Ford Go Bikes in The Bay Area: An Analysis

Introduction

Ford Go Bike is a bike sharing service that is in current operation in the San Francisco/ Oakland bay area. What got me interested in analyzing a bike service near me was a trip to Sacramento, where I met an employee of the transportation department for the state, and was working with Jump Bikes to access their data. This gave me an itch to begin looking into my local bike sharing service to better understand the data. After opening this dataset and understanding what data was available, I form the following questions

- What are the most common routes?
    - What are distance of these routes?
    - What is the incoming and outcoming traffic delta for the routes
- What is the difference in behavior between Customers and Subscribers of the services?
    - How many subscribers vs. one-off customers?
    - How many occurances of late fees?
    - How much was made from each group on these late fees?
- What were the most individually ridden bikes?
    - What total time were they ridden?
    - How many total trips were ridden?
    - What bikes are mostly likely still in the system?

The following report will show my finding for these questions and what that means for the service.

The Data

The data is the 2017 Ford Go Bikes data set that starts in June 28 and ends December 31. The data is recorded by individual trips. The variables for the data are:

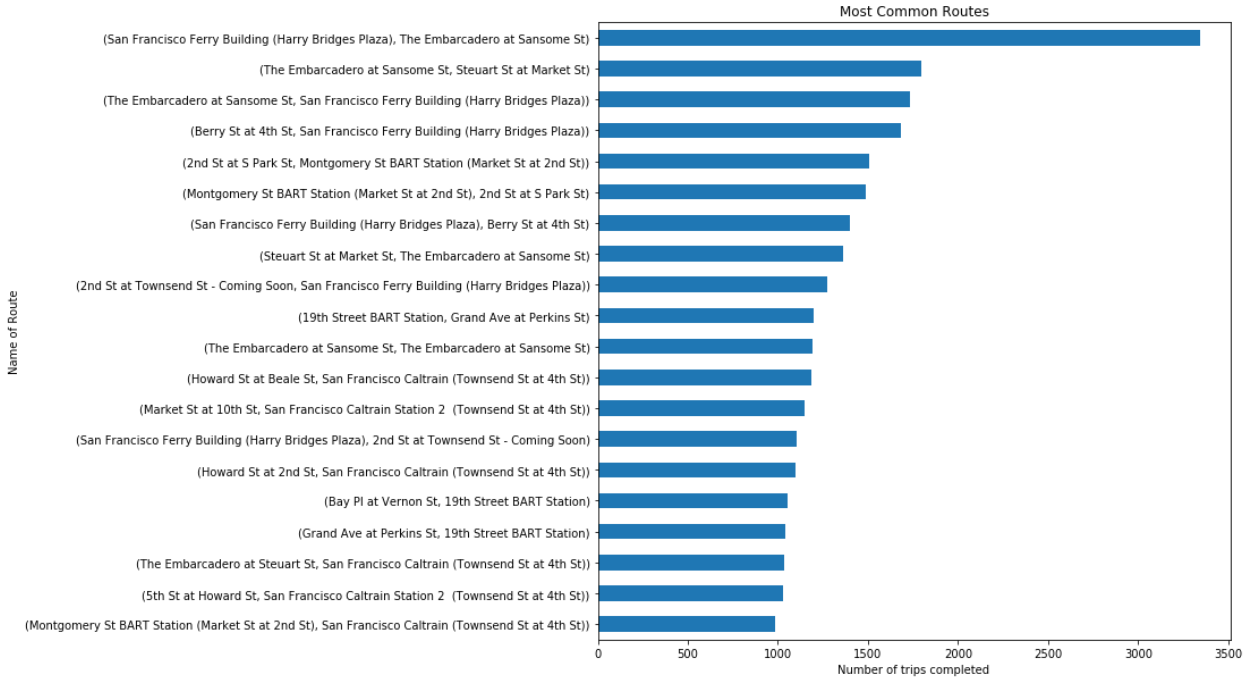
1. Duration_sec : the length of the trip in seconds
2. Start_time:
3. End_time:
4. Start_station_id:
5. Start_station_name:
6. Start_station_Longitude:
7. Start_station_Latitude:
8. End_station_id:
9. End_station_name:
10. End_station_Latitude:
11. End_station_longitude:
12. Bike_id:
13. User_type:
14. Member_birth_year

15. Member_gender

Constraints around the data include missing data for birth years and gender since only subscribers are recorded, and not having gps information.

1. What are the most common routes?

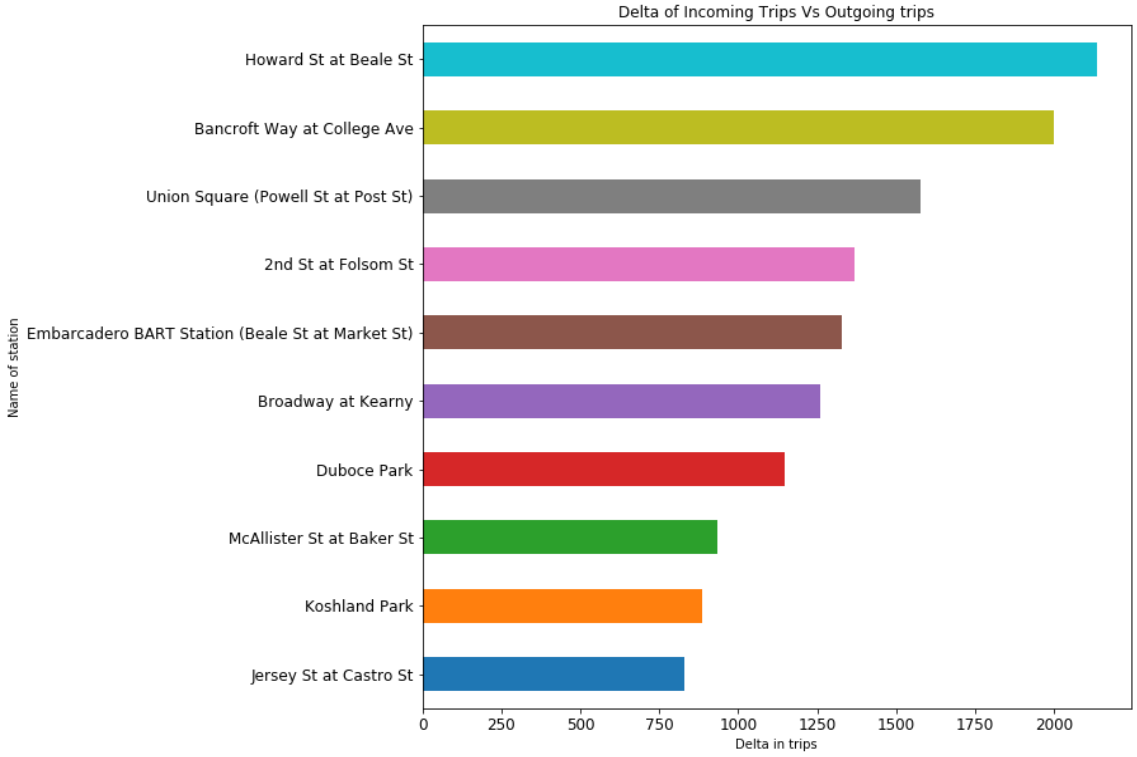
The routes of this bike service range all over the bay area. They are critical to understand in order to know where to put in more resources and what the service is used for. In the graph below you can see that the Ferry terminal to Embarcadero is San Francisco is the most popular route, speaking to both commuters and the tourism of San Francisco.

