New York CitiBike Trip Histories

Sajjad Pakdaman Sovoji
York University
Toronto, Ontario, Canada
savoji@yorku.ca

Amin Fadaeinejad
York University
Toronto, Ontario, Canada
afadaei@yorku.ca

Pouya (Adrian) Firouzmakan
York University
Toronto, Ontario, Canada
Adrianfi@yorku.ca

Figure 1: Citibike in New York City

ABSTRACT

Bicycling is an activity which yields many benefits: Riders improve their health through exercise, while traffic congestion is reduced if riders move out of cars, with a corresponding reduction in pollution from carbon emissions. In recent years, Bike Sharing has become popular in a growing list of cities around the world. The NYC "CitiBike" bicycle sharing scheme went live (in midtown and downtown Manhattan) in 2013, and has been expanding ever since, both as measured by daily ridership as well as the expanding geographic footprint incorporating a growing number of "docking stations" as the system welcomes riders in Brooklyn, Queens, and northern parts of Manhattan which were not previously served. One problem that many bikeshare systems face is money. An increase in the number of riders who want to use the system necessitates that more bikes be purchased and put into service to accommodate them. Heavy ridership induces wear on the bikes, requiring more frequent repairs. However, an increase in the number of trips does not necessarily translate to an increase in revenue because riders who are clever can avoid paying surcharges by keeping the length of each trip below a specified limit (either 30 or 45 minutes, depending on user category.) We seek to examine Citibike trip data to extract meaningful relations and conclusions which will provide insight for executive sectors. Our findings can be categorized under three titles: usage trends, common locations, expansion policy. In the first category, we study usage trends on multiple scales ranging from hourly to yearly. The second category is dedicated to finding geographical relations between commonly used stations and the last category suggests an expansion policy based on the current station density and frequency of trips made.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/1122445.1122456

KEYWORDS

Data Analytics, Data Visualization, Bike Trips History,Bike Stations, New York City

ACM Reference Format:
Sajjad Pakdaman Sovoji, Amin Fadaeinejad, and Pouya (Adrian) Firouzmakan. 2022. New York CitiBike Trip Histories . In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 INTORDUCTION AND MOTIVATION

Since 2013 a shared bicycle system known as CitiBike is available in New York City. The benefits of having such a system include reducing New Yorkers' dependence on automobiles and encouraging public health through the exercise attained by cycling. Additionally, users who would otherwise spend money on public transit may find bicycling more economical – so long as they are aware of Citibike's pricing constraints.

There are currently about 33000 shared bikes which users can rent from about 4594 docking stations located in Manhattan and in western portions of Brooklyn and Queens. A rider can pick up a bike at one station and return it at a different station. The system has been expanding each year, with increases in the number of bicycles available and expansion of the geographic footprint of docking stations. For planning purposes, the system operator needs to project future ridership to make good investments.

The available usage data provides a wealth of information which can be mined to seek trends in usage. With such intelligence, the company would be better positioned to determine what actions might optimize its revenue stream.

- In section 2 we review literature on bike share systems, some of which examine the relationship of weather on ridership.
- In section 3 we examine the methodology from the standpoint of data sources, collection, cleaning, exploratory data analysis (EDA), limitations, aggregation, and statistical analysis.
- In section 4 we discuss results from our modeling.
- In section 5 we present a conclusion and summary of the results and discuss proposed future research and enhancements.
- In section 6 we list the references used in this study.

