementioned variables with the demand. Pearson's coefficients and multiple regression analysis were applied to test the linear correlations between them. It was found that the number of POIs, regardless of their types, the supply level, and the availability of public transport have had positive impacts on the ridership, to a different extent, ranging from 0.18 to 0.92. The factors (temperature, precipitation and humidity) under the weather group, however, showed a different pattern: the temperature was found to positively correlate with the demand with the Pearson's coefficient at 0.16 while precipitation and humidity were correlated with the demand in a negative way, with the Pearson's coefficient at -0.03 and -0.31. Moreover, the multiple regression presented the relative importance to the demand of the factors under the same group: humidity exerted the most significant effects on the ridership in the weather groups and the amount of sustenance and offices were the essential ones in the POI category. In addition to the correlation analysis, the functional analysis of different units in The Hague was also done, with the findings that the units in the central area are prone to be recreational-oriented while the office-dominated units are situated across the whole city.

**Demand pattern and development of operational strategies:** secondly, the demand pattern analysis, including the derivation of the operational strategies, was directed. A descriptive analysis was firstly done, revealing that there is only one peak period of bondi's e-bike sharing service in The Hague, different from the widely-observed two peaks in other studies. It is followed by general demand pattern analyses as well as the temporal clustering analyses of the demand pattern, on two different units, neighbourhoods and 400m overlapping circles, via agglomerative hierarchical clustering. The central units were found to be the most popular places, in both general and clustering analyses. Besides, 5 periods emerged from the hourly clustering, which are the first peak hour (16:00 to 16:59), the second peak hour (17:00 to 17:59), the first transition hour (18:00 to 18:59), the second transition hour (19:00 to 19:59) and the off-peak period (20:00 to 15:59). It was observed that from the second peak hour, people moved towards the recreational-oriented zones from the office-oriented ones and the station unit became prevalent with the most rides from the 2nd peak hour until the end of the transition hours. Furthermore, the popularity of the station unit implies that people have used this service as a supplement to their train trips for the first and the last mile. Following this, it was found that the outskirts in the southwest of The Hague were lack in usage, from the supply efficiency and average travel distance/duration analyses. Thereon, three sets of rebalancing strategies and the reduction in the service area were proposed based on these findings.

**The execution and evaluation of the operational strategies:** turning to the last stage, the operation. On top of this, the suggested strategies were applied in different time scopes. Two classes of KPIs were then proposed and computed to assess the effects of these strategies, from the operators' and the users' perspectives. All strategies were proven to improve the service levels. Among these, the third run of reallocations based on hourly clustering on the circle level was proved to improve the service to the largest degree, out of all strategies, with a ridership ratio at 0.67 and a decrease in the origin-based and destination-based vehicle idle time at around 12 and 19 hours, compared to the ID verification period. Besides, the adjustment in the operational zone had a relatively moderate beneficial impact on the service, but it decreased the operational efforts to a large extent. After applying the strategies, the level of ridership ratio is quite decent compared with the other schemes with a similar supply level around the world. Moreover, it was also found that users became more satisfied with the service after the implementation of the suggested strategies, indicated by a increased net retention rate ranging from 52.10% to 86.87% and a grown average user expenditure of around 15 from 9 euros per user. It was suggested multiple operational strategies should be applied together, complementary to each other.

## 8.2 Answer to research questions

The research objective of this study is to develop a data-driven approach to derive beneficial operational strategies, in order to improve the service of e-bike sharing, by conducting data analysis, and demand pattern analysis and the follow-up examination of proposed strategies.

To answer the main research question, sub-questions are firstly answered in a sequence.

1. *What are the determinants from internal trip data and external from other sources affecting the ridership, empirical analysis methods, and widely-applicable operation strategies for e-bike sharing according to literature?*

There are 6 main types of influential factors, spatial and infrastructure, weather-related, mobility and trip characteristics, temporal factors, sociodemographic factors and safety factors, influencing the ridership. In the existing research, demand pattern analysis, involving spatial or/and temporal clustering, was widely applied to understand the demand for shared bike projects. Turning to the operational strategies, the most frequent category in the operational strategies is rebalancing actions, either static or dynamic depending on the methods where the former one is usually taken during nighttime when the demand level is very low and the effects on the ridership are omitted and the latter is executed during day time, taking the dynamics of demand into account.

2. *What are the correlations of the influential factors with the ridership of dockless e-bike sharing?*

In this study, only four categories of independent variables were examined with the correlation to the ridership, which are supply, weather, POIs and the availability of public transport namely. Supply, POIs and the availability of public transport were proven to have a positive correlation with the demand by Pearson's coefficient, and the same holds for the temperature. However, precipitation and humidity were found to correlate with the demand, in a negative way. Additionally, multiple regression analyses were also conducted, complementary to Pearson's coefficient. It was observed that POIs led to a higher model fitness and points of sustenance, office and recreation have positive influences on the ridership in the case study. Consistently, the availability of PT has a beneficial effect on the ridership. Unexpectedly, the coefficient of temperature was found to be negative and the precipitation is insignificant in the regression. The selected aggregation level, 3-hours, is a reason behind it.

3. *What is the demand pattern of e-bike sharing and can it be clustered into different temporal clusters?*

There is only one peak hour in the early evening, from 16:00 to 18:00 and followed by another two transition hours in the e-bike sharing project in the case study, different from the widely-observed two peaks corresponding to two commute stages in the other studies. It implies that people tend to use e-bike sharing when they are not hurried and it is supported by the insights from separate analyses of different hourly clusters where during the peak hours people tend to leave from either the office-dominated areas or the station-nearby unit to recreational areas. Additionally, the popularity of station units also infers the fact that e-bike sharing serves the first mile and the last mile of rail trips.

Based on the 4-months data, temporal clustering was conducted by agglomerative clustering. It was found there are 5 distinguishing hourly clustering of explainability, corresponding to the peak hours (16:00 to 16:59, and 17:00 to 17:59), the transition hours (18:00 to 18:59, and 19:00 to 19:59) and the others (20:00 to 15:59), while the daily clustering, in terms of day of a week, was unclear in the case study. The demand pattern during peak hours, transition hours and the off-peak period is quite different, with the central areas as the hot spots all the time, though. The ambiguity in daily clustering is attributed to the fact that this period is the first four months of the operation of this e-bike sharing project and thereby people did not adopt a habit in terms of day of a week yet.

4. *How can these insights facilitate operational strategies?*

The demand pattern provides the insights of imbalance in different units and rebalancing strategies were applied to alleviate the gap between the demand and the supply accordingly. There were three runs of reallocation strategies, targetting on the neighbourhood level and 400m overlapping circles respectively, consecutive to each other in the timeline. Additionally, insights from vehicle idle time and average trip distance/duration on the unit level suggests a reduction in the service area, by leaving out the areas with the lowest usage.

5. *What key performance indicators can be applied to examine the effects of operational strategies?*

There are two groups of KPIs to evaluate the effects of operational strategies, by the operators' side and the users' side. The first one consists of general daily ridership, ridership ratio, and average vehicle idle time on the vehicle level and the other is composed of net retention rate and average expenditure. The first group considers the service level from the operational side while the latter focuses more on the users, managing to understand the user satisfaction by the data directly available to the provider.

6. *To what extent the proposed strategies can improve the service of e-bike sharing?*

The reallocation strategies have more significant effects on the improvement of the e-bike sharing project. Among the three runs of reallocation, the reallocations considering the hourly dynamics of demand has better results, and the reallocation based on the hourly dynamics on the circle level has the best performance out of three, increasing the ridership ratio from 0.31 to 0.67, and decreasing the vehicle idle time by 12 hours (origin-based) and 19 hours (destination-based) namely.

The reduction in the service zone has also exerted a positive effect on the service, in a moderate way compared to the reallocation, though. The ridership ratio during this period is around 0.5 and there is a 10-hour decline in the vehicle idle time in this period, compared to the ID verification period.

It is no doubt that reallocation strategies have facilitated a better service level while it requires more effort, and the shrinks in the service area have decreased these efforts, maintaining a decent ridership at the same time. These two types of operational strategies are complementary to each other and should be applied together.

The answer to the main research question is stated as follows:

***How can data-driven methods be used for improving operations of e-bike sharing services?***

The proposed data-driven method connects several methods in a sequential way, involving the determination of contributing factors of e-bike sharing, basic data analysis, demand pattern, development of operational strategies and the follow-up assessment of these strategies.

The demand pattern analysis is the vital part to generate the operational strategies. The first two approaches serve as the prerequisite for the demand pattern analysis, and the operational strategies are developed based on the insights gained from this core section. Afterwards, a follow-up evaluation is conduct to assess the effects of the proposed strategies.

The central stage, the demand pattern analysis, composing several steps, and each of them cultivates different operational strategies. The general demand pattern and the demand pattern by different temporal groups reveal the magnitude of flow, and the relationship between the supply and the demand in each unit, facilitating the fosterage of reallocation strategies to mitigate the imbalance. Moreover, the vehicle idle time and average trip duration/distance unveil the popularity of different units. The results support the adjustment on the service area.

In this research, the proposed methods were applied in the case study of bondi's e-bike sharing in The Hague, and a set of operational strategies were obtained based on these data-driven methods. It is proved that the proposed operational strategies obtained from this data-driven method contribute to the service to different degrees, which almost doubled the ridership ratio, and decreased the vehicle idle time by circa 10 hours, with the reference of the ID verification period.

## 8.3 Contributions

The contributions of this research are suggested as follows, in both the scientific and practical ways:

- **The scientific contributions**

1. An innovative spatial analytical unit, the overlapping circles, was proposed in this study, with better performance on the development of cost-effectively reallocation strategies based on the demand pattern analysis, compared to its counterpart on the neighbourhood units;
2. Additional to the proposal of a novel unit as mentioned above, this study compared the results of different units, bridging the gap of the comparison of effects of different spatial units on the development of reallocation strategies;
3. It added a study in the field of e-bike sharing service, and also compared the findings in this work with the existing studies, revealing similarities and dissimilarities. It is a reference of future work in (e-)bike sharing service;
4. It developed a framework to evaluate operational strategies in the real-life context, taking both operators and users into account, which is missing in the existing studies;
5. It introduced two novel KPIs to the evaluation of operational strategies, which are net retention rate, which is widely applied in the business world, and average user expenditure. These two KPIs help understand the user satisfaction and success of shared-mobility services.

