TRANSPORTES
revistatransportes.org.br
anpet
associação nacional de pesquisa e ensino em transportes
anpet.org.br

Analysis of the temporal profile of use of bikesharing stations in the Bikesampa system

Análise temporal do perfil de uso das estações de bicicletas compartilhadas do sistema Bikesampa

Thais Malta Baracat[1], **Orlando Strambi**[2], **Patrícia Lavieri**[3]

[1]University of São Paulo, São Paulo – Brazil, thais.baracat@usp.br
[2]University of São Paulo, São Paulo – Brazil, ostrambi@usp.br
[3]University of Melbourne, Victoria – Australia, patricia.lavieri@unimelb.edu.au

Submitted:
1 de agosto de 2022
Accepted for publication:
23 de novembro de 2022
Published:
27 de fevereiro de 2023
Editor in charge:
Antônio Nélson Rodrigues da Silva

Keywords:
Bikesharing.
Temporal pattern.
Clusters.
Spatial correlation.

Palavras-chave:
Bicicletas compartilhadas.
Perfil de uso.
Clusterização.
Autocorrelação espacial.

DOI: 10.58922/transportes.v31i1.2825

OPEN ACCESS
CC BY SA

ABSTRACT
Bikesharing systems have gained popularity over the years and now face the challenge of being responsive and of meeting the growing demand. Thus, understanding the temporal pattern of bikesharing trips is paramount. This study examined data from Bikesampa (a fixed-station system operating in the Brazilian city of São Paulo) and applied k-means clustering to stations according to the hourly demand of pickups and returns. The results revealed three clusters: (i) balanced, (ii) unbalanced, with higher rates of bike pickup in the morning and (iii) unbalanced, with higher rates of bike return in the morning. A spatial autocorrelation analysis showed that cluster membership was not randomly distributed over space, suggesting an association with characteristics of the urban environment, and indicating that the system may require different rebalancing strategies depending on the stations location. Such understanding can help guide the development of operational strategies and user incentive policies to improve the efficiency of bikesharing systems.

RESUMO
Os sistemas de bicicletas compartilhadas ganharam popularidade nos últimos anos, de forma que precisam ser responsivos à demanda. Logo, é necessário entender o comportamento temporal destas viagens. Esta pesquisa estudou o sistema de estações fixas Bikesampa, da cidade de São Paulo, e teve como objetivo classificar as estações segundo a demanda horária de retiradas e devoluções. Utilizou-se um agrupamento k-médias, resultando em três *grupos* de estações: (i) balanceado, (ii) desbalanceado, com maior número de retiradas na manhã, e (iii) desbalanceado, com maior número de devoluções na manhã. Verificou-se por meio da análise de autocorrelação espacial que os grupos não se distribuem aleatoriamente no espaço, sugerindo uma associação com características do espaço urbano e a necessidade de diferentes estratégias de rebalanceamento entre estações dependendo da localização. O conhecimento do comportamento temporal das viagens de bicicletas compartilhadas permite o desenvolvimento de políticas de operação e de incentivo ao usuário para melhorar a eficiência desses sistemas.

1. INTRODUCTION

Bicycle use has grown steadily, be it for the increase in recreational and sports use (Levy, Golani and Ben-Elia, 2019), or for the benefits to public health and to the environment (Shaheen, Cohen and Martin, 2013; Zhang et al., 2015), or even for being economically accessible to different social classes (Chardon, Caruso and Thomas, 2017; Pritchard et al., 2019).

Bikesharing systems – service offering temporary loan of bicycles in strategically located regions – explain part of this increase (Faghih-Imani et al., 2017; Fishman, 2016; Wang et al., 2016).

These systems provide convenience to users and make this means of transport more accessible and flexible, since they dispense with bicycle ownership and reduce the entrance barrier, introducing new cyclists to the system (Benedini, Lavieri and Strambi; Eren and Uz, 2020; Fishman, Washington and Haworth, 2013). Bikesharing systems are divided into two types: (i) fixed-stations, in which the user picks up and returns the bicycle at a station, and (ii) without stations, which provides a door-to-door service (Wu, Kim and Chung, 2021). This article focuses on the fixed-station systems.

With the growing popularity of bikesharing, understanding its dynamic is fundamental to develop systems that are efficient and responsive to the demand (Shaheen, Cohen and Martin, 2013). Some recent studies have been dedicated to understanding the behavior of the hourly demand for bikesharing along the day. Most of the trips are taken in peak periods in the morning and in the evening, frequently for pendular first and last mile trips (Gu, Kim and Currie, 2019; Wu, Kim and Chung, 2021). There are also more diffuse uses along the day, especially for education and leisure reasons (Faghih-Imani et al., 2017; Gu, Kim and Currie, 2019).

The general aim of this study is thus to complement the literature knowledge on the hourly variation in demand of fixed-station bikesharing systems, aiming to guide the elaboration of policies for improving operations and for stimulating and attracting users. The object of study was the Bikesampa system in the city of São Paulo. Figure 1 presents the knowledge gaps and hypotheses, as well as the specific objectives and contributions of the present research.

Gaps	Recent studies generally employ aggregate metrics for analyzing demand, and do not assess the hourly variation in demand
Hypotheses	The stations present different average temporal demand profiles; these temporal profiles are not randomly distributed in space
Objectives	- Identifying clusters of stations from the temporal demand profiles - Analyzing whether the clusters of stations are randomly distributed in space, suggesting some degree of association with characteristics of the urban space
Contributions	- Complementing the studies on the hourly variation in bikesharing demand - Understanding whether the temporal behavior of demand allows increasing the operational efficiency of redistributing bicycles among stations - Improving the decision-making process for bikesharing system operators, especially from the viewpoint of the critical process of repositioning bicycles among stations and of rebalancing the pickup and return flows. Stations are thus prevented from being saturated or depleted, for increasing the system reliability for the user and, therefore, making the bicycles use more attractive - There are few studies on bikesharing-related behavior in Latin America and especially in a megalopolis such as São Paulo, which underwent a fast urbanization process and has an eccentric trip dynamic

Figure 1. Gaps, hypotheses, objectives and contributions

2. LITERATURE REVIEW

There was a significant increase in the number of researches about bikesharing systems in recent years. Si et al. (2019) analyzed the most recurrent topics in the literature between 2010 and 2018 and identified the most studied themes regarding the so-called 3rd generation systems, which count on docks or devices that prevent bicycle thefts, such as telecommunication systems and other technologies to identify the user and the location of the bicycles (Demaio and Gifford, 2004). These authors particularly point out that the most urgent problem of this generation is the imbalance between bicycle demand and supply at the station.

