

Exploratory Data Analysis of CitiBike System Data in New Jersey and Hoboken: Insights and Trends from the NYC Citi Bike Data [2021 - 2023].

Ivan Francis

AIT 580

Abstract:

Since its launch in 2013, New York City's Citibike system has been a well-liked means of transportation for commuters and visitors alike. Since then, the system has been extended to more regions, including Hoboken and New Jersey, giving its consumers even more convenience. This project report focuses on investigating the Citibike system statistics for the years 2021 to 2023 in the cities of New Jersey and Hoboken that were taken from the bigger NYC dataset.

The purpose of the research questions covered in this study is to shed light on the utilization trends and patterns of the Citi Bike system in Hoboken and New Jersey. Initially, we look into the weekly and seasonal patterns of Citi Bike usage in the focus areas. Second, we investigate any demographic factors—such as age and gender—that might have an impact on these usage patterns. Finally, we look into whether member riders and casual riders have any different usage habits. To gain useful insights from the information, a number of data visualization and statistical methodologies were employed. According to the statistics, there are significant seasonal and weekly consumption patterns seen in the Hoboken and New Jersey regions, with usage obviously rising throughout the summer and on weekends. Males and younger age groups tend to utilize the Citibike system more frequently than females and older age groups, and there were also other noticeable demographic disparities in utilization patterns. Our data also showed that compared to casual riders, members use the Citi Bike system more frequently and for longer periods of time. For stakeholders and decision-makers in the transportation sector, the findings from this project report may be valuable in making wise choices regarding the operations and regulations of the Citi Bike system. This project report offers insightful information about the usage trends and patterns of the Citi Bike system in the Hoboken and New Jersey regions. The results point to significant seasonal and weekly usage patterns, demographic usage pattern disparities, and different usage patterns between casual and member riders. In order to improve the Citibike system's operations and the services it offers to its consumers, stakeholders can make data-driven decisions by having a greater understanding of these trends and patterns.

Introduction:

With a combined population of more than 10 Million, New Jersey and Hoboken are two of the most inhabited regions in the country. Since the populace mainly relies on public transportation, biking, and walking, transportation is essential in these places. To provide residents and visitors with an affordable and practical means of getting around the city, the Citibike system has been introduced as an alternate mode of transportation. For stakeholders to make educated decisions about the system's operations and rules, they need to understand the data from the Citi Bike system in New Jersey and Hoboken. This research report intends to offer insightful information about the trends and usage patterns of the Citi Bike system in these locations.

The data for the Citi Bike system used in this study was taken from the wider NYC dataset and spans the years 2021 to 2023. A number of columns are included in the information, including "journey duration," "start and end times," "start and end stations," "user type," "gender," and "birth year." The information in each column value can be utilized to study the utilization trends of the Citi Bike system in the Hoboken and New Jersey regions.

Although the Citibike system in New York City has been the subject of a sizable amount of study and analytics, the statistics from the system in Jersey City and Hoboken have received much less attention. This initiative intends to close this research gap and offer insights on usage patterns and trends of the Citi Bike system in various regions.

The sections that follow in this project report will give a thorough analysis and insights by answering the following research questions :

- 1.) What are the busiest bike rental locations in New Jersey and Hoboken City and how do usage patterns change over the day, week, and season?
- 2.) Are there any demographic factors/patterns that are visible when considering the Age and Gender of the riders utilizing the CitiBike transport system ?
- 3.) Are there any patterns or differences that can be observed between users or riders who are of two categories that are casual and members.

Literature Review :

With its efficient and environmentally friendly alternative to more conventional modes of transportation like cars and buses, the Citibike system has grown in popularity as a means of mobility in urban areas. In contrast to New York City, where the Citibike system has been thoroughly researched, Hoboken and New Jersey, where the system is most prevalent, have seen comparatively little research on it. Policymakers and transportation planners must comprehend the usage trends and effects of the Citi Bike system in these locations in order to develop and put into practice efficient transportation strategies.

"Big Blue Comes to Jersey City: A Study of Citi Bike Expansion into New Jersey " is one research that has looked into the extension of the Citi Bike system into New Jersey[1]. The Alan M. Voorhees Transportation Center at Rutgers University completed this study, which was then published in the Journal of Transport Geography in 2018. The study examined the usage trends and effects of the Citi Bike system in Jersey City and discovered that it offered a different mode of transportation and had a beneficial effect on the neighborhood's economy. There is still a need for more research on the Citibike system in other parts of New Jersey, such as Hoboken, even if the study offers insightful information on its effects in Jersey City.

2017 saw the publication of a paper titled "Exploration Of A Public Bike Sharing Program In Hudson County" in the Journal of Urban Planning and Development[2]. The study examined the usage trends and user traits of the New Jersey's Hudson County Hudson Bike Share program. According to the survey, the system was largely utilized for quick excursions and commuting to work or school, and that the majority of users were young, male, and highly educated. The study also discovered that the system improved local transportation and gave locals and visitors a different mode of transit. The study's main finding is that bike-sharing programs have the potential to provide effective and environmentally friendly transportation in metropolitan settings.

Another similar research was done for the bike sharing program for NYC rather than Jersey City and Hoboken. However, as the research objective and motivation align with the topic of this project report it was important to analyze the content provided in the research which was "Weather and cycling in New York: The case of Citi Bike". The study examines the connection between weather and the use of Citi Bikes in New York City[3].The authors use CitiBike and meteorological data to investigate elements such as temperature,topography,land use, precipitation and wind speed affecting cycling, specifically the Citi Bike System in New York. What are the busiest bike rental stations in New York City, and how do utilization patterns alter over the day, week, and season? are the research questions that the study is pertinent which confirms the direction of my own research questions as well. According to the survey, the weather has a big impact on how often people use Citi Bikes, with clear and mild days showing greater usage rates. According to this data, the busiest bike rental locations may be affected by usage patterns that

are influenced by variables like the season and weather. For instance, throughout the summer, usage patterns may alter as a result of the warmer weather, resulting in more bikes being rented in specific locations or at particular times of the day. The paper also emphasizes how crucial it is to comprehend how weather patterns affect bike-sharing usage because doing so can help with decisions about the placement of bike stations and maintenance schedules. Overall, by examining a particular component that influences usage patterns, the report offers insights that supplement my first study question. It implies that weather conditions should be taken into account when examining utilization patterns and the location of bike stations.

