

Ford Gobike EDA



Investigation Overview

The goal of the analysis is to find the key insights from the data, identifying business questions and providing solutions, and extract helpful information that helps the business in several fields like the marketing strategies they make.

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.

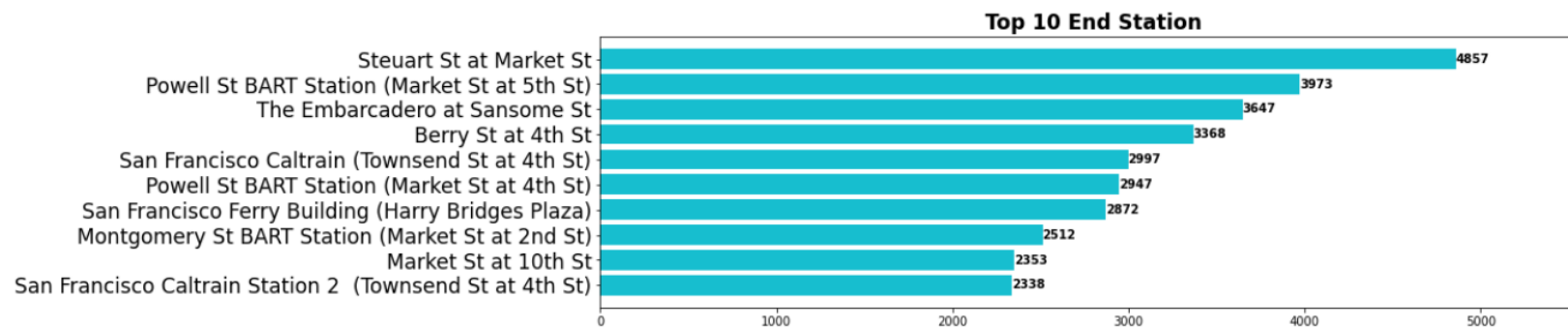
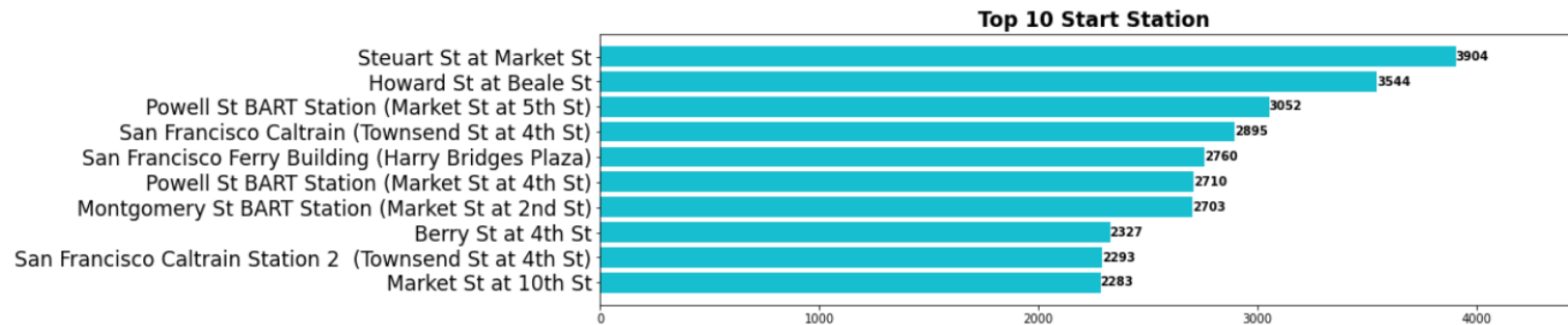


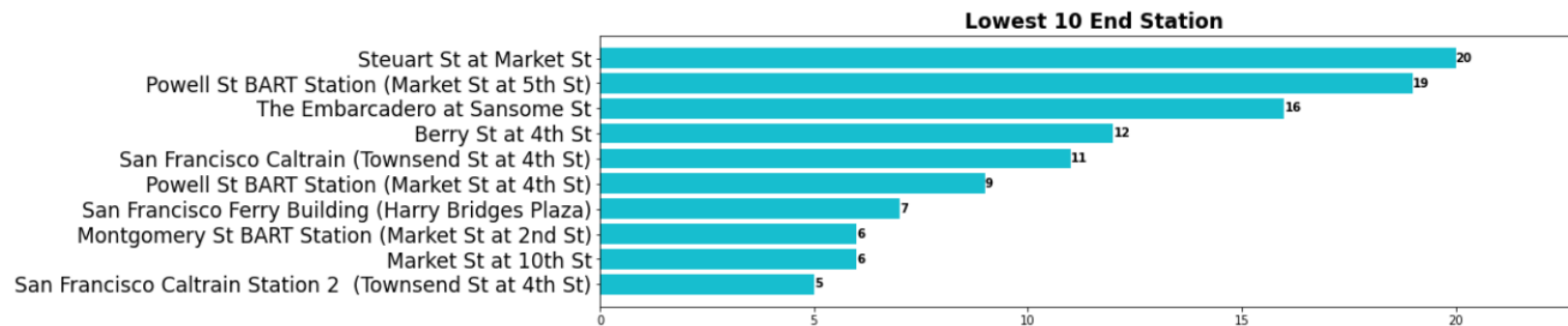
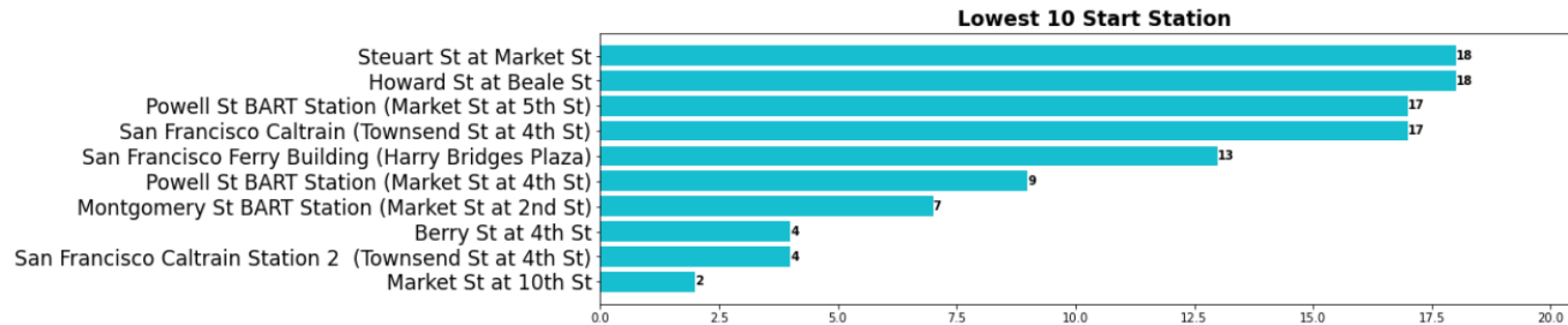



Question 1: What is the top 10 start and end stations for our trips?

Top 10 Start and End Stations










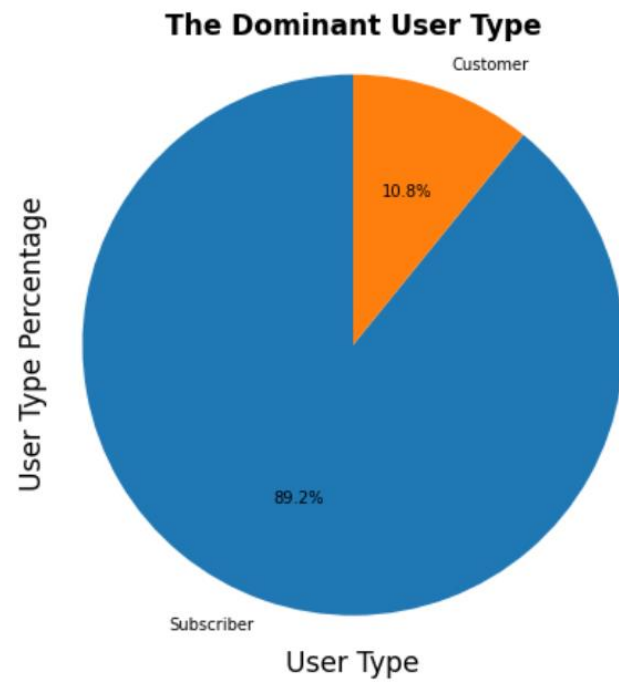
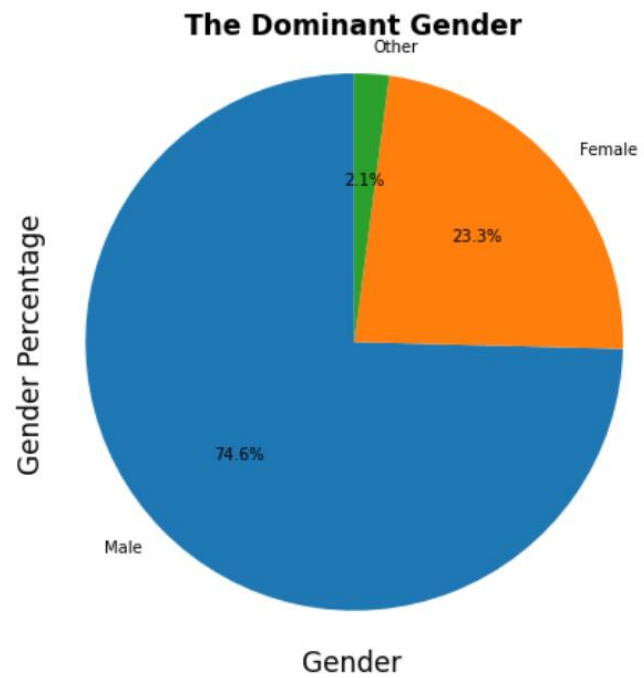
We have determined the top 10 start and end stations. This information tells us how to distribute the bikes as we might put more bikes in these stations than other. And also increasing the marketing campaign for the less used stations to grab more users.






