Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in
Data Analysis Exam
Google

Ford GoBike

Introduction:

The dataset is about Ford GoBike from 2017 June to 2018 July, including attributes like trip duration, 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. This dataset required merging data from multiple sources, cleaning, creating calculated fields and more. Analysis provide few key take-aways as described in the following sections.

Trend Patterns:

This is a high-level Trend Patterns for the Management.

The first 3 patterns show the overall growth patterns on daily bases & monthly bases in terms of Total Number of Rides & Total Duration. In the first plot, we can observe growth in trend, with the weekends hitting the business. In 2nd & 3rd plot, we observe the continuous growth, with seasonality hitting the growth pattern during winter season.

Finally, the HeatMap shows the popular or busy time of the day & even day of the week. The peak hours being 7:00 am to 10:00 am & 4:00 pm to 7:00 pm, they are only during the WeekDays. This gives us the hint that, during the peak hours many customers use Ford GoBikes to commute to their work. But further analysis shows that the Average Trip Duration is concentrated towards the mean of 16 minutes & standard deviation being 48 minutes. With this we can conclude that most of the commuters use Ford GoBike to commute to the nearest locations such as from train station to their office location.

Recommendations -

1- For Customers, we can charge more cost during the peak time for a single ride than for the regular times. Given a relation between Customer behaviour upon adding cost price, we can use optimization algorithms & maximize the revenue for the company & retaining customers at the same time. (This applies to customers only, not for subscribers)

2- We need to work on providing some promotional benefits during winter season, to avoid the mentioned seasonality. We can use similar optimization algorithms given the customer behavior upon provided promotional benefits.

For the mentioned recommendations, we need to observe & formulate customer behhavior, & then optimize the business as mentioned above.

Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in
Data Analysis Exam

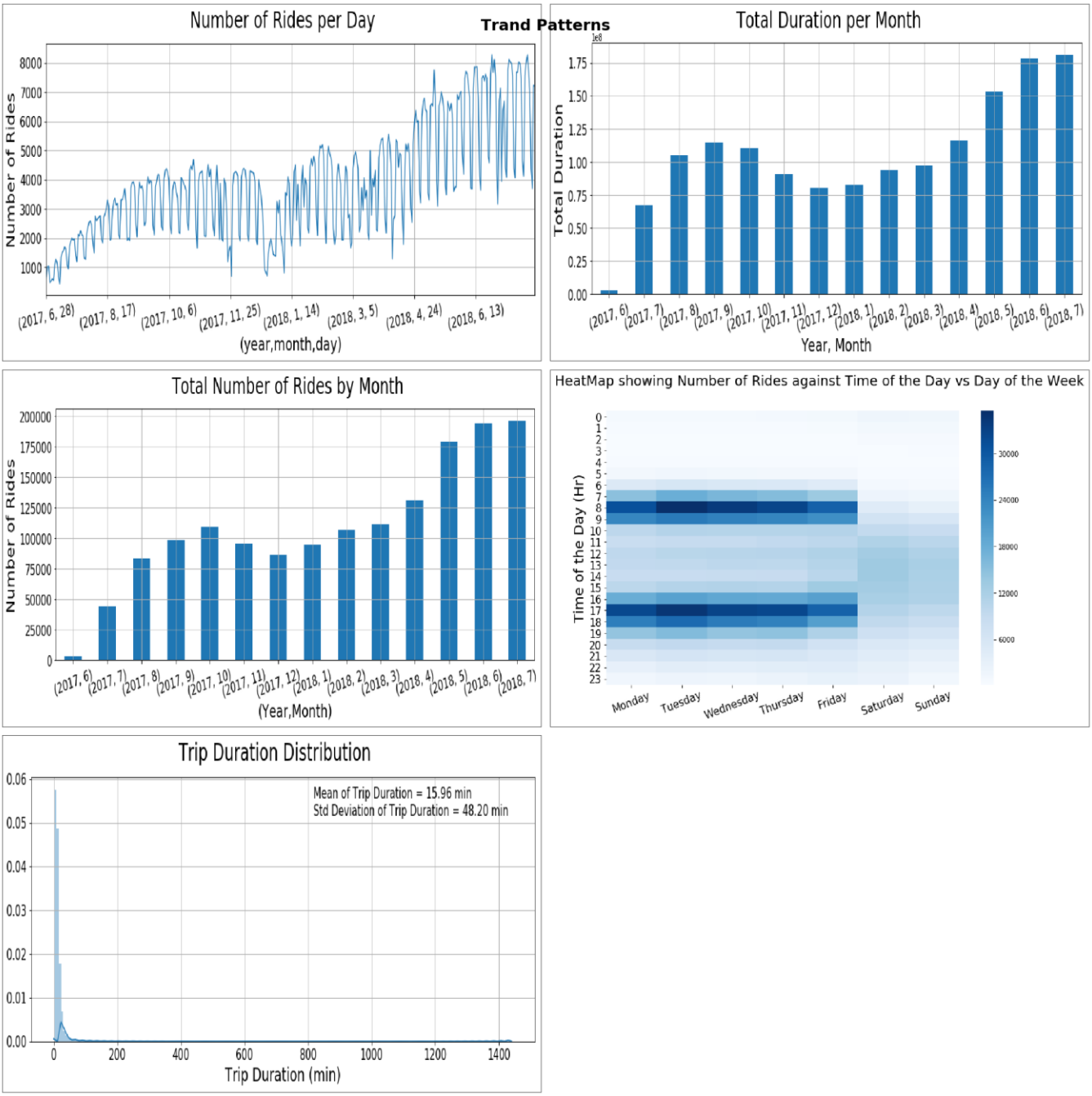

Fig. 1 – Trend Patterns

Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in
Data Analysis Exam

Understand your Customers & Target your Market:

Understanding the Users takes Targeted Marketing way ahead. This dashboard tells about our Users.

The first 2 plots informs us about the gender of our customers. It is weird that, California having more Female population than Male the Genderwise Usage Variation shows that 68% of the rides have been used my Males to that of 22% by Females. (https://www.states101.com/gender-ratios/california) Although the overall growth rate for females is higher; there is a huge scope for gaining more female users.

The next 2 plots explain about Usage variation by User Type. Subscribers use our service more than regular customers. California definitely has a good tourist market, hence we need to work on attracting tourists to use Ford GoBike services.

If we look at the Usage variation by Member's Age, we see that Users are mostly from range of Age 25 to 35. Although there is potential market with Age between 35 & 100; Age group between 18 to 25 are our potential future market.

The last graph shows that male subscribers as well as customers use Ford GoBike service more compared to that of Female and Other genders.

Recommendations-

1- As mentioned before, a huge portion of Usage being done by commutors like from Rail Station to their office; it is hard to gain Female customers due to their attire for their work making it uncomfortable for the bike rides. If we can overcome this obstacle in some way, that would raise our users to a new level.

2- Given further information about Subscribers & Customers including User_ID, their permanent address & more we can draw further conclusions. Depening on who gives more profit for the company, we can put more efforts in marketing or promotional benefits accordingly.

3- Concentrating more on out potential future market (Age Group 18-25, & possibly 35-40) can help growing our market reach.

Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in
Data Analysis Exam

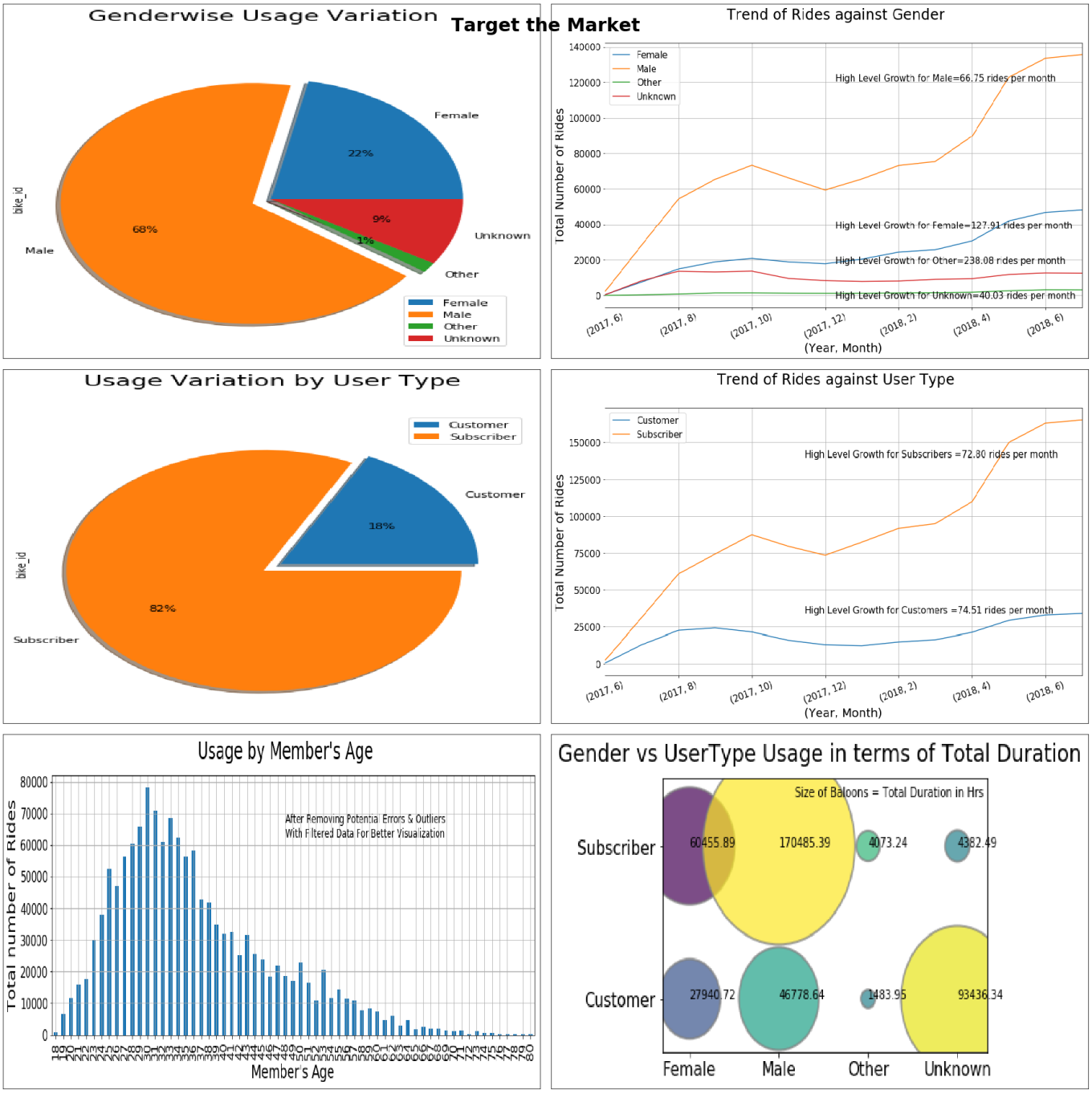

Fig. 2 – Target the Market

Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in
Data Analysis Exam

Station Utilization:

This dashboard tells about our Station Utilization. These graphs show some of the top busy stations & least busy stations.

If we further study the station location we observe that many of the busy stations are either near to the Bay water or Train Station. And it is the other way around for the non-busy stations.

The last plot here shows Popular Combinations of Stations. Most of these are beside the Bay water. So users hang out beside the Bay Water, somewhere between Golden Gate Bridge and San-Fransisco-Okland Bay Bridge. Train station are common also, but do not come under popular combinations as the users spread out to multiple work locations as they take bike from a Train Station.

Recommendations-

1- When Ford GoBike is planning on expanding it's business to other states, this data is highly important. Understand what types of locations are popular & what are the potential customers. Based on this Ford GoBike can decide on station locations & the target market.

2- Further analysis could be Network Balancing to serve Users better and maximize the profit for the company. As the users pick bikes from one location & drop it at another, it requires a lot of efforts in Handeling the Network. If a station has a capacity of 10 bikes, and if a user wants to drop a bike, but what if there is not space remaining on the station??? If a user wants a bike from station 'A', and there are no bikes available at the station??? So given some more time, we can create an algorithm that helps us handle this network at its best. But to start of with, we can use these graphs to optimize the netowrk to certain extent. We know the peak timings are 7:00am to 10:00 am & 4:00 pm to 7:00pm. We also know the top busy stations & least busy stations. So during the peak time, we can keep a lot of bikes available near these top busy stations than the least busy stations.

3- We can also offer some promotional benefits on least busy stations to attract some market there. And upon given some financial information, we can further analyze if these non-busy stations have the potential to make profit for the company and hence decide on keeping or removing the stations.

Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in

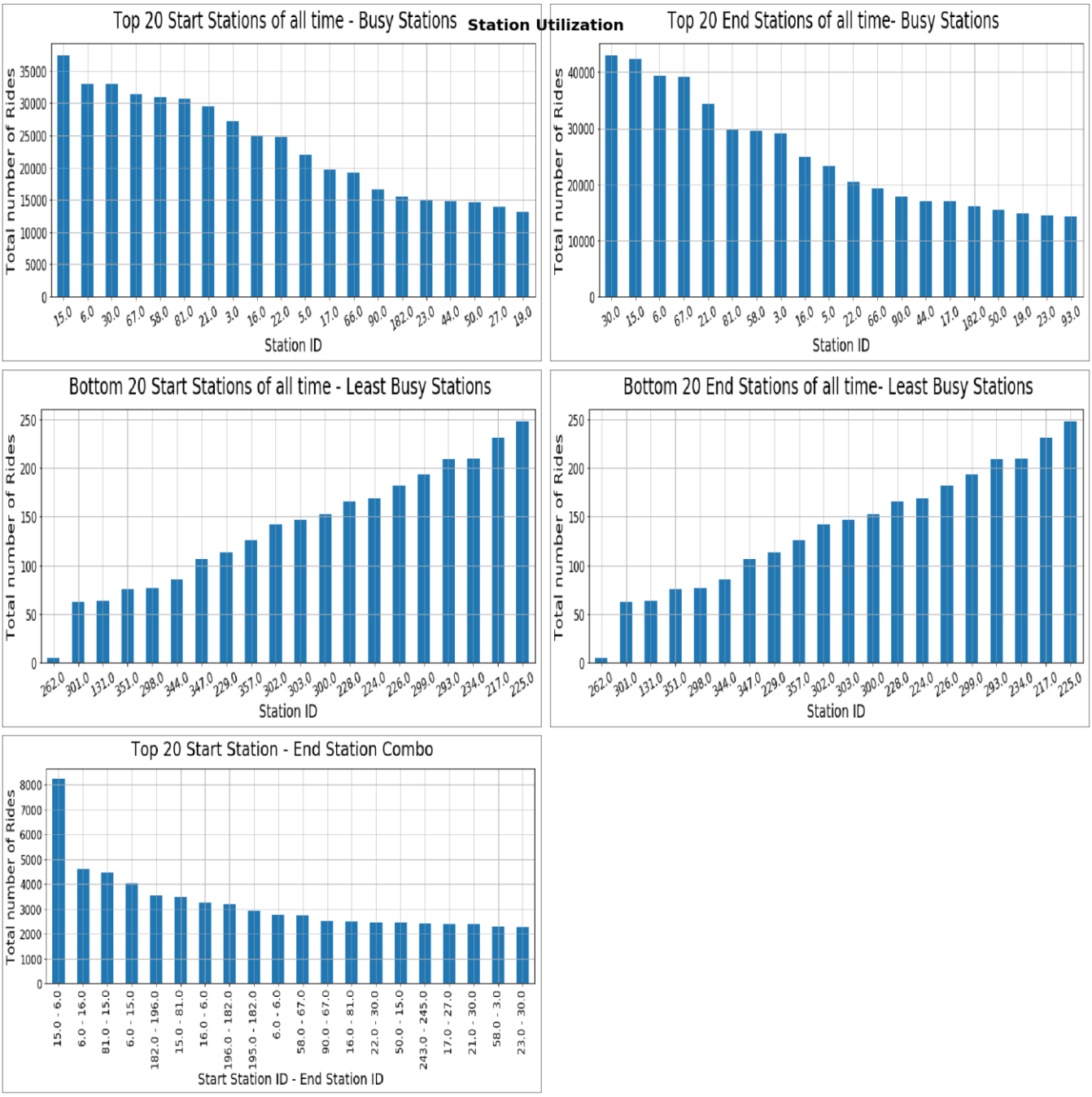

Fig. 3 – Station Utilization

Kartik Chavan
chavankartik94@gmail.com
chavankartik94@yahoo.in

Further Ideas:

1- **Churn Rate/Potential Bike Repair Needed**- If a customer takes a bike from station 'A' and returns it back with a span of 0 to 3 minutes to the same station or the nearest station (displacement as calculated in the script), then it is a prime suspect. Is it because the bike requires some maintenance service? Or is it because the user did not like our service. We can build a Churn Model using Machine Learning Algorithm to predict the churn rate. Also, we can conduct further analysis to understand the reason for churning.

2- **Network Balancing** - Further analysis could be Network Balancing to serve Users better and maximize the profit for the company. As the users pick bikes from one location & drop it at another, it requires a lot of efforts in Handling the Network. If a station has a capacity of 10 bikes, and if a user wants to drop a bike, but what if there is not space remaining on the station??? If a user wants a bike from station 'A', and there are no bikes available at the station??? So, given some more time, we can create an algorithm that helps us handle this network at its best. But to start off with, we can use these graphs to optimize the network to certain extent. We know the peak timings are 7:00am to 10:00 am & 4:00 pm to 7:00pm. We also know the top busy stations & least busy stations. So, during the peak time, we can keep a lot of bikes available near these top busy stations than the least busy stations.

3- **IOTs**- With users' permission, if can obtain the usage of bike including the distance travelled, or locations visited using an IOT system, we can further analyze the customer behavior & improve our business model.

4- **Per User Analysis** - Given further information about Subscribers & Customers including User_ID, their permanent address & more we can draw further conclusions. Depending on who gives more profit for the company, we can put more efforts in marketing or promotional benefits accordingly.

5- **Money** - Given further information about Subscribers & Customers including User_ID, monthly or annual subscription, date of subscription and more, we can conduct further analysis on revenue made till the date, or by month and more.