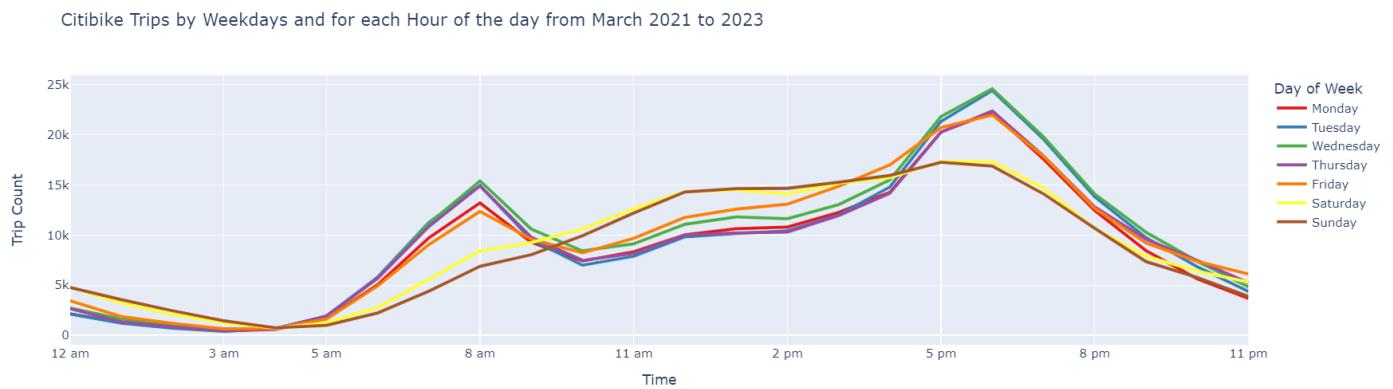
Data Source and Methodology:

The NYC Citi Bike System Data[4] is the source of the data for this project, with an emphasis on the information pertinent to the NJ and Hoboken region from March 2021 to 2023. The data was downloaded as CSV files submitted for each month from the official Citibike website, which provided a hurdle because the files needed to be combined to create a comprehensive dataset. To do this, a total of 25 CSV files were combined using the Python "pandas" and "os" packages. Each CSV file was cleaned using Excel by filtering null values, especially for the birth year, before the data was combined. Since there were few null values relative to the total number of data points, it was decided to eliminate the null data. The final dataset had 1.6 million data points after cleaning.

Python, R, and PySpark were used to undertake exploratory data analysis (SQL). In order to address research questions about seasonal and weekly usage trends, age distribution of riders, trips distribution based on gender and age, popular stations, and variations in usage by different types of riders, Python was used to add age values from the customers' birth years, create visualizations, and summarize statistics. The massive dataset was queried using PySpark, and SQL queries were used to extract summary statistics and evaluate data according to the most popular stations. The utilization density for each start station and finish station in the dataset was then visualized more effectively and interactively using R. In conclusion, the process entailed combining and cleaning the data before performing exploratory data analysis with Python, R, and PySpark.

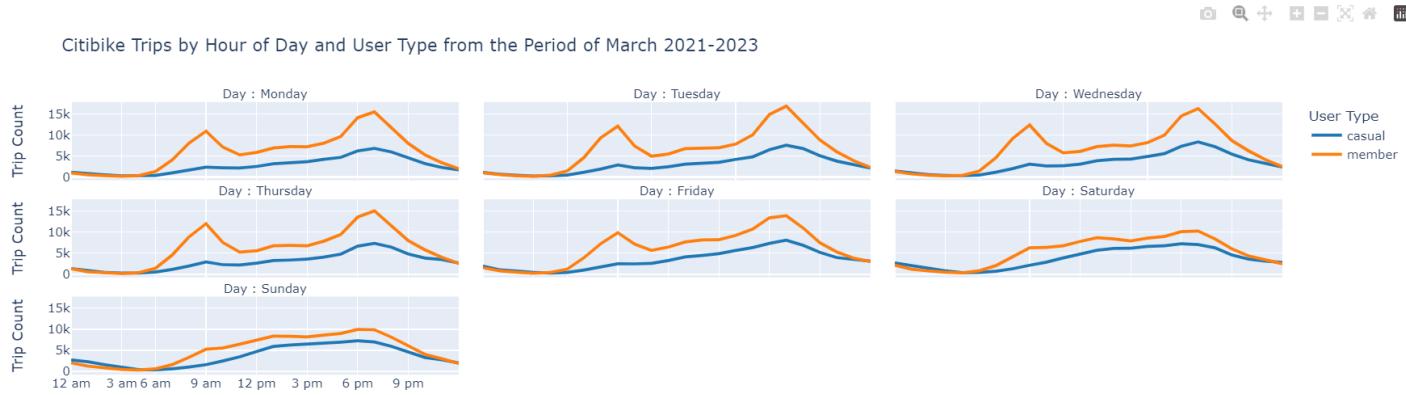
Analysis and Results :

A.) Understanding the busiest stations and usage patterns among Citibike Riders in NJ and Hoboken City.

