Citi Bike Analysis

Alexey Pilyugin

Citi Bike Project

1

Number of Bikes Hired : 32,361
Period : 9/1/2013 12:00:00 AM to 9/30/2013 11:54:00
Average Time per Trip : 739.7
Average change in Latitude : 0.0004458
Number of Bikes Hired : 10,658
Period : 9/1/2013 12:09:00 AM to 9/30/2013 11:58:00
Average Time per Trip : 864.4
Average change in Latitude : 0.0002932

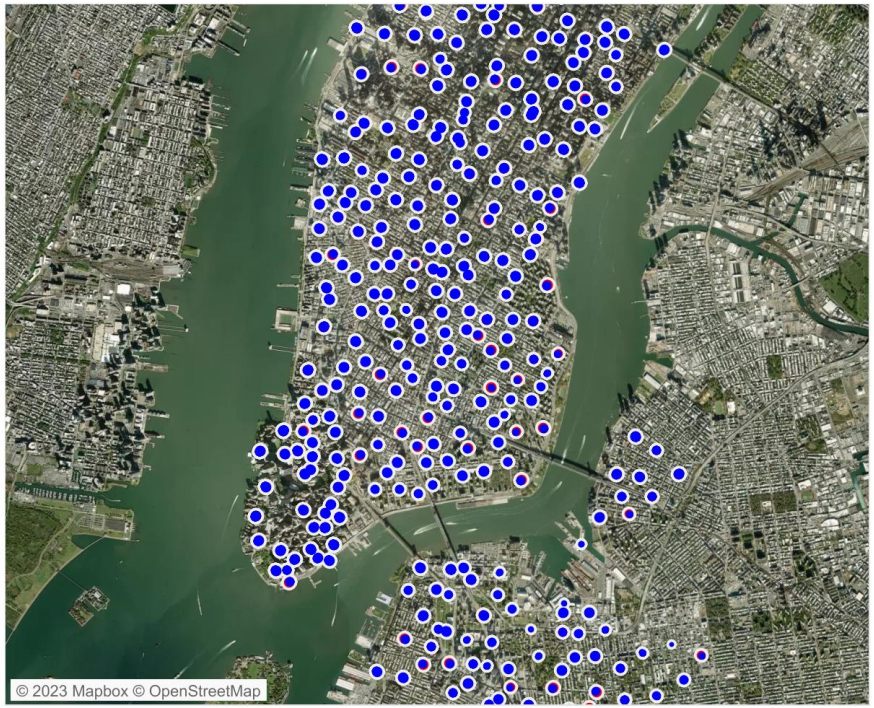
Trip Duration
60
500
1,000
1,500
2,000
2,697

Gender
1
2

Gender
1 2

Age
16 to 114
and Null values

Startpoint



Introduction

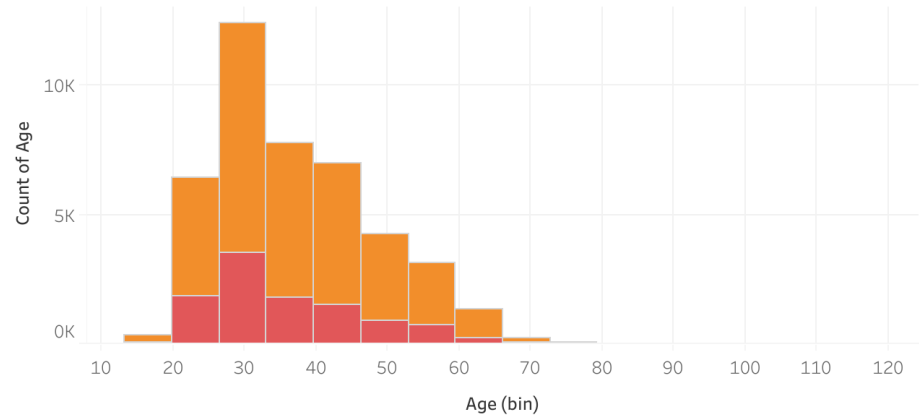
Citibike, a public bicycle sharing system serving the New York City area, has grown to become an integral part of the city's public transportation landscape. As urban areas continue to explore sustainable and convenient transportation options, analyzing Citibike usage patterns provides valuable insights into urban mobility and commuter behavior.

