

An Inclusive and Justice Bike-sharing: What Socio-economic factors Influence the Usage of NYC Citi Bike?

Proposal

Jiaxuan Tang and Muwen (Wen) He

I. Introduction (100 words): A basic overview of the research topic and why it is important to study. What will you and others learn from this study? (10 pts)

Planning and promoting sustainable mobility is one of the solutions to many urban challenges, such as, but not limited to, global warming led by massive greenhouse gas emissions, resource and land taken due to urban sprawl, road congestion caused by overpopulation, and transportation inequality. Among the sustainable transit options, the affordable and flexible bicycle-sharing provides an additional option for users and is widely used globally (DeMaio, 2009). Citi Bike has grown rapidly in New York City in recent years and covers all of Manhattan, including Central Park, Western Queens, and Northwest Brooklyn. We hope to study Citi Bike usage data compared to other NYC social and economic data, analyze and visualize Citi Bike usage and hotspots, then further analyze factors that influence usage and build predictive models that help predict the usage in the future, and can be applied to other cities or regions.

II. Problem statement (200 words): A clear statement of your particular research question within the topic. How does your question address the perspectives and approaches that are missing in the study of this topic? Where are your site(s)? What is the hypothesis, based on your research question, you are studying? (20 pts)

The main objective of this paper is to model the factors that influence Citi Bike usage within the range of New York City, to examine the impact of multiple socioeconomic factors on bike-sharing usage in different parts of the city. Among the existing studies on bike-sharing (especially Citi Bike in New York City) systems, there are few such socioeconomic-based studies (Guo et al., 2022). We believe that this may be due to the fact that extensive and valid socioeconomic data are difficult to obtain in other cities; however, in New York City, we have access to a variety of socioeconomic data containing geographic information using numerous official sources, allowing our study to proceed (Kranish, 2021a, 2021b, 2021c, 2021d, 2021e, 2021e, 2021f, 2021g). We

expect that a variety of spatially heterogeneous factors, including population density, income structure, public transportation availability, motor vehicle ownership rates, and neighborhood accessibility, will influence people's propensity to use the bike-sharing system, in addition to the effects of weather factors and bicycle station density that have been extensively studied. Using a spatial statistical approach to model these factors, we expect to identify how these factors affect bike-sharing usage rates. Also, using visualization tools and machine learning methods, we can derive a system for predicting future Citi Bike usage in New York City, which can also predict the usage of potential bike-sharing systems in other cities based on local data.

III. Background (300 words): The current key debates and latest developments in the topic. This section should justify and support your problem statement. (30 pts)

The factors affecting bike-sharing usage patterns are complex and multifaceted. Most of the existing studies focus on (1) modeling bike-sharing usage based on individual influences such as weather, including using machine learning approaches (AitBihiOuali & Klingen, 2022; Hamad et al., 2021; Wang, 2016; Xin et al., 2023); (2) analyzing spatial characteristics of bike-sharing usage, including time and space lag effects (Faghih-Imani & Eluru, 2016); (3) discussion of how to locate the bike-sharing stations (Mix et al., 2022); (4) discussion of the complementarity of bike-sharing with other transportation modes (Cui et al., 2023); and (5) some other topics of the characteristics of the system itself (Zhang et al., 2022). Few socioeconomic and land use features are mentioned, although they are important for setting the location of bike-sharing stations, planning for community access to bike-sharing, and using bike-sharing to reach urban planning concepts, such as, the “last mile”, “15-minute cities”, that are essential to the development of a city, which are valuable to be considered during urban planning and design. These social characteristics are likely to have a significant impact on bicycle sharing, for example, studies have shown that the younger generation is more likely to support sustainability goals and change their lifestyles to do so (Yamane & Kaneko, 2021).

Secondly, much of the research in recent years has been influenced by the pandemic, for example, the shift in the purpose of trips from work to leisure due to the impact of working-from-home. Also, Citi Bike has expanded significantly in recent years to nearly double the number of base stations and has spread from lower Manhattan and a small part of Brooklyn to all of Manhattan and into Queens since 2016. We have more and more diverse and up-to-date usage data than other studies.