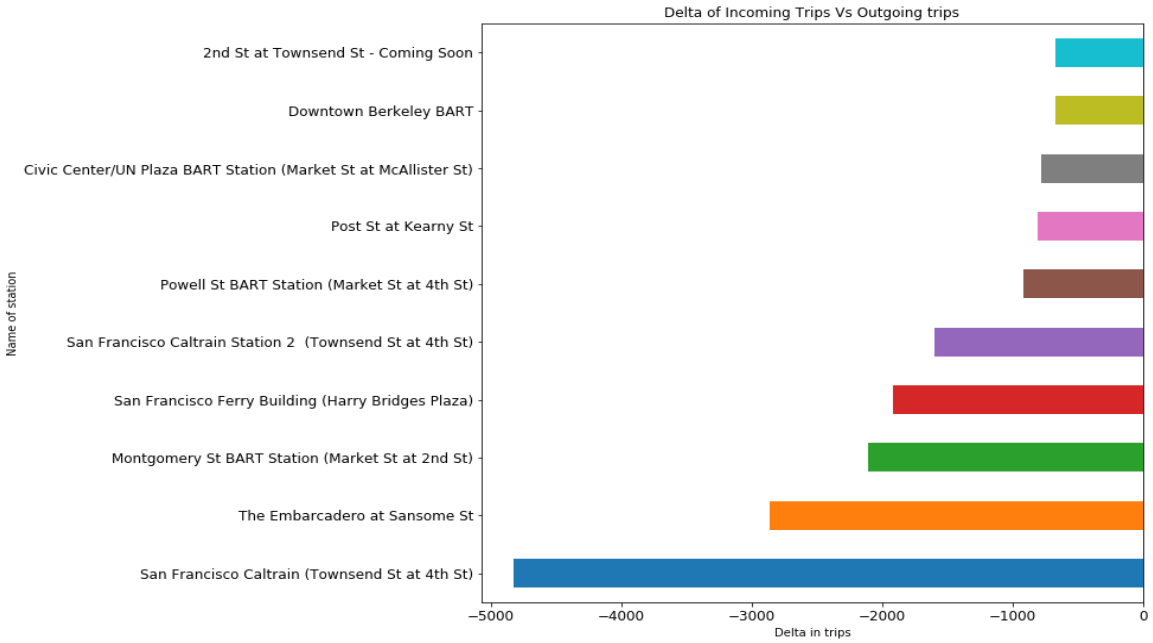


Then we can break down the distance of the most popular routes to understand if the majority of users are using the bikes for short distance or long distance trips. And what we see is that the most popular routes are all under 2.5Km (1.55 miles), showing that this service is becoming a alternative form of short distance transportation. The rest of the routes are mostly below 2.5 kilometers, except for a loop at Embarcadero and Sansome St., show that most users use them for short distance travel.

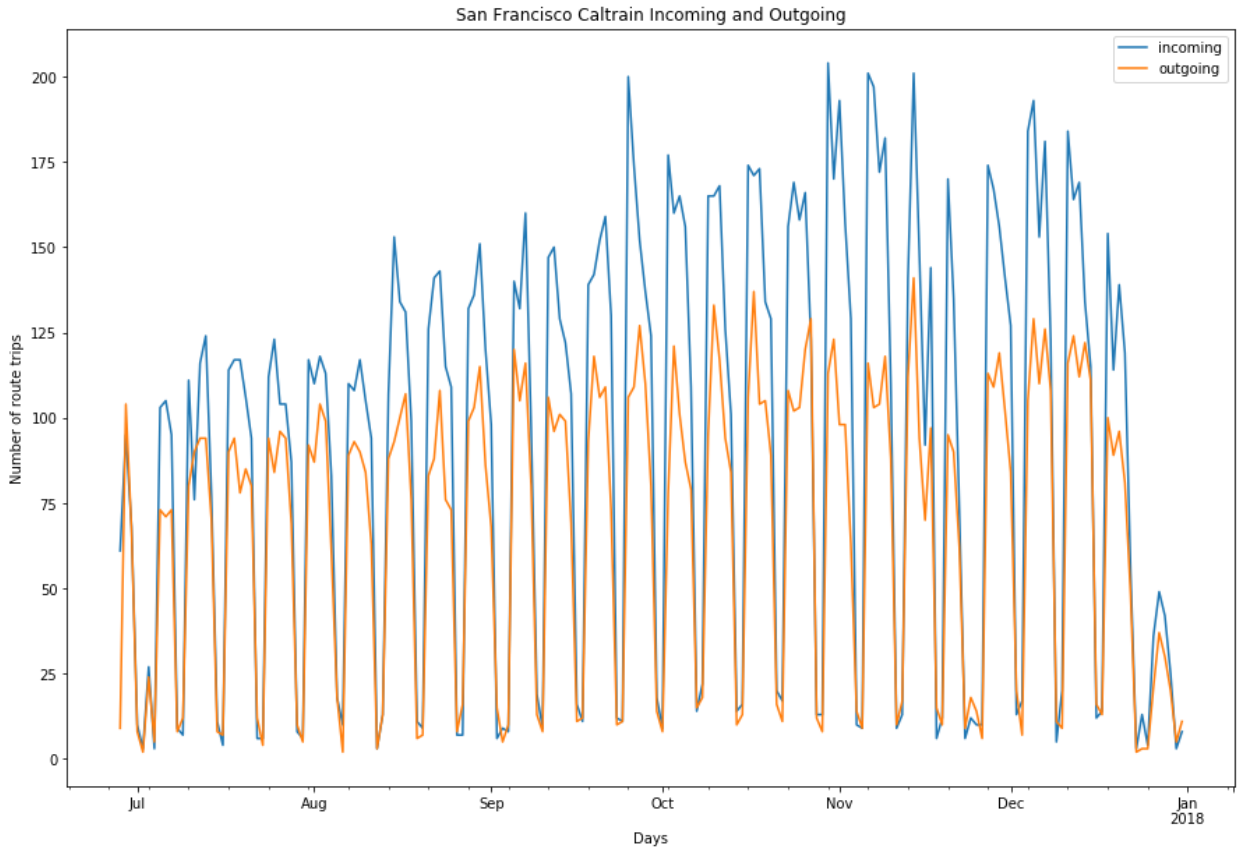
Even analyzing the names on these routes brings forth some information, how most routes go towards or from another form of transportation. Caltrain, Bart, and the Ferry building are noted.

Another aspect that needs to be understood is that of incoming versus outgoing traffic.

