

Article

Addressing Vehicle Sharing through Behavioral Analysis: A Solution to User Clustering Using Recency-Frequency-Monetary and Vehicle Relocation Based on Neighborhood Splits

Nicolo' Brandizzi ¹, Samuele Russo ², Gaspare Galati ² and Christian Napoli ^{1,*}

¹ Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto 25, 00185 Roma, Italy

² Department of Psychology, Sapienza University of Rome, Via dei Marsi 78, 00185 Roma, Italy

* Correspondence: cnapoli@diag.uniroma1.it

Abstract: In many developed cities around the world, vehicle sharing is becoming an increasingly popular form of green transportation. While such services are associated with lower emissions and easier mobility, their management poses a significant challenge. In this paper, we examine a dataset collected in Barcelona during the months of august and september 2020 in order to investigate relocation strategies and user clustering. By proposing a neighborhood area split and relating it to user demand, we propose two different areas based on majority demand and users' requests and provide interpretations of both. We then aim to identify groups of similar users using a variant of Recency Frequency Monetary/Duration (RFM or RFD) clustering that extends to GPS coordinates of voyages in order to differentiate scores based on economic and geographical factors; furthermore, a user-based clustering approach was used to maximize client preferences. As a result of our analysis, the sharing company may be able to make more informed decisions regarding where to focus its resources. In fact, we find that the majority of the demand is concentrated in an area that represents 7.47 percent of the city's area. Additionally, we propose a discount-based approach in order to influence the user's behavior in parking the vehicle where it is most needed.

Keywords: green mobility; vehicle relocation; user clustering



Citation: Brandizzi, N.; Russo, S.; Galati, G.; Napoli, C. Addressing Vehicle Sharing through Behavioral Analysis: A Solution to User Clustering Using Recency-Frequency-Monetary and Vehicle Relocation Based on Neighborhood Splits. *Information* **2022**, *13*, 511. <https://doi.org/10.3390/info13110511>

Academic Editors: Miguel Félix Mata Rivera and Roberto Zagal Flores

Received: 3 October 2022

Accepted: 20 October 2022

Published: 25 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Despite pandemic conditions and mobility restrictions, car-sharing company formed registered several million users in 2021 in Europe, a 30 percent increase over the previous year.

However, traffic congestion in 2021 was worse than in previous years, resulting in 140 h lost in traffic [1]: these effects are not only detrimental to the environment, but also to people's mental and physical well-being.

The sharing of vehicles (bikes, scooters, and cars) can contribute to the solution and has been an active field of research with numerous publications [2]. Firstly, it reduces the number of vehicles owned in total, thereby reducing the demand for raw materials. Secondly, it can reduce total emissions by investing in low-polluting vehicles such as electric engines. As an example, in [3] the authors state that each vehicle of *Share Now* can replace up to twenty vehicles, however, they admit that the study is not without controversy: the study is limited to only one company's customers, residents of only eleven cities in Europe. As a final point, it has been shown to increase the net benefit to society in a similar, if not greater, manner than public transportation schemes [4].

The operation of a vehicle-sharing company presents a number of challenges, one of which is the relocation issue, which entails finding a disposition of the vehicles so that, from the provider's perspective, profits are maximized while avoiding losses to the customer.

In order to implement new marketing strategies aimed at capturing the attention of their clients, companies must analyze customer data in order to identify relationships between different elements and then develop personalized marketing campaigns based on the findings. Based on geographic, demographic, psychographic, and behavioral variables, these strategies identify customer segments [5] that can be used to analyze customers' behaviors. Moreover, in light of the wide variety of consumer behaviors, tastes, and needs, marketing strategies must be customized.

By clustering, users can be divided into groups in a manner that makes them more similar to each other. Clustering consists of grouping objects in such a way that they are more similar to each other in the group. Each of these groups is called a cluster, and it is defined as a region in which the density of objects is higher than in other regions.

The most frequently used clustering algorithm based on the minimization of an objective function is *K-means clustering*; in the presence of n data points in a d -dimensional space, R^d , and an integer K , K -means determines a set of K points-centers based on the mean squared distance to the nearest center [6] so that the mean squared distance between each data point and center is minimized. There are many applications of clustering, including pattern recognition, image processing, machine learning, and statistics [6].

Customers may be clustered according to the parameters suggested by the *RFM* (Recency-Frequency-Monetary) model or by its variant *RFD* (Recency-Frequency-Duration). *RFM/RFD* are popular techniques used for customer segmentation according to their purchase behaviors. Using these two models, customers are ranked on the basis of three factors: the period since the last purchase (Recency), the number of purchases made within a specific period (Frequency), and the amount (Monetary) or time (Duration) spent on a product [7]. Each group is represented with a three-digit *RFM/RFD* combination that summarizes a purchase pattern typical within members' transaction history [5].

In this paper, our contribution is two-fold:

- In the first step, we address the relocation problem by analyzing the provided dataset and providing neighborhood divisions according to vehicle location throughout the day.
- Secondly, we implement a clustering technique based on recency-frequency-monetary data and identify groups of customers with similar characteristics.

Paper Structure

In the following paper, we present a brief overview of the state of the art, Section 2, in the vehicle relocation problem, Section 5.1, and user clustering, Section 5.2, together with a brief theoretical background, Section 3. Then, in Section 4, we analyze the vehicle sharing dataset collected over the city of Barcelona. We provide time, money and location information which we later use in order to implement our solution. In Section 5, we provide a custom clustering and relocation algorithm taking into account the previous data analysis. Finally, we provide our conclusions and future works in Section 6.

2. Related Work

The first studies in vehicle sharing systems perform analysis in a simulated environment. Barth et al. [8] focus on average customer wait time and the number of relocations required to keep this time low. The system analyzed is station-based, i.e., the vehicle can be picked up and returned only in predefined spaces.

Ferrero et al. [2] reviewed the most important papers in this area and gives a comprehensive definition of different aspects of vehicle-sharing services, in particular the taxonomy of their operating modes. The type of systems are:

- *Two-Way System (Station-Based)*: in this mode, the vehicles are in predefined parking lots (the stations), where the user can pick them up, and then return to the same space. From a service point of view, this situation has no particular critical issue, given that the station will always have the same number of vehicles. A total of 19% of analyzed papers focus on this mode.

- *One-Way System (Station-Based)*: this mode is similar to the previous one, with the difference that the destination parking lots can be different from the parking lots where the vehicle has been picked up. This case is more complex than the previous one due to the different amount of vehicles in between stations. This is the most studied mode with 50% of papers.
- *Free-Floating*: this is the newer mode, the company that owns the vehicles defines an operating area in which they can be freely parked, and the trip can have any starting and finishing point in this area. This is the most complex one in terms of relocation, and we will focus on this. A total of 19% of papers studied this mode.

2.1. Relocation Problem

As previously stated, one of the main challenges for vehicle-sharing systems is the relocation problem, i.e., placing the vehicle in such a way that the service provider's profits and customer satisfaction are maximized. This is extremely challenging due to the high inherent unpredictability of the incoming users' requests, more precisely the two questions that one has to answer are: where and when the next request will come from? The two perspectives to be considered are those of the vehicle manager and the customers.

First of all, in [9], where the authors reviewed the main issue concerning electric vehicles and their relocation, they distinguished between strategic and tactical problems (e.g., the location of stations), which have long-term objective and high-cost, and operational problems, which include within-day optimal relocation of vehicles and battery exchange to restore vehicle autonomy. Our focus will be on the latter.

There are essentially two ways to relocate vehicles. The first one is called operator-based, and the burden is on the provider which will have to execute the relocation. Weikl et al. [10] propose two possibilities for the system provider: on one hand the vehicles can be moved by maintenance personnel or additional staff, and this can happen for each vehicle singularly or more than one at the same time, but the latter is easy with bikes or scooters, more complicated with cars. These rides can be combined with maintenance operations, such as gas filling or battery charging. On the other hand, the provider could arrange buffer depots with a specific number of vehicles at high-demand locations, such that during the period with a lot of requests nothing has to be performed. The main issue for both strategies is the additional cost, and for the former also the ecological impact of additional trips. For example in [11], the authors developed a Mixed Integer Linear Programming formulation of the Electric Vehicle Relocation Problem and an operator-based method to relocate the vehicle, which tries to forecast the unbalancing of electric vehicles in the system. Such a method is successfully applied to the data gathered in Milan, and despite the additional trip necessary for the staff to optimally relocate the fleet vehicles, it showed only a small impact on urban traffic. Gambella et al. [12] analyze a station-based one-way electric vehicle sharing system, from a provider perspective, proposing an operator-based exact relocation Mixed Integer Programming model for operating hours, also considering vehicles battery consumption and recharge. They also propose two heuristics: the reduction in relocation density with respect to time distance (e.g., every 30 min instead of every 15 min) and the gradual inclusion of relocation arcs over time.

User-Based Relocation

The second method to relocate vehicles is called user-based, and, as the name suggests, the task shifts from provider to user. More precisely the system manager has to find the right incentives to persuade the users to change some characteristics of the trip. What are these characteristics? Todd et al. [13] deployed a user-based relocation algorithm at the University of California Riverside, which proposed a discount trip price to customers that were willing to share the ride, i.e., *trip joining*, or, in the presence of two or more customers, that were willing to split for the ride, *trip splitting*. They found that the user-based relocation mechanism is highly dependent on user participation: they assumed 100% of participation, which showed an overall 42% decrease in the number of relocations. The

majority of the literature focuses on incentives to convince the user to change the origin or the destination. Stokkink et al. [14] proposed a user-based relocation strategy based on the current distribution of vehicles and expected future demands, also at the single-user level. The customers are stimulated to relocate the vehicles from over-saturated locations to under-saturated locations, through discounted prices. The algorithm is applied in a station-based one-way car-sharing system, with data collected in Grenoble. Notice that although technically feasible, the prediction at the single-user level can give rise to some privacy concerns.

Similarly, in [15] we can see a two-step relocation algorithm, where the first step determines the optimal drop-off locations by only considering the future demand satisfaction, and the second one determines the discounts that minimize the request rejection ratio and maximizes the operator profit. The scenario is a one-way car-sharing system, the results show that the system is effective and it can also be applied in areas not characterized by asymmetric mobility demand, increasing the profit significantly. Instead, Clemente et al. [16] used timed Petri net to provide to the user, in real time, indications based on actual vehicle allocation and if the customer follows the given directions, he has a right to a discount on the rental. The difference with other approaches is that customers starting or arrival times are also taken into account and can be modified. In [17], the authors used a model-based decision support system, in particular a closed-loop control scheme, which is based on Particle Swarm Optimization algorithm, with the aim of incentives the users to drop off the vehicles in precise locations. They show that the incentive mechanism is very effective if the fleet size is coherent with the demand, and the benefits decrease if the demand increases too much.

There are not many studies that cover both free-floating scenarios and user-based relocation. An interesting analysis of this particular case is present in [18], where the case study is Cologne, which is divided into two areas (central and peripheral) and the trips were classified into four types (the combination of the previous two areas). Then, they find out that the incentive should be offered to relocate the vehicles from the peripheral to the central zone. Free-floating case and user-based relocation is also the subject in [19]. The authors use integer programming and a graph representation that reformulate the problem as a K-disjoint shortest path. They propose an alternative origin or destination, or start or arrival time, to the customer, which, in exchange, will receive a discount and they also manage to solve the problem in polynomial time.

There are also studies that use a mixed approach, such as [20] that propose a mesoscopic approach, i.e., there is a macroscopic-level operator-based algorithm, that acts at the zone level, and a microscopic-level user-based algorithm, which acts at the vehicle level.

2.2. User Behavior

At this point, it is clear that if one wants to apply a user-based relocation algorithm has to have a deep knowledge of users' behavior and preferences. Barth et al. [21] are among the first to study user behavior in this domain, through surveys and operation analysis, and most importantly they estimated the distance that users are willing to walk to reach the vehicle, that is 400 m. In [22], the authors provide a business perspective on Vehicle-Sharing Systems. Through 34 interviews with car-sharing providers, they tried to determine what the success factors are for this type of business, e.g., the influence of environmental context, and the characteristics of the typical customer. For our interest, the most interesting factors are the ones about users, in particular, that customers expect cars to be within 300–500 m, and for the free-floating scenario, some reserved parking spaces in dense areas, information on parking, and free-of-charge search for parking-spaces are expected. Focusing mainly on distance, [23] try to determine how the distance to an available vehicle influences the users' behavior, and if there exists a threshold distance such that the customers are satisfied if a vehicle is available within this distance, and are unsatisfied otherwise. The results, based on a free-floating car-sharing system in Munich, show that the distance of the nearest available car is the most important decision criterion for users and the probability of booking a car

linearly depends on the distance to the nearest available car, instead a threshold distance has not been detected.

2.3. Operating Area Clustering

A necessary step in the successful application of the algorithm is the clustering of the operating area. For example [10] propose a two-step model for relocation, and the first step consists of an offline demand clustering for prediction of the optimal future state of spatially available vehicles (to be performed occasionally, e.g., once a year). Practically they found the first two principal components and then they combined them into groups by applying K-means clustering. Instead of doing a relocation once a year, the authors of [24] provide a clustering method for the operating area that works online, also changes during the operating time and performs only the necessary and effective repositioning tasks. They propose as a performance indicator of their system the Zero-Vehicle-Time (ZVT), which means that a certain zone is without vehicles and then the requests at that time in that zone cannot be satisfied. Obviously, both service providers and customers want to avoid this situation, hence their strategy is to globally minimize the ZVT. An alternative approach can be seen in [25], where the authors proposed a methodology that exploits a potential function, expression of user's interest concerning different areas of the city, using in particular Gaussian Mixture Models and Theory of mind [26], to split a metropolitan area into sectors, that will be used in relocation. For simplicity, they took into account only the multivariate 2D Gaussian that describes the vehicle distribution and also the position of points of interest and points of disinterest. First, they estimate Gaussian mean and covariance based on the position of each vehicle at the end of the day i (the procedure is then repeated for n days), then, using the probability density function of the estimated above, they divide the area into sectors.

2.4. User Clustering

Customer diversification is a key task to accomplish for the above-mentioned reasons. With Collaborative Filtering, a reasonable strategy is to find similar users to a given one and create clusters of clients to suggest new items. Most of the researchers focused on how to apply a Collaborative Filtering (CF) method to datasets made by users that have expressed preferences on the items they have tested through the assignment of ratings. The main goal of the state-of-the-art research is to find a way to optimize the accuracy of their models: Haifeng Liu et al. [27] tried to combine the local context for common ratings of each pair of users and the global preference of each user's ratings. Differently, Nadia Fadhil AL-Bakri et al. [28] applied a common recommended system to a movie-reviews dataset. Moreover, a hybrid approach between User-Based CF and Item-Based CF has been explored by SongJie Gong [29] in which users have been clustered on users' ratings and then an item clustering has been applied to produce the recommendations.

All these approaches and many more [30–33] have all the peculiarity of clustering people based only on the ratings they have assigned to items. In this paper, ratings are not available for the vehicles users have used, so the state-of-the-art approach of CF cannot be applied. Thus, the common clustering based on ratings has been replaced by K-means based on the Recency-Frequency-Monetary features extracted by analyzing the dataset and expanded by also including the duration of the analysis. To do so, the techniques described by Yen-Liang Chen et al. [7], where RFM sequential patterns from customers' purchasing data have been applied to electronic commerce, and İ. Kabasakal [5], where RFM is presented along with a case study of an e-retailer from Turkey have been applied. The application of a common RFM/RFD classification is, however, limited because it does not consider other information that may be helpful, such as the places customers have frequented. In our work, users' locations are fundamental features to take into consideration for segmenting, especially for economic purposes. For this reason, a good approach is suggested by Schmidt et al. [34] that implemented a Spectral Clustering to find Interest-Based Communities of Twitter users according to their preferences. This approach

of segmentation may be applied to vehicles-users according to the places they are used to traveling more. Finally, a User-Based CF technique has been considered to recommend vehicles to the company's users according to the maximization of Pearson's correlation metric, as suggested by Leily Sheugh [35].

