

Bay Wheel's trip data exploration - Report

The accompanying exploration_template. ipynb file contains the detailed programmatic analysis of the dataset. The dataset was obtained from the Ford Go Bike System Data. The data pertaining to the year of 2017 was analyzed. The data are of trips made during 2017 via the Bay Wheels public bicycle sharing system in the San Francisco Bay Area.

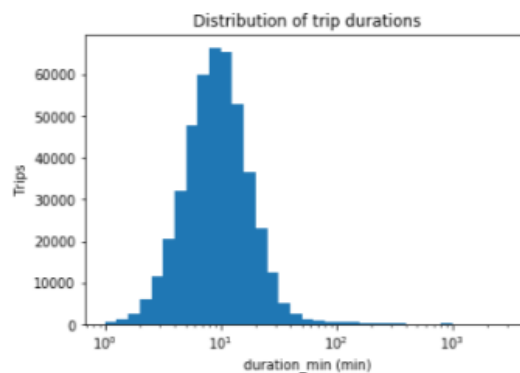
Data Wrangling and Cleaning

After the data was received, it was subjected to data wrangling and cleaning to remove null records and/or mis-recorded data. There were no duplicates in the dataset but after removal of the nulls, some feature engineering was performed to come up with more insightful features that were used in the subsequent analysis. The feature engineering included:

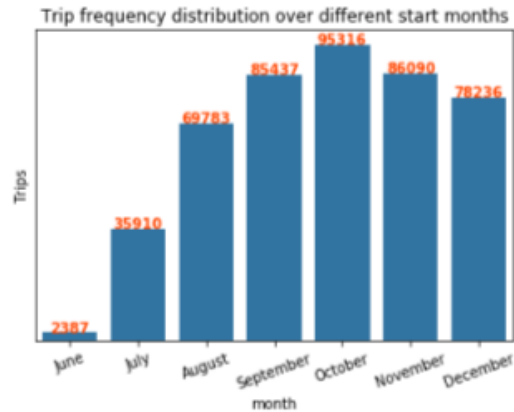
- Conversion of features to their appropriate datatypes like datetime.
- Feature engineering to extract new insightful features like extracting months, days and hours from datetime objects.
- Conversion of birth years to member ages and so on.

Summary of Findings

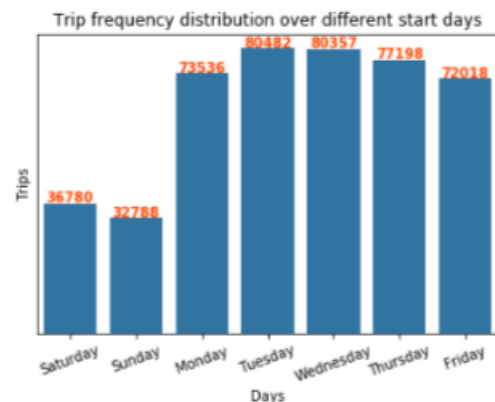
The analysis was performed on the dataset envisioning myself as business analyst for the company that manages this public bike sharing service and the aim was such that the company ought to increase its market share and revenue. Based on the exploratory analysis performed on the dataset, the following summarized findings and conclusions were reached at:



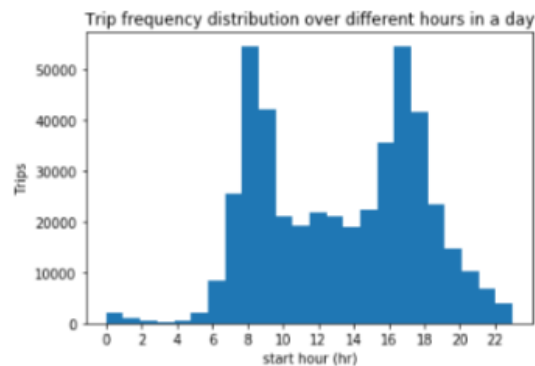
* The most frequent trip duration was found to have a median around 10 minutes and hence it would be advisable for the company to bolster advertising for those people who make trips around the most frequent duration of 10 minutes trip duration.



* October and surrounding months show the highest frequency of bike share trips with the lowest occurrences during the months of June (or July). Probably during the summer months, the number of people using bike sharing platforms dwindles. It will thus be advisable for the company to target customers on or near the month of October for the second half of the year.



