Ford Gobike EDA

Investigation Overview

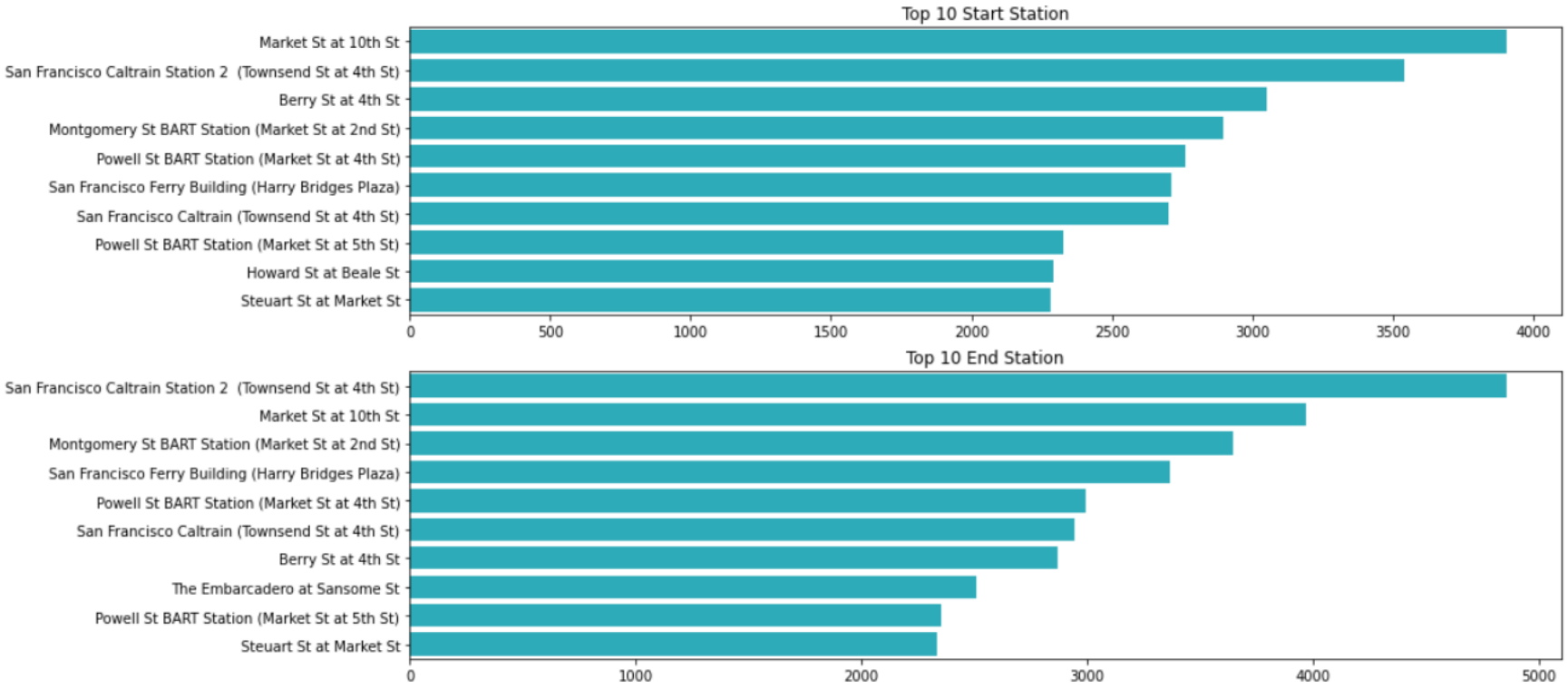
In this investigation, I wanted to look at the specifications of trips that have a correlation with it. The main focus was on: member age, user type and duration.

Dataset Overview

There are 183412 fordgobike trips in the dataset with 16 specifications (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). Out of 16 specifications 9 are numerical, 2 are datetime and 5 are object type.

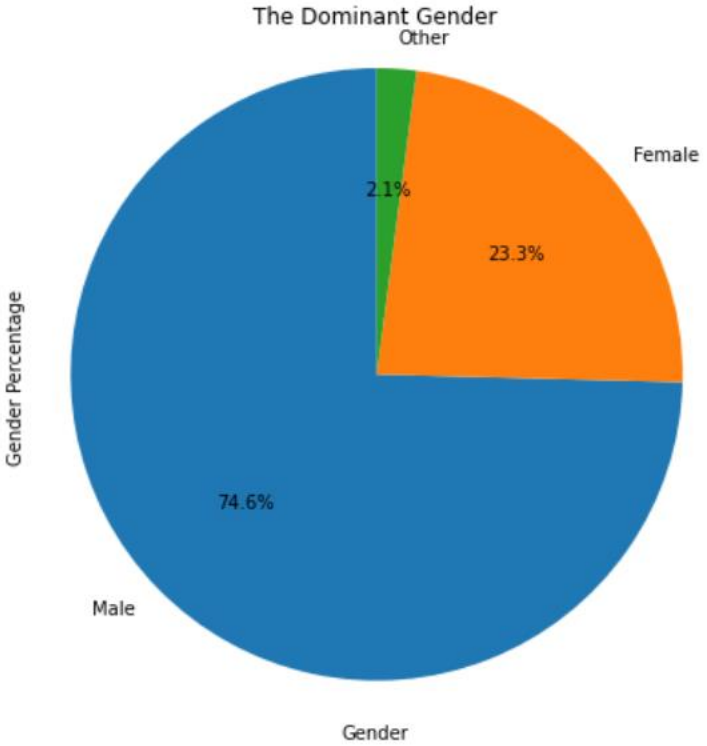
Top 10 Start and End Stations

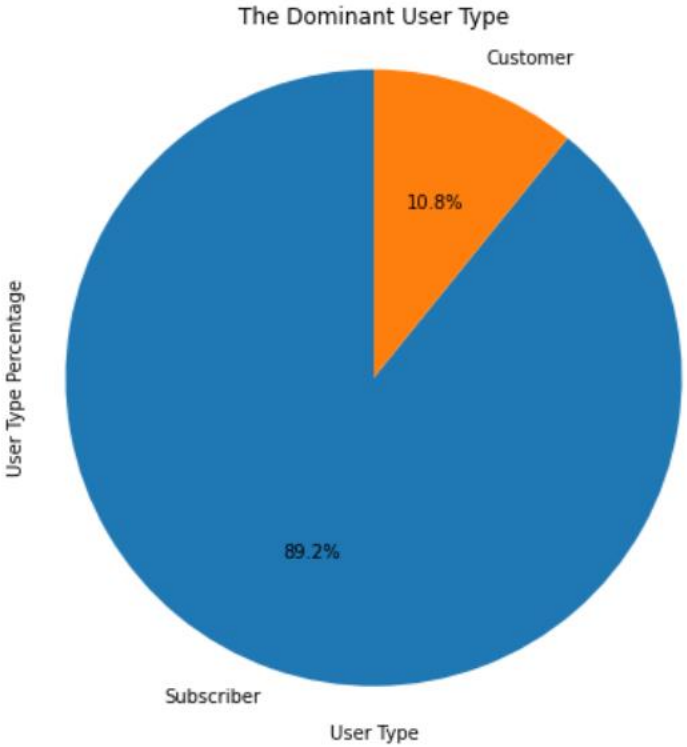
We can see the top Start and End stations that is favorite for our users.