Question 2: What is the dominant gender and 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%.

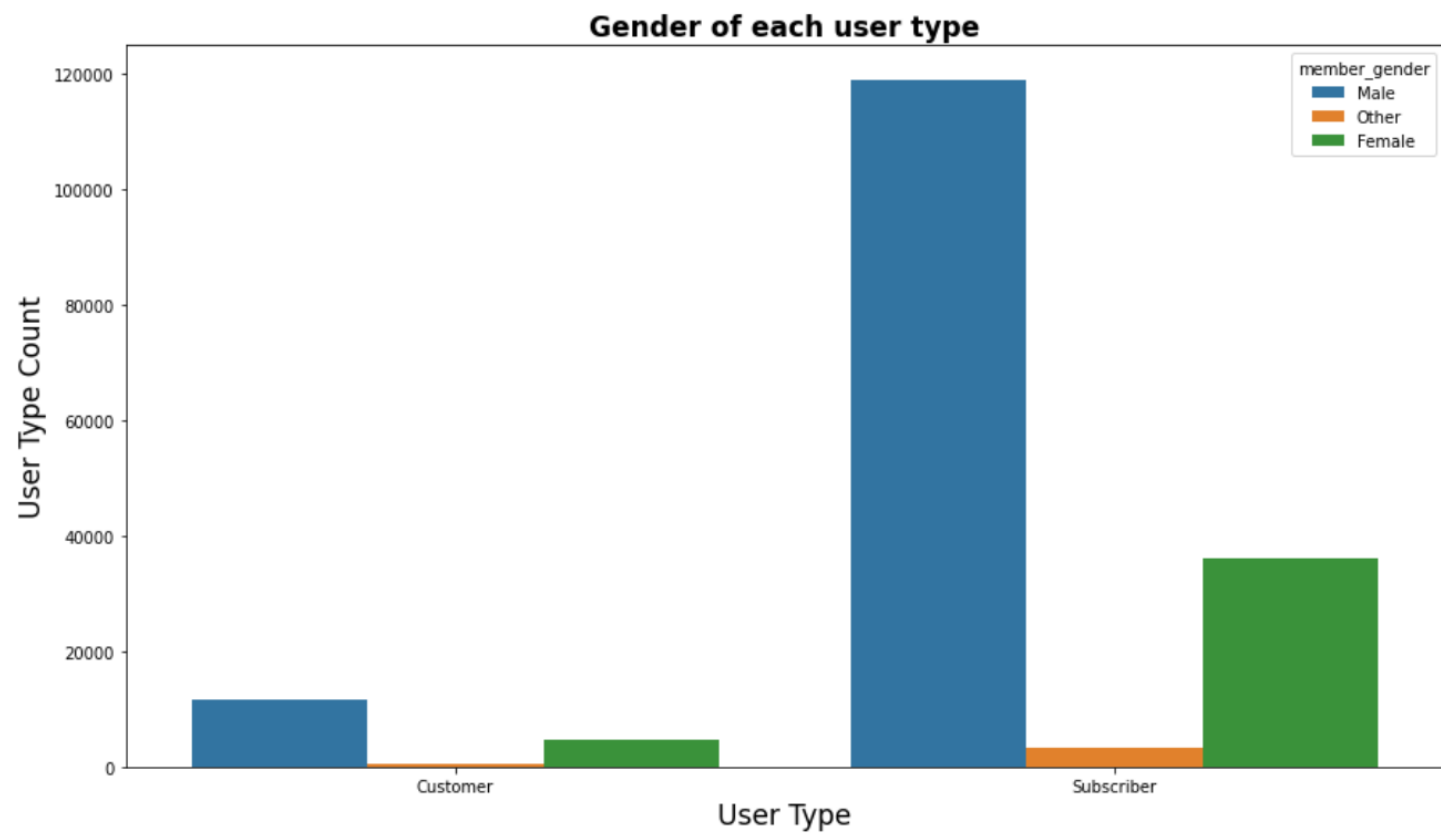
The analysis of the gender shows that males rent bikes with almost triple the frequency of the females which also can be included in marketing campaign to target the females to grab more females to the company.


I recommend make discounts for new customers or offers to grab the irregular customers.



Question 3: What are the proportion of genders for each 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%

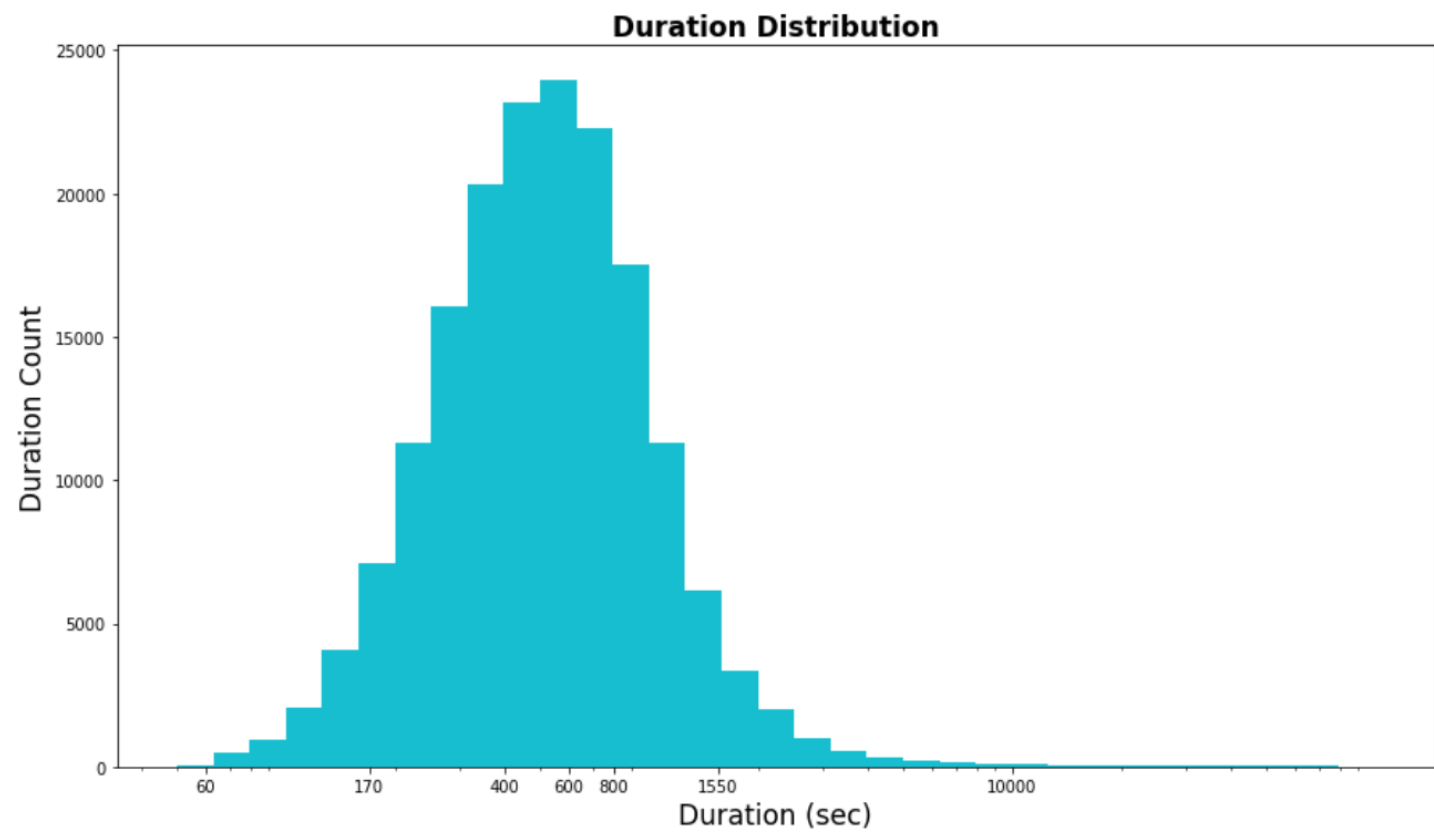
The proportions are so close which tells us that there is no relation between the gender and a user type.



Question 4: What is the most durations we have?


We will determine that with the distribution of duration.





