

The dataset I chose was the Ford GoBike System. This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area. I used the included demographics like gender and age to find out who is using the bikes the most often. I also extracted distance traveled by taking the longitude and latitude of the start and end stations and the earth's radius, and then calculated using the haversine formula how far a bike travels. I also took duration in seconds and converted it into minutes for better analysis. I then used these two to figure out how far users were traveling and for how long.

Some of my main findings were that females between the ages of 26 and 34 traveled more than males. I also found that the age group 26 - 35 traveled the most in general. Another thing I found was that most trips were no more than 1.5 miles and 20 minutes no matter the age group or gender. Additionally, although there were more males in this dataset that traveled further and longer than females, on both average distance and duration, females traveled longer and further than males. Lastly, when it comes to when users are traveling, I found that more users travel during weekdays than on weekends if they are subscribed to the system.

References used in project:

- <https://stackoverflow.com/questions/33029396/using-pandas-to-calculate-distance-between-coordinates-from-imported-csv>
- <https://www.kaggle.com/residentmario/bivariate-plotting-with-pandas>