Dataset:

Ford GoBike System Data. This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area.

Findings:

- Trip duration and distance depend on the user's age, gender and type.
- Rental counts follow the rush-hour patterns during weekdays, indicating usage of GoBike for commuting.
- During weekdays, we see an increase in the number of rentals compared to weekends.
- The longest trip distances are from 8am and 5pm, whereas the longest trip durations are usually seen at noon.
- Different regions in the Bay area share similar patterns in terms of trip duration and distance.

Presentation Motivation:

I chose some representative features in the presentation. First, since the data is from a bike rental company, the most interesting thing would be the statistics on the ride itself, such as how long the trip takes, how far the trip goes. Then, I looked at the user groups including their age, gender, this information could be useful because it might help the rental company make better decisions to improve the product. Next I decided to study some relationships between different variables, because there might be some hidden features that we didn't see from single variables. I chose some of my interesting findings. For example, based on the usage during weekdays and weekends, on each hour of the day, and on different age groups, I found a large portion of the rentals goes for commuting. I also found that the max duration (during lunchtime) and max distance (morning and afternoon commuting) on each hour of the day are not the same.

Resources:

Geopy: https://geopy.readthedocs.io/en/stable/