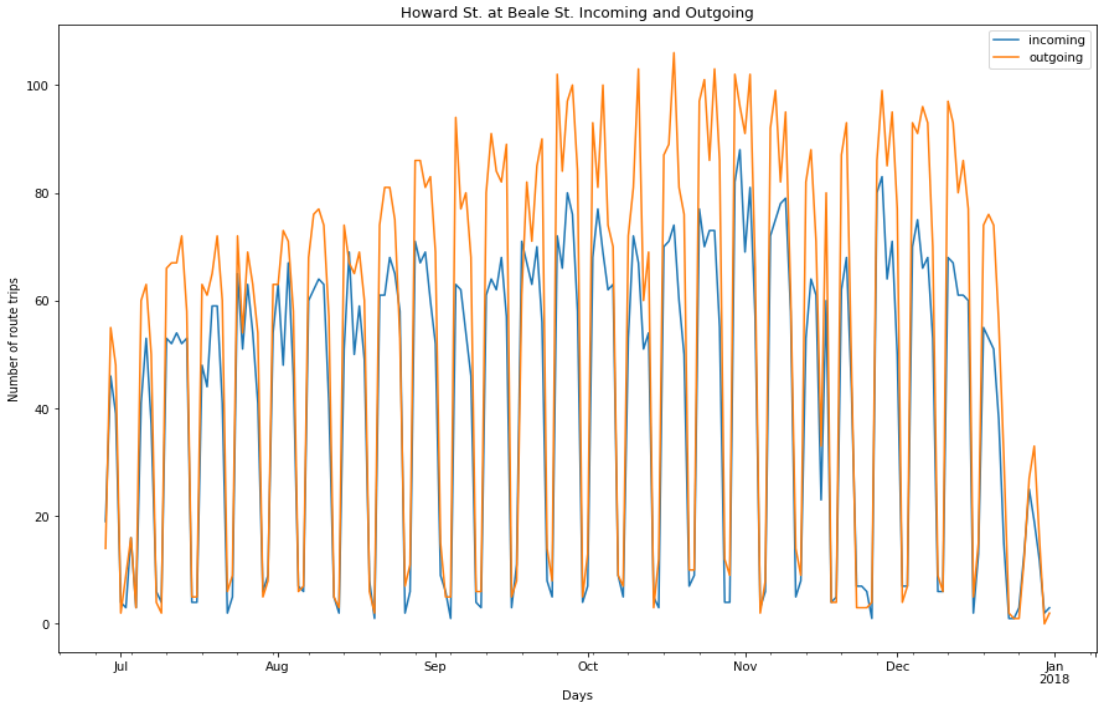




Graphed above are the top 10 stations for abundance in outgoing traffic and incoming traffic respectively. This creates a very interesting view on how stations are used. For instance the San Francisco CalTrain Station has the biggest incoming traffic, meaning more bikes should be at the station each day. Below is that very trend, where we see how trips are distributed throughout the day.

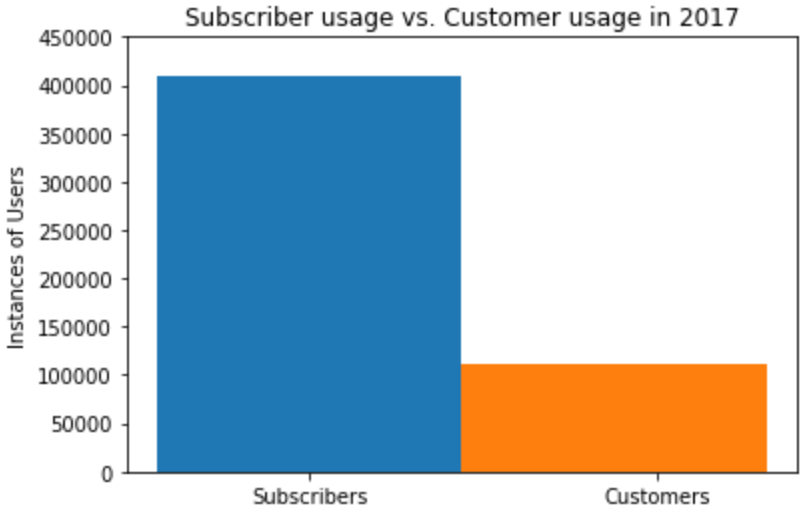


Then the same can be said for the Howard st. at Beale St. station.This brings the conversation to that of inventory management. This means some form of displacement and transferring must be occurring in order for these stations to have such a net loss or gain without facing shortages of some kind. Similar conclusions can be reached for every station that leans towards either outgoing trips or incoming trips. Meaning inventory management is happening at some point on a system wide scale that allows for this kind of displacement.

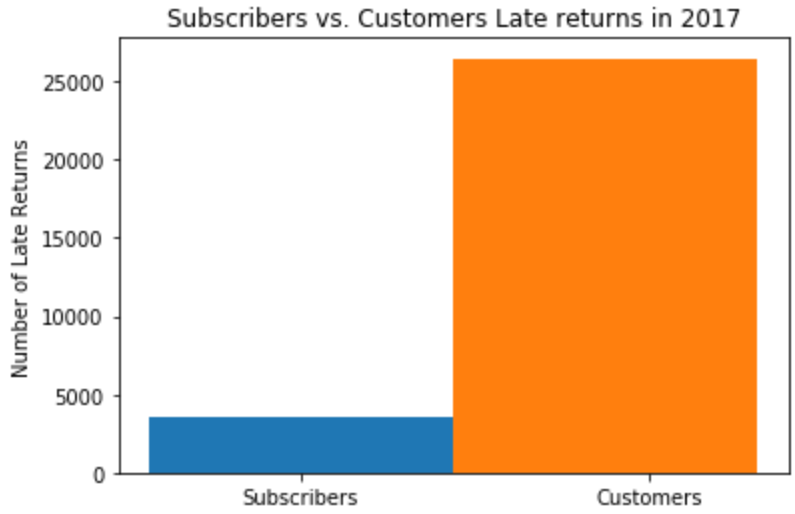


2. What is the difference in behavior between Customers and Subscribers of the service?

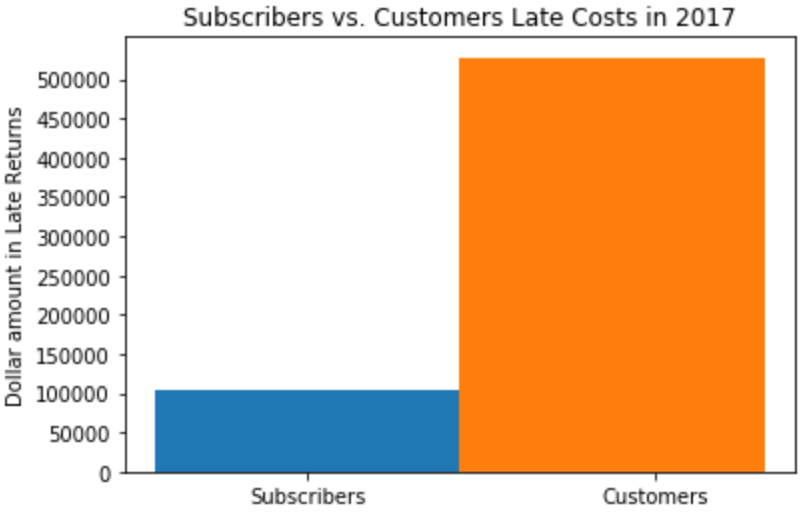
Ford Go Bikes offers two forms of purchases plans when going to rent your bike. First is the costumer, which is a one time use of 30 minutes payment. Then there is the subscriber which pay monthly and is allowed unlimited 45 minute trips. As seen above, the Subscribers are more than 78 percent of the unique rides that were recorded in 2018. This means that the company has in some form a steady stream of income per month. Which cannot be calculated due to not having access to the unique IDs of members, which for privacy issues, makes sense.
An interesting part of this service is that regardless of which tier you pay, late fees are still possible.