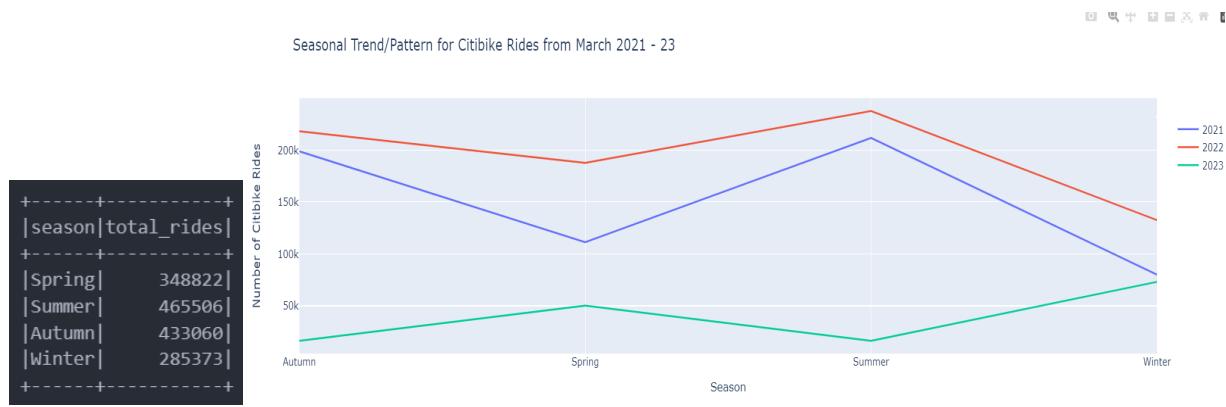


First, we look at how frequently the Citibike Riders in New Jersey and Hoboken are using the bikes on a day to day basis by tracking the trip count for each hour of the day. Looking at the visualization, we can say that in the weekdays, which is Monday-Friday, usage of CitiBikes peaks in the early mornings from 7 am to 9 am which makes sense as most of the population are commuting towards their workplaces. The graph also displays that in the

evenings which is 5pm to 8pm utilization of CitiBikes again peaks or increases as during this time of the day most of the population return home. However, we can see an interesting trend going on where in the mornings the peak isn't on the same level as in the evenings. This can be explained, as the data is recorded from March 2021 - 2023 which means that post-covid there were flexible schedules that were allotted to people working full time jobs. As a result, when such a factor is taken into consideration we can explain the lower level of peak in the morning due to people working from home or working hybrid schedules during this time. Moreover, we can also observe that Sunday's and Saturday's do not observe the same peaks in the morning and evenings compared to the weekdays which makes sense as most of the population are on holidays on the weekend.

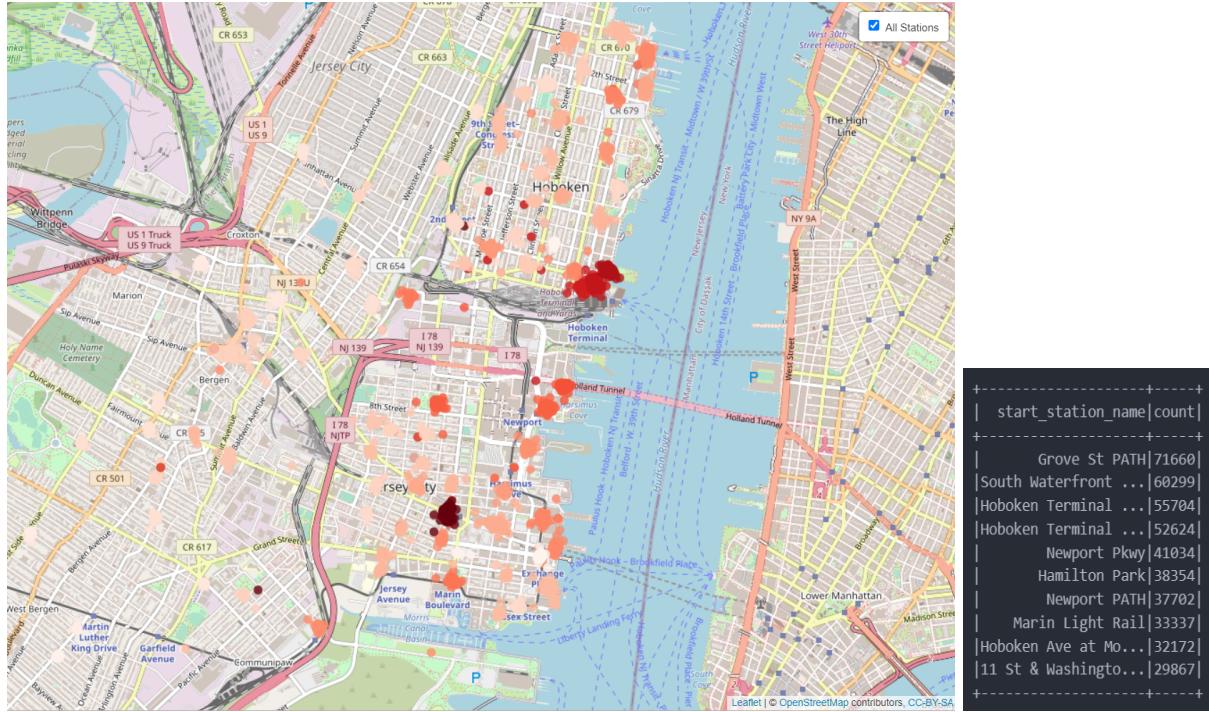


When we plot the CitiBike utilization per hour of the day graph for casual and member riders we surprisingly observe that CitBike members travel and utilize the Bikes more when compared to the casual riders which is not the same when you compare the studies on CitiBike based on NYC from 2013 to 2017. This can be explained as New Jersey and Hoboken City are less densely populated compared to NYC, and hence the demand for CitiBikes could be relatively less which means that Citibike Member riders do not face any competition in that regard.