- **The practical contributions**

1. The service of e-bike sharing was improved by a set of operational strategies derived from the data-driven approach developed in this study;
2. The proposed operational strategies achieved a better service of e-bike sharing, without consuming too many resources, which improved the operation of bondi e-bike sharing while decreasing the efforts required to operate this service;
3. This study set up a successful example for other bike sharing providers, especially for small companies.
4. The improved service level is beneficial for users, and the whole society. It also benefits the image of the city and contributes to the future of transport.

## 8.4 Limitations, future work and recommendations

➢ **Limitations and future work**

**Limitations:** apart from the main findings stated in 8.1 and the answers to the research questions as mentioned in 8.2, there are still some limitations that need to be addressed in this study. The time scope and availability of data resultes in several problems. The essential one is the unavailability of supply data and thereby it is inferred by the ride records, leading to the unprecise fleet size, to an acceptable degree, though. Additionally, the applied spatial analytical units, overlapping circles, are still quite novel and thereby some data, such as population, income level, ownership of cars/motors, and the number of residential units are still unavailable on this level and thereby is not included in the following analysis, which is a pity. Another important limitation worthy mentioning is the evaluation. The applied evaluation method does not distinguish the effects caused by the operational strategies and other factors and thereby it attributes all changes in service to the results of the operational strategies. However, there are much more factors affecting the service level and to tackle this issue, better reference should be introduced to the evaluation.

**Improvements of data:** as discussed in chapter 7, the data input needs to be further improved, and a longer time period is desirable for this type of research. There is always a time lag between the research and the real-life application while a longer research time scope mitigates this limitation to some degree. For instance, a more accurate predictive model based on a larger dataset is a recommendation for future, helping develop more useful operational strategies. Another important point is the supply data obtained from the time-series vehicle records. It will facilitate the following steps involving the supply data to a large extend.

**Improvements of methods:** furthermore, there are two main recommendations to improve the methods. The first one targets supply inference, as mentioned above, which will be addressed directly by better data input. The other is about the evaluation method. It should only consider the effects caused by the operational strategies, preventing the disturbances from other factors. Predictive models with acceptable accuracy will help to simulate the service level without the influences of operational strategies on the same period, and these then serve as the reference to be compared when assessing the strategies.

- **Future work**

In terms of future work, the overlapping circles, is suggested to be applied continually since the operational strategies on this level are proven to have better effects on the service in this study. More data sources from the contributing factors, are suggested to be incorporated on this level. Apart from that, there should be more work targeting on the short-term demand pattern of small bike sharing projects, with limited ridership levels and records. It will provide additional insights to develop operational strategies, without any doubt. The final remark is that more research is needed to study the evaluation of operational strategies of e-bike sharing projects in a real-life setting and better reference should be incorporated to the current framework applied in this study, eliminating the influences of other factors.

- **Recommendations for bondi**

For the development of bondi, there are also some recommendations, which are divided into suggestions of the construction of reward systems and advices of other possible operational actions.

**Reward system**

**Dynamic pricing:** besides the reallocation conducted by bondi itself, it is suggested to introduce some initiatives for users to relocate the bikes by themselves, which mitigates the operational efforts to a large extent. Some existing schemes are recommended, such as dynamic pricing, which rewards users with lower ride cost if they return the bike in the desirable locations.

**Discounts for frequent users:** to retain existing users, it is beneficial to provide discounts for future rides for them. A point-collection is interesting for this service: users collect one point when they spend one minute for the service, and they will be rewarded by certain minutes after they spend, for instance, 100 minutes in the service. Alternative to reward of free minutes, discount of future rides is also a choice.

**Operational strategies**

**Reallocation before the regular morning:** there is no morning peak observed in the service. One reason for this is that the bikes are not well-reallocated to these locations people prefer to start their morning trips. Thereby, night-time reallocations or early morning reallocations help to tackle this drawback, by providing bikes in those residential zones as well as the station area, serving the last mile to train trips.

# APPENDICES

## Appendix A – Demand pattern clustering analysis on the neighbourhood level

➢ Hourly cluster

It is found from the dendrogram that similar time periods cluster together as Figure A-1 shows. 5 is chosen as the cluster number and the corresponding classes are defined as follows: Midday (10:00 to 15:59), 1st peak (16:00 to 16:59), 2nd peak (17:00 to 17:59), transition hours (18:00 to 19:59) and sleeping hours (20:00 to 9:59).

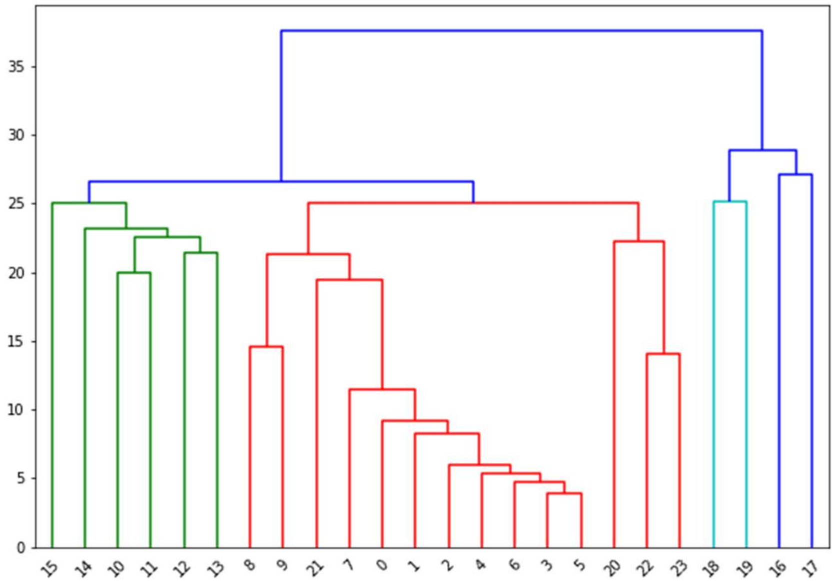



**Figure A-1 Dendrogram of hourly clustering on the neighbourhood level**

1. Midday from 10:00 to 15:59：

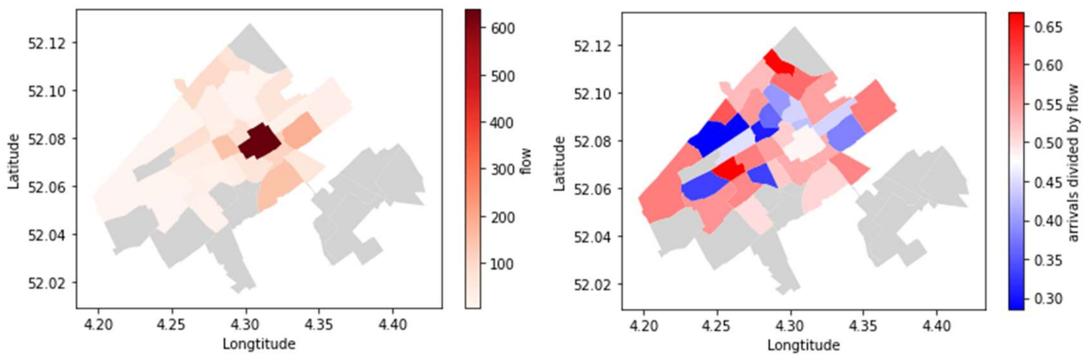



**Figure A-2 Heatmaps of total flow and arrival ratio of the 1st cluster on the neighbourhood level**

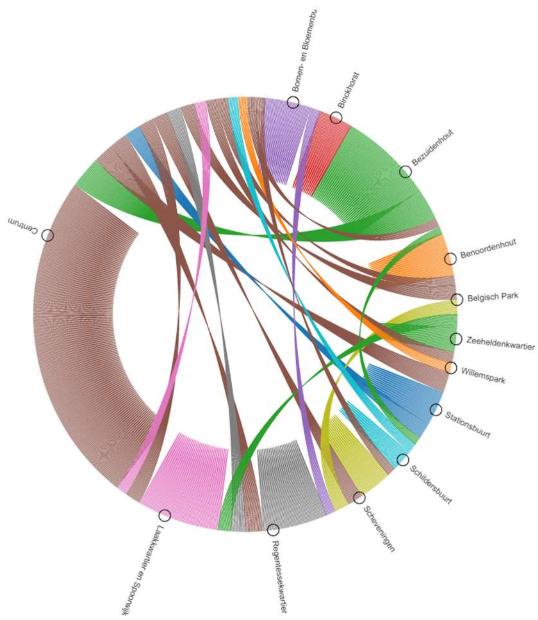



**Figure A-3 Chord diagram of the 1st cluster on the neighbourhood level**

During midday, the densest flow is in Centrum, followed by Bezuidenhout, an office-oriented zone and Laakkwartier en Spoorwijk, a multi-functional zone. Regentessekwartier, which is also a multifunctional one, also catches the attention in terms of the magnitude of flow.

When tracking the chord diagram, Centrum is the most balanced area in the hotspots, the main destinations from Centrum are Zeeheldenkwartier, Stationsbuurt, Laakkwartier en Spoorwijk (mixed), Westbroekpark en Duttendel (office), and Scheveningen (recreation). For Bezuidenhout, people tend to go to the central areas, nearby recreational zone and the beach area. For Laakkwartier en Spoorwijk, however, people only go to the central zones such as Centrum and Stationsbuurt.

During midday, the flow generally comes from multifunctional and office-oriented zones to multifunctional and recreational zones.

2. The first peak from 16:00 to 16:59：

For this first peak hour, the flow in Bezuidenhout and Laakkwartier en Spoorwijk increases, where Bezuidenhout is dominant by departures while the latter is quite a balance. In this period, there are still slightly more arrivals in the centrum, and the beach area start to attract more departures compared to the midday. Besides, the office-oriented zones, such as Bezuidenhout, also begin to witness a departure flow.

Additionally, it is noticeable that Bezuidenhout, Laakkwartier en Spoorwijk and Stationsbuurt become attractive in terms of the flow. Stationsbuurt, as the neighbourhood accommodating two main railway stations (i.e. The Hague Centraal and HS) experiences more arrivals than departures with an increasing flow, implying the potential of e-bikes as the mode to serve the first mile, supplementing the rail trips.

