Research in Transportation Economics 101 (2023) 101340

Contents lists available at ScienceDirect

Research in Transportation Economics

journal homepage: www.elsevier.com/locate/retrec

Research paper

Improving the service of E-bike sharing by demand pattern analysis: A data-driven approach

Ziru Zhang [a,*], Panchamy Krishnakumari [a], Frederik Schulte [b], Niels van Oort [c]

[a] P.O. Box 5049, 2600 GA, Delft, the Netherlands
[b] P.O. Box 5031, 2600 GA, Delft, the Netherlands
[c] P.O. Box 2600, 2600 GA, Delft, the Netherlands

ARTICLE INFO

JEL classification:
R41
R48
R49

Keywords:
E-Bike sharing system
Temporal clustering
Demand pattern analysis
Operational strategy
Service level evaluation

ABSTRACT

In recent years, there has been a surge in the popularity of free-floating e-bike sharing. However, the shared mobility sector is fiercely competitive demanding, efficient operations and high-quality service to cater to user expectations.

We propose several data-driven methods that apply demand pattern analysis. We suggest the use of a new spatial unit (i.e., overlapping circles) to enhance the cost-efficiency and user-friendliness of e-bike sharing. Moreover, temporal clustering is employed to develop operational strategies that counter the imbalance in supply and demand in recurrent clusters.

To evaluate the impact of these strategies, we introduce a framework and apply it in a case study of an e-bike sharing project in The Hague, The Netherlands. We identify 5 hourly clusters which enable reallocation strategies to alleviate the imbalance among spatial units in these clusters.

The results demonstrate that the derived operational strategies improve the service significantly, with almost 1.5 times increased ridership, an approximately 20% decrease in vehicle idle time, and a decent monthly net retention rate of around 60%.

1. Introduction

Shared mobility has become a major trend since 2010, aiming to improve the sustainability of the transport sector and alleviate traffic congestion (European Commission, 2011). However, the shared mobility market is highly competitive, requiring providers to achieve efficient operations and high service quality (Beirigo, Negenborn, Alonso-Mora, & Schulte, 2022). Among various shared mobility options, shared bikes have gained widespread popularity due to their active mode of transport with the associated health benefits (Barbour, Zhang, & Mannering, 2019; DeMaio, 2009). With the introduction of electrification in the mobility sector, e-bikes, which offer higher travel speeds and reduce physical efforts, have been gradually incorporated into bike-sharing schemes (Fishman & Cherry, 2016).

Bike-sharing projects can be classified into two operational types, viz., station-based and free-floating schemes (see e.g. Ma, Ji, et al. (2020)). The station-based scheme relies on pre-defined stations for users to pick up and return bikes, while the free-floating offers flexibility, allowing users to drop bikes at various locations within designated operational zones (B. Beirigo, Schulte, & Negenborn, 2018). The latter eliminates the constraints associated with station availability in station-based systems, contributing to the growing popularity of free-floating bike sharing in recent years during the past years (Chen, van Lierop, & Ettema, 2020; Fishman, 2016).

Regardless of the operational type, understanding user travel behaviour is crucial for matching supply and demand in bike-sharing systems (Hua, 2020; A. Li, Zhao, Huang, Gao, & Axhausen, 2020). Extensive research has been conducted on different aspects of bike-sharing systems, including determinants of bike-sharing demand, the interaction with public transport (Montes, Geržinic, Veeneman, van Oort, & Hoogendoorn, 2023; van Marsbergen, Ton, Nijënstein, Annema, & van Oort, 2022; van Mil, Leferink, Annema, & van Oort, 2021), demand pattern analyses, prediction of demand in different time scopes, and optimization of reallocation of shared bikes (Albuquerque, Sales Dias, & Bacao, 2021; Eren & Uz, 2020; Fishman, 2016; Fishman, Washington, & Haworth, 2013; Galatoulas, Genikomsakis, & Ioakimidis,

* Corresponding author.
E-mail addresses: ziruzhang1101@gmail.com (Z. Zhang), N.vanOort@tudelft.nl (P. Krishnakumari), F.Schulte@tudelft.nl (F. Schulte), P.K.Krishnakumari@tudelft.nl (N. van Oort).

https://doi.org/10.1016/j.retrec.2023.101340
Received 30 November 2022; Received in revised form 23 July 2023; Accepted 29 July 2023
Available online 26 August 2023

2020; Ma, Yuan, Van Oort, & Hoogendoorn, 2020). These studies can be categorized into three phases: 1) identifying determinants of shared bike usage; 2) analysing datasets, identifying demand patterns and predicting future demand; 3) devising optimal strategies to reallocate bikes.

However, most of the research has focused on station-based bike-sharing systems and limited attention has been paid to free-floating e-bike sharing. Substantial differences exist between station-based shared bikes and free-floating shared bikes, as well as between regular bikes and e-bikes (Chen et al., 2020; Galatoulas et al., 2020; Gu, Kim, & Currie, 2019). Free-floating bike-sharing systems offer users a higher degree of freedom by eliminating the need to rent and return bikes at designated stations. However, this flexibility increases the complexity of modelling, as demand cannot be attributed to specific station units as in station-based systems. Therefore, spatial analytical units, such as virtual stations or traffic area zones, need to be defined for free-floating bike--sharing to model trip generation and attractions (S. Liu, Hou, Liu, Khadka, & Liu, 2018). Additionally, e-bikes have distinct trip characteristics, such as travel distance, which vary from regular bikes due to reduced physical effort and the presence of batteries, which exert an influence on people's travel decisions (Galatoulas et al., 2020).

Moreover, the existing studies on bike-sharing operations lack experiments and evaluations of different operational strategies in real-life contexts. Current approaches often rely on dedicated but complicated mathematical models to determine the optimal strategies with either static or dynamic demand input. However, these methods can be quite time-consuming and unrealistic for small and medium-sized shared mobility operators, considering the limited resources and uncertainties of operational actions (Alvarez-Valdes et al., 2016, p.; Angelopoulos, Gavalas, Konstantopoulos, Kypriadis, & Pantziou, 2018; Chemla, Meunier, & Wolfler Calvo, 2013; Dell'Amico, Hadjicostantinou, Iori, & Novellani, 2014; Gavalas, 2016; Raviv, Tzur, & Forma, 2013).

Based on the existing literature, there are 3 scientific gaps: 1) a spatial analytical unit which is friendly for both the operators and the users, especially its efficiency for operators; 2) studies targeting free-floating e-bike sharing projects; 3) a well-rounded evaluation approach of the real-life effects of the proposed operational strategies.

Considering both the scientific gaps and the needs of operators, the objective of this work is to develop a data-driven approach to derive beneficial operational strategies. Those should then be deployed to improve the service of e-bike sharing by conducting a data analysis, a demand pattern analysis and a follow-up examination of the proposed strategies in reality.