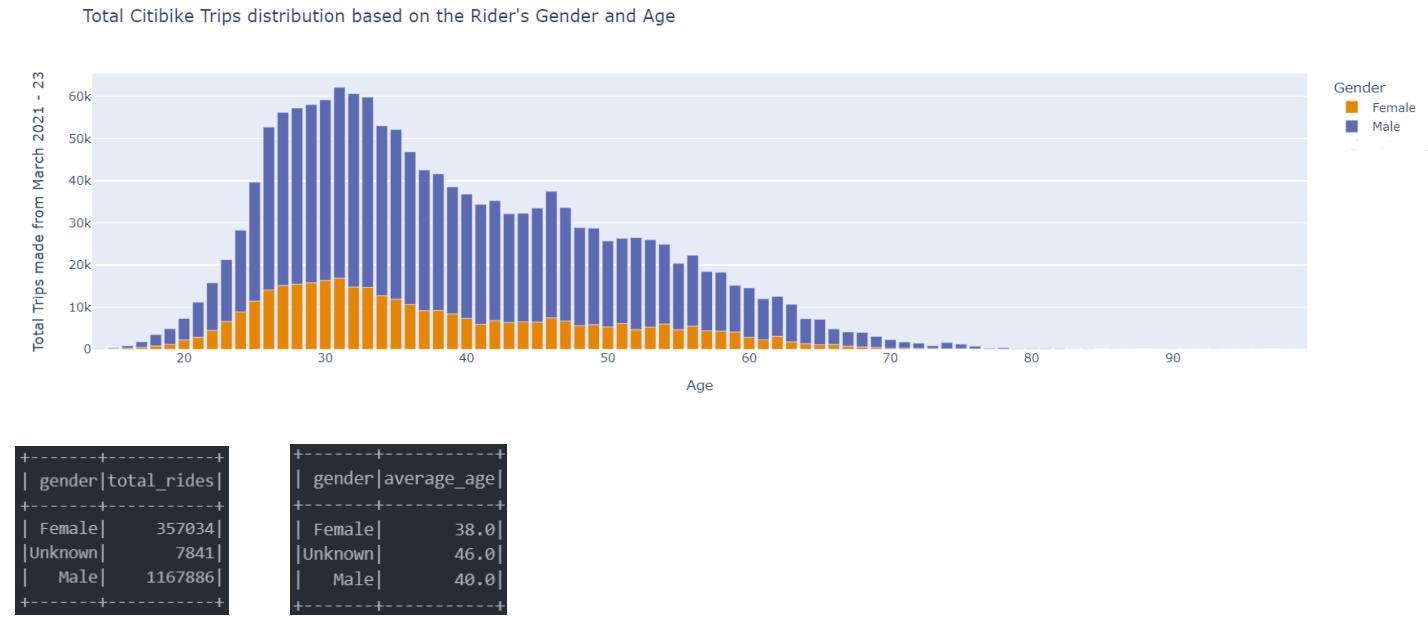
The Seasonal Trend / Pattern for Citi Bike Rides is another interesting infographic which tells us when commuters in the city prefer riding or using a Citi Bike. When we take the total rides per season into consideration we can see that commuters use it the most in the Summer followed by Autumn , Spring and Winter. The graph for 2021 is lower as explained above due to the pandemic which resulted in less commuting numbers. However, 2021 and 2022 follow the same pattern. The reason for the year 2023 to have such low users and a different pattern is because it only consists of data from January to March and not the whole seasonal year.

To better understand which CitiBike Stations are being utilized the most, a map plot was used by taking the stations and their total count in the dataset to comprehend the busiest stations in New Jersey and Hoboken City.



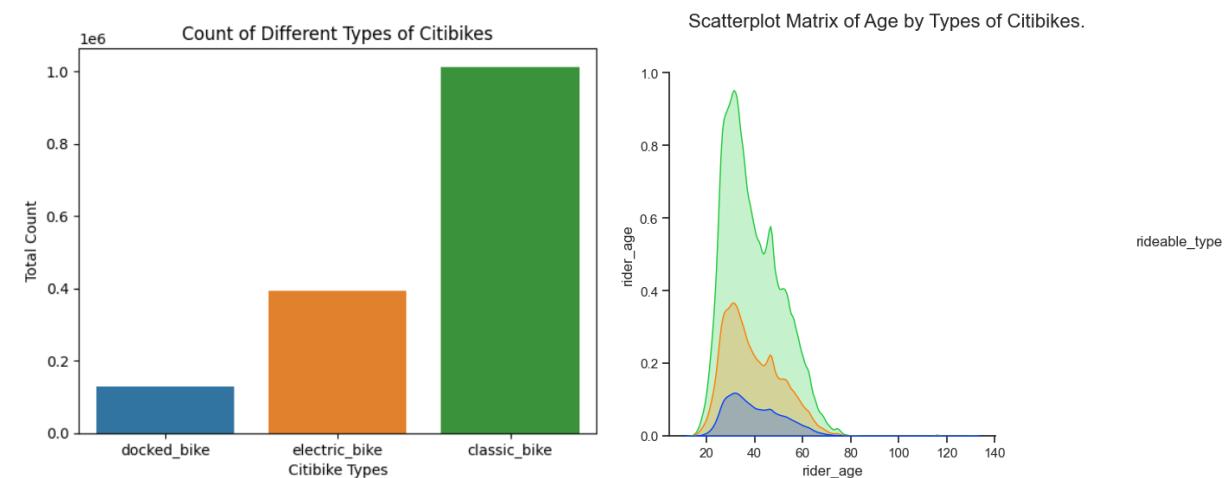
We observe that the busiest CitiBike Stations are the Grove St Path , Hoboken Terminal and Newport Pkwy and much more. The high utilization of these stations can be mainly seen at Grove St. Path and the Hoboken Terminal. A significant transportation hub in Jersey City is Grove St. Path, which links numerous means of transportation including ferries, buses, and trains. The station is a well-liked destination for commuters and locals alike due to its placement in a heavily populated region with numerous residential and business buildings. Moreover, Grove St. Path station offers easy access to well-known locations including Newport Mall and Newport Waterfront. Grove St. Path in New Jersey is a very popular destination for Citibike users due to a combination of all these characteristics. The same can be seen with Hoboken Terminal, It functions as a transit hub for access to various regions of New Jersey and New York City. It is therefore a well-liked site for commuters and tourists, making it an ideal place for Citibike users.

B.) Analysis based on Citi Bike user's Age and Gender.



