Investigating the Citi Bike Usage and User Type Surrounding the COVID-19 Pandemic

Analysis by Isabella Hidalgo

Abstract

This investigation of Citi Bikes in the New York area focuses on the dynamic between two main variables - the type of user and the duration of these bike rides - and whether or not the COVID-19 pandemic has affected them in drastic ways. People's lives are constantly being changed in relation to the current status of the world and their communities. Citi bikes have become an essential part of the active lifestyle of New York, as they provide an easy and accessible way for many different users to travel across the state. The program RStudio was used to conduct this research. To begin in this analysis, I narrowed down the dozens of datasets, which were split into months and years, to just those immediately surrounding the peak of the pandemic: June of 2018, 2019, 2020, and 2021. Note: 2022 data was not available at the time of this study. As the first method of comparison, bar graphs were composed to compare the total number of casual vs member riders in their respective year. Then, the average trip duration was calculated for each year as well, although the 2021 dataset did not have an explicit trip duration variable. Casual riders were seen to have an increase over the years, as well as the trip duration having an increase. This was not as expected in the initial hypothesis of shorter trip durations and less casual riders. With this being concluded, the general need for these Citi Bikes has not waivered even with the impact of the pandemic.

Background

The COVID-19 pandemic affected all aspects of everyone's lives for years, and still is; from daily commutes becoming as short as walking from one side of one's room to the other, to millions of people's routines being compromised because of restrictions placed on certain areas. This investigation dives into how the type of Citi Bike rider has been affected preceding, during, and after the peak of the pandemic. The datasets utilized included those from citibikenyc.com, with features including:

- Variables Used in the Analysis:
    - User Type (Customer/Casual or Subscriber/Member)

*Only for the 2018-2020 datasets

    - Trip Duration (seconds)*
- Other Variables:
    - Start and End Location/Time
    - ID Number for the Rider/Bike
    - Gender*
    - Year of Birth*

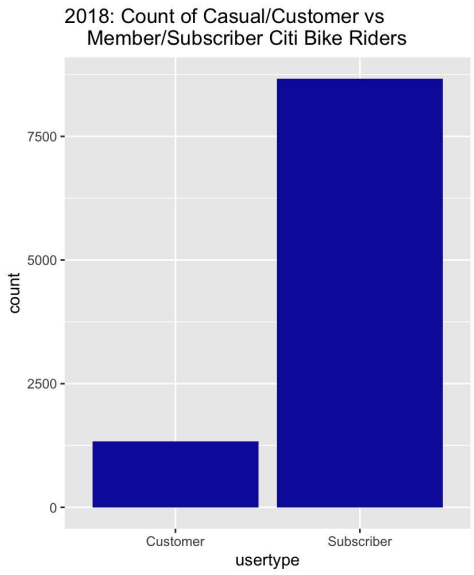
Research Question

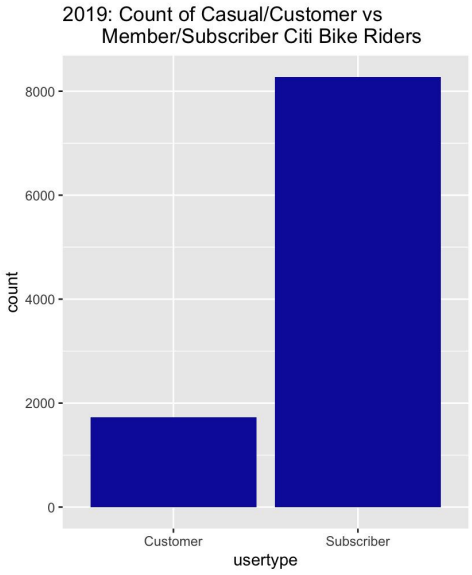
Surrounding the COVID-19 pandemic, how has the number of casual vs member Citi Bike riders changed in the New York area? How have various features of the rides been affected?

