Report on total ridership for Citi Bike Program
by Steven Green

The first thing that needed to be done to the dataset was to clean and narrow data that we could then create visualizations for. The data was taken for the year 2019 represents 5% of the actual total for that year. Meaning that every number in the dataset and visualization could be multiplied by 20 to get the correct amount. I also disregarded any date before 1921 due to the outliers that would be there. Someone turns 100 and wants to be taken on a bike ride is possible just not something that can happen in our dataset a lot.

The first graph which is a bar shows the number of total riders in our clean dataset broken down by gender. The total number is 411,032 broken down to be 280,665 men, 99,201 women and 31,166 unknown riders. The second bar graph shows the increases and decreases of rides by gender throughout the dataset. This graph shows that the summer months increase ridership in all areas but unknown ridership the biggest category from April through October. The third graph is a line graph, and it shows the age of the riders throughout the dataset. This shows that our greatest number of riders who subscribe to our service for males and females are people in their mid to late 20s. A suggestion would be to use this data to create rewards for this group or market to others in this group specifically. My guess is that it's the young urbanites who would like a more sophisticated way to move around NYC. The unknown we will not worry about but notice how the unknown customers have the unknown date. Also notice that we have unknown subscribers, like people just gave us their credit card information but nothing else...we may need to follow up on that one. One way is to reach out to people using their card information to allow them to the option to continue to subscribe. This may cut down on cancellations. The fourth graph is the same information as shown on slide two but just broken down by gender and in a comparative way so that the data can show what increases and decreases we have easier to visualize. Men are the group that most use our bikes in all areas. Women increased every month from February to October. January to February of 2019 showed a decrease in all groups of subscribers. Customers shows a different story as ridership increased during that same period (slightly).

After digging into the dataset, I wanted to check and see if the unknown ridership was just bad data or not. Bad data represents pieces of data that could be overlooked because no discernable pattern can be extracted from them. There are 31,166 in this sample dataset which would mean that over the course of the year there would be over 620,000 people that were overlooked. That's a big number and that's just from one year. I took the top 25 areas that had unknown customerships. I filtered the date to the default year of 1969. And What I found out is that most of unknown customers along 2 areas which are visualized in the following map. I wanted to make a map that shows the data from the previous slide. It's easier to see a pattern than a table. The top 25 places that have the most unknown customers are along the Hudson River AND pretty much all-around Central Park. When you realize that both spots are both big tourist attractions and places that have bike trails. A good thing for this would be to maybe in a few of these places have just customer only bike racks so that we can truly capitalize on this phenomenon. This last map shows the most popular areas of ridership. I put a filter for the months so that we could see our burgeoning expansion efforts and how that's going. Over the time frame, the data shows that the expansion is moving along into Brooklyn and Queens, just very slowly. Maybe a bigger campaign of awareness would allow more people to ride bikes in the area that we hope to expand into. I wanted to include most and least popular ridership areas, but the new areas are the least popular but that may just be that the data is not in this dataset to show that as of yet.