3. Theoretical Background

In the following section, we provide the necessary theoretical background to better understand the subsequent sections. Section 3.1 introduces the two main models for segmentation, RFM and RFD, and outlines their characteristics in behavior-based clustering. Our next Section 3.2 defines two main clustering algorithms, K-means and spectral clustering, which we will use in our experiments.

3.1. RFM and RFD Models for Segmentation

According to the introduction, RFM and RFD are methods used to identify the characteristics of customers and then segment them into groups based on some common characteristics. When segmenting data, data analysts generally consider different types of patterns such as a customer's occupation, gender, age, status, and geographical distribution, as well as psychographic characteristics such as social class, personality, and behavior. By analyzing previous basket transactions, the RFM and RFD use these behavior-based approaches to group customers into clusters. In addition, they can be used to identify customer groups whose members are similar in terms of their purchase decisions. In both RFM and RFD, the focus is on how recently, how often, and how much customers purchase, with the only difference being that RFD focuses on time rather than money.

An RFM and RFD represent purchase behavior based on a combination of three dimensions, represented by the three values of Recency, Frequency, and Monetary or Duration:

- Recency: how many days passed since last purchase?
- Frequency: what is the total number of purchases?
- Monetary: what is the total amount of money spent?
- Duration: what is the total time spent?

During RFM/RFD, the first step is to score and rank customers based on these three attributes by dividing their distribution into quartiles.

In each dimension, the highest 25 percent is ranked as 1 (best), and the lowest 25 percent is ranked as 4 (worst). It is important to note that high values in Recency indicate an undesirable condition. Therefore, the lowest 25% is ranked as 1 and the highest 25% is ranked as 4 [5]. Following the above quartile division, these scores are juxtaposed in order to create a segment bin value. As a final step, customer RFM/RFD scores are arranged in ascending or descending order; the top clients are those with the lowest recency and highest frequency-monetary/duration amounts, which means they have used the vehicles recently and frequently.

By assigning scores and values to each individual user, the algorithm clusters them into four distinct groups according to their RFM/RFD values. These four groups are named *Platinum*, *Gold*, *Silver*, *Bronze* and represent the groups of customers from the top ones (Platinum) to the worst ones (Bronze). This clustering is based on *K-means*.

3.2. Clustering Techniques

Generally, clustering technologies are associated with unsupervised problems in the literature, and a wide variety of approaches can be found in this area. This paper introduces the most common clustering approach, *K-means clustering*, as a baseline for the later, spatial-based *spectral clustering*.

3.2.1. K-Means

The K-means clustering method is an unsupervised classification method for identifying patterns. It works by defining some unlabeled objects as vectors in a multidimensional

space. The vectors are grouped into K different clusters, each with common characteristics with the other vectors in the same cluster.

In order to assign clusters, a similarity measure must be defined. In general, K-means partitions objects of interest into K clusters by minimizing the similarity between the target object and the cluster centers; the similarity metric is typically *Euclidean distance*:

$$\operatorname{argmin}_{c_i \in C} d(c_i, x) \tag{1}$$

where $d(c_i, x)$ is the distance between the centroid of the selected cluster C and one point x in this cluster. More specifically, given two points x_1 and x_2 , their Euclidean distance is:

$$d(x_1, x_2) = |x_1 - x_2|. \tag{2}$$

The K-means algorithm can be summarized into four steps [28], as shown by the pseudocode in Algorithm 1.

Algorithm 1 K-means algorithm

- 1: *Starting centroids* \leftarrow Choose randomly K items as initial centers
 - 2: **repeat**
 - 3: *assign to centroids* \leftarrow according to a metric, i.e., minimization of euclidean distance
 - 4: *mean computation* \leftarrow The mean is computed for all the items within one cluster. The new center is the average of each cluster.
 - 5: *Calculation of new centroids* \leftarrow The difference between the newly calculated center and the previous center in the same cluster is computed
 - 6: **until** centroids do not change (difference between old and new is zero) or the maximum number of repetitions has reached
-

3.2.2. Spectral Clustering

In graph theory, spectral clustering is a technique for identifying groups of nodes based on the edges connecting them within a graph. This method relies on eigendecomposition of the matrix associated with the users (represented by the nodes of the graph) and their relationships (given by the edges) [34]. The matrix is called *Adjacency matrix* $W = (w_{ij})_{i,j=1,\dots,n}$, where w_{ij} represents the connections between the vertices v_i and v_j and it is equal to zero when no edges are present to connect them. Moreover, it is possible to define the degree of a vertex $v_i \in V$ as:

$$d_i = \sum_{j=1}^n w_{ij} \tag{3}$$

representing how many edges are connected to a given node. From this result, it is possible to define the *Degree matrix* D as the diagonal matrix having the degrees d_1, \dots, d_n on its diagonal.

Clusters are usually extracted using a standard clustering algorithm, such as K-means, on the relevant eigenvectors extracted from a *Laplacian matrix*. The most common type of Laplacian matrix is *unnormalized graph Laplacian*, which can be calculated by simply subtracting the adjacency matrix from the degree matrix:

$$L := D - A \tag{4}$$

This matrix has important properties [36], such as:

1. For every vector $f \in \mathbb{R}$ we have:

$$f'Lf = \frac{1}{2} \sum_{i,j=1}^n w_{ij}(f_i - f_j)^2 \tag{5}$$

2. L is symmetric and positive semi-definite.
3. The smallest eigenvalue of L is 0, the corresponding eigenvector is the constant one vector 1.
4. L has n non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

The last step is to compute the clusters for n based on the vectors [34] using a clustering algorithm.

4. Dataset

Our dataset consists of 13,289 different trips for 4246 customers over a period of five months. Two sources are provided: one containing requests for trips from users, and the other containing periodic status updates sent by vehicles (a mix of scooters and bikes), both in *csv* format. Among the features of the first one are: *start*, *end*, *SiteId*, *VehicleId*, *startLatitude*, *startLongitude*, *startAltitude*, *finishLatitude*, *finishLongitude*, *finishAltitude*, *DateTime*.

According to the coordinates, the data was collected in Barcelona, but subsequent data processing and analysis can be applied to other cities as well.

As a first step, we arbitrarily defined the operating area (Latitude: 41.300–41.500, Longitude: 2.050–2.225), and dropped all entries outside of the operating area. Using Python Client for Google Maps Services, we were able to add two additional columns to the data set: *startNeighborhood* and *finishNeighborhood*. Using the Google API, we gathered additional information: given the latitude and longitude coordinates, we chose to save the neighborhood of the starting and finishing points.

Our analysis of the distribution of requests in Barcelona was based on the 73 neighborhoods that exist in the city. Accordingly, we retrieve from the Barcelona open data repository ([bcn-geodata](#) (accessed on 1 May 2022) a map of the city with subdivision into neighborhoods in *geojson* format that can be processed using *geopandas* library.

We chose to analyze:

- Time distribution throughout the day, grouped hourly.
- Time distribution throughout the weekdays, grouped daily.
- Spatial distribution on plain map, in particular, scatter plot and heat map.
- Spatial distribution on map with neighborhood division, also presenting the most common starting/finishing neighborhood combination.
- Both previous spatial distribution throughout the hour of the day.

In the following sections, we present relevant results, obtained from data collected between August 2020 and September 2020. In this case, we are dealing with a tourist city during a period of high touristic demand (data were collected in August, the month with the highest touristic demand in 2019 according to ([37], p. 40)). The data collected has a strong bias due to the pandemic effect on tourist demand, see -91.5% for overnights in hotels.