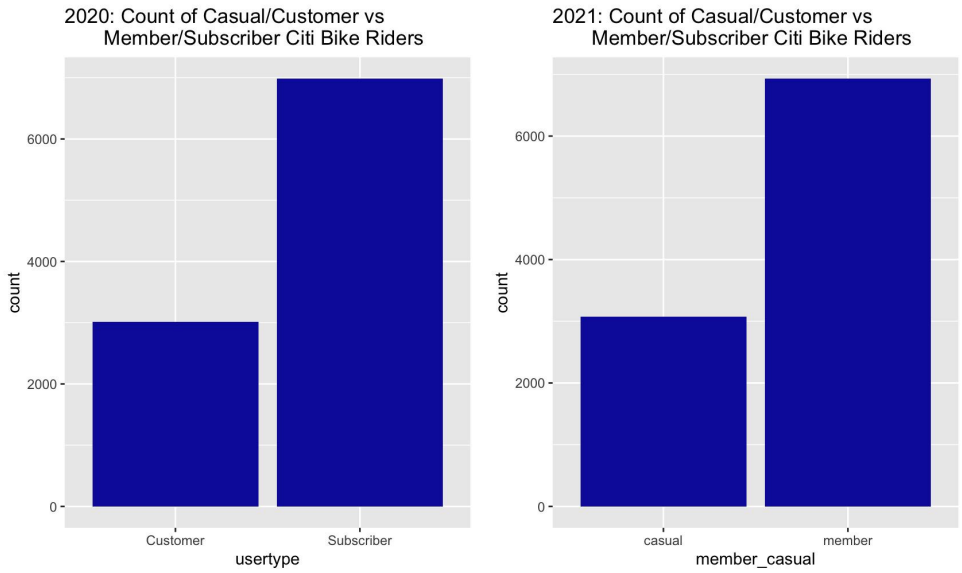
Hypothesis

Due to the COVID-19 pandemic, the number of casual riders (tourists, one-time riders, etc) have decreased in comparison to member riders.

Figures 1-4







These bar plots represent a random sample of 10,000 riders from each year, with each bar having the count of either a customer/casual or member/subscriber Citi Bike rider. In all four years, the member Citi Bike riders consisted of ~70-80% of the users. However, there seems to be an increase of casual riders as the years progressed, which was not expected in the original hypothesis.

Figure 5

```
#### averages ####
mean_times2018m <- summarise(data2018, m = mean(tripduration)/60)
mean_times2019m <- summarise(data2019, m = mean(tripduration)/60)
mean_times2020m <- summarise(data2020, m = mean(tripduration)/60)

mean_times3yrsm <- (mean_times2018m + mean_times2019m + mean_times2020m)/3

mean_times2018m
#above mean is 20.18866
mean_times2019m
#above mean is 18.48931
mean_times2020m
#above mean is 27.31818
mean_times3yrsm
#above mean is 21.99872
#note that there is no official trip duration collection method
# for 2021 and beyond since the reformatting of the citi bike
# data, so the mean may be off
```

To find the average trip durations of each year, the above code was produced. Using the summarise() and mean functions in R, the mean times were calculated. Since the datasets included these times in seconds, the result was converted to minutes. Also contradicting the initial hypothesis, there was an increase in trip times during the peak of the pandemic, whereas a shorter time was expected.

Future Usage

Further analysis could be done to narrow down specific reasons why these trends surfaced. For instance, the reason for the general increase of casual riders could have many explanations, whether the Citi Bike system just became more popular or there were actually just more onet-time riders using the bikes at the time. Comparisons to other high traffic areas would be helpful to determine whether the same trends would appear. Incorporating more variables into the investigation could also have resulted in different conclusions, assessing various VARIABLES in other situations.

Code and Data Accessibility

The code and graphs created throughout these analyses are located in the GitHub repository custom-ds-projects, where my information can be found under the folder isabellas-proj.