In general, this time period witnessed a flow from office-oriented zones to multifunctional, recreational and residential zones.

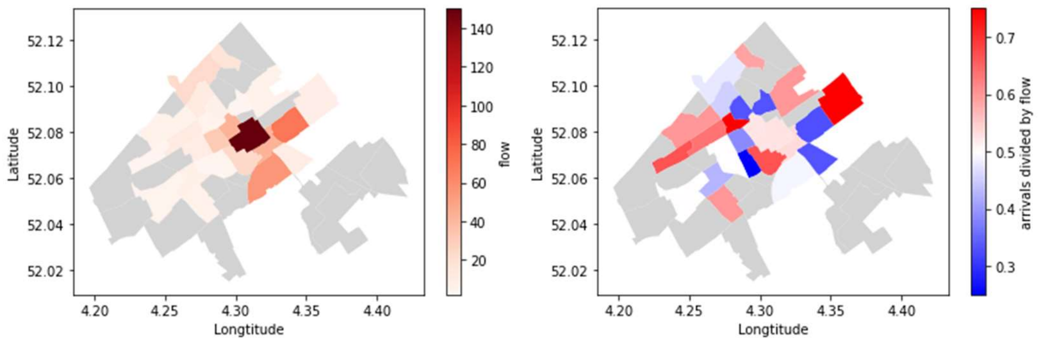



**Figure A-4 Heatmaps of total flow and arrival ratio of the 2nd cluster on the neighbourhood level**

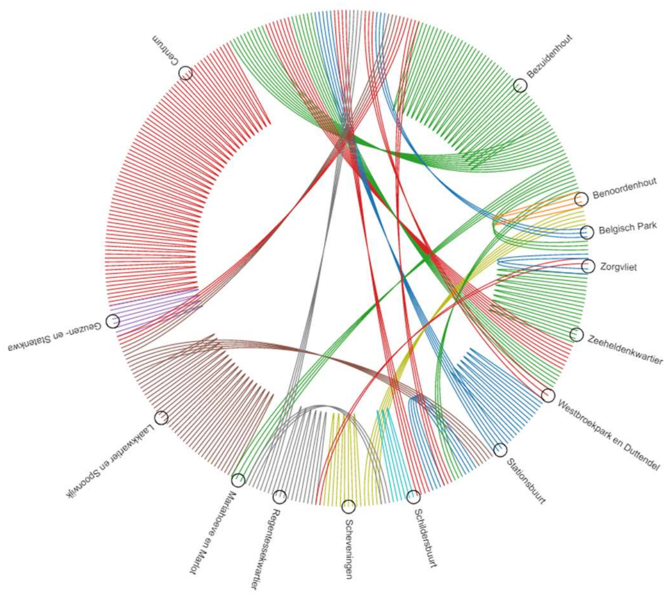



**Figure A-5 Chord diagram of the 2nd cluster on the neighbourhood level**

3. The second peak from 16:00 to 16:59：

From the second peak hour, Centrum begins to accommodate more departures than arrivals and Stationsbuurt continues catching riders though with more departures during this time period, indicating possible last-mile trips complementary to rail trips.

In this time period, the departures are fundamental compositions for flows in the central areas while the flow along the beach is quite balanced. The recreational areas occupied by sports facilities attract lots of arrivals.

In summary, this time period witnesses a flow from multifunctional, office-oriented and recreational zones to recreational, especially those are dominated by sports facilities and residential zones.

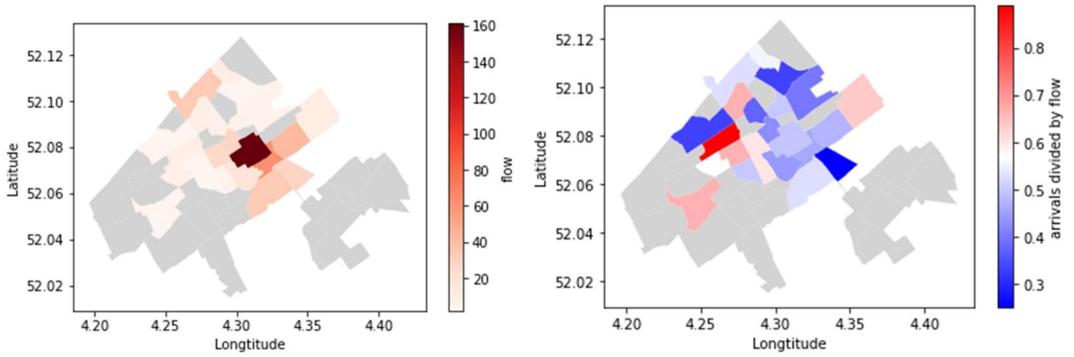


**Figure A-6 Heatmaps of total flow and arrival ratio of the 3rd cluster on the neighbourhood level**

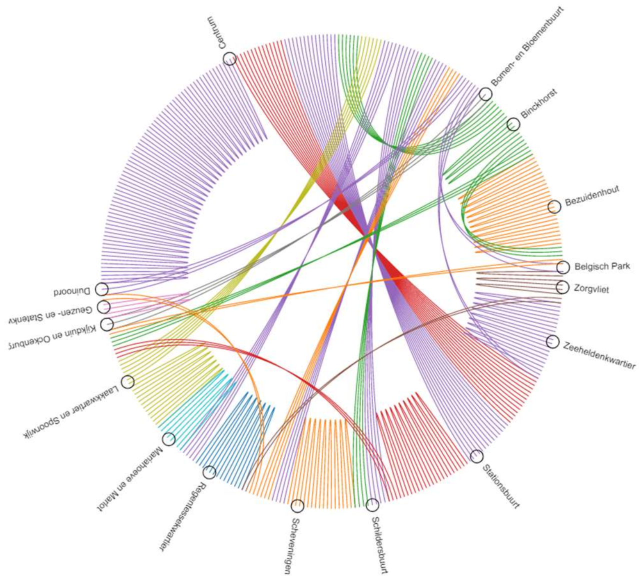


**Figure A-7 Chord diagram of the 3rd cluster on the neighbourhood level**

4. Transition hours from 18:00 to 19:59：

In this transition period, Centrum debounce to a balance status again, with roughly equal arrivals and departures, and the same holds for Stationsbuurt, implying the potential that people use shared bikes as both first mile and last mile trips connecting to the train. Besides, there are slightly more arrivals in Laakkwartier en Spoorwijk and the highest arrival ratio occurs in Benoordenhout, which is a residential zone in Haagse Hout.

In conclusion, although the departures are relatively sparse during this time period, the flow rushes to recreational and residential zones in these two hours.

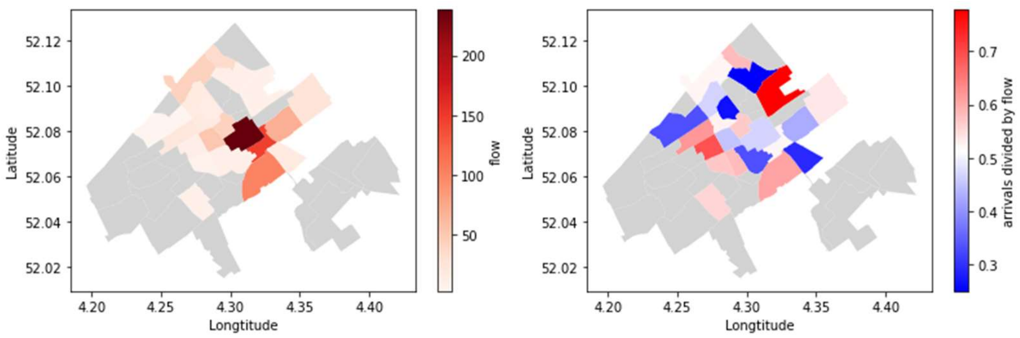



**Figure A-8 Heatmaps of total flow and arrival ratio of the 4th cluster on the neighbourhood level**

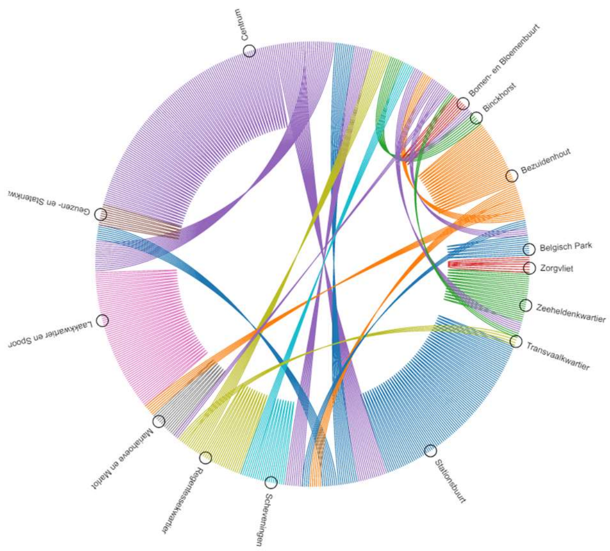



**Figure A-9 Chord diagram of the 4th cluster on the neighbourhood level**

5. Sleeping hours from 20:00 to 8:59：

During this long sleeping period, the hourly flow is inactive, witnessing a decrease in general. The main flow comes towards residential zones and non-residential zones witness an overwhelming departure ratio, especially for the recreational zones.

