DOI: 10.21120/LE/18/1/3

Acta Geographica Debrecina
Landscape & Environment series 18 (1) 2024. 28-35

DEVELOPING A STRATEGY OF DATA COLLECTION AND PRE-PROCESSING TO ASSESS BIKE SHARING SYSTEM STATION PLACEMENTS WITH THE HELP OF GIS

TAMÁS MIHÁLY VERES

Doctoral School of Informatics, Faculty of Informatics, University of Debrecen, Hungary
Email: veres.tamas@inf.unideb.hu

Received 13 January 2024 accepted in revised form 28 February 2024

Abstract
Our research presents a methodological framework for analyzing bicycle-sharing systems, using the self-service bike operations of JCDecaux in Toulouse as a case study. The objective was to identify a method for obtaining a cleansed and structured attribute list that could be useful in evaluating and optimizing the placement of bicycle rental docks. Utilizing open data, our approach involves developing a Python script within QGIS to create new layers around each of the 288 studied bicycle rental stations, based on a selected 100-meter buffer. This buffer size is chosen to reduce data overlap in dense urban settings. The script is designed to collect urban features within these buffers that register as multipolygons (mainly buildings) or points (amenities, transportation features), moreover it applies categorization of data, such as identifying and marking the different building types. The method includes a basic visualization of potential data in QGIS using OpenStreetMap.

Keywords: Bank Erosion, Sediment Deposition, Channel Shifting, Jiadhal River, Remote Sensing

1. Introduction

Bicycle-sharing systems are a convenient mode of transportation found in most major European cities.

The establishment and optimization of bicycle-sharing systems contribute significantly to urban sustainability and pollution reduction. By strategically placing bike-sharing stations, these systems encourage a shift from short car trips to cycling, directly reducing vehicle emissions, a dominant source of urban air pollution. This change does not only reduce traffic jams but also supports environmental goals by encouraging eco-friendly transportation. The success of such initiatives would be even more significant in major cities (Park et al., 2022).

Park and Sohn conducted a study in Seoul, discovering that similar strategies could lead to substantial environmental and urban mobility improvements. Being relevant to our study, it is also important to understand the data collection method of Park and Sohn. The authors utilized taxi trajectory data as a proxy for identifying potential high-demand areas for bicycle stations. They chose taxi data for its availability and similarities to individual travel patterns, despite some differences from private vehicle data. The researchers extracted pick-up locations from taxi trajectories to reflect typical bicycle riding patterns. This approach helped them identify potential high-demand sites for bicycle-sharing stations, contributing to the research's goal of efficiently reducing short automobile trips (Park – Sohn, 2017).

Identifying high-demand areas is crucial when optimizing bicycle-sharing station placement. High-demand areas are typically locations where a significant number of short trips originate or terminate, such as business districts, public transportation hubs, shopping areas, or educational institutions. Placing bike-sharing stations in these areas ensures they are conveniently accessible to a larger number of potential users, thereby encouraging the use of bicycles over cars for short trips (Torrisi et al., 2021, García-Palomares et al., 2012).

An essential factor in planning and establishing such a transportation network is the location of docking stations. It is important to maximize their usage, which can be greatly supported by a well-executed, strategic placement (Eren – Uz, 2019, Zochowska et al. 2021). Major bike rental operators, like JCDecaux, offer extensive, openly accessible data on their stations; however, we could not identify any official rule or guideline published about the planning process of their bicycle-network station arrangement. The more direct and indirect factors we take into consideration, the better utilization we can achieve; therefore, it can be an important pre-measure to identify those entities and their properties, that can be used to further optimize the emplacement of these rental stations (Kull et al., 2015).

Wei and Zhu noted that the use of bike sharing networks is affected not just by the physical environment, such as land use, urban design, and transportation systems but also by the social economy. This emphasizes the importance of identifying these elements and using them to optimize these systems. Moreover, their paper presents a comprehensive collection of various studies that are using governmental or enterprise-sourced open data to validate their findings. This, to an extent, confirms the viability and credibility of this (commonly available) data source (Wei – Zhu, 2023).

Additionally, when researching any system that has a territorial or environmental relationship, GIS-based research provides an effective approach for both visualizing and analyzing the collected data. As pointed out by Hervai et al., even complex data structures, like the spatial and temporal variations of soil moisture, can be a fitting subject for this kind of research toolkit (Hervai et al., 2017).

Non-material information can also be visualized, thus underlining the capabilities of GIS systems. Digital forensics, when analyzed through geographic visualization, shows another great example of the usability and effect of implementing such methods. Furthermore, this finding further emphasizes the importance of patterns and relationships that might be hidden in raw data (Zichar, 2016).