Another very important thing to consider is understanding the demography of the population utilizing the CitiBike service. This was done by deconstructing the Total CitiBike trips based on the riders gender and age. We can see that there are more male riders compared to female riders with an approximate ratio of 1:3 female to male riders. Also the average age for the female population is 38 years whereas the average age for the male population using the CitiBikes is 40 years. Which means that the CitiBikes are not only being utilized by one specific age group on the younger side or older side. People from all types of age groups are utilizing them for reasons such as commuting to work , school and colleges.

C.) Different Types of Citi Bikes and how are they utilized by the population.



CitiBike offers three types of Bikes to the users which are : docked bikes , electric bikes and the classic bike. From the top left bar chart based on the total count of Citi Bikes users v/s CitiBike Types we can clearly observe that classic bikes are used the most by users followed by electric bikes. While the docked bikes are utilized the least. We can also understand how the age can be involved in the former infographic by creating a scatterplot matrix of the

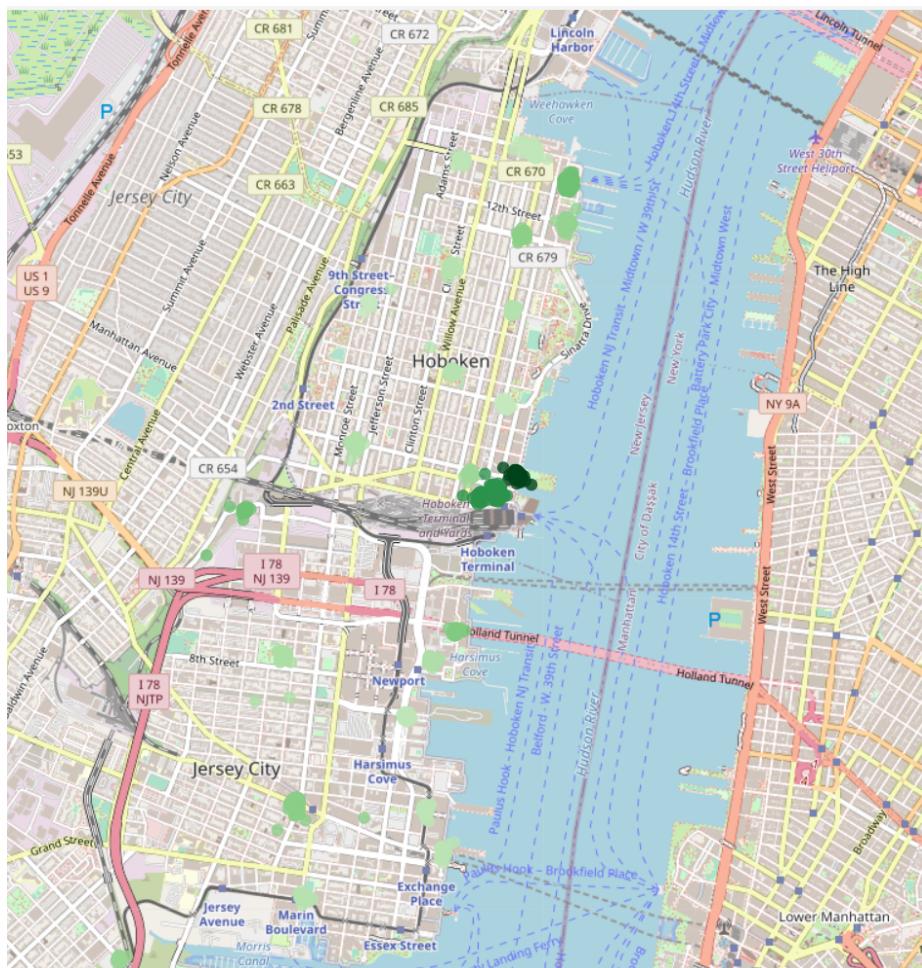
User's age and the types of Citibikes. We observe that users over the age of 35 Years and above prefer to use the classic bike while users over the age of 10-12 Years of age tend to use the electric bikes which makes sense because the younger usebase are mostly teenagers who have a preference for speed , better mobility , comfort and convenience as it requires less effort to ride the electric bikes.

D.) Understanding Citi Bike usage patterns and trends based on customer types [casual and member riders].

