

Ford GoBike 2017

In this data set we can find a report of all the bike rides for the city of San Francisco

The structure of the dataset

There are 519700 bike rides in the dataset with 19 features. Most variables are categorical.

The main features of interest in my dataset.

I'm most interested in figuring out how does every different biker group (gender, age, customer-subscriber) behave and how the bike use change through time.

Some questions i would like to answer are:

- 1) How does our clients separate by gender and age?
- 2) What are the riding habits of each gender? Are there any differentiations or all the genders have the same behavior in a day?
- 3) Which month, day, hour has the more rentals?
- 4) What are the riding habits of subscribers and ordinary customers?
- 5) Why does our clients rent our bikes? Is it for fun or for everyday activities? Is it the same both for "subscribers" and "customers"?
- 6) Which are the most popular starting end destination stations?

The analysis will answer all the above questions

Plots used

I use various plots to explore my data. Taking on mind that the majority of data are 'categorical' I choose mainly 'countplots' and heatmaps. I also use boxplots and violin plots mainly to study the 'Duration' variable.

Key Insights for Presentation

For the presentation, at the beginning i focus on the different riding habits of every gender and age group. At second time i study the bike usage for every month- day and hour. There are same important conclusions based on my analysis.

We can see the difference between Subscribers and common customers, between the various genders etc. I tried to make more obvious what our clients riding habits are, what is the purpose of bike riding for every group.

I discover that bikes are more popular to male, especially at the age of 30s. The better month was October.

Some important conclusions are:

- It is obvious that at weekends the rides are generally longer especially for the Customers. The subscribers seem to keep the same duration mean for every day in a week. I can assume that this is happening because subscribers are using the bikes for a specific reason and not for leisure. That is why we do not see longer rides at the weekend by them. At the other hand, the Customers are using the bikes mainly at weekends.
- We can see that the more rides are on working days at 8:00 and 16:00-17:00. This can lead us to the conclusion that the riders are using their bikes to go to work at working days. At weekends the most popular hours are 11:00-15:00 and as we saw previously more riders are 'Customers' that means that at weekend the bikes are used mainly for fun rides around the city.