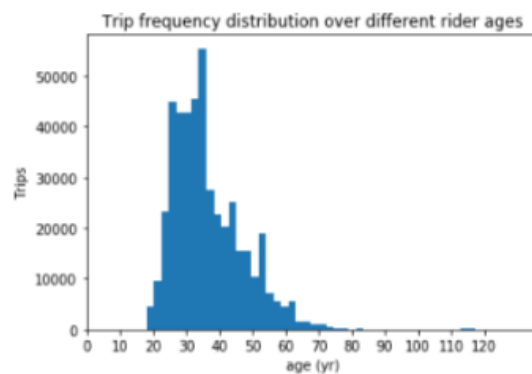
* The number of trips made on weekdays far outnumbers those made on weekends. Weekday trips are almost more than double of those made on weekends on average. The company should cater its service towards weekday trips if resources are constrained.



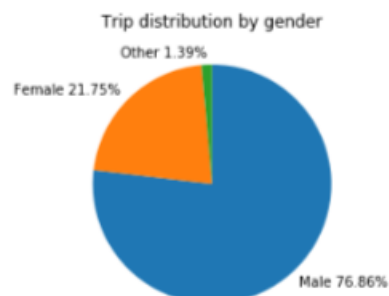
* Most of the commutes on bikes would happen when people are trying to go to work in the morning (peak at 8 hrs.), and then subsequently returning home during the evening (peak at 17 hrs.). The company should target these hours in a day to tailor its services for better customer support.



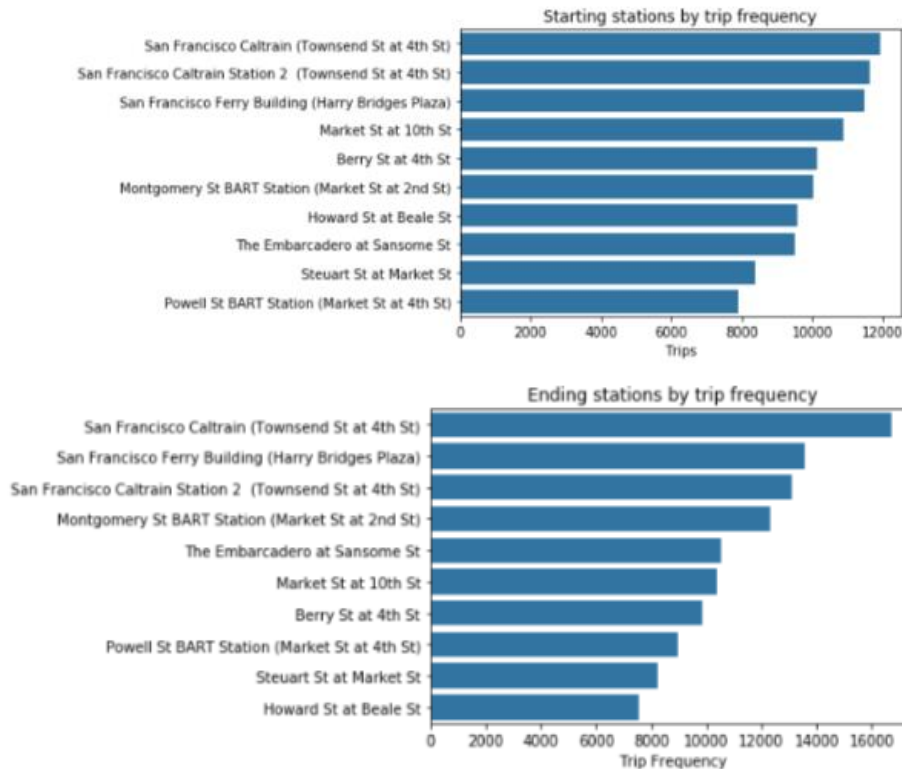
* Almost 90% of the bike riders are subscribers while a meagre 10% are customers who do one off purchases. It would thus be advisable for the company to concentrate and provide better service to its loyal subscribers who create the bulk of its business.



