Citi Bike Data Analysis

For my data analysis, I chose to limit the data source to a six month period between May and October, 2019.

The first phenomenon that I wanted to look into was the popularity of the stations. From the initial dynamic maps, we are able to see where the stations are located and the size and color of their respective bubbles are related to their popularity (with the stations with higher counts of Bike IDs appearing larger and either more green or red, for starting and ending locations). I used a bar chart to represent the top 10 stations in a more straightforward manner. As mentioned on my story, the top three stations are the same for both starting and ending bike trips. It’s also very apparent that Grove St Path is exceedingly popular with nearly twice the number of start station records as Hamilton Park (ranked second) and two and a half times the end station records.

Next, I chose to take a look at the popular start times for each bike trip. The first bar chart depicts the number of records for each start time by the hour. I chose to differentiate between Usertype to see if Customers and Subscribers had similar habits. There are two major peaks: 7AM-9AM and 5PM-7PM, which could likely be due to bike usage to and from work. It could be worth noting that Customer usage seems to progress mid-morning and then is consistently 2,200-2,400 until the surge at 5PM-7PM where we can see an addition 400+ trips begin each hour. I also compared the start times across the six month span and they appear to be consistent from month to month.

Focusing on the top 10 starting locations, I looked at the number of records for each Usertype and calculated the average trip duration in minutes to compare how customers and subscribers use the bikes. Our subscribers tend to use the bikes for shorter trips, between 6-12 min on average, while our customers take the bikes out 26-42 min on average. It could be helpful to consider how subscribers would view the Citi Bikes (i.e. daily transportation for work) in contrast to customers who may be using them more at leisure. Increasing marketing attention toward single-use customers touring the city or using the bikes for leisure could help increase customer engagement as part of a public health initiative as well.

Since we were only given minimal demographic information of each user (i.e. gender and birth year), I wanted to utilize the dimensions for another phenomenon. The next visualization represents each individually logged trip. The three concentric groups of bubbles are grouped by gender, where 0 = Undefined, 1 = Male, and 2 = Female. The size of each bubble is determined by the length of the trip duration, so larger bubbles are longer trips. Hovering over each bubble will provide more information for the record, including: Bike ID, the user’s birth year, the start and end station, and the user type.

The last phenomenon I wanted to explore was millennial engagement. I feel like the millennial age group is a key demographic for services like Citi Bike, so I wanted to take a closer look at who, specifically, in that age group is using the service and how. I filtered the records to only include those records in which the user selected either male or female as their gender, excluding those who did not wish to disclose. The highest engagement within this age group is from females born in 1988 and females also consistently were more engaged every year, though it is important to designate that engagement is as a subscriber, not a customer. Males slightly outnumbered their female counterparts as users every year. By the end of the millennial age group, it appears all four dimensions were converging between 500-1,500 records each.

Lastly, I included a map of every start station with the Average Household Size by Zip Code as a visual layer as well. I mostly just wanted to play with the maps a bit more, but I do think it’s interesting that many of the stations, especially those which were more popular on the original maps, were located in zip codes with an average household size of 1.020-2.260.