This report aims to analyze a dataset from Citibike to uncover various aspects of rider behavior and system usage.

Age and Trip Duration Distribution

2

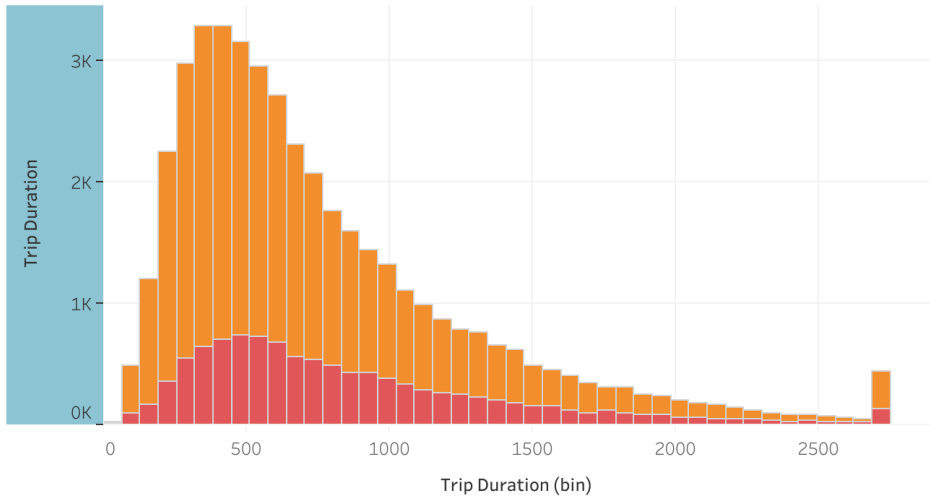
Histogram of Age



Gender
1
2

Age
16 to 114
and Null values

Histogram of Trip Duration



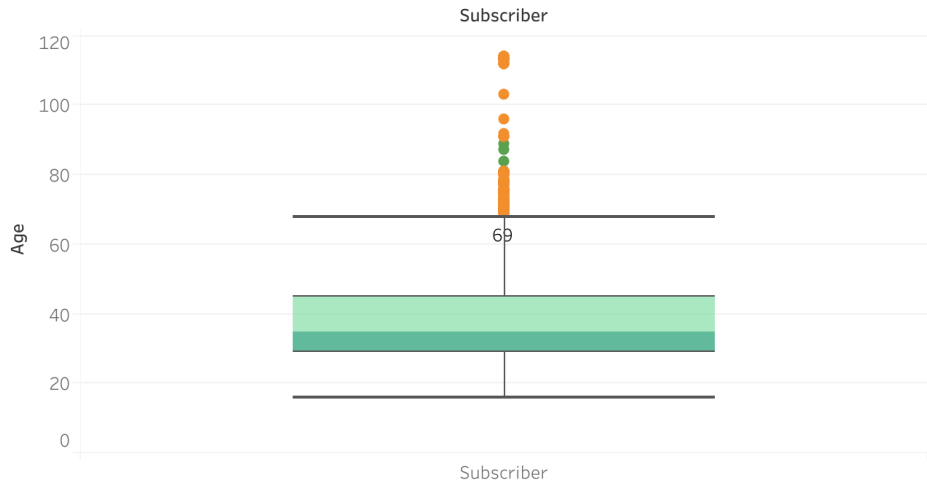
The age distribution of Citibike users, reveals distinct patterns for male and female riders. For both genders, the *majority of users* fall within the age range of **25 to 45 years**, showcasing a pronounced peak in this interval. While the distribution for males and females follows a similar trend, males exhibit a slightly broader distribution, indicating a more **diverse age range** among *male* riders. Both distributions exhibit a right-skewed pattern, indicating a gradual decrease in the number of users as age increases, with fewer users in the higher age brackets.

The distribution of trip durations indicates that the *majority* of trips for both genders are of **shorter duration**, with a high frequency of trips lasting **less than 1000 seconds**. The distribution exhibits a sharp decline as the trip duration increases, indicating that longer trips are less common among Citibike users. Both male and female users follow a *similar* pattern in trip duration, with the distributions being right-skewed, signifying a higher concentration of shorter trips and a tail extending towards longer trip durations.