2 RELATED WORKS

Westland et al [17]. examined consumer behavior in bike sharing in Beijing using a deep-learning model incorporating weather and air quality, time-series of demand, and geographical location; later adding customer segmentation. Jia et al[10]. performed a retrospective study of dockless bike sharing in Shanghai to determine whether introduction of such program increased cycling. Their methodology was to survey people in various neighborhoods where the areas were selected by sampling, and the individuals were selected by interviewing individuals on the street. Jia and Fu[11] further examined whether dockless bicycle-sharing programs promote changes in travel mode in commuting and non-commuting trips, as well as the association between change in travel mode and potential correlates, as part of the same Shanghai study. Dell'Amico et al[2]. modeled bike sharing rebalancing programs initially in Reggio Emilia, Italy using branch-and-cut algorithms. In a more recent paper, Dell'Amico et al[3]. examined the bike-sharing rebalancing problem with Stochastic Demands, aimed at determining minimum cost routes for a fleet of homogeneous vehicles in order to redistribute bikes among stations. Zhou[18] analyzed massive bike-sharing data in Chicago, constructing a bike flow similarity graph and using a fast-greedy algorithm to detect spatial communities of biking flows. He examined the questions 1. How do bike flow patterns vary as a result of time, weekday or weekend, and user groups? 2. Given the flow patterns, what was the spatiotemporal distribution of the over-demand for bikes and docks in 2013 and 2014? Hosford et al[9]. surveyed participants in Vancouver, Canada and determined that public bicycle share programs are not used equally by all segments of the population. In many cities, program members tend to be male, Caucasian, employed, and have higher educations and incomes compared to the general population. Further, their study determined that the majority of bicycle share trips replace trips previously made by walking or public transit, indicating that bicycle share appeals to people who already use active and sustainable modes of transportation. In another paper, Hosford et al[8]. determined that that the implementation of the public bicycle share program in Vancouver was associated with greater increases in bicycling for those living and working inside the bicycle share service area relative to those outside the service area in the early phase of implementation, but this effect did not sustain over time. Schmidt[13] observed that the number of bike-sharing programs worldwide grew from 5 in 2005 to 1,571 in 2018. He further noted that disparities in bike-sharing usage are evident around the country, with users skewing towards younger white men. Wang et al[16]. examined the rebalancing problem and determined that the fluctuation of the available bikes and docks is not only caused by the user but also by the operators' own (inefficient) rebalancing activities; they propose a data-driven model to generate an optimal rebalancing model while minimizing the cost of moving the bikes. Vogel and Mattfeld [14] observe that Short rental times and one-way use lead to imbalances in the spatial distribution of bikes at stations over time, and present a case study demonstrating that Data Mining applied to operational data offers insight into typical usage patterns of bike-sharing systems and is used to forecast bike demand with the aim of supporting and improving strategic and operational planning. They analyze both operational data from Vienna's shared bike rental system as well as local weather data over the period. Fuller et al[6]. examined the impact of a public transit strike (November 2016 in Philadelphia) on usage of the bike share service in that city. In an earlier study, Fuller et al[5]. examined bikeshare in Montreal by collecting samples prior to the launch of the program, and following each of the first two seasons. [Unlike other cities such as New York, the Montreal bike share system does not operate year-round. Rather, because of the especially harsh winters, their bikeshare system is dismantled each fall and reinstalled each spring.] Fuller's methodology incorporated a 5-step logistic regression in which the weather variables entered at step 4; this rendered nonsignificant the differences between the three survey periods.

Faghih-Imani and Eluru[4] study the decision process involved in identifying destination locations after picking up the bicycle at a BSS station. In the traditional destination/location choice approaches, the model frameworks implicitly assume that the influence of exogenous factors on the destination preferences is constant across the entire population. They propose a finite mixture multinomial logit (FMMNL) model that accommodates such heterogeneity by probabilistically assigning trips to different segments and estimating segment-specific destination choice models for each segment. Unlike the traditional destination-choice-based multinomial logit (MNL) model or mixed multinomial logit (MMNL), in an FMMNL model, we can consider the effect of fixed attributes across destinations such as users' or origins' attributes in the decision process.

An et al[1]. examine weather and cycling in New York City and find that weather impacts cycling rates more than topography, infrastructure, land use mix, calendar events, and peaks. They do so by exploring a series of interaction effects, which each capture the extent to which two characteristics occurring simultaneously exert a combinatorial effect on cycling ridership – e.g, how is cycling impacted when it is both wet and a weekend day or humid day in the hilliest parts of the cycling network?

Heaney et al[7]. examine the relation between ambient temperature and bikeshare usage and to project how climate change-induced increasing ambient temperatures may influence active transportation in New York City.

In the 1990s, Nankervis [12] examined the effect of weather and climate on university student bicycle commuting patterns in Melbourne, Australia by examining counts of parked bicycles at local universities and correlating with the weather for each day, finding that the deterrent effect of bad weather on commuting was less than commonly believed (though still significiant.)

3 METHODOLOGY

In this section we present our general methodology towards solving this problem. To that end we will explicitly go over the data analytics architecture consisting of the steps one needs to take such as data collection/ingestion process, data storage process, data analysis, data serving, and data visualization. We will also describe the algorithms used while addressing the limitations and difficulties with the respective approaches.

3.1 Data Sources

Citibike makes a vast amount of data available regarding system usage as well as sales of memberships and short-term passes.
For each month since the system's inception, there is a file containing details of (almost) every trip. (Certain "trips" are omitted from the dataset. For example, if a user checks out a bike from a dock but then returns it within one minute, the system drops such a "trip" from the listing, as such "trips" are not interesting.)
There are currently 103 monthly data files for the New York City bikeshare system, spanning July 2013 through December 2021. Each file contains a line for every trip. The number of trips per month varies from as few as 200,000 during winter months in the system's early days to more than 2 million trips this past summer. The total number of entries was more than 100 million, resulting in 200GB of data. To perform the aimed statistical analysis on this dataset we deployed a distributed general framework when applicable. In cases in which our distributed approach was not practical we used subsampling.