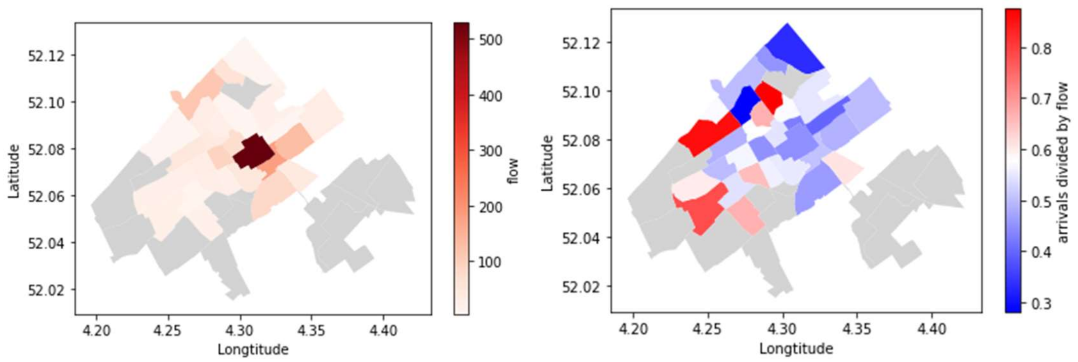



**Figure A-10 Heatmaps of total flow and arrival ratio of the 5th cluster on the neighbourhood level**

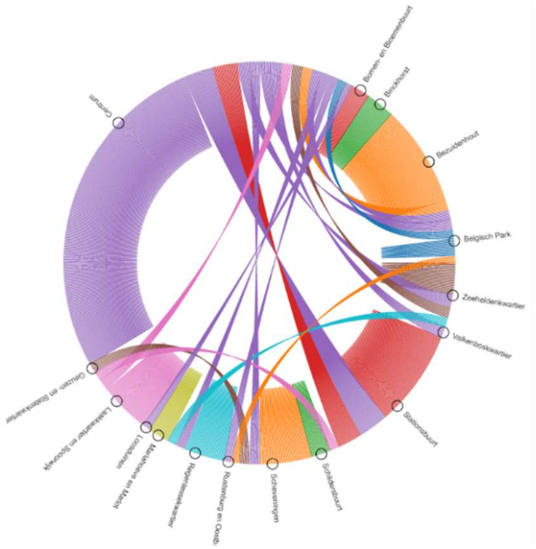



**Figure A-11 Chord diagram of the 5th cluster on the neighbourhood level**

In conclusion, almost half of trips start and end in the same neighbourhood and Centrum is always the most popular neighbourhood no matter what the period is. For Centrum, there are only more arrivals than departures during the midday and the first peak (i.e. 10:00 to 16:59) and the departures are prevalent during other periods. The beach areas attract more arrivals during the midday (10:00 to 15:59) and transition hours (18:00 to 19:59).

Additionally, there are two phenomena parallel with people's daily travel behaviour: For those two peak hours (16:00 to 17:59), there is an obvious departure flows from office-oriented neighbourhoods to zones with other functions, such as residential and recreational functions; During the midday and sleeping periods, flows in residential zones are more tangible, especially for the sleeping period.

- Daily clustering

The first run of daily clustering is based on 42-days (19/06/21-30/07/21) of ride records, several clustering methods and metrics are applied to figure out explainable groups, illustrated in Figure A-12.

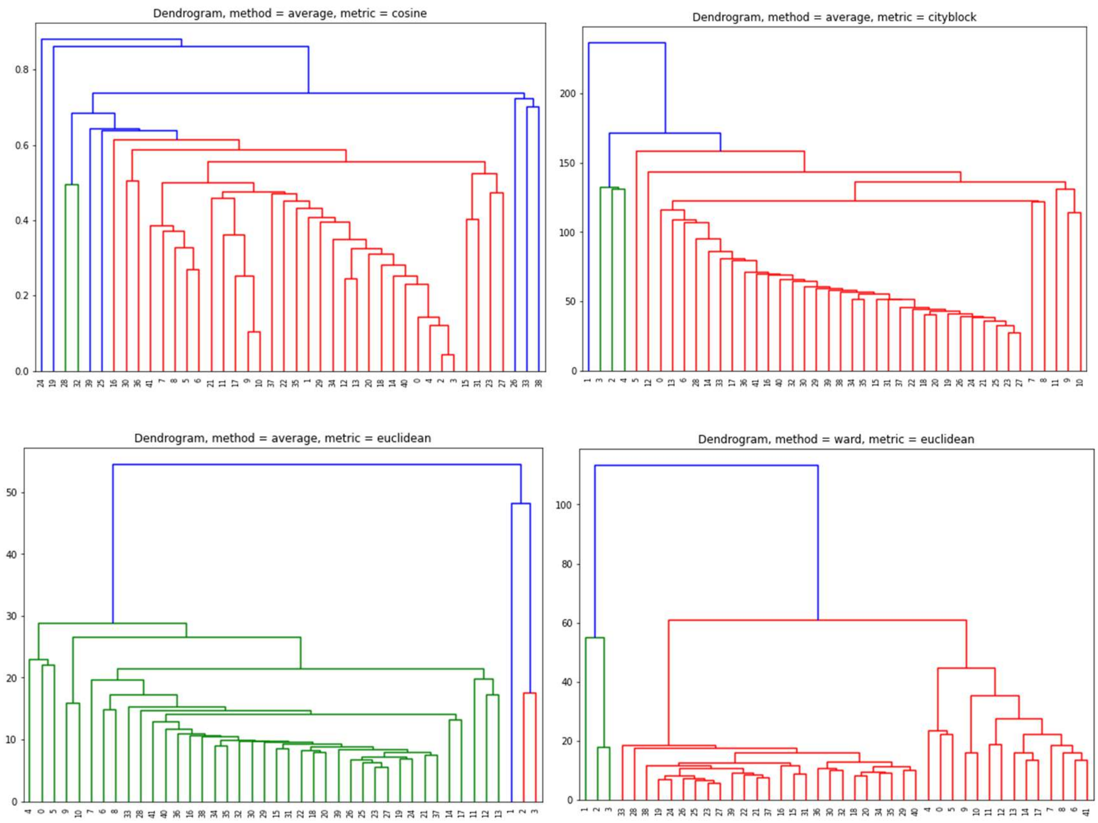



**Figure A-12 Dendrograms of daily clusters based on various metrics and methods of 42 days**

It is found that the free ride days are always clustered together. However, they belong to different days of the week and therefore the agglomerative clustering is redone by rides of 38 days, from 23/06/21 to 30/07/21, excluding the days of free rides as Figure A-13 presents.

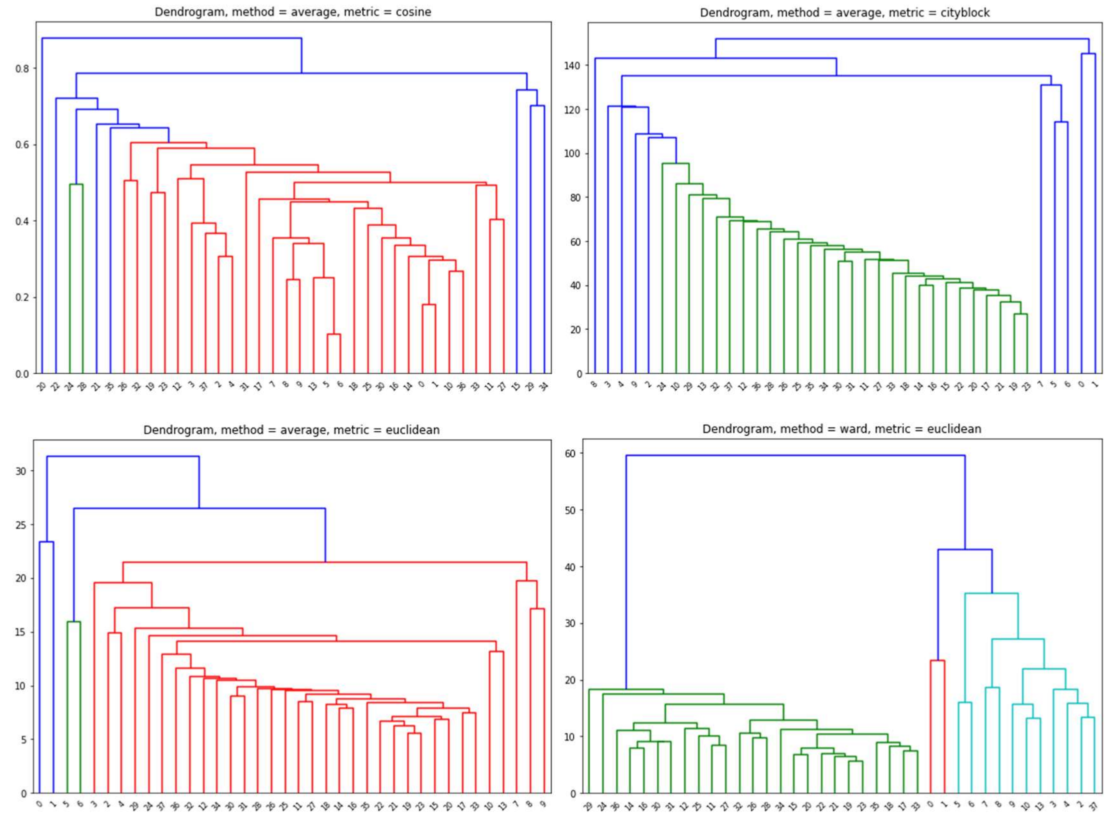



**Figure A-13 Dendrograms of daily clusters based on various metrics and methods of 38 days**

However, the results still reveal unclear daily clustering, at least there are no clear clusters in terms of day of the week. Instead, the days merge into different clusters in their stage of the operational period. For example, given 5 clusters, the first two days cluster together, and the same hold for the other clusters in general. This implies that different days of a week do not have a clear pattern in the given data within 42 or 38 days while the process that people get used to this service play an important role in the clustering of the demand pattern.

## Appendix B – Short-term demand prediction

With the purpose of short-term demand prediction, a deep neural network, LSTM was applied according to the findings in 2.4, allowing to capture the temporal dependency within the dataset.

To set up the data, ride records are firstly aggregated into the 3-hours interval, with corresponding spatial analytical units. Further, additional attributes related to temporal characteristics are included, which are the type of day of a week as a binary variable where 1 represents weekend and 0 represents weekday, day of a week, and month of a year. Besides, weather information involving temperature, precipitation and humidity are added to each data point.

Considering the purpose of the prediction, the model is constructed on a single-unit basis, providing insights of further development in the unit of interest, complementary to the strategies derived from the demand pattern phase. Since the target is the ridership (which is indicated by a number of departures or number of arrivals) for one specific unit, spatial and infrastructure-related attributes are excluded.

Thereby, each data point has the following form:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| timestamp | #departures | #arrivals | weekend | day | month | supply | temperature | precipitation | humidity |

The dataset for predictive models ranges from 19/06/2021 to 12/09/2021, derived from 6066 ride records.

Unit, 105, where Laan van Noi station and lots of offices are situated, is selected as the targeted unit. There are 688 data points in total. 90% of them are training sets and 10% is used as the test set. A number of departures are selected as the dependent variable. A robust scale is applied to standardize the independent variables.

