**Citi Bike Analytics**

The Citi Bike Program in New York and Jersey City has implemented a robust infrastructure for collecting data on the program's utilization beginning in 2013. Through the programs efforts, each month bike data is collected, organized, and made public on the [Citi Bike Data](https://www.citibikenyc.com/system-data) webpage. One of the shortfalls of the program is that have not taken the time to completely analyze the data as it is collected. Therefore, the officials have asked that several questions of the program, so my task is to build a set of dashboards along with a story to answer their questions.

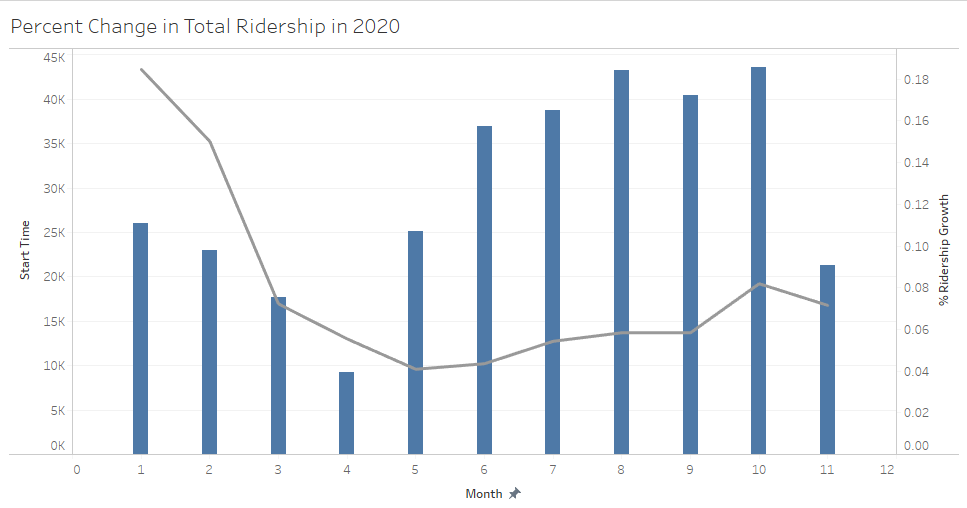
Citi Bike Trip History data for Jersey City was collected for these reports as CSV files. The timeline used for this analysis is January 2020 – November 2020. The following steps outline the Data Extraction, Transformation and Loading procedures and tools utilized in this project:

* To streamline the process, instead of saving files, I wrote python code to scrape the Citi Bike website for the files within the timeframe set. After the site was scraped for the files and combined into a single dataframe, it was then cleaned to be more manageable.
* Nulls were dropped and an extra column was added that represents the age of each rider. Dates and times for the “starttime” and “endtime” were stored as timestamps.
* Once the csv’s were joined into one dataframe and cleaned, one csv was saved locally and then uploaded to the Tableau workbook.

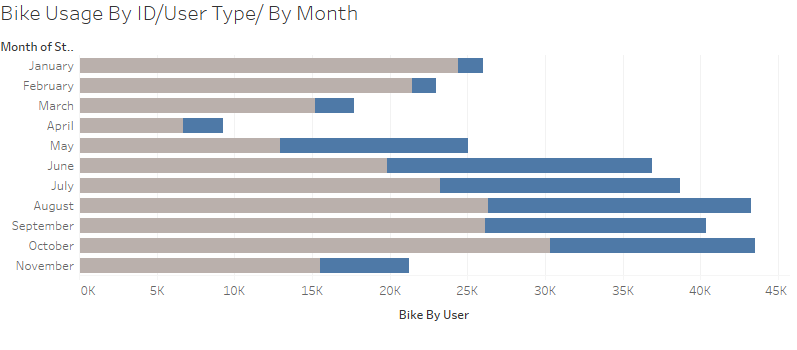
**Analysis and Conclusions:**

Below is just a few of the questions that were answered in the story that is accompanied with the two dashboards. The questions below are a few of the phenomenon that I found the most interesting while conducting my research on the Citi Bike data for Jersey City in 2020.

