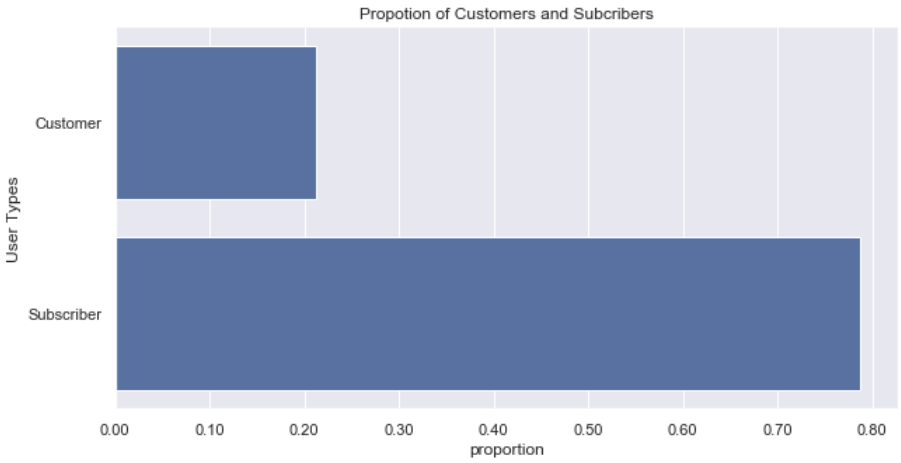
1. Choice of DataSet:

Ford GoBike System Data – DATASET.

Here you'll find Bay Wheel's trip data for public use.
The Data
Each trip is anonymized and includes:

- 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).

2. Main findings from the exploratory data analysis.



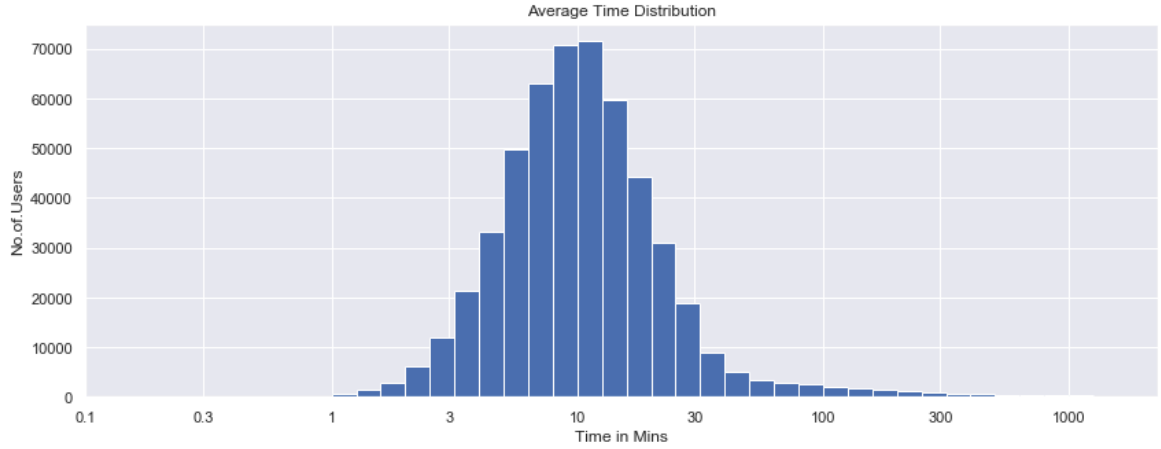
78 % Subscribers

22 % Customers

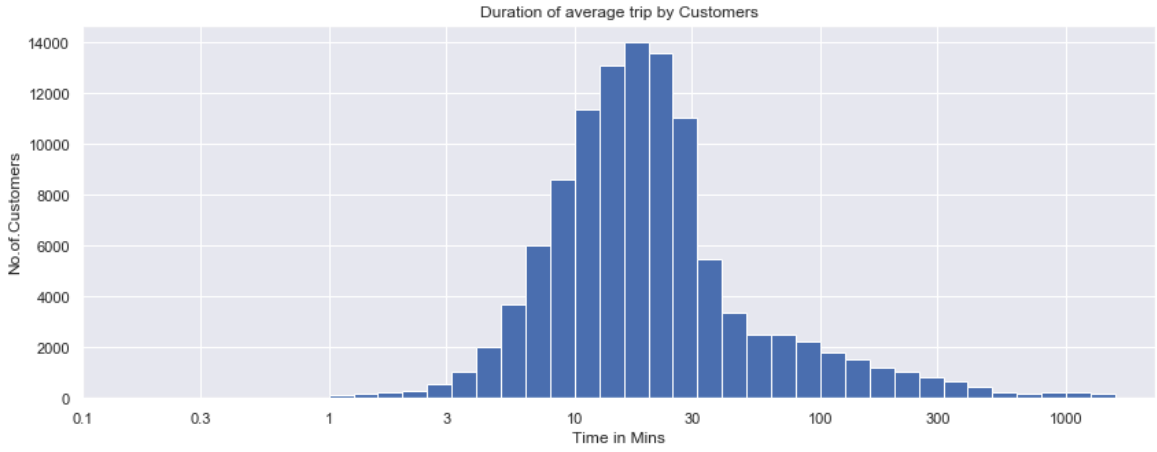
As from the above bar chart we can say that the Company has around 78% of Subscribers and 22% Customers.

