*Project Description*

I was tasked with the project of performing data analysis on Citi Bike data from their website <https://www.citibikenyc.com/system-data>. The main story can be found on my Tableau project “Jersey City citibike use” under the story title of “Jersey City Citi Bike Story”: <https://public.tableau.com/profile/diane.scherpereel#!/vizhome/JerseyCitycitibikeuse/JerseyCityCitiBikeStory>

*Collecting data*

I downloaded all the data from the Citi Bike website but realized when trying to join all the files in SQL that this was too much data to work with. I was wondering what “JC” meant in some of the file names and thought it might be Jersey City. Putting some of the latitudes & longitudes from those JC files into Google Maps showed that they were from Jersey City so I joined all the JC files using SQL and pulled that csv file into Tableau. Plotting all the start stations in Tableau confirmed that every start station was in New Jersey (See tab “Jersey City Start Stations”). I also verified that this JC data was NOT duplicated in the same time-frame files that do not include “JC” in their titles.

*Phenomena 1*

I was wondering whether people are allowed to take bikes in the Holland Tunnel to get from New Jersey to New York so I searched that on Google and found that it is allowable. Then I plotted all the end stations from these JC files and found that there are many stations in New York that these bikes, originating in New Jersey, were returned to. This can be viewed in Tableau on “Jersey City End Stations,” and both the start stations listed above in *Collecting Data* and these end stations are shown in the “Jersey City Citi Bike Story.” For these plots, I used all the data from 2015 – 2019. Citi Bikes started in 2013, but it wasn’t until 2015 that they added the stations in Jersey City.

*Phenomena 2*

The summary of phenomena 2 can be found in the story titled “1969?” on Tableau and is also included the “Jersey City Citi Bike Story.”

By making bar graphs of the number of subscribers compared to customers, the data shows that there are more subscribers (annual members) than customers (24-hour pass or 3-day pass users) and that subscriber’s average trip durations are shorter than customer’s average trip durations while the subscriber’s sum of trip durations are larger than the customer’s sum. (See “Customer vs. Subscriber Trip Duration Dashboard” or individual sheets: “Customer vs. Subscriber Avg Trip Duration” and “Customer vs. Subscriber Sum Trip Duration.”) This makes sense if subscribers tend to use the bikes to commute to work (shorter duration on each trip but taking more trips) compared to customers, like visitors to Jersey City, for example, who may be using the bikes for recreational use (longer individual trips but shorter total sum).

When adding the birth year (age) to the analysis of subscribers and customers, the birth year 1969 stands out on the “Customer vs. Subscriber by Age” Tableau sheet, the “Trip Duration-Sum by Age” sheet, and to a slightly lesser extent on the “Trip Duration-Avg by Age” sheet.

Capturing the details on “Birth-Year-1969 long trips” reveals that one customer took the bike for 11 days (the 6th to the 17th). Another person of the same birth year took a bike for 6 days (21st to 27th), and a third single rider of the same birth year took the bike for 5 days (22nd to 27th). It can’t be determined from this data how many miles they were riding the bike but hovering on the blue box for the customer who kept the bike for 11 days starting on the 6th of the month, shows the start and ends stations. Plotting those on Google Maps shows that the bike was returned to an end station only 0.4 miles from the start station where it was picked up (See Google Map below).

A close up of a map

Description automatically generated

*Phenomena 3*

While trying different things in Tableau, I also plotted the increase in ridership by year from 2015 – 2019, which has grown, as we would expect since Citi Bikes is still in business. (See “Yearly increase in ridership.”) A similar plot displaying the breakout by month shows that ridership is highest in the summer and early fall. (See “Monthly highs in ridership.”) It makes sense that summer and early fall would be good times to use a bike. From the data in the earlier years, I wondered why spring ridership was lower than fall ridership (compare April & May 2016 to October and November 2016). From 2017 – 2019, October is still higher than April and May, but May is higher than November. I imagine that “April showers bring May flowers” might contribute to higher ridership in fall than in the spring month of April. The dashboard “Increases in ridership” shows how these yearly and monthly increases occur together.

*City Official Map*

For city officials, I was also tasked with plotting popular start and end stations on a map that shows the zip codes. I included both the specific zip codes and the zip code boundaries. For this task I used only the more recent data of 2019. On the individual sheets (“Jersey City Start Stations 2019” and “Jersey City End Stations 2019”) plus on the dashboard titled “City Official Map,” the darker colored circles indicate more popular stations. Hovering over the circle shows the latitude, longitude, station name, and the number of records which is the number of bikes that were taken from starting stations or returned to ending stations. The most popular start station is the same as the most popular end station: Grove St.PATH in Jersey City, so though many stations in New York received bikes that started in Jersey City, the one single end station that received the most bikes originating in Jersey City is also in Jersey City.

*Final Comments*

I would have been interested in exploring more of this data, but Tableau Public would freeze regularly on my computer and required restarting (sometimes just Tableau Public and sometimes having to restart my whole computer) between every save. This continued to happen even after shutting down all other things on my computer, and a classmate using a smaller data set had the same problem. It convinced me that if this were a tool I was using regularly, I would need to purchase one of Tableau’s products rather than relying on the free Tableau Public.