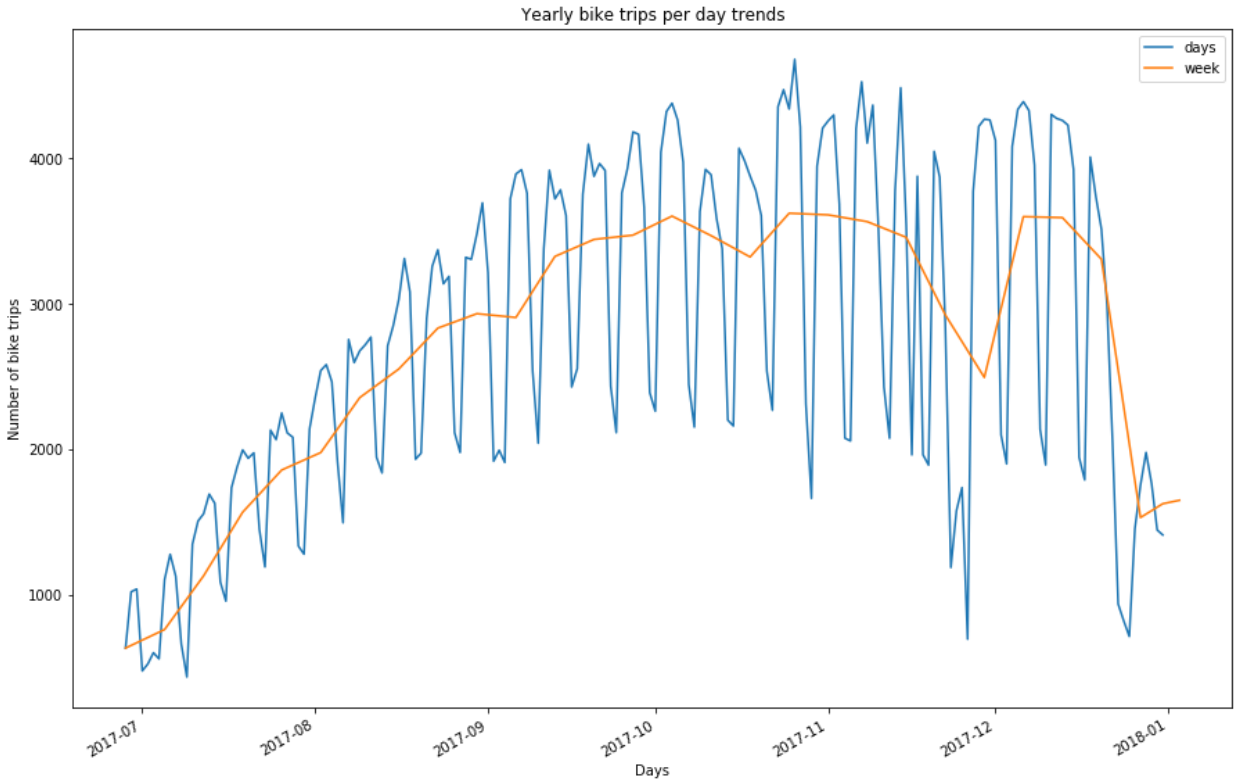
Thus we obtain the list of those who went over their limits, and no surprise, 5% of the all the costumer trips had late fees, while only .69% of subscribers were late. This allows us to determine a little bit of consumer habit and the monetary gain of having late fees, especially on services that can apply to anyone. Taking a closer look at the late fees, since it is expressed that for every 15 minutes that one goes over, 3 dollars will be added to the bill.



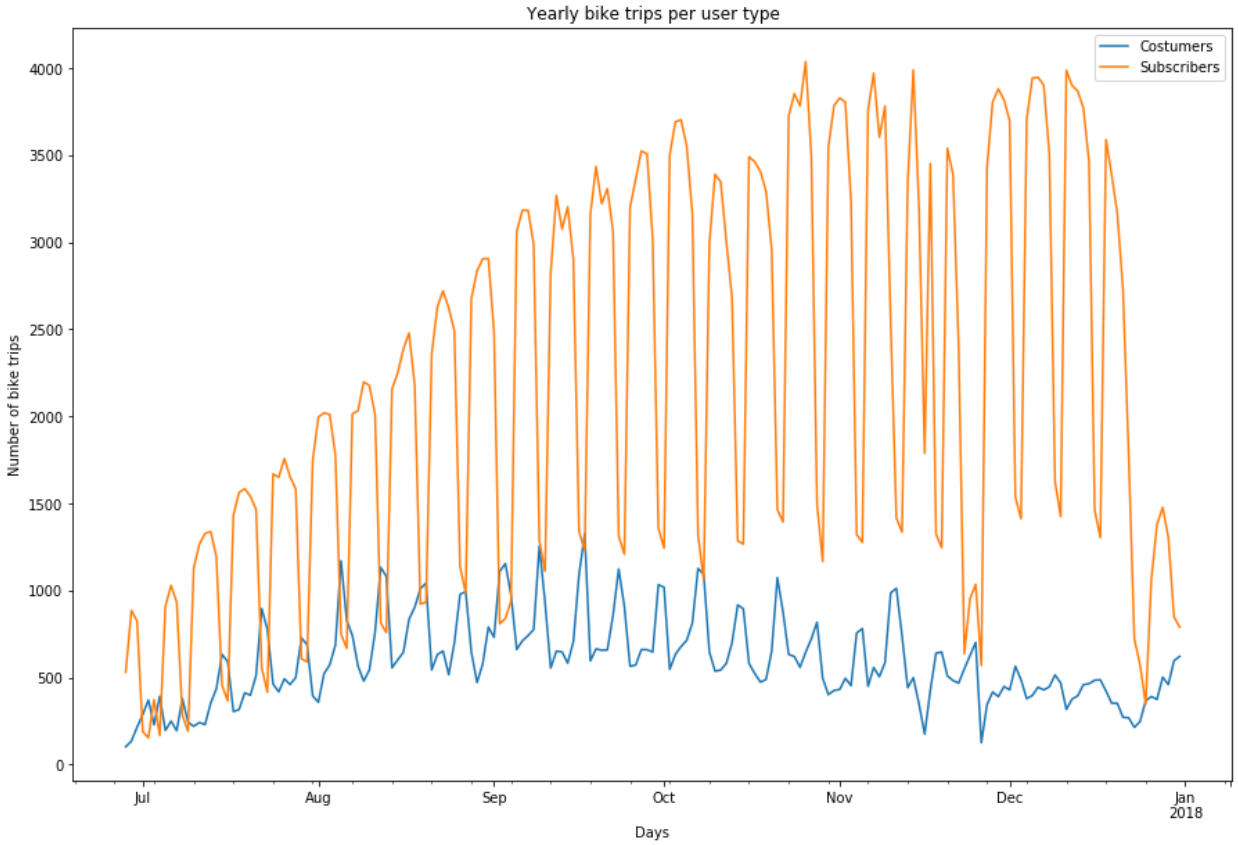
Over $630,000 dollars were made last year on late fees alone. With the costumer making 5 times the cost of the member. Again this give us insight into how the monetary mechanics of the service works, showing that even though customers are not frequent or stable as members, either through not knowing or not caring, the customer has made Ford Bike Go a pretty penny.



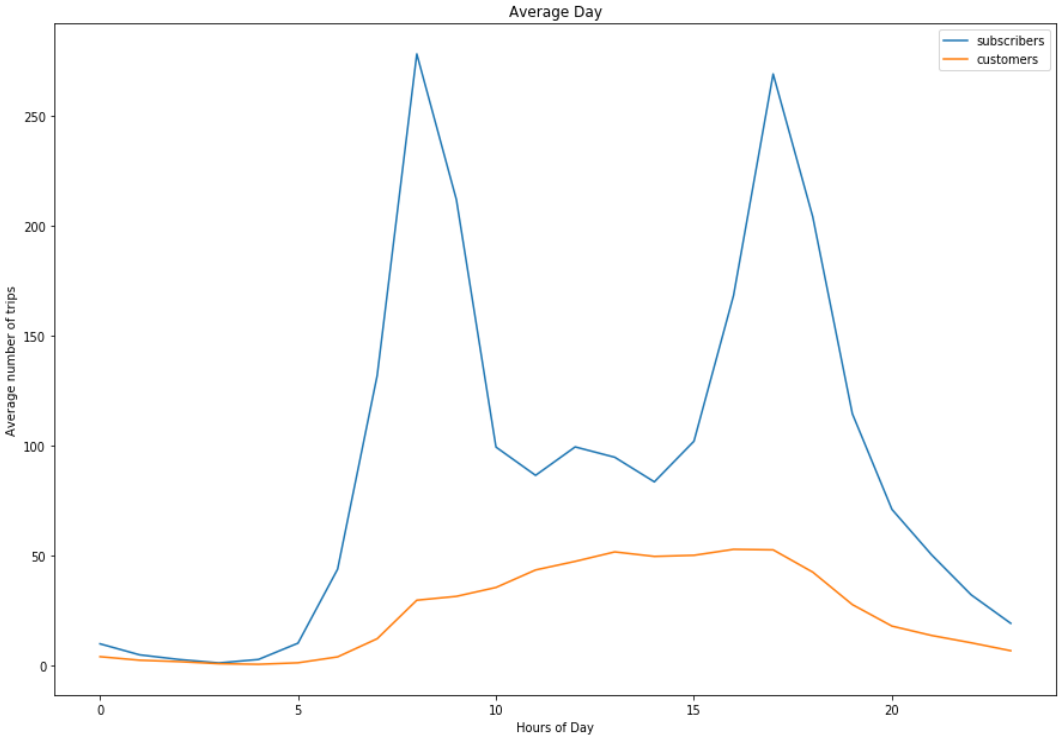
Yet late fees would account for only a tiny fraction of the overall revenue created for the service. While it is not possible to calculate said revenue, we can use usage of the service to see how customers and subscribers are using the service.



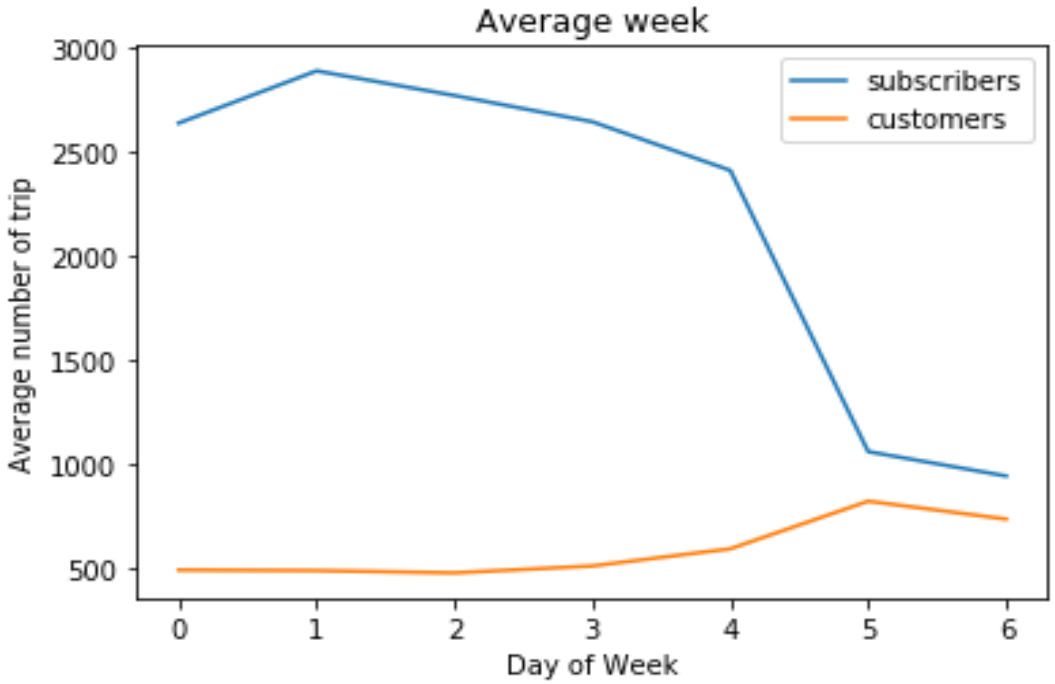
Here we see both the cyclical day trends and the averaged out weekly trends, showing that the service is indeed increasing in trips each week, with some drop offs happening around the winter holidays. Yet we can break this down even more, into the user types, to show how the subscribers and customers differ, noting how the first graphed showed the majority of rides coming from subscribers.



This affirms our first graph, showing how and when each type of user uses the services. Note the intersecting lines and how the valley of the subscriber tends to be the peak of the user. To truly see we need to manipulate the value to get an average day and average week to show the underlying patterns of the user types.



The average day again shows the cyclical nature of the subscriber, and the almost docile nature of the customer, noting the only subtle rise near midday. What this begins to point to is that of subscribers being workers in the city, while customers leaning to being the tourists that are free during the middle of the day.