4.1. Time Distribution

First, we analyze vehicle usage according to the time of day and day of the week. Figure 1 illustrates the hourly distribution of requests throughout the day, with a spike between 16:00 and 18:00, with over 1500 requests each. We speculate that the vehicles are used mostly by locals at the end of their working shift since there is no clear peak associated with morning working hours. Taking into account the spatial distribution in the following section, we believe that the vehicles are being used by customers to hang out near the city center, where most of the attractions are located.

For a personalized experience, it is essential to take into account this information. We also report the relationship between the time spent on the vehicles and the amount spent. Two trends can be seen in Figure 2: the vertical trend on the leftmost side shows that a large number of users did not spend any money on the trip (usually, free trips are included with app installation); furthermore, the diagonal line indicates that the duration of a journey is linearly correlated with the amount of money spent.

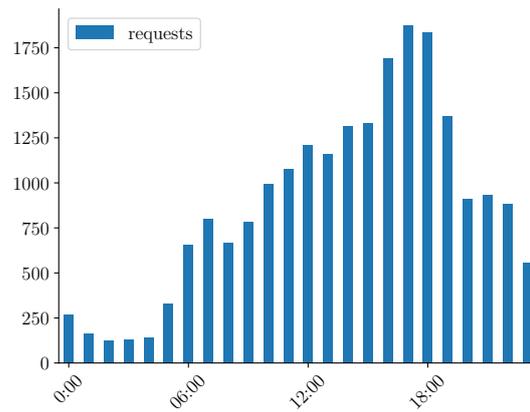


Figure 1. Hourly request throughout the day.

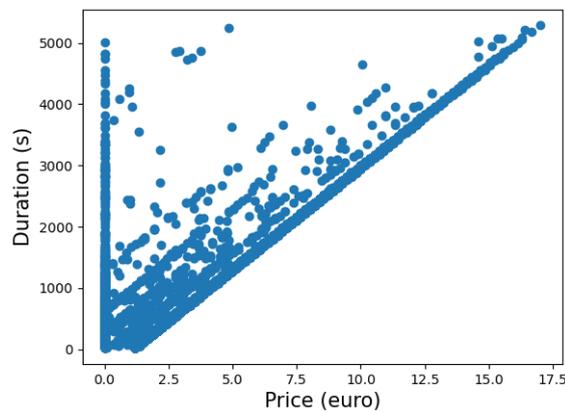


Figure 2. Time spent on vehicles (y) vs. money paid (x).

As can be seen from Figure 3, the days with the most requests are Thursday and Friday, not the weekends. Compared to the first three days of the week, Saturday and Sunday have comparable if not lower usage. It is possible that the majority of users may not be tourists, but city residents and the service may be primarily used for leisure, or it may be that residents are not in town on the weekend. In accordance with the speculation above, the last days of the working week are often associated with an increase in the value of the market stock, an effect known as the *weekend effect* [38,39]. It is our belief that this behavior is similar if not related to the *weekend effect*, thus identifying the user base as being primarily local.

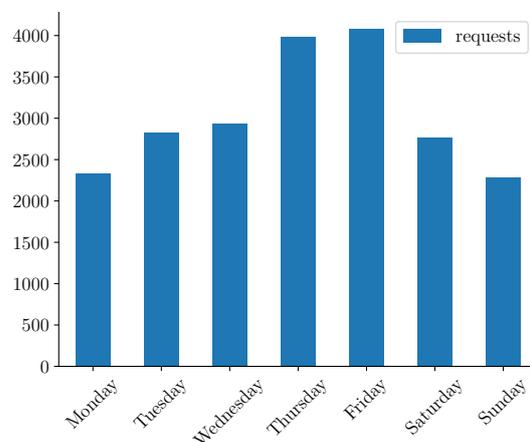


Figure 3. Daily request throughout the week.

4.2. Spatial Distribution

Our next step is to report the location of the requests based on the time distribution of the requests. Our heat map illustrates the actual density of requests starting, Figure 4a, and finishing points, Figure 4b. On both maps, there appears to be a high concentration area that gradually decreases with distance, with the exception of the eastern seaside area. In comparison to the number of trips that started in peripheral areas, the main difference between the two maps is the area from the center northward. There is no certainty as to whether the surplus of trips originated in the city center or in another peripheral area. A major issue that can arise from these distributions, namely a concentrated demand for requests in the central area and a more widespread distribution of returned vehicles, is that eventually there may be a shortage of vehicles in areas that are in high demand. In spite of the fact that this analysis does not take into account user behavior outside of location history, it is still essential to determine which areas require the most vehicles in order to identify where they are needed.

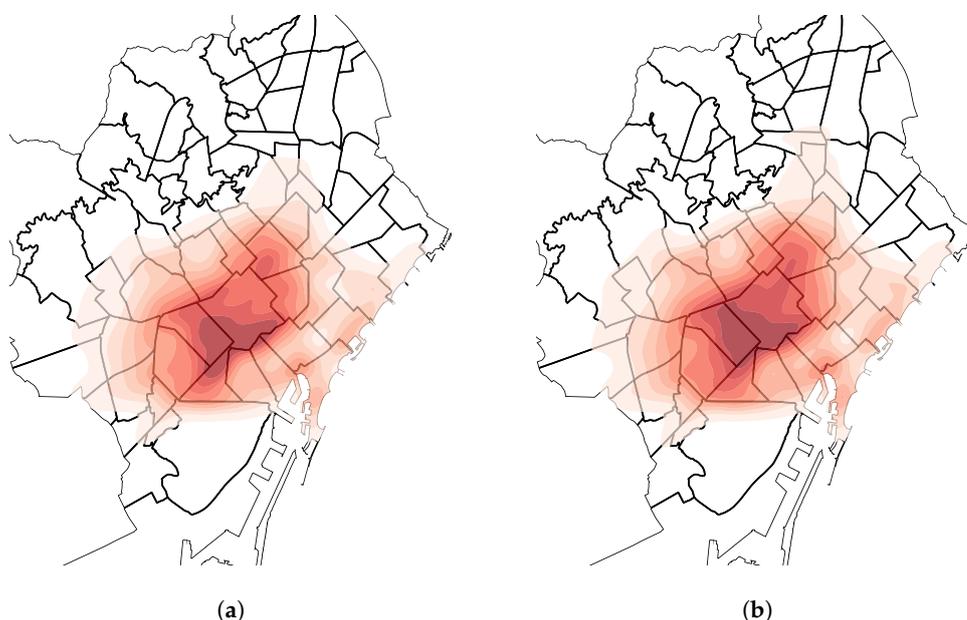


Figure 4. Location heat maps over the data set: (a) Starting points density; (b) Finishing points density.

4.3. Neighborhood Distribution

Section 4.2 shows a simple and raw analysis of the requests, so we group them according to their neighborhood. In addition to 73 neighborhoods, Barcelona is divided into 10 districts. Figure 5a,b illustrate the fact that the majority of requests originate and end in a neighborhood in the center of the city known as *La Dreta de l'Eixample*, where the majority of requests are initiated and completed. Among the most important places in the city are, for example, *Plaça de Catalunya*, *La Rambla* and *Passeig de Gràcia*, which are located in this neighborhood). Figure 6a,b show that the first three neighborhoods are the same for both starting and finishing points, while *La Sagrada Família* becomes the fourth neighborhood and *Sant Antoni* no longer appears in the top-5. This finding may be attributed to a lack of touristic presence in the city of Barcelona as a result of the pandemic situation and the hypothesized factors discussed in the previous section. Because of the high amount of marketplaces in the neighborhood, *Sant Antoni* is considered to be a popular destination, whereas *Sant Gervasi—Galvany* is mainly residential. Therefore, the migration from a work-related neighbor to a residential neighbor may indicate that the majority of users are locals rather than tourists in the considered time period. As a result of this correlation, the sharing company may be able to better target user segments in the future, for example, by partnering with local industries and offering free rides on the same line as food stamps.