member_casual	User_count
casual	542642
member	990119

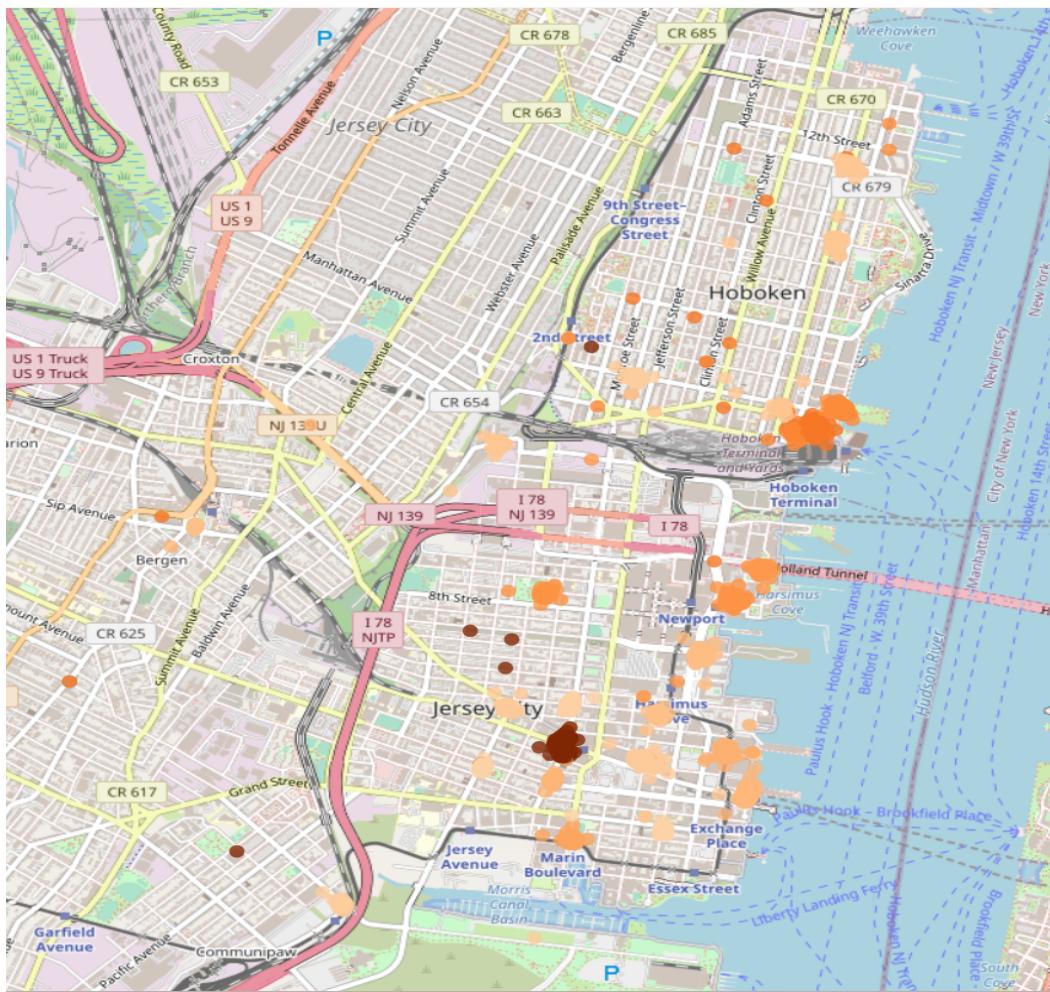
In the infographic for CitiBike Trips per hour of the day based on casual and member riders we observed that member riders are more. With the above query we can exactly see that the users who have a Citi Bike membership are 46 percent more in population when compared to casual users. However, a deep dive can be done with respect to the popular stations that are used by casual and member riders while also taking the popular routes into consideration to see if there are any differences.

D. 1.) Top 25 popular stations that are utilized by casual Citi Bike riders.



From the map, we can observe that the popular Citi Bike stations which are used by casual members don't really show a strong insight except for the Hoboken Terminal. Now as explained above, Hoboken terminal is a major transportation hub for commuters, with connections to trains, buses and ferries. This suggests that many people may be using CitiBike as a last-mile transportation option to go from the station to their final destination or as a method of transportation to get to the terminal from their home or place of employment. Second, Hoboken Terminal is situated in an area that is very crowded and has a lot of homes and businesses. Because casual customers are more likely to use bike share for quick, one-way excursions, there may be a strong demand for CitiBikes in the neighborhood. Last but not least, Hoboken Terminal is close to a number of well-known locations, including parks, restaurants, and entertainment centers. This suggests that a sizable number of infrequent users may be using CitiBike to explore the neighborhood and visit these attractions.

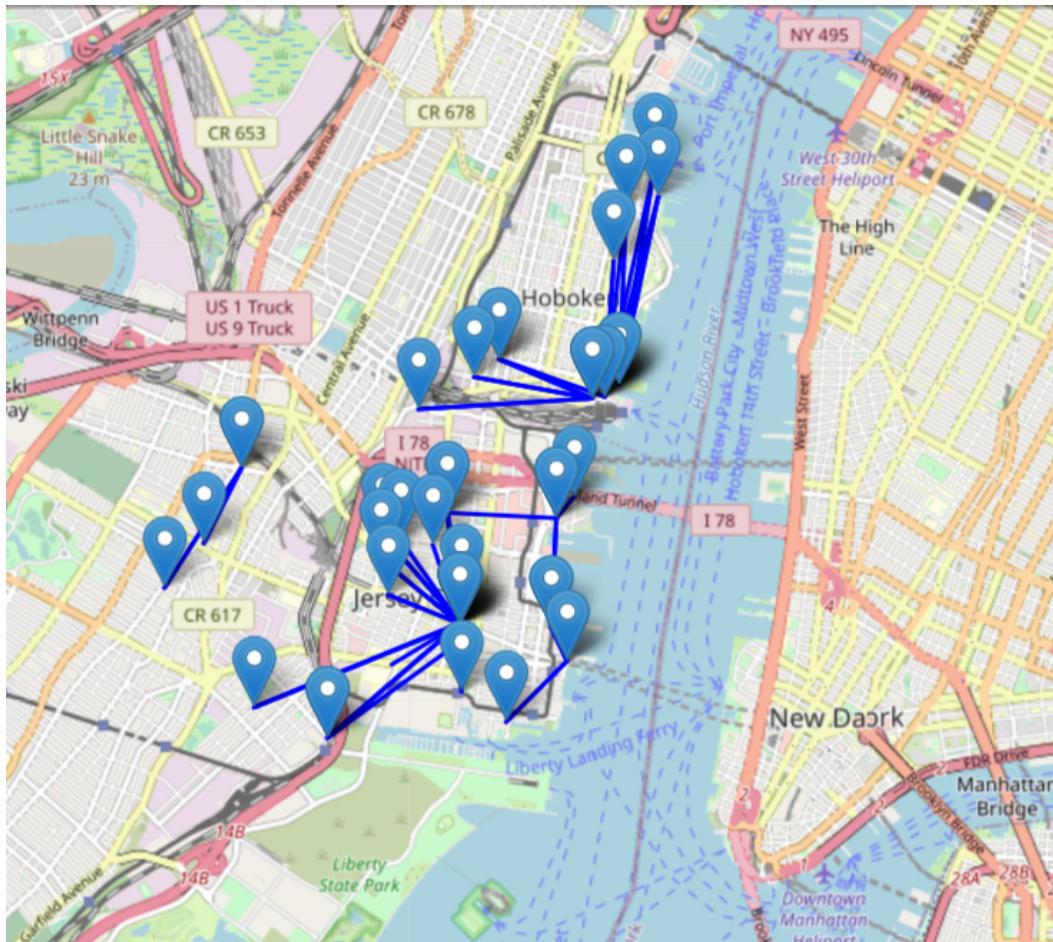
D. 2.) Top 25 popular stations that are utilized by riders having a Citi Bike Membership.