3.2 Limitations and Challenges

Because there is so much data, it is difficult to analyze the entire universe of trip-by-trip data unless one has high-performance computational resources. The following are some of the challenges we faced whilst conducting this study:

- Data column names change slightly from month to month.
- In some months, Citibike specifies dates as YYYY-MM-DD, while in other months, dates are MM/DD/YYYY.
- In certain months, the timestamps include HH:MM:SS (as well as fractional seconds) while in other months, timestamps only include HH:MM , as seconds are omitted entirely.
- We encountered an unusual quirk which manifests itself just once a year, on the first Sunday of November, when clocks are rolled back an hour as Daylight Savings time changes to Standard time:
  - The files do not specify whether a timestamp is EDT or EST. On any other date, this is not a problem, but the hour of 1am-2am EDT on that November Sunday is followed by an hour 1am-2am EST.
  - If someone rents a bike at, say, 1:55am EDT (before the time change) and then returns it 15 minutes later, the time is now 1:10am (EST).
  - The difference in time timestamps suggests that the rental was negative 45 minutes, which is of course impossible.
- Sometimes there is an unusually long interval between the start time of a bicycle rental and the time at which the system registers such rental as having concluded.

3.3 Overall Analytics Structure

The overall analytics pipeline that we follow consists of 6 major blocks. This pipeline starts with data cleaning and preprocessing and continues with Feature extraction, Feature Engineering, Distributed Head, Analysis, Visualization (see figure 2).

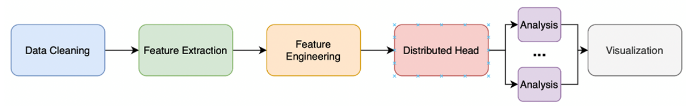

Figure 2: Overall analytics structure

3.4 Algorithms and Implementation

This project involves visualization of this rich dataset therefore the algorithms are also adjusted to such purpose. As the data is provided in CSV format for each month, we used Pandas to read and clean the dataset. For feature extraction, feature engineering we used NumPy, and at the end for visualization we took advantage of numerous platforms including Matplotlib, Seaborn, Tableau, and Folium.
As for the Distributed head we use a divide and concur strategy in which each analysis is performed on a subset of dataset and the result of them will be aggregated in a submodule. Although we did not use parallel processing in this study due to limited computation resources, such a structure fully supports papalism. In figure 3 the exact process of distributing an analysis is depicted. In this scheme one should divide a conventional analysis into two submodules. First one should decide how to perform the analysis on a subset of the data, and in the end how to aggregate the results in a way that the final statistics holds for all splits. For instance, for creating a histogram of data, given that the bin edges are placed the same for all data splits, the analysis part is creating the histogram for that split and result aggregation is summing the corresponding counts from all data splits.

3.5 Interactive Data Visualization

One of the challenges of visualizing rich data is the static scale of figures. For a data set as huge as the one we are working with, there is useful information at different scales. To address this problem, we used an interactive visualization library called Folium. A demo of such interactive visualization can be seen here.

- The top 50 stations for starting and ending.
- The trend of flow for each station, this shows the number of bikes that entered and left that station, color red indicates that the number of bikes left is greater than the number of bikes that have entered. Color green indicates that the number of bikes entered is greater than the number of bikes that have left. (Demo Link)
- This shows the trend of magic-trips (The trips that have been done by the organization itself since no trip records were available for them). Color red indicates that the number of bikes left is greater than the number of bikes that have entered. Color green indicates that the number of bikes entered

is greater than the number of bikes that have left.(Demo Link)
- Top 50 trips that were made. (Demo Link)

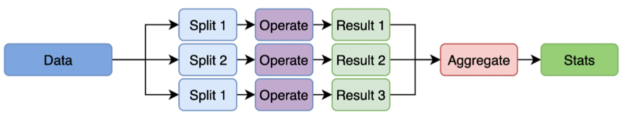

Figure 3: Distributed processing scheme used.

4 RESULTS

The results of this study can be categorized under four components: Usage Trend, Commonly Used Location, Allocation Strategy, and Expansion Policy. In Usage Trend we extracted the basic statistics and trends in the data such as trip start time, end time, duration, and it occurrence with respect to weekdays. For Most Commonly Used Locations we investigate how often different stations are used, and what this implies in terms of geographical relations in NYC. As for the Allocation Strategy we explore the difference between departed and arrived trips to each station. Finally, we will investigate how commonly used stations are distributed with respect to the overall density of stations. By doing so we will suggest an expansion policy which will increase the throughput of the network to the best of our speculations.

4.1 Usage Trend

The bar graphs in figure 5 are Illustrating the cumulative number of bikes, the cumulative number of stations, and the number of trips in each month since June 2013, respectively. AS it can be seen, the number of stations has increased significantly since 2013. However, regarding the number of bikes, there is no considerable change since December 2020. On the other hand, by taking the number of trips per month into consideration, it is noticeable that the same pattern has been repeated each year because of changing the seasons and general weather conditions despite the fact the number of trips has increased.

