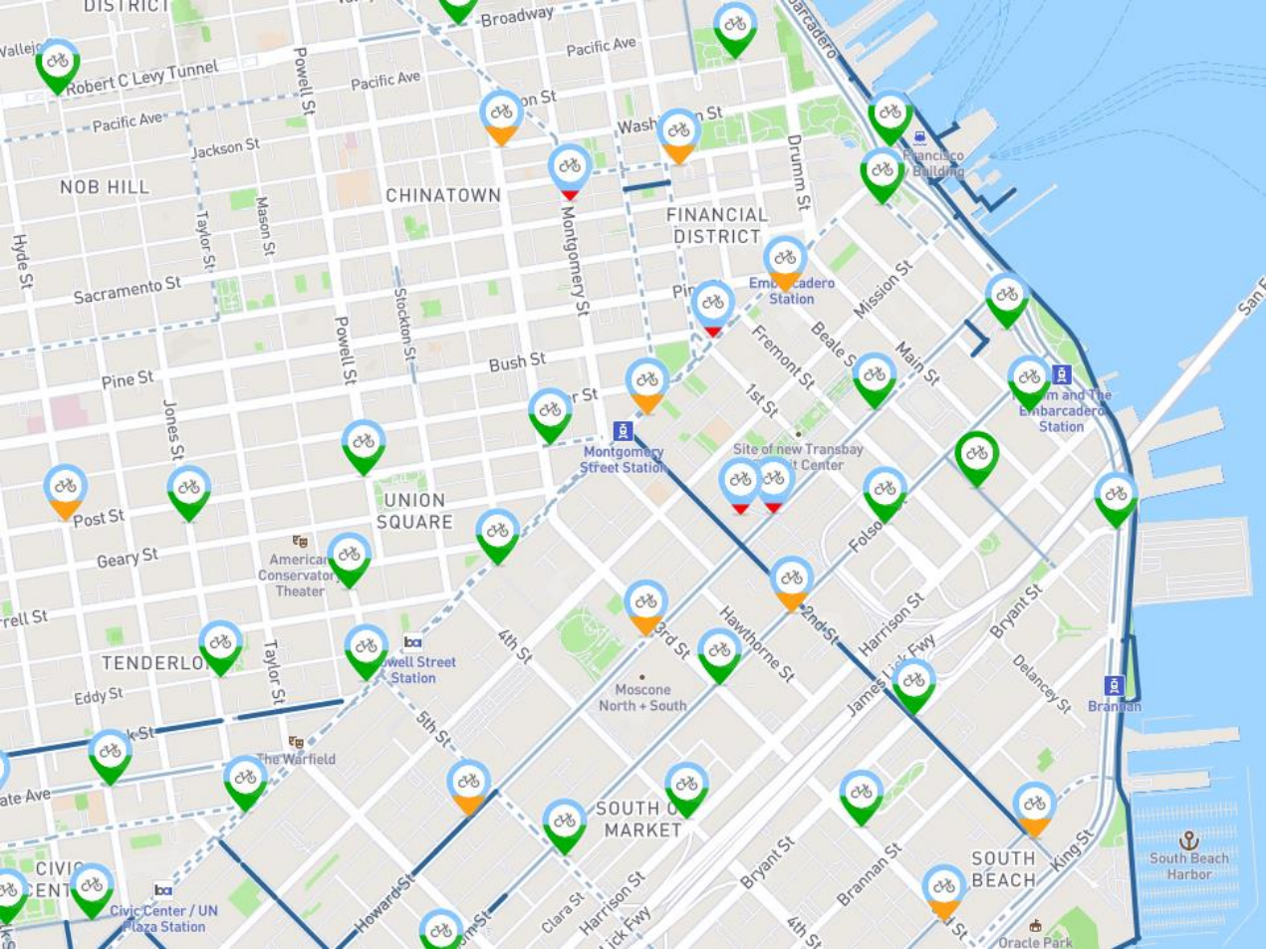


Data Analysis
Ford GoBike Explanatory Report

BY: ABEBE TAREKEGNE

PUBLISHED: ON MAY 10, 2019

What is Ford GoBike and How it Works?

Ford GoBike bike share is ***a convenient, healthy, fun and affordable way to get around.*** It's great for short/one-way trips as well as longer recreational rides. With annual membership or purchase a single ride or access pass you'll get access to thousands of bikes across San Francisco, the East Bay and San Jose.

Bike share involves a fleet of specially designed bikes that are locked into a network of docking stations. People use bike share to commute to work or school, run errands, get to appointments, and more.

Sources of Ford GoBike datasets for Investigation

We use Ford GoBike's trip data available for public use from here

The main attributes includes:

- ✓ *trip duration,*
- ✓ *start/end time and date,*
- ✓ *start/end station id, name, lat/long,*
- ✓ *bike id,*
- ✓ *user type*
- ✓ *member year of birth, and gender*

What are the main feature(s) of interest in the dataset?

I am most interested in the *`duration_sec`* but as the start and end locations or coordinates are provided it will be more interesting to find the distance and the average speed of each trip with that new features can be added.