Top 10 Start Station
Market St at 10th St
San Francisco Caltrain Station 2 (Townsend St at 4th St)
Berry St at 4th St
Montgomery St BART Station (Market St at 2nd St)
Powell St BART Station (Market St at 4th St)
San Francisco Ferry Building (Harry Bridges Plaza)
San Francisco Caltrain (Townsend St at 4th St)
Powell St BART Station (Market St at 5th St)
Howard St at Beale St
Steuart St at Market St
0
500
1000
1500
2000
2500
3000
3500
4000
Top 10 End Station
San Francisco Caltrain Station 2 (Townsend St at 4th St)
Market St at 10th St
Montgomery St BART Station (Market St at 2nd St)
San Francisco Ferry Building (Harry Bridges Plaza)
Powell St BART Station (Market St at 4th St)
San Francisco Caltrain (Townsend St at 4th St)
Berry St at 4th St
The Embarcadero at Sansome St
Powell St BART Station (Market St at 5th St)
Steuart St at Market St
0
1000
2000
3000
4000
5000

Dominant Gender & Dominant User Type

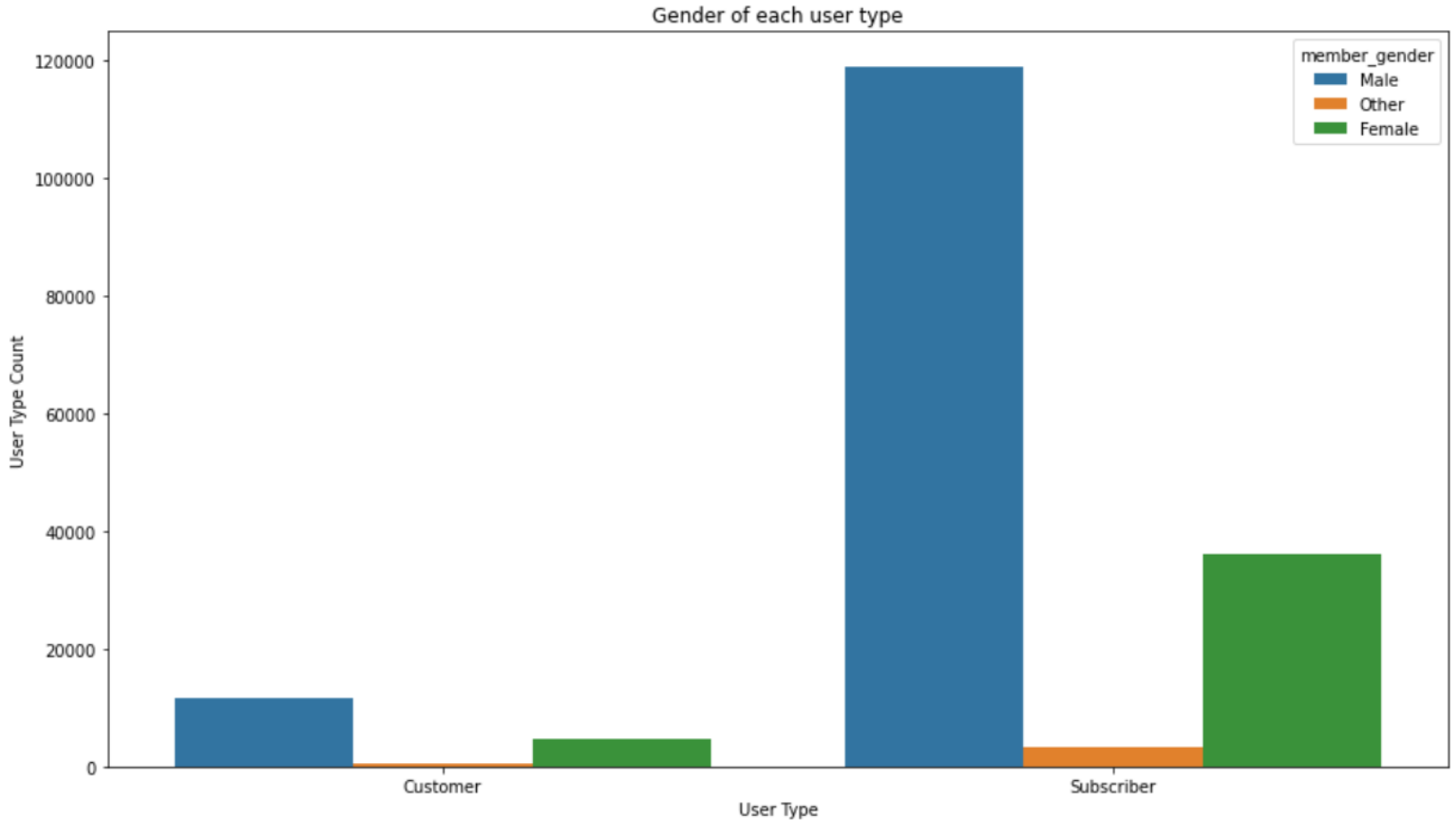
From these plots, we can determine that males are the dominant gender by the percentage of 74.6% and the dominant user type is the subscribers by the percentage of 89.2%.


The Dominant Gender
Other
Female
2.1%
23.3%
Gender Percentage
74.6%
Male
Gender


The Dominant User Type
Customer
10.8%
User Type Percentage
89.2%
Subscriber
User Type

Gender vs. User Type

From this plot, we can determine the number of each gender in each user type and the ratio


Gender of each user type
member_gender
Male
Other
Female
120000
100000
80000
60000
40000
20000
0
User Type Count
Customer
Subscriber
User Type

