**Introduction:**

I chose the Ford GoBike Dataset from the following link:

<https://video.udacity-data.com/topher/2020/October/5f91cf38_201902-fordgobike-tripdata/201902-fordgobike-tripdata.csv>

This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area. The dataset contains 183412 rows and 16 columns.

The columns are:

duration\_sec : The trip duration in seconds

start\_time : The start date and time of the trip

end\_time: The finishing date and time of the trip

start\_station\_id : A unique id given to each start station

start\_station\_name : The name of the start station

start\_station\_latitude : The latitude coordinates of the start station

start\_station\_longitude : The longitude coordinates of the start station

end\_station\_id : A unique id given to each end station

end\_station\_name : The name of the end station

end\_station\_latitude : The latitude coordinates of the end station

end\_station\_longitude : The longitude coordinates of the end station

bike\_id : The unique id of the bike (4646 bikes in total)

user\_type : Either Customers or Subscribers

member\_birth\_year : The birth year of the member

member\_gender : The gender of the member (Male, Female or Other)

bike\_share\_for\_all\_trip : Did the user make the whole trip using bike share

From this initial overview i would modify the start and end time of the trips into an appropriate format, drop the station id columns, station coordinates, and bike id. The main features I would concentrate on are the user gender, age, type, the most frequented stations (start and stop).

I started First by checking if the station id count matches with the station names so we can drop one of the two columns and they matched which led me to drop one of the columns. I then counted the types and number of users and found out that there are 163544 Subscriber and 19868 Customer.

I then counted the genders and number of users and found out that there are 130651 Male users, 40844 Female users, and 3652 Other users. From this initial overview I would modify the start and end time of the trips into an appropriate format, drop the station id columns, station coordinates, and bike id. The main features I would concentrate on are the user gender, age, type, the most frequented stations (start and stop).

I proceeded to drop these columns from the dataset ['start\_station\_id' ,'start\_station\_latitude' ,'start\_station\_longitude' ,'end\_station\_id','end\_station\_latitude','end\_station\_longitude','bike\_id']. And then I dropped the NaN values.

**I would also ask the following questions:**

Q1 Does the frequency of trips depend on if a user is a Male, Female or Other?

Q2 How long does the average trip take?

Q3 Does the frequency of trips depend on the age of the user?

Q4 Does the frequency of trips depend on if a user is a subscriber or customer?

Q5 Does the Duration of a ride depend on the Age of the User?

Q6 How many users choose to use Bike Share for all Trip?

Q7 what is the relation between the Gender of the user and their age?

Q8 What are the top ten start and end stations?

Q9 The Relation between gender, age, and duration of trip.

From these questions I found the following conclusions:

* Males in general use the Ford GoBike service approximately 3 times as Females and orders of magnitudes more than Other, the marketing department should focus more on Females to increase their numbers.
* The average ride time is around 5 to 10 minutes which means that most users approximately 65000 take 5 to 10 min trips while a smaller number of users (approx. 39000) take the bikes for 5 minutes and a similar number take the bike for more than 10 minutes, and a minority of users take the bikes for more than an hour.
* Most of the users are between the age of 25 to 40 and decreasing after 40 till we reach a few users by the age of 80.
* The number of Subscribers surpass the number of Customers, and that Customers form about 1/8 from the users of the Ford GoBike service.
* Users of age 120 years and durations or rides that exceed 3 hours to better focus on the users and their ride times i choose to view users from age 20 to 70 and time durations of 2 hours. it is also observed that most users take bikes for the duration of around 15 minutes and users from ages of 20 to 40 use the bike for an extended period of about 30 minutes.
* The number of users who chose not to use the bike share for all the trip is significantly higher than users how chose to use it.
* The mean is around the age of 35 for males and slightly lower for females while others have a slightly higher age mean. we can also observe outliers above the age of 60 ~ 65 which corresponds well with previous findings.
* The top 10 stations in the start and end of a trip have a lot in common (most stations are in both lists) this means that the general area between these stations contains the highest traffic, and we should invest more in this area.
* The male and female users under the age of 30 usually have the same trip duration of about 50 minutes and stay almost consistent for all ages. However, we noted from a graph we explored before that the number of female users is quite smaller than that of male users which reinforces the finding that marketing team should focus more on campaigns targeted towards female users.