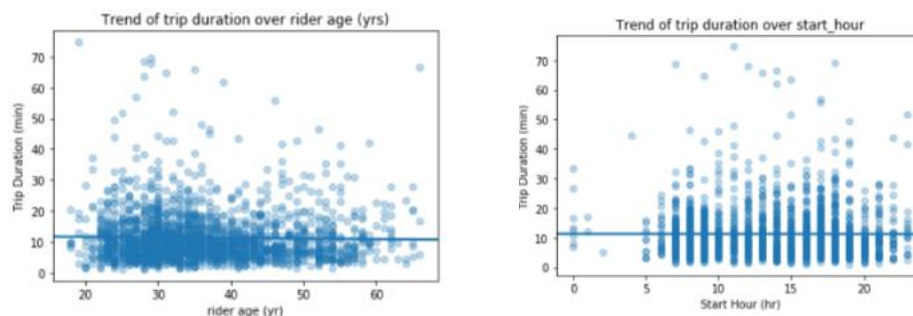
* The distribution of trip frequency over different rider ages is unimodal, with a peak at around the age of 35 years and is right skewed. It would thus be advisable for the company to target its services towards the 20-40 years old age and profession group which has the potential of bringing it the largest amount of revenue.



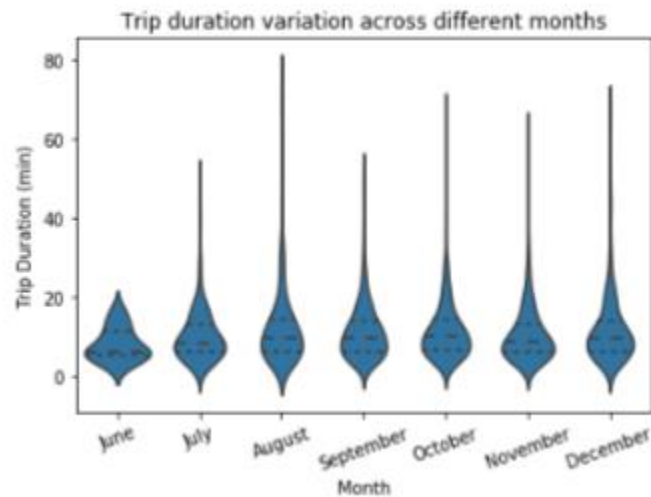
* Males use the bike sharing service almost overwhelmingly over females and other genders identities. It will be perhaps advisable for the company to device services and plans that would encourage females and other gender identities to avail the bike sharing system even more, possibly having potentials of opening avenues of whole new market sectors.



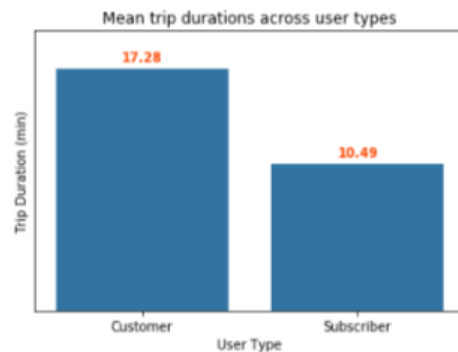
* The most popular starting and ending destinations is San Francisco Caltrain Station. The company can target these popular starting and ending stations by making their services more convenient and improving customer experience such that the high volume of trips along these routes are maintained.



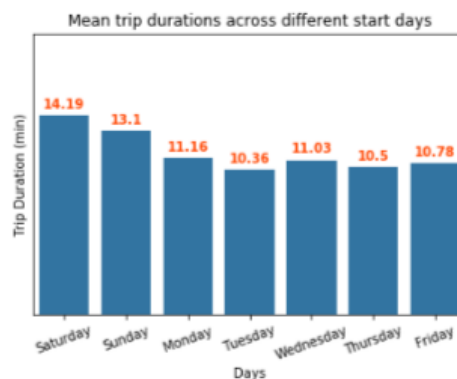
* There is poor/weak correlation between the rider age and trip duration, likewise between trip duration and start hour. No decision should be made based on correlation of these features.