To better understand the behavior of the cyclists in these systems, numerous studies seek to establish an association between this behavior and characteristics of the system operation, of the urban space, of social economy and of land use. These researches frequently use an aggregate metrics in time (day, month or year) to represent the demand for bikesharing, hindering the understanding of its hourly variation (Faghig-Imani et al., 2014; Tran, Ovtracht and D’Arcier, 2015).

The knowledge of hourly variation is relevant to identify factors that foster the use of bicycles at specific times, helping to minimize the risk of saturating or depleting the stations and reducing the cost of redistributing bicycles, improving the quality and availability of the service (Fricker and Gast, 2016; Tran, Ovtracht and D'Arcier, 2015).

In places in Europe, Asia and America, demand peaks at bikesharing stations were verified in the morning and in the evening, and occasionally at lunchtime (Faghih-Imani et al., 2017; Gu, Kim and Currie, 2019; Hu et al., 2021; Schimohr and Scheiner, 2021; Zhu et al., 2020). This behavior is typical of pendular trips (Faghih-Imani et al., 2014; Mix, Hurtubia and Raveau, 2022; Tran, Ovtracht and D’Arcier, 2015).

The use in the afternoon and evening period is particularly greater as compared to the other periods of the day, probably due to additional trips made at this time for leisure and gastronomic reasons. This hypothesis was confirmed by the studies of Faghih-Imani et al. (2014) and of Tran, Ovtracht and D'Arcier (2015).

In places with mixed land use or with different types of activities, the demand for both pickups and returns is more diffuse in time (Faghih-Imani et al., 2017). The literature also attributes this behavior to the existence of academic hubs, once the schedules are not generally fixed (Faghih-Imani et al., 2014; Gu, Kim and Currie, 2019).

Aiming to complement the studies on temporal demand at a disaggregated level, this research identified clusters of temporal profiles for pickup and return at stations.

3. MATERIALS AND METHODS

For contextualization purposes, we first provide an overview of bike use in the city of São Paulo. Then, the data employed in the research regarding the Bikesampa bikesharing system are presented. Next comes the method for constructing the average temporal profiles of use of the stations that served as a basis to form the station clusters. Lastly, a spatial autocorrelation analysis is presented to assess whether the clusters are randomly distributed in the urban space or whether they have a pattern that can be associated to the characteristics of the station’s surroundings.

3.1. Bicycle use in the city of São Paulo

The city of São Paulo spreads over 1,521 km^2 and has about 12 million inhabitants (IBGE, 2021), thus configuring the most densely populated city in Brazil and in Latin America. The trips in the São Paulo Metropolitan Region are distributed as follows: 36% in public transport, 31% in individual motorized vehicles, 32% on foot and 1% on bicycles (Metrô-SP, 2020b). As compared to the use of private bicycles, a larger proportion of bikesharing trips is made in a combination with other modes, particularly public transport. This is an expected result, since it is not possible, with few exceptions, to transport the bicycle inside buses, trains or subway (Benedini, Lavieri and Strambi, 2019).

Regarding the cycling infrastructure, São Paulo counts on about 700 kilometers of bicycle lanes, bike paths and cycleroutes (CET, 2020). However, 72% of the trips made on bicycles do not use segregated lanes (Metrô-SP, 2020a). The main purposes for using a bicycle are: work (69%), school (14%) and leisure (6%). Bicycle trips are mostly made by men (90%), between 30 to 49 years old (41%), who have completed secondary education (40%) and having a family income of about R$1,908 to R$3,816 (in Brazilian reais of 2018) (43%) (Metrô-SP, 2020a).

3.2. The Bikesampa system

There are currently two fixed-station bikesharing systems in the city of São Paulo, both considered to be third-generation. About 93% of the stations belong to the Bikesampa system (BikeItau, 2022), associated to the Itaú Bank and operated by the Tembici company. Given its strong representativeness, this system was chosen as the object of study.

The reference period of the study considered working days between February 01, 2020, and March 15, 2020 (Tembici informed a marked drop in the number of trips as from March 16, 2020, due to the restrictive measures taken to control the Covid-19 pandemic). During this period, Bikesampa operated with 279 stations, concentrated on the West, South and Central zones; no stations were located in the North and East zones of São Paulo. It is worth highlighting that rainy or cold days were not disregarded, as due to the concentration of stations in space, this effect is believed to be transversal for all the stations.

For the analysis period, Tembici made available a database on trips consisting of day, time and pickup station, and the same information for the return station. Note that there is not an identifier for individuals, so that individuals may appear with distinct frequencies in the databases, depending on the number of trips made during the period. Also, the following characteristics of the stations were available: location coordinates, number of docks and the existence of a pocket (a station reserve capacity). Of the 279 stations, only 239 were selected for analysis (illustrated in Figure 2), the following being excluded:

- 35 stations that operated in only one of the months (February or March). Since the method for building the temporal profiles considers the average flow in the period under analysis, this exclusion prevents the flows from being underrepresented, since the period is smaller than that under analysis;
- 1 station that did not have the information on the coordinates or number of docks;
- 2 stations located in the campus of the University of São Paulo, where the spatial and socioeconomic characteristics differ from the remainder of the city;
- 2 stations located in Largo da Batata, which have flows and number of docks much higher than the remaining stations, and may be considered discrepant observations.