Females ratio of customer user type is: 23.4%

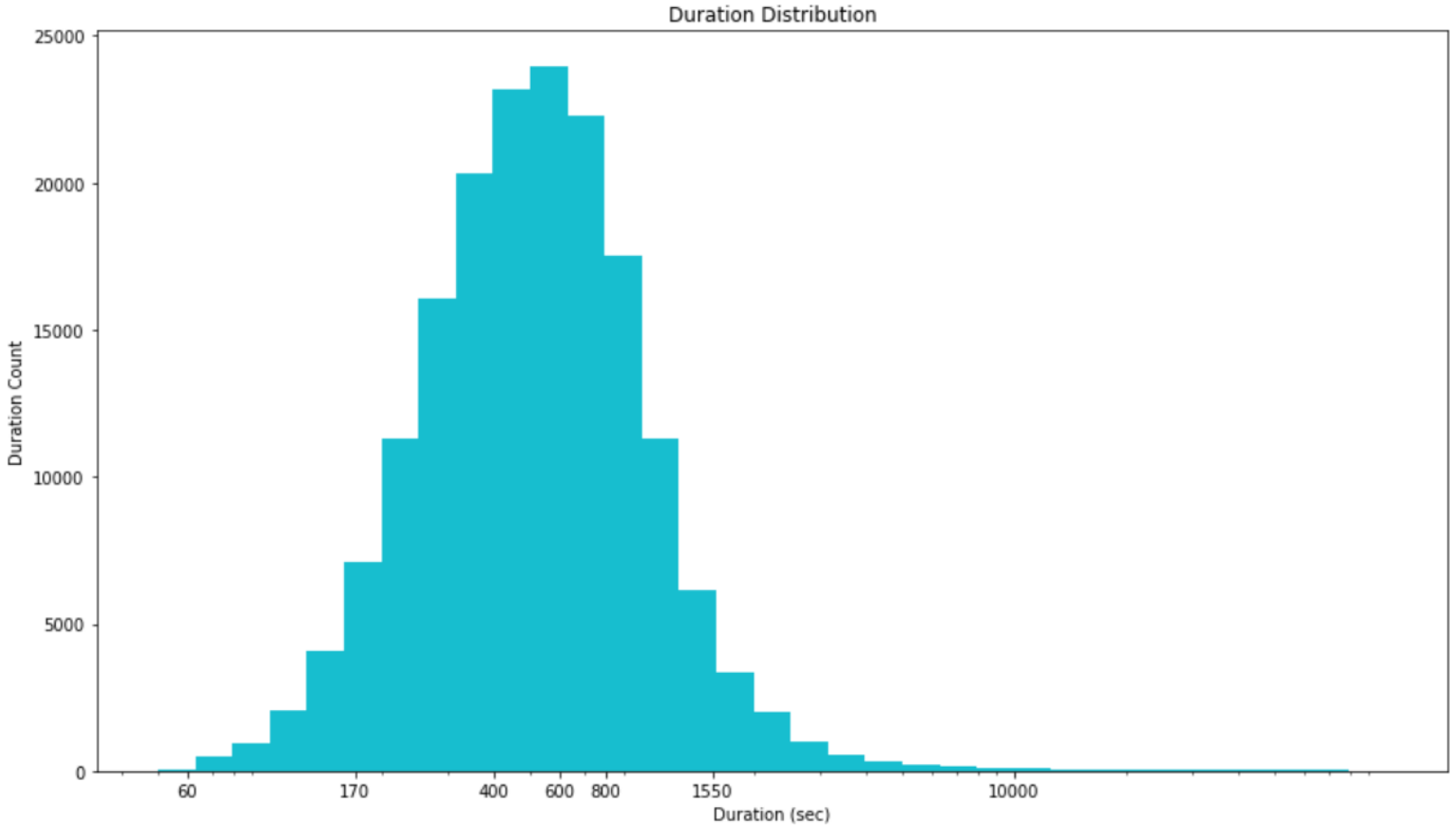
Females ratio of Subscriber user type is: 22.1%

