NYC Citi Bike Trip Analysis

Prior to Analysis

When first choosing which time interval to form the basis of my analysis, I came across an issue. My desire was to cover an entire year's worth of Citi Bike Rides, but this goal was impossible due to Tableau's maximum data file size when connecting to a local data source. I found a workaround this problem in the form of creating a Python script to loop through each month in my desired year, and to extract a random 1% from each month's ridership data. I then appended each month's random sample set into one dataframe, which I then converted to a CSV file to connect to Tableau.

From the data that I decided to extract and utilize in this analysis, I felt that it would be fascinating to explore the dynamic between the 'casual' and 'member' rider as documented in the Excel files that I obtained from Citi Bikes System Database. Further on, I was particularly interested to see where the top 10 stations were located within New York City in which there was the highest percentage of riders starting trips as a 'member' and as a 'casual'. This would assist Citi Bike in discovering the hubs from which users initialize their experience with the service. For purposes of this objective, I excluded stations which had 100% representation of one group (member of casual) because I felt that this would have improperly skewed my data and I only included stations in which there were at least 305 rides started from to highlight the most popular stations.

Analysis

What first struck me about the 'member' representation data was that the top 10 stations were all positioned along the eastern side of Manhattan. Immediately, I began exploring the census data (2018) that is built into the Tableau platform and I was able to see that the vast majority (9/10) of these stations were embedded within or on the border of sections of the city in which there is the highest population density category as defined in the census (as shown by the darkest color). From this, one can figure that the more people there are in a space, the fewer available area there can be for residents' cars. Therefore, other modes of transport need to be adopted, in this case Citi bikes.

When looking at the 'Rides Taken by Week' chart, I was left both surprised but also unsurprised. Obviously, it should be no surprise to anyone that the number of rides taken during the winter months would be less than the summer, spring, or fall. However, due to the busyness and non-stop lifestyle that characterizes the streets of New York one would think that the streets are well maintained by city officials during times of inclement/poor weather, thus enabling the use of bicycles still. In addition to this, by more people resorting to enclosed (roofed) modes of transport during the colder months, this would only prolong the busy traffic in the city and entice more people to seek other modes of transport such as by bicycle. Although I don't have pure evidence at my disposal right now to back up these thoughts, I do indeed find it surprising to see the fluctuation in rides by week between December 12th 2021 and March 13th 2022. It would be fascinating to see weather patterns layered onto this chart to see if that would provide more evidence to justify the fluctuations as expressed in the graph. Incorporating weather in accompaniment with the Citi Bike dataset would enable one to grasp a better understanding of the performance of the Citi Bike during certain weather conditions which could enable significant R&D into the creation of a more durable and reliable bike suitable to combat unfavorable weather.