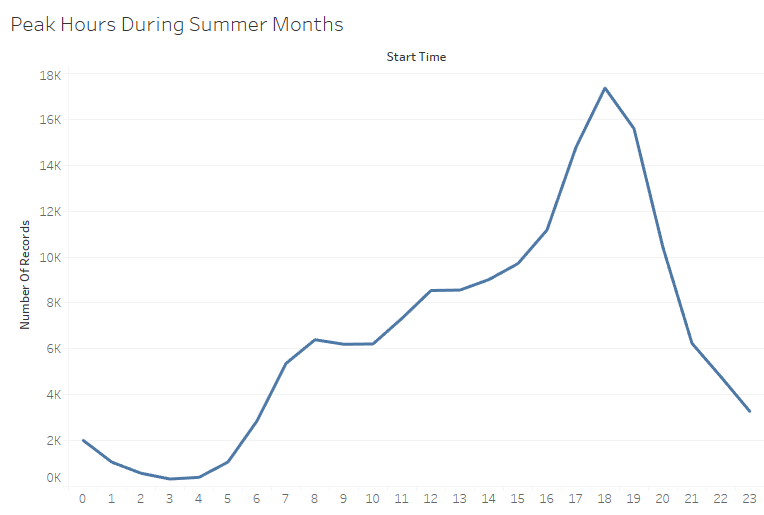
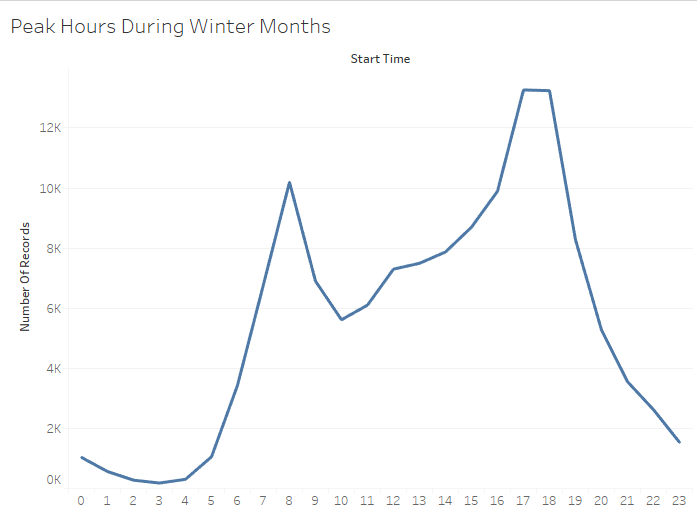
* How many trips have been recorded total during the chosen period and by what percentage has total ridership grown?
  + There were at total of 666,061 trips taken during the January 2020 - November 2020 time period. In the months analyzed, 325,083 total trips were taken. Having only picked one year, to answer the question of rider growth, the data was broken down by the month. As can be seen in the visuals, rider growth has grown over 60% throughout the year. I believe that the rise in ridership could be due more to the warmer months. Additional analysis by comparing two to three years would better show a trend of rider growth.



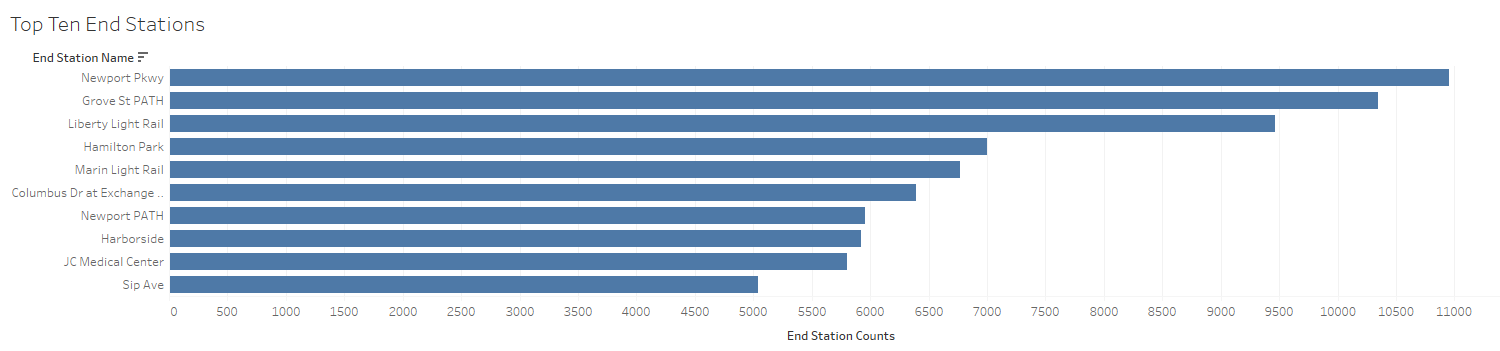
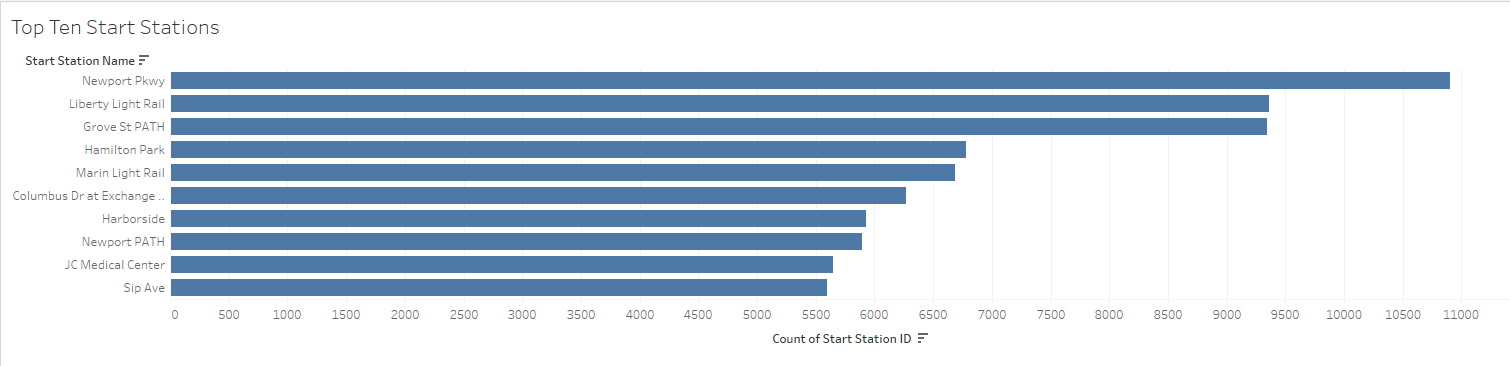
* How has the proportion of short-term customers and annual subscribers changed?
  + It appears that the share of short-term customers and annual subscribers tends to fluctuate based on the time of year and season. Although subscribers make up the largest share of users, during the Spring and Summer the total number of short-term customers tends to increase. During January, February and March subscribers make up over 90% of total users. However, perhaps as the temperature and weather improves, the total share of customers increases, and the share of subscribers drops slightly. It appears that outreach is critically important during warmer months to optimize the increase in short-term customers.