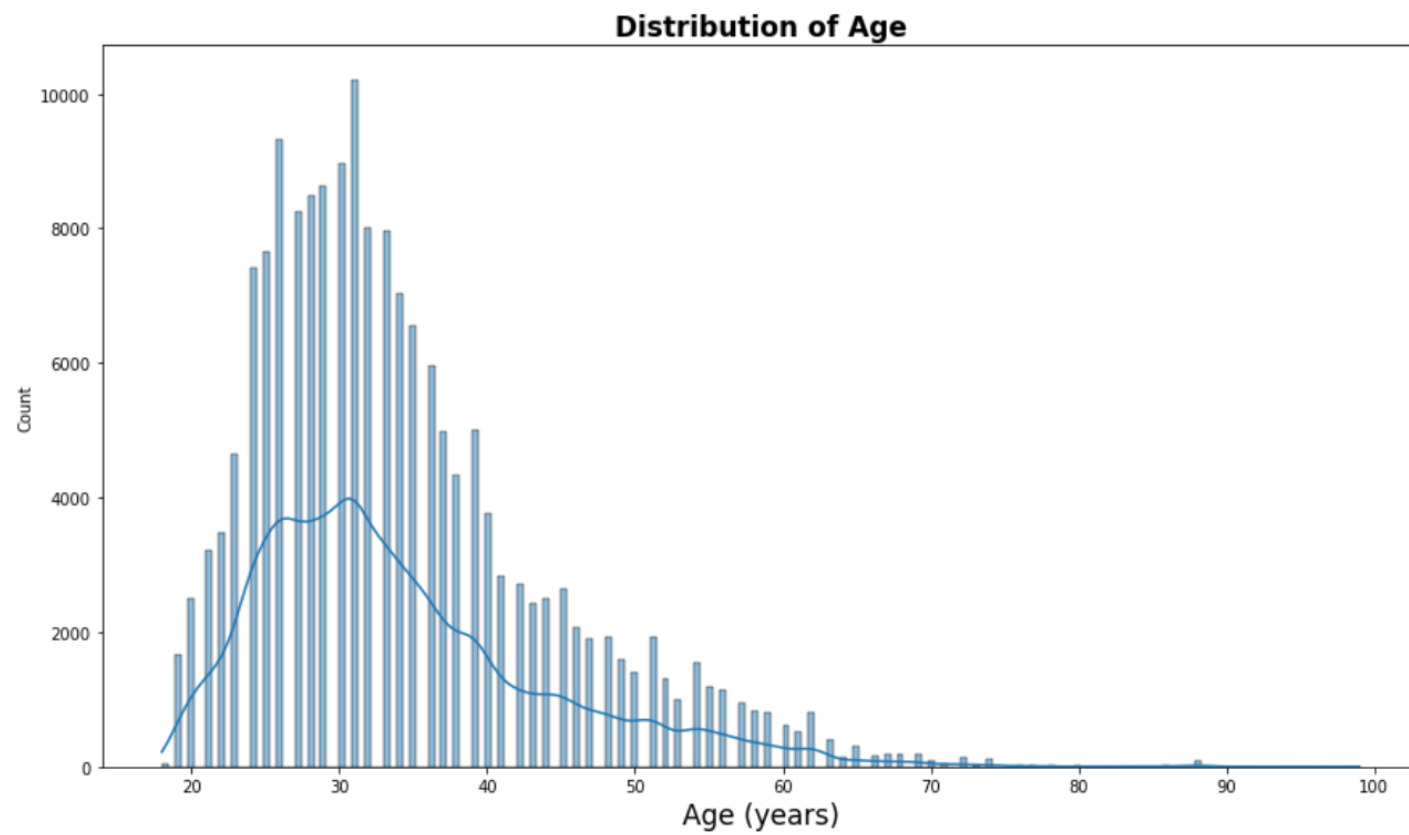
most of the trips is between 170sec(3 minutes) to 1550sec(25 minutes)
and they centered around 600sec(10 minutes).

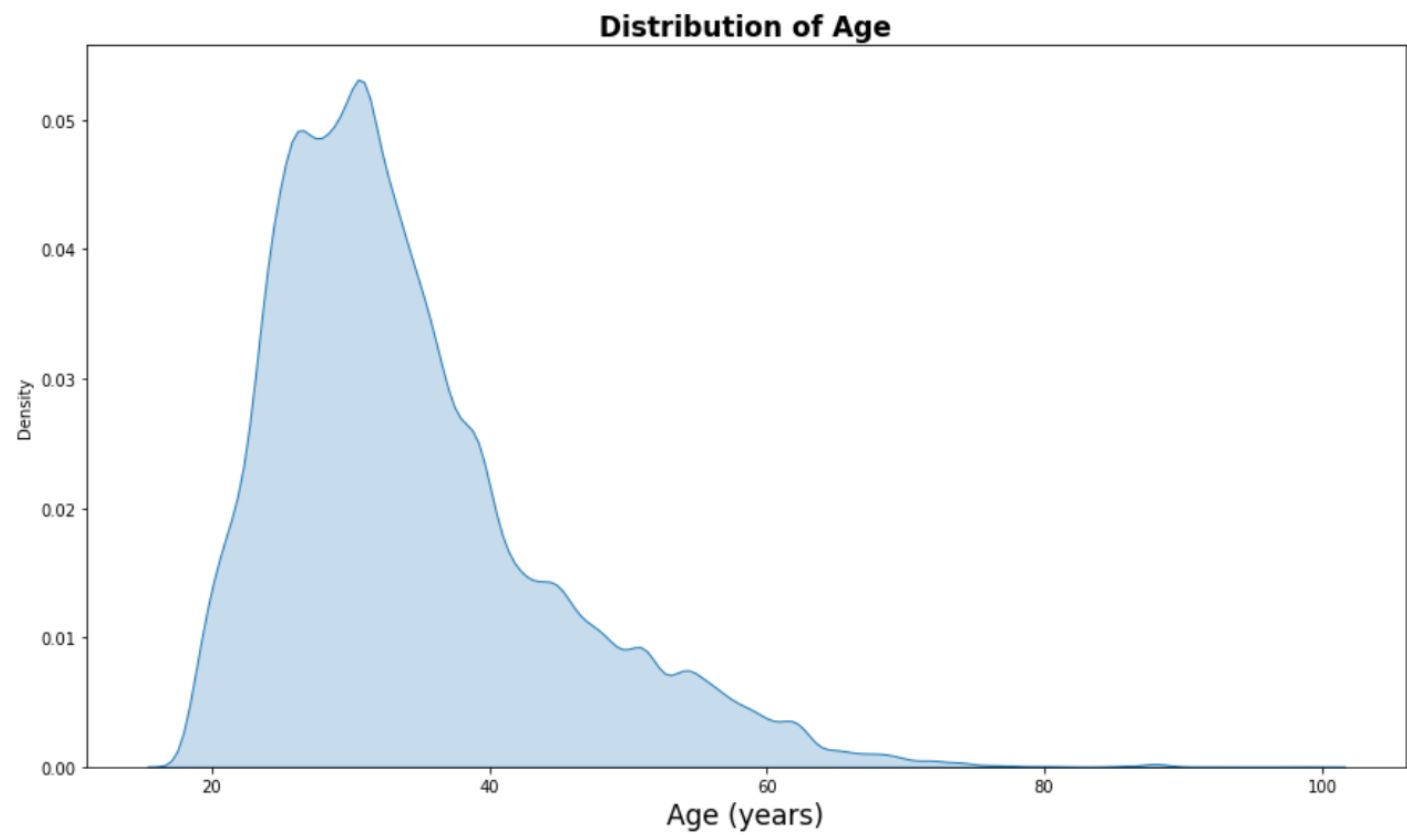



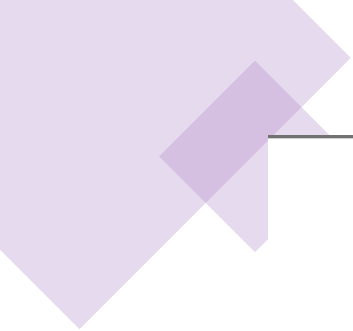


Question 5: What is the age of most of our users?