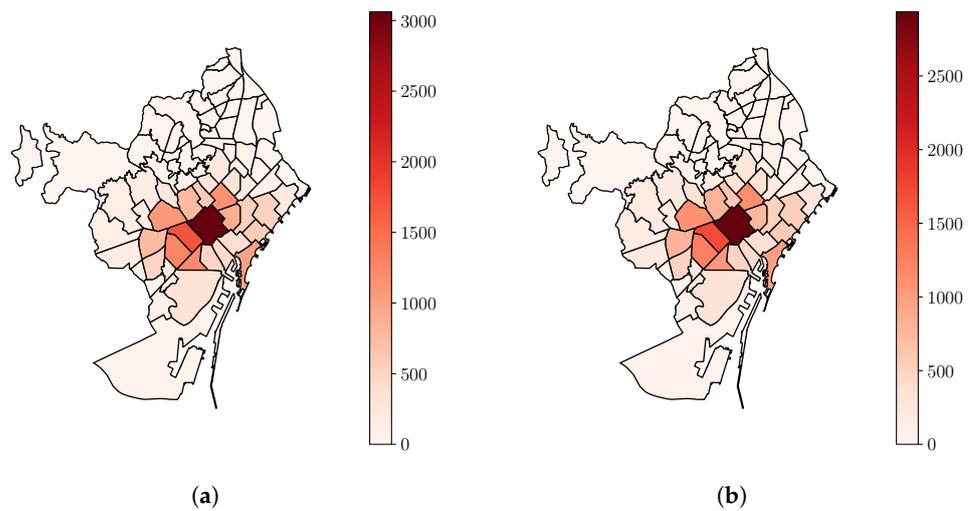


Figure 5. Neighborhood distribution maps of the whole data set: (a) Starting point per neighborhood; (b) Finishing points per neighborhood.

The previous analyses can be combined to obtain the plot in Figure 7, which shows the most common combination of the starting and finishing neighborhood. Clearly, the destinations do not change much between the start and end of the trip. It seems that users tend to use the vehicle-sharing service for short trips, often ending up in the same neighborhood or in a nearby neighborhood. Moreover, a majority of requests are received within *La Dreta de l’Exaïmple* which is considered to be the wealthiest area of Barcelona. Considering that the *La Dreta de l’Exaïmple* district covers 2% of Barcelona’s land (*La Dreta de l’Exaïmple* occupies 2.12 km² out of 101.9 km² for Barcelona), despite accounting for 32.7% of the vehicles sharing usage, this analysis can assist the sharing company in concentrating its efforts more locally rather than uniformly.

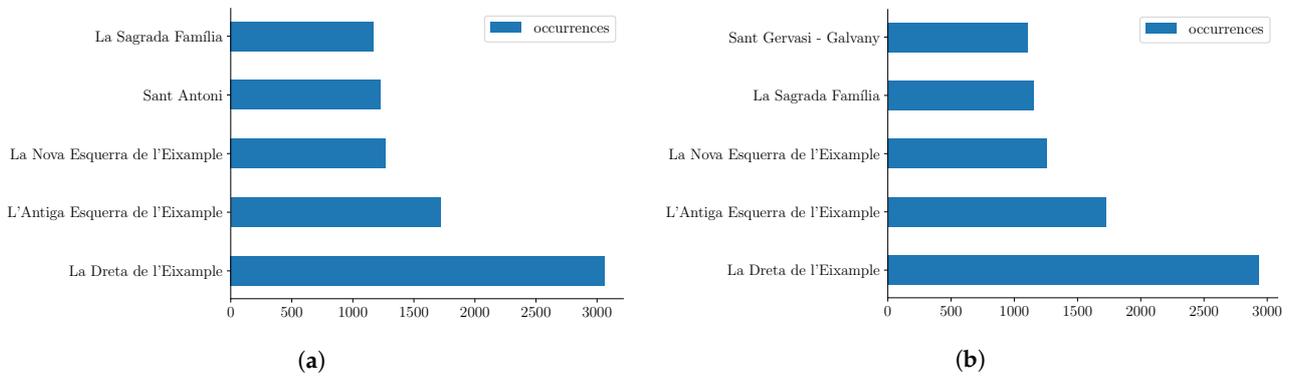


Figure 6. Five most frequent neighborhoods: (a) Starting points per neighborhood; (b) Finishing points per neighborhood.



Figure 7. Ten most common combinations of starting/finishing neighborhoods for trips.

4.4. Time and Spatial Distribution

We conclude by showing the distribution of requests throughout the day, combining what we have performed in Section 4.1 with Sections 4.2 and 4.3. As a result of the large volume of plots produced, we have presented only a small portion of them.

As a first step, we analyzed the peak hour, which is between 17:00 and 17:59. Specifically, the starting and finishing neighborhoods distribution maps are very similar to those in Figure 5, as well as the neighborhoods with the greatest number of requests (Figure 6). There are also similar results in the other two intervals of the peak hours, namely 16:00–16:59 and 18:00–18:59. It appears that these requests do not follow a particular pattern in time, and we can only make some assumptions about the reason: the city structure, i.e., most of the interests of the user are located in areas with high demand, and these demands are determined by the user's characteristics, such as age, occupation, and so on.

In addition, it is interesting to observe the request trend during the nighttime hours. It is expected that users' behavior differs in the morning or afternoon, specifically during the time interval of 22:00–22:59. Nevertheless, there are no significant differences in the trend, with the only change being that the third most frequent neighborhood is *La Barceloneta*, a neighborhood defined by the presence of wealthy complexes such as yachts and sunbathing beaches. As a result, the requests do not appear to be grouped according to a particular time period. It is also noteworthy that late night and early morning hours, despite their smaller demand, have a similar spatial distribution.

5. Experiments and Results

This section proposes two approaches to the problems discussed so far. Section 5.1 discusses the relocation problem, while Section 5.2 discusses the clustering of users.

5.1. Relocation Problem

The purpose of this section is to present a methodology to persuade users to adopt certain behaviors with the aim of relocating the vehicle from zones with high demand to zones with lower demand. Normally, this objective is achieved by offering some type of incentive to the customer: it could be a discounted rate for the trip or a bonus time that can be used in the future. Due to the lack of data, we are unable to investigate the economic aspect of the situation. As a result, we propose two subdivisions of the city, based on the analysis presented in Section 4. Eventually, a manual relocation of the vehicle may be needed, given that it appears too optimistic to achieve the desired relocation without the intervention of an operator for an extended period of time.

According to a procedure similar to that described in [18], we arbitrarily defined a central area, then categorized the neighborhood according to the number of requests received. As we have not found a particular time pattern, the subdivisions are based on all the collected data, without regard to the hour of the day.

Initial partitioning was determined as being the district containing the majority of the neighborhood, i.e., the *Example* district, in which most requests are generated, while the rest are classified as peripheral. In contrast, the second subdivision is based on the number of

requests (see Table 1 for the exact threshold values). The next step was to classify the origin neighborhood of the trips according to this subdivision in order to verify that it accurately reflects the demand for zones. The two partitioning methods are shown in Figure 8, with only two neighborhoods belonging to different categories.