2. Materials and Methods

There are multiple factors to take into consideration when analyzing the placement of common bicycle sharing docks. Since our aim was to develop a strategy that can help in understanding (and potentially replicating) the general placement of such stations, we based our study on the following aspects:

JCDecaux has a significant presence in urban mobility, complementing public transportation networks with its self-service bikes (SSB). Today, it operates over 25,000 bikes in 73 cities across 10 countries, of which France can be singled out, where the company has a particularly strong presence. Their business model and commitment to eco-friendly mobility align with the European Commission's "Green Taxonomy", emphasizing low-carbon transition activities.

Toulouse, known for its vibrant culture and as a hub for aerospace innovation, also features a significant bike-sharing system managed by JCDecaux. The company reported a 2% increase in bicycle rentals in the city in 2022 compared to 2019, and a 14% increase compared to 2021 (JCDecaux, 2023).

Toulouse also came to our attention because of the seemingly random placement of bicycle rental docks. This may refer to the hypothesis that there can be a specific logic behind their placements, making the network a good base to develop our analytical approach on. Additionally, the city boasts an expanded public transportation network, including trams, buses, and a well-developed metro system, which might be useful in the future for more extensive research or comparison.

(Real-time) Data is available through JCDecaux's developer API: Stations with static as well as dynamic data such as coordinates and contract (city) name. This includes several key attributes that make it possible to analyze station usage patterns, such as identifying the most frequently used stations:

- Status (status, OPEN/CLOSED): Indicates operational status, helping identify active stations.
- Connection Status (connected): Shows if the station is connected to the central system, reflecting its operational availability.
- Total Stands (totalStands): Gives the capacity of the station, useful for understanding its size and potential usage.
- Main and Overflow Stands (mainStands, overflowStands): Provide detailed capacity information, indicating how many bikes can be accommodated.
- Availabilities (availabilities): Indicates the number of spaces available and the number of bikes by type of bike.
- Last Update Timestamp (lastUpdate): Indicating when the last update was made.

These data points collectively could help us identify usage patterns as well as trends in availability (JCDecaux developer, 2023).

Python is widely used in geoinformatics, mainly for its versatility in handling spatial data, enabling tasks such as data analysis, visualization, and automation. It supports a range of geospatial libraries and tools, such as PyQGIS (Zichar, 2014). The method also involves a Python script developed for use within the QGIS environment. The primary objective of the script is to process spatial data by creating new layers based on specific geographic criteria. The script operates on two layers: A point layer acting as a fixed layer, containing precise locations (points) for all 288 bicycle rental stations operated by JCDecaux in the city. A data layer, containing either point or multipolygon features. Multipolygons mainly represent the buildings in Toulouse. Features of this geometry are particularly useful for understanding the layout and distribution of urban infrastructure around the rental stations. By analyzing these building features, we can gain insights into the urban density or the types of buildings (such as residential, commercial, or public facilities). Points on the other hand can represent a variety of relevant urban elements, such as public amenities like parks or tourist attractions. It can also mark transportation-related features like bus stops or parking areas, as well as commercial entities such as shops, cafés, or restaurants.

The procedure can be structured as follows:

1. Buffer Creation Around Points: For each selected feature in the point layer, the script creates a circular buffer zone. The radius of this buffer is set to 100 meters.
1. Feature Collection Within Buffer: The script then identifies and collects all features from the data layer that intersect with each buffer.
1. New Layer Generation: For each buffer zone, the script creates a new layer. This layer is populated with the features from the data layer that fall within the corresponding buffer zone. These new layers inherit the Coordinate Reference Systems (CRS) of the data layer and retain all attribute fields and values from the original data layer features.

1. Layer Addition to QGIS Project: Each new layer is added to the current QGIS project, allowing for immediate visualization and further analysis within the QGIS interface.

This script provides an automated solution to our data processing, particularly useful for proximity-based analysis. The choice of a 100-meter radius is a decision that helps reduce the noise in the data around each point, providing more precise and relevant insights for urban planning. Using a larger radius, such as 300 meters, could result in significant overlap of buffers, especially in areas with high densities of stations. This overlap can lead to overestimating the availability of features (study attributes) in an area that we want to narrow down as precisely as possible. The newly created layers around the rental station points can be further analyzed by grouping data based on specific attributes like 'amenity' or 'building type'. This grouping, which can also be visually distinguished through color coding, allows for more advanced exploration.