We also wanted to address the question of what we could do with the findings. While the Citi Bike in NYC is successful and typical as a case study, we wanted to offer help to other cities to replicate New York's success. Most of the other studies made very short textual recommendations based on the results, we, on the other hand, wish to use the results of the study to transfer them into a system that could be applied.

IV. Methodology plan (300 words): The data and analysis methods you plan to use. Detail all the data sources that you plan to use. If it is open data, include the links and provide a short explanation of how you plan to map this. If this is data you want to collect, detail a plan and timeline for this process. For the analysis, explain which methods from class you will use, why you are choosing these, and what you think they will reveal. What is the unit of analysis? Who or what specifically are you studying? (people, towns, plans, policies, etc.) What are the key concepts involved in your study? How will you define and operationalize the key concepts? What are the specific relationships that you hope to examine? (30 pts)

1. Data

The main data used in this research is the usage data of Citi Bike in New York City, which is available for download on the official Citi Bike website¹. The data covers the period from 2013 to the present, and includes order start and end times, start and end coordinates, and user profiles, etc. We also need to obtain the distribution of Citi Bike's stations.

In addition, we need to obtain socio-economic data for New York City, including but not limited to: (1) Population density; (2) Population structure; (3) Income structure; (4) Number of public transportation stops in the area/average distance to public transportation stops in the area; (5) Motor vehicle ownership rate; (6) Parking index (measures the difficulty of street parking in the area); (7) Neighborhood function (commercial, residential, etc., can be obtained through official data or POI analysis); (8) Topography.

These data are usually available from two sources: (1) public APIs from public map providers (e.g., Google Maps); and (2) open data portals from the City of New York and affiliated agencies (e.g., NYC Open Data² and more^{3,4}). For the latter, the information obtained will generally be

¹ Citi Bike System Data | Citi Bike NYC, <https://citibikenyc.com/system-data>.

² NYC Open Data, <https://opendata.cityofnewyork.us/>.

³ Zola | NYC's Zoning & Land Use Map, <https://zola.planning.nyc.gov/>.

⁴ U.S. Census Bureau, <https://www.census.gov/>.

generalize to a certain administrative level, which requires us to weigh the geospatial division in this study after collecting the information.

2. Structure

It is currently expected that the research in this paper will consist of two parts: visualization of the data, and modeling analysis.

The first part is the initial processing and visualization of the data, primarily to demonstrate the usage of Citi Bike and socioeconomic characteristics of New York City, as well as to prepare for the second part. In this part, we need to first clean the data and then select (or generate) quantifiable metrics (e.g., frequency, duration, distance, or a weighted average of the three, etc.) based on the Citi Bike data that can be used to measure its usage. We can also use some visualization tools (e.g., cluster analysis, etc.) and create human-machine interfaces that can be interacted with to help the reader better understand the geographic characteristics of the data.

The second part will be the modeling of socio-economic factors influencing Citi Bike usage. By collecting, cleaning and normalizing the socio-economic data listed above, we will be able to build a spatial statistical model to study the impact of different factors. Also, we must consider the possible spatial spillover effects in Citi Bike usage. Therefore, it is worth considering a regression model that includes spatial lags. The selection of a specific model needs to be decided based on a preliminary exploration of the data. In addition, an optional option is to use machine learning methods for this part of the analysis. Given our current lack of understanding of machine learning methods and the limited relevant research, we will determine when to use such methods after further study and discussion with our mentor.

At the end of Part II, we will argue that our model has predictive capabilities, including (1) predicting new Citi Bike demand when socio-economic characteristics change somewhere in New York City in the future, and (2) predicting potential demand distribution in U.S. cities that do not yet have a public bike system in operation. We will select one city as a case study for practice.

In addition, a further factor to consider is time. So far, all our models have been based on static (at a given moment in time) New York City socio-economic data, which requires that we also use static data when analyzing bike orders (e.g., the sum/average of orders over a certain period of time). However, dynamic orders also contain numerous hidden information that can be studied, such as: the difference in usage between different times of the day; the difference in usage between workdays and weekends, and the impact of COVID-19. To utilize this component would

require us to further refine the research question or expand it (e.g., examine the factors influencing Citi Bike usage during commuting time periods). This will be the next step we need to explore (while conducting the initial study) and may influence our outcome.