Figure 5 shows the trend for using the Citibikes in New York on each day of the week between 2020 and 2022. It is obvious that there is not any remarkable difference between the days of the weeks. The reason would be understandable by considering the main factors for determining the location of the stations. This part would be explained in Section Stations' Location Trends.

4.2 Stations' Location Trends

After analyzing the trend for allocating the bike stations using figure 6 showing the location of the bike stations in 2013 and 2022, it turns out that, regardless of several potential aspects, e.g., the distribution of population and the age of people, the main factors to determine the coordinates of the stations are the Number of House Units and White-Collar Occupations. In this figure,there are two set of maps related to 2013 and 2022. from left to right, each set is representing the location of stations, location of the stations considering the house units, and location of stations by taking the

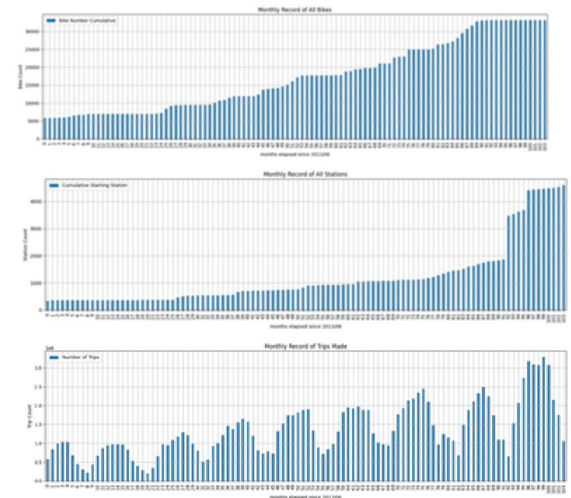

Figure 4: Monthly usage trends from 2013 to 2022

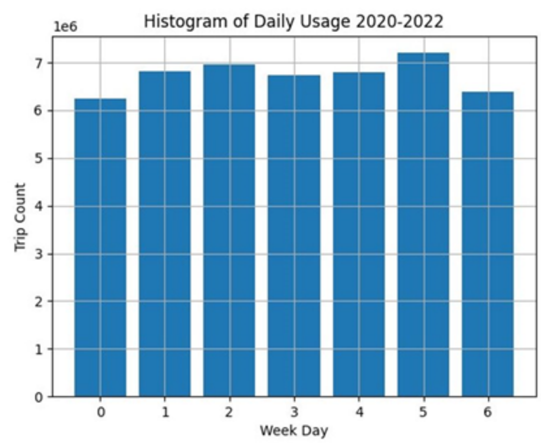

Figure 5: Weekdays trips trend from 2013 to 2022

White Collar Occupations into account, respectively. It is obvious that increasing the number of stations from Manhattan to the north and east is a solution to deal with the heavy traffic and rush hours in New York by encouraging people to use bikes instead of private cars and even other types of public transportation. Accordingly, it is reasonable that the number of bikes used on each day of the week is similar to the other weekdays as a large number of the trips have been devoted to commuting to work.

Another interesting point is that the number of new stations and the area they are covering is suddenly boosted in 2020 which is the year that the pandemic started (This fact can be confirmed by Phase 3 of the Major Citibike Expansion link). In this regard, it is safe to say that the pandemic was a fortune for the authorities to improve the infrastructures faster and increase the number of stations and cover more areas including Queens and a part of Bronx.

In figure 7, the start time and end time of the trips between 2020 and 2022 are illustrated. The ironic point of this diagram is that it

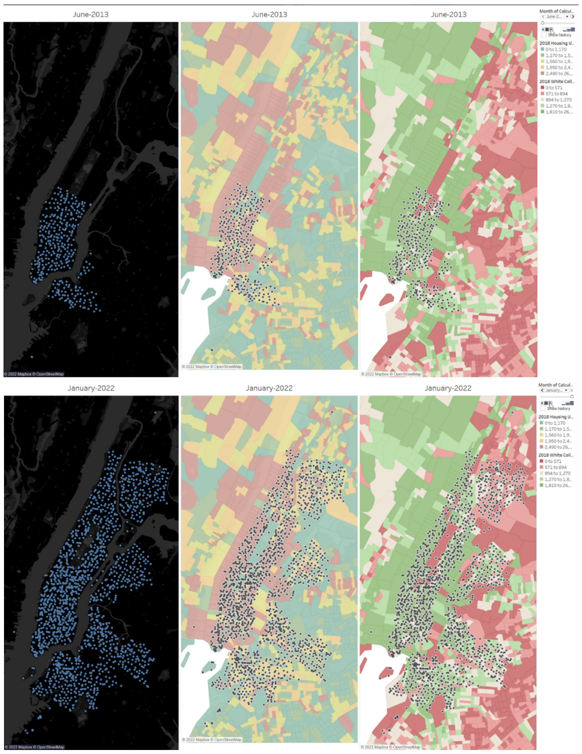

Figure 6: Locations of the bike stations in 2013 and 2022.

