Citi Bike trip data Analysis

The data used for this project was all the observations from January 1st, 2020 to January 31st, 2021, this data was collected on a CSV format from the original “Citi Bike” data collector. These files were read and joined to consolidate all the data in one data frame, a total of 20,602,202 cells formed the overall data.

The first step into data cleanup was to change the columns that involved time data to a “datetime” , this was made for “starttime” and “stoptime” columns. A look for “NA” or “null” values was made to ensure the data was complete. The third step was to eliminate any outlier in the data, for this the calculation of the quantiles was made to the column “tripduration” and with this a calculation of the Tinter Quartile Range and Boundary Values took place resulting in the next values: IQR= 967, LBV=-1032.5, UBV= 2835.5. With this information all the data that was over the UBV was removed from the study. The resulting data frame was exported in a CSV format to be read by “Tableau Public”.

## Tableau Data Visualizations.

**Total Recorded trips:** Count of all the trips in the Data.

**Ridership Growth:** This visualization is plotting the total trips made in each month of the study. It´s clear that the usage of bikes peaks on the spring and summer months.

**User Type Analysis:** Total trips made by month with the separation of the User Type that makes the trip. There is a clear difference on the utilization of bikes, customers seems to use the service more than the subscriber, maybe this is the result that the bikes are being used as a tourist transportation option, with this information a reaccommodation of the resources (bikes) could be made to have more availability on the tourist sites of NYC.

**Top stations to start:** Shows a list of the top stations were a trip begins

1. 2006 – Central Park S & 6 Ave
2. 499 – Broadway & W 60 St
3. 3423 – West Drive & Prospect Park West
4. 3881 - 12 Ave & W 125 St
5. 514 - 12 Ave & W 40 St
6. 426 – West St & Chambers St
7. 281 – Grand Army Plaza & Central Park S
8. 3724 – 7 Ave & Central Park South
9. 3282 – 5 Ave & E 88 St
10. 3374 – Central Park North & Adam Clayton Powell Blvd

**Top stations to end:** Shows a list of the top stations were a trip ends

1. 499 – Broadway & W 60 St
2. 3423 – West Drive & Prospect Park West
3. 426 – West St & Chambers St
4. 2006 – Central Park S & 6 Ave
5. 514 - 12 Ave & W 40 St
6. 3881 - 12 Ave & W 125 St
7. 3165 – Central Park West & W 72 St
8. 217 – Old Fulton St
9. 3724 – 7 Ave & Central Park South
10. 3137 – 5 Ave & E 73 St

**Gender Participants over time:** This plot shows the relationship between the gender information of the users over the months in study. The graph shows that both genders have similar behavior over time, with “female” having more participation in the trips.

**Trip Duration per Age:** This bar chart shows the user average time of a trip, for this the information was separated into 7 groups in bins on 10 years, starting from 20 and under the consideration that since it´s acknowledged that most of the users are using the bikes for tourism transportation, a user under 20 years will be considered that this user is accompanied by an older family member. Interestingly the group of 20 years or less has and the group of 51-60 has a similar time trip average, this matches the hypothesis that users under the age of 20 years are accompanied by an older user.

**Average Distance per Trip:** For this chart an assumption of an average velocity of 5 mph for a bike trip on NYC ( <https://www.bicycling.com/news/a28928975/bikes-or-cabs-faster-in-city/#:~:text=Which%20Is%20Faster%20for%20Getting%20Through%20the%20City%3A%20Bikes%20or%20Cabs%3F&text=According%20to%20the%20Mobility%20Report,go%20about%2030%20percent%20faster.>), with this information an estimated distance per trip was calculated and grouped this information per bike id. Adding the amount of times that a particular bike ID started from the Top 10 stations. With this information it´s possible to address the bikes that requires maintenance. It´s assumed that maintenance is needed on any bike that has over the 75% quantile of the total trip distance.

**Top 10 Starting / Ending Stations Map:** Map showing the top 10 Starting / Ending stations.