When considering the 25 most utilized Citi Bike stations by users who have membership we see that the stations near Grove Street Path north of Marin Boulevard and Newark Avenue are being the most used by riders who have a CitiBike membership. Also Citi Bike Stations at Newport and Hoboken Terminal are also really popular stations for CitiBike Members. It is because Grove Street Path, Newport, and Hoboken Terminal all have Citi Bike stations nearby with access to New York City's public transportation systems. The PATH train system, which includes stops

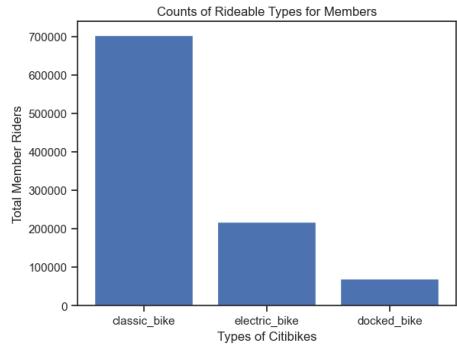
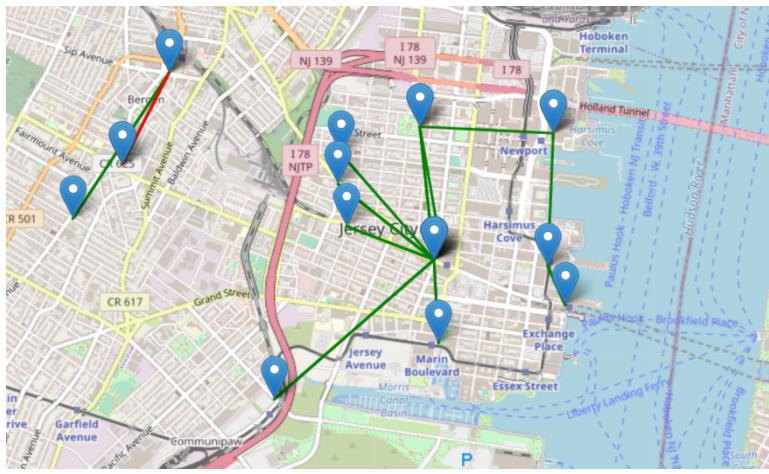
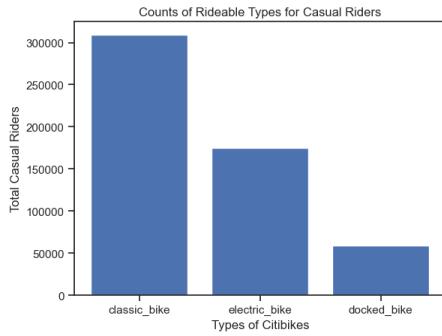
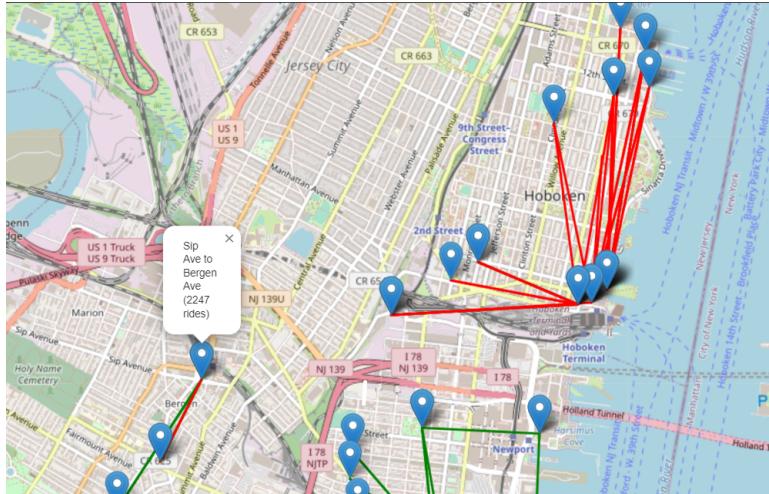
throughout New York City, is connected to the Grove Street Path station. The Hudson-Bergen Light Rail system, which also includes stops in New York City, is connected to the Newport and Hoboken Terminal stations by these lines. Due to this, these stations are well-liked by CitiBike members who must travel to New York City for employment or other purposes.

D. 3.) Top 25 Popular routes taken by Citi Bike Riders :



When we compute the most popular routes it affirms the observations that we have deduced thus far. We observe that stations in and around Grove Street Path, Hoboken Terminal, Newport and Columbus drive are the most utilized starting/ending stations. Most of the routes are converging to these stations as they are the main access to different transportation systems that might want to use to go to New York.

D. 4.) Popular routes and CitiBike Types distribution for casual users and users who have a Citibike Membership.



The popular routes for casual riders are marked in red whereas the popular routes for members are marked in green. Instantly, we see a difference between the popular routes segregated by the types of riders. We observe that the routes north to the Hoboken Terminal are the ones that are most utilized by the casual riders whereas the routes in Jersey City are utilized by riders who have Citibike membership. [5] The difference can be explained as Jersey City is home to the financial district Exchange Place , with companies near Newport and Harborside which means that with residential areas it is home to a lot of corporate companies , organizations and offices. Due to which, CitiBike commuters need to do their daily up/downs to workplace and home, consistently tending for riders in this area to get a Citibike membership which may prove to be cheap and efficient compared to cabs or other transportation. Now while Hoboken city also consists of many businesses and corporate offices, we saw from the previous analysis that people tend to use the Hoboken Terminal for transportation services mainly connecting to the City of New York. These observations can be further affirmed as we can see that the busiest location in Hoboken is situated at the coast of the hudson river while the busiest location in Jersey city is located at the center. Also when we compare the graphs for Citibike user types for different types of Citibikes we can see that casual riders tend to use the electric bike more when compared to the riders who have a CitiBike membership. A good assumption for this might be that as they are not committed to the long-term membership costs that CitiBike users pay, casual riders may be more ready to pay the additional expense for an electric bike. Finally, it's also conceivable that CitiBike users prefer the traditional bike over the electric bike out of pure preference. This might be the case if users are more seasoned bikers or prefer the workout provided by pedaling a traditional bike. However, we do not have the data to back this point.

E.) Limitations and Further research needed :

There were a few limitations that I faced during this study. First, there is a lack of research or analysis done on CitiBikes for New Jersey and Hoboken City. Without this data, it may be challenging to decide how to improve the service, focus marketing efforts, and pinpoint potential growth areas. The absence of statistics may also make it difficult to compare Hoboken and New Jersey's CitiBike usage trends to those of other cities or areas that have more established bike-sharing programs, which makes it harder to draw conclusions that are more broadly applicable.

Additionally, an accurate trip duration value of each Citibike ride in New Jersey and Hoboken City would've been quite helpful to understand behavioral patterns between Casual riders and riders having a membership .Accurate trip duration data can be used to spot trends in when and how individuals use the bikes, including peak usage hours and routes. By doing this, CitiBike can improve the way it distributes bikes and plans routes so that they are always available where and when they are most required. We could've also been able to answer the question if casual riders tend to take shorter trips, based on the answer we can derive insights which can be useful to CitiBike to offer shorter rental periods at a lower cost to attract more casual riders.

F.) Discussion and Conclusion :-

The report and discussion successfully manages to answer the following research questions:-

What are the busiest bike rental locations in New Jersey and Hoboken City and how do usage patterns change over the day, week, and season?

According to the analysis, weekday mornings (7 to 9 am) and evenings (5 to 8 pm) are the busiest times for using CitiBikes, perhaps as a result of commuters going to and from work. The morning peak is lower than the evening peak, which is an intriguing pattern that can be attributed to post-COVID flexible hours and remote work arrangements. Due to the fact that most individuals are away on weekends, peak usage periods differ from those on weekdays.The Summer and Autumn seasons see the peak of Citibike riders. Grove St. Path, Hoboken Terminal, Newport Pkwy, among other stations, are the busiest. Due to their locations in densely populated areas with easy access to multiple modes of transportation, Grove St. Path and Hoboken Terminal are well-liked destinations for commuters and residents, making them perfect locations for CitiBike users.

Are there any demographic factors/patterns that are visible when considering the Age and Gender of the riders utilizing the CitiBike transport system ?

By breaking down the total trips depending on the riders' gender and age, the investigation examined the demographics of CitiBike users. According to the findings, there are around three times as many male as female riders, and the average age of female users is 38 compared to 40 for male users. This implies that individuals of various age groups use CitiBikes to get to work, school, and college. Three different bike kinds are available from CitiBike: classic bikes, electric bikes, and docked bikes. According to the data, electric bikes are used the most frequently by users, while docked bikes are used the least frequently. By constructing a scatterplot matrix of the users' ages and preferred CitiBike kinds, the study also investigated the relationship between age and bike preference. According to the findings, people over 35 prefer to ride traditional bikes, while kids between the ages of 10 and 12 choose electric bikes. This might be as a result of the younger age group preferring the speed, better mobility, comfort, and convenience of the electric bikes, which take less effort to ride.

Are there any patterns or differences that can be observed between users or riders who are of two categories that are casual and members ?

Yes we can see two different trends when observing the casual riders and riders with membership based on the most popular stations that they use and the most popular routes that they take. In contrast to members who utilize routes in Jersey City, which is home to the financial area and corporate buildings, casual riders typically take lines north to the Hoboken Terminal. This indicates that CitiBike membership is more appealing to commuters who must regularly travel to work. Whereas the busiest spot in Hoboken is near the Hudson River's shore, the busiest spot in Jersey City is in the city's core. Additionally, non-CitiBike members utilize electric bikes more frequently than casual users.

Conclusion: Notwithstanding the restrictions and difficulties encountered throughout this study, we have learned important information on the usage patterns of CitiBike in Hoboken and New Jersey. The survey has provided information on the popularity of specific CitiBike stations and the most popular bike types among various user categories. Also, we have noted prospective areas for development, such as the requirement for precise trip duration data, which can aid CitiBike in better comprehending customer behavior and streamlining its offerings. To present a more thorough picture of CitiBike utilization in Hoboken and New Jersey and to make comparisons with other cities and regions that have created bike-sharing programs, additional research and analysis are required. Overall, this research lays the groundwork for further investigation and can assist in decision-making by CitiBike and other bike-sharing programs in the region.

References and Citations :-

- [1]. "Big Blue Comes to Jersey City: An Analysis of Citi Bike Expansion into New Jersey." *TRID Database*, 1 April 2017, <https://trid.trb.org/view/1438465>.
Accessed 1 May 2023.
- [2]. Schwartz, Sam. "Exploration Of A Public Bike Share Program In Hudson County." *Hudson County*, https://www.hcnj.us/wp-content/uploads/2021/09/Pages-from-Hudson_Co_Report_FINAL_150526red_Ch1-3.pdf.
Accessed 25 April 2023.
- [3]. Pojani, Dorina. "(PDF) Weather and cycling in New York: The case of Citibike." ResearchGate, 8 May 2019,
https://www.researchgate.net/publication/332932885_Weather_and_cycling_in_New_York_The_case_of_Citibike.
Accessed 10 April 2023.
- [4]. "Index of bucket "tripdata."" *Amazon S3*, <https://s3.amazonaws.com/tripdata/index.html>.
Accessed 5 April 2023.
- [5]. Baer, Marilyn, and Daniel Israel. "Business and commercial development." *Hudson Reporter*, 10 June 2018,
<https://hudsonreporter.com/2018/06/10/business-and-commercial-development/>.
Accessed 5 May 2023.