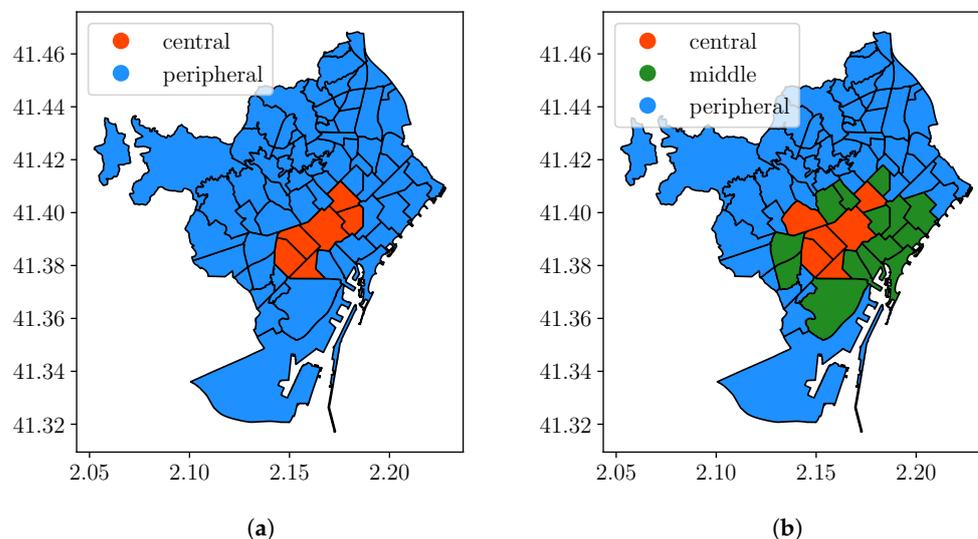


Figure 8. The two types of city’s areas division. Central, middle and peripheral neighbors are defined by the number of requests on the right and by the presence of most neighborhoods on the left: (a) Two-zones type, the central area is represented by *Example* district; (b) Three-zones type, the neighborhoods are classified according to the number of requests.

Table 1. Three-zones division resulting areas and relative thresholds.

Zone	Threshold (# of Request)	Area (km ²)	% Total Area
Central	n > 999	8.21	8.07
Middle	299 < n < 999	18.56	18.24
Peripheral	n < 299	75.00	73.69

5.1.1. Partition by Majority

Regarding the first type of subdivision, Table 2 shows a significant difference in the percentage of total area: neighborhoods classified as central account for only 7.34% of the total area. We can see in Figure 9 that despite this disproportion, 44.1% of requests originate and end in the central zones, while 21.3% originate and end in peripheral zones. The aim of the relocation algorithm is to ensure that there are vehicles available in the central area during peak hours, by offering incentives to routes that fall into the peripheral-central category (which is actually the least represented type of run).

Table 2. Two-zones division resulting areas.

Zone	Area (km ²)	% Total Area
Central	7.47	7.34
Peripheral	94.30	92.66

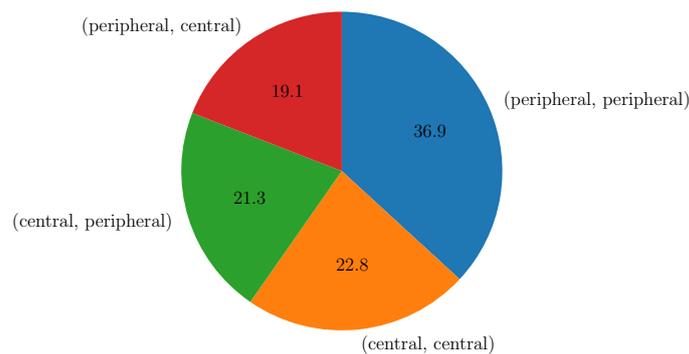


Figure 9. Number of trips classified according to two-zones division with respect to total number of trips.

5.1.2. Partition by Request

As for the second type of subdivision, in Table 1, once again the zones’ area and their relative percentage are shown. The situation for peripheral zones, which can be seen in Figure 10, is even more extreme: they account for 73.69% of the total area, but only 16.2% of trips originate from here, as opposed to 56% previously. According to this representation, the majority of trips are of the central-central type (24%) and the middle-middle type (16.6%). According to this subdivision, peripheral-central, middle-central, and peripheral-middle (although slightly fewer) trips should be encouraged.

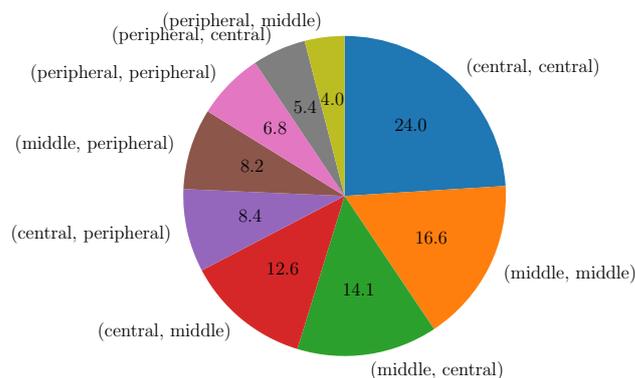


Figure 10. Number of trips classified according to three-zone division with respect to total number of trips.

5.1.3. Comparison

To test whether real users respond to incentives, some practical experiments in real scenarios need to be conducted, but since this is not possible, we can only make assumptions. The second method is more beneficial both to the service provider, who has a more refined picture of the customer on which to base the discount, and to the users, who will have a broader selection of options.

Since we cannot conduct experiments on real customers, we compared these results to those collected in Cologne in [18]. First, the two operating areas are comparable: for Barcelona is 101.77 km² and for Cologne is 94.2 km²; but the type of mobility analyzed in [18] refers to automobiles rather than scooters and bikes. Nevertheless, the coverage and number of passengers are similar to what is presented in this paper, so we can draw some conclusions from the comparison of [18] with ours. Based on two-zone subdivisions, Barcelona’s central area (7.34%) is half the size of Cologne’s (14.5%), yet more requests (44.1%) originate in Spanish cities compared to 35.24% in German cities. There is a higher proportion of requests coming from a smaller geographic area: this is an advantage for the provider, who is able to manage the fleet of vehicles easily.

5.2. User Clustering

It is now time for us to focus our attention on the analysis of user clusters.

5.2.1. RFMD Analysis

The first step in this process is the transformation of the dataset into a Recency-Frequency-Monetary-Duration dimension. The RFM values are calculated for each single user:

- Recency: The difference between the current and last day of vehicle use.
- Frequency: total number of times a customer used the vehicle.
- Monetary/Duration: total count of the euros/time spent.

For K-means to work, two conditions must be met: the distribution of the data should not be skewed (i.e., long-tail distribution) and it should be standardized (i.e., mean 0, standard deviation 1). The resulting distribution is shown in Figure 11.

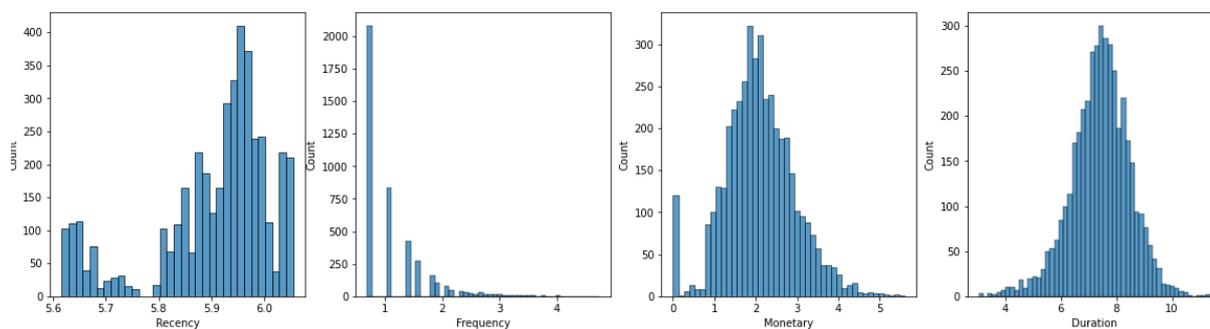


Figure 11. Plot of the scaled distribution of the RFMD values. From left to right: Recency, Frequency, Monetary, Duration.

5.2.2. Quantiles Buckets

First division can be estimated by discretizing the RFM variables into four equal-size buckets based on the sample quantiles (So that groups of users fall, respectively, into the 25%, 50%, 75% or 100% intervals of the distribution plots). In descending order of importance, these buckets represent the loyalty of the customers and their aptitude for spending money, beginning with the best users (Platinum) and ending with the worst users (Bronze). Accordingly, we report the number of users per level in Table 3 and it can be seen that most clients are at the top level.

Table 3. User Level clustering.