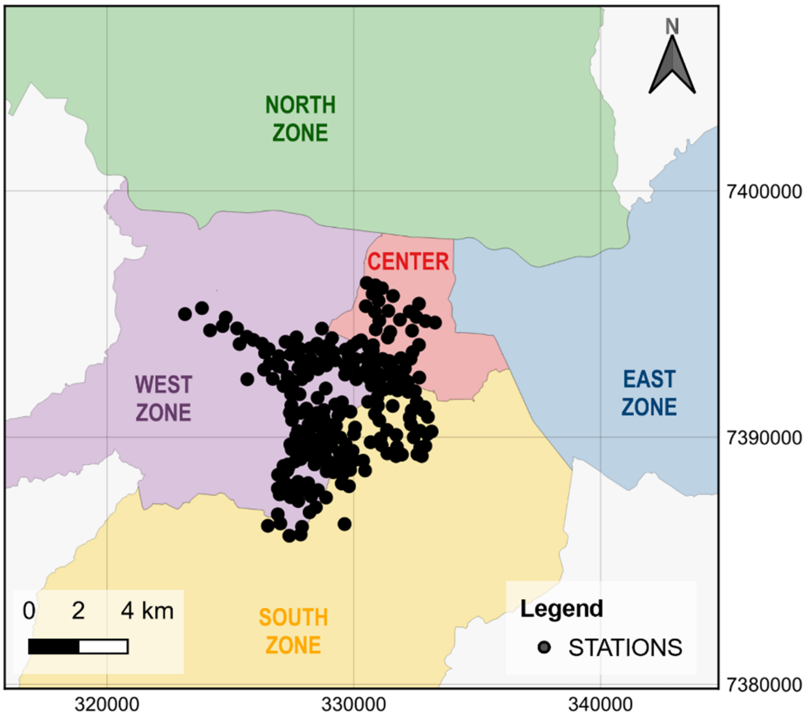

Figure 2. Location of the 239 Bikesampa stations selected

3.3. Clusters of stations

3.3.1. Construction of the average temporal profiles of use of the stations

To understand the temporal behavior of the stations, the concept of average hourly profiles of use of the stations was employed, according to the methodology proposed by Gu, Kim and Currie (2019) for the city if Suzhou (China), which served as a reference for the present research. Firstly, the average hourly flow (f) is calculated, which corresponds to the average of the number of bicycles returning to ($fi_{m,t}$) or leaving ($fo_{m,t}$) a given station for each hour of the day along the period of analysis, by the equations:

$$fi_{m,t} = \frac{1}{k} \sum_1^k fi_{m,t,k} \quad (1)$$

$$fo_{m,t} = \frac{1}{k} \sum_1^k fo_{m,t,k} \quad (2)$$

where: k: number of days in the analysis period;
t: time of day;
m: station;
i/o: return (in) or pickup (out).

Since the stations present different orders of magnitude of flows, to compare the profiles of the variation in average hourly flows along the day, a standardization is carried out, dividing the average hourly flow by the maximum average daily flow (f_{max}), be it of return or pickup, as represented in the following formulation:

$$f_{max,m} = max_m[fi_{m,t=0}, fi_{m,t=1}, \ \dots, fi_{m,t=23}, \ fo_{m,t=0}, fo_{m,t=1}, \ \dots, fo_{m,t=23}] \quad (3)$$

Lastly, the NF_m vector is created with 24 pairs of normalized average hourly flows (totaling 48 variables), varying between 0 and 1

$$NF_m = \left[\frac{fi_{m,t=0}}{f_{max,m}}, \frac{fi_{m,t=1}}{f_{max,m}}, \ \dots, \frac{fi_{m,t=23}}{f_{max,m}}, \frac{fo_{m,t=0}}{f_{max,m}}, \frac{fo_{m,t=1}}{f_{max,m}}, \ \dots, \frac{fo_{m,t=23}}{f_{max,m}}\right] \quad (4)$$

3.3.2. Classification of stations according to their temporal profiles

Once the NF_m vectors were defined, Gu, Kim and Currie (2019) applied a cluster analysis technique (clusterization), aiming to identify groups with similar temporal profile of use of the stations. Clusterization is a multivariate, unsupervised technique that uses continuous or binary variables to cluster observations, aiming to increase the internal uniformity of each cluster and the heterogeneity among clusters (Fávero and Belfiore, 2017). Particularly, the k-means clusterization technique aims to minimize the intracluster variance (MacQueen, 1967).

In the non-hierarchical clustering techniques, such as k-means, the number of clusters is defined *a priori*. For this definition, the silhouette coefficient, a measure of cohesion, is frequently used; it compares the average distance among all the elements in their cluster and the average distance to all the elements in each of the other clusters (Rousseeuw, 1987).

This procedure was applied to the average temporal profiles of the Bikesampa stations. Given the reduced use of the bicycles during the overnight period, only the pickups and returns occurring between 06:00 and 22:59 were considered, reducing the size of the NF_m vector to 34. The k-means clusterization and the silhouette curve were computed in the R language with the *kmeans* and *silhouette* functions of the *cluster* package.

3.4. Spatial autocorrelation analysis of the clusters of stations

To assess whether the clusters of stations are randomly distributed in space, Moran's Index was employed; it is an autocovariance measure given by Equation 5 (Almeida, 2012).

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{5}$$

where: n: number of regions;
z: values of the standardized variable of interest;
w_{ij}: values of the weight matrix W regarding region i and region j;
S_0: sum of the elements of weight matrix W.

According to Almeida (2012), the value expected from Moran's I is I_{exp} = $-(n-1)^{-1}$. The null hypothesis (H_0) is not rejected when $I \leq I_{exp}$. In this scenario, the spatial pattern observed is equally likely to any other (spatial randomness). If this condition is not satisfied, the alternative hypothesis (H_1) is considered; it indicates positive spatial autocorrelation (similarity between the attribute value and the spatial location) or negative (otherwise). To visualize the spatial autocorrelation in each region, LISA cluster maps are commonly used. Each region with significant local statistics is colored according with the following convention:

- High-High: the region and the neighboring ones present high concentration of the phenomenon;
- Low-Low: the region and the neighboring ones present low concentration of the phenomenon;
- Low-High: the region has low concentration of the phenomenon, whereas the neighboring ones present inverse behavior;
- High-Low: the region has high concentration of the phenomenon, whereas the neighboring ones present inverse behavior.

Moran's I was independently calculated for each cluster of stations in the *GeoDa* software. The variable of interest used was the number of stations in each of the 54 zones of the

Origin-Destination Survey (ODS) of the São Paulo Metropolitan Region (Metrô, 2020b). These 54 zones cover the area of operation of the 239 Bikesampa selected for the analysis (see Figure 3). This choice of variable of interest is due to the limitation of Moran's Index, which requires numerical variables.

The closeness between regions can be represented by distance or proximity matrices. The latter alternative considers only regions sharing borders, under the hypothesis that contiguous areas have greater spatial interaction (Almeida, 2012). The proximity matrix used was the contiguity one, of the *Queen Contiguity* type, an analogy to the moves allowed to the "Queen" chess piece, in which the borders with an extension greater than zero are considered to be the contiguous vertices.

