Analyzing Ford GoBike Data for (December 2019)

Project overview:

This project has two main parts (Exploratory and Explanatory), In the exploratory part, I will do an exploratory analysis for Ford GoBike Data for (December 2019). I will use Python visualization libraries to systematically explore a selected dataset in order to have clear image about the data structure, patterns and relationships between the variables. Starting with univariate relationships plots of single variables and building up to plots of multiple variables relationships.

In the explanatory part, I will take the main findings from the exploratory part and communicate my results in short presentation that illustrates interesting properties, trends, and relationships that I discovered in Ford GoBike Data for (December 2019). The explanatory part of the project should make heavy use of the first part of the project.

Choosing a data set:

I choose Ford GoBike bike data (December 2019) for my project. It's about bicycle sharing system in the San Francisco Bay Area, the bicycles are available 24 hours a day, seven days a week with data about (duration, bike id, start time, end time, rental access method, user type).

Data Exploration findings:

Starting with univariate plots of single variables and building up to plots of multiple variables relationships. I found that there is relationship between four variables in my data set (duration sec, start time hour, start time weekday, user type).

Key insights:

The center of attention was the influence of user type, start time hour, start time weekday and duration in seconds. I began by exploring the duration in seconds, then start time hour, then I started to plot two variables then three together. I used different plots and add more variables in each step in order to get the key insights, and make sure it is clear and well-polished and the points can be easily understood by the reader.

Summary:

In the exploration part, I discovered that there are two types of users using the system subscribers and customers. The data shows that the subscribers are using the system more often with probability of having short trips to and from work or school with peak hours between 8 to 9 a.m. then followed by similar peak volume between 4 to 6 p.m.

Throughout the week peak days throughout a week with most peek days are Monday & Tuesday, then followed by little lower peak on Thursday & Friday. By adding duration in seconds variable I start to notice that actually customers are using it more than the subscribers. The average for customer trip duration its 900 seconds (15 min) and for subscriber its 700 seconds (11.6 min), the difference is 3.4 min for the customer.

This is because the subscribers are using it for short trips such as work or school, on the other hand the Customers are using it more for longer trips around the bay area. Even on the start time hour (24 hours) customers are more likely to use it with higher trip duration as the plot shows, which prove that the customers are using the bicycle sharing system more than the subscribers in duration in seconds.