Using public data to assess operational strategies of bicycle-sharing systems

Gustavo Gonçalves Oliveira[1[0000-0002-9282-7157]] and
Antônio Nélson Rodrigues da Silva[2[0000-0003-0996-8600]]

University of São Paulo - São Carlos School of Engineering, São Carlos SP 13561-410, Brazil
gustavogoncalves@usp.br, anelson@sc.usp.br

Abstract. Considering that users of bicycle-sharing systems (BSS) are directly affected by the operational decisions implemented regularly, we believe this information should be open to the public. As this is not usually the case, the objective of this study is to develop a strategy to assess some of the unrevealed operational decisions using public data. We used Python scripts to process and analyze data provided by government agencies (e.g., census data and a digital elevation model), open access datasets from online platforms (e.g., Open Street Map and Google Maps), and the bicycle-sharing system operator's API. We used the proposed approach to analyze the relatively small (i.e., 20 stations) BSS currently in operation in Vila Velha, Brazil. As a first result, we confirmed that factors usually considered when planning a BSS, such as the distribution of population and activities, the topography, and the existence of bikeway infrastructure were associated with the location of the stations. In addition, the procedure used to create a demand profile based on the production and attraction of trips also produced reasonable estimates for utilitarian trips, even though it underestimated leisure trips. We also showed that various stations have level of service issues, mainly because of insufficient numbers of bicycles. Our approach is limited by the absence of a validation procedure, which could be easily implemented if we had access to a sample of the records containing the individual movements of bicycles.

Keywords: Assessment of Bicycle-sharing Systems, Public Data, Level of Service of Bicycle-sharing Systems.

Introduction

Bicycle-sharing systems (BSSs) combine two potential contributions to sustainable mobility: the cycling mode of transport and shared mobility. This helps explain why they are present in more than 400 places worldwide, including big cities [1, 2]. The most common strategies originally used to operate BSSs were based on docked (i.e., station-based systems) bicycles. These systems have progressed and undergone updates since their creation, and dockless programs have gained attention since 2016 [3]. However, experiences from cities indicate that docked (i.e., station-based systems) and dockless systems can be complementary, covering different kinds of trips.

Docked systems can also be more suitable for small cities and places with low bicycle use [4]. Nonetheless, the large number of systems already built with docking infrastructure means that research on the operation of these systems can still be constructive.

Users of docking systems have the possibility to get or return a bicycle at any station, as long as they find available bicycles and docks, respectively. Although this characteristic benefits the users, it can be a problem for the operator due to the variation and uncertainty of the times and locations of users' entry and exit points. This means that, depending on the system's utilization, the number of bicycles in each station varies, without any direct form of demand control by the operator. Sometimes, this can limit the system's operation, resulting in empty stations and displeased users, who have to travel to another station using other means of transport or even be forced to cancel their trips [5].

Different strategies have been developed for bicycle-sharing system planning and operations [e.g., 6, 7, 8, 9, 10, 11 and 12]. However, as most of the operators are private businesses, for commercial reasons they do not share operational data or any information about their planning and management choices. Considering that users are directly affected by operational decisions, we believe these choices should be open to the public. As this is not usually the case, this study aims to develop a strategy to assess some of the operators' unrevealed strategies by means of public data. We consider this would help to better understand the problem of unavailable bicycles or docks and guide the public discussion regarding the matter. We used the proposed approach to analyze the bicycle-sharing system in Vila Velha, which is a coastal city with an estimated population of 501,000 in 2020 located in the state of Espírito Santo, Brazil. The BSS has 20 stations and was chosen for its relatively small size. To achieve our goal, the following research questions were addressed: (RQ1) Are the stations' locations compatible with the geographical and demographic characteristics of the city? (RQ2) Is the distribution of docks among the stations compatible with their region's potential for production and attraction of trips? (RQ3) Is the system's level of service compatible with recommended values? If not, what are the possible reasons for that?

Methodology

The procedures we developed for this study are summarized in Figure 1 and detailed next.

The first step of our approach to assessing bicycle-sharing system operational strategies using public data was to identify and analyze factors that operators usually consider when planning a BSS, such as: population density, topography, and existing bikeways (according to The Bikeshare Planning Guide [2]). Regarding population density, we used grid data from 2010 [13], which is the latest national census with available data. For topography, we used a digital elevation model provided by INPE - The National Institute for Space Research [14]. For geolocating bikeways, we used Open Street Map data [15]. Finally, for the locations of the stations, we used the bicycle-sharing system operator's API [16]. The different datasets were combined and mapped using Python scripts (main packages: Kepler.gl [17] and GeoPandas [18]).