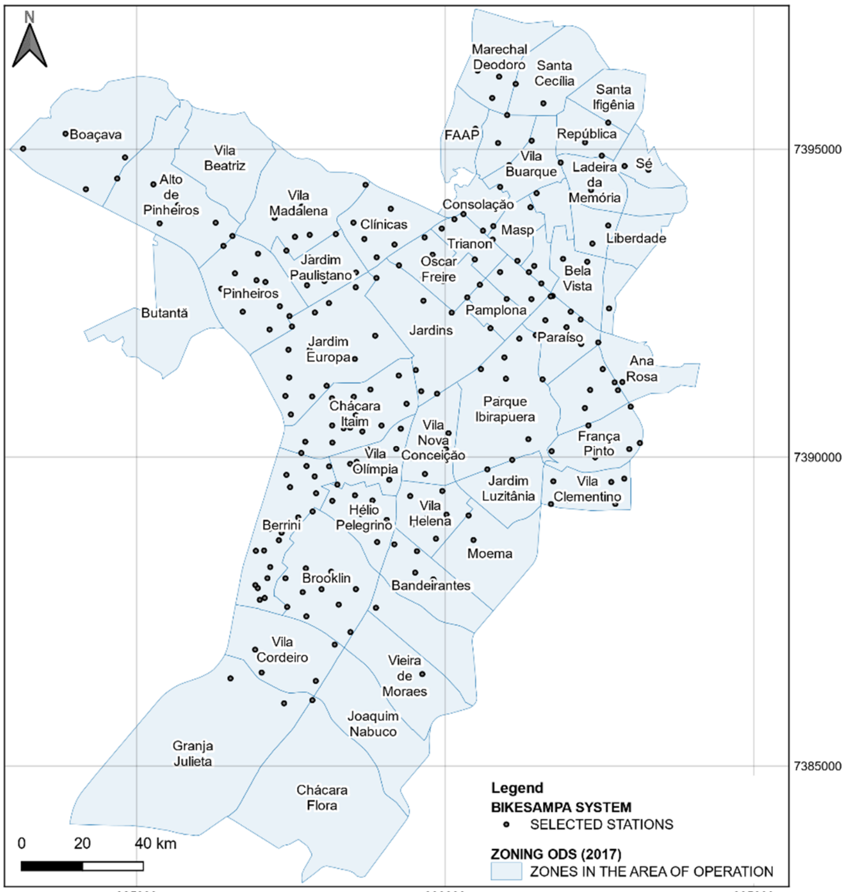

Figure 3. Zoning adopted for calculating Moran's Index

4. RESULTS AND DISCUSSION

The results and the discussion on the study are presented as follows. Firstly, the clusters of stations are presented, followed by their spatial autocorrelation analysis.

4.1. Clusters of stations

Different values for the number of clusters were used in the k-means procedure. For choosing the most adequate value, the silhouette curve was employed (Figure 4). The measure of cohesion is maximum with three clusters, this being the value chosen for clustering the profiles.

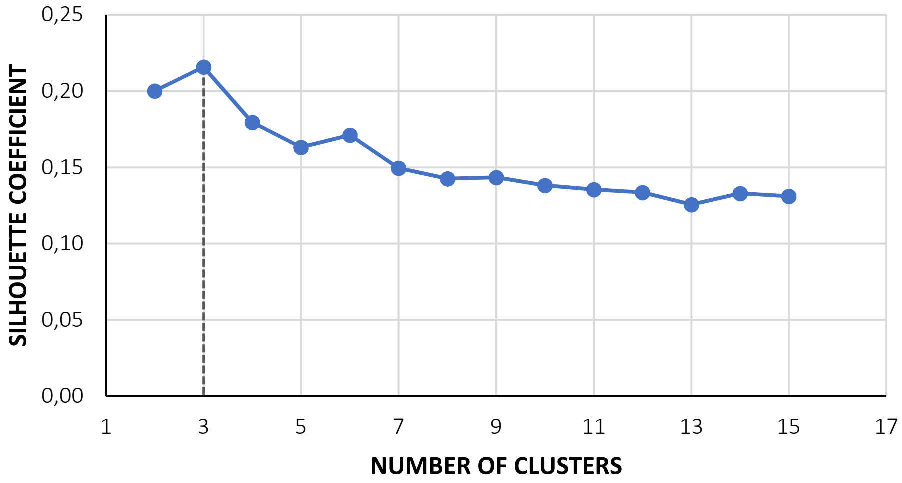

Figure 4. Silhouette curve for different values of number of clusters

Therefore, the 239 Bikesampa stations were divided into three clusters of temporal profiles of use (see Figure 5). The location of the stations of the different clusters in the urban space of the region of interest is presented in Figure 6. The clusters characteristics are:

Cluster 1 – Balanced: normalized average flow of pickups and returns more balanced along the day, except in the morning peak period, when the pickup flow is slightly higher. There is an increase in flow between 12:00 and 13:00, and especially in the late afternoon, early evening. The higher normalized average flow is equal to 0.80, at 18:00. 95 stations (39.7% of the system) belong to this cluster. The stations are concentrated in the axis of Av. Paulista and Av. Rebouças, and near the Ibirapuera Park, regions with supply of services and other activities, in addition to residences;

Cluster 2 – Morning Pickups: normalized average flow of pickups and returns unbalanced between the morning and evening periods. In the morning, there is a greater pickup flow, reaching a peak of normalized flow equal to 0.62 at 08:00. In the evening, the return flow is more expressive, reaching a peak of 0.73 at 18:00. 87 stations (36.4% of the system) belong to this cluster. These stations are on the boundaries of the system, in residential regions;

Cluster 3 – Morning returns: normalized average flow of pickups and returns unbalanced between the morning and evening periods. In the morning, there is a greater flow of returns, reaching a peak of 0.88 at 08:00. In the evening, the pickup flow is greater, with a maximum of 0.62 at 18:00. 57 stations (23.9% of the system) belong to this cluster. These stations are concentrated on Av. Faria Lima and on Av. Berrini, regions of commerce and services.

A visual inspection of the distribution of stations in the urban space suggest that stations in each cluster do not seem to be randomly distributed. The stations in **Cluster 2 – Morning Pickups** seem to be located on the boundaries of the system. In turn, the stations in **Cluster 3 – Morning returns** seem to be more concentrated on the main commercial axes of the city. The spatial randomness hypothesis is verified by the spatial autocorrelation analysis in the next subitem.

As compared to the reference work by Gu, Kim and Currie (2019), the clustering obtained shows some similarity with the results for the city of Suzhou, in China; in that study, the measure of cohesion is also maximum with three clusters. However, the temporal patterns are not different. Although both systems present a cluster with gradual increase of flows along the day, in Suzhou this cluster is unbalanced and counts with one station only. In turn, in São Paulo, **Cluster 1 – Balanced** has a balanced profile and comprises about 40% of the stations.