For the long short term memory model, the time step is set up to 8, which means the data from the previous day are also regarded as the input when predicting the dependent variable. The dropout rate is set as 0.2. Besides, to overcome the small size of data points, 10-fold cross-validation is also applied in both models.

Firstly, 500 epochs are used to train the data while it is observed from Figure B-1 that the model starts to overfit since 20 epochs where the value of loss function for cross-validation data starts to increase and the value of loss function for training data starts to decrease. Thereby, 20 epochs are applied to train the model.

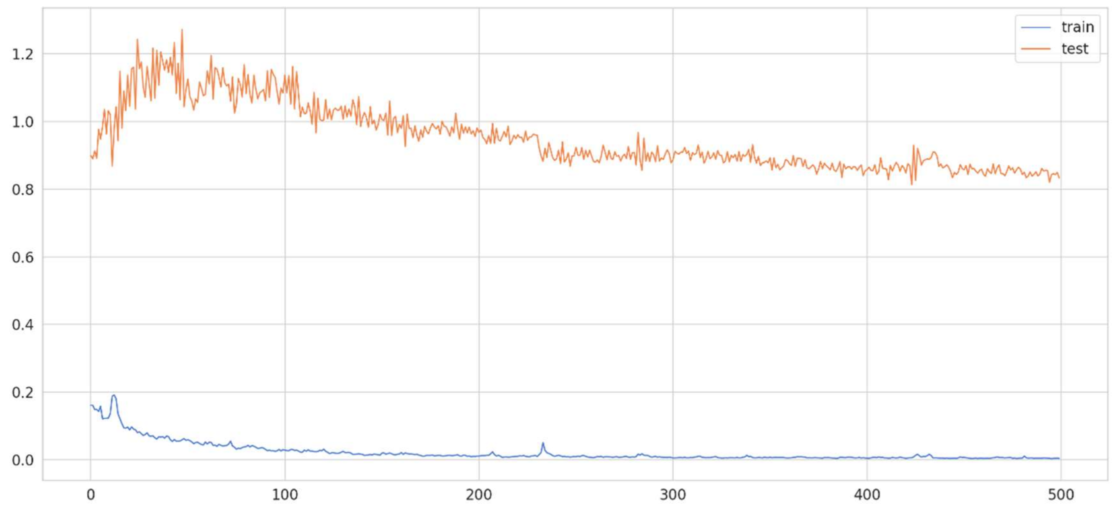



**Figure B-1 Value of loss function in the train set and test set of 500 epochs**

The results are presented in Figure B-2 and Figure B-3. The root-mean-square deviation of the test set is 1.14, indicating the inaccuracy of this model since this value is significant considering the maximal number of departures in the test set is only 3. The poor performance is also found from the fact that the model even predicts the number of departures lower than 0 in the test set.

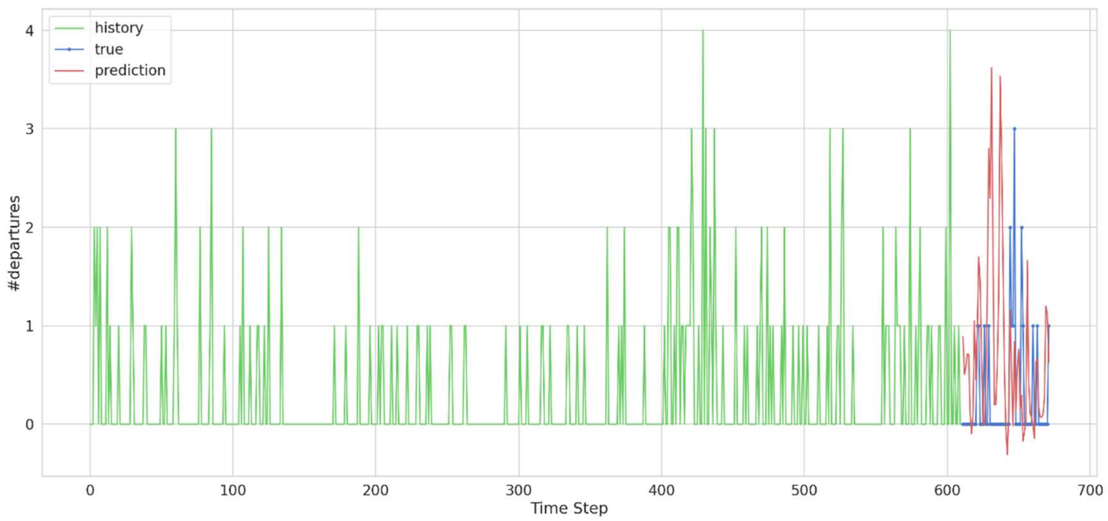



**Figure B-2 Number of departures by time step**

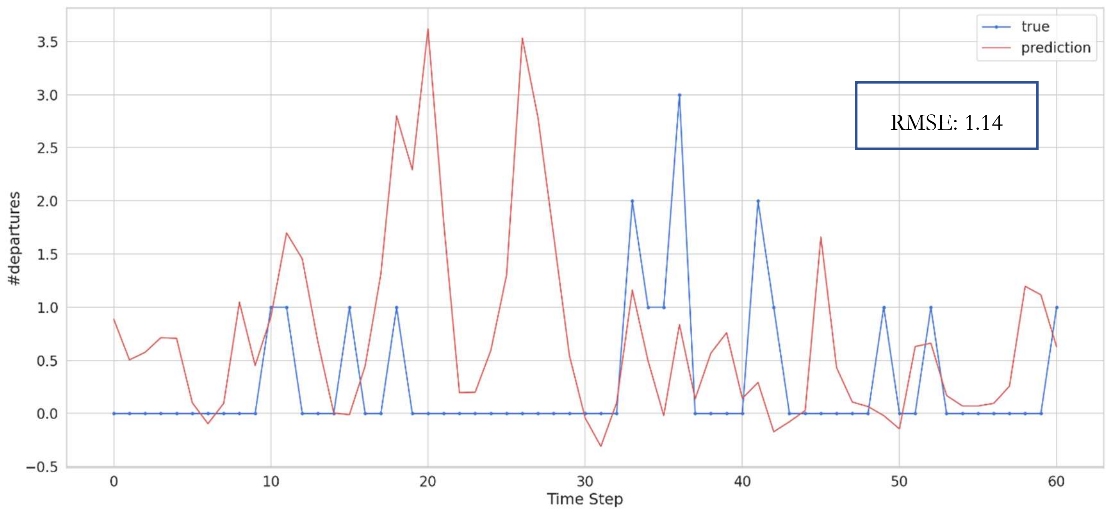



**Figure B-3 Comparisons between the true value and predictive values on the test set**

This model is found to have poor performance in predictive accuracy. Thereby, it has limited use to provide insights for operational strategies and thereby is omitted from the main content of this research.