Males ratio of customer user type is: 76.6%

Males ratio of Subscriber user type is: 77.9%

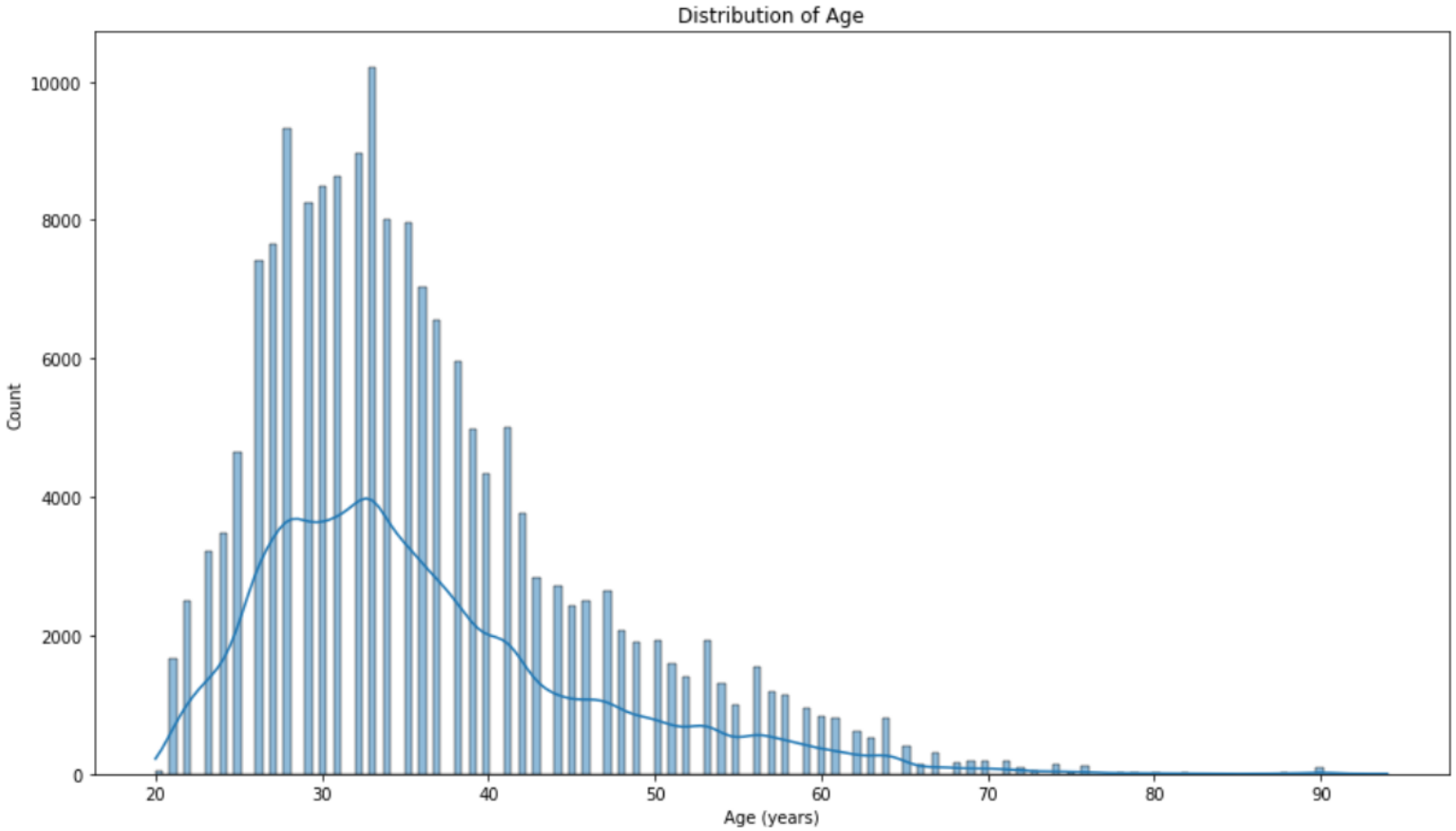
Distribution of Duration

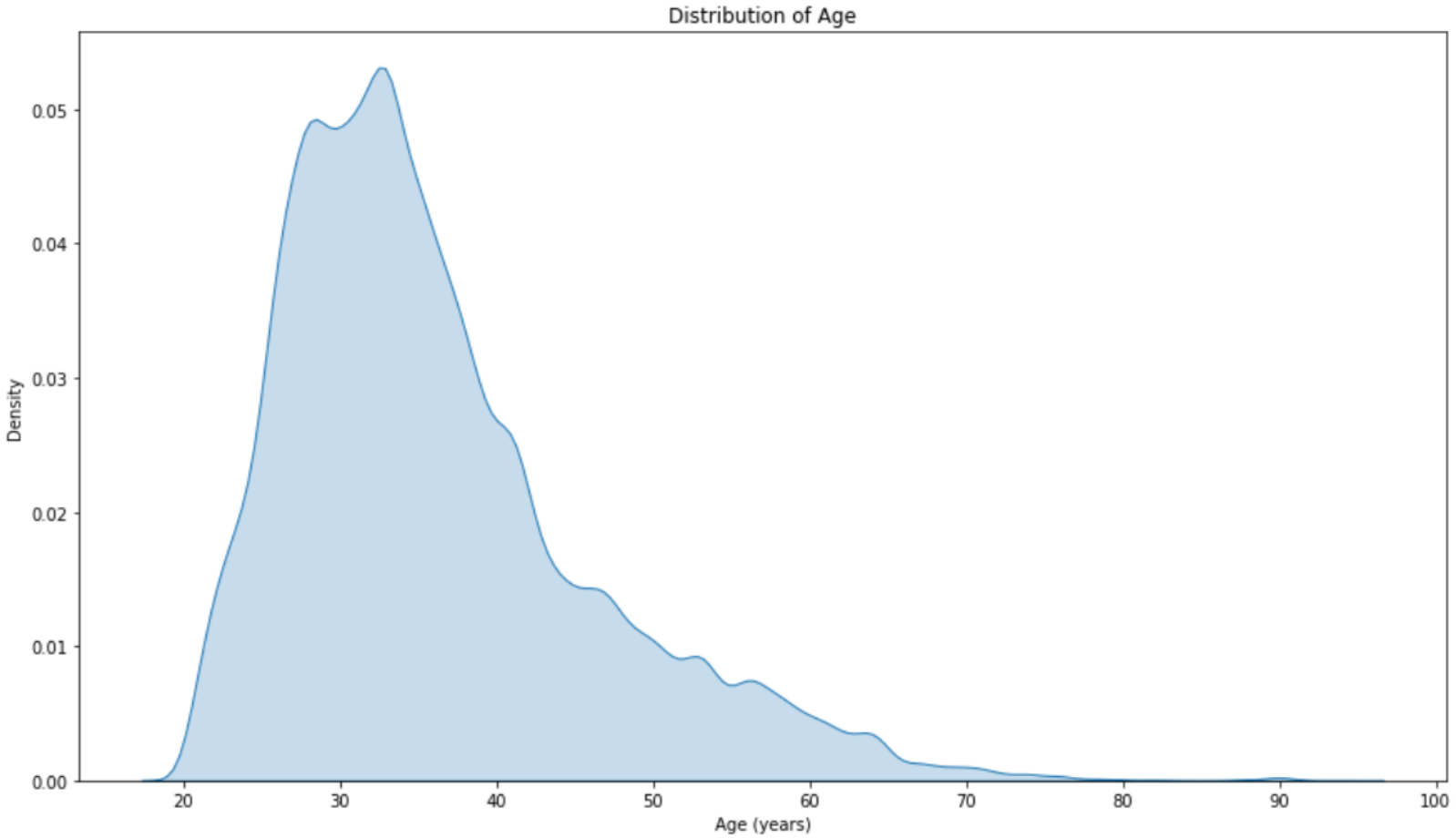
most of the trips is between 170sec(3 minutes) to 1550sec(25 minutes) and they centered around 600sec(10 minutes)


Duration Distribution
25000
20000
15000
10000
5000
0
Duration Count
60
170
400
600
800
1550
10000
Duration (sec)

Distribution of Age

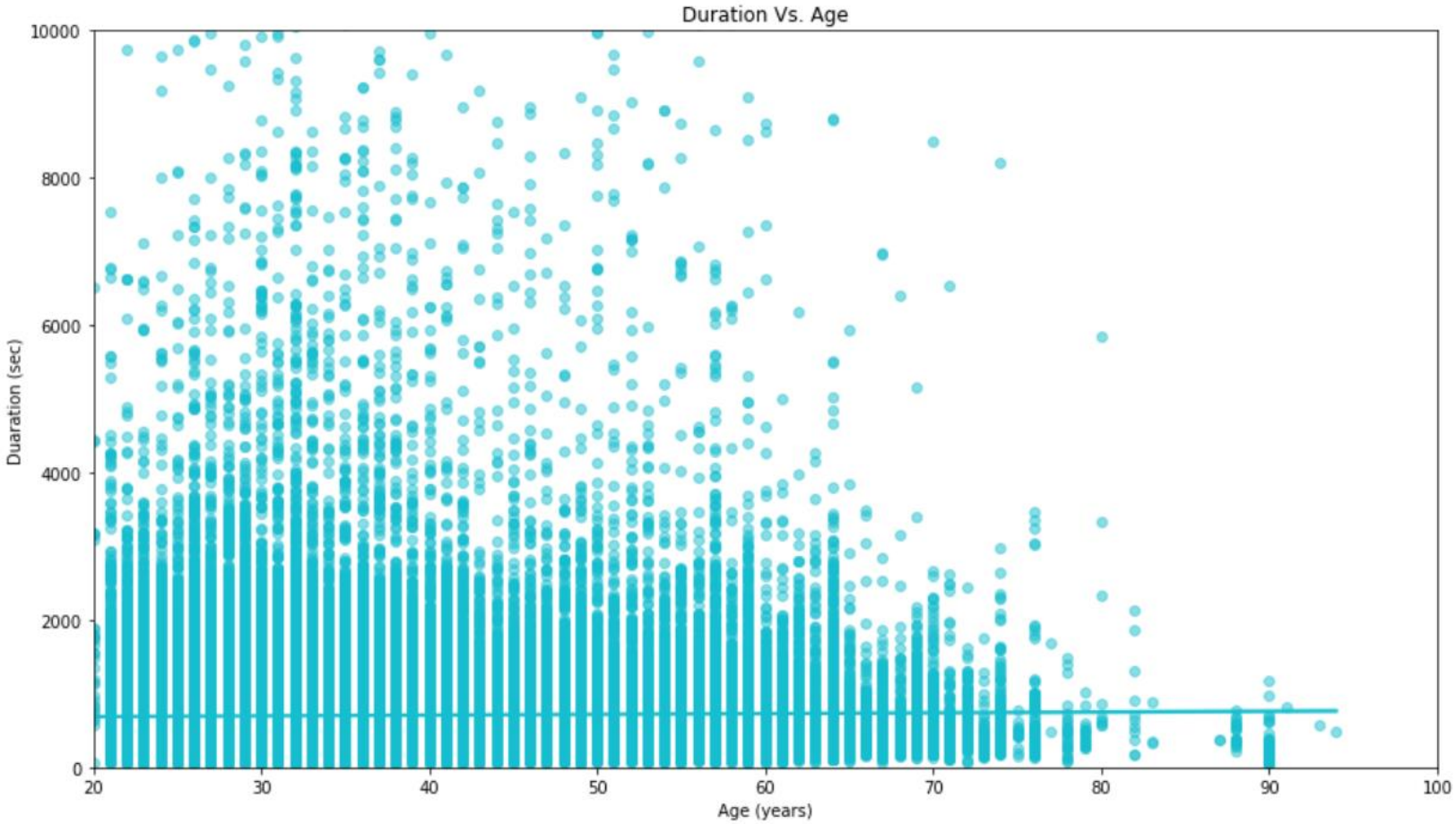
From these plots, we can determine the most age range is between 20 and 50 and the most one is around 30 and 40.