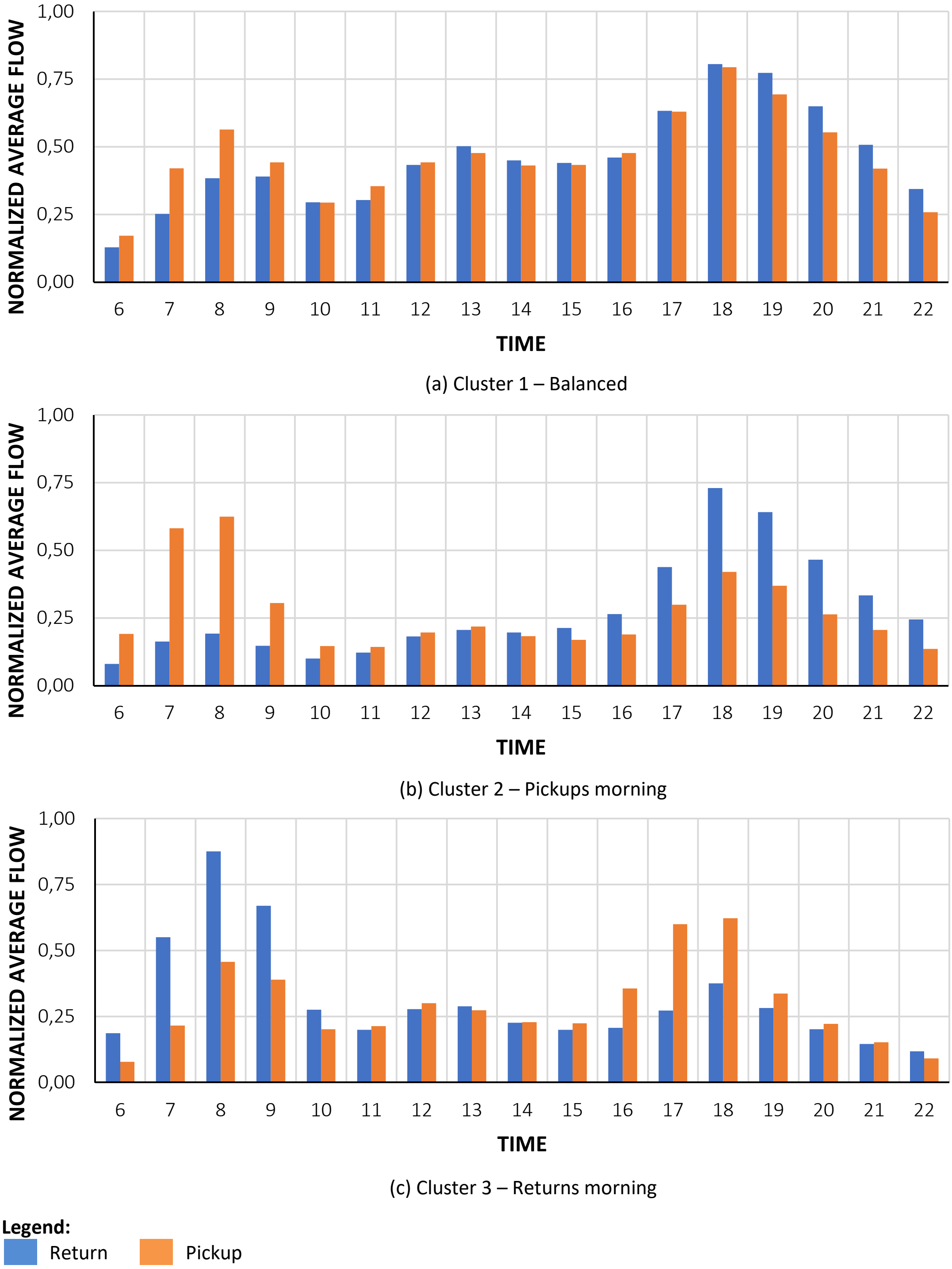

Figure 5. Temporal profile of use for each station cluster

The two other clusters present two-peak temporal patterns, as observed in Suzhou. However, in São Paulo, both patterns are unbalanced between pickups and returns. This behavior can be explained by the urban structure of São Paulo, known for having work and residential poles concentrated in specific areas of the city.