# Appendix C – Ridership ratio of bike-sharing schemes around the world

**Table C-1 Ridership ratio of different bike sharing projects around the world (Médard de Chardon et al., 2017)**

| | *City* | *Country* | *Brand name* | *Operator* | *\|stations\|* | *\|bikes\|* | *Trips/day/bike* |
|---|---|---|---|---|---|---|---|
| *1* | Barcelona | Spain | Bicing | BSM | 420 | 4852 | 8.4 |
| *2* | Ljubljana | Slovenia | Bicike (LJ) | JCDecaux | 33 | 252 | 8.2 |
| *3* | Dublin | Ireland | dublinbikes | JCDecaux | 49 | 584 | 8 |
| *4* | Turin | Italy | [TO]BIKE | Comunicare | 136 | 495 | 7.9 |
| *5* | Zaragoza | Spain | Bizi | Clear Channel | 130 | 1211 | 7.3 |
| *6* | Valencia | Spain | Valenbisi | JCDecaux | 276 | 2403 | 6.6 |
| *7* | Vilnius | Lithuania | Cyclocity Vilnius | JCDecaux | 33 | 245 | 6 |
| *8* | Lyon | France | Vélo'v | JCDecaux | 346 | 3301 | 5.3 |
| *9* | Paris | France | Vélib' | JCDecaux | 1228 | 17,151 | 5.2 |
| *10* | Milan | Italy | bikeMi | Clear Channel | 187 | 2832 | 5.1 |
| *11* | Tel Aviv | Isreal | Tel-O-Fun | FSM GS Ltd. | 177 | 1411 | 4.9 |
| *12* | Oslo | Norway | Oslo Bysykkel | Clear Channel | 100 | 882 | 4.8 |
| *13* | New York City | US | CitiBike | ABS/Motivate | 357 | 5208 | 4.7 |
| *14* | Bordeaux | France | VCub | Keolis | 139 | 1279 | 4.7 |
| *15* | Boston | US | Hubway | ABS/Motivate | 115 | 1037 | 4.2 |
| *16* | Seville | Spain | Sevici | JCDecaux | 260 | 2203 | 3.9 |
| *17* | Nantes | France | bicloo | JCDecaux | 102 | 887 | 3.8 |
| *18* | Toulouse | France | VélÔToulouse | JCDecaux | 256 | 2193 | 3.8 |
| *19* | Lille | France | V'lille | Keolis | 214 | 2038 | 3.6 |
| *20* | Montreal | Canada | Bixi | PBSC/Bixi | 421 | 4044 | 3.6 |
| *21* | Nancy | France | vélOstan'lib | JCDecaux | 29 | 245 | 3.1 |
| *22* | Washington DC | US | Capital Bikeshare | ABS/Motivate | 297 | 2278 | 3 |
| *23* | La Rochelle | France | Yélo | RTCR | 57 | 210 | 2.9 |
| *24* | Marseille | France | Le Vélo | JCDecaux | 123 | 661 | 2.9 |
| *25* | Chicago | US | Divvy | ABS/Motivate | 300 | 2191 | 2.8 |
| *26* | Gothenburg | Sweden | Styr & Ställ | JCDecaux | 57 | 728 | 2.7 |
| *27* | Miami | US | DecoBike Miami Beach | decobike | 94 | 601 | 2.6 |
| *28* | Nice | France | Vélo Bleu | Veolia Transdev | 178 | 1401 | 2.4 |
| *29* | Rennes | France | Le vélo STAR | Keolis | 83 | 779 | 2.4 |
| *30* | Rio | Brazil | Bike Rio | Serttel | 46 | 280 | 2.4 |
| *31* | Valladolid | Spain | Vallabici | Ingenia Soluciones | 29 | 181 | 2.2 |
| *32* | London | UK | Santander Cycles | Serco | 748 | 11,864 | 2 |
| *33* | Toronto | Canada | Bike Share Toronto | PBSC/Bixi | 80 | 769 | 2 |
| *34* | Rouen | France | cy'clic | JCDecaux | 21 | 193 | 1.9 |
| *35* | Calais | France | Vel'ln | Veolia Transdev | 36 | 213 | 1.9 |
| *36* | Montpellier | France | Vélomagg' | Veolia Transdev | 49 | 280 | 1.9 |
| *37* | Orleans | France | vélo'+ | keolis | 33 | 309 | 1.8 |
| *38* | Vienna | Austria | Citybike Wien | Gewista | 95 | 1072 | 1.8 |
| *39* | San Francisco | US | Bay Area Bike Share | ABS/Motivate | 68 | 611 | 1.8 |
| *40* | Mulhouse | France | Vélocité | JCDecaux | 40 | 245 | 1.7 |
| *41* | Besancon | France | Vélocité | JCDecaux | 30 | 203 | 1.5 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 42 | Delft | Netherlands | Mobike | Mobike | | 800 | 1.5 |
| 43 | Denver | US | Denver B-cycle | Denver B-cycle | 80 | 569 | 1.5 |
| 44 | Belfort | France | Optymo | Optymo | 25 | 201 | 1.3 |
| 45 | Amiens | France | Velam | JCDecaux | 26 | 240 | 1.2 |
| 46 | Madison | US | Madison B-cycle | B-cycle | 32 | 245 | 1.1 |
| 47 | Columbus | US | CoGo | ABS/Motivate | 30 | 225 | 1.1 |
| 48 | Brussels | Belgium | VillO! | JCDecaux | 323 | 3708 | 1.1 |
| 49 | Sao Paulo | Brazil | Bike Sampa | Serttel | 95 | 571 | 1 |
| 50 | Minneapolis | US | Nice Ride Minnesota | NRM | 169 | 1399 | 1 |
| 51 | Saint Etienne | France | VéliVert | Veolia Transdev | 33 | 229 | 0.92 |
| 52 | Ottawa | Canada | Capital BIXI | PBSC/Bixi | 25 | 244 | 0.89 |
| 53 | Namur | Belgium | Li Bia Velo | JCDecaux | 24 | 190 | 0.86 |
| 54 | Houston | US | Houston B-cycle | Houston B-cycle | 28 | 200 | 0.8 |
| 55 | Nashville | US | Nashville B-cycle | Nashville B-cycle | 21 | 166 | 0.79 |
| 56 | Melbourne | Australia | Melbourne Bike Share | ABS/Motivate | 51 | 546 | 0.71 |
| 57 | Caen | France | V'éol | Clear Channel | 40 | 350 | 0.69 |
| 58 | Luxembourg | Luxembourg | vel'oh! | JCDecaux | 72 | 684 | 0.67 |
| 59 | Pau | France | IDECycle | keolis | 22 | 199 | 0.66 |
| 60 | **Den Haag** | **Netherlands** | **bondi** | **bondi** | | **130** | **0.66** |
| 61 | Alacant | Spain | Alabici | Tevaseñal SA | 24 | 120 | 0.62 |
| 62 | Charlotte | US | Charlotte B-cycle | Charlotte B-cycle | 21 | 164 | 0.58 |
| 63 | Dijon | France | Vélodi | Clear Channel | 40 | 401 | 0.56 |
| 64 | Boulder | US | Boulder B-cycle | Boulder B-cycle | 22 | 132 | 0.55 |
| 65 | Avignon | France | VéloPop | TCRA | 20 | 173 | 0.54 |
| 66 | Fort Lauderdale | US | Broward B-cycle | Broward B-cycle | 25 | 154 | 0.54 |
| 67 | Cergy-Pontoise | France | vélO2 | JCDecaux | 43 | 318 | 0.54 |
| 68 | Chattanooga | US | Bike Chattanooga | ABS/Motivate | 33 | 262 | 0.47 |
| 69 | Santander | Spain | TusBic | JCDecaux | 15 | 175 | 0.46 |
| 70 | Valence | France | Libélo | Veolia Transdev | 20 | 164 | 0.43 |
| 71 | Clermont-Ferrand | France | C.vélo | Vélogik | 10 | 104 | 0.42 |
| 72 | San Francisco | US | San Antonio B-cycle | B-cycle | 52 | 388 | 0.42 |
| 73 | Brisbane | Australia | CityCycle | JCDecaux | 151 | 1856 | 0.32 |
| 74 | Bari | Italy | BariinBici | Comunicare | 32 | 44 | 0.29 |
| 75 | Fort Worth | US | Fort Worth B-cycle | FW B-cycle | 34 | 267 | 0.28 |
| 76 | Vannes | France | Vélocéa | Veolia Transdev | 25 | 153 | 0.26 |
| 77 | Perpignan | France | BIP! | Clear Channel | 15 | 123 | 0.22 |

# BIBLIOGRAPHY


Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., & Ju, Y. (2019). A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Computing and Applications*, *31*(5), 1665–1677. https://doi.org/10.1007/s00521-018-3470-9

Albuquerque, V., Sales Dias, M., & Bacao, F. (2021). Machine Learning Approaches to Bike-Sharing Systems: A Systematic Literature Review. *ISPRS International Journal of Geo-Information*, *10*(2), 62. https://doi.org/10.3390/ijgi10020062

Almannaa, M. H., Elhenawy, M., & Rakha, H. A. (2020). Dynamic linear models to predict bike availability in a bike sharing system. *International Journal of Sustainable Transportation*, *14*(3), 232–242. https://doi.org/10.1080/15568318.2019.1611976

Alvarez-Valdes, R., Belenguer, J. M., Benavent, E., Bermudez, J. D., Muñoz, F., Vercher, E., & Verdejo, F. (2016). Optimizing the level of service quality of a bike-sharing system. *Omega*, *62*, 163–175. https://doi.org/10.1016/j.omega.2015.09.007

Angelopoulos, A., Gavalas, D., Konstantopoulos, C., Kypriadis, D., & Pantziou, G. (2018). Incentivized vehicle relocation in vehicle sharing systems. *Transportation Research Part C: Emerging Technologies*, *97*, 175–193. https://doi.org/10.1016/j.trc.2018.10.016

Bao, J., Yu, H., & Wu, J. (2019a). Short-term FFBS demand prediction with multi-source data in a hybrid deep learning framework. *Iet Intelligent Transport Systems*, *13*, 1340–1347.

Bao, J., Yu, H., & Wu, J. (2019b). Short-term FFBS demand prediction with multi-source data in a hybrid deep learning framework. *IET Intelligent Transport Systems*, *13*(9), 1340–1347. https://doi.org/10.1049/iet-its.2019.0008

Barbour, N., Zhang, Y., & Mannering, F. (2019). A statistical analysis of bike sharing usage and its potential as an auto-trip substitute. *Journal of Transport & Health*, *12*, 253–262. https://doi.org/10.1016/j.jth.2019.02.004

basics: Walking distance to transit. (2011, April 24). *Human Transit*.

https://humantransit.org/2011/04/basics-walking-distance-to-transit.html

Bicycle-sharing system. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Bicycle-sharing_system&oldid=1047900760

Böcker, L., & Thorsson, S. (2014). Integrated Weather Effects on Cycling Shares, Frequencies, and Durations in Rotterdam, the Netherlands. *Weather, Climate, and Society*, *6*(4), 468–481. https://doi.org/10.1175/WCAS-D-13-00066.1

Boor, S. (2019). *Impacts of 4th generation bike-sharing*. 92.

Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., & Borecki, N. (2013). Are Bikeshare Users Different from Regular Cyclists?: A First Look at Short-Term Users, Annual Members, and Area Cyclists in the Washington, D.C., Region. *Transportation Research Record: Journal of the Transportation Research Board*, *2387*(1), 112–119. https://doi.org/10.3141/2387-13

Caggiani, L., Camporeale, R., Ottomanelli, M., & Szeto, W. Y. (2018). A modeling framework for the dynamic management of free-floating bike-sharing systems. *Transportation Research Part C: Emerging Technologies*, *87*, 159–182. https://doi.org/10.1016/j.trc.2018.01.001

Caggiani, L., Ottomanelli, M., Camporeale, R., & Binetti, M. (2017). Spatio-temporal Clustering and Forecasting Method for Free-Floating Bike Sharing Systems. In J. Świątek & J. M. Tomczak (Eds.), *Advances in Systems Science* (pp. 244–254). Springer International Publishing.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Campbell, A. A., Cherry, C. R., Ryerson, M. S., & Yang, X. (2016). Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transportation Research Part C: Emerging Technologies*, *67*, 399–414. https://doi.org/10.1016/j.trc.2016.03.004

Cats, O., Krishnakumari, P., Arbez, N., Chiabaut, N., & van Lint, H. (2020, January). *Empirical Evaluation of the Efficiency and Effectiveness of a Ride-Sourcing Service*. 99th Transportation Research Board Annual Meeting, Washington DC.

Cervero, R., & Duncan, M. (2003). Walking, Bicycling, and Urban Landscapes: Evidence From the San

Francisco Bay Area. *American Journal of Public Health*, *93*(9), 1478–1483. https://doi.org/10.2105/AJPH.93.9.1478

Chemla, D., Meunier, F., & Wolfler Calvo, R. (2013). Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, *10*(2), 120–146. https://doi.org/10.1016/j.disopt.2012.11.005

Chen, L., Zhang, D., Wang, L., Yang, D., Ma, X., Li, S., Wu, Z., Pan, G., Nguyen, T.-M.-T., & Jakubowicz, J. (2016). Dynamic cluster-based over-demand prediction in bike sharing systems. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 841–852. https://doi.org/10.1145/2971648.2971652

Chen, M., Wang, D., Sun, Y., Waygood, E. O. D., & Yang, W. (2020). A comparison of users' characteristics between station-based bikesharing system and free-floating bikesharing system: Case study in Hangzhou, China. *Transportation*, *47*(2), 689–704. https://doi.org/10.1007/s11116-018-9910-7

Chen, Z., van Lierop, D., & Ettema, D. (2020). Dockless bike-sharing systems: What are the implications? *Transport Reviews*, *40*(3), 333–353. https://doi.org/10.1080/01441647.2019.1710306

Chen, Z., Wang, B., & Gorban, A. N. (2020). Multivariate Gaussian and Student-t process regression for multi-output prediction. *Neural Computing and Applications*, *32*(8), 3005–3028. https://doi.org/10.1007/s00521-019-04687-8

Daddio, D. (2012). *Maximizing Bicycle Sharing: An Empirical Analysis of Capital Bikeshare Usage*. 45.