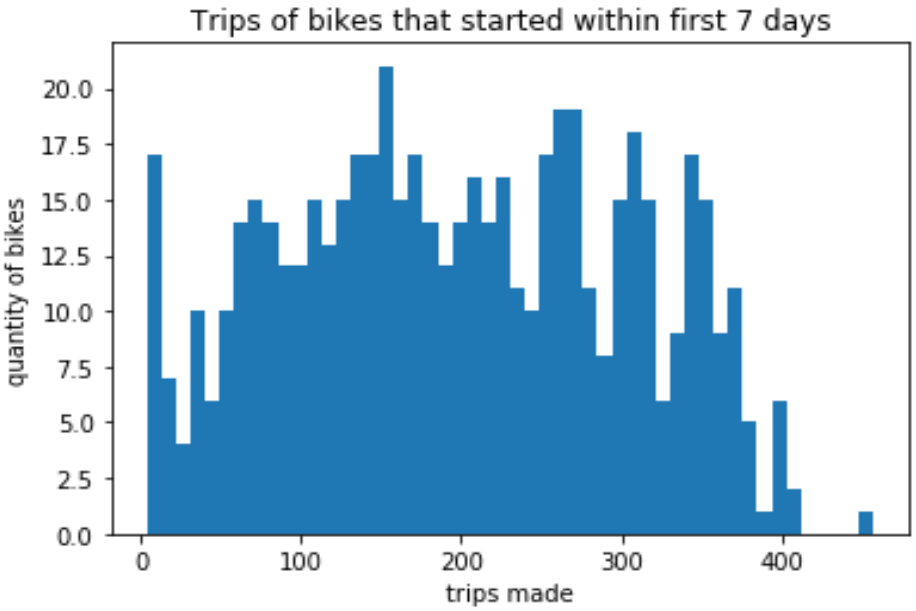


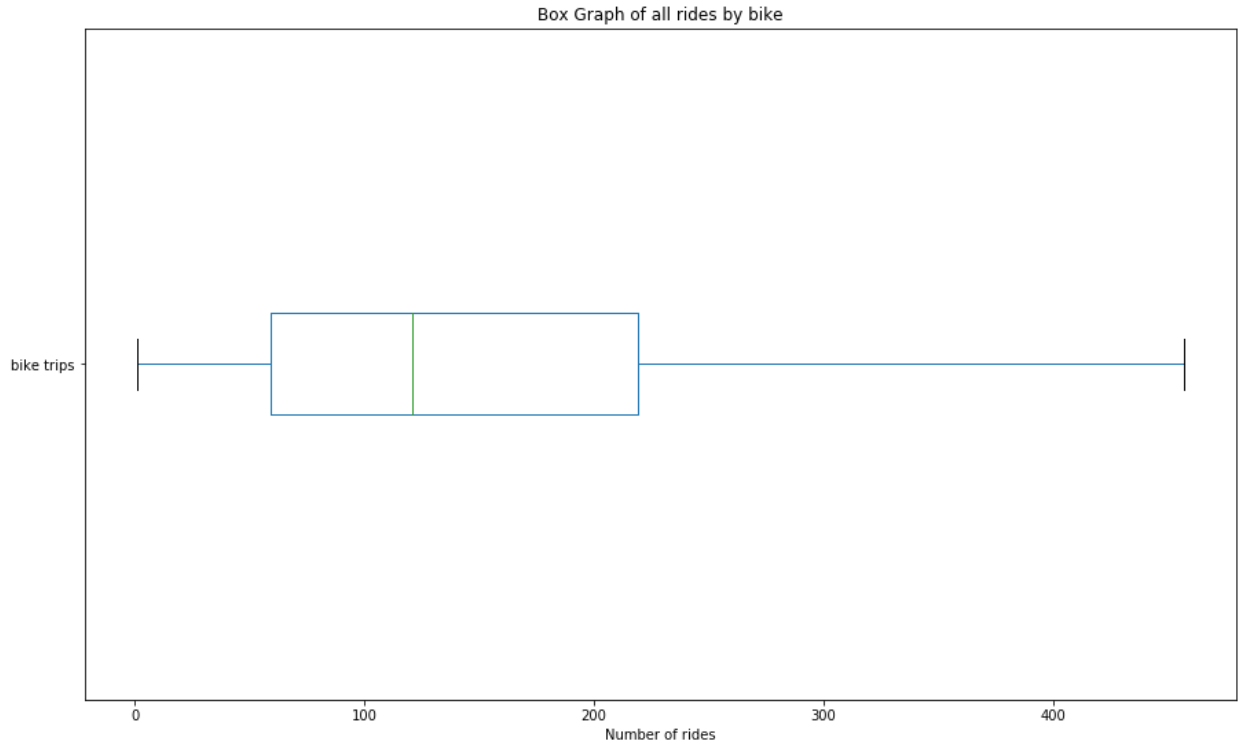
The average week shows a very similar picture, with a sharp drop off during the weekend for subscribers, and a subtle jump for customers on the weekend. The small drops throughout the week for subscribers are also interesting, though I am not entirely certain what it could mean.

3. How often are individual bikes ridden, and for how long?

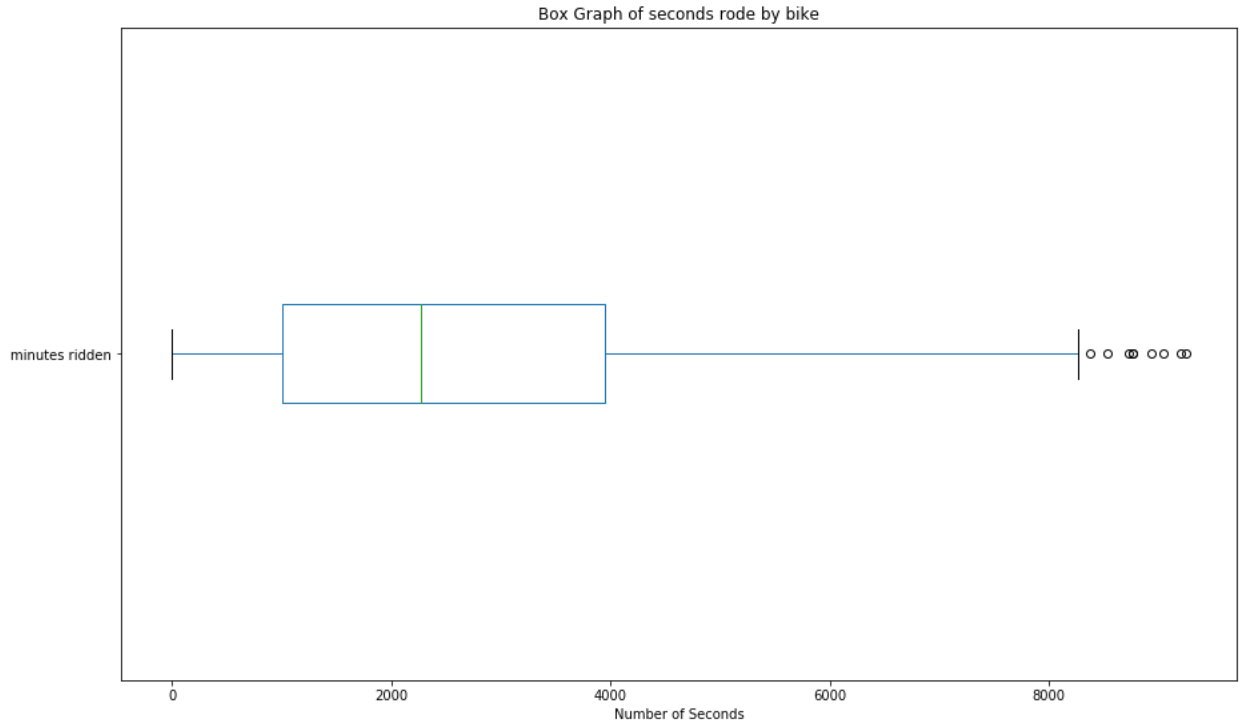
The last topic that must be included is that of the bikes themselves, since it can be very easily argued that this is the soul of the bike sharing service. We begin with understanding how often these bikes are used

The graph above denotes the number of trips made by bikes that started service within the first week of the dataset. This information needs to be built upon in order to answer the question above.