To this end, we 1) introduce an innovative spatial analytical unit, with the overlapping circles, and prove that the reallocation strategies derived based on this unit, are more cost-effective, requiring only one relocation operation per period; 2) we add insights to the field of e-bike sharing, taking a different angle than the current studies; 3) we develop a framework to evaluate the operational strategies and experiments in real-life settings, considering both operators and users. Results indicate that the proposed cost-effective operational services contribute to a positive effect of increasing ridership by roughly 1.5 times.

The remainder of this paper is structured as follows: Section 2 describes the methodology, which is then applied to a case study presented in Section 3. The results and corresponding discussion are provided in Section 4, along with a comparison to parallel work. Finally, Section 5 presents the conclusions, including the main findings, contributions, limitations, and recommendations for future research.

Through this study, we introduce an innovative spatial analytical unit to investigate the demand pattern of e-bike sharing services and propose cost-effective operational services contributing to a positive effect of increasing roughly 1.5 times ridership.

2. Methodology

In this work, we propose a four-step approach, as depicted in Fig. 1 to address the research objective. Firstly, the literature review is conducted to determine the contributing factors to the bike-sharing demand. The preliminary data analysis is sequentially done. These results are input to the demand pattern analysis, which investigates the pattern of bike-sharing services in depth. The operational strategies are derived based on these insights accordingly. Finally, the strategies are implemented in a real-life setting and evaluated systematically.

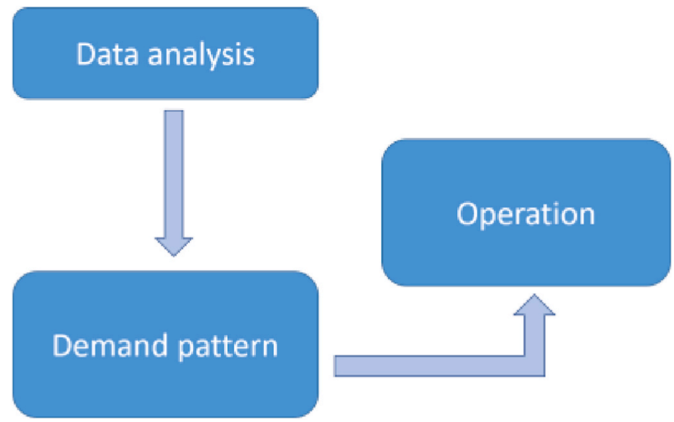

Fig. 1. The four-step framework of this study.

2.1. Data analysis

The data analysis is the prerequisite of the demand pattern analysis. This phase involves three steps, which are data description, correlation analyses between determinants and the demand, and land use pattern analysis.

2.1.1. Data description

The data description phase encompasses the acquisition of available data, data cleaning, and data processing/aggregation. The primary dataset used in this research consists of ride records, as exemplified in Table 1. Each record contains ride start/end time, ride start/end location, trip duration, and trip distance with a unique ride ID, and rider ID. Additionally, data pertaining to six other groups are included. These data are obtained from open-source databases, such as Google Maps for spatial and infrastructure factors, by points of interest, meteorological institutions for weather-related data and the government for socio-demographic factors; trip characteristics and temporal data can be retrieved from ride records and safety factors are usually omitted in the most studies due to their reliance on deliberative interviews. Following data obtainment, a data cleaning is employed to eliminate incomplete or faulty data records and align the data with the research scope. Moreover, the statistical characteristics of those data are described.

2.1.2. Correlation analysis

Following the data description, linear correlations are conducted between the determinants and the demand (i.e., the ridership). Only linear correlations are examined in this study to test the hypothesis of whether the determinants found in the literature indeed impact the demand considering the research scope. The methods used in this study are Pearson's coefficient and multiple linear regression. They are chosen

Table 1
Ride record sample.

Ride id	Rider id	Start time	End time
2356	3556	16:00:02 July 24, 2023	16:26:24 July 24, 2023
Start location 52.0579935, 4.2638107	End location 52.055401, 4.268105	Trip duration 00:26:22	Trip distance 7.839 km

because of their simplicity and the power of revealing the correlations.

Pearson's coefficient is computed for each variable, and it indicates the normalized covariance between two variables as shown in Equation (1).

$$r_{xQ} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Q_i - \overline{Q})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2}} \quad \text{Equation 1}$$

Where n is the sample size; x_i and Q_i are the individual sample points, indexed with i; $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the sample mean) and analogously for $\overline{Q}$; in this paper, 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 is used when there is more than one variable under a determinant group to see what the essential factor under the determinant group is.

2.1.3. Land use pattern analysis

In addition to data description and correlation analysis, a land use pattern analysis is conducted to gain insights into the functional composition of different locations based on the distribution of POIs. This analysis is performed using two selected spatial analytical units: neighbourhood and 400 m overlapping circles.

First, the distributions of POIs are visualized to reveal the general pattern of different facilities. Then, the function p (i,c) is defined in Equation (2), representing the POI distribution. If p exceeds 0.5 then the unit would be defined with the corresponding function. Only workplaces and recreations (including sustenance and entertainment amenities) are considered in this study due to their most relevancy to traffic demand.

$$p(i,c) = \frac{|i_c|}{\sum_{i=1}^{n}|i_c|}, \forall i \in I, c \in C \quad \text{Equation 2}$$

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.

The outputs would be considered in the following demand pattern analysis, to understand the prediction of flow.

2.2. Demand pattern analysis

The demand pattern analysis serves as the focal point of this research. First, a descriptive analysis of crucial trip characteristics is conducted. Second, the demand pattern is performed, incorporating temporal clustering methods. The insights from demand pattern analysis support the development of reallocation strategies as they help mitigate the imbalance between the supply and the demand in different units. Third, supply efficiency is examined by the distribution of vehicle idle time per unit. Similarly, the average trip time and distance are explored on the unit level. These analyses, in conjunction with the supply efficiency assessment, provide valuable insights into the operational efficiency and popularity of different spatial units. This information facilitates prioritizing operational strategies and making necessary adjustments to the service area.

A. Spatiotemporal aggregation

After the general demand pattern analysis, the spatial analytical units are determined, specifically the neighbourhood and 400 m overlapping circles. Data are aggregated and analysed on these two distinct levels.

The temporal clustering is then conducted, based on similarities and dissimilarities in demand of different periods. The purpose of temporal clustering is to investigate if the demand pattern of different periods emerges and the insights are studied in the next step, aiding the development of operational strategies (T. L. K. Liu, Krishnakumari, & Cats, 2019).

The complete procedures are as follows: firstly, ride records are aggregated in the spatial units determined in the last step and then OD (origin-destination) matrix is computed accordingly; it is followed by the temporal clustering based on OD matrices, gathering different periods with similar features together. They are used to capture essential demand peculiarity.

OD matrix presents the flow between different locations with insights on how trips are attracted and generated at zonal levels.