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



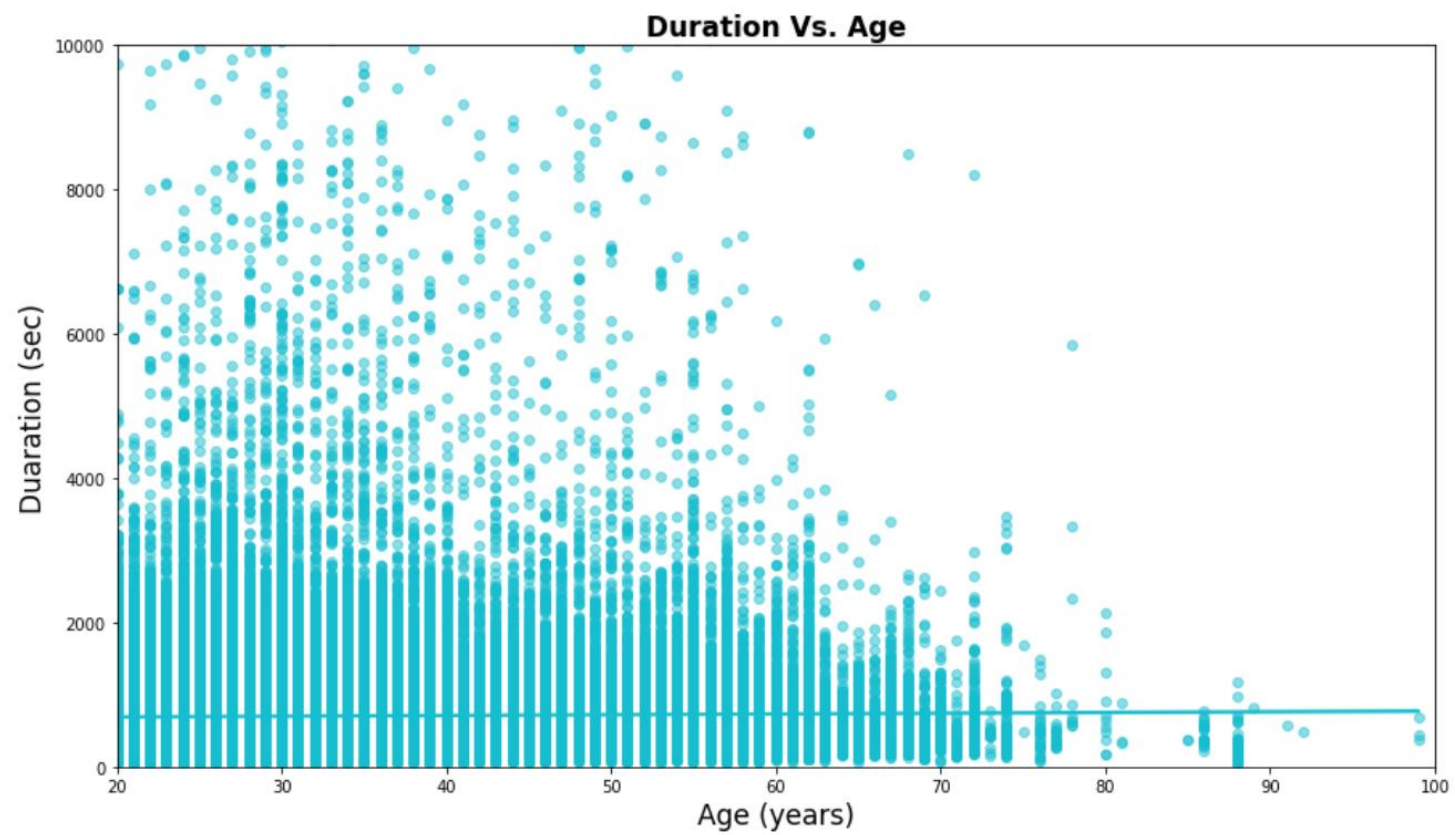


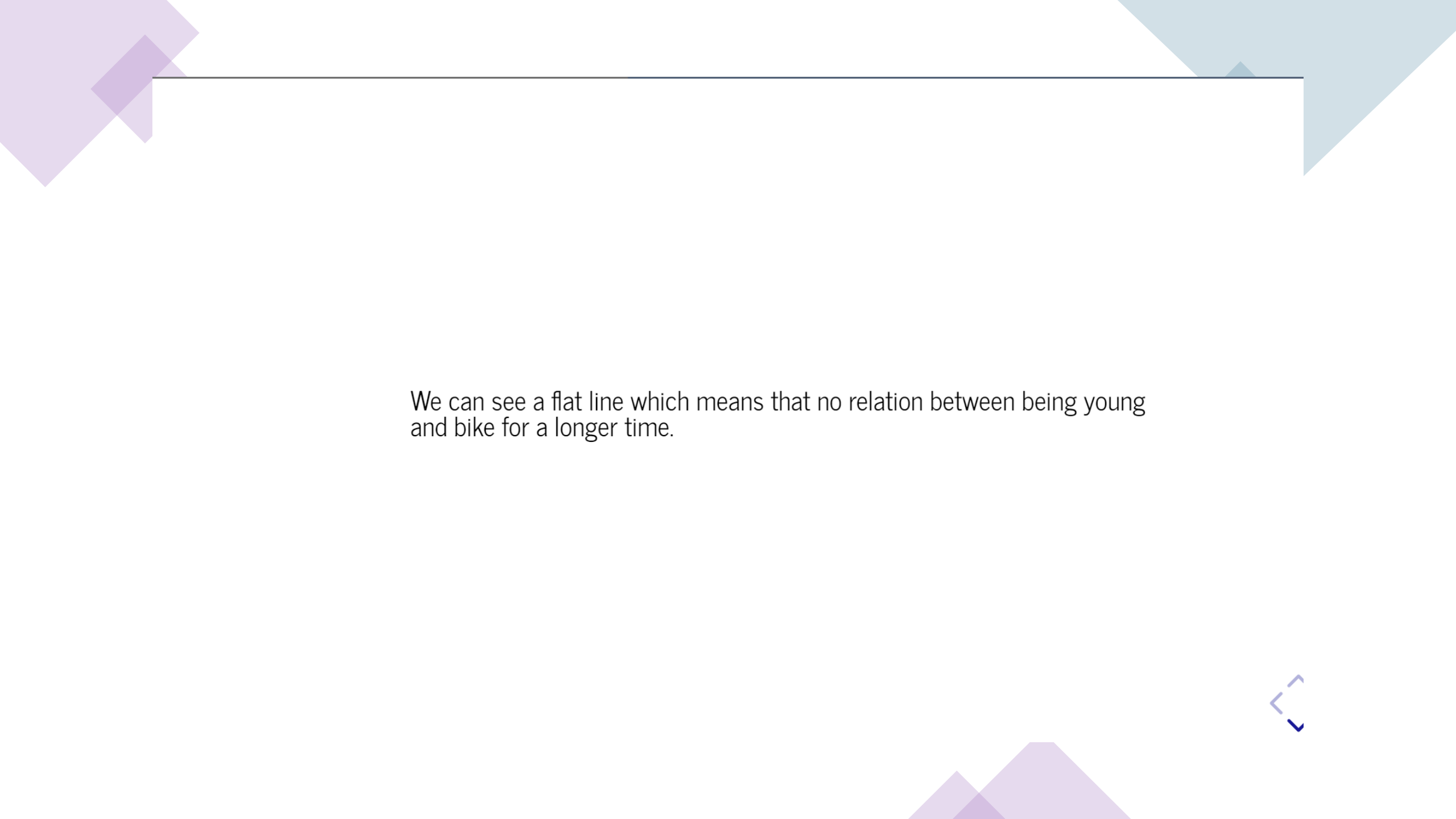
Question 6: Is there any relation between the age and trip duration (Do you think younger people bike longer than older people)?

We will determine that with several plots.


Lets see the scatter plot and heat map for ages vs durations.

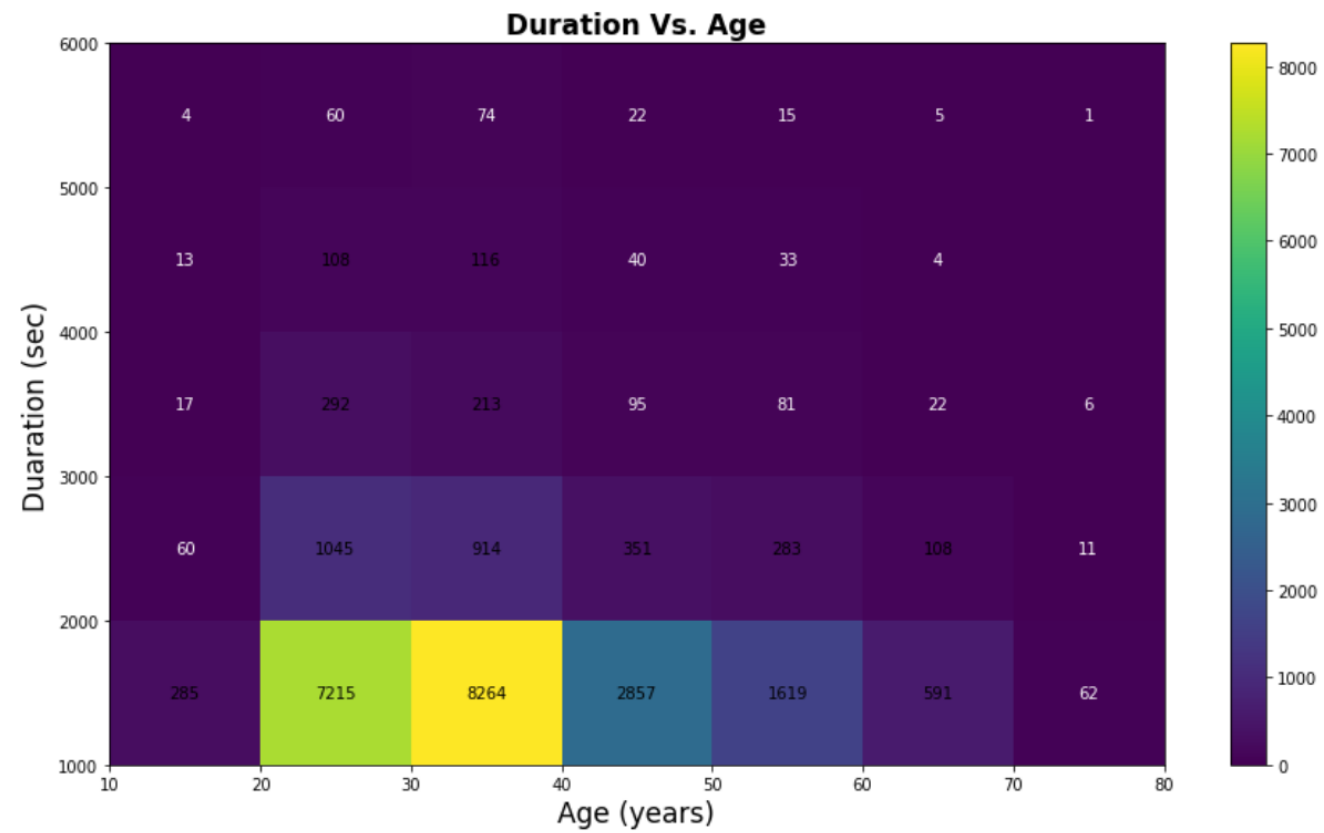







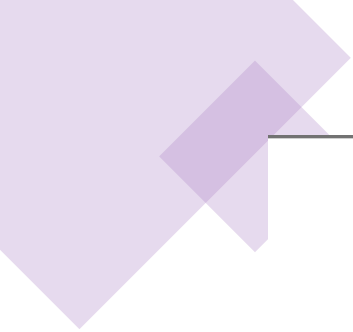
We can see a flat line which means that no relation between being young and bike for a longer time.





