

Project: (Ford GoBike Data Exploration)

Dataset

I have used Ford GoBike Dataset, which are a dataset contains 183412 bike with 16 feature most of the variables are numeric, and some of them categorical variables and time series.

This data set includes information about individual rides made in a bike-sharing system for 2019 covering the greater San Francisco Bay area.

The Data

- Duration Sec
- Start Time
- End Time
- Start Station ID
- Start Station Name
- Start Station Latitude
- Start Station Longitude
- End Station ID
- End Station Name
- End Station Latitude
- End Station Longitude
- Bike ID
- user_type
- member_birth_year
- member_gender
- Bike Share for All Trip

which tracks members who are enrolled in the Bike Share for All program for low-income residents.

- User Type
(Subscriber or Customer – “Subscriber” = Member or “Customer” = Casual)

Summary of Findings

• Univariate Exploration:

1. the majority in the dataset are not enrolled in "Bike Share for All Program" and the user type majority are subscriber.
2. The distribution is more concentrated between 20 to 40 years old.
3. Number of males is higher than female and other

• Bivariate Exploration:

1. There is a high correlation between start station latitude and end station longitude
2. There is a positive correlation between the start and end latitude as we can see from the visualization.
3. Though quantity of male riders are very high then other and female but we can see that higher percentage of female and other rides longer trips then males.

• Multivariate Exploration:

1. The two categorical measures gender and user type play into the relationship between trip duration and age.
2. There is not a single user who's a customer and enrolled in bike for all program in 2019. We noticed that the subscribers have the greatest number of trips on October.