All the analyzed data, including the buffered areas and the categorized urban features can be effectively visualized within QGIS using OpenStreetMap (OSM) as a base map. By overlaying our data layers on top of OSM, we can gain a clearer visual understanding of how the script classifies our data (Fig.1, Fig.2).

Figure 1 shows the categorized multi-polygons within the 100m radius of Station No. 25. This station has 3 types of multipolygons registered as buildings in its proximity: Apartments, Retail and common buildings (Fig.1).

Similarly, Figure 2 shows the categorized points within the 100m radius of the same station, having quite various values including bank, bar, café, dentist, fast_food, motorcycle_parking, nightclub, pharmacy, post_office, pub, restaurant (Fig.2).

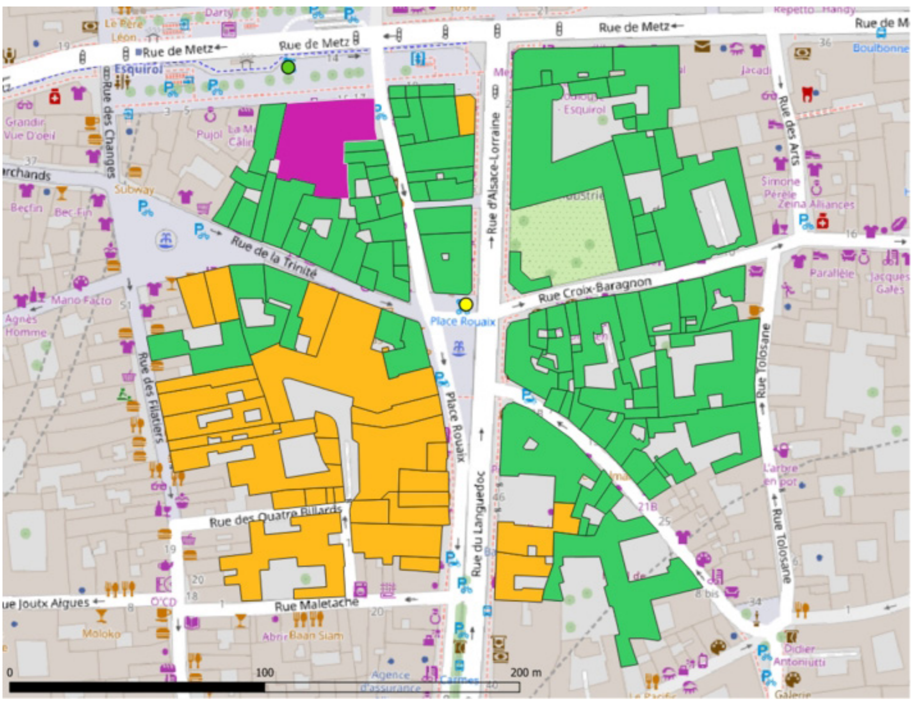

Fig. 1. Multipolygons data visualized using OpenStreetMap inside QGIS

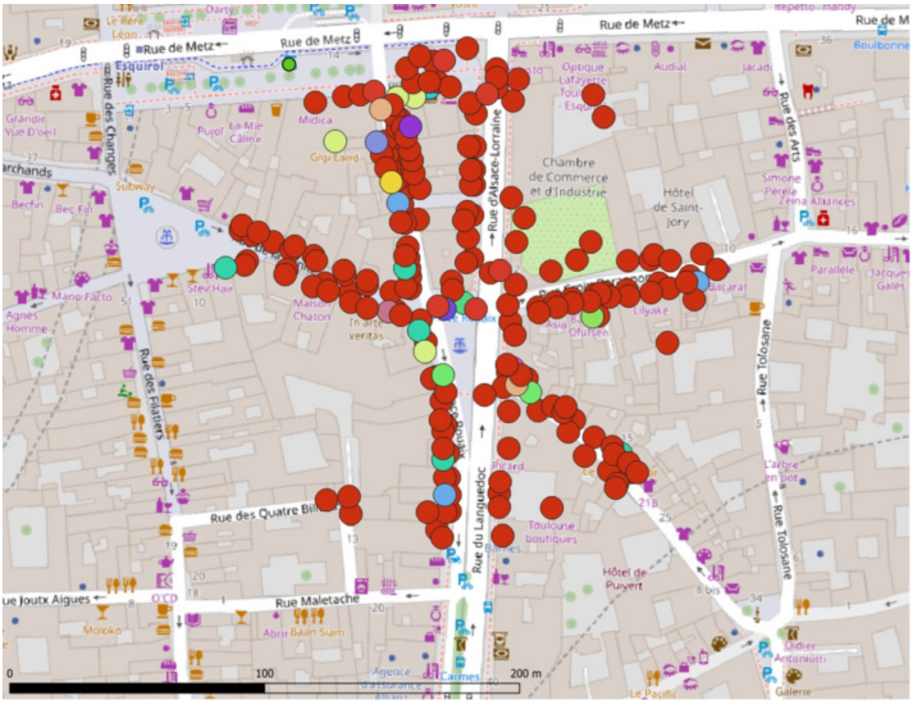

Fig. 2. Points data visualized using OpenStreetMap inside QGIS

3. Results

Based on the discovered data and data gathering methods, the assessment of bike-sharing stations can be summarized in the following steps:

1. Formulating the Hypothesis: Beginning with the assumption that the placement of SSB docking stations in a city can be logically assessed with the help of GIS- as well as open-sourced data.
2. Data collection and processing: Gathering real-time data about bicycle rental station usage in the city. If real-time data is not available, looking for reliable statistics to understand the usage patterns and peculiarities of each station. Gathering data within QGIS to identify a limited set of attributes (about 5-10) that can effectively indicate station usage and consequently, validate the placement of each station.
3. Application and Comparison in a Different City: Finding a city with similar urban characteristics and an SSB network. Apply the identified attributes from Step 2 to propose a hypothetical distribution of bike stations in this new city. Compare this proposed distribution with the existing one in the city to evaluate the effectiveness of the identified attributes in predicting logical station placement.

4. Discussion

In the current study, we focused on using only the static data coming from JCDecaux, which details the locations of bicycle rental stations in Toulouse. This decision was made to concentrate first on understanding the fixed aspects of this kind of network. However, to fully validate the earlier mentioned hypothesis assuming a certain logic behind SSB dock placements, real-time data needs to be included in the framework of future work. This will allow us to check if the attributes identified by our script align with actual usage patterns observed in real-time. By doing so, we can better evaluate the effectiveness of the bike-sharing system and understand more about user behavior.

One important aspect that can be also identified throughout the analysis is that working on an optimization method that will be used to maximize the usage of shared means of transportation –with slight modifications–, could be applied to other businesses, especially those that involve resource allocation, demand prediction, and spatial distribution. This makes this topic a viable base for future research. Potential applications include:

- Carsharing and Taxi Services: Optimizing pick-up and drop-off points based on demand patterns.
- Logistics and Delivery Companies: Efficiently locating distribution centers or parcel lockers in high-demand areas.
- Retail Businesses: Identifying strategic locations for new stores or service centers.
- Public Transport Systems: Planning bus routes and stops according to commuter flow.
- Urban Planning: Developing public amenities and services like parks, libraries, or community centers based on population density and usage patterns.

Zafar et al. similarly prepared a GIS-based approach to plan the location framework for EV charging stations. They formulated the location problem as a Maximum Coverage Location Problem (MCLP), which considers existing locations of petrol and fuel stations as potential sites for electric vehicle charging stations (Zafar et al., 2021).

As shown by Diethe et al., analyzing Self-Service Bike (SSB) networks and their data also presents an excellent opportunity to incorporate various predictive and, potentially machine learning techniques. The nature of the data structure in these networks makes them particularly suitable for this kind of analysis. Forecasting of bike availability, usage patterns, and station occupancy all fit within this category, which is also crucial for optimizing the efficiency and user experience of bike-sharing systems (Diethe et al., 2015).

While spatial analysis and open data provide valuable insights into the physical layout and usage patterns of bicycle rental systems, based on the location, they might need to be complemented with additional environmental data to fully understand the dynamics of bike rental usage. Incorporating weather, seasonal changes, and other environmental factors when studying urban cycling might have a significant influence on the results (Saneinejad et al., 2012, Sears et al., 2012). In our case, this is a limitation, as the toolset we use does not include these natural variables.

In the context of our research focusing on Toulouse's seemingly random bike rental dock placement, it is important to note that in some cities (for example in Valencia), these stations are arranged more symmetrically (Valencia City Council, Open Data, 2024). This might suggest a simpler underlying logic, like maintaining equal distances between stations. This implies that our analytical approach, designed to discover and optimize complex placement patterns, may be less applicable or relevant in these cities, where station placement follows a more grid-like distribution strategy.

To summarize, our approach represents a quite simplified method for validating the placement of bicycle rental stations, excluding several factors like natural impacts and socio-economic influence. However, as several other studies cited in this paper pointed out, the use of open data as well as geospatial information is still able to contribute positively to research on similar topics. This approach, while having its limitations, can underscore the importance of leveraging diverse data sources in urban planning.