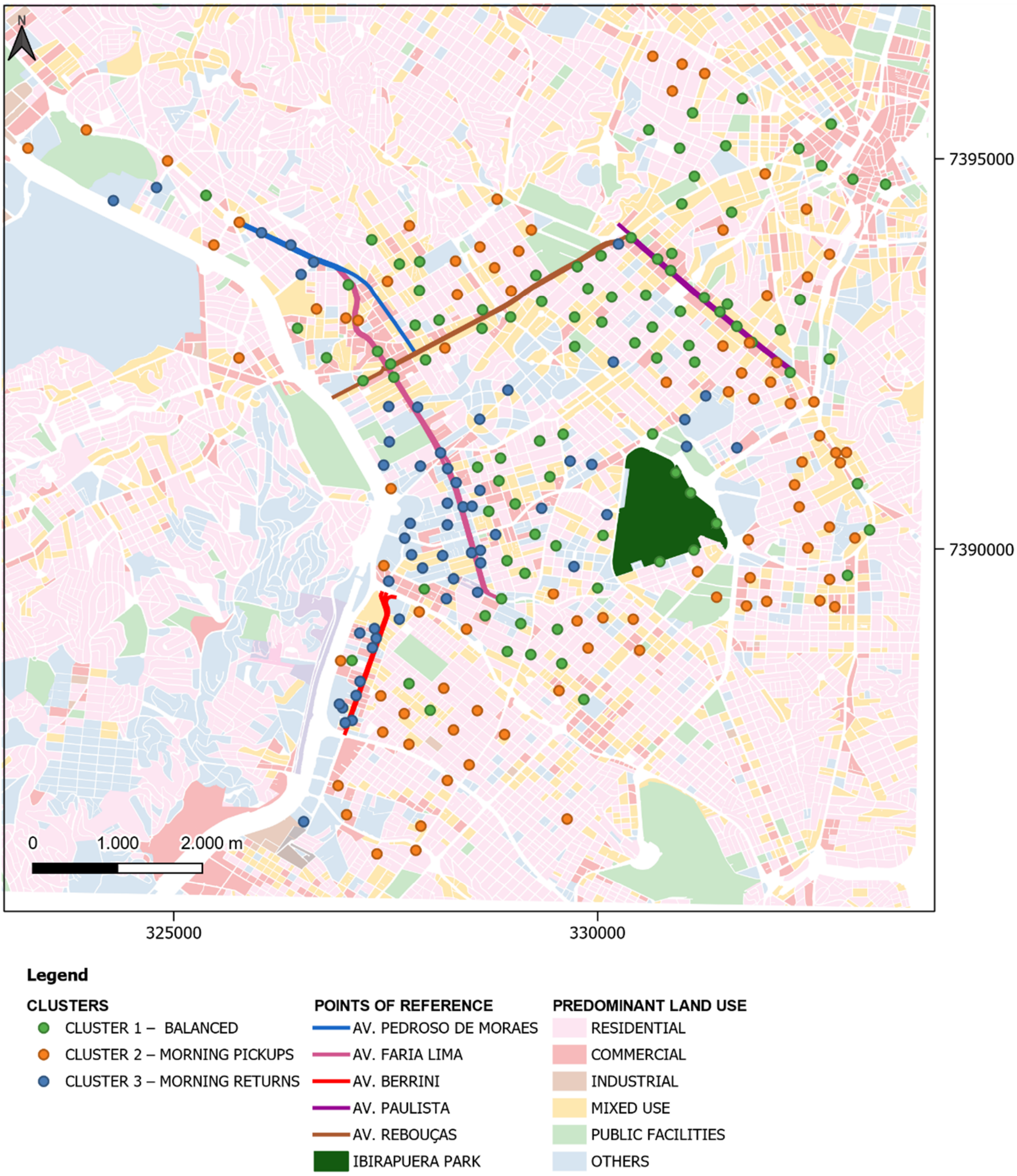

Figure 6. Spatial distribution of the temporal profile clusters

The results indicate peak periods in the morning and in the early evening, between 07:00 and 09:00 o'clock and between 17:00 and 19:00, respectively, in São Paulo. For a comparison, the peak demand periods for trips in the São Paulo Metropolitan Region in motorized vehicles occur between 06:00 and 07:00 o'clock in the morning, and between 17:00 and 18:00 in the early evening (Metrô-SP, 2020b). Thus, the bikesharing peak demand periods occur slightly after the motorized mode peak periods.

Compared to the results of the international literature, the peak demand periods found for the Bikesampa system coincide with the studies conducted in Santiago and in Lyon (Mix, Hurtubia and Raveau, 2022; Tran, Ovtracht and D'Arcier, 2015). In turn, in Suzhou, the evening peak period occurs earlier, between 16:00 and 19:00 (Gu, Kim and Currie, 2019), whereas the cities of Barcelona and Sevilla present longer peak demand periods, which extend until 10:00 in the morning and up to 22:00 in the evening (Faghih-Imani et al., 2017).

Even though the international literature shows higher flows in the evening period, only two clusters present this behavior in São Paulo. The stations in **Cluster 3 – Morning returns** present a high flow of returns in the morning, exceeding the maximum values of the other clusters at any other time of day. A hypothesis for this behavior would be the spatial closeness of these stations, allowing a larger number of shorter trips during a period.

Understanding the time pattern of the pickup and return flows, especially in peak periods, is therefore important to prevent the stations from saturating or from being depleted, discouraging or inhibiting bikesharing use.

In practical terms, the clusters with unbalanced pickups and returns may be the target of a special operation. In this context, the stations in **Cluster 2 – Morning** p**ickups** could receive relocated bicycles in the morning period, when the demand is higher and there is a greater probability that stations in this cluster be emptied. Inversely, the stations in **Cluster 3 – Morning** r**eturns** could receive relocated bicycles in the early evening period. The stations in **Cluster 1 – Balanced** are less dependent on bicycle relocation operations.

Relocation policies aimed at users could also be adopted between stations, stimulating cyclists to reposition the bicycles. An example would be implementing a reduced price for those that take trips in the counterflow in the peak period, thus helping to mitigate the lack of bicycles or of docks at the stations of unbalanced clusters.

4.2. Spatial autocorrelation of the clusters of stations

The null hypothesis of random spatial distribution can be rejected for the three clusters of stations, since calculated Moran's I exceeds its expected value (Table 1).

Table 1 – Moran's Global Index

Cluster	Calculated Value	Expected Value
Cluster 1 – Balanced	0.172	
Cluster 2 – Morning Pickups	0.100	-0.019
Cluster 3 – Morning Returns	0.373	

For understanding the local behavior of this statistic, Figure 7 illustrates the LISA maps, which indicate the areas with significance equal to or less than 5%. The results indicate that:

Cluster 1 – Balanced: according to Figure 7a, the Jardim Europa, Jardins, Pamplona, Vila Nova Conceição and Vila Olímpia areas, regions in which residences, commerce and services coexist, present a "High-High" behavior. Conversely, the Granja Julieta, Chácara Flora and Joaquim Nabuco areas, in the southernmost, with a smaller number of jobs in services and greater residential use, present "Low-Low" behavior. The Clínicas and Campinas areas are classified as "Low-High", regions in which other clusters predominate;

Cluster 2 – Morning pickups: Figure 7b shows that the Jardim Luzitania area presents "High-High" behavior, very likely due to the smaller supply of jobs and larger number of residences. In turn, the Masp area (on Av. Paulista) has a "Low- Low" value, due to the major presence of stations of **Cluster 1 – Balanced** in the region. Classified as "Low-High" regions are the Ibirapuera Park and Ana Rosa areas, whereas the Marechal Deodoro, Ladeira da Memória and Bexiga areas are classified as "High-Low". Since the stations in this cluster present a more dispersed pattern, there are more areas classified as having transition behavior (High-Low or Low-High);

Cluster 3 – Morning returns: Figure 7c demonstrates that this cluster presents exclusively "High-High" behavior in the Berrini, Chácara Itaim, Helio Pelegrino, Jardins, Vila Nova Conceição and Vila Olímpia areas. These regions are known for being poles of jobs in services, hosting office buildings.

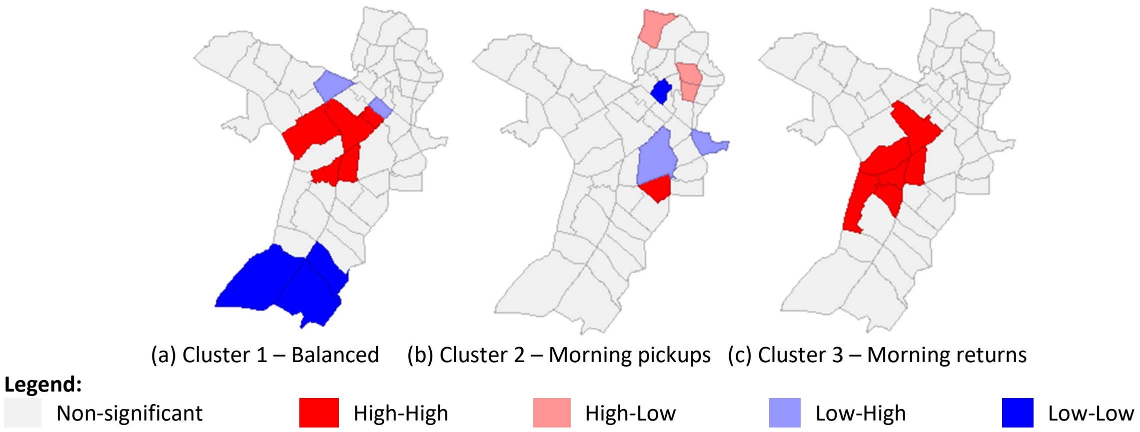

Figure 7. LISA cluster map

It is worth noting that the system stations do not result from a natural, non-controllable process, but rather from the operator decision-making, based on demand expectations, the characteristics of road geometry for positioning the stations and the type and intensity of activities in the neighborhood of the station. The location of the stations thus results from an anthropic, non-random event, a fact corroborated by the results of the spatial autocorrelation analysis.