Subscriber Status

3

Boxplot Age by Subscribers

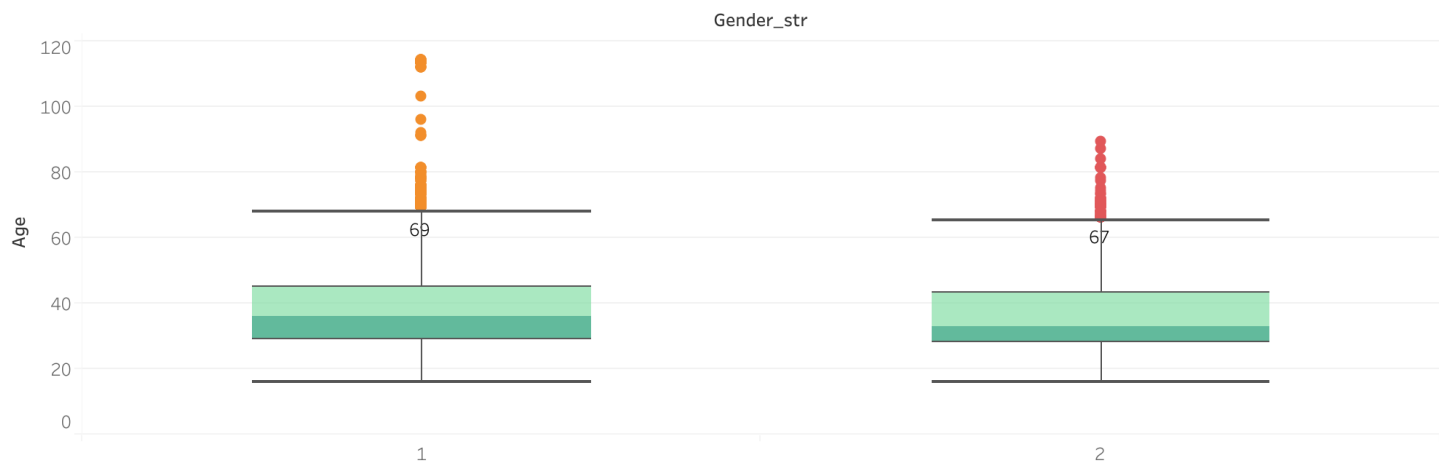


Gender
1
2
Age
16 to 114
and Null values

Both genders exhibit a number of outliers, primarily on the higher age end, highlighting the presence of older users in the dataset.

Subscribers exhibit a lower median age and a broader interquartile range this could imply that **younger individuals** are *more likely to* subscribe, possibly due to factors such as commuting patterns, fitness habits, or pricing considerations.