First of all, the ride records are aggregated with the predefined spatial units as the origin and destinations, and the flow $q(x, y, t, z)$ then corresponds to a specific origin x, a destination y, a ride date z, 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-h intervals 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 z, as indicated in Equation (3).

$$Q_t(c_o, c_d, t) = \sum_{x \in c_o}\sum_{y \in c_d}\sum_{z} q(x, y, t, z) \quad \text{Equation 3}$$

$$Q_\tau(c_o, c_d, z) = \sum_{x \in c_o}\sum_{y \in c_d}\sum_{t} q(x, y, t, z) \quad \text{Equation 4}$$

Similarly, daily OD matrices are on a daily basis, generating totally D 24*k*k OD matrices where each cell corresponds to the flow increment of a 1-h interval between a given OD pair (c_o, c_d) for a given date z; 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, shown in Equation 4.

Thirdly, temporal clustering is employed with the aggregated OD matrices as the feature vectors. Each data point is composed of a corresponding OD matrix. Agglomerative hierarchical clustering is applied because of its loose prerequisite of the number of clusters, and its dendrogram to assist in the determination of the optimal number of clusters (Rokach & Maimon, 2005).

In this research, Euclidean distance is used to compute the dissimilarity metric and the ward method is applied to combining the clusters by the variance of clusters which is found to be the most suitable method for quantitative variables (Calinski & Harabasz, 1974).

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 number of clusters can then be decided based on the dendrogram considering the interpretability.

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

2.3. Supply efficiency analysis

The supply efficiency analysis is performed by examining the vehicle idle time per spatial unit. Vehicle idle time is 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, Krishnakumari, Arbez, Chiabaut, & van Lint, 2020).

Commonly, the vehicle idle time only refers to the time interval between two rides while sometimes there are fewer than 2 ride records. The assignment of the location corresponding to the idle time is tricky.

Thereby, the vehicle idle time is adapted to tackle this problem.

To keep consistency and include all possible idle time records, vehicle idle time is separated into two types: one corresponding to the origins of rides as $VIT^v(c_o, \tau)$; the analogous for the destinations as $VIT^v(c_d, \tau)$ as Equation (5) and Equation (6).

$$VIT^v(c_o,\tau)=\begin{cases} t, \left|R_v^\tau\right|=0 \\ ts_r^v-ts_\tau, \left|R_v^\tau\right|=1 \\ \begin{cases} ts_r^v-ts_\tau, \textit{if r is the first ride in } R_v^\tau \\ ts_{r+1}^v-te_r^v, \forall r\in R_v^\tau \end{cases}, \left|R_v^\tau\right|>1 \end{cases} \quad \text{Equation 5}$$

$$VIT^v(c_d,\tau)=\begin{cases} t, \left|R_v^\tau\right|=0 \\ te_\tau-te_r^v, \left|R_v^\tau\right|=1 \\ \begin{cases} ts_{r+1}^v-te_r^v, \forall r\in R_v^\tau \\ te_\tau-te_r^v, \textit{if r is the last ride in } R_v^\tau \end{cases}, \left|R_v^\tau\right|>1 \end{cases} \quad \text{Equation 6}$$

$\forall v \in V$ 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 and 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^τ 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 period, the vehicle idle time is separated into two sub-vehicle idle times, corresponding to the original unit and the destination unit separately.

For example, if vehicle v only has 1 ride record during time period τ, with the attributes ride start time ts_v, the original 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 would be 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_τ 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_τ.

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 h. 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 assist the reallocation, avoiding low usage in general. Thereby, this time interval should also be compatible with the frequency of reallocation.

Unit-based vehicle idle time is calculated based on the formulas. To reallocation geographically, this vehicle idle time per spatial unit per period is computed as Equation (7) and Equation (8).

$$AVIT_{c_o}(T_t)=\frac{\sum_{i\in I_{c_o}^{T_t}} VIT^v(c_o,\tau)\forall v\in V,\tau\in T_t}{\left|I_{c_o}^{T_t}\right|}, \forall c_o\in C, T_t\in T \quad \text{Equation 7}$$

$$AVIT_{c_d}(T_t)=\frac{\sum_{i\in I_{c_d}^{T_t}} VIT^v(c_d,\tau)\forall v\in V,\tau\in T_t}{\left|I_{c_d}^{T_t}\right|}, \forall c_d\in C, T_t\in T \quad \text{Equation 8}$$

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, $\left|I_{c_o}^{T_t}\right|$ and $\left|I_{c_d}^{T_t}\right|$, is the number of all vehicles idle time belonging to unit c_o/c_d during the time period T_t.

Based on the magnitude of $AVIT_{co}(T_t)$ and $AVIT_{cd}(T_t)$ in different units, a heatmap would be visualized, describing which unit(s) vehicles encounter shorter vehicle idle time and thereby those locations are appealing to reallocate bikes.

2.4. Average trip time/distance analyses

Similarly, average travel distance and duration are computed on the spatial unit level. These two metrics are origin-oriented since the destination is less relevant from the operator's perspective. The calculations are conducted as Equation (9) and Equation (10).

$$ATD_c(T)=\frac{\sum_{i\in R_c^T} \textit{travel distance}_i}{\left|R_c^T\right|}, \forall c\in C \quad \text{Equation 9}$$

$$ATT_c(T)=\frac{\sum_{i\in R_c^T} \textit{travel time}_i}{\left|R_c^T\right|}, \forall c\in C \quad \text{Equation 10}$$

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 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.

2.5. Development of the operational strategies

In this study, two primary types of operational strategies are identified: reallocation strategies and adjustment in the service area.

General demand patterns and temporal clusters provide insights into the departures, arrivals and total flow of each spatial unit, as well as how the e-bike sharing traffic flows between different OD pairs. These analyses serve the purpose of informing the development of reallocation strategies, which aim to mitigate the imbalance between the supply and the demand.

Furthermore, the assessment of supply efficiency and average trip distance/duration analyses is performed at the spatial unit level, indicating the popularity and operational efficiency geographically. Based on these findings, recommendations in the operational areas can be formulated.

2.6. Operation

During the operational stage, the strategies derived from the previous phase are implemented in the real-life context and systematically examined.

To evaluate these strategies, two sets of KPIs are introduced. The first group focuses on the operational aspect, while the second category evaluates user satisfaction from a business perspective. The first group includes metrics such as daily ridership, ridership ratio and average vehicle idle time per vehicle; the second is presented by net retention ratio and average user expenditure.