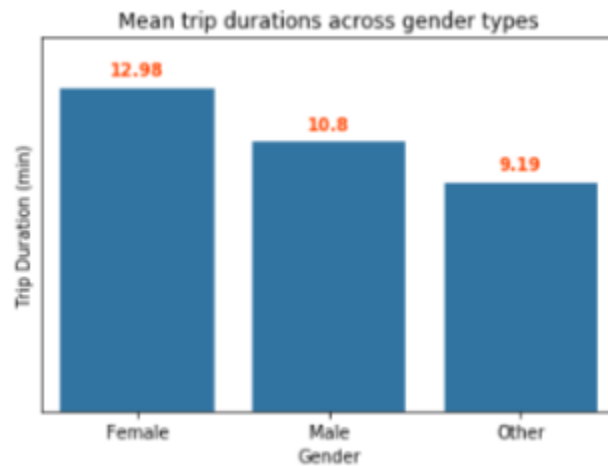
* Across all months, the median trip duration seems to be around 10 minutes like we have seen earlier. However, trip duration variability increases largely after July and continues up to December. In October, trip durations can go as high as up to approximately 80 minutes. This gives valuable information to the company that the months of October and the months surrounding it are when it can anticipate high business volumes.



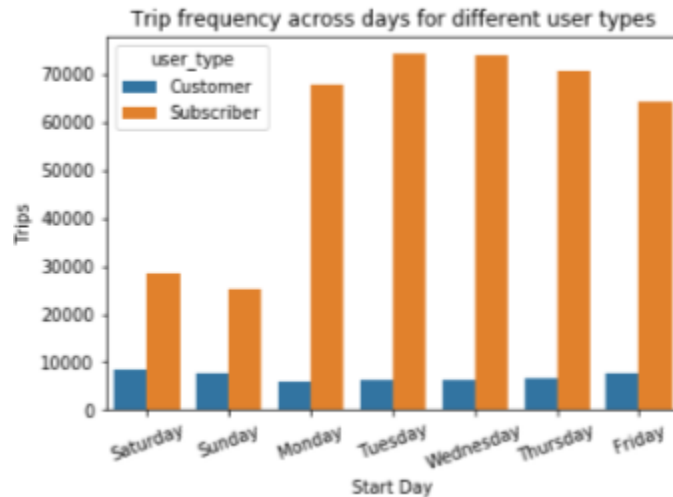
* Customers on average spend more time in trips as compared to subscribers although around 90% of the users are actually subscribers. Decisions should not be made thus, just on the basis of trip duration.



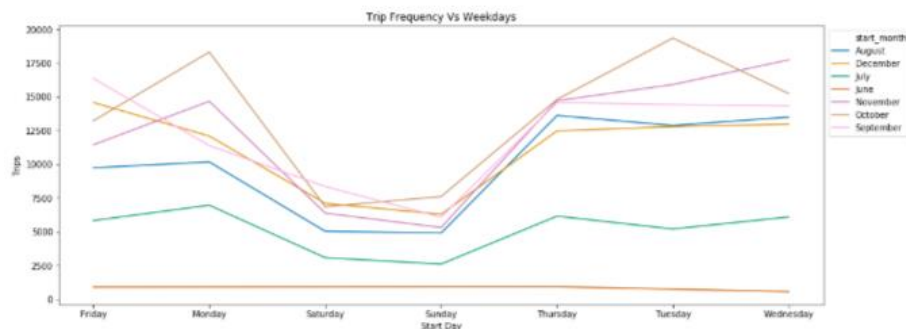
* Although the trip numbers on weekdays outnumber those on weekends like we found in univariate exploration, but the mean trip durations on weekends are larger than those on weekdays. So, people make fewer but much longer trips on weekends. The company thus can cater services to users on weekends that allow them to go for longer rides without having to pay too much. This can potentially increase the number of customers.



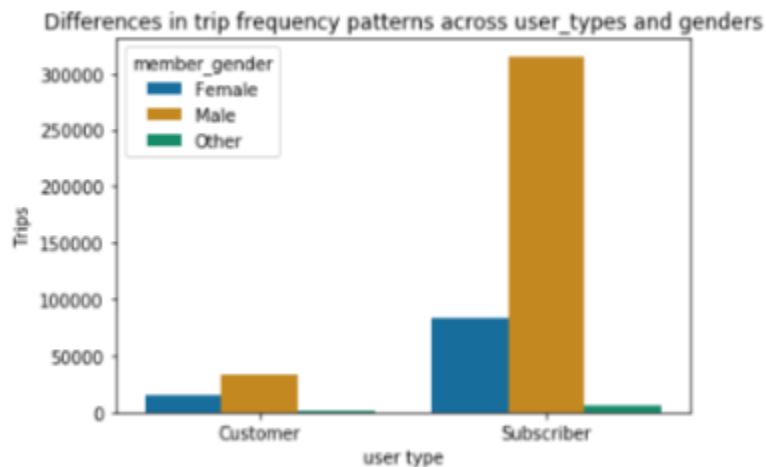
* Although the trip numbers by males outnumber those of other genders like we found in univariate exploration, but the mean trip durations for other genders are larger when compared to that of males. So, males make more trips but of usually shorter durations.



