Ford GoBike System Data Exploration

by Amanda Zhou
September 05, 2020

Dataset

This data set includes information about 192,082 individual rides made in a bike-sharing system covering the greater San Francisco Bay area for January 2019. 248 data points were removed from the analysis due to inconsistencies or missing information. Below are the attributes of each trip. In addition, I will extract the hours and days from Start Time, and caculate the trip distance based on its start and end station's latitude and longitude.

Each trip is anonymized and includes:

- Trip Duration (seconds)
- Start Time and Date
- End Time and Date
- Start Station ID
- Start Station Name
- Start Station Latitude
- Start Station Longitude
- End Station ID
- End Station Name
- End Station Latitude
- End Station Longitude
- Bike ID
- User Type (Subscriber or Customer – “Subscriber” = Member or “Customer” = Casual)
- Bike Share for All Trip

Summary of Findings

In the exploration, I found:

1. On average, customers spent more time than subscribers on each trip.
2. Trip duration in 0-3 am time period is longer than other hours of the day for both users.
3. Subscribers use bike share a lot during rush hours in weekdays and not much in the weekend.
4. Customers use bike share much more during weekend, especially on Saturday afternoon, probably for the purpose of entertainment, and their trip durations are longer than other days.
5. Trip distance doesn't have a strong correlation with its duration.
6. Users who are subscribers occupied a large proportion in this dataset, which is 7 times more than non-subscribers.
7. Overall, the usage amount for Start Station looks roughly the same as End Station based on their IDs.
8. Only 8.4% of users used bike share for all trip.

Key Insights for Presentation

For the presentation, I focus on the relationship between trip duration, time and the user type.

I start by introducing the trip duration, followed by pairing it up with Start Time (includes Hour of the Day and Day of the Week) and user type to observe the pattern in their distribution plots.

Afterwards, I combine these three elements by using heatmap. So we can clearly see that subscribers spent less time than customers on each trip. Also, they use bike share a lot during rush hours in weekdays and not much in the weekend. However, Customers use bike share much more during weekend, especially on Saturday afternoon.