➢ Daily ridership

This is represented 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 would be assigned to the rides belonging to 10/09/21.

$$Q_\tau = \sum_x \sum_y \sum_t q(x, y, t, \tau), \forall \tau \in T \qquad \text{Equation 11}$$

where t is the ride start time. It sums of all the rides in the given day τ, if 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.

$$r_\tau = \frac{Q_\tau}{supply_\tau} \qquad \text{Equation 12}$$

➢ Average vehicle idle time

The vehicle idle time is computed in the same way. 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, in a different aggregation way, though. The object of this metric corresponds to each vehicle, and the average idle is computed based on all corresponding vehicle idle time records. Afterwards, the average vehicle idle time is computed by taking the average of all average idle time corresponding to the available fleet during this period. The origin-based and the destination-based indicators apply in a similar way.

$$AVIT_v^{c_o}(T_t) = \frac{\sum_{i \in I_v^{c_o}} VIT_i^v(c_o, \tau) \forall c_o \in C, \tau \in T_t}{|I_v^{c_o}|}, \forall v \in V^{T_t}, T_t \in T \qquad \text{Equation 13}$$

$$AVIT_v^{c_d}(T_t) = \frac{\sum_{i \in I_v^{c_d}} VIT_i^v(c_d, \tau) \forall c_d \in C, \tau \in T_t}{|I_v^{c_d}|}, \forall v \in V^{T_t}, T_t \in T \qquad \text{Equation 14}$$

where $I_v^{c_o}$ is the set of origin-based vehicle idle records belonging to the vehicle v, and analogously applies for 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 period T_t, the denominator is the fleet size belonging to this period within the operation zone.

$$AVITV_{c_o}(T_t) = \frac{\sum_{v \in V^\tau} AVIT_v^{c_o}(T_t)}{|V^{T_t}|}, \forall T_t \in T \qquad \text{Equation 15}$$

$$AVITV_{c_d}(T_t) = \frac{\sum_{v \in V^\tau} AVIT_v^{c_d}(T_t)}{|V^{T_t}|}, \forall T_t \in T \qquad \text{Equation 16}$$

➢ 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 grows or churns monthly (Guide to Net Dollar Retention (NDR) - Definition, Calculation, Tips, 2021). It indicates how much customers spend and their expenditure changes on the service over time and is a way to understand customers' satisfaction: if they are satisfied with the service, they would keep a subscription and continue spending money on it.

$$NRR = \frac{Starting\ MRR + Expansion\ MRR - Contraction\ MRR - Churn\ MRR}{Starting\ MRR} \qquad \text{Equation 17}$$

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

➢ Average user expenditure

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

$$average\ user\ expenditure_m = \frac{total\ expenditure_m}{|users_m|}, \forall m \in I \qquad \text{Equation 18}$$

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

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

3. Case study

The case study is conducted based on an e-bike sharing project in The Hague. The project was launched in mid-June 2021 and the research focuses on a 4-month duration following the project's launch. The data from the first three months are utilized to develop operational strategies while the entire period is considered for evaluating the effects of these strategies. The e-bike sharing project operates with a fleet size of a few hundred e-bikes, with dynamic adjustments made to the fleet size based on operational conditions in the subsequent days.

The Hague, the third-largest city in The Netherlands, serves as the context for this case study. It has a population of approximately 550,000 inhabitants as of 2021, consists of 8 administrative districts and 44 neighbourhoods, referred to as Wijken in Dutch (The Hague, 2021; The Hague in Numbers, 2021; Wijken en buurten in Den Haag, 2021). In this research, the neighbourhood units and the overlapping circles are treated as the spatial analytical units, as the visualization in Fig. 2.

Neighbourhoods are a convenient unit for research purposes, but their substantial heterogeneities can pose challenges to operational efficiency. Also, for the users, the operations on the neighbourhood level are still too coarse and thereby cannot capture their needs in specific locations. To address this limitation, an innovative unit, overlapping circles, is introduced. The radius of overlapping circles is set to 400 m: on one hand, 400 m approximately correspond to 5-min walking based on the Dutch average walking speed of 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

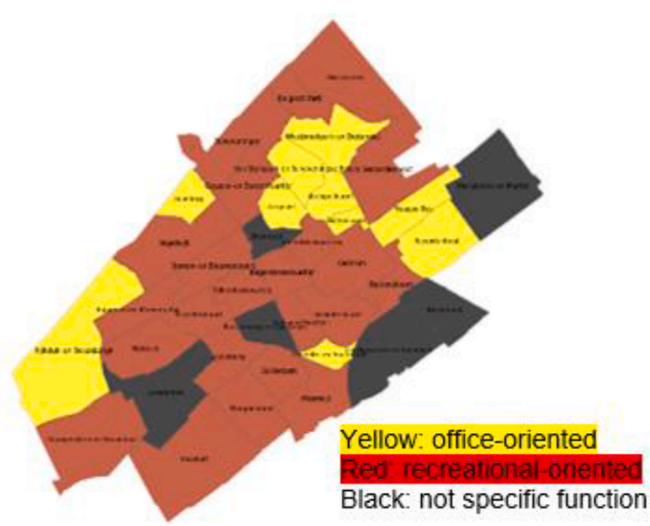

Fig. 2. Map of functions on the neighbourhood level.

the other hand, a sensitivity analysis was conducted to ensure the robustness of 400 m in terms of revealing demand patterns in hourly clustering.

The main function of neighbourhoods is illustrated in Fig. 4. It is evident that recreational areas are predominantly concentrated in the central and southwestern part of The Hague, particularly in Centrum and along the beach area. For workplace function, it plots relatively sparsely. There are quite some office-oriented areas along the beach and in the north-central areas.

Turning to the descriptive statistics of the demand for the first 1.5 months this timeframe is chosen as it represents the only available data during that specific research phase. To account for the unstable ridership caused by the effects of the project's launch, the ridership is normalized. The highest ridership is 1 and the rest is the ratio between the ridership and the highest value. It is observed for the first 3 days, the ridership was exceptionally high because the service was free of charge during this period. Afterwards, the ridership continually declined, with several fluctuations though. The ID verification, aiming to avoid misbehaved rides, was installed on the 19th day after launch. It has exerted a negative effect on ridership with a decline to circa 0.15 per day. This also arises from the bad weather condition. Followingly, reallocation strategies obtained from the data-driven methods have been implemented, which would be discussed in detail in the later sections, and they contribute to the rebounce of ridership to 0.3 per day approximately, as presented in Fig. 3.

For ridership in terms of the day of the week, there is no obvious difference between the ridership per day, indicating that the ride day does not insert a noticeable effect on ridership. Additionally, the demand does not either present the typical morning peak pattern, seen in Fig. 4, contrary to the previous work (S. Li et al., 2021; Miranda-Moreno & Nosal, 2011; Tin Tin, Woodward, Robinson, & Ameratunga, 2012; Xing, Wang, & Lu, 2020). One reason is that people are still unfamiliar with this service which prevents them from using it during the morning peak, to avoid being late for work. Similar to the literature, there is an evening peak between 16:00 to 19:59, which reaches its climate at around 17:00 to 17:59. It implies the service attracts people after their work, consistent with the assumption under the fact of no morning peak in this case study.

The histogram of ride distance shown in Fig. 5, illustrates that the majority of trips are shorter than 10 km and the peak is between 0 and 2 km. The inspection of trip duration indicates that most rides are shorter than 50 min and the peak is between 0 and 10 min. The average travel duration is 18.2 min and the median is 11.5 min, next to the peak range.

4. Results and discussion

This section presents the results of the application of the methods described in Section 2. First, the general determinants for e-bike sharing services are introduced with the specification of the determinants in this case study. Second, the results from the demand pattern are presented, followed by a summary of the derived operational strategies. Last, the operational strategies are appraised using the proposed methods.