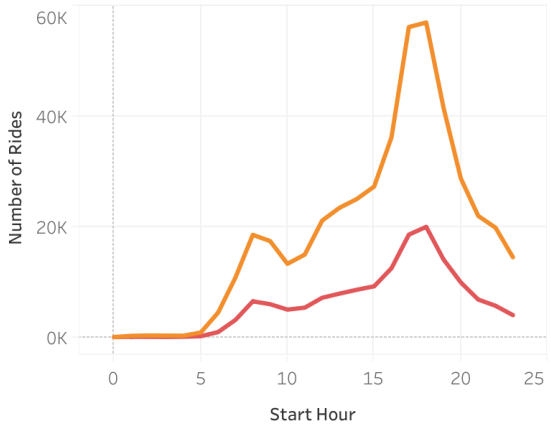
Boxplot Age by Gender



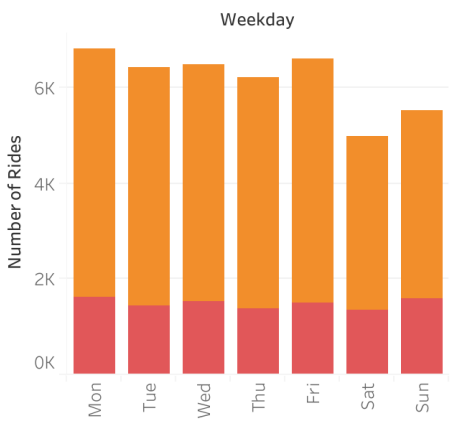
Ride Frequency

4

Ride per Hour



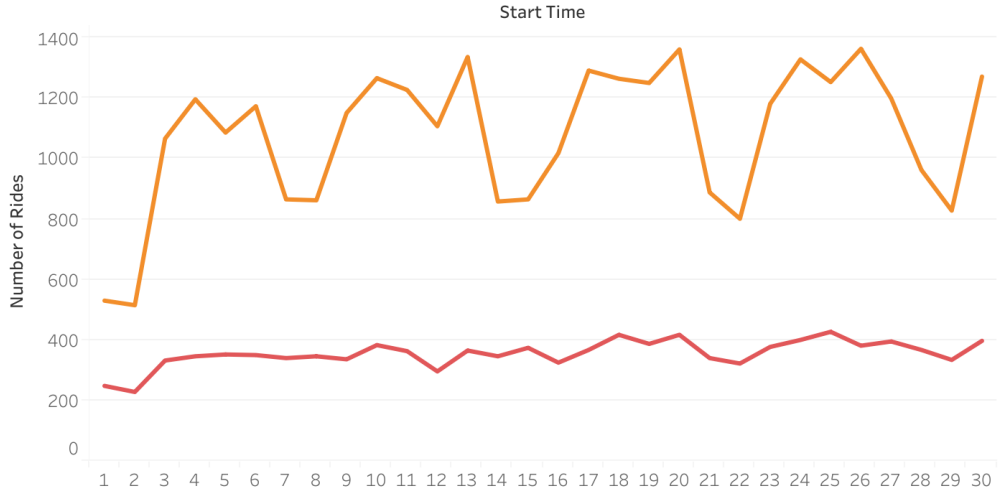
Ride per Weekday



The bar plot *"Ride per Weekday"* reveals a variation in the number of rides across different weekdays. The weekdays, particularly **Tuesday through Thursday**, exhibit *higher* ridership compared to the *weekends* (Saturday and Sunday). This trend may suggest that Citibike is primarily used for commuting purposes during the workweek, with decreased usage for leisure or other activities on the weekends.

The line plot *"Ride per Hour"* showcases distinct *peaks* during the morning and evening hours, typically around **8-9 AM** and **5-6 PM**. This **bimodal pattern** aligns with traditional commuting hours, reinforcing the hypothesis that Citibike is heavily used for work-related travel. The troughs in the mid-day and late-night hours indicate lower ridership during non-peak times.

Number of Rides in per day in September

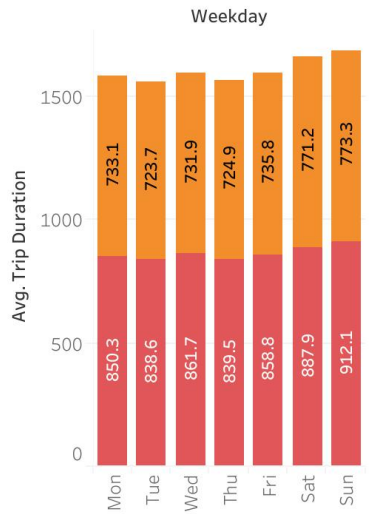


The *"Number of Rides in per day in September"* plot displays fluctuations in the number of rides, with *some days* experiencing **higher ridership than others**. While the plot provides an overview of the daily variations, additional contextual information (such as weather conditions, holidays, or special events) would be helpful to explain specific peaks or troughs. A more detailed analysis could explore patterns related to seasonality, weather, or other external factors influencing daily ridership.