Firstly, Figure 6 suggests a possible association between the land use characteristics surrounding a station and its pertinence to a cluster of temporal profile of demand. Knowing or inferring a given temporal profile is an information that allows orienting the operator decision-making as regards the possible need to reposition bicycles to prevent saturating the stations, especially for the clusters with unbalanced pickup and return flows (**Cluster 2 – Morning pickups** and **Cluster 3 –Morning returns**).

Secondly, the results contribute to the discussion of new possibilities to treat the repositioning of bicycles. In areas classified as "High-High", for example, there is a greater concentration of stations of a given cluster in the area and in its neighboring zones, there consequently being a greater chance for the occurrence of bicycle or dock deficit for the clusters of stations with unbalanced pickup and return flows.

In such cases, the operator could analyze at least four distinct operation policies to reduce the risk of not meeting the demand for bicycles or for free docks in a station, namely: (i) increasing the number of docks (and bicycles) at the station, (ii) implementing pockets, (iii) implementing new stations in the region, and (iv) repositioning bicycles coming from other stations.

Figure 8 shows an example of the complementary character of the distinct station clusters in neighboring areas. The Berrini area counts on more stations with a prevalence of morning returns (**Cluster 3 – Morning returns**). The surplus of bicycles returned in this period can be relocated to stations with a greater morning pickup flow (**Cluster 2 –Morning pickups**) in the Brooklin and Vila Cordeiro areas. In the early evening, the process is inverted.

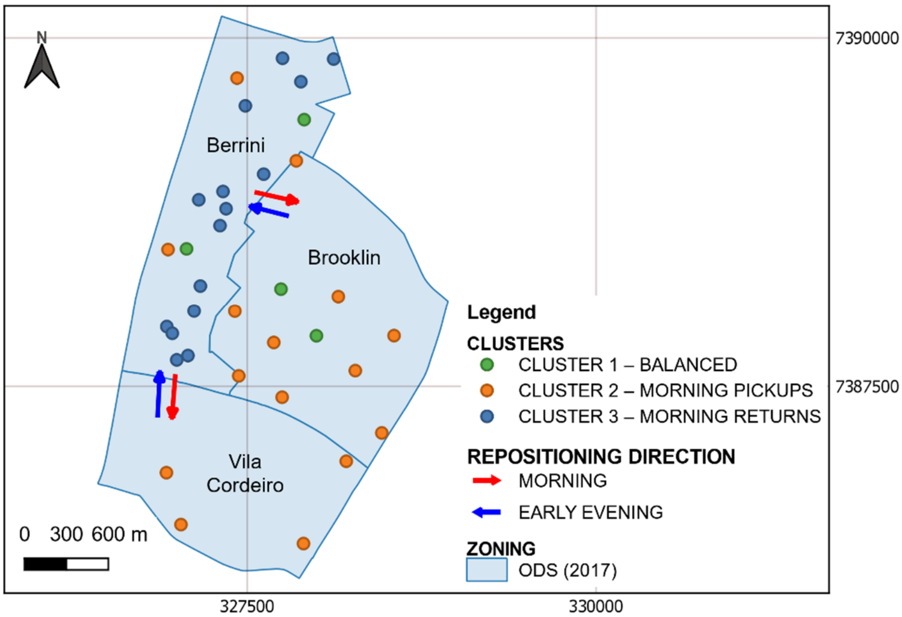

Figure 8. Suggestion for rebalancing in the Berrini-Brooklin/Vila Cordeiro area

5. CONCLUSION

This research investigated the Bikesampa bikesharing system, which operates in the city of São Paulo. The stations were classified into three clusters based on the average hourly flows of pickups and returns: Balanced, Morning pickups and Morning returns. These clusters are not randomly distributed in space, there being a pattern for each cluster, suggesting an association with the land use characteristics.

Cluster 1 – Balanced has a balanced pattern between pickups and returns, with a gradual increase in demand along the day. This behavior is probably due to their being located in areas with a larger proportion of mixed land use. In turn, **Cluster 2 – Morning pickups** has an unbalanced two-peak pattern, with a larger number of pickups in the morning and returns in the early evening, which is possibly explained by the predominance of residential land use around these stations. Lastly, **Cluster 3 – Morning returns** has an unbalanced two-peak pattern, yet with a larger number of returns in the morning and pickups in the early evening, possibly explained by the greater commercial land use.

The results contribute to the design of new incentive and operation policies for a better management of the system by the operator, to make it more efficient and friendly to the user. To circumvent the critical problem of bicycle or dock deficit in peak periods, some recommendations are made, especially for the stations with an imbalance between the pickup and return flows.

The first suggestion is that the operator considers the creation of an incentive policy to reward the cyclists that take trips in the counterflow in the peak period, since these users indirectly participate in the bicycle repositioning process. In this sense, the operator could assess the impact of implementing reduced prices or other types of advantages in the subscription plan.

After that, knowing the time of day in which there is a greater probability of bicycle or dock deficit at the stations, the bicycles could be repositioned in the peak periods of demand for each type of cluster. The stations with greater pickup in the morning are thus expected to receive bicycles redirected from other stations or count on a reserve capacity in that period, whereas the stations with greater return in the morning, receive them in the afternoon.

Complementarily, given that the station clusters are not randomly distributed in space, it is reasonable to admit that there is an association between the temporal profile of demand of a station and the characteristics of its surroundings. Therefore, as a recommendation for future works, the suggestion is studying the relationship between the clusters and the characteristics of the system operation, of the urban space and the socioeconomic attributes, using confirmatory modeling techniques.