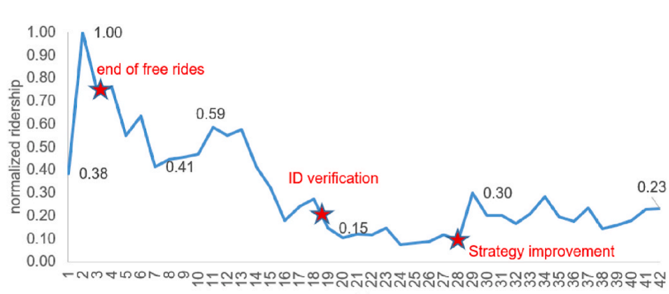

Fig. 3. Ridership overview.

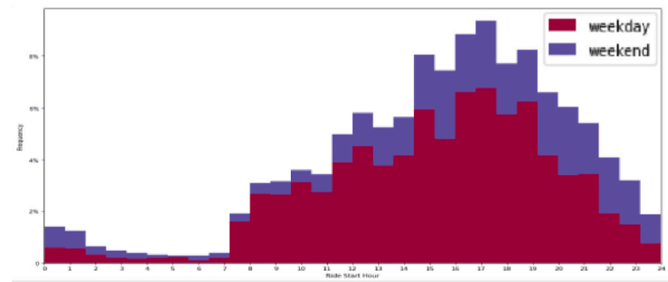

Fig. 4. Temporal distribution of the demand by ride start hour of a day.

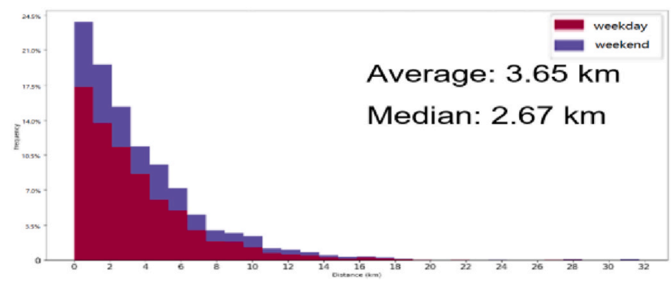

Fig. 5. Distribution of the distance.

4.1. *Determinants for the demand for e-bike sharing services and correlation analysis*

Based on the literature review, there are 6 types of influential factors for e-bike sharing service, viz., spatial and infrastructure factors, weather-related factors, mobility and trip characteristics, temporal factors, sociodemographic factors and safety factors (Daddio, 2012; Fishman, Washington, & Haworth, 2012, 2013; Hampshire, 2012; Ji, Cherry, Han, & Jordan, 2014). However, only the first four groups are applied in this study due to the availability of data.

Correlation analyses are then conducted to see whether these general determinants do exert effects as expected in this case study. Most of these determinants are proven to have a Pearson's coefficient higher than 0.3 with the demand. Unexpectedly, the precipitation level and temperature are found to correlate with the demand at a low level, −0.03 and 0.16 respectively, stemming from their subtle variations of them under the time scope of this research. Another study also presents a similar result where precipitation is found to be insignificant to affect the demand level (He et al., 2019).

In harmony with the previous studies, the number of POIs has positive effects on the ridership level (Faghih-Imani, Hampshire, Marla, & Eluru, 2017); the humidity, proved from both Pearson's coefficient and the multiple linear regression model, however, impacts the demand negatively (El-Assi & Mahmoud, 2015).

4.2. *Demand pattern analysis and temporal clustering*

There are two demand pattern analyses, with the dataset of two periods. First, general demand pattern analysis is conducted on the neighbourhood level, using the rides of the first month; then, temporal clustering is done on both the neighbourhood and circles level, using the dataset of the first 1.5 months.

As seen in Fig. 6, it is found 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 catches attention, standing out as the heat spots. In addition to that, the arrivals outweigh departures in most neighbourhoods, despite two central ones.

Hence, it is recommended to rebalance the bikes from the beach areas to Centrum and the south of Centrum. Another operational

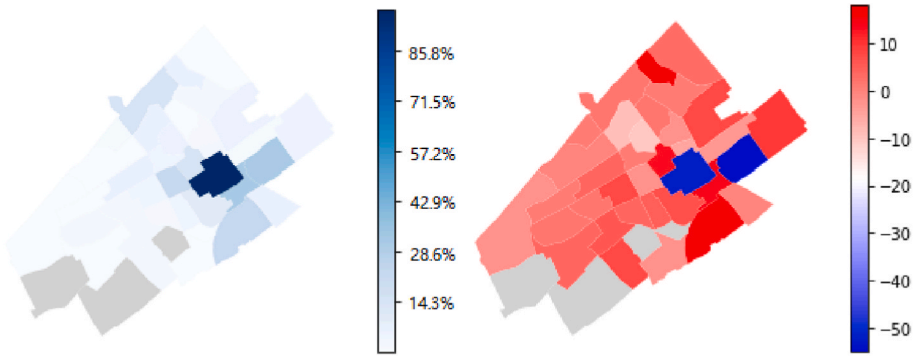

Fig. 6. Heatmaps of normalized departures (left) and the difference between arrivals and departures per neighbourhood (right).

strategy is to place bikes in batches when rebalancing the bikes, making the bikes more noticeable.

Afterwards, temporal clustering is conducted to see if different periods share similar demand patterns. Both spatial units are used in this analysis Only the clustering on the overlapping circles is present, as Fig. 7.

As observed from Fig. 8, the peak hours and transition hours distinguish themselves to be separate clusters, and the rest hours emerge together, given 5 clusters. Thereby, there are 5 hourly clusters with their characteristics, shown in Fig. 9 as follows.

i. The first peak hour (16:00–16:59): for this period, the rides gather in Centrum and the beachside area. Besides these two areas, the rides are relatively sparse in the outer units at a low magnitude. The most prevailing spots are in the centre which belongs to Centrum and are recreational-oriented areas, located in the centre of Fig. 9.1. It is noteworthy that the other units of attention are generally recreational-oriented zones despite the units located to the east of the centrum with office functions. 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 varies per unit.

ii. The second peak hour (17:00–17:59): 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 near the train station and this unit also reveal a slight departure tendency. The alike phenomenon appears in another unit close to the central railway station, with a higher departure ratio. It suggests that travellers prefer to pick up the bikes near the train station during this period, which is conceivable to serve the last mile of their trips, compensating for 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.

