**Citi Bike Analytics Report**

**Data Source**

The data used for this analysis was collected from the Citi Bike Data webpage. In order to determine the effect of season on bike use, the data from June, July, and August 2019 were compiled to use as the summer months and the data from December 2019, and January and February 2020 were compiled as the winter months.

**Data Cleaning**

The data was cleaned using pandas in Jupyter Notebook. The 6 CSV files (one for each month) were combined into one data frame. A new Season column was created based on month and a new Age column was created by subtracting the Birth Year column from the year 2020.

**Excluding Outliers**

Once the data was imported into Tableau, several outliers became obvious. A significant number of users classified their Gender as “Unknown” and their age as “50”. This data was likely fabricated so all Gender Unknown datapoints were excluded. The Age range also included several outliers, with a particularly long bike trip being found in the age 16 category and several users claiming to be over 100 years old. So, the data was filtered to only include users whose age fell between 17-80.

**Dashboard 1: The Effect of Season on Citi Bike Use**

**Question:** Does the number and duration of Citi Bike trips vary between summer and winter months?

**Observation 1**: There are a larger number of bike trips made during the summer months and trip duration tends to be longer.

The number of trips made using Citi bikes is higher during the summer months than the winter months (Fig 1). The trip duration also tends to be longer in the summer months (Fig 1). So cyclists are using Citi bikes more frequently and for longer periods of time in the summer when the weather is more favorable for outdoor activities like biking.

**Observation 2**: Many cyclists are using bikes to commute to and from work regardless of season.

In terms of peak hours of use, cyclists appear to be using Citi bikes most frequently during “rush hour” aka. 7-9 am and 4-6 pm (Fig 2). This suggests that a large proportion of cyclists are using the bikes to commute to and from work. This trend can be seen in both the summer and winter months which also indicates cyclists are using the bikes for commuting rather than recreational activities, as biking is an unfavorable recreational activity during the cold winter months but people must commute to work year-round.

**Observation 3**: Stations further away from the city center are used more frequently in the summer months.

Since trip duration is longer in the summer, it is unsurprising that bike stations further away from the city center see more use during the summer months (Fig 3 & 4). In particular, bike stations in the Tremont area north of the city are exclusively used in the summer months. Interestingly, one deviation from this trend is that stations in the Bushwick area east of the city exclusively see use during the winter months. Further investigation should be done to see if this is a yearly trend or if it was only observed in the Winter of 2019/2020 for temporary reasons (such as construction, detours, etc.).

**Dashboard 2: Difference in Citi Bike Use Between Subscribers and Customers**

**Question**: Do annual subscribers and short-term customers vary in their use of Citi bikes?

**Observation 1**: Subscribers use Citi bikes more frequently, but customers use Citi bikes for longer periods of time.

Overall, subscribers made far more trips using Citi bikes than customers in both summer and winter months (Fig 1). Interestingly, customers made trips of longer duration than subscribers in both summer and winter months (Fig 1). This means that subscribers are making larger numbers of shorter trips, while customers are making fewer trips but are using the bikes for a longer duration and presumably going longer distances.

**Observation 2**: Subscribers use Citi bikes for commuting to and from work. Customers use Citi bikes for recreation.

Peak hours for bike use for subscribers are during “rush hour” periods. Meanwhile, customers do not show any peak hours, with bike use occurring steadily throughout the day. This suggests that subscribers are using the bikes to commute to and from work, while customers are using the bikes for recreation. This would also explain why subscribers make a greater number of shorter trips (since they likely are not travelling far for work) while customers make fewer longer trips (since they are likely travelling for exercise/enjoyment or to explore the city).

**Observation 3**: There are more male subscribers and customers than female, but there is a much larger difference between the number of male and female subscribers than customers.

There are a larger number of male Citi bike users across all age groups, in both the subscriber and customer groups. However, there is a much larger proportion of male users in the subscriber group than the customer group, particularly in the 25-35 age range. The gender gap in bicycle use is a well-documented phenomenon (<https://genderpolicyreport.umn.edu/bicycles-gender-and-risk/>) and further research should be conducted to explore ways to encourage more women to use bicycles as a means of transportation.

**Dynamic Map Analysis**

The dynamic map of the locations of all bike stations and their popularity across months clearly shows that the most popular bike stations are all in the city center. Bike stations outside of the city center are used less frequently, particularly in the winter months. This is understandable as the largest concentration of people as well as the largest concentration of businesses/points of interest are in the city center. Also, since most Citi bike users are subscribers using the bikes to commute to and from work (see dashboards) it makes sense that most of them would be located in the city center.