seems, in general, the end time curve is almost a shift form of the start time. To understand the cause of the effect, the Histogram of Trip Duration (figure 8) is employed. It is significantly noticeable that most trip duration are between 0 and 30 minutes. One reason is that the number of stations is mostly allocated for White Collar Occupations and, most people have been using bikes for commuting. However, another considerable reason can be found by checking out the fees for renting the bikes. In this regard, based on figure 9, after 30 minutes, the renter would be charged $ 0.23 per minute. Hence, clearly, people prefer to end the trip before 30 minutes and start it again as it is more reasonable from the price point of view if they were to want to use the bike for another 30 minutes.

4.3 Effect of Pandemic

From the figure 10 showing the number of trips per day from January 1st, 2019 to January 1st, 2021 it is clear that the general pattern in 2020 is almost the same as in 2019 except a period of time between March 14th and April 30th in 2020 due to the increasing the number of deaths caused by Covid-19. In addition to this, there is another study published in June 2021 that confirms the fact that the usage of the Citibikes has recovered very fast even during the peaks of the pandemic in comparison with other types of public transportation such as subway (Wang 2021 [15]). The reason is in fact not surprising as during the pandemic people prefer to, for example, commute alone for the sake of safety rather than being on a bus with other people.

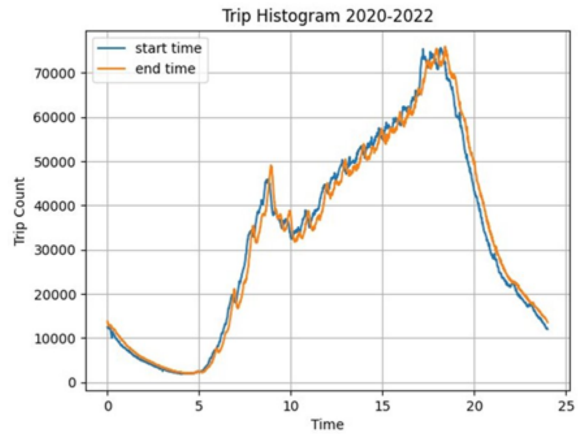

Figure 7: Start and end time histogram between 2020-2022.

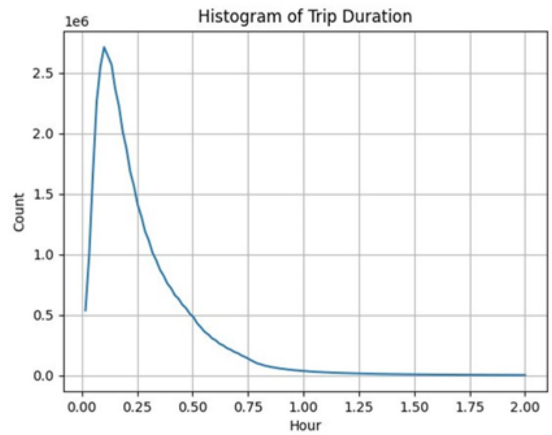

Figure 8: Trip duration histogram from 2013 to 2022

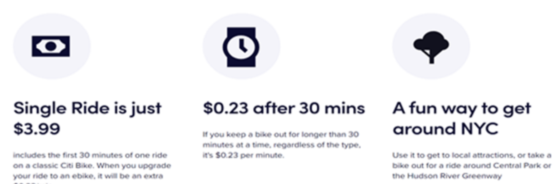

Figure 9: Citibikes rental fees (link)

4.4 Commonly Used Locations

In figure 11, maps are showing the top 50 most used starting and ending stations in 2021. As can be seen, all of them are located in Manhattan although there are stations in other areas such as Queens, just to name. The information obtained from these maps have been used in the next subsection to demonstrate interesting discoveries.