ACKNOWLEDGEMENTS
The authors thank CNPq - The National Council for Scientific and Technological Development for the financial support under the grant 312016/2018-2. In addition, the authors would like to thank the Tembici team, in particular Renata Rabello, who provided the data and many insightful comments regarding the bikesharing system analyzed. We also acknowledge the contributions from Mariana Giannotti, Fernando Gomes and Flavio Soares for their constructive feedback.

REFERENCES

Almeida, E. (2012). *Econometria espacial aplicada* (1ª ed.). Campinas: Alínea.

Benedini, D.; P. Lavieri e O. Strambi (2019) Understanding the use of private and shared bicycles in large emerging cities: The case of São Paulo, Brazil. *Case Studies on Transport Policy*, v. 8, p. 564-575. DOI: 10.1016/j.cstp.2019.11.009.

BikeItau (2022). Aluguel de Bicicleta em São Paulo. Disponível em: < https://bikeitau.com.br/> (Acesso em: 29/09/2022).

Chardon, C. M. D.; G. Caruso e I. Thomas (2017) Bicycle sharing system ‘success’ determinants. *Transportation Research Part A*, v. 100, p. 202-2014. DOI: 10.1016/j.tra.2017.04.020.

CET (2020). Mapa de Infraestrutura Cicloviária. Disponível em: <http://www.cetsp.com.br/consultas/bicicleta/mapa-de-infraestrutura-cicloviaria.aspx> (Acesso em: 05/05/2022).

Demaio, P. e J. Gifford (2004) Will Smart Bikes Succeed as Public Transportation in the United States? *Journal of Public Transportation*, v. 7, n. 2, p. 1-15. DOI: 10.5038/2375-0901.7.2.1.

Eren, E. e V. E. Uz (2020) A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, v. 54, p. 1-12. DOI: 10.1016/j.scs.2019.101882.

Faghih-Imani, A. et al. (2017) An empirical analysis of bike sharing usage and rebalancing: evidence from Barcelona and Seville. *Transportation Research Part A*, v. 97, p. 177-191. DOI: 10.1016/j.tra.2016.12.007.

Fávero, L. P. e P. Belfiore (2017) *Manual de Análise de Dados: Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®* (1ª ed). São Paulo: GEN LTC.

Fishman, E. (2016) Bikeshare: A Review of Recent Literature. *Transport Reviews*, v. 36, p. 92-113. DOI: 10.1080/01441647.2015.1033036.

Fishman, E.; S. Washington e N. Haworth (2013) Bike share: A synthesis of the literature. *Transport Reviews*, v. 33, p. 148-165. DOI: 10.1080/01441647.2013.775612.

Fricker, C. e N. Gast (2016) Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *EURO Journal on Transportation and Logistics*, v. 5, n. 3, p. 261-291. DOI: 10.1007/s13676-014-0053-5.

Gu, T.; I. Kim e G. Currie (2019) Measuring immediate impacts of a new mass transit system on an existing bike-share system in China. *Transportation Research Part A*, n. 124, p. 20-39. DOI: 10.1016/j.tra.2019.03.003.

Hu, S. et al. (2021) Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic. *Journal of Transport Geography*, v. 91, p. 16. DOI: 10.1016/j.jtrangeo.2021.102997.

IBGE (2021) Cidades. Disponível em: <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama> (Acesso em: 28/09/2022).

Levy, N.; C. Golani e E. Ben-Elia (2019) An exploratory study of spatial patterns of cycling in Tel Aviv using passively generated bike-sharing data. *Journal of Transport Geography*, v. 76, p. 325-334. DOI: 10.1016/j.jtrangeo.2017.10.005.

Macqueen, J. B. (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, v. 1, p. 281-297. Califórnia: University of California Press.

Metrô-SP (2020a) Caracterização socioeconômica dos deslocamentos. Disponível em: <http://www.metro.sp.gov.br/pesquisa-od/arquivos/Caracteriza%C3%A7%C3%A3o_Socioecon%C3%B4mica_dos_Deslocamentos_2017.pdf> (Acesso em: 15/09/2020).

Metrô-SP (2020b) Pesquisa Origem e Destino de 2017. Disponível em: <http://www.metro.sp.gov.br/pesquisa-od/arquivos/Ebook%20Pesquisa%20OD%202017_final_240719_versao_4.pdf>. (Acesso em: 01/07/2020).

Mix, R.; R. Hurtubia e S. Raveau (2022) Optimal location of bike-sharing stations: A built environment and accessibility approach. *Transportation Research Part A*, v. 160, p. 126-142. DOI: 10.1016/j.tra.2022.03.022.

Pritchard, J. P. et al. (2019) Potential impacts of bike-and-ride on job accessibility and spatial equity in São Paulo, Brazil. *Transportation Research Part A*. 386-400. DOI: 10.1016/j.tra.2019.01.022.

Schimohr, K. e J. Scheiner (2021) Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption. *Journal of Transport Geography*, v. 92. DOI: 10.1016/j.jtrangeo.2021.103017.

Shaheen, S. A.; A. Cohen e E. Martin (2013) Public Bikesharing in North America: Early Operator Understanding and Emerging Trends. *Transportation Research Record*, v. 2387, n. 1, p. 83-92. DOI: 10.3141%2F2387-10.

Si, H. et al. (2019) Mapping the bike sharing research published from 2010 to 2018: A scientometric review. *Journal of Cleaner Production*, v. 213, p. 415-427. DOI: 10.1016/j.jclepro.2018.12.157.

Tran, D. T.; N. Ovtracht e F. B. D'Arcier (2015) Modeling Bike Sharing System using Built Environment Factors. *Procedia CIRP*, v. 30, p. 293-298. DOI: 10.1016/j.procir.2015.02.156.

Wang, X. et al. (2016) Modeling bike share station activity: Effects of nearby business and jobs on trips to and from stations. *Journal of Urban Planning and Development*, v. 142, n. 1, p. 04015001. DOI: 10.1061/(ASCE)UP.1943-5444.0000273.

Wu, C.; I. Kim e H. Chung (2021) The effects of built environment spatial variation on bike-sharing usage: A case study of Suzhou, China. *Cities*, v. 110. DOI: 10.1016/j.cities.2020.103063.

Zhang, L. et al. (2015) Sustainable bike-sharing systems: characteristics and commonalities across cases in urban China. *Journal of Cleaner Production*, v. 97, p. 124-133. DOI: 10.1016/j.jclepro.2014.04.006.

Zhu, R. et al. (2020) Understanding spatio-temporal heterogeneity of bike-sharing and scooter-sharing mobility. *Computers, Environment and Urban Systems*, v. 81. DOI: 10.1016/j.compenvurbsys.2020.101483.