Average Time of the Trips.(Uni-Variate Visualizations)

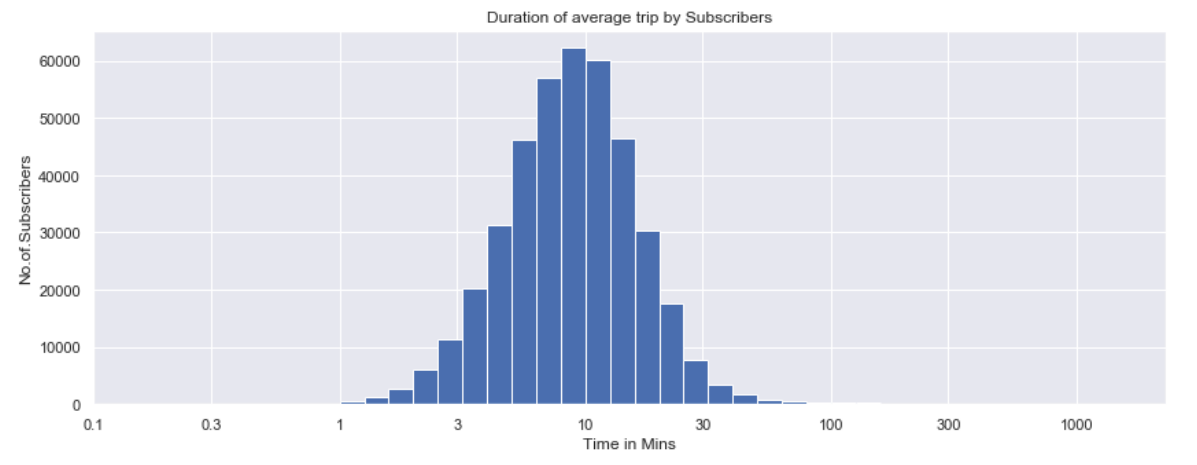
- The Average time of all the Users of the Company is around 3 – 300 mins.



- But we can see some variations in the users behavior.