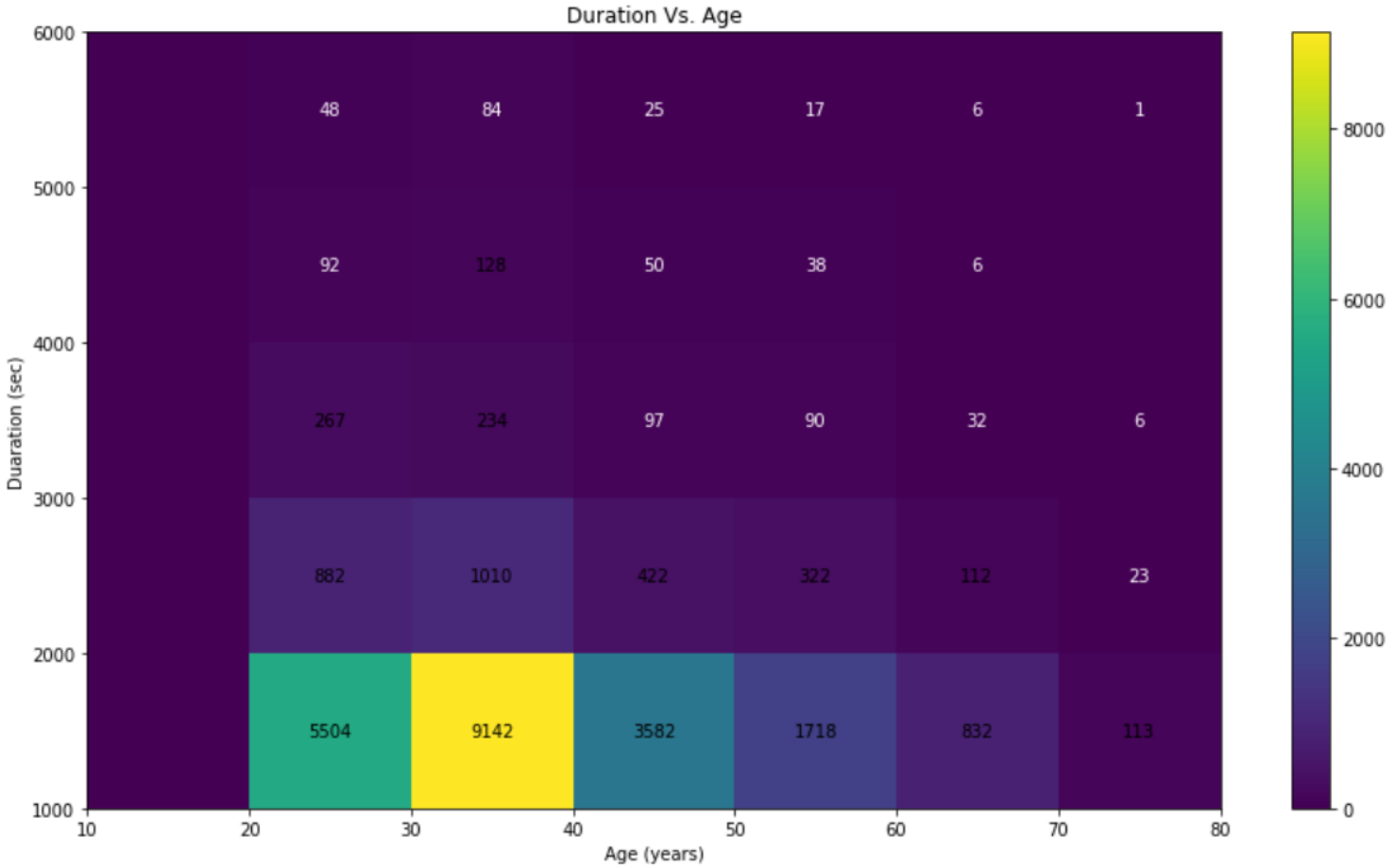

Distribution of Age
Count
Age (years)
0
2000
4000
6000
8000
10000
20
30
40
50
60
70
80
90


Distribution of Age
Density
0.00
0.01
0.02
0.03
0.04
0.05
20
30
40
50
60
70
80
90
100
Age (years)

Duration vs. Age

From these plots, we can determine the correlation between the age and duration as it a relatively stable relationship, then a slight negative relationship, and then strengthened with age, so the duration decreases.

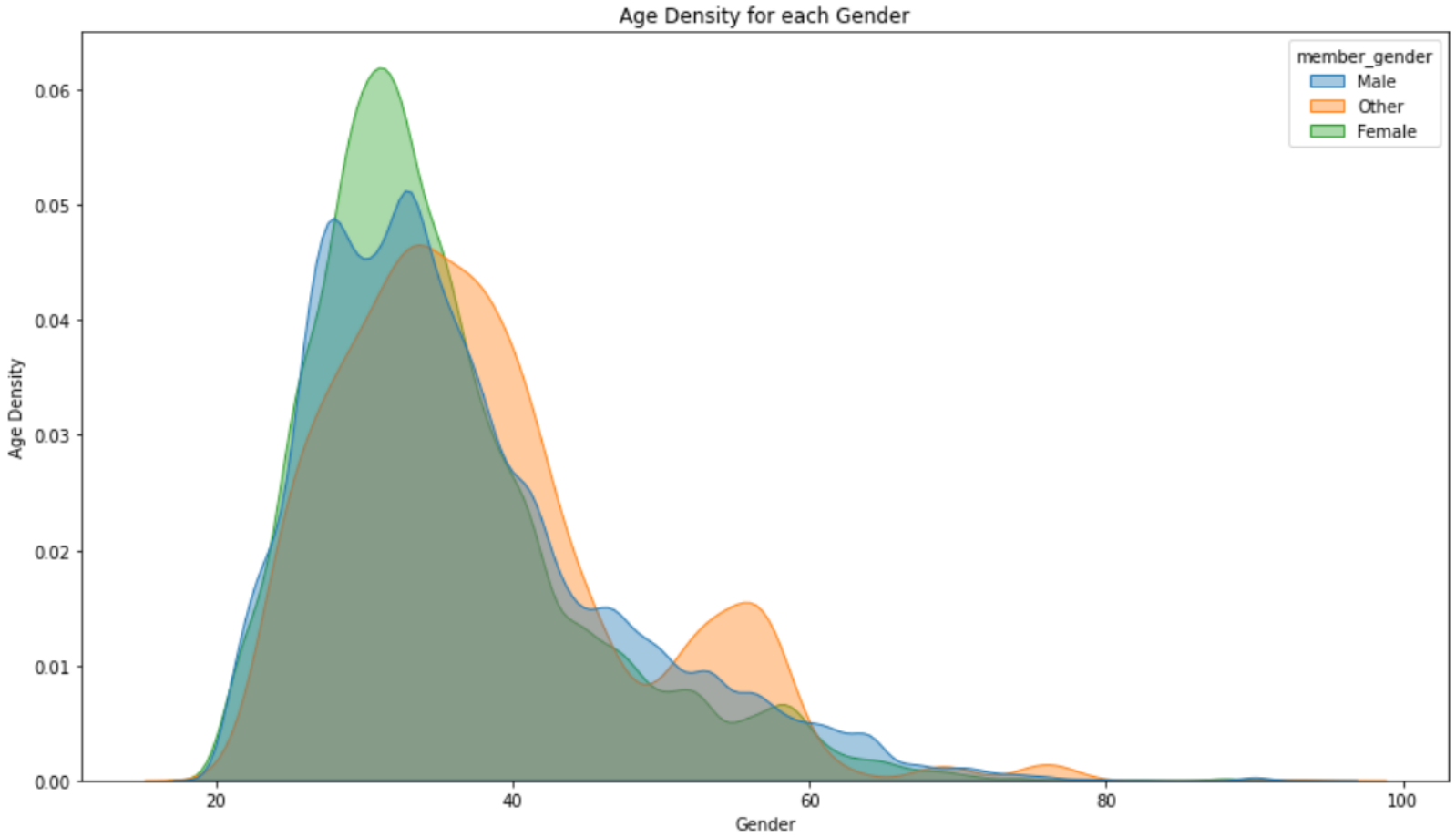

Duration Vs. Age
Duaration (sec)
Age (years)
0
2000
4000
6000
8000
10000
20
30
40
50
60
70
80
90
100