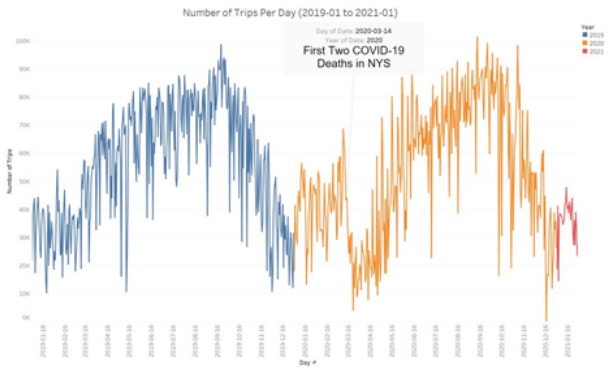

Figure 10: Number of trips per day during pandemic

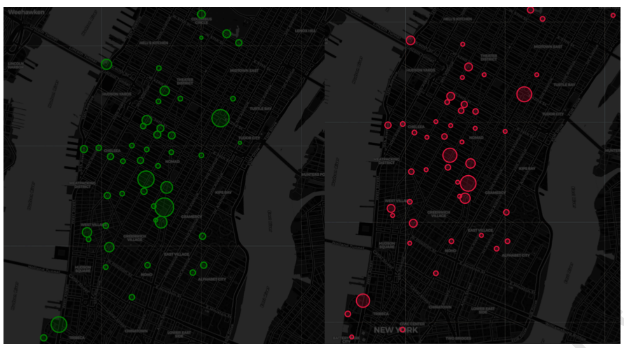

Figure 11: Most used stations for starting and ending trips from 2013 to 2022 (green: entered, red:exited)

4.5 Proposed Expansion Policy

As mentioned before, the expansion of the bike stations was from Manhattan to the north and east to cover more areas. However, in this section a visualization analysis for expanding the bike stations to support the most used stations is proposed. By putting figure 11 into a heat map illustrating the density of stations in different area of New York and zooming in, it can be seen a number of commonly used stations are located in areas that there are not enough number of stations there. In this regard, these areas have the potential to have more stations and as a result more bikes (see figure 12).

4.6 Allocation Strategy

There is no guarantee that the number of bikes available at a specific location is going to remain constant, because there is going to be some stations that the number of bikes that have left would be far more than the number of bikes that have entered. Figure 13 shows the flow of the trips. It could be noticed that the bikes are moving from the center of the city(Manhattan) towards the north and east side of the city, Since the color of the circles in the center of the city red it shows that more trips are moving outwards from the center and the color green in the figure 13 indicates that the number of bikes coming in that station is more than the number of bikes leaving that station. For more details you can see our web Demo.

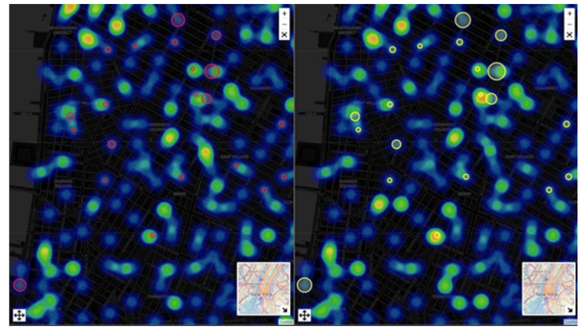

Figure 12: Proposed expansion policy using heat-map and most used stations from 2013 to 2022

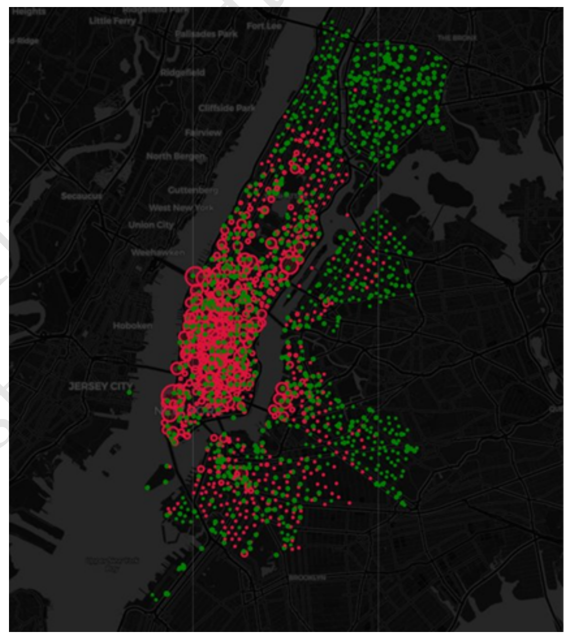

Figure 13: Flow of trips from 2013 to 2022 (green: entered, red:exited)

This issue is a problem since after a while there is going to be one station that will run out of bikes. In order to solve this problem, Citibike, moves the bikes from the station that are receiving more bikes to the stations that are losing bikes. The way detected this phenomenon is that we find out that if stating station is different from the last position that the bikes previously had. We called this phenomenon as **Magic Trip**, figure 14 shows the flow of the magic trips. We also have prepared a web Demo.

Figure 15 shows the duration of the magic trips, this parameter is measured by computing the time where the bikes has not been used. According to figure 15 there are some conclusions about that fact:

- Main peak is under 5 hours.
- More peaks happen periodically 24 hours apart (Daily operations).