Dell'Amico, M., Hadjicostantinou, E., Iori, M., & Novellani, S. (2014). The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, *45*, 7–19. https://doi.org/10.1016/j.omega.2013.12.001

DeMaio, P. (2009). Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*, *12*(4), 41–56. https://doi.org/10.5038/2375-0901.12.4.3

*Den Haag in Cijfers—Average—Districts*. (2021). https://denhaag.incijfers.nl/Jive?sel_guid=233d9374-cb29-44eb-bc6a-c42762ffa2f2

*Den Haag Tram & RandstadRail*. (2021). http://www.urbanrail.net/eu/nl/dhg/den-haag.htm

El-Assi, W., & Mahmoud, M. S. (2015). *Effects of Built Environment and Weather on Bike Sharing Demand: A*

*Station Level Analysis of Commercial Bike Sharing in Toronto*. 39.

Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, *54*, 101882. https://doi.org/10.1016/j.scs.2019.101882

European Commission. (2011). *Transport White Paper* (p. 31). European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0144&from=EN

Faghih-Imani, A., Hampshire, R., Marla, L., & Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice*, *97*, 177–191. https://doi.org/10.1016/j.tra.2016.12.007

Fishman, E. (2016). Bikeshare: A Review of Recent Literature. *Transport Reviews*, *36*(1), 92–113. https://doi.org/10.1080/01441647.2015.1033036

Fishman, E., & Cherry, C. (2016). E-bikes in the Mainstream: Reviewing a Decade of Research. *Transport Reviews*, *36*(1), 72–91. https://doi.org/10.1080/01441647.2015.1069907

Fishman, E., Washington, S., & Haworth, N. (2012). *Understanding the fear of bicycle riding in Australia*. *23*, 9.

Fishman, E., Washington, S., & Haworth, N. (2013). Bike Share: A Synthesis of the Literature. *Transport Reviews*, *33*(2), 148–165. https://doi.org/10.1080/01441647.2013.775612

Fuller, D., Gauvin, L., Kestens, Y., Daniel, M., Fournier, M., Morency, P., & Drouin, L. (2011). Use of a New Public Bicycle Share Program in Montreal, Canada. *American Journal of Preventive Medicine*, *41*(1), 80–83. https://doi.org/10.1016/j.amepre.2011.03.002

Galatoulas, N.-F., Genikomsakis, K. N., & Ioakimidis, C. S. (2020). Spatio-Temporal Trends of E-Bike Sharing System Deployment: A Review in Europe, North America and Asia. *Sustainability*, *12*(11), 4611. https://doi.org/10.3390/su12114611

Gavalas, D. (2016). *Chapter 13—Design and management of vehicle-sharing systems: A survey of algorithmic approaches*. 29.

Gu, T., Kim, I., & Currie, G. (2019). To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. *Transportation Research Part A: Policy and Practice*, *119*, 122–147. https://doi.org/10.1016/j.tra.2018.11.007

*Guide to Net Dollar Retention (NDR)—Definition, Calculation, Tips*. (2021, September 19). UserGuiding. https://userguiding.com/blog/net-dollar-retention-ndr/

Guido, C., Rafal, K., & Constantinos, A. (2019). A low dimensional model for bike sharing demand forecasting. *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1–7. https://doi.org/10.1109/MTITS.2019.8883283

Guidon, S., Reck, D. J., & Axhausen, K. W. (2019). *Expanding a (electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests*. 23.

Hampshire, R. C. (2012). *An Analysis of Bike Sharing Usage: Explaining Trip Generation 2 and Attraction from Observed Demand*. 17.

He, Y., Song, Z., Liu, Z., & Sze, N. N. (2019). Factors Influencing Electric Bike Share Ridership: Analysis of Park City, Utah. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(5), 12–22. https://doi.org/10.1177/0361198119838981

*Holland-Cycling.com—Finding your way*. (n.d.). Retrieved 21 October 2021, from https://www.holland-cycling.com/the-hague/getting-around-the-hague/finding-your-way-the-hague?highlight=WyJoYWd1ZSIsImhhZ3VlJ3MiXQ==

Hsu, C.-C., Liou, J. J. H., Lo, H.-W., & Wang, Y.-C. (2018). Using a hybrid method for evaluating and improving the service quality of public bike-sharing systems. *Journal of Cleaner Production*, *202*, 1131–1144. https://doi.org/10.1016/j.jclepro.2018.08.193

Hu, Y., Zhang, Y., Lamb, D., Zhang, M., & Jia, P. (2019). Examining and optimizing the BCycle bike-sharing system – A pilot study in Colorado, US. *Applied Energy*, *247*, 1–12. https://doi.org/10.1016/j.apenergy.2019.04.007

Hua, M. (2020). *Forecasting usage and bike distribution of dockless bike-sharing using journey data*. https://doi.org/10.1049/iet-its.2020.0305

Huang, F., Qiao, S., Peng, J., & Guo, B. (2019). A Bimodal Gaussian Inhomogeneous Poisson Algorithm for Bike Number Prediction in a Bike-Sharing System. *IEEE Transactions on Intelligent Transportation Systems*, *20*(8), 2848–2857. https://doi.org/10.1109/TITS.2018.2868483

Ji, S., Cherry, C. R., Han, L. D., & Jordan, D. A. (2014). Electric bike sharing: Simulation of user demand and system availability. *Journal of Cleaner Production*, *85*, 250–257. https://doi.org/10.1016/j.jclepro.2013.09.024

Jia, W., Tan, Y., & Li, J. (2018). Hierarchical Prediction Based on Two-Level Affinity Propagation Clustering for Bike-Sharing System. *IEEE Access*, *6*, 45875–45885. https://doi.org/10.1109/ACCESS.2018.2865658

Jin, X., & Tong, D. (2020). Station-Free Bike Rebalancing Analysis: Scale, Modeling, and Computational Challenges. *ISPRS International Journal of Geo-Information*, *9*(11), 691. https://doi.org/10.3390/ijgi9110691

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, *6*(4), 455–466. https://doi.org/10.1016/j.pmcj.2010.07.002

Kloimüllner, C., Papazek, P., Hu, B., & Raidl, G. R. (2014). Balancing Bicycle Sharing Systems: An Approach for the Dynamic Case. In C. Blum & G. Ochoa (Eds.), *Evolutionary Computation in Combinatorial Optimisation* (Vol. 8600, pp. 73–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44320-0_7

Kong, H., Jin, S. T., & Sui, D. Z. (2020). Deciphering the relationship between bikesharing and public transit: Modal substitution, integration, and complementation. *Transportation Research Part D: Transport and Environment*, *85*, 102392. https://doi.org/10.1016/j.trd.2020.102392

Li, A., Zhao, P., Huang, Y., Gao, K., & Axhausen, K. W. (2020). An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China. *Journal of Transport Geography*, *88*, 102828. https://doi.org/10.1016/j.jtrangeo.2020.102828

Li, D., & Zhao, Y. (2019). A Multi-Categorical Probabilistic Approach for Short-Term Bike Sharing Usage Prediction. *IEEE Access*, *7*, 81364–81369. https://doi.org/10.1109/ACCESS.2019.2923766

Li, S., Zhuang, C., Tan, Z., Gao, F., Lai, Z., & Wu, Z. (2021). Inferring the trip purposes and uncovering spatio-temporal activity patterns from dockless shared bike dataset in Shenzhen, China. *Journal of*

*Transport Geography*, *91*, 102974. https://doi.org/10.1016/j.jtrangeo.2021.102974

Li, Y., & Zheng, Y. (2020). Citywide Bike Usage Prediction in a Bike-Sharing System. *IEEE Transactions on Knowledge and Data Engineering*, *32*(6), 1079–1091. https://doi.org/10.1109/TKDE.2019.2898831

Li, Y., Zheng, Y., Zhang, H., & Chen, L. (2015). Traffic prediction in a bike-sharing system. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1–10. https://doi.org/10.1145/2820783.2820837

Lin, J.-R., & Yang, T.-H. (2011). Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E: Logistics and Transportation Review*, *47*(2), 284–294. https://doi.org/10.1016/j.tre.2010.09.004

Lin, L., He, Z., & Peeta, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies*, *97*, 258–276. https://doi.org/10.1016/j.trc.2018.10.011

*Line network map—HTM*. (2021). https://www.htm.nl/reizen/lijnennetkaart

Liu, J., Sun, L., Chen, W., & Xiong, H. (2016). Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1005–1014. https://doi.org/10.1145/2939672.2939776

Liu, S., Hou, J., Liu, Y., Khadka, A., & Liu, Z. (2018). OD Demand Forecasting for the Large-Scale Dockless Sharing Bike System: A Deep Learning Approach. *CICTP 2018*, 1683–1692. https://doi.org/10.1061/9780784481523.167

Liu, T. L. K., Krishnakumari, P., & Cats, O. (2019). Exploring Demand Patterns of a Ride-Sourcing Service using Spatial and Temporal Clustering. *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1–9. https://doi.org/10.1109/MTITS.2019.8883312

Ma, X., Cao, R., & Jin, Y. (2019). Spatiotemporal Clustering Analysis of Bicycle Sharing System with Data Mining Approach. *Information*, *10*(5), 163. https://doi.org/10.3390/info10050163

Ma, X., Yuan, Y., Van Oort, N., & Hoogendoorn, S. (2020). Bike-sharing systems' impact on modal shift: A case study in Delft, the Netherlands. *Journal of Cleaner Production*, *259*, 120846.