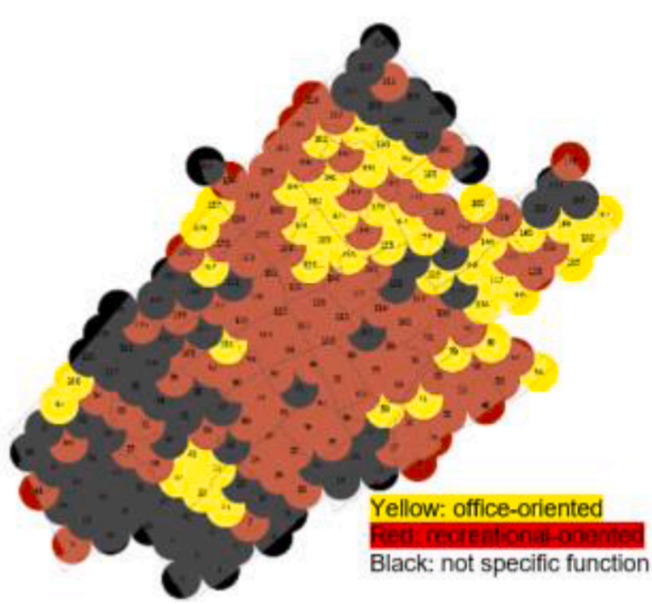

Fig. 7. Map of functions on the circle level.

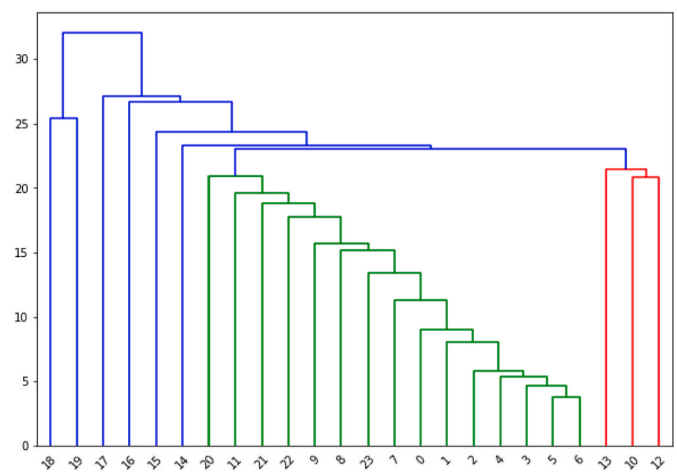

Fig. 8. Dendrogram of hourly clustering on the circle level.

iii. The first transition hour (18:00–18:59): Office-oriented units present a predilection for 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.

iv. The second transition hour (19:00–19:59): The unit near the train station is still the most popular in this period, with almost equal departures and arrivals, echoing the deduction that these trips serve the first/last mile supplementary train trips. Rides are relatively distributed to other units while the central areas are still prevalent all the time. In this period, the central units are generally dominated by more departures while the arrivals are towards the outer units.

1. The off-peak (20:00–15:59): invariably, the central areas are still quite essential, followed by units next to the south and east of the Centrum district. Besides, the outer residential units are either balanced with similar departure and arrival ratios, or prone to more arrivals in this period.

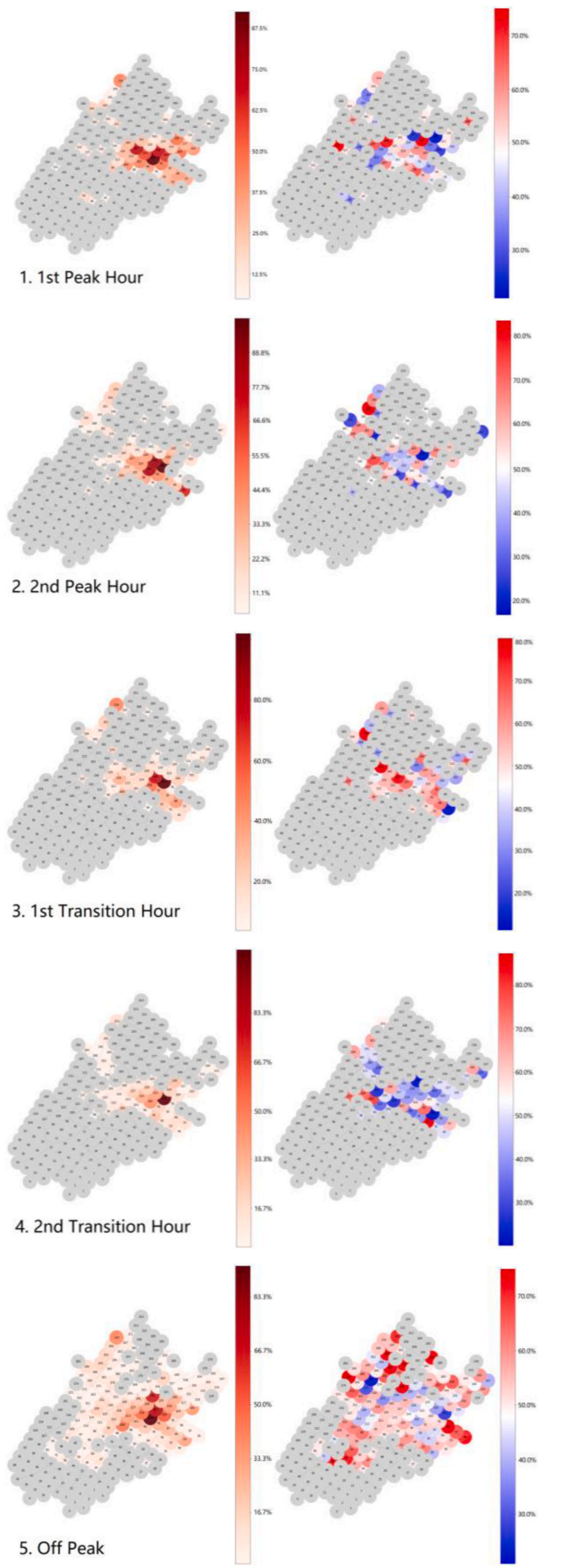

Fig. 9. Heatmaps of flow (proportional to the maximal flow) and arrival ratio for 5 hourly clusters.

Correspondingly, the rebalance suggestions are constructed in Table 2. The repositioning strategies on the neighbourhood levels are also derived in the same way while the details are omitted in this paper. The comparisons between these two sets of strategies would be presented in the evaluation section.

4.3. Supply efficiency analysis and average trip duration/distance analysis

This series of analyses provide an overview of the usage over different spatial units.

It is found from the supply efficiency analysis that the outer units usually experience a longer vehicle idle time, especially those located in the southwest, as seen in Fig. 10.

Average travel distance and duration indicate that the southwest part witnesses a lack of use as presented in Fig. 11, aligned with the observations in the supply efficiency analysis. 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.

Thereby, it is suggested to adjust the service area, leaving out the southeast part as well as placing those bikes in the hotspots of this service. By doing so, efforts of battery swaps can also be significantly alleviated.

4.4. Evaluation of the operational strategies

There are in total 5 operational strategies with various time scopes. 3 out of 5 are reposition strategies, consecutive in the timeline of two months from July to September, in which the first set is based on the general demand pattern on the neighbourhood level, and the second and the third are developed from temporal clustering on the neighbourhood and overlapping circle levels. The details of temporal clustering are omitted in this study while the derived strategies are still being assessed, compared with their counterparts on the overlapping circles.