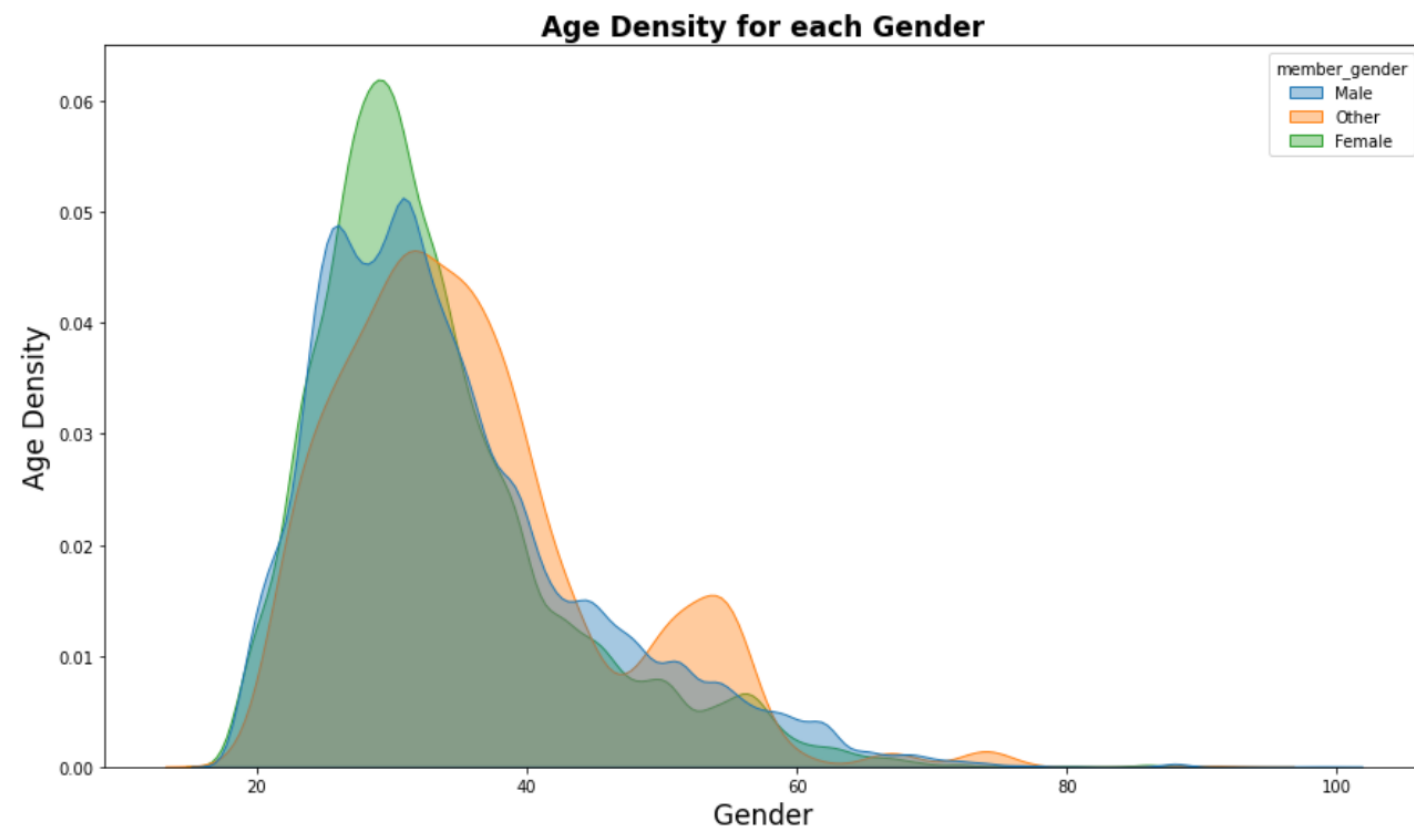
We can see that the higher duration trips occurred in the age range of 20 to 40 and under 2000sec(33 Minutes).





Question 7: What is the higher number of users ages for each gender (Is there a specific gender that has a higher count in a specific age range)?






The Surprise!

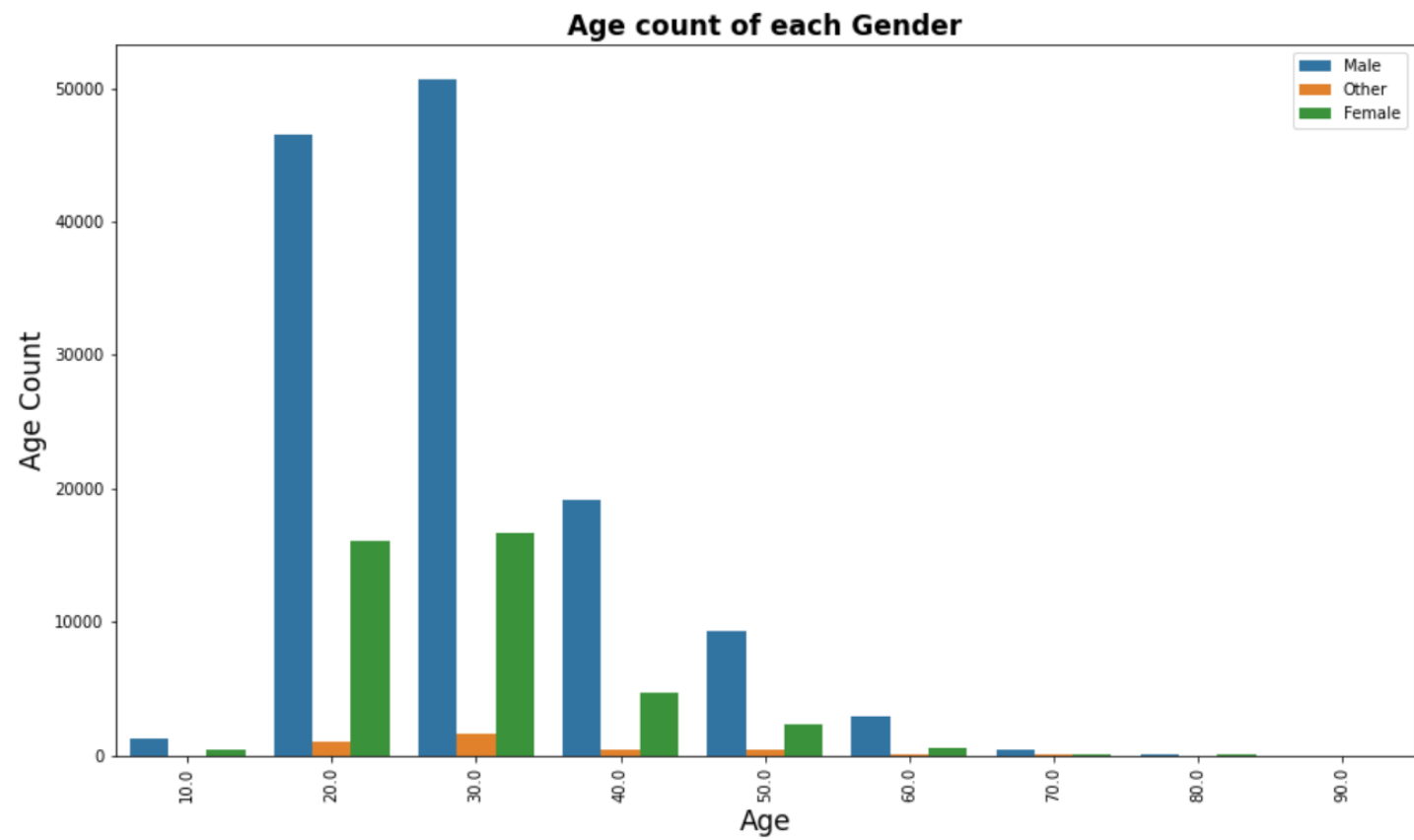
We can see that the density of ages for females hit the higher values and comes after females the male's one then the other gender and noticed that the other gender age density is jumped again in range 46 to 58 years old which it is a surprise!.





Question 8: What is the dominant gender for each age group?







Males dominate in all age groups and we can see that around the age of 30 is the highest one between all genders.