Platinum	Gold	Silver	Bronze
1097	1492	1004	653

5.2.3. Similarity Clustering with K-Means

According to the previous phase, users were tagged as Platinum, Gold, Silver, or Bronze based on their contribution to the company’s earnings, but the similarity between users was not considered. In fact, a clustering approach is necessary.

The K-means method identifies points in space that have a smaller distance from their neighbors and clusters them together. Having an understanding of how many clusters (K) there are before clustering is essential. This method involves plotting different values of cost (inertia) with changing K in order to determine the optimal number of clusters. Increasing the value of K will result in fewer elements in the cluster, as shown in Figure 12.

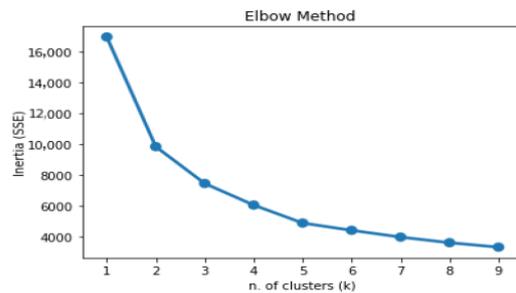


Figure 12. Elbow Plot for K-means clustering.

As an additional input, latitude and longitude coordinates are combined to account for position data. For each possible trip represented in the dataset, a spectral clustering algorithm has been applied. Since users may make multiple trips and end up in different locations, it is important to select only one cluster for each trip. We consider a user location as the one that is most frequently visited by the customer in this paper. Consequently, two additional features are included in K-means for each user (in addition to the RFMD values), and the results are shown in Figure 13.

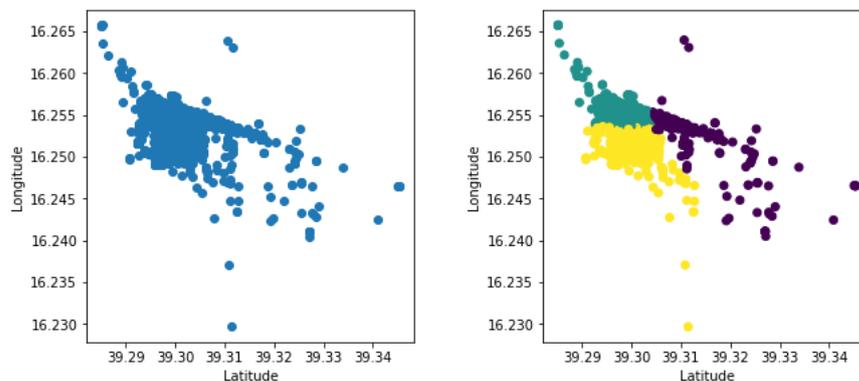


Figure 13. Spectral clustering for all the initial trip positions in the dataset. It should be noted that different trips have been made by the same user and that in some cases the user associated with a specific trip has been clustered in different groups. (Left): Latitude vs. longitude plot of all trips without clustering. (Right): three clusters of the same position trips, each color refers to a different spectral clustering.

In a subsequent step, the most relevant cluster for each user is identified in a table along with the RFMD values found previously and the K-means method is applied to $K = 3$ possible clusters. In Figure 12, it can be seen how the best possible solution is to choose $K = 3$. We report the resulting clustering in Table 4, showing how most users belong to cluster number 0, which stores the top users.

Table 4. Number of users for final clustering.

Cluster 0	Cluster 1	Cluster 2
2200	980	1004

6. Conclusions

This paper addresses two key issues faced by vehicle-sharing companies: vehicle relocation and user segmentation.

First, we analyze the provided dataset, Section 4, to infer user behavior with respect to vehicle arrangement and money spent over different time frames. It appears that users are most likely to rent vehicles in the late evening, probably because of working hours, and

that a large percentage of users do not pay for their trips, which may be due to the fact that the first trip is discounted.

In the next section, we present an implementation of a user-based relocation algorithm that can be further enhanced to consider the characteristics of each customer in a particular location. The first method is based on a *partition by majority*. Due to the number of neighborhoods in the city, the area has been arbitrarily divided into two zones, resulting in a reasonable distribution of requests. A *partition by request* was used to cluster the area into three zones in the second method. Through this approach, we are able to define two threshold values based on the number of requests made by users and obtain a more detailed picture of the demand. In order to determine where the vehicle-sharing company should focus its efforts, both methods can be used. This allows the company to focus on an area that accounts for 7.47% of the total area of the city, thereby reducing the relocation problem to the most affected areas of the city. Additionally, based on the financial ability of the company, different thresholds can be set to divide the area into smaller clusters. By doing so, a proportional effort can be made in the areas with a higher need, rather than a distributed effort across the entire city.

After addressing the problem of user clustering, we develop a system that combines spatial and K-means clustering to customize the experience of vehicle-sharing customers. Our approach is two-fold:

- We deploy a Recency-Frequency-Monetary-Duration analysis extended with latitude/longitude positions in order to extract user behavior from the data.
- Then we apply a user-based clustering algorithm to suggest similar users vehicles strategies.

In this paper, we are interested in offering discounts based on the RFMD values in order to influence user behavior. As a result of such discounts, users can be motivated to end their trip at a location of high demand, thus assisting with relocation.

Limitations and Future Works

As a result of a lack of data, our study presents some important limitations that could not be addressed. For example, it would be interesting to analyze the behavior of pre-pandemic and post-pandemic situations in relation to the current analysis. As a result of the restrictions present throughout most of the world, it is not surprising that COVID-19 had a great influence on vehicle sharing. Accordingly, it may be necessary to analyze how the post-pandemic situation is evolving and whether a convergence with pre-pandemic behavior is evident.

Furthermore, the data provided covers only one month, August to September, which is a holiday period. It is true that our data shows counterintuitive trends that can be attributed to the combination of restrictions and the holiday season, as well as to the nature of the user segment as a whole. In addition, we are unable to analyze the entire history of users based on only one month's worth of data. Most of the trips analyzed relate to novel users who have not previously interacted with the company. As a result of inspecting this additional data, we can gain a deeper understanding of the nature of users' behavior, allowing us to provide them with more personalized service.

Finally, we are not able to test our hypotheses in a real-world environment. We present some solutions to the relocation problem and the segmentation of users, however, these solutions remain theoretical since they cannot be tested. We believe, however, that vehicle-sharing companies can still benefit from our findings.

Author Contributions: All authors have equally contributed to this work. All authors read and agreed to the published version of the manuscript.

Funding: This work has been supported by the project "HERMES WIRED" funded by Sapienza University of Rome within the Sapienza founding scheme 2020, and by the P.O.N. "Ricerca e innovazione 2014–2020—azione IV.5" funding schema.

Data Availability Statement: Not applicable.