The exact action of temporal-based reallocations is dependent on the execution time of the action, which is usually between 10:00 to 16:00 during the daytime. It is noted that only one relocation operation is needed per period, requiring only a few efforts by the operator.

The KPIs from the operational sides are presented in Table 3. It is obvious that the period of the implementation of ID verification performs the worst out of all the periods. During the periods of rebalancing, all KPIs are improved to different degrees. Among the three rebalancing periods, the repositioning 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 in the improvement of the service level.

Though the ridership ratio seems too low, under 1 ride per vehicle per day, it is acceptable compared to the other schemes with similar supply levels around the world. For instance, Alacant, Boulder, Clermont-Ferrand and Perpignan (marked as yellow dots in Fig. 12), have a similar supply level. The first two schemes present alike ridership ratios and the latter two have a lower ridership ratio at around 0.4 and 0.2 respectively (Médard de Chardon, Caruso, & Thomas, 2017). Considering Mobike in Delft, which offered a much higher supply, only saw the ridership ratio at around 1.5 in 2018 (Ma, Yuan, et al., 2020). All

Table 2
Rebalance suggestions based on temporal clustering in circles.

	period	From	To
1st **peak**	**16:00–16:59**	Centrum	Station
2nd **peak**	**17:00–17:59**	Centrum	South of Centrum
1st **transition**	**18:00–18:59**	South of Centrum	South
2nd **transition**	**19:00–19:59**		
Off-peak	**20:00–15:59**	beach	Pier of beach

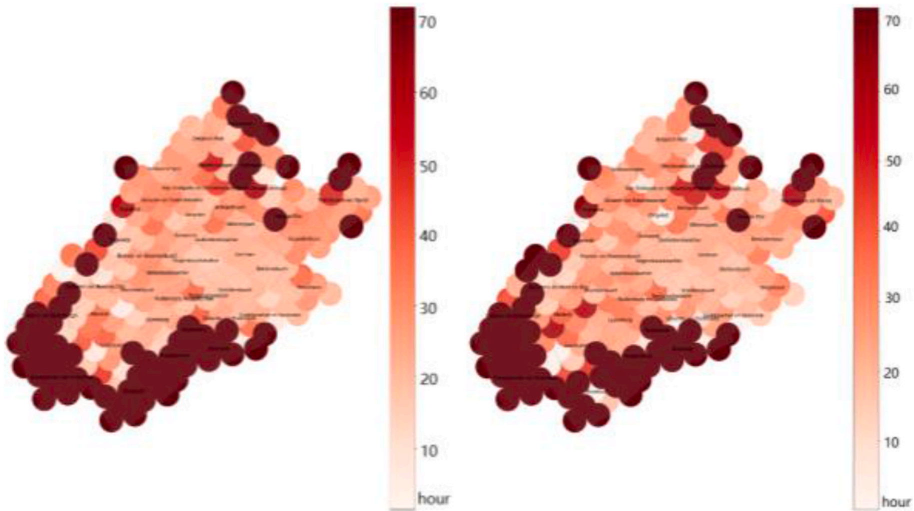

Fig. 10. Heatmaps of origin-based (left) and destination-based (right) vehicle idle time.

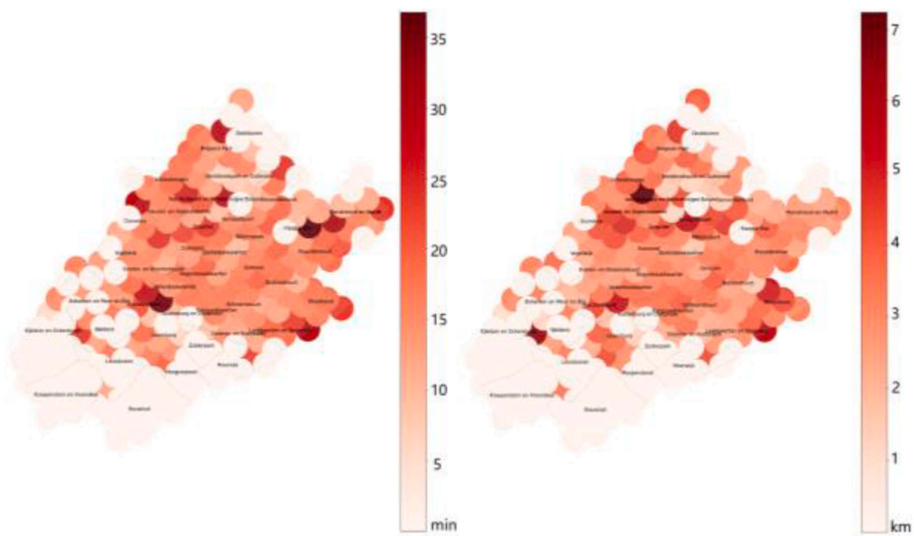

Fig. 11. Heatmaps of average trip time (right) and distance (left).

Table 3
Summary of the operational KPIs.

periods	date	average ridership ratio	average vehicle idle time per vehicle (h)	
			origin-based	destination-based
free rides	Day 1–3	2.11	8.81	17.82
adopting period	Day 4–18	0.98	47.28	32.14
ID verification	Day 19–28	0.31	59.51	52.66
first-round rebalance	Day 29–43	0.53	51.84	43.25
second-round rebalance on the neighbourhood level	Day 44–50	0.59	57.3	41.08
second-round rebalance on the circle level	Day 51–82	0.67	47.3	33.07
without specific strategy	Day 83–97	0.64	50.16	41.08
reduction in the operational area	Day 98–122	0.52	50.9	43.63
overview		0.66	40.77	33.05

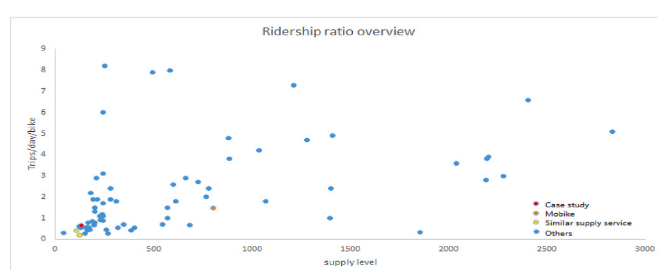

Fig. 12. Ridership ratio overview.

these cases show the relevance of the results and the efficacy of operational strategies in this work.

From the users' perspective, the figures related to the monthly net retention rate are illustrated in Table 4. The third month from Mid-August to Mid-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 new users, as seen in Table 5.

Table 4
Summary of NRR.

Period	NRR
Mid-June to Mid-July	
Mid-July to Mid-August	15.78%
Mid-August to Mid-September	86.87%
Mid-September to Mid-October	52.10%

Table 5
Summary of average user expenditure.