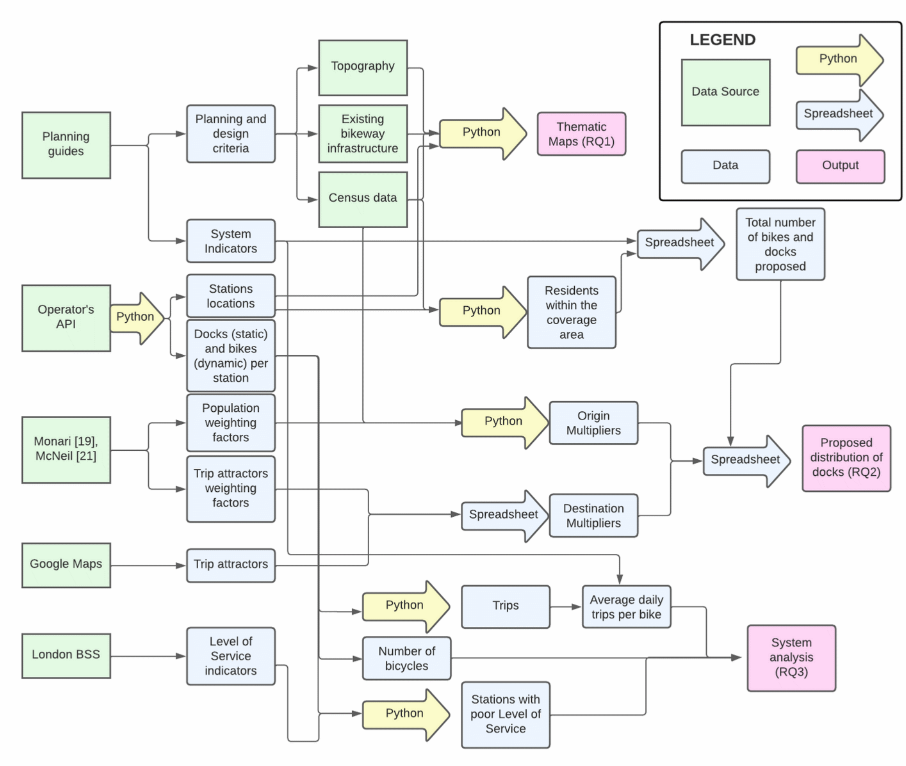

Fig. 1. Summary of the procedures developed using public data to assess strategies adopted by bicycle-sharing system operators.

To identify a demand profile for the studied city, we mixed the methods described by The Bikeshare Planning Guide [2] and Monari [19]. First, we calculated the population potentially served by the system, overlaying the census grid [13] with 300-meter bands around the stations. Next, we combined that potential demand with performance metrics from the planning guide [2] to find indicator values that match the actual number of docks. We used the upper and lower limits presented as Scenarios A and B in Table 1 to search for intermediate values that result in an estimate close to the actual total number of docks.

Table 1. Indicators selected to analyze the system (adapted from The Bikeshare Planning Guide [2]).

Indicators	Scenario A	Scenario B
Average daily trips per resident	0.05	0.025
Average daily trips per bike	4	8
Docks per bike	2.5	2

To find the number of docks per station, we used the trip generation approach proposed by Monari [19], who used multipliers to estimate trip productions and attractions. The trip production estimates (Equations 1 and 2) were based on the population distribution weighted by a factor that combines income and age, as summarized in Table 2.

$$q_i = \sum_{k=1}^{12} y_k \times p_{i,k} \tag{1}$$

$$M_i = \frac{q_i}{\sum_{i \in O} q_i} \tag{1}$$

Where:
q_i is the potential cycling demand at origin i;
y_k is the weighting factor for age-income combination k;
$p_{i,k}$ is the population belonging to age-income combination k at origin i;
M_i is the multiplier for origin i.

Table 2. Population weighting factors (y_k) based on combinations of age and income level (k), with values in %.

Income level (in the number of times the minimum wage)	Age (in years)			
	10-29	30-49	50-69	≥ 70
Less than 2	27.5 (1)	20.5 (4)	13.6 (7)	2.2 (10)
From 2 to 5	6.7 (2)	5 (5)	3.3 (8)	0.5 (11)
More than 5	8.9 (3)	6.7 (6)	4.4 (9)	0.7 (12)

Source: Monari [19]

The multipliers for each destination (Equations 3 and 4) were based on identified trip attractors around the stations, using Google Maps [20], weighted by the factors presented in Table 3.

$$a_j = \sum_{l=1}^{16} y_l \times u_{j,l} \tag{1}$$

$$N_j = \frac{a_j}{\sum_{j \in l} a_j} \quad (1)$$

Where:
a_j is the cycling attractiveness of destination j;
y_l is the weighting factor for trip attractor l;
$u_{j,l}$ is the number of trip attractors l at destination j;
N_j is the multiplier for destination j.

Table 3. Weighting factors (y_l) for trip attractors (l), with values in %.

General classification	Trip attractor (l)	yl
Industry/Factory	Any industry or factory (1)	20
Educational institutions	Child care (2)	2.5
	Preschools (3)	2.5
	Elementary schools (4)	5
	Middle schools (5)	5
	University (6)	5
Leisure	Fitness locations/General entertainment (7)	10
	Parks and open public spaces (8)	10
Retail	Specialty grocery stores (9)	2.5
	Beauty salons, barbers, etc. (10)	2.5
	Clothing stores (11)	5
	Restaurants, cafes and snacks (12)	5
	Supermarkets (13)	5
Other	Services in general (banks, post office) (14)	5
	Religious organizations (15)	5
	Emergency services, hospitals (16)	10
Total		100

Sources: Monari [19] and McNeil [21]

The data regarding the system's usage was gathered using the operator's API, which shows real-time station occupation. Using Python, we collected the data every 30 seconds from 3:27 a.m. to 12:23 p.m. on September 30th, 2022. We started the data collection in the early hours for capturing how the system was set up before the first users arrive. However, not all records are users' interactions with the system. To avoid empty stations throughout the day, operators perform rebalancing operations, sometimes transferring large numbers of bicycles between stations. As these operations cannot be explicitly identified using the API data as separate interactions at the stations [22], we plotted the data to identify the rebalancing operations and filter them out of actual user trips by analyzing unusual patterns, as shown in Figure 2.

At Station 18, for example, a big number of interactions was observed in a very short period around 6:40 a.m., when the station had a poor level of service (i.e., it had no bicycles). In addition, those interactions followed an inverse trend from previous records, with bicycle returns opposed to pickups. We assumed that the counterintuitive sequence of interactions is a result of a rebalancing operation and should be filtered out from user trips. A similar analysis can be made for Station 2, with opposite operations (see also Figure 2).

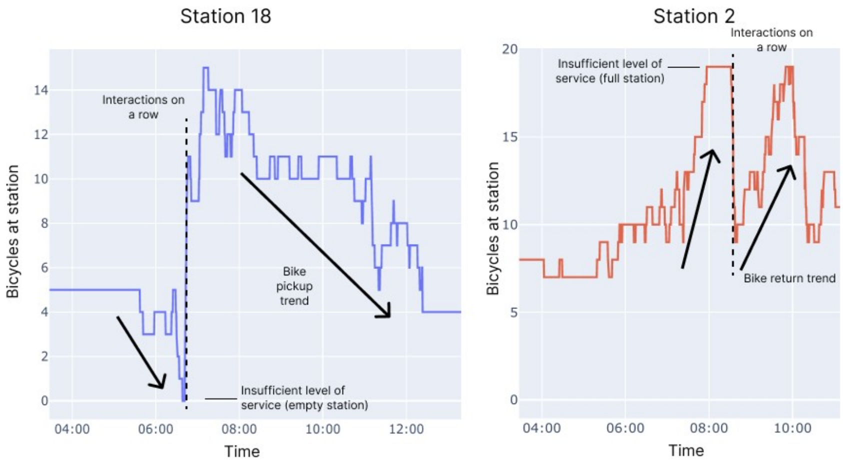

Fig. 2. Variation in the number of bicycles at stations 18 and 2 of the BSS in Vila Velha, ES, Brazil, on September 30th, 2022, according to API data. Comments highlight criteria used to identify rebalancing operations.

The trip data was used to calculate the number of trips per dock and per bike. For the level of service indicators, we chose the ones used by the London system, as shown in Table 4. As the system is small, with only 20 stations, and there is no indication of priority among them, all of the stations were considered as high priority.

Table 4. Redistribution Indicators (based on The Bikeshare Planning Guide [2]).

Time	Percentage of time that stations are empty (in %)	
	High-priority stations	Low-priority stations
Peak hours (7-10 a.m. and 4-7 p.m.)	6	23
Off-peak hours	3	8

Results

The maps in Figure 3 show how the locations of the stations in the BSS are compatible with the characteristics of the town (RQ1). Nearly half of the stations (from 1 to 8) are located on the coastal part of the city, taking advantage of areas with high values of population density, relatively flat terrain, and existing bikeways, including one along the seafront. This can be an indication of the relevance of leisure trips in the planning process, particularly in the case of stations 5 to 8.

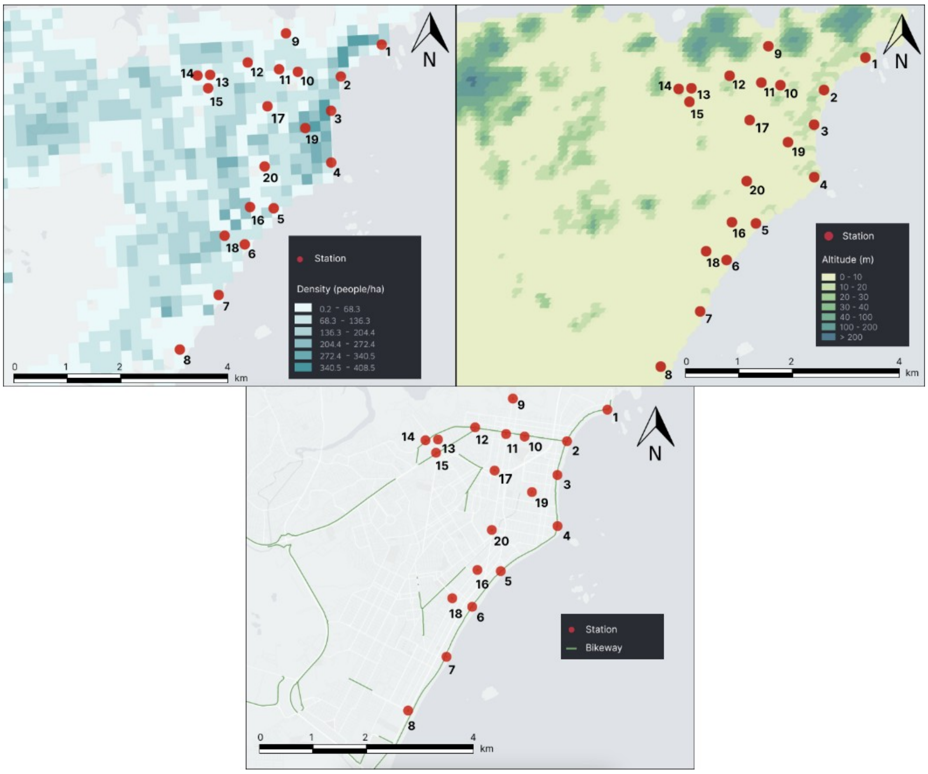

Fig. 3. Maps showing details of the bicycle-sharing system in Vila Velha, ES, Brazil. Upper left: distribution of population density in 2010 [13]. Upper right: topography [14]. Lower: distribution of the cycling infrastructure (i.e., bikeways) in 2022 [15].

Considering the population density values within the 300-meter bands around the stations, the system could potentially serve 43,468 people. We used this value, along with the indicator values presented in Table 1, to calculate bicycles and docks matching the actual city values. The results are shown in the lower part of Table 5.

The search for trip attractors resulted in areas such as the seafront with prevalence of touristic activities and residential buildings, and regions with intense commercial activities, such as the city's commercial zone, where stations 13, 14 and 15 are located, and a shopping mall next to station 20 (Figure 4). Using this data, the multipliers for origins and destinations were calculated, and the total number of docks found (see Table 5) was divided among the stations. Regarding RQ2, the comparison of these capacities, with the actual number of docks installed at each station, has shown that seafront stations (i.e., stations 1 to 8), had more docks actually in place than our estimates, while stations in commercial regions have fewer (Figure 5). The same pattern was also observed when we analyzed the average number of trips per dock (Figure 6).

Table 5. Estimates of the number of bicycles and docks for the BSS in Vila Velha based on the indicators suggested by The Bikeshare Planning Guide [2] - see also Table 1.

Indicators	Scenario A	Best matching values for the actual scenario	Scenario B
Average daily trips per resident (I)	0.05	0.025	0.025
Average daily trips per bike (II)	4	6	8
Docks per bike (III)	2.5	2.2	2
Estimates			
Residents (Census data) (IV)	43468	43468	43468
Trips/day (V = I * IV)	2173.4	1086.7	1086.7
Bicycles (VI = V / II)	543.4	181.1	135.8
Docks (VII = VI * III)	1358.4	398.5	271.7

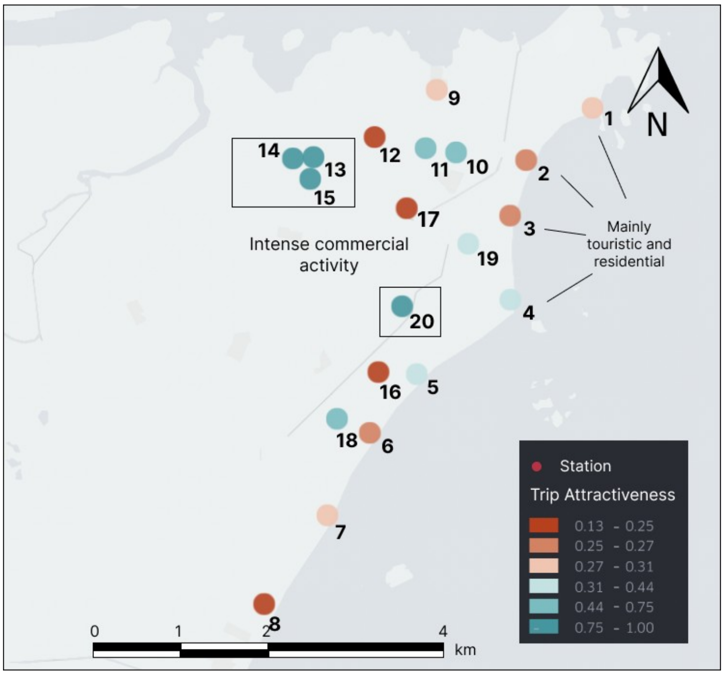

Fig. 4. Map showing the BSS in Vila Velha including the trip attractiveness of each station.

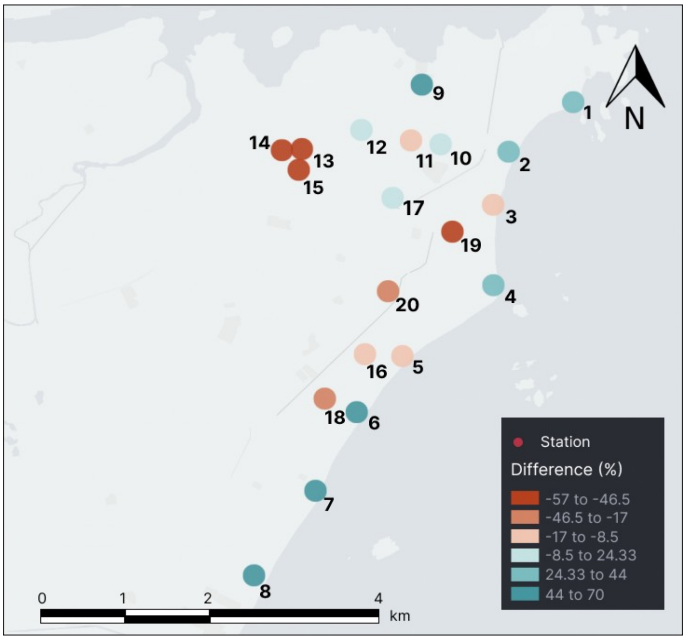

Fig. 5. Map showing the BSS in Vila Velha including the differences between the number of docks we estimated and those actually in operation.