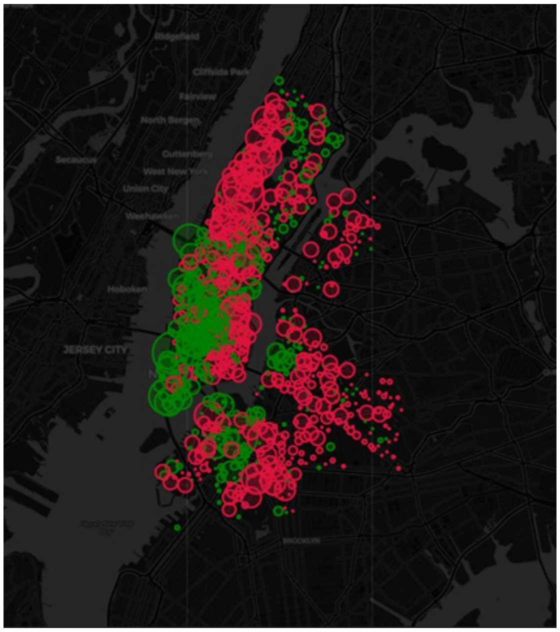

Figure 14: Flow of magic trips from 2013 to 2022 (green: entered, red:exited)

- Long duration could mean that the bike has gone to maintenance.

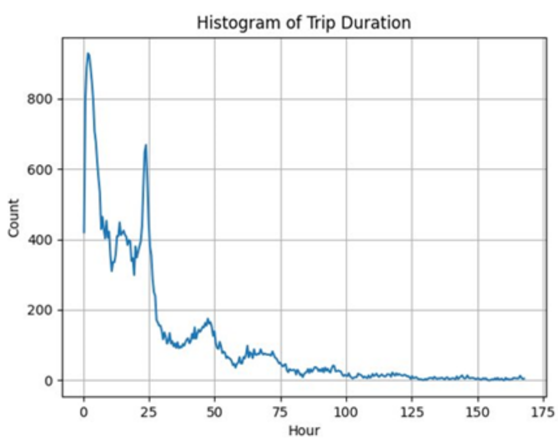

Figure 15: Magic Trip duration from 2013 to 2022

Figure 16 shows the trend of magic trips during the week days. Most of the magic trip have been occurring during the weekdays while in the weekends the system is less active. This makes sense because Citibike similar to other companies in USA, work less during the weekends.

4.7 Common Trips

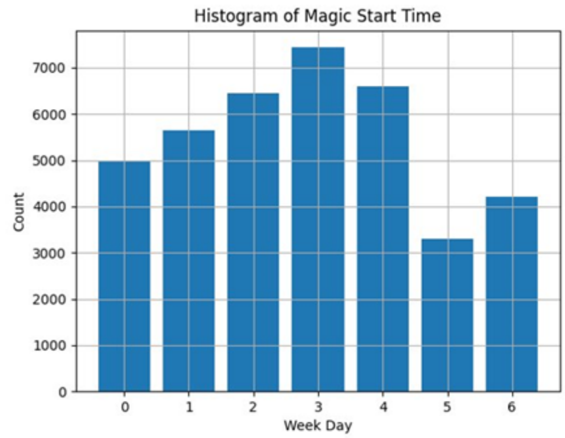

Figure 16: Flow of magic trips from 2013 to 2022

Figure 17 shows the most common trips that have been occurred from 2013 to 2022. the interesting discovery that we have is that the top most common trip has a same staring and ending station, which means that the rider has returned the bike to the same position after his/her ride. The interesting part is that the location of this station is beside the central park at New York, which means that the reason for this is that a lot of users where using the bikes for exercising. In can be seen in figure 17 there are a few dots near the central park which represent this phenomenon. It is worth mentioning that the green color represents that the number of bikes exited that station is more than the bikes that have entered, color red represents that the number of bikes exited that station is less than the bikes that have entered. We also have prepared a web Demo. for this, where you see the details about it.

Figure 17: Most Common trips from 2013 to 2022

5 CONCLUSION

In conclusion, we state that there is indeed a wealth of information within this data set many of which is yet to be explored. In our study which was performed withing the scope of a graduate-level course, we found the following:

- The number of active bikes and stations have seen constant growth throughout these 103 months of data collection with an exceptionally fast rate in 2020.
- The number of trips made follows a seasonal trend in which more trips are made during warm seasons in contrast with cold seasons.
- Most Trips start and end within the traffic rush hours meaning that this bike sharing system is used as a solution for NYC traffic congestion.
- Most trips take more than 15 minutes, while a standard ride is expected to be 30 minutes in the company's policy. This suggests that administrators can adjust the trip standard duration to further increase overall profitability. This also indicated a need for more flexible usage plans for customers who use these bikes for longer duration.
- These bikes are used the same withing a week which means that the maintenance schedule should be adjusted to other factors such as traffic congestion in streets rather than bike usage in general.
- The network expansion policy is correlated with housing units' density and office locations.
- Although the pandemic imposed a radical drop in the number of trips made in 2020, but bikeshare has recovered its usual trend whereas many public transportation systems still lack safety standards.
- Commonly used stations are located within the financial district of Manhattan.
- The most used trip is a path around Central Park.
- Stations which are in the central region of Manhattan sink (bikes end up there) while the stations in the western area of Queens are source (bikes start from there).
- Bikes are relocated via network admins for allocation purposes as well as maintenance.
- Bike re-locations follow the reversed trend in terms of allocation compared to usual trips.
- Interactive visualization recovers some information that is hidden with static visualization.