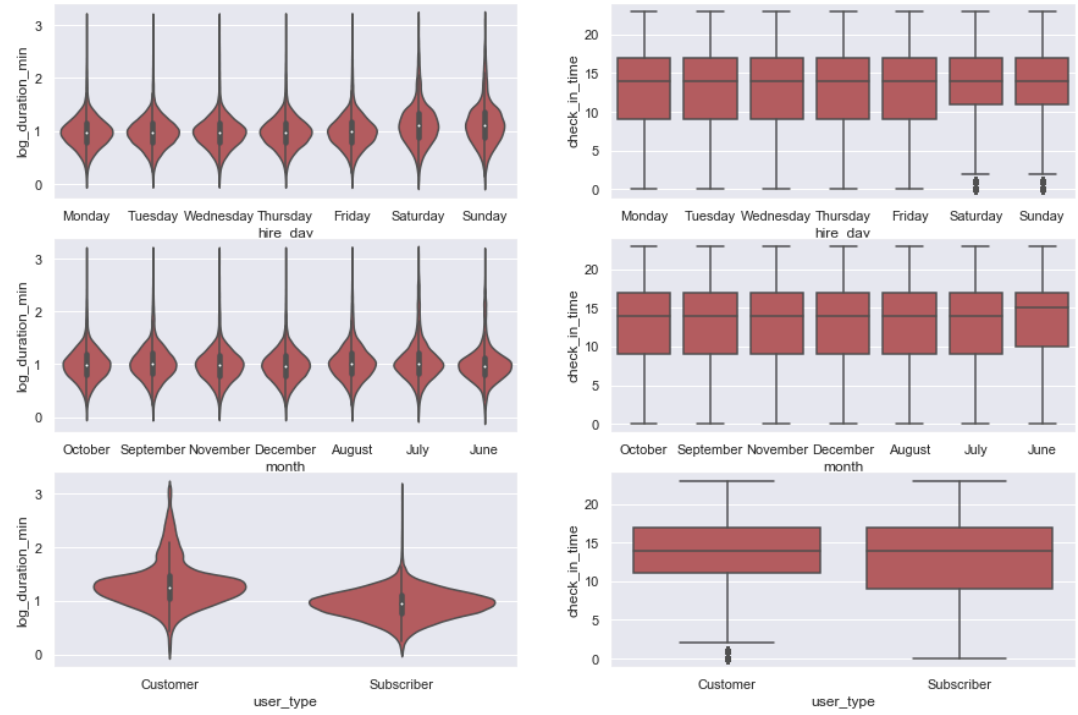


- As we can see the Above Graph the Average time of the trip varies to customers as they tend to use for longer time as the graph indicates the time longs from 5 – 300 mins.



- But this changes in case of Subscribers as we can see the time varies from 1 – 80mins

When are the most trips taken in terms of time, Day, Month.(Bivariate Visualizations)



- From the above charts the transformation of the average depicts clearly that subscribers use the service less than the customers. (i.e) Their average time is less when compared to the Customers.
- In my point of view the violin plot say that the average time for all month are almost equal.
- Thirdly the average time for the weekdays is same when compared to the weekends there is slight variation which is more than the weekdays.

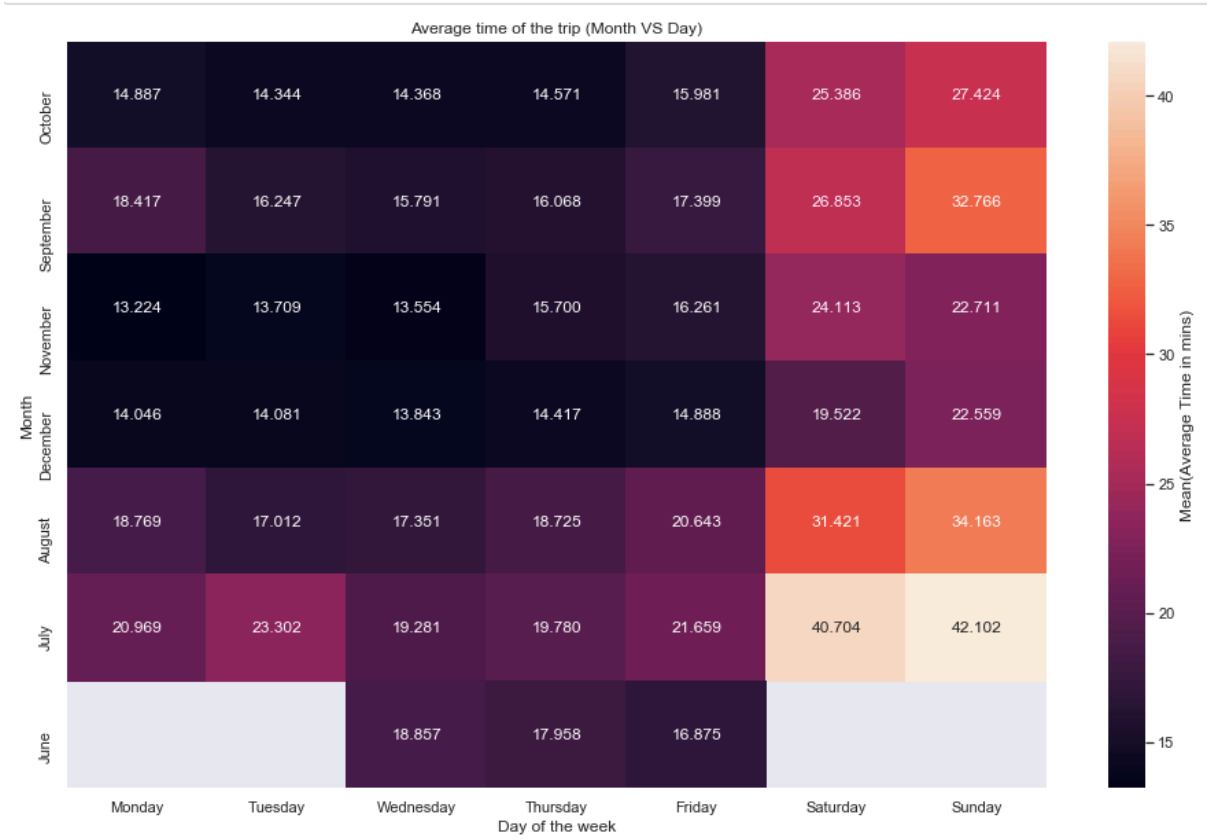
- Speaking about the with respect to the Check in time it is clear that Customers tend to use the service between 11:00 hrs - 17:00 hrs while the subscribers tend to use the service between 8:00 hrs - 17:00 hrs.

