FordGo Bike Data Exploration

by Syed M Kamran

Dataset

The dataset I've chosen for analyses is FordGO Bike sharing data. This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area. For the purpose of this project, I will be analyzing and visualizing FordGo Bike data for the years 2017 and 2018. Dataset with feature documentation is available here.

Summary of Findings

- October was the most popular month for the rides followed by September then August. Most of trips peaked around 8-9 AM and 5-6 PM in the evening indicating most of these users commute to their offices using FordGo Bikes. Moreover, there are a greater number of trips during weekdays (82%) than weekends (18%).

- I have plotted the distribution of trip duration in minutes by taking log10 transformation and concluded most of the people use bikes for short trips (under 17 minutes) while average trip duration was around 11 minutes.

- Usage for subscribers drop significantly on weekends. On the other hand, Usage of customers is pretty equal on weekdays (Mon to Fri) but there is a slight increase in the volume of customers during weekends.

- We have also discovered that following are the most popular stations.
    - San Francisco Caltrain (Townsend St at 4th St)
    - San Francisco Ferry Building (Harry Bridges Plaza)
    - San Francisco Caltrain Station 2 (Townsend St at 4th St)

Key Insights for Presentation

For presentation, I started by analyzing the bike usage for different months followed by the bike usage for different days and then finally I plotted usage of bikes by hour. I also identified average trip duration (after removing outliers) and predicted that most of the trips are of short durations.

In the second part of the presentation, I have analyzed how different user types influence bike usage patterns. I started by plotting number of subscriber and customers followed by their weekly usage and then hourly usage. I have also plotted trip duration based on user type in the form of violin plots and finally for little deeper analysis, I have included a heatmap depicting the usage patterns for subscribers and customers for different hours and trip durations.

In the last part of the presentation, I have plotted popular start stations (top 20), predicted start stations that are near to the homes, and end stations that are near to the offices.

Resources

- https://www.lyft.com/bikes/bay-wheels/system-data
- https://github.com/