REFERENCES

[1] Ran An, Renee Zahnow, Dorina Pojani, and Jonathan Corcoran. Weather and cycling in new york: The case of citibike. *Journal of Transport Geography*, 77:97–112, 2019.

[2] Mauro Dell'Amico, Eleni Hadjicostantinou, Manuel Iori, and Stefano Novellani. The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, 45:7–19, 2014.

[3] Mauro Dell'Amico, Manuel Iori, Stefano Novellani, and Anand Subramanian. The bike sharing rebalancing problem with stochastic demands. *Transportation Research Part B: Methodological*, 118:362–380, 2018.

[4] Ahmadreza Faghih-Imani and Naveen Eluru. A finite mixture modeling approach to examine new york city bicycle sharing system (citibike) users' destination preferences. *Transportation*, 47:529–553, 2018.

[5] Daniel Fuller, Lise Gauvin, Yan Kestens, Mark Daniel, Michel Fournier, Patrick Morency, and Louis Drouin. Impact evaluation of a public bicycle share program on cycling: A case example of bixi in montreal, quebec. *American Journal of Public Health*, 103(3):e85–e92, 2013. PMID: 23327280.

[6] Daniel Fuller, Hui Luan, Richard Buote, and Amy H. Auchincloss. Impact of a public transit strike on public bicycle share use: An interrupted time series natural experiment study. *Journal of Transport Health*, 13:137–142, 2019.

[7] Alexandra K. Heaney, Daniel Carrión, Katrin Burkart, Corey Lesk, and Darby Jack. Climate change and physical activity: Estimated impacts of ambient temperatures on bikeshare usage in new york city. *Environmental Health Perspectives*, 127(3):037002, 2019.

[8] Kate Hosford, Daniel Fuller, Scott A. Lear, Kay Teschke, Lise Gauvin, Michael Brauer, and Meghan Winters. Evaluation of the impact of a public bicycle share program on population bicycling in vancouver, bc. *Preventive Medicine Reports*, 12:176–181, 2018.

[9] Kate Hosford, Scott Lear, Daniel Fuller, Kay Teschke, Suzanne Therrien, and Meghan Winters. Who is in the near market for bicycle sharing? identifying current, potential, and unlikely users of a public bicycle share program in vancouver, canada. *BMC Public Health*, 18, 11 2018.

[10] Yingnan Jia, Ding Ding, Klaus Gebel, Lili Chen, Sen Zhang, Zhicong Ma, and Hua Fu. Effects of new dock-less bicycle-sharing programs on cycling: a retrospective study in shanghai. *BMJ Open*, 9(2), 2019.

[11] Yingnan Jia and Hua Fu. Association between innovative dockless bicycle sharing programs and adopting cycling in commuting and non-commuting trips. *Transportation Research Part A: Policy and Practice*, 121:12–21, 2019.

[12] Max Nankervis. The effect of weather and climate on bicycle commuting. *Transportation Research Part A: Policy and Practice*, 33(6):417–431, 1999.

[13] Lucas Van der Meer. *Spatio-Temporal Forecasts for Bike Availability in Dockless Bike Sharing Systems*. PhD thesis, 02 2019.

[14] Patrick Vogel and Dirk C. Mattfeld. Strategic and operational planning of bike-sharing systems by data mining – a case study. In Jürgen W. Böse, Hao Hu, Carlos Jahn, Xiaoning Shi, Robert Stahlbock, and Stefan Voß, editors, *Computational Logistics*, pages 127–141, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[15] Haoyun Wang and Robert B. Noland. Bikeshare and subway ridership changes during the covid-19 pandemic in new york city. *Transport Policy*, 106:262–270, 2021.

[16] Shuai Wang, Tian He, Desheng Zhang, Yuanchao Shu, Yunhuai Liu, Yu Gu, Cong Liu, Haengju Lee, and Sang Son. Bravo: Improving the rebalancing operation in bike sharing with rebalancing range prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2:1–22, 03 2018.

[17] James Christopher Westland, Jian Mou, and Dafei Yin. Demand cycles and market segmentation in bicycle sharing. *Information Processing Management*, 56(4):1592–1604, 2019.

[18] Xiaolu Zhou. Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in chicago. *PLOS ONE*, 10:1–20, 10 2015.