Period	New user average spent (€)	Retained user average spent (€)	total user average spent (€)
Mid-June to Mid-July			
Mid-July to Mid-August	7.83	15.78	9.24
Mid-August to Mid-September	12.80	20.87	15.30
Mid-September to Mid-October	9.91	19.24	14.45

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

From 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 partially overlapped with the second month 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 customer loyalty is not fully established. It is common for a new service. Additionally, this level is quite acceptable compared with the other bike-sharing providers in which the retention rate is also below 100%, ranging from 20% to 70%, even the famous project, such as Lime, claims a national retention rate of 60% (Shaheen, Martin, Chan, Cohen, & Pogodzinski, 2014; SmartCitiesWorld, 2017).

The operational strategies are proven to be beneficial for improving the service, observed by an increase in NRRs and average user expenditure, with substantial fluctuations though.

The fourth month experiences a decline in NRR, while the average expenditure of new users and all users do not witness a considerable decrease, even with the shrinkage 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 although it requires more effort in operation. However, a reduction in the service area, 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.

5. Conclusion

This work addresses several scientific gaps in free-floating e-bike sharing research by introducing an innovative spatial analytical unit, the overlapping circles, and demonstrating that the reallocation strategies derived based on this unit, are more cost-effective because only one relocation operation per period is needed. Moreover, this work adds new insights to the scarcely investigated, proposes a framework to evaluate the operational strategies in a real-life context, while considering the perspectives of both operators and users.

We introduce data-driven methods such as demand pattern analysis to design new operational strategies. In addition to that, we conduct a descriptive analysis, revealing a single peak period of the e-bike sharing service in The Hague, contrary to the widely observed two peaks in other studies. Building on this initial analysis, we perform general demand pattern analyses as well as temporal clustering analyses of the demand pattern via agglomerative hierarchical clustering. The central units are found to be the most popular places, in both general and clustering analyses. Furthermore, 5 periods emerged from the hourly clustering, which is the first peak hour, the second peak hour, the first transition hour, the second transition hour, and the off-peak period. We observe 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. Additionally, the analysis indicates a lack of use in the outskirts in the southwest of The Hague, based on the supply efficiency and average travel distance/duration analyses. Based on these findings, three sets of rebalancing strategies and the reduction in the service area were proposed.

Subsequently, these strategies were implemented in a real-life context and their effects are examined using two categories of KPIs that assesses the perspectives of operators and users. All strategies demonstrate improvements in service levels. Among these, the third set of repositions based on hourly clustering on the circle level showed to improve the service to the largest degree, out of all strategies, with a ridership ratio of 0.67 and a decrease in the origin-based and destination-based vehicle idle time at around 12 and 19 h, compared to the ID verification period. Besides, the adjustment in the operational zone has a moderate beneficial impact on the service, but it decreases the operational efforts to a substantial 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 is also found that users became more satisfied with the service after the implementation of the suggested strategies, indicated by an increased net retention rate ranging from 52.10% to 86.87% and a grown average user expenditure of around 9 to 15 euros per user. Therefore, we advise to implement multiple operational strategies in combination to complement each other.

However, there are still some limitations. The explicit supply data is unavailable thereby the supply level was inferred from the ride records. 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 repeats the research with better input data, including the precise supply data, and incorporate the relationship of public transport into analyses, on the overlapping circle units. Future work should furthermore investigate predictions of short-term e-bike sharing demand, adapt the evaluation method of the operational strategies which isolates the effects stemming from the strategies, by providing better reference cases for the predictive models.

CRediT authorship contribution statement

Ziru Zhang: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Panchamy Krishnakumari:** Supervision, Writing – review & editing. **Frederik Schulte:** Supervision, Writing – review & editing. **Niels van Oort:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

During this research, the main author has done an internship at a company operating e-bike sharing in Europe, who provides the main data input.

Data availability

The data that has been used is confidential.

Acknowledge

This research uses data provided by an anonymous shared-mobility operator. The authors would like to thank Jam and Mickey Mouse NL for their help with data acquisition.

References

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

Alvarez-Valdes, R., Belenguer, J. M., Benavent, E., Bermudez, J. D., Muñoz, F., Vercher, E., et al. (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

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

Beirigo, B. A., Negenborn, R. R., Alonso-Mora, J., & Schulte, F. (2022). A business class for autonomous mobility-on-demand: Modeling service quality contracts in dynamic ridesharing systems. *Transportation Research Part C: Emerging Technologies, 136*, Article 103520. https://doi.org/10.1016/j.trc.2021.103520

Beirigo, B., Schulte, F., & Negenborn, R. (2018). *Dual-mode vehicle routing in mixed autonomous and non-autonomous zone networks. 2018 21st international Conference on intelligent transportation systems (ITSC), 1325–1330*. https://doi.org/10.1109/ITSC.2018.8569344

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

Cats, O., Krishnakumari, P., Arbez, N., Chiabaut, N., & van Lint, H. (2020). Empirical evaluation of the efficiency and effectiveness of a ride-sourcing service. In *99th transportation research board annual meeting, Washington DC*.

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, 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

Daddio, D. (2012). *Maximizing bicycle sharing: An empirical analysis of capital bikeshare usage* (Vol. 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

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* (Vol. 39).

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

European Commission. (2011). European commission. Transport White Paper (p. 31) 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* (Vol. 23, p. 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

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* (Vol. 29).

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

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

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

Hampshire, R. C. (2012). *An analysis of bike sharing usage: Explaining trip generation 2 and attraction from observed demand* (Vol. 17).

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

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

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. In *2019 6th international conference on models and technologies for intelligent transportation systems (MT-ITS)* (pp. 1–9). https://doi.org/10.1109/MTITS.2019.8883312

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*, Article 102828. https://doi.org/10.1016/j.jtrangeo.2020.102828

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*, Article 102974. https://doi.org/10.1016/j.jtrangeo.2021.102974

Ma, X., Ji, Y., Yuan, Y., Van Oort, N., Jin, Y., & Hoogendoorn, S. (2020). A comparison in travel patterns and determinants of user demand between docked and dockless bike-sharing systems using multi-sourced data. *Transportation Research Part A: Policy and Practice, 139*, 148–173. https://doi.org/10.1016/j.tra.2020.06.022

van Marsbergen, A., Ton, D., Nijënstein, S., Annema, J. A., & van Oort, N. (2022). Exploring the role of bicycle sharing programs in relation to urban transit. *Case Studies on Transport Policy, 10*(1), 529–538. https://doi.org/10.1016/j.cstp.2022.01.013

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*, Article 120846. https://doi.org/10.1016/j.jclepro.2020.120846

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

van Mil, J. F. P., Leferink, T. S., Annema, J. A., & van Oort, N. (2021). Insights into factors affecting the combined bicycle-transit mode. *Public Transport, 13*(3), 649–673. https://doi.org/10.1007/s12469-020-00240-2

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

Montes, A., Geržinic, N., Veeneman, W., van Oort, N., & Hoogendoorn, S. (2023). Shared micromobility and public transport integration - a mode choice study using stated preference data. *Research in Transportation Economics, 99*, Article 101302. https://doi.org/10.1016/j.retrec.2023.101302

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

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.

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* (Vol. 234).

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

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

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

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.

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*, Article 102787. https://doi.org/10.1016/j.jtrangeo.2020.102787