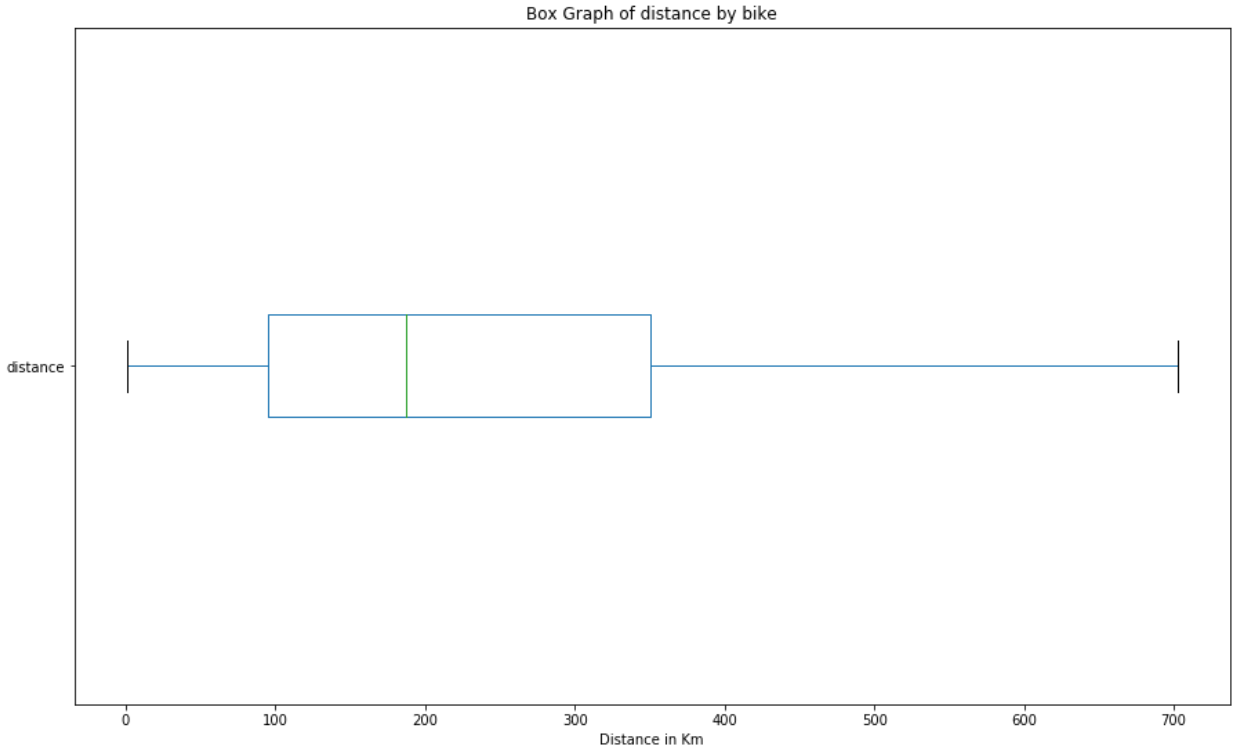




This box graph shows the range of trips taken by each one individually, from 5 to over 400 trips, with the median being around 130 trips, and the 75 percentile above 200 trips. This show a disparity in how often a bike is used. To try and get a better picture, let us look at the time that is spent on each bike.

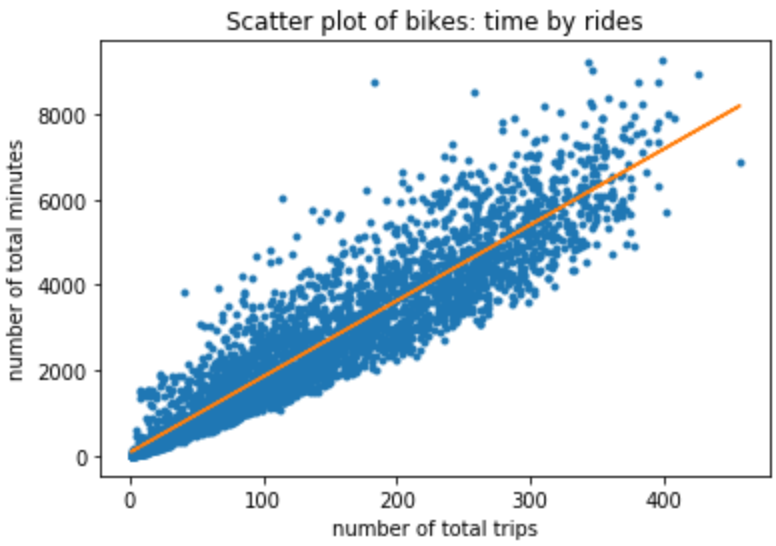


Again the box graphs shows a large range, that there are a gross number of bikes that are overused and some that are very underused. Even after noting some outliers above 8000 minutes.



Finally we have the distance of each individual bike. With again a wide range from 5Km to over 700Km.

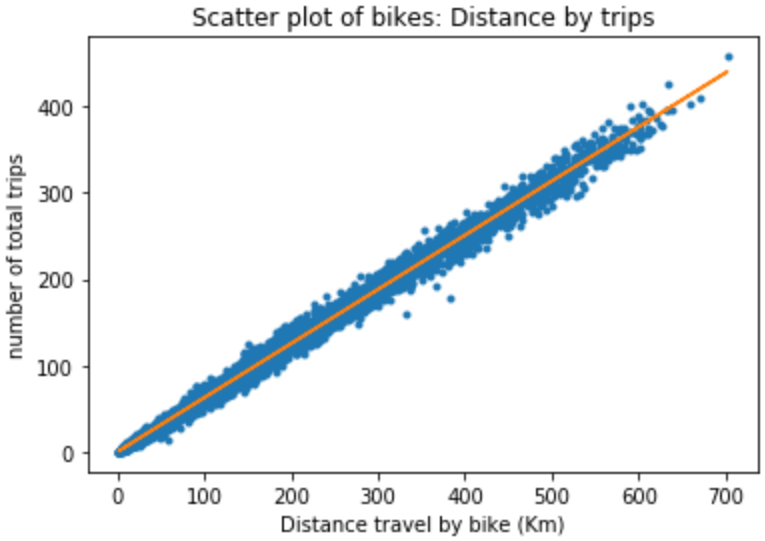
In order to make more sense of this, I created a scatter plot where all the bikes are paired with their corresponding trips and time and graph a regression line to note the general trend. This shows while a general trend is seen and that most bikes strongly follow this trend, that there are outliers and that there is a group of bikes that are being underutilized by the user.