Acknowledgments: The preliminary studies were performed in cooperation with the students Lorenzo De Rebotti and Alessandro Reali.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. TomTom. *Traffic Congestion Ranking | TomTom Traffic Index*; TomTom: Amsterdam, The Netherlands, 2022.
2. Ferrero, F.; Perboli, G.; Rosano, M.; Vesco, A. Car-sharing services: An annotated review. *Sustain. Cities Soc.* **2018**, *37*, 501–518. [[CrossRef](#)]
3. Jochem, P.; Frankenhauser, D.; Ewald, L.; Ensslen, A.; Fromm, H. Does free-floating carsharing reduce private vehicle ownership? The case of SHARE NOW in European cities. *Transp. Res. Part A Policy Pract.* **2020**, *141*, 373–395. [[CrossRef](#)] [[PubMed](#)]
4. Fellows, N.; Pitfield, D.E. An economic and operational evaluation of urban car-sharing. *Transp. Res. Part D Transp. Environ.* **2000**, *5*, 1–10. [[CrossRef](#)]
5. Kabasakal, İ. Customer segmentation based on recency frequency monetary model: A case study in E-retailing. *Bilişim Teknol. Derg.* **2020**, *13*, 47–56. [[CrossRef](#)]
6. Likas, A.; Vlassis, N.; Verbeek, J.J. The global K-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461. [[CrossRef](#)]
7. Chen, Y.L.; Kuo, M.H.; Wu, S.Y.; Tang, K. Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. *Electron. Commer. Res. Appl.* **2009**, *8*, 241–251. [[CrossRef](#)]
8. Barth, M.; Todd, M. Simulation model performance analysis of a multiple station shared vehicle system. *Transp. Res. Part C Emerg. Technol.* **1999**, *7*, 237–259. [[CrossRef](#)]
9. Brandstätter, G.; Gambella, C.; Leitner, M.; Malaguti, E.; Masini, F.; Puchinger, J.; Ruthmair, M.; Vigo, D. Overview of Optimization Problems in Electric Car-Sharing System Design and Management. In *Dynamic Perspectives on Managerial Decision Making: Essays in Honor of Richard F. Hartl*; Dawid, H., Doerner, K.F., Feichtinger, G., Kort, P.M., Seidl, A., Eds.; Dynamic Modeling and Econometrics in Economics and Finance; Springer International Publishing: Cham, Switzerland, 2016; pp. 441–471. [[CrossRef](#)]
10. Weikl, S.; Bogenberger, K. Relocation Strategies and Algorithms for Free-Floating Car Sharing Systems. *IEEE Intell. Transp. Syst. Mag.* **2013**, *5*, 100–111. [[CrossRef](#)]
11. Bruglieri, M.; Colorni, A.; Luè, A. The Vehicle Relocation Problem for the One-way Electric Vehicle Sharing: An Application to the Milan Case. *Procedia-Soc. Behav. Sci.* **2014**, *111*, 18–27. [[CrossRef](#)]
12. Gambella, C.; Malaguti, E.; Masini, F.; Vigo, D. Optimizing relocation operations in electric car-sharing. *Omega* **2018**, *81*, 234–245. [[CrossRef](#)]
13. Todd, M.; Xue, L.; Barth, M.J. *User-Based Vehicle Relocation Techniques for Multiple-Station Shared-Use Vehicle Systems*; Transportation Research Board: Washington, DC, USA, 2004; p. 17.
14. Stokkink, P.; Geroliminis, N. Predictive user-based relocation through incentives in one-way car-sharing systems. *Transp. Res. Part B Methodol.* **2021**, *149*, 230–249. [[CrossRef](#)]
15. Di Febbraro, A.; Sacco, N.; Saeednia, M. One-Way Car-Sharing Profit Maximization by Means of User-Based Vehicle Relocation. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 628–641. [[CrossRef](#)]
16. Clemente, M.; Fanti, M.P.; Mangini, A.M.; Ukovich, W. The Vehicle Relocation Problem in Car Sharing Systems: Modeling and Simulation in a Petri Net Framework. In *Proceedings of the Application and Theory of Petri Nets and Concurrency*; Colom, J.M., Desel, J., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; pp. 250–269. [[CrossRef](#)]
17. Clemente, M.; Fanti, M.P.; Iacobellis, G.; Nolich, M.; Ukovich, W. A Decision Support System for User-Based Vehicle Relocation in Car Sharing Systems. *IEEE Trans. Syst. Man, Cybern. Syst.* **2018**, *48*, 1283–1296. [[CrossRef](#)]
18. Lippoldt, K.; Niels, T.; Bogenberger, K. Analyzing the Potential of User-Based Relocations on a Free-Floating Carsharing System in Cologne. *Transp. Res. Procedia* **2019**, *37*, 147–154. [[CrossRef](#)]
19. Schiffer, M.; Hiermann, G.; Rüdell, F.; Walther, G. A polynomial-time algorithm for user-based relocation in free-floating car sharing systems. *Transp. Res. Part B Methodol.* **2021**, *143*, 65–85. [[CrossRef](#)]
20. Weikl, S.; Bogenberger, K. A practice-ready relocation model for free-floating carsharing systems with electric vehicles—Mesoscopic approach and field trial results. *Transp. Res. Part C Emerg. Technol.* **2015**, *57*, 206–223. [[CrossRef](#)]
21. Barth, M.; Todd, M. User Behavior Evaluation of an Intelligent Shared Electric Vehicle System. *Transp. Res. Rec.* **2001**, *1760*, 145–152. [[CrossRef](#)]
22. Seign, R.; Bogenberger, K. Prescriptions for the Successful Diffusion of Carsharing with Electric Vehicles. In *Conference on Future Automotive Technology*; Springer: Berlin/Heidelberg, Germany, 2012; p. 8.
23. Niels, T.; Bogenberger, K. Booking Behavior of Free-Floating Carsharing Users: Empirical Analysis of Mobile Phone App and Booking Data Focusing on Battery Electric Vehicles. *Transp. Res. Rec. J. Transp. Res. Board* **2017**, *2650*, 123–132. [[CrossRef](#)]
24. Caggiani, L.; Camporeale, R.; Ottomanelli, M. A dynamic clustering method for relocation process in free-floating vehicle sharing systems. *Transp. Res. Procedia* **2017**, *27*, 278–285. [[CrossRef](#)]

25. Brandizzi, N.; Russo, S.; Brociek, R.; Wajda, A. First Studies to Apply the Theory of Mind Theory to Green and Smart Mobility by Using Gaussian Area Clustering. In Proceedings of the ICYRIME Conference, Online, 9 July 2021; Volume 3118, pp. 71–76.
26. Frith, C.; Frith, U. Theory of mind. *Curr. Biol.* **2005**, *15*, R644–R645. [[CrossRef](#)]
27. Liu, H.; Hu, Z.; Mian, A.; Tian, H.; Zhu, X. A new user similarity model to improve the accuracy of collaborative filtering. *Knowl.-Based Syst.* **2014**, *56*, 156–166. [[CrossRef](#)]
28. AL-Bakri, N.F.; Hashim, S.H. Collaborative Filtering Recommendation Model Based on K-means Clustering. *Al-Nahrain J. Sci.* **2019**, *22*, 74–79. [[CrossRef](#)]
29. Gong, S. A collaborative filtering recommendation algorithm based on user clustering and item clustering. *J. Softw.* **2010**, *5*, 745–752. [[CrossRef](#)]
30. Liao, Q.; Yang, F.; Zhao, J. An improved parallel K-means clustering algorithm with MapReduce. In Proceedings of the 2013 15th IEEE International Conference on Communication Technology, Guilin, China, 17–19 November 2013; pp. 764–768.
31. Khurana, P.; Parveen, S. Effective hybrid recommender approach using improved K-means and similarity. *Int. J. Comput. Trends Technol.* **2016**, *36*, 147–152. [[CrossRef](#)]
32. Kumar, M.; Yadav, D.; Singh, A.; Gupta, V.K. A movie recommender system: Movrec. *Int. J. Comput. Appl.* **2015**, *124*, 7–11. [[CrossRef](#)]
33. Wu, Z.; Chen, Y.; Li, T. Personalized recommendation based on the improved similarity and fuzzy clustering. In Proceedings of the 2014 International Conference on Information Science, Electronics and Electrical Engineering, Sapporo, Japan, 26–28 April 2014; Volume 2, pp. 1353–1357.
34. Schmidt, A.; Fink, C.; Barash, V.; Cameron, C.; Macy, M. Using spectral clustering of hashtag adoptions to find interest-based communities. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–7.
35. Sheugh, L.; Alizadeh, S.H. A note on pearson correlation coefficient as a metric of similarity in recommender system. In Proceedings of the 2015 AI & Robotics (IRANOPEN), Qazvin, Iran, 12 April 2015; pp. 1–6.
36. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
37. Observatori del Turisme a Barcelona. Tourism Activity Report. 2021. Available online: <https://www.observatoriturisme.barcelona/> (accessed on 17 February 2022).
38. French, K.R. Stock returns and the weekend effect. *J. Financ. Econ.* **1980**, *8*, 55–69. [[CrossRef](#)]
39. Miller, E.M. Why a weekend effect? *J. Portf. Manag.* **1988**, *14*, 43. [[CrossRef](#)]