Reference

- AitBihiOuali, L., & Klingen, J. (2022). Inclusive roads in NYC: Gender differences in responses to cycling infrastructure. *Cities*, 127, 103719.
<https://doi.org/10.1016/j.cities.2022.103719>
- Cui, Y., Chen, X., Chen, X., & Zhang, C. (2023). Competition, Integration, or Complementation? Exploring Dock-Based Bike-Sharing in New York City. *The Professional Geographer*, 75(1), 65–75. <https://doi.org/10.1080/00330124.2022.2081224>
- DeMaio, P. (2009). Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*, 12(4), 41–56. <https://doi.org/10.5038/2375-0901.12.4.3>
- Faghih-Imani, A., & Eluru, N. (2016). Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography*, 54, 218–227. <https://doi.org/10.1016/j.jtrangeo.2016.06.008>
- Guo, Y., Yang, L., & Chen, Y. (2022). Bike Share Usage and the Built Environment: A Review. *Frontiers in Public Health*, 10, 848169. <https://doi.org/10.3389/fpubh.2022.848169>
- Hamad, S. Y. Y., Ma, T., & Antoniou, C. (2021). Analysis and Prediction of Bikes sharing Traffic Flow – Citi Bike, New York. *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1–8.
<https://doi.org/10.1109/MT-ITS49943.2021.9529290>
- Kranish, C. (2021a, March 22). *Use Weather History with NYC Bike Share Trip Data*. Medium.
<https://towardsdatascience.com/use-weather-history-with-nyc-bike-share-trip-data-51fcb89297d>
- Kranish, C. (2021b, March 29). *Reverse Geocoding with NYC Bike Share Data*. Medium.
<https://towardsdatascience.com/reverse-geocoding-with-nyc-bike-share-data-cdef427987f8>
- Kranish, C. (2021c, April 4). *Visualizing NYC Bike Share Trips with a Chord Diagram*. Medium.
<https://towardsdatascience.com/visualizing-nyc-bike-share-trips-with-a-chord-diagram-eb4c8e14366>

- Kranish, C. (2021d, April 8). *Exploring the Effects of the Pandemic on NYC Bike Share Usage*. Medium. <https://towardsdatascience.com/exploring-the-effects-of-the-pandemic-on-nyc-bike-share-usage-ab79f67ac2df>
- Kranish, C. (2021e, April 9). *Estimating Bike Availability from NYC Bike Share Data*. Medium. <https://towardsdatascience.com/estimating-bike-availability-from-nyc-bike-share-data-7cfc4655d5f6>
- Kranish, C. (2021f, April 13). *Interpolating NYC Bike Share Data to Discover Rebalancing Movements*. Medium. <https://towardsdatascience.com/interpolating-nyc-bike-share-data-to-discover-rebalancing-movements-6cf8a80eb902>
- Kranish, C. (2021g, May 14). *Exploring Bike Share Data*. Medium. <https://towardsdatascience.com/exploring-bike-share-data-3e3b2f28760c>
- Mix, R., Hurtubia, R., & Raveau, S. (2022). Optimal location of bike-sharing stations: A built environment and accessibility approach. *Transportation Research Part A: Policy and Practice*, 160, 126–142. <https://doi.org/10.1016/j.tra.2022.03.022>
- Wang, W. (2016). Forecasting Bike Rental Demand Using New York Citi Bike Data. *Dissertations*. <https://arrow.tudublin.ie/scschcomdis/79>
- Xin, R., Ding, L., Ai, B., Yang, M., Zhu, R., Cao, B., & Meng, L. (2023). Geospatial Network Analysis and Origin-Destination Clustering of Bike-Sharing Activities during the COVID-19 Pandemic. *ISPRS International Journal of Geo-Information*, 12(1), Article 1. <https://doi.org/10.3390/ijgi12010023>
- Yamane, T., & Kaneko, S. (2021). Is the younger generation a driving force toward achieving the sustainable development goals? Survey experiments. *Journal of Cleaner Production*, 292, 125932. <https://doi.org/10.1016/j.jclepro.2021.125932>
- Zhang, L., Xiao, Z., Ren, S., Qin, Z., Goh, R. S. M., & Song, J. (2022). Measuring the vulnerability of bike-sharing system. *Transportation Research Part A: Policy and Practice*, 163, 353–369. <https://doi.org/10.1016/j.tra.2022.05.019>