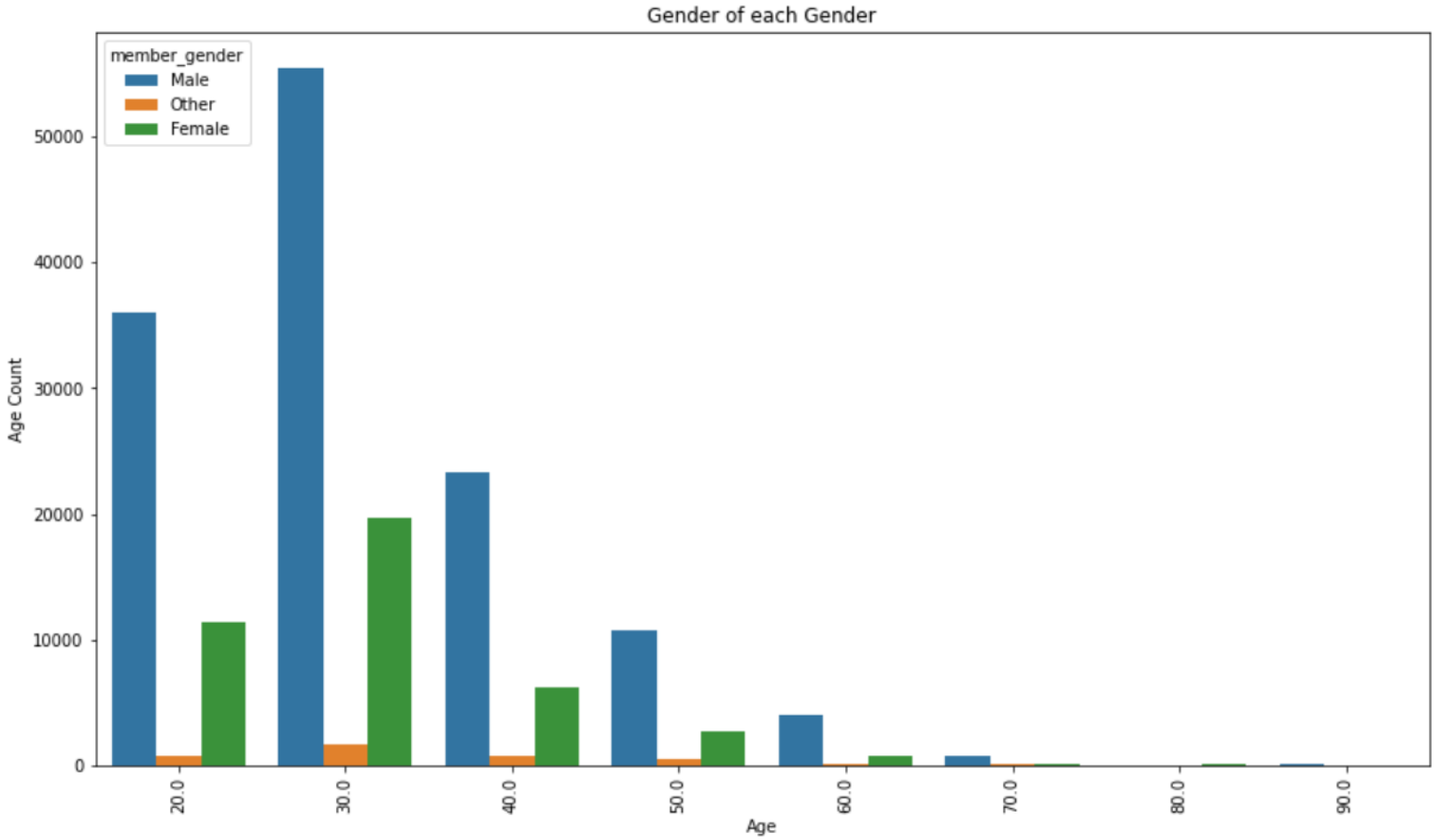

Duration Vs. Age
Duaration (sec)
Age (years)
6000
5000
4000
3000
2000
1000
10
20
30
40
50
60
70
80
48
84
25
17
6
1
92
128
50
38
6
267
234
97
90
32
6
882
1010
422
322
112
23
5504
9142
3582
1718
832
113
8000
6000
4000
2000
0

The Surprise!

Gender Vs. Age

We can see that the density of ages for females hit the higher values and comes after females the male's one then the other gender and noticed that the other gender age density is jumped again in range 50 to 60 years old which it is a surprise!.