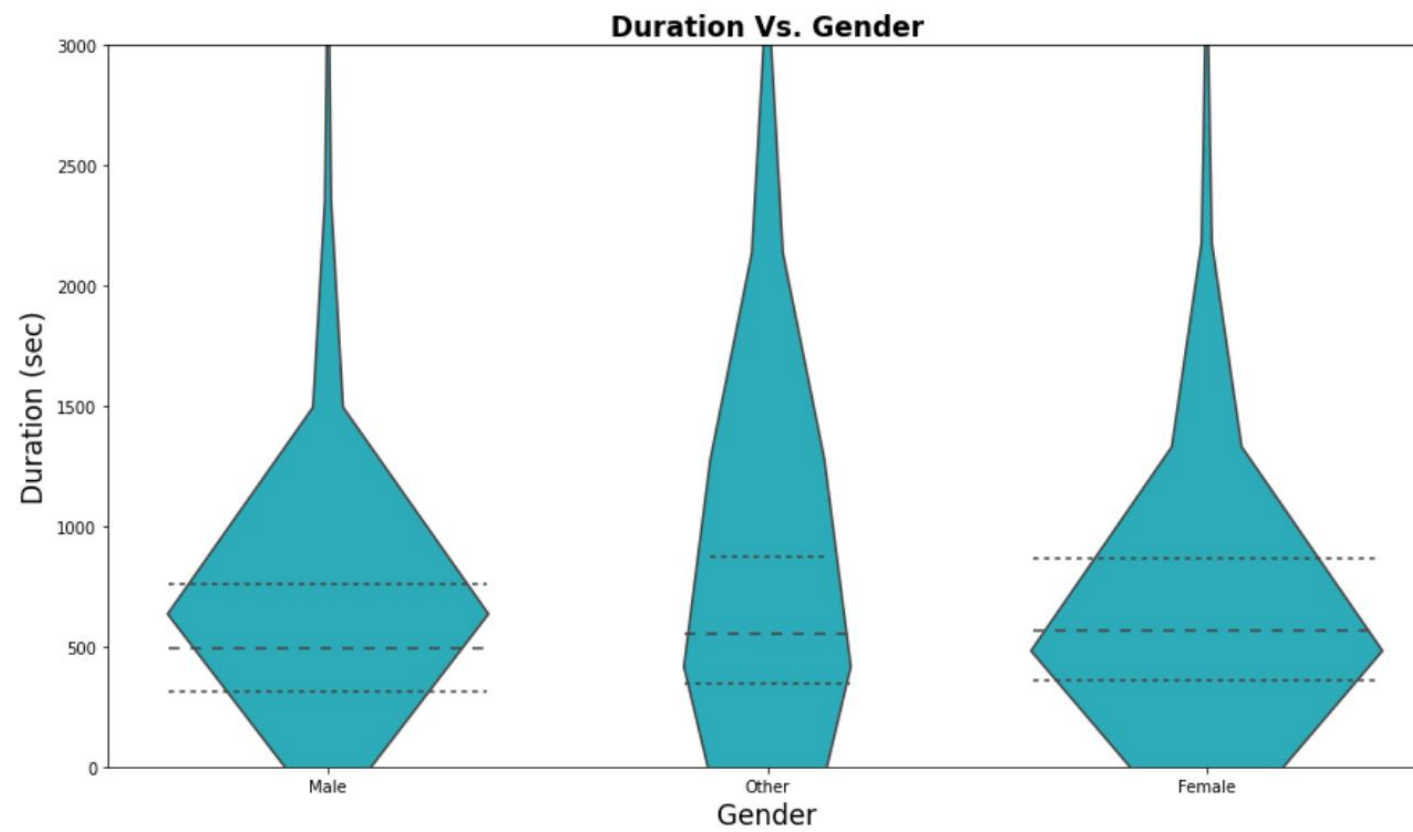
Females count considers being the same in 2 age groups (20 and 30).






Question 9: Are all genders have the same
duration distribution and duration medians?








The other gender has a wide distribution which means that they changing smoothly over all durations and I think that is back to the high jump they made from age 46 to 58.

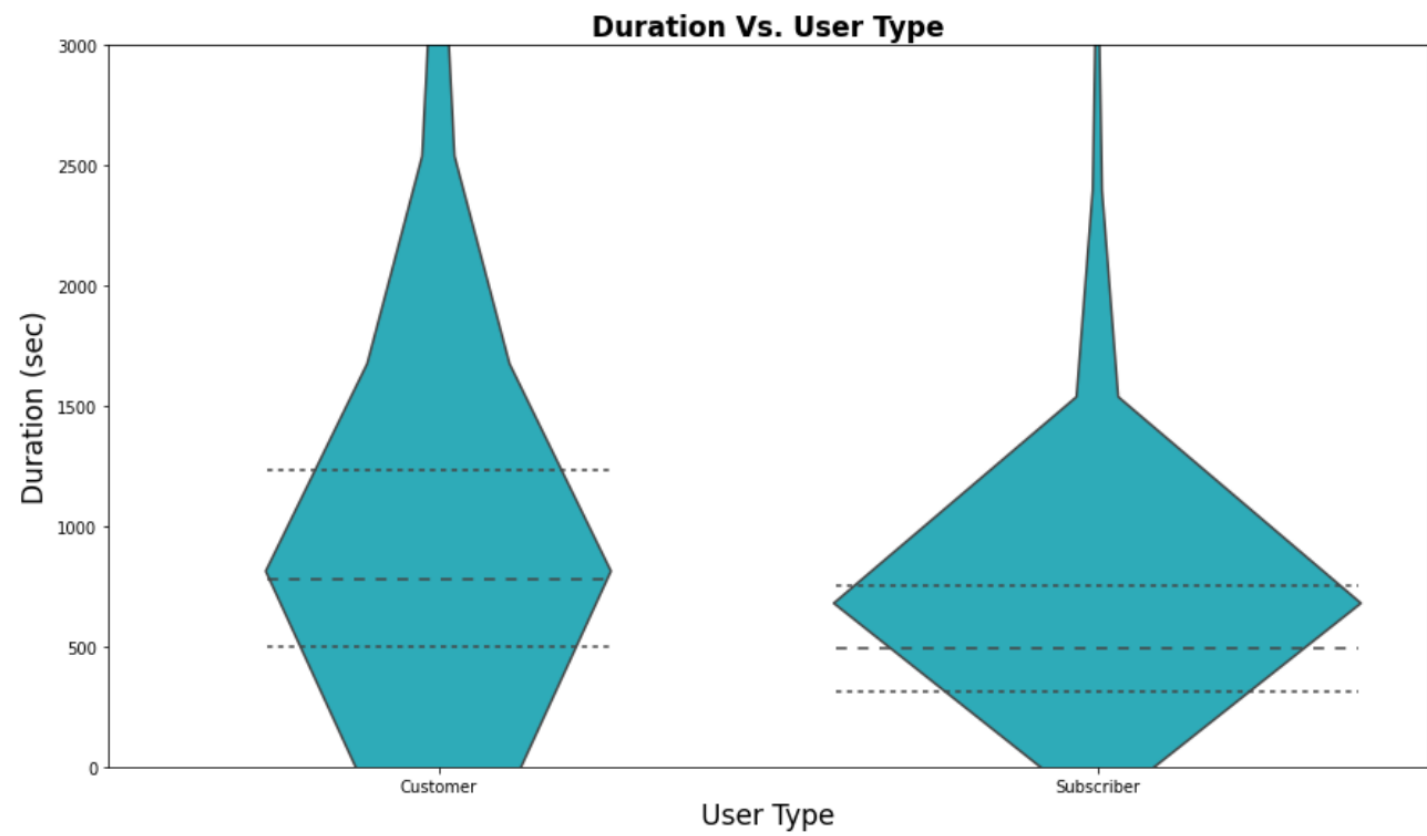
we can see that the female's median duration is higher than the males and have longer trips than males although the number of males is the higher, and the other gender has the longest trips overall.


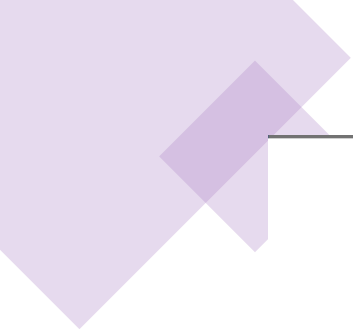




Question 10: Is there a user type that has higher duration rates?








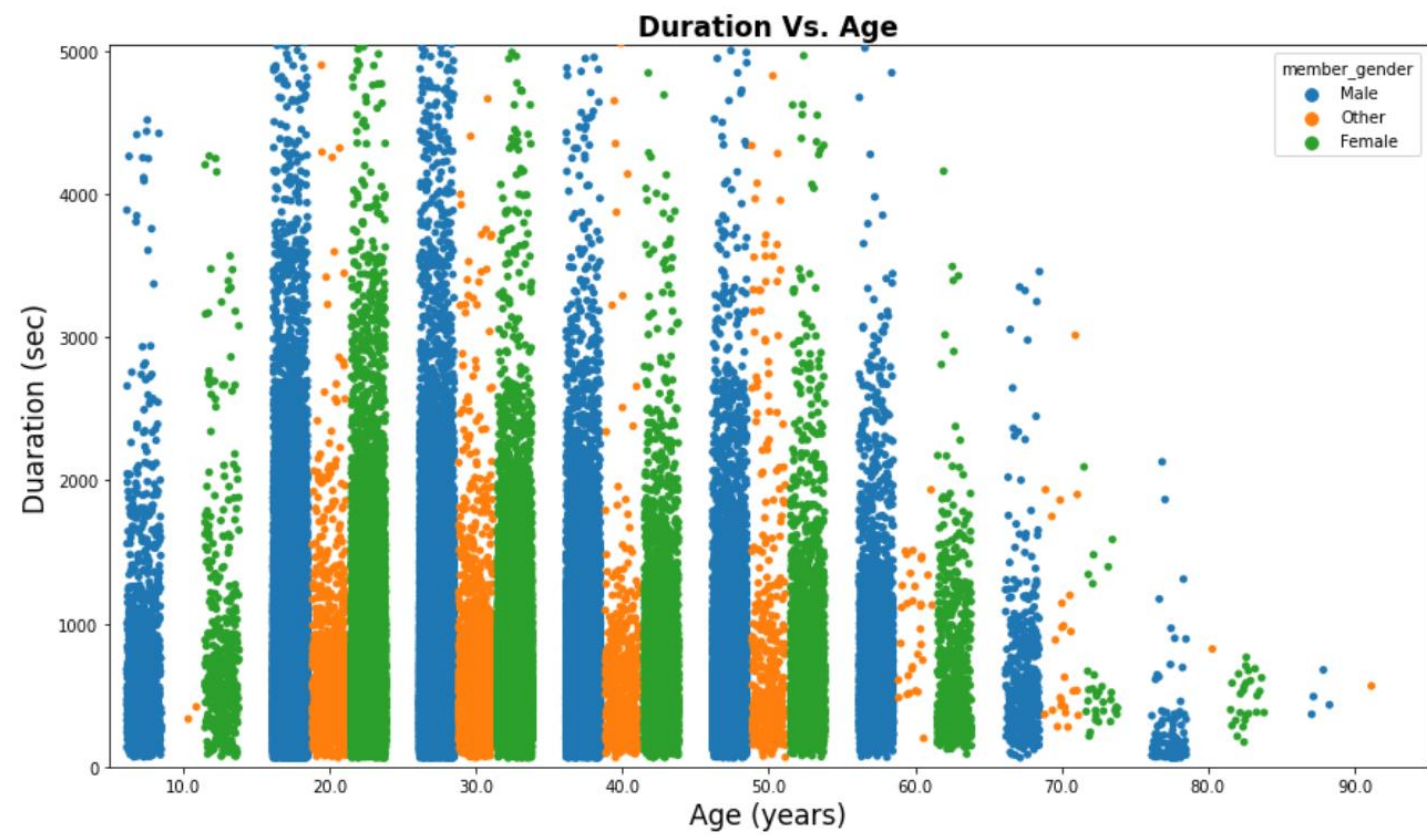
We can see that the Customer has a longer trip duration and a higher average duration than the subscriber.






Question 11: In different age groups, does the average trip duration differ for each gender?










It is obvious that the 20 to 50 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.

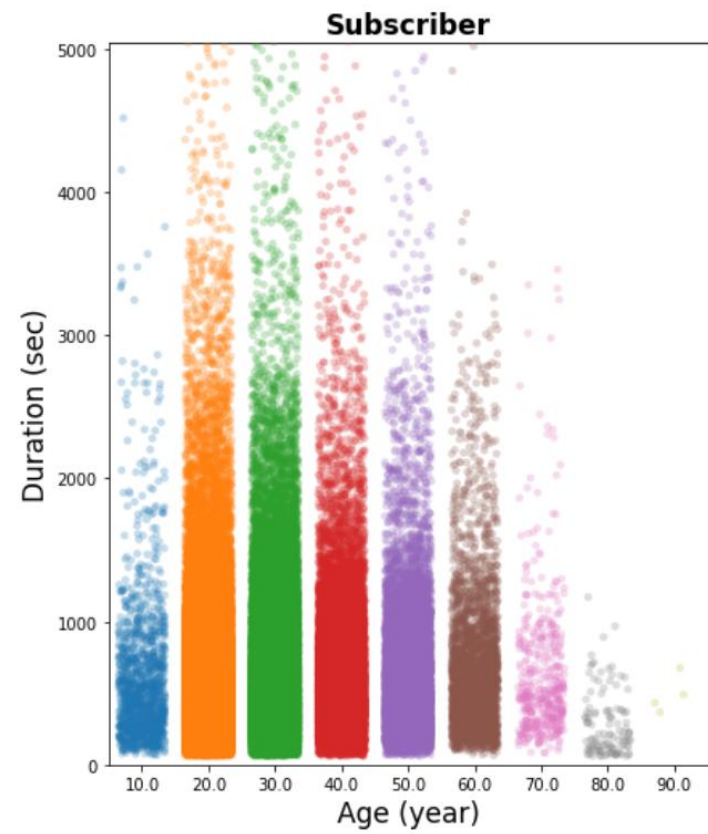
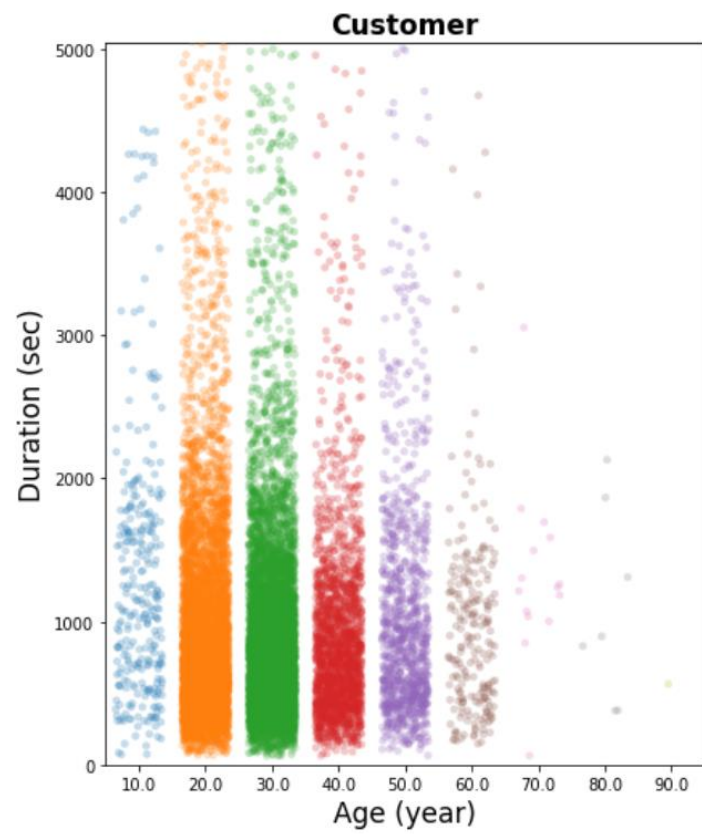
Also we noticed that the other gender is not active in ages of 10,80,90 and the females disappear at the age of 90.





Question 12: In different age groups, does the average trip duration differ for each user type?







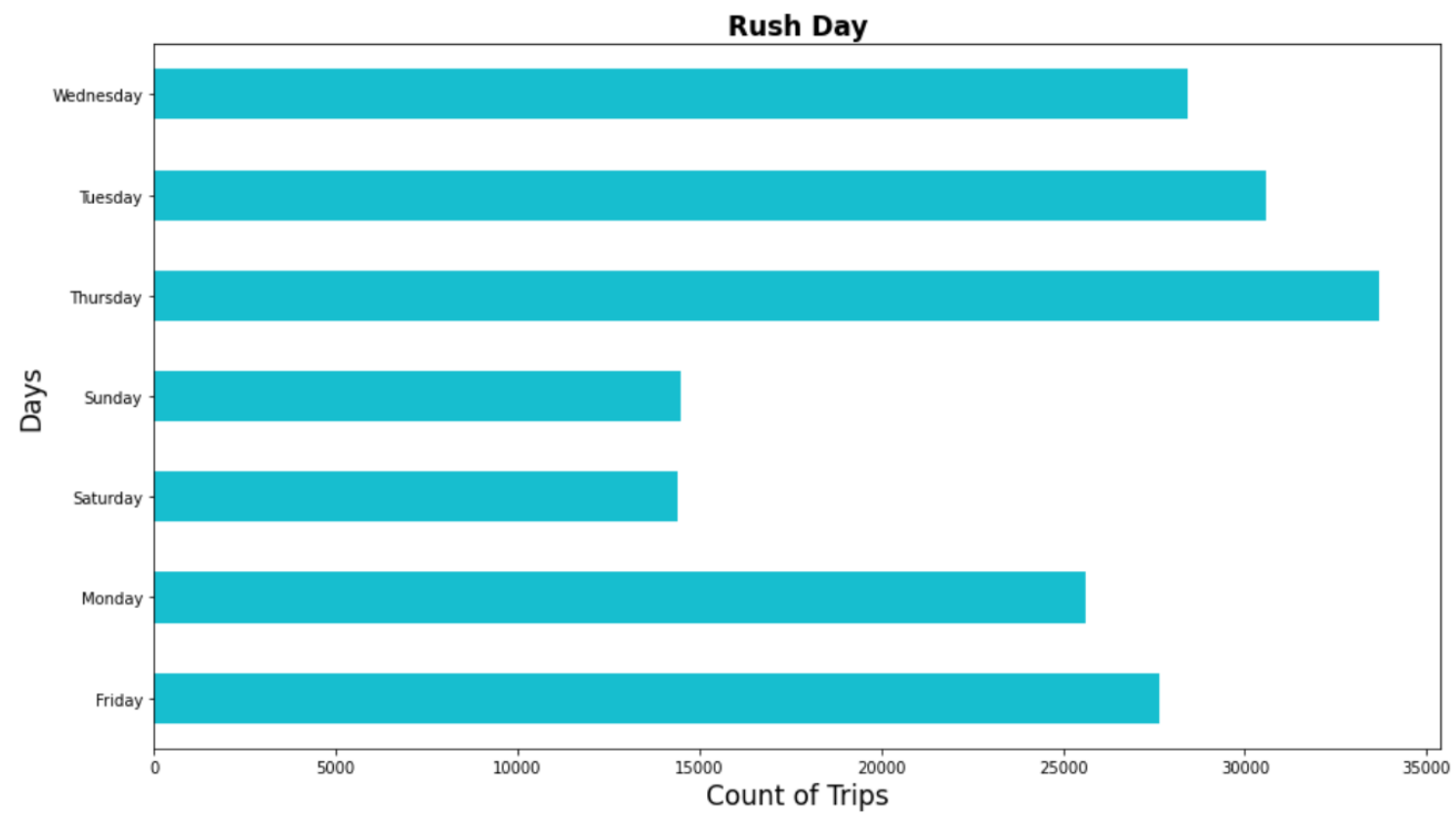
We can determine that the subscribers have a higher density in the low and mid duration, and the high durations is for the customers.


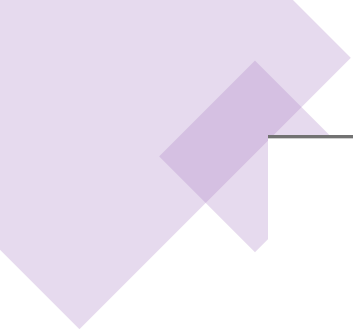




Question 13: What is the rush day?







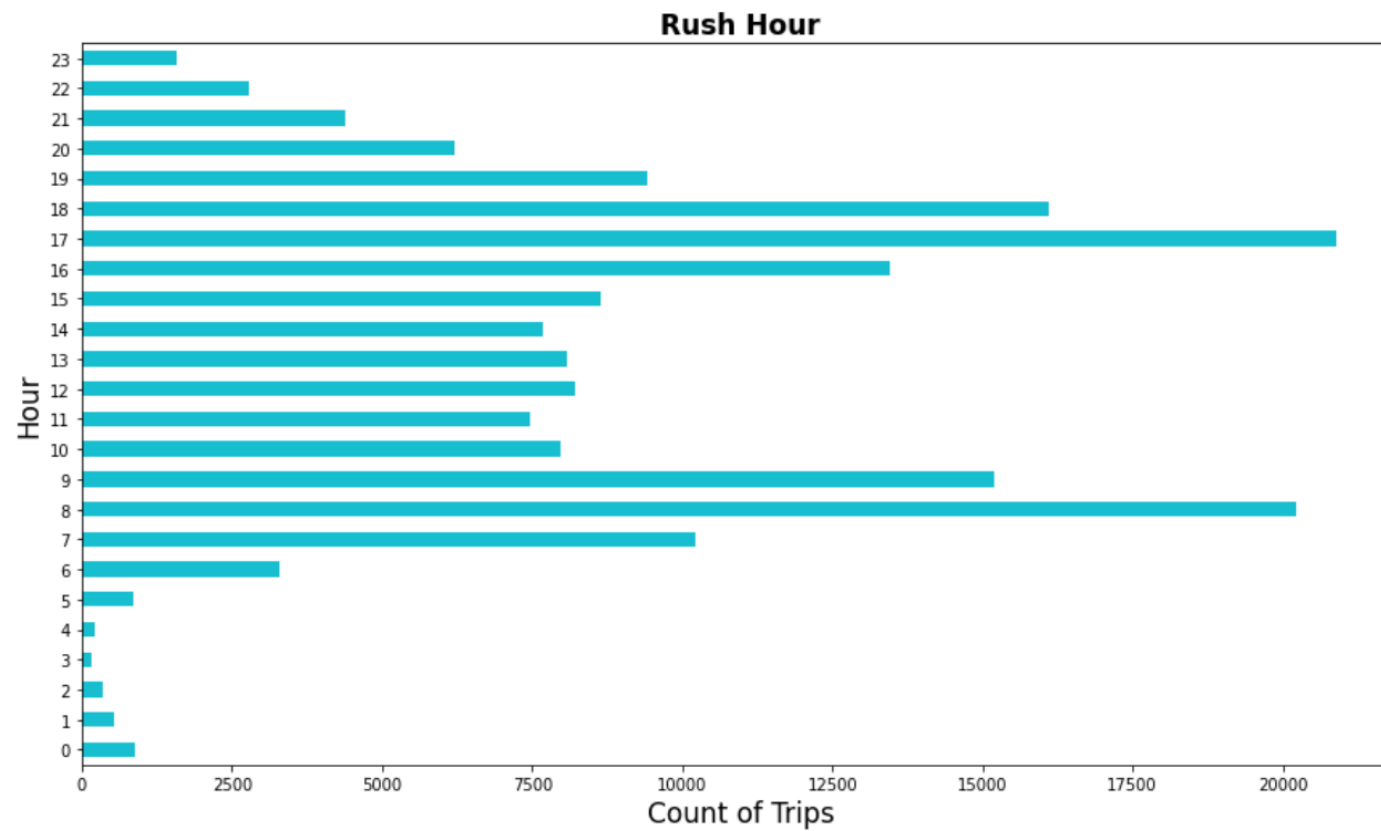
We can see that our rush day is Thursday. Sunday and Saturday is the lowest one because they are the weekend.






Question 14: What is the rush hour?





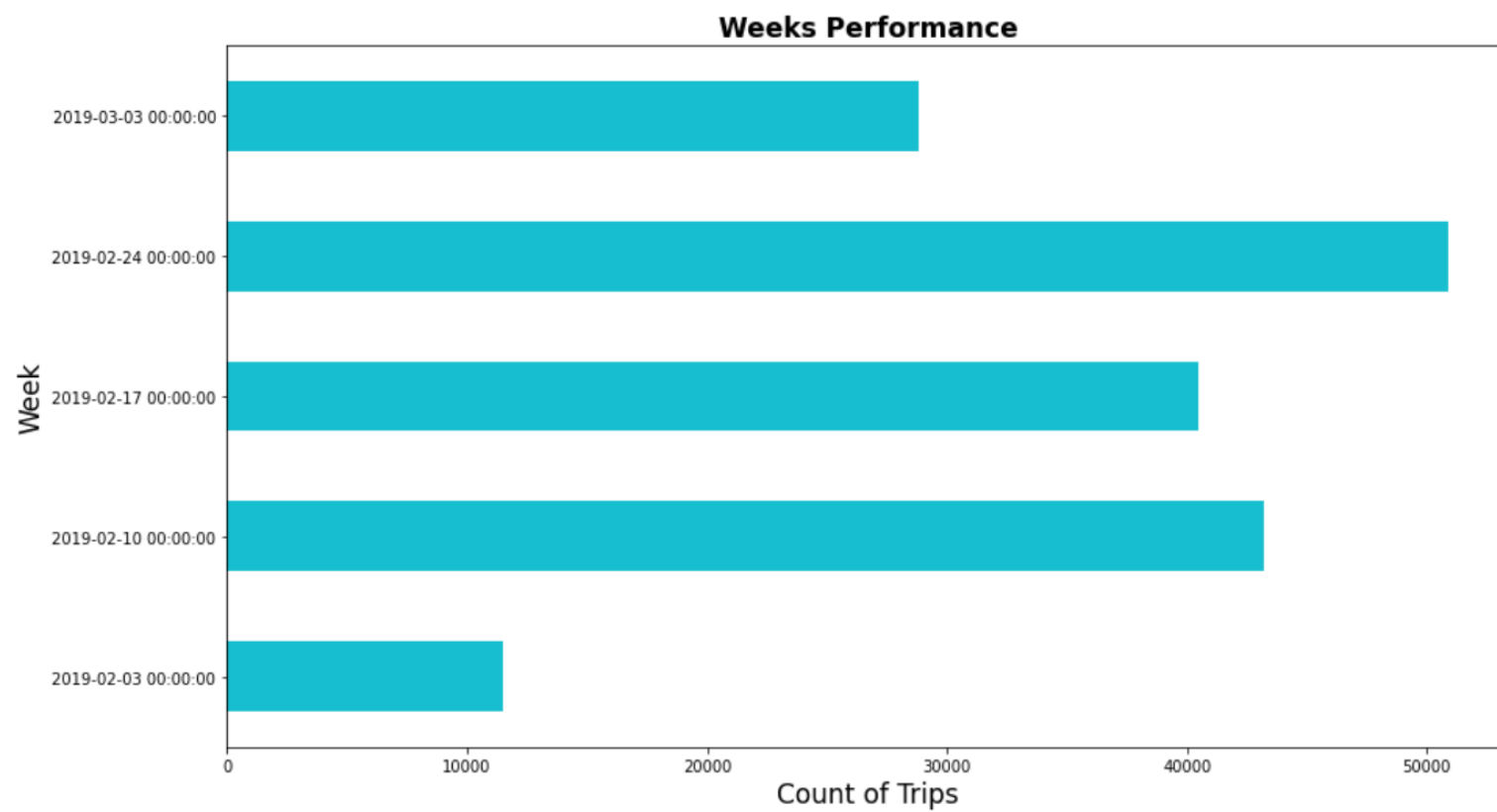
We can see that our rush hour is 7 to 9 which makes sense as they are the start work hours and also 16 to 18 which the end of the working day.






Question 15: Is there any helpful info about the weeks' performance?







We can see that the first week haven't a great performance, but the performance increased onwards until the last week is came down again, not like the first week but it still a downwards, so we recommend making promotions to increase the subscribers again taking in count that 2 of March is Texas Independence Day and 3 of march is National Anthem Day which could the reason of the downward.

