**Citi Bike Analysis:**

I have used two data sets to do the visualization and analysis.

**JC Citi Bike:**

For the purpose of this visualization, I decided to use the dataset from the year 2017-01 to 2019-05. The other files were too big and greatly reduced my performance issues. The reason for selecting data starting at 2017-01 is because I wanted to do forecast and wanted to show seasonality while identifying trends.

I downloaded the data from Citi Bike website and combined them using Pandas.

**Membership:**

The purpose of this dataset it to see the 3-day and 24-hr passes bought by riders.

**Citi Bike Visual Analysis Using Tableau.**

**JC Citi Bike Findings:**

**No. of rides recorded:** The total number of records collected from 2017-01 to 2019-05 is 779,858. However, there seems to be some records that are missing. These are Birth Year and some start station ID. It is understandable that people usually do not like to give their ages

**Total Ridership:** The number of rides peaked during August and October. There is also a seasonality of rides peaking during the warmer days.

**Usertype Proportion:** There is definitely a big gap between subscribers and customers. There does seem to be a consistent proportion along the years though.

**Top Bottom Stations:** The top start and end station is the Grove St. Path. The number of rides at this station is also very high compared to other stations. This is the Port Authority of NY&NJ.This makes sense.

**Gender Breakdown:** The number of male riders (71.5%) is overwhelmingly higher than female riders (21.31%)! The gender outreach program doesn’t seem to have worked because female riders don’t show a spike. They follow the same trend as male riders.

**Trip vs Age:** 25-45-year age group makes for most of the riders. After that age group there are fewer riders upto the age of 75. There a lot of outliers in this graph. This is probably because people don’t like to disclose their ages and must have entered incorrect year of birth.

**Peak Hours:** The peak hours during both summer and winter are 6-8:30 pm. But the number of rides is much less in the winter days. Also, the duration of trip in winder are lesser than those during summer months.

**Forecast:** The numbers of riders are expected to be up and show the same pattern for the rest of 2019.

**Membership:**

The average 24-hour passes are increasing year over year whereas the average 3\_day passes are decreasing year over year.

**Data Limitations:**

It would have been interesting to see the revenue generated by subscribers and customers. We could compare the percentage of revenue generated by customers and subscribers to the whole. That would throw some light on whether subscribers are using more rides just because they have a subscription and are trying to get much out of their subscription fees.

This would also help to analyze if the revenue are growing proportionately to usage.

Age = There were a lot of birth years missing. There were some anomalies where ages came to be more than 100!

It would have been interesting to find out the usage during days of the week. Like weekdays vs weekends but unfortunately the data only had month and year.