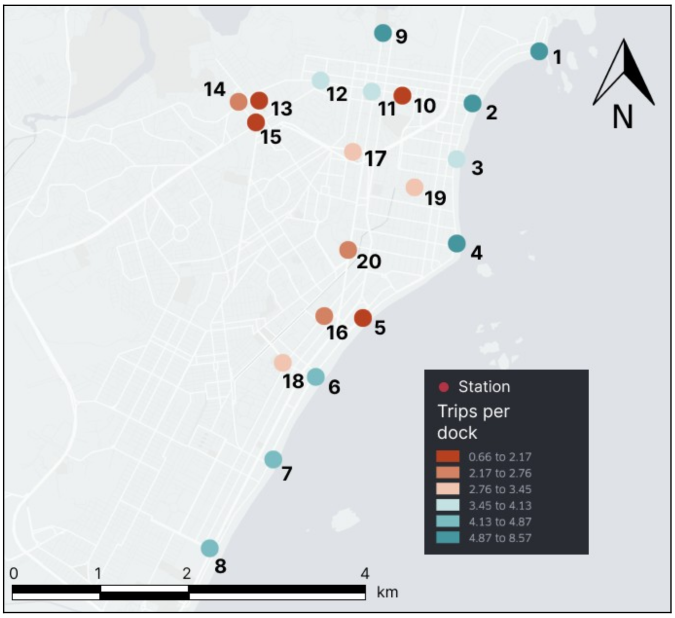

Fig. 6. Map showing the BSS in Vila Velha including the average number of trips per dock in each station, considering the proposed dock distribution.

Regarding the system's level of service (RQ3), the API data showed that the percentage of time the stations were empty was larger than the reference levels (as in Table 4) for 9 of the 20 stations. Although the operator does not inform the actual number of bicycles in the system on a regular basis, API data in the interval between 4 a.m. and 5 a.m. - during which the number of trips was only 4, ensuring most bicycles are docked - showed 159 available bicycles. This is less than 200, which was the latest figure released by the operator to the public [23]. This can be an indication that bicycles with problems are not being promptly fixed or replaced. Instead, they are being taken out of circulation, which reduces the number of bicycles available at stations and, therefore, negatively affects the level of service.

Besides that, the indicator 'Average daily trips per bike' for the morning period was close to 4. Therefore, considering an entire day, the value is likely to exceed 8, which is the recommended maximum (see Table 1). Such an intensive use of bicycles obviously requires additional maintenance measures to avoid reductions in the operational fleet having negative impacts on the level of service.

Conclusions

In general, we believe we developed a reasonable strategy to assess some of the operators' unrevealed strategies by means of public data. At the beginning, we were able to identify some of the main factors affecting the overall planning and design of the studied BSS. As expected, the distribution of population and activities, along with the topography and the existence of bikeway infrastructure, seem to match the criteria used by the operator to locate stations and bicycles. Furthermore, the procedure used to create a demand profile, which was based on the production and attraction of trips, also produced reasonable estimates for utilitarian trips, even though it apparently underestimated leisure trips.

More importantly, regarding the system's level of service, our approach was able to show that various stations have problems, mainly because of an insufficient number of bicycles. This is an important outcome of our study because it can not only stimulate discussions to improve the system with the operator but also guide them to critical aspects. Among the limitations of our approach, we believe that the most important one is the absence of a validation procedure. This is a direct consequence of the fact that we were not able to follow the individual movements of bicycles and users, even though the information is available to the operator. We believe that even a sample of that dataset could substantially improve our estimates, generating valuable information not only for the users but also for the operator.

References

1. DeMaio, P.: Bike-sharing: history, impacts, models of provision, and future. Journal of Public Transportation 12(4), 41-56 (2009). https://doi.org/10.5038/2375-0901.12.4.3.