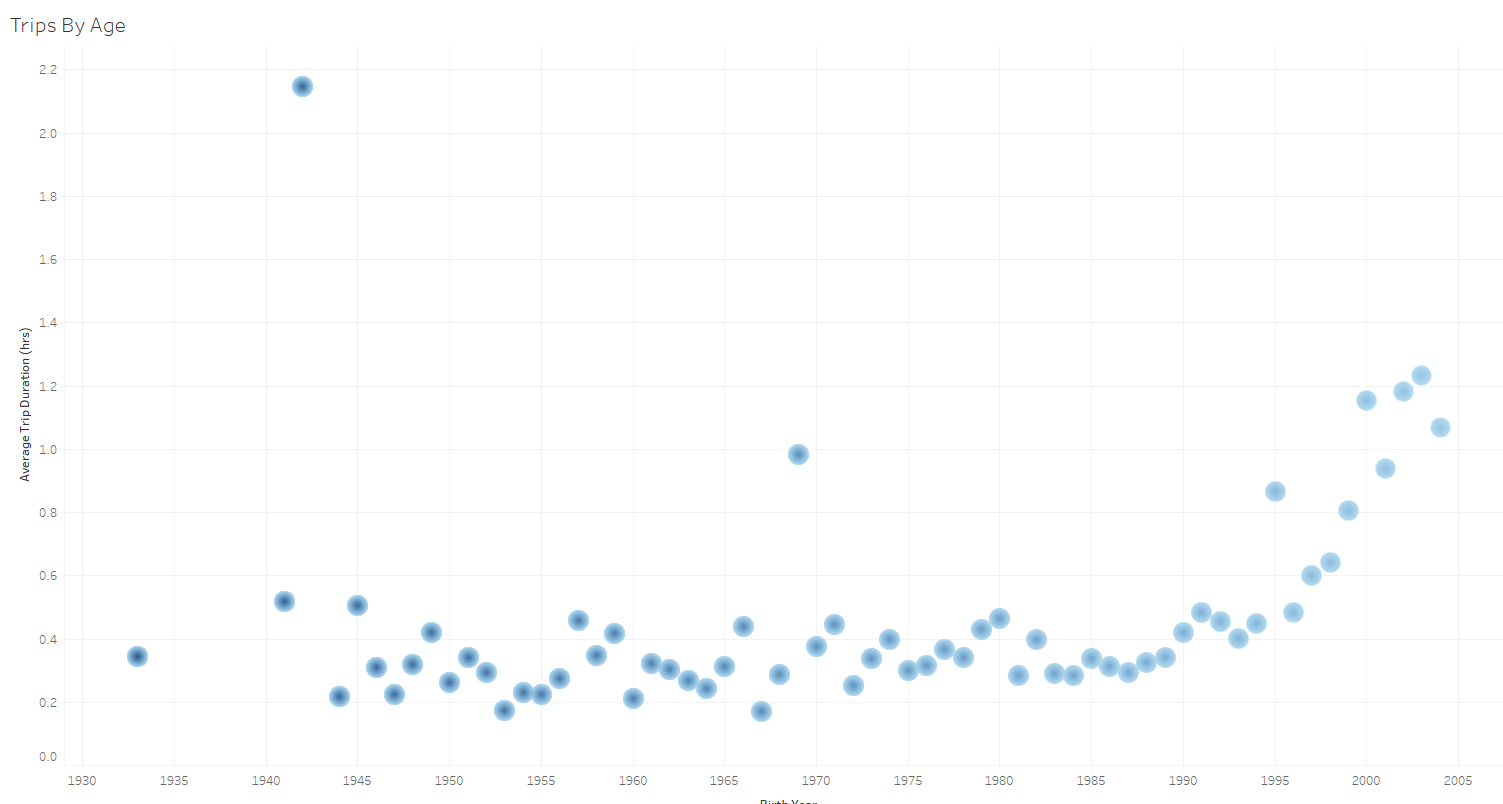
* What are the peak hours in which bikes are used during summer and winter months?
  + Although peak hours for bike usage during the winter and summer appear to be in the mid-morning and mid-afternoon. The data shows that the peak times begin to increase at the time of the morning that would represent travel to work and then peaks again when people would typically return from work.



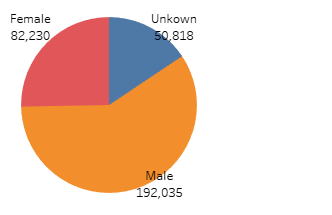
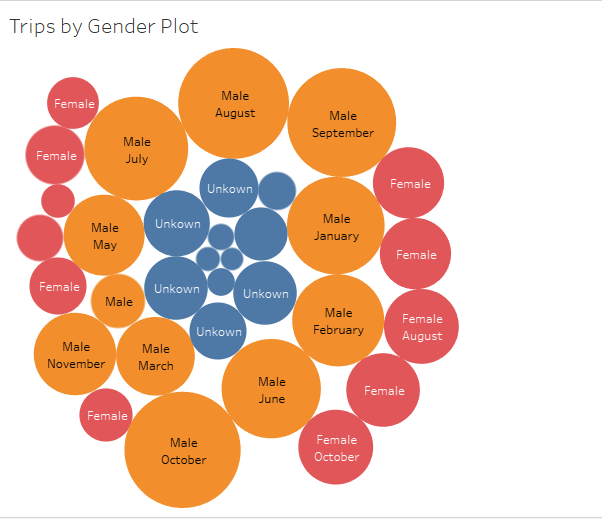
* + Today, what are the top 10 stations in the city for starting and ending a journey?
    - Many of the same stations are the most popular to start and end a journey. You can see by the dashboard maps that all the top ten stations to end and start are in an area of lesser income per household. This makes total sense; lower income families rely more on public transportation than higher income families.



* How does the number of total trips and trip duration change by age?
  + Age does appear to have an impact on the total number of trips and their duration. Individuals in their late teens and early 20's tend to take the most and longest trips. However, two phenomena that I discovered during my analysis is that 51- and 71-year old’s actually have the highest and longest number of trips. There is a sharp increase in total customers at this age that I find curious. Due to the duration and number of trips, these values are clearly outliers but is it due to actual involvement in the bike share program or were there issues for these customers. More research should be conducted on these two age groups to see if the values are correct.



* Today, what is the gender breakdown of active participants (Male v. Female)? How effective has gender outreach been in increasing female ridership over the timespan?
  + Although there does appear to be some improvements in the increase of female ridership over time, males still make up the largest share of users. Further progress could be made to ensure that active female participants and total ridership between both genders continues to increase. Additionally, I think that a further look into this over a larger period of time could yield a better understanding of this phenomenon. A study should also investigate why ridership in women is not as high as men. Maybe there are factors that are not considered that influenced female ridership for Citi Bike. It could be as simple as safer bike lanes, lighter bike weight, and more convenient trip times that are affecting female ridership.



* Which bikes (by ID) are most likely due for repair or inspection in the timespan?
  + Currently, there are quite a few bikes that are likely due for repair and inspection. Based on total time duration over the 11 month period, there are 10 bikes that stand out. The bike ids and total trip duration are as follows:

