

# Analysis of San Fransisco Bay Area Ford Gobike Service

Adegbite Paul Adekunle

## About Dataset

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

## Columns

- Trip Duration (seconds)
- Start Time and Date
- End Time and Date
- 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 (Subscriber or Customer – “Subscriber” = Member or “Customer” = Casual)

## Summary of findings

- The majority of the users are male
- The majority of the users are subscribers
- The service is mostly used during working hours around 7 - 9 am and 4 - 6 pm
- This trip service is mostly used during the weekdays as seen that it was mostly used during work hours
- weekdays are the busiest between 7 am and 8pm but there isn't much trips between 10 am and 3 pm which should be work hours ideally.
- on weekends, it is busy between 11pm and 2am
- across all week days and hours, subscribers have higher trips than customers
- Bikers who do not use bike sharer for all trips engage in more trips across all hours of day and days of week compared to bikers that do.
- No customers have used bike share for all trips while some Subscribers use bike share for trips.
- Male bikers have more trips compared to female bikers across all hours of the day and all days of the week
- Bikers who do not specify their gender is a very small portion of Customer and subscriber while male bikers carry the majority for both user type
- Bikers use the service mostly on Thursday for all gender type.
- 8am and 5pm are the hours that bikers of both gender spend most time on trips.

## **steps taken during exploration:**

firstly, I gathered the data which has been provided to us by Udacity, and then i cleaned the data (removed duplicates, treated nulls, changed datatypes etc),I then did some feature engineering, then I started the exploratory data analysis starting with univariate exploration on some of the categorical variables generating some insight, I then moved to bivariate exploration checking the relationship between two variables generating more insight, after that i concluded with some multivariate exploration ,doing some log transformation as I have been taught. it was a good project.

## **Key insight for Presentation**

- Distribution of gender use of service
- Distribution of user type (customers/subscriber) use of service
- Most use of service for hours, day and distribution of bike share for all trips
- Relationship between duration and age
- Relationship between hour and user type
- Relationship between day of week and use type
- Relationship between day of week and hour
- Relationship between user type and bike share for all trips
- Relationship between day of week and bike share for all trips
- Relationship between hour of use and gender
- Relationship between ay of week and gender
- Relationship between gender user type