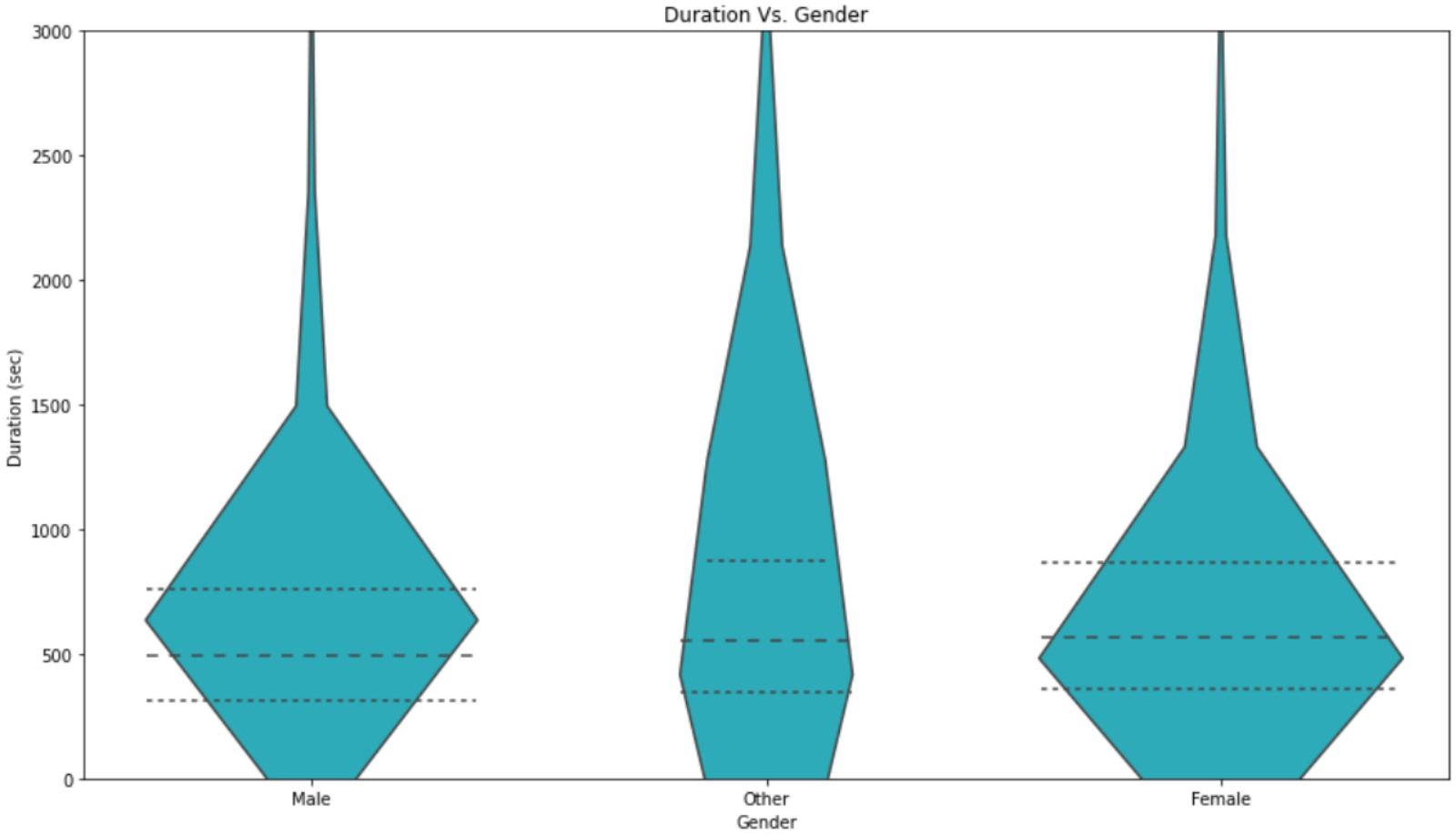

Age Density for each Gender
member_gender
Male
Other
Female
Age Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
20
40
60
80
100
Gender


Gender of each Gender
member_gender
Male
Other
Female
Age Count
50000
40000
30000
20000
10000
0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
Age

Males dominate in all age groups and we can see that around the age of thirty is the highest one between all genders.

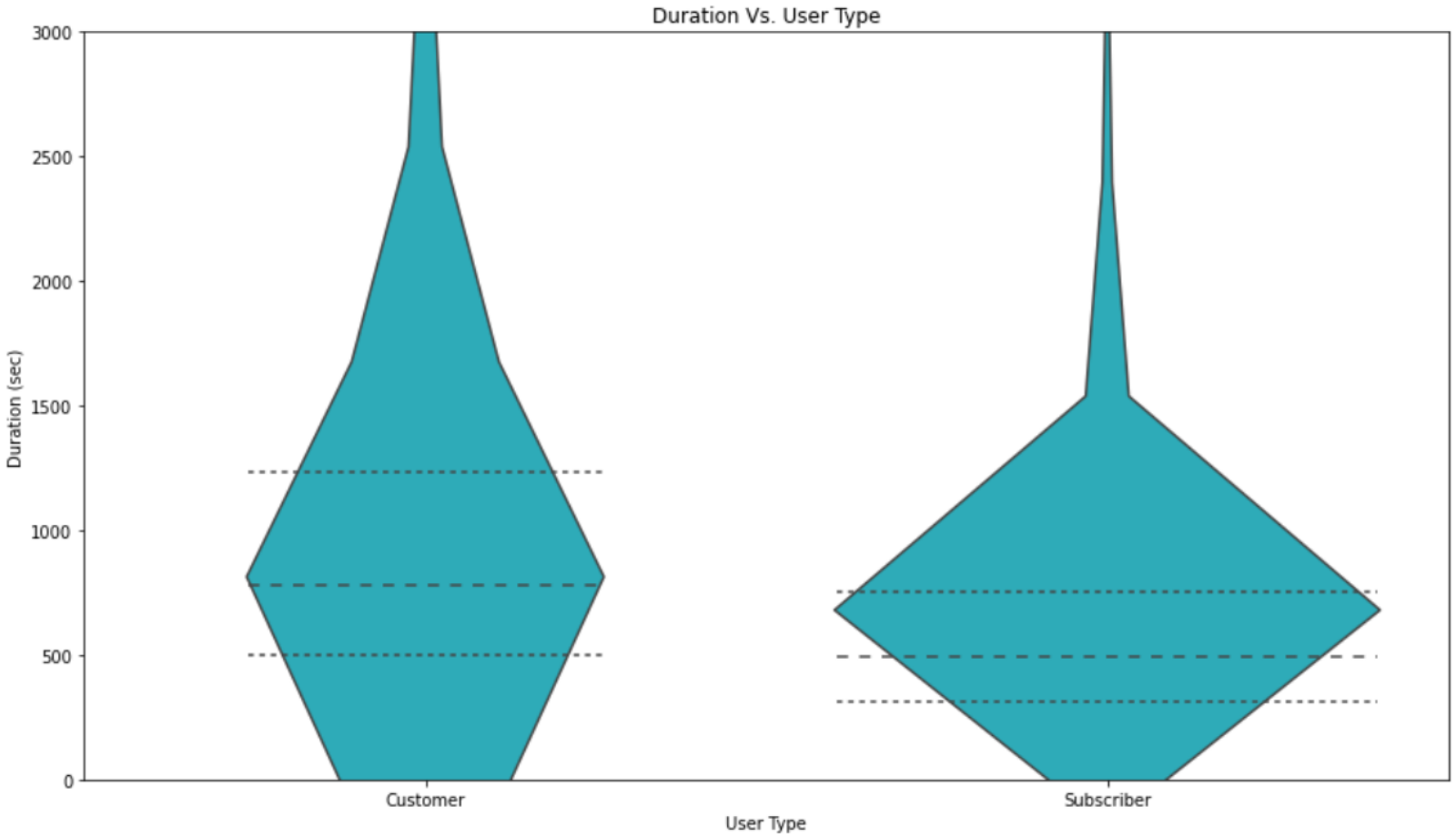
Duration vs Gender

we can see that the female's median duration is higher than the males and have the longer trip than males although the number of males is more, and the other gender has the longest trips.


Duration Vs. Gender
3000
2500
2000
1500
1000
500
0
Duration (sec)
Male
Other
Female
Gender