2. ITDP - Institute for Transportation & Development Policy: The Bikeshare Planning Guide. ITDP, Rio de Janeiro (2018). https://itdpbrasil.org/guia-de-sistemas-de-bicicletas-compartilhadas/
3. Si, H., Shi, J., Wu, G., Chen, J.: Mapping the bike sharing research published from 2010 to 2018: A scientometric review. Journal of Cleaner Production, v. 213, p. 415-427 (2019). https://doi.org/10.1016/j.jclepro.2018.12.157.
4. Mix, R.; Hurtubia, R., Raveau, S.: Optimal location of bike-sharing stations: A built environment and accessibility approach. Transportation Research Part A, v. 160, p. 126-142 (2022). https://doi.org/10.1016/j.tra.2022.03.022.
5. Bernstein, A.: Survey: 64% of Citibike users unhappy about full or empty docks. WNYC Transportation Nation, 2013. https://www.wnyc.org/story/survey-64-citibike-users-unhappy-about-full-and-empty-docks/, last accessed 2022/05/23.
6. Frade, I., Ribeiro, A.: Bike-sharing stations: A maximal covering location approach. Transportation Research Part A: Policy and Practice 82: 216-227 (2015). https://doi.org/10.1016/j.tra.2015.09.014.
7. Lin, J. J., Lin, C. T., Feng, C. M.: Locating rental stations and bikeways in a public bike system. Transportation Planning and Technology 41(4): 402-420 (2018). https://doi.org/10.1080/03081060.2018.1453915.
8. Banerjee, S., Kabir, M., Khadem N. K., Chavis, C.: Optimal locations for bikeshare stations: A new GIS based spatial approach. Transportation Research Interdisciplinary Perspectives 4: 100101 (2020). https://doi.org/10.1016/j.trip.2020.100101.
9. Bahadori, M. S., Gonçalves, A. B., Moura, F.: A systematic review of station location techniques for bicycle-sharing systems planning and operation. International Journal of Geo-Information 10(8): 554 (2021). https://doi.org/10.3390/ijgi10080554.
10. Fazio, M., Giuffrida, N., Le Pira, M., Inturri, G., Ignaccolo, M.: Bike oriented development: Selecting locations for cycle stations through a spatial approach. Research in Transportation Business and Management 40(6): 100576 (2021). https://doi.org/10.1016/j.rtbm.2020.100576.
11. Guler, D., Yomralioglu, T.: Location evaluation of bicycle sharing system stations and cycling infrastructures with best worst method using GIS. The Professional Geographer 73(3): 535-552 (2021). https://doi.org/10.1080/00330124.2021.1883446.
12. Caggiani, L., Camporeale, R., Hamidi, Z., Zhao, C.: Evaluating the efficiency of bike-sharing stations with Data Envelopment Analysis. Sustainability 13(2), 881 (2021). https://doi.org/10.3390/su13020881.
13. The Brazilian Institute of Geography and Statistics: 2010 Brazilian Census. IBGE, Rio de Janeiro (2011).
14. The National Institute for Space Research. TOPODATA. INPE, São José dos Campos. http://www.dsr.inpe.br/topodata/index.php, last accessed 2022/09/22.
15. Open Street Map. www.openstreetmap.org, last accessed 2022/11/17.
16. Bike Share Research. Open Bike Share Data. https://bikeshare-research.org/#bssid:vila_velha, last accessed 2022/10/06.
17. He, S., & et al. 2021. kepler.gl: A powerful open source geospatial analysis tool for large-scale data sets. GitHub. https://github.com/keplergl/kepler.gl, last accessed 2023/01/09.
18. Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, … François Leblanc. (2020, July 15). geopandas/geopandas: v0.8.1 (Version v0.8.1). Zenodo. http://doi.org/10.5281/zenodo.3946761, last accessed 2023/01/09.
19. Monari, M.: Bicycle level of traffic stress and open data geoprocessing combined for the definition of cycling networks in small-sized cities. Ph.D. thesis, São Carlos School of En-

gineering, University of São Paulo, São Carlos, Brazil (2022). doi:10.11606/T.18.2022.tde-05092022-145612, last accessed 2023/01/09.
20. Google. Vila Velha - Google Maps. 2022. https://www.google.com/maps/place/Vila+Velha+-+ES/@-20.4273698,-40.5069309,47867m/data=!3m2!1e3!4b1!4m5!3m4!1s0xb81635625b6987:0x7bc9ac2d2fe35a1c!8m2!3d-20.3478055!4d-40.294933, last accessed 2022/09/22.
21. McNeil, N. (2011). Bikeability and the 20-min neighborhood: How infrastructure and destinations influence bicycle accessibility. Transportation Research Record, 2247(1), 53–63. https://doi.org/10.3141/2247-07, last accessed 2022/09/22.
22. Chardon, C. M. de., Caruso, G.: Estimating bike-share trips using station level data. Transportation Research Part B: Methodological 78: 260-279 (2015). https://doi.org/10.1016/j.trb.2015.05.003.
23. Bike VV. Página Inicial | Bike Vila Velha - Tembici. 2022. https://www.bikevv.com.br, last accessed 2022/10/06.