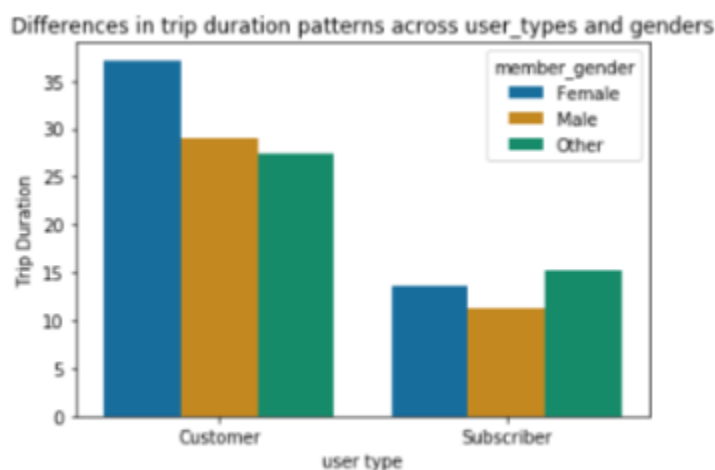
* The number of customers remains almost constant throughout all days, however, that for subscribers dwindles on weekends. It goes to show that weekend customers are usually casual or one-off customers. The company could devise plans and services targeted at these people to convert them to full time subscribers.



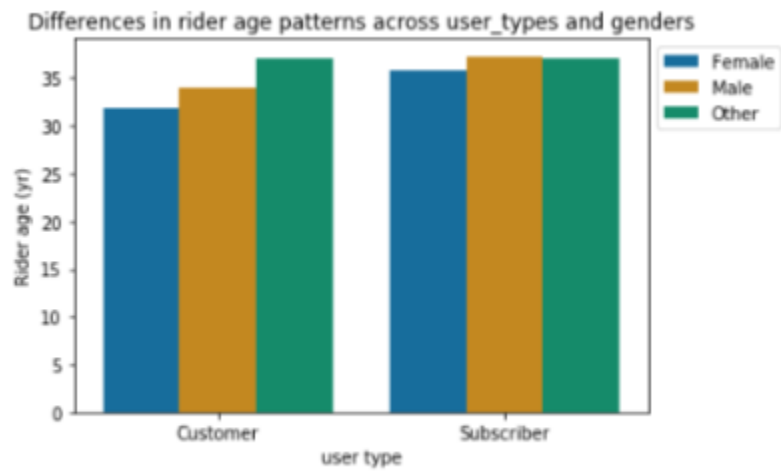
* Almost all of the months show similar patterns of trip frequencies over the weekdays, with significant dips in trip numbers on weekend days. October shows the highest overall trips, with June the lowest. Marketing strategy should be targeted in increasing revenue in the lull summer months like June or July and in the weekends.



* It looks like male customers and male subscribers make most of the trips. In case of customers, males are almost double in number of females while for subscribers, they are more than 3 times the number of female subscribers. This shows that there is a lot of room of improvement for the company in attracting more loyal female subscribers. There are no customers from other genders and there are very little from among the subscribers. This is another possible area of improvement, where the company can attempt to appeal to of other genders to use its bike sharing service and prove to be an all gender friendly service system.



* In case of mean trip duration, across all user types, female travel for longer mean periods of time as compared to males and other genders. An interesting finding is that, males travel for the shortest trips from among subscribers. This possibly is an area of improvement where the company can encourage its male subscribers to go for longer journeys on weekends at reduced prices.



* The mean age group across all user types and genders fall above the 30 yrs. line and is below the 40 yrs. line. It is this group that the company should aim at.