Ford GoBike System Data

Dataset

- The data we do our analysis on it is “Ford GoBike System Data” in San Francisco Bay area.

Summary of Findings

- In the exploration, I found:

• The dominant gender is male by the percentage of 74.6%, females are 23.3% and the other gender are 2.1%.

• The dominant user type is the subscriber by the percentage of 89.2% and the customers are 10.8%.

• The most users are Subscribers and male.

• The second most users are Subscribers and female.

• females’ ratio of customer user type is: 23.4%.

• females’ ratio of Subscriber user type is: 22.1%.

• We can determine that they are similar but do not have the same order and there is only one in the end and the start stations is different.

• The most trips is under 1800sec (30min) and the most trip duration is around 600sec (10min).

• The most age range is between 20 and 40 and the most one is around 25 and 35.

• The correlation between the age and duration as it a relatively stable relationship, then a slight negative relationship, and then strengthened with age, so the duration decreases.

• The female's median duration is higher than the males and have the longer trip than males although the number of males is more, and the other gender has the longest trips.

• The Customer has a longer trip duration than the subscriber.

• The other gender density jumped again in the age of 50 and 60 which is it a surprising thing.

• There is slight tilt to higher age with high trip duration for subscribers.

Key Insights for Presentation

- For the presentation, I focus on the correlation and the density between age, user type, and gender with the trip duration and extract some information about the top stations.
- I start by introducing the top 10 Start and End stations, followed by the Dominant gender and user type, then plot each gender count of each user type.
- Afterward, I introduce the distributions and density plots of the main specifications I focus on one by one. To start,
- I used different plots and heat maps of trip Duration and Age then gender.
- I'm only looking at the clarity grade plot here since it's the clearest example of how Gender, User Type, and Age affect the trip duration.

Resources user in this project:

- https://www.python-graph-gallery.com/71-density-plot-with-shade-seaborn
- https://www.python-graph-gallery.com/density-chart-multiple-groups-seaborn
- Matplotlib documentation
- seaborn documentation
- StackOverFlow