https://doi.org/10.1016/j.jclepro.2020.120846

Martens, K. (2004). The bicycle as a feedering mode: Experiences from three European countries. *Transportation Research Part D: Transport and Environment*, *9*(4), 281–294. https://doi.org/10.1016/j.trd.2004.02.005

Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A: Policy and Practice*, *41*(4), 326–338. https://doi.org/10.1016/j.tra.2006.09.010

Mátrai, T., & Tóth, J. (2016). Comparative Assessment of Public Bike Sharing Systems. *Transportation Research Procedia*, *14*, 2344–2351. https://doi.org/10.1016/j.trpro.2016.05.261

Mattson, J., & Godavarthy, R. (2017). Bike share in Fargo, North Dakota: Keys to success and factors affecting ridership. *Sustainable Cities and Society*, *34*, 174–182. https://doi.org/10.1016/j.scs.2017.07.001

Médard de Chardon, C., Caruso, G., & Thomas, I. (2017). Bicycle sharing system 'success' determinants. *Transportation Research Part A: Policy and Practice*, *100*, 202–214. https://doi.org/10.1016/j.tra.2017.04.020

Miranda-Moreno, L. F., & Nosal, T. (2011). Weather or Not to Cycle: Temporal Trends and Impact of Weather on Cycling in an Urban Environment. *Transportation Research Record: Journal of the Transportation Research Board*, *2247*(1), 42–52. https://doi.org/10.3141/2247-06

Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-Scale Vehicle Sharing Systems: Analysis of Vélib'. *International Journal of Sustainable Transportation*, *7*(1), 85–106. https://doi.org/10.1080/15568318.2012.660115

Noussan, Carioni, Sanvito, & Colombo. (2019). Urban Mobility Demand Profiles: Time Series for Cars and Bike-Sharing Use as a Resource for Transport and Energy Modeling. *Data*, *4*(3), 108. https://doi.org/10.3390/data4030108

Ogilvie, F., & Goodman, A. (2012). Inequalities in usage of a public bicycle sharing scheme: Socio-demographic predictors of uptake and usage of the London (UK) cycle hire scheme. *Preventive Medicine*, *55*(1), 40–45. https://doi.org/10.1016/j.ypmed.2012.05.002

*OpenStreetMap*. (2021). OpenStreetMap. https://www.openstreetmap.org/

Pal, A., & Zhang, Y. (2017). Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies*, *80*, 92–116. https://doi.org/10.1016/j.trc.2017.03.016

Politis, I., Fyrogenis, I., Papadopoulos, E., Nikolaidou, A., & Verani, E. (2020). Shifting to Shared Wheels: Factors Affecting Dockless Bike-Sharing Choice for Short and Long Trips. *Sustainability*, *12*(19), 8205. https://doi.org/10.3390/su12198205

Rainer Lesniewski, A. (2021). *Administrative and political map of the Dutch city The Hague Stock Vector Image & Art—Alamy*. https://www.alamy.com/administrative-and-political-map-of-the-dutch-city-the-hague-image150311031.html

Raviv, T., Tzur, M., & Forma, I. A. (2013). Static repositioning in a bike-sharing system: Models and solution approaches. *EURO Journal on Transportation and Logistics*, *2*(3), 187–229. https://doi.org/10.1007/s13676-012-0017-6

Reck, D. J., Haitao, H., Guidon, S., & Axhausen, K. W. (2021). Explaining shared micromobility usage, competition and mode choice by modelling empirical data from Zurich, Switzerland. *Transportation Research Part C: Emerging Technologies*, *124*, 102947. https://doi.org/10.1016/j.trc.2020.102947

*Reizigersgedrag | NS Dashboard*. (2021). https://dashboards.nsjaarverslag.nl/reizigersgedrag/den-haag-centraal

Rixey, R. A. (2013). *Station-Level Forecasting of Bike Sharing Ridership: Station Network Effects in Three U.S. Systems*. 17.

Rokach, L., & Maimon, O. (2005). Clustering Methods. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 321–352). Springer-Verlag. https://doi.org/10.1007/0-387-25465-X_15

Sabir, M. (2011). *Weather and Travel Behaviour*. 153.

Sarker, R. I., Mailer, M., & Sikder, S. K. (2019). Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany. *Smart and Sustainable Built Environment*, *9*(1), 38–53. https://doi.org/10.1108/SASBE-07-2017-0031

Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future. *Transportation Research Record: Journal of the Transportation Research Board*, *2143*(1), 159–167. https://doi.org/10.3141/2143-20

Shaheen, S. A., Martin, E. W., Chan, N. D., Cohen, A. P., & Pogodzinski, M. (2014). *Public Bikesharing in North America During a Period of Rapid Expansion: Understanding Business Models, Industry Trends and User Impacts*. 234.

SmartCitiesWorld. (2017). *The benefits of smart bike-sharing*. Smart Cities World. https://www.smartcitiesworld.net/news/the-benefits-of-smart-bike-sharing-2420

Sohrabi, S., Paleti, R., Balan, L., & Cetin, M. (2020). Real-time prediction of public bike sharing system demand using generalized extreme value count model. *Transportation Research Part A: Policy and Practice*, *133*, 325–336. https://doi.org/10.1016/j.tra.2020.02.001

Staples, H. (2017, November 17). *Holland-Cycling.com—Do share bikes have a future? Part 3*. Holland-Cycling. https://www.holland-cycling.com/blog/288-do-share-bikes-have-a-future-part-3

Statistiek, C. B. voor de. (2017). *CBS Urban Data Center/Den Haag* [Webpagina]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/onze-diensten/cbs-urban-data-centers/locaties/cbs-urban-data-center-den-haag

Świątek, J., & Tomczak, J. M. (Eds.). (2017). *Advances in Systems Science* (Vol. 539). Springer International Publishing. https://doi.org/10.1007/978-3-319-48944-5

Tang, Y., Pan, P. H., & Shen, P. Q. (2011). *Bike-Renting Systems in Beijing, Shanghai and Hangzhou and Their Impact on Travel Behavior*. 20.

The Hague. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=The_Hague&oldid=1048037264

*The Hague by the sea | DenHaag.com*. (2021). https://denhaag.com/en/see-and-do/the-hague-by-the-sea

*The Hague Historical Weather*. (2021). WorldWeatherOnline.Com. https://www.worldweatheronline.com/the-hague-weather/south-holland/nl.aspx

*The Hague in numbers*. (2021). https://www.denhaag.nl/en/municipality-of-the-hague/the-hague-in-numbers.htm

*The Meddin Bike-sharing World Map*. (2021). https://bikesharingworldmap.com/

Tin Tin, S., Woodward, A., Robinson, E., & Ameratunga, S. (2012). Temporal, seasonal and weather effects on cycle volume: An ecological study. *Environmental Health*, *11*(1), 12. https://doi.org/10.1186/1476-069X-11-12

van Oort, N., & Drost, M. (2015). *Data-driven public transport ridership prediction approach including comfort aspects*. 13.

Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. *Procedia - Social and Behavioral Sciences*, *20*, 514–523. https://doi.org/10.1016/j.sbspro.2011.08.058

Wang, S., Chen, H., Cao, J., Zhang, J., & Yu, P. S. (2020). Locally Balanced Inductive Matrix Completion for Demand-Supply Inference in Stationless Bike-Sharing Systems. *IEEE Transactions on Knowledge and Data Engineering*, *32*(12), 2374–2388. https://doi.org/10.1109/TKDE.2019.2922636

Wang, X., Lindsey, G., Schoner, J. E., & Harrison, A. (2016). Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and from Stations. *Journal of Urban Planning and Development*, *142*(1), 04015001. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000273

Waterstaat, M. van I. en. (2019, October 3). *Walking facts* [Publicatie]. Ministerie van Infrastructuur en Waterstaat. https://doi.org/10/3/walking-facts

Wijken en buurten in Den Haag. (2021). In *Wikipedia*. https://nl.wikipedia.org/w/index.php?title=Wijken_en_buurten_in_Den_Haag&oldid=60056431

Wu, X., Lyu, C., Wang, Z., & Liu, Z. (2019). Station-Level Hourly Bike Demand Prediction for Dynamic Repositioning in Bike Sharing Systems. In X. Qu, L. Zhen, R. J. Howlett, & L. C. Jain (Eds.), *Smart Transportation Systems 2019* (pp. 19–27). Springer Singapore.

Xiao, G., Wang, R., Zhang, C., & Ni, A. (2020). Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-020-08803-y

Xin, F., Chen, Y., Wang, X., & Chen, X. (2018). Cyclist Satisfaction Evaluation Model for Free-Floating Bike-Sharing System: A Case Study of Shanghai. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(31), 21–32. https://doi.org/10.1177/0361198118770193

Xing, Y., Wang, K., & Lu, J. J. (2020). Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai, China. *Journal of Transport Geography*, *87*, 102787. https://doi.org/10.1016/j.jtrangeo.2020.102787

Xu, C., Ji, J., & Liu, P. (2018). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies*, *95*, 47–60. https://doi.org/10.1016/j.trc.2018.07.013

Xu, M., Liu, H., & Yang, H. (2020). A Deep Learning Based Multi-Block Hybrid Model for Bike-Sharing Supply-Demand Prediction. *IEEE Access*, *8*, 85826–85838. https://doi.org/10.1109/ACCESS.2020.2987934

Yang, H., Zhang, X., Zhong, L., Li, S., Zhang, X., & Hu, J. (2019). Short-term demand forecasting for bike sharing system based on machine learning. *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, 1295–1300. https://doi.org/10.1109/ICTIS.2019.8883732

Yang, Z., Chen, J., Hu, J., Shu, Y., & Cheng, P. (2019). Mobility Modeling and Data-Driven Closed-Loop Prediction in Bike-Sharing Systems. *IEEE Transactions on Intelligent Transportation Systems*, *20*(12), 4488–4499. https://doi.org/10.1109/TITS.2018.2886456

Zhao, J., Deng, W., & Song, Y. (2014). Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transport Policy*, *35*, 253–264. https://doi.org/10.1016/j.tranpol.2014.06.008

Zhou, Q., Wong, C. J., & Su, X. (2020). Machine Learning Approach to Quantity Management for Long-Term Sustainable Development of Dockless Public Bike: Case of Shenzhen in China. *Journal of Advanced Transportation*, *2020*, 1–13. https://doi.org/10.1155/2020/8847752