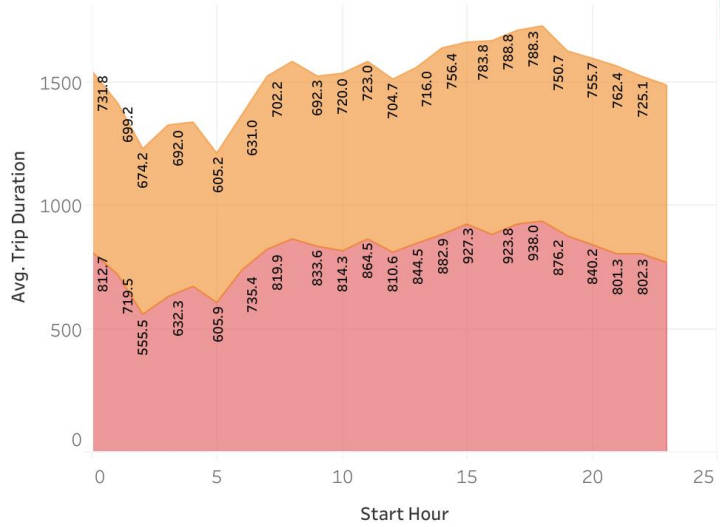
Time Distribution

5

Duration by Weekday



Average Trip Duration per Hour

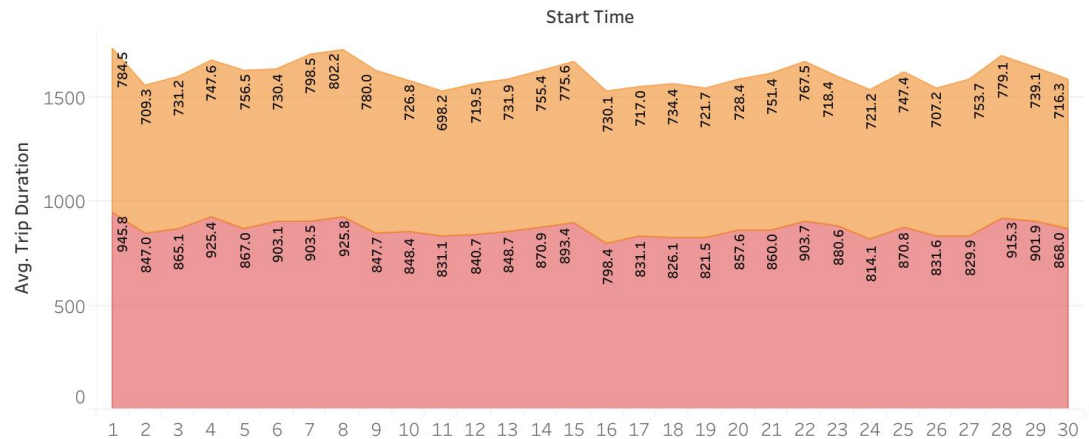


Gender
1
2

The bar graph titled *"Duration by Weekday"* depicts the average trip duration for each day of the week. It is observed that the average trip duration is relatively *higher* during the **weekends**, specifically on Saturdays and Sundays, compared to the weekdays. This suggests that users tend to take longer trips on weekends, possibly for recreational activities or leisure, whereas, during the weekdays, the trips are likely shorter and more utilitarian, potentially for commuting. This pattern is common for both male and female riders.

The line plot titled *"Average Trip Duration per Hour"* showcases the mean trip duration across different start hours of the day, separated by gender. For both males and females, there is a noticeable *increase* in average trip duration during the **early morning hours**, followed by a *decrease* as the day progresses. This pattern might be attributed to recreational or exercise-related trips taken in the early hours. The average trip duration then exhibits a *slight increase* in the **evening hours**, potentially corresponding to leisure trips or non-commute related activities. Similar to the daily plot, the hourly plot also does not show a significant difference in trip durations between genders.

Average Trip Duration per Day



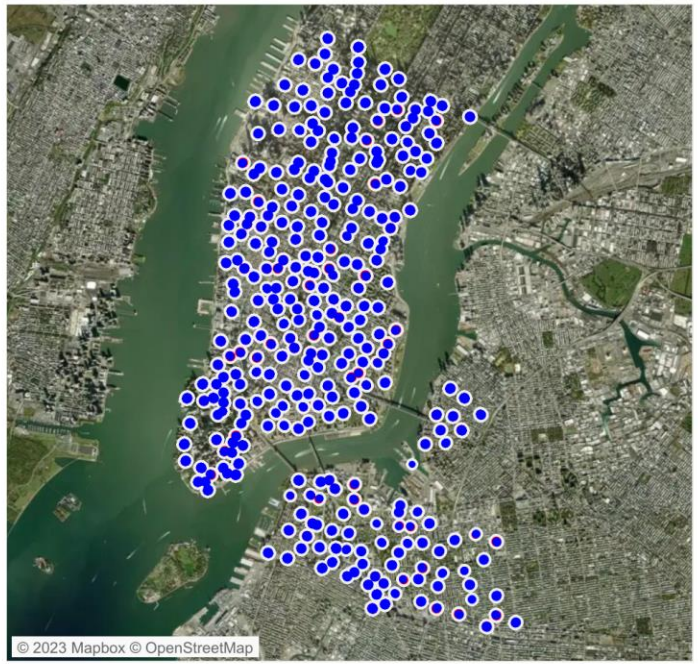
The line plot titled *"Average Trip Duration per Day"* is grouped by gender, and reveals a fluctuating pattern for both males and females over the observed dates. **Both genders** exhibit similar trends, with peaks and troughs occurring around the *same dates*, indicating that external factors, such as weather conditions or holidays, might infl..

