Analysis of Ford GoBike System in 2017-2018-2019

Sandhya Dhanikonda
May-2019

Dataset

The data consist of approximately 2,75M bike rides from year , 2017, 2018 and first 3 months of 2019. The attributes include the trip start/end time, start/end station, duration in seconds as well as additional information such as user type, gender, and birth date. The data that was not considered for analysis was that of the location details of the start stations and end stations. The data I took up for analysis can be found in the website: https://s3.amazonaws.com/fordgobike-data/index.html

Summary of Findings

In general exploration, I found that there are two types of clients using the system- subscribers and customers. Subscribers are mainly daily commuters, having short trips to and from work, who rent a bike on weekdays at 7-10am and 4-7pm. Customers, usually tourists or occasional riders who use the system mainly on weekends to explore the Bay Area. The data shows that there are two peaks in the demand of bikes in the initial two months and then high demand being in the months from May-october.
But from March there is a clear drop in usage, probably due to the weather condition. More over I analysed the data for if there is a different trends of usage as per the age groups of the users.
As I explored the data for the categories of subscribers and customers separately, and found that the age groups 20-40 are the main users of the system, they took the more than 75% of bike rides. Among those, 30 to 40 years old people's rides account almost %40 of all bike rides. People who have subscribed to the system use this service on weekdays more than weekends. 8am and 5pm are the peak hours for this service. Also, people use this service when they are in lunch time as well. Percentage of

subscribers is almost %88.7. Percentage of customers is almost %11.3. 90% of bike rides take place on weekday. The peak bike rides time for all members is around commute time. Customers use this service mainly in the summer months on for mostly weekends or after work hours.

Key Insights for Presentation

For the presentation, I focus on the influence of age, timeframe, weekday, age group of bike hiring data. I start by introducing the age distrubition, monthly bike hiring trend, followed by age group distribution, then plot the weekday, timeframe with age data. I tried to use different color palettes for each variables to make sure it is clear.