5. Conclusion

Our study presents a simplified method for validating the placement of bicycle rental stations, focusing on Toulouse's network. We used static data from JCDecaux, acknowledging the need to incorporate real-time data in future research for a comprehensive evaluation. The potential of our approach extends beyond bicycle-sharing to other areas like carsharing and urban planning. However, our method excludes factors like environmental and socio-economic impacts, a limitation for a complete understanding. The study underscores the value of open and geospatial data in urban planning, despite these limitations, and sets a foundation for future research to optimize urban mobility.

6. References

Park J et al. (2022): Air Pollution and Public Bike-Sharing System Ridership in the Context of Sustainable Development Goals. Sustainability. 14(7):3861. https://doi.org/10.3390/su14073861

Wei, B, - Lei Z. (2023): Exploring the Impact of Built Environment Factors on the Relationships between Bike Sharing and Public Transportation: A Case Study of New York" ISPRS International Journal of Geo-Information 12, no. 7: 293. https://doi.org/10.3390/ijgi12070293

Hervai, A. et al. (2017): Interpolation and 3D visualization of soil moisture, Landscape & Environment 11 (1) 2017. 23-34, DOI: 10.21120/LE/11/1/3

Zichar, M. (2016) Guidelines for cost-effective geovisualization in digital forensics, Landscape & Environment 10 (3-4) 2016. 117-122, DOI: 10.21120/LE/10/3-4/3

Torrisi, V. et al. (2021): Exploring the factors affecting bike-sharing demand: evidence from student perceptions, usage patterns and adoption barriers, Transportation Research Procedia, Volume 52, Pages 573-580, https://doi.org/10.1016/j.trpro.2021.01.068.

García-Palomares, J. C. et al. (2012) Optimizing the location of stations in bike-sharing programs: A GIS approach, Applied Geography, Volume 35, Issues 1–2, Pages 235-246, https://doi.org/10.1016/j.apgeog.2012.07.002.

Eren, E, - Uz VE, (2019): A Review on Bike-Sharing: The Factors Affecting Bike-Sharing Demand, Sustainable Cities and Society, doi: https://doi.org/10.1016/j.scs.2019.101882

Zochowska, R. et al. (2021): A GIS-Based Method of the Assessment of Spatial Integration of Bike-Sharing Stations Sustainability 13, no. 7: 3894. https://doi.org/10.3390/su13073894

Kull, M., et al. (2015): Bike rental and weather data across dozens of cities. In ICML 2015 Workshop on Demand Forecasting (pp. 687-694).

Park, C. – Sohn S. Y. (2017): An optimization approach for the placement of bicycle-sharing stations to reduce short car trips: An application to the city of Seoul, Transportation Research Part A: Policy and Practice, Volume 105, Pages 154-166, https://doi.org/10.1016/j.tra.2017.08.019.

Zichar, M., (2014): Python in geoinformatics, In: Az elmélet és a gyakorlat találkozása a térinformatikában V.: Térinformatikai konferencia és szakkiállítás / Balázs Boglárka (szerk.), Debreceni Egyetemi Kiadó, Debrecen, 429-433, 2014. ISBN: 9789633184349

Zafar, U. et al. (2021): A GIS-based optimal facility location framework for fast electric vehicle charging stations. In: The 30th International Symposium on Industrial Electronics, 2021-06-23 - 2021-06-26. (In Press)

Diethe, T. et al (2015): Gaussian Process Model Re-Use.

Record-breaking +21% global increase in self-service bike rentals operated by JCDecaux, Paris, https://www.jcdecaux.com/press-releases/record-breaking-21-global-increase-self-service-bike-rentals-operated-jcdecaux, 18.01.2023, last accessed: 02.01.2024.

JCDecaux developer, Real-time data, Station and Parks resources, https://developer.jcdecaux.com/#/opendata/vls?page=dynamic&contract=toulouse, last accessed: 02.01.2024.

Saneinejad, S. et al. (2012): Modelling the impact of weather conditions on active transportation travel behaviour, Transportation Research Part D: Transport and Environment, Volume 17, Issue 2, Pages 129-137, https://doi.org/10.1016/j.trd.2011.09.005.

Sears, J., Flynn, B. S., Aultman-Hall, L., & Dana, G. S. (2012). To Bike or Not to Bike. Transportation Research Record: Journal of the Transportation Research Board, 2314(1), 105–111. doi:10.3141/2314-14

Valencia City Council, Open data portal of València, https://valencia.opendatasoft.com/pages/home/, last accessed: 02.01.2024.

QGIS.org (2024). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.org

Python Software Foundation. Python Language Reference, version 3.12.1 Available at http://www.python.org