6 Conclusion

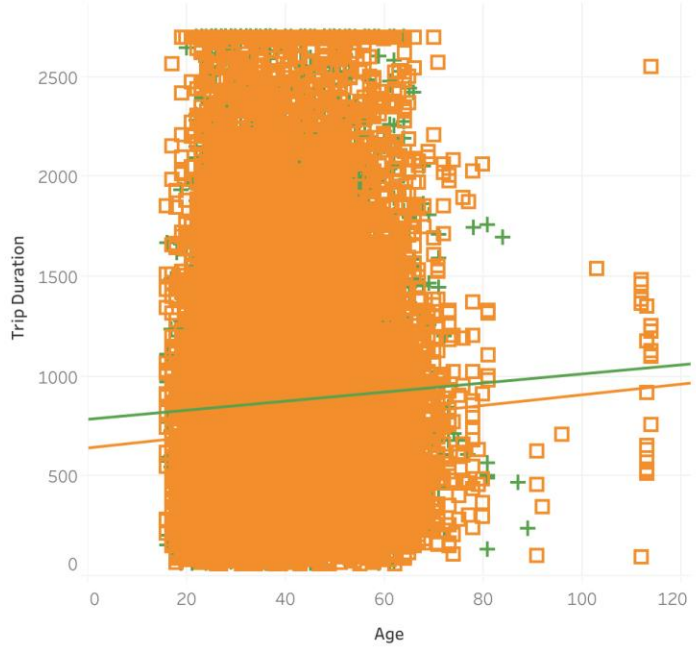
The *"Scatter plot of Age and Trip Duration"* plot depicts a subtle **negative correlation** between *age and trip duration* for both male and female users of CitiBike, suggesting that **younger users** tend to embark on marginally *longer trips* compared to their older counterparts. However, the extensive spread of data points indicates significant variability, implying that while age may influence trip duration, the relationship is not pronounced, and other factors likely contribute to the observed diversity in trip lengths.

In conclusion, the analysis of the Citibike dataset reveals insightful patterns about urban mobility and commuter behavior in New York City. The dataset indicates a diverse user base, predominantly aged between 25 and 45, with both genders exhibiting similar trends in usage and trip duration. The service experiences heightened ridership during weekdays and traditional commuting hours, suggesting a significant role in facilitating work-related travel. The presence of a subtle negative correlation between age and trip duration, along with variations in average trip duration across different times and days, highlights the multifaceted nature of user engagement with the system.

Endpoint



Scatter plot of Age and Trip Duration



7 Recommendations

1. Target Age-Specific Marketing: Due to distinct age patterns, with the majority falling between 25-45, and less diverse female usage, Citi Bike should target various age groups and genders. Promotions can emphasize convenience for youth and the health benefits of cycling for older riders to expand the user base.

2. Weekend Leisure Packages: To leverage the observed longer weekend rides, Citi Bike can offer special leisure packages, partnering with local attractions and businesses to provide weekend discounts, encouraging riders to explore the city leisurely

3. Real-time Data Integration: It was observed that ride frequency fluctuated throughout the month of September, this may have been due to external factors. To enhance user experience, and decrease unpredictability, Citi Bike can integrate real-time data like weather, traffic, and events into the app, providing riders with timely information for better trip planning.

4. Introduce Dynamic Pricing for Peak Hours: Citi bike could use dynamic pricing strategies to adjust the cost of rides during peak hours/ locations. They could do this by increasing prices slightly during high-demand times (early morning, evening rush hours).

5. Target Non-Commuter Audiences: To boost off-peak revenue, Citi Bike should target non-commuters by promoting our service as ideal for evening entertainment, offering discounts and incentives during those hours.