Yet the regression line only states the general trend, in order to obtain the Pearson coefficients. These simply state correlation the the two sets of data are and the probability of an uncorrelated sets of data reach the same extreme as our data.

(0.9258288177133038, 0.0)

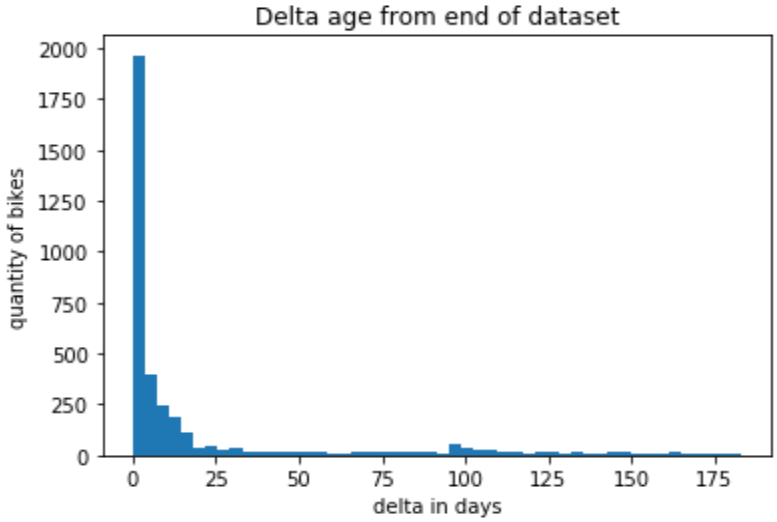
This confirms a strong correlation between the two data sets and that there is no chance that this can be duplicated by uncorrelated datasets. Meaning that we can estimate the distance or time of a newly acquired bike when we have only one value.



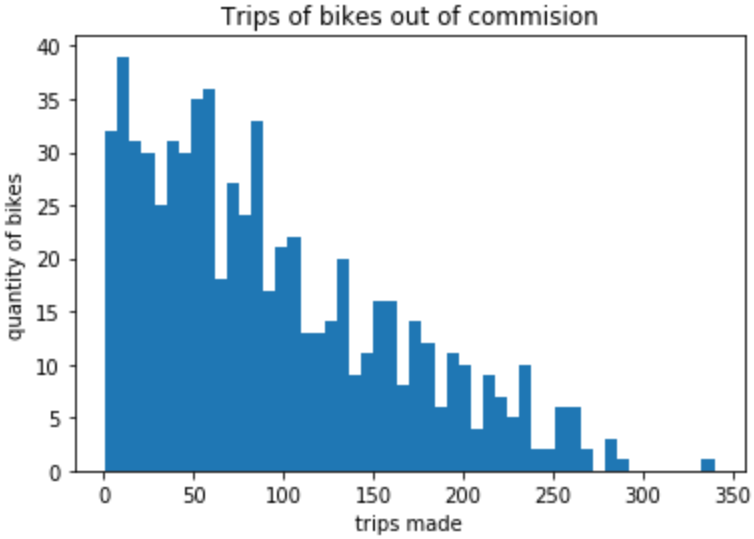
(0.9961652160552849, 0.0)

An even stronger correlation can be seen with Distance and trips and confirmed with the Pearson Coefficient of over .99. An almost linear scatter plot notes that this graph can easily be applied to all bikes and an possible accurate approximation can be made when you know only one value. The slope is .6238299811970669,which means nothing since it notes trip per distance, so inverting it makes it become 1.603 Km per trip.

How long do bikes last before going out of commission?



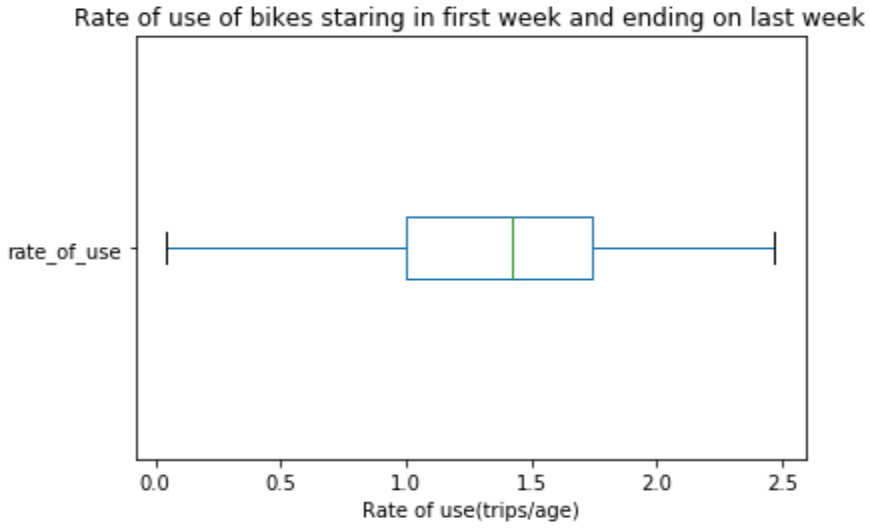
Using all bikes,we ask how do they behave at the end of the dataset. Here we see the delta from the end date of the individual bike to the end date of the dataset, noting the positive of the majority being zero, means that most were used on the last day of the data set. But there is again a sliver that reach pass the 100 days mark, meaning that these bikes were not used at all during the fall.



This graph was created by cutting right above the mean of 25 for the delta graph and seeing how many trips were made by these bikes. The reason for noting them as out of commision is because of the there underutilized nature, meaning there is a much lower likelihood of them

being picked up again or being discontinued. And as we can see, the trips are much lower than the average noted earlier, with a few exceptions being above 200 trips

Now let's also account for the fact that this service was here before the data, but that the bikes are long lasting, noting the bikes that started in the first 7 days and ended in the last 7 days. By finding a rate of use from the number of trips divided by the age of the bikes of this set. Thus giving us a range for the average rate of trips for long lasting bikes.



Giving us the the mean rate of use below 1.5, but we are more interested on a higher rate in order to truly determine whether the bikes noted as out of commision can be proven. So we will use the rate at the 75 percentile mark, 1.8.

The percentage of bikes with this rate of use in the out of commision category is about 42%. Meaning that it cannot be truly shown that these bike are out of commision and are just simply out of use. Which means that it is possibly that a few bikes go out of commision in this timeline, but it is not possible to note which ones or even how long others can last.

Conclusion

The Ford Go Bike service is a bike sharing service that is growing, though it is unclear if this is due to seasonal causes or a more general overall trend, with unique rides on the rise for both user types, but a heavy lean towards subscribers and the commuters. Marketing would likely be most effective at high volume stations during commuter hours. Along with implementing marketing for new stations around established mass transportation. Route analysis also shows that some method of transferring bikes to certain stations need to be implemented or eventually commuter will have no means of traveling, thus lowering potential profits. We assume this is already in place.

Continuing analysis can be done on current 2018 dataset to see current trends and how the service is doing when asking the same questions. Finding if there are bikes that have no rides, since they would technically not have a record, and if so finding out how many. Also further analysis into out of commision bikes and figuring out the precise lifetime use of a bike.