Read Me

The chosen data set is Ford GoBike Data System for the year 2017 provided by the Udacity group. After performing a few cleaning steps, the data were statistically described using 'df. describe ( )' code. The results had shown insights to think about the variability in the duration time and how it is related to other variables. Also, by looking at the data, many queries may come to your head regarding user type or stations names variables. In the exploratory section of the project, in the beginning, knowing that it is hard to interpret all station names because of the huge number of records, I have chosen to focus my studies on the top five start station names and top-five end station names. Hence, the univariate plots showed those top-five start and top-five end station names. The data was adjusted by dividing it into two sub-sets, the first sub-set of the data included the top 5 start stations names, and the second sub-set of the data includes the top 5 end stations names. Then a bivariate plot is used to plot the top 5 start station name user type. The results showed that the top 5 start stations are mostly represented by subscribers more than normal customers. This means that subscribers usually subscribe to the system for frequent routine rides while customers usually tend to use Ford GoBike rides for the unplanned ride. The results for the top 5 end stations have shown that most users were represented as subscribers as well. At this point, it was clear that there is a relationship between the frequent number of the stations and the user type. More frequent stations in the data belong to subscribers. Now, to be able to answer the third question: Which user type takes a long time for a ride in the top 5 start stations and top 5 end stations? A third variable has been added to the plot converting it to a multivariate plot to investigate the relationship more. The findings show that customers use Ford GoBike for longer durations than subscribers. This is because the top five start stations and end stations are attractive sites for tourists. This explains that customers might be tourists or people who like enjoying the weather with a long ride. Not like subscribers who might follow routine time using GoBike.