- Preferred time to use the service for all the month is almost equal while in June we can see a slight variation.

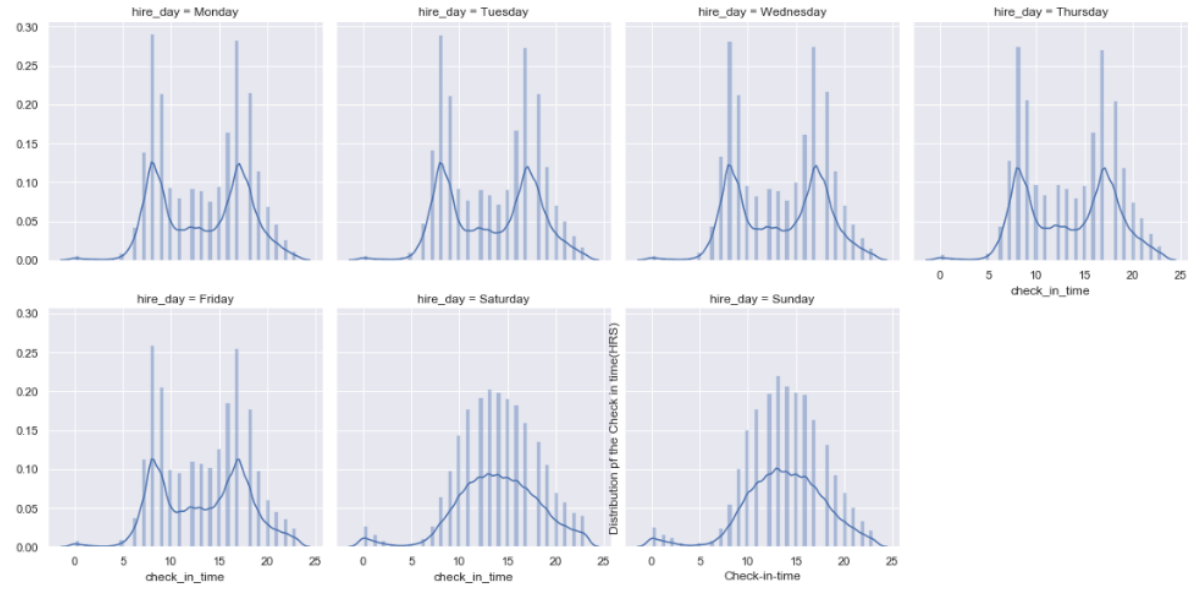
- Preferred time to use the service for the weekdays is almost equal but we can see a variations in Sunday and Saturday.

Multi variate Visualizations.

What is the Average time per day for each month, and Which time do the Users prefer to use the services of the company each day.



- Even though the users tend to use the services mostly on Tuesday and Wednesday, from the above heat map we can say that users on Sunday and Saturday they tend to use the services for longer time as the Average Time of the trip is more.
- Another interesting fact is that, in week days users tend to use the services in the mornings and when compared to the weekends they are not as we can see that they prefer to use the services of the company during the mid-day (i.e.) from 14:00 to 18:00 hrs. which is quite interesting.



Conclusion.

1. we can say that the time utilised by the users is fairly right skewed. We have right skew issue which should be solved by using logarithmic scale. Bike durations range from less than 1 minute to 1400+ minutes with median at around 9 min and mean at around 12 min. We have to do some data transformation to make data visualization and data interpretation easier.

2. After transformation we can say definitely that both the customers and the subscribers use the service for average time is between 3 - 30 mins from the above histogram.

3. From the above the analysis we can come to know that the company has more than 78% Subscribers and around 22% customers.

4. From the analysis we have a surprising details that although the company has only 22% customers they are the users who tend to use the service of the company for longer time when compared to the subscribers. There are less subscribers who tend to use the service for longer time is less. To be precise we can say that the Average time of the subscribers Fall in the range between (1 -50 mins) while the customers tend to use the service for longer time (i.e) the average time fall between (1- 350 mins).

5. We can say now confidently that the most preferred time for the user to use the service provided by the company is Morning (i.e) 05.00 hrs to 8.00 hrs. Followed by Mid-day (i.e) from 15.00 hrs to 18.00hrs.The least preferred is the Evenings people never use the service the most during 18.00 hrs to 23.00 hrs.

6. As we already know that the months October and September are the most preferred months for the users but every month they tend to use the services for longer times on Saturdays and Sundays.

7. From the Distplot we can see that the Saturday and Sunday the users tend to use the services between 10:00 hrs - 20:00 hrs. And in weekdays the users tend to use the service more between 6:00 hrs - 10:00 hrs & 15:00 hrs - 19:00 hrs.