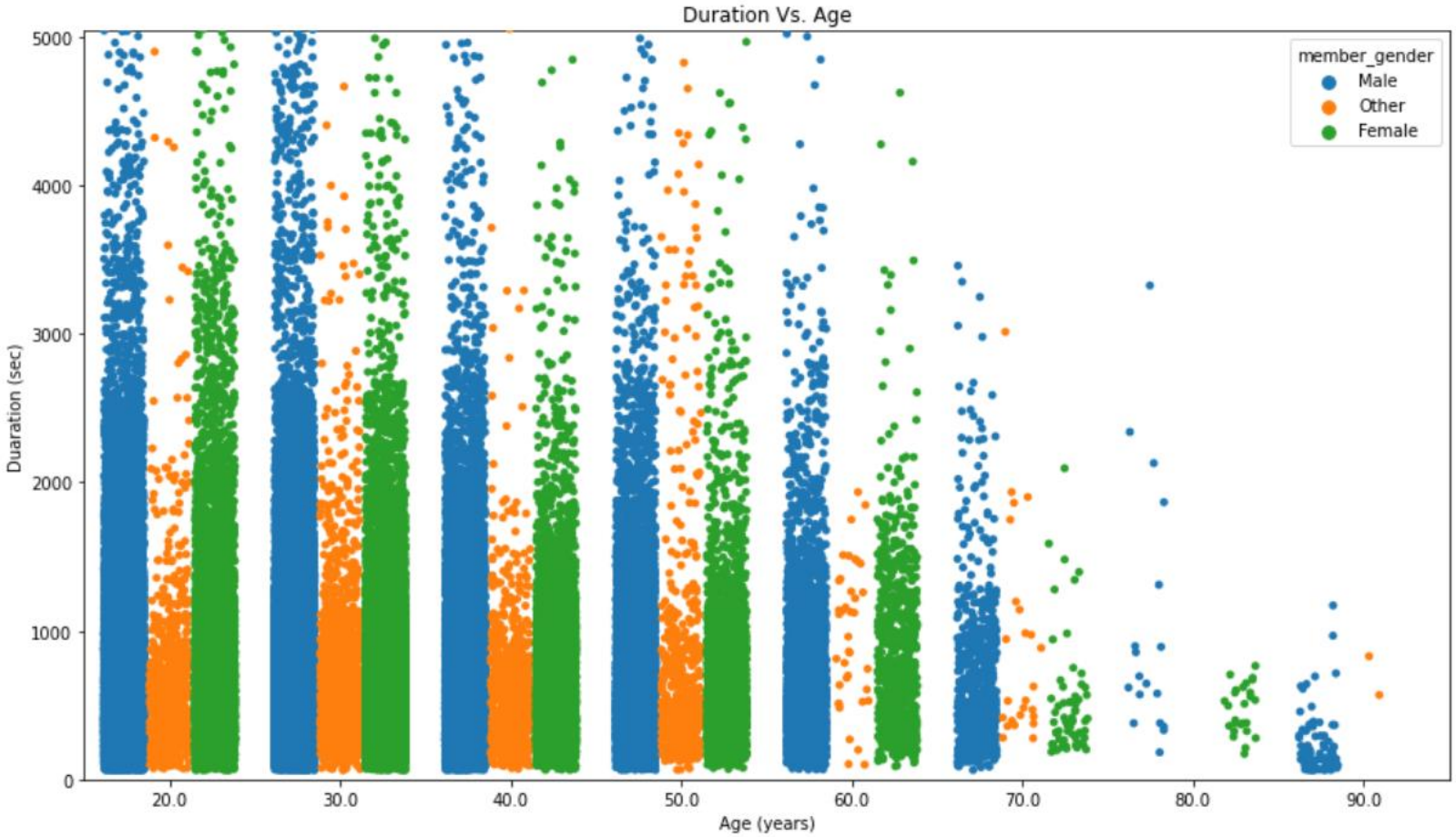
Duration vs. User Type

We can see from this plot that the Customer has a longer trip duration than the subscriber.


Duration Vs. User Type
3000
2500
2000
1500
1000
500
0
Duration (sec)
Customer
Subscriber
User Type

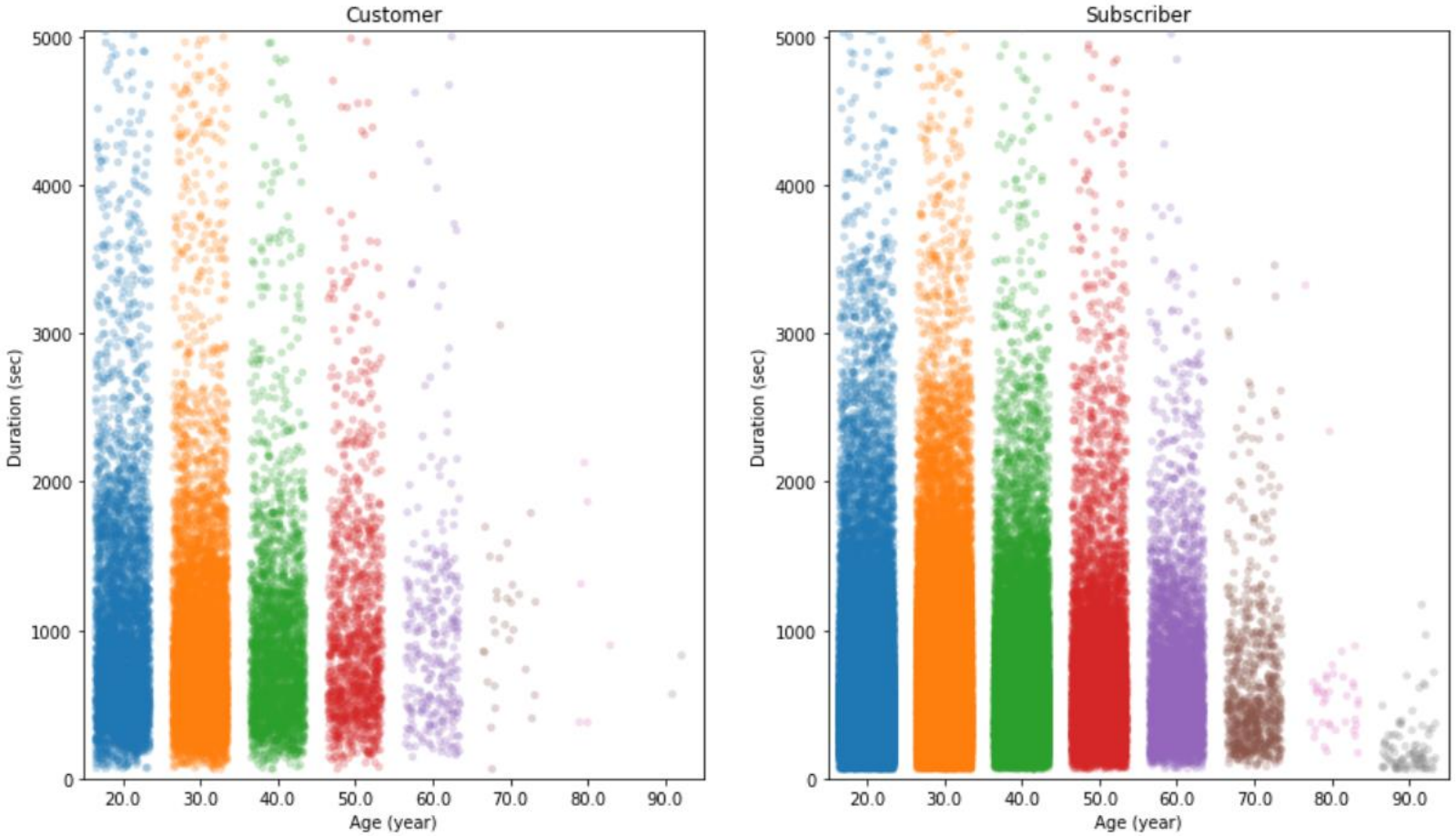
Duration vs. Age to each Gender

It is obvious that the 20 to 40 age range is dominant for all genders in all duration levels and also we noticed that the males have the highest trip durations and females come after, and finally the other gender.


Duration Vs. Age
member_gender
Male
Other
Female
Duaration (sec)
0
1000
2000
3000
4000
5000
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
Age (years)

Duration vs. Age for each User Type

We can see from these plots that there is slight tilt to higher age with high trip duration for subscribers.


Customer
Subscriber
Duration (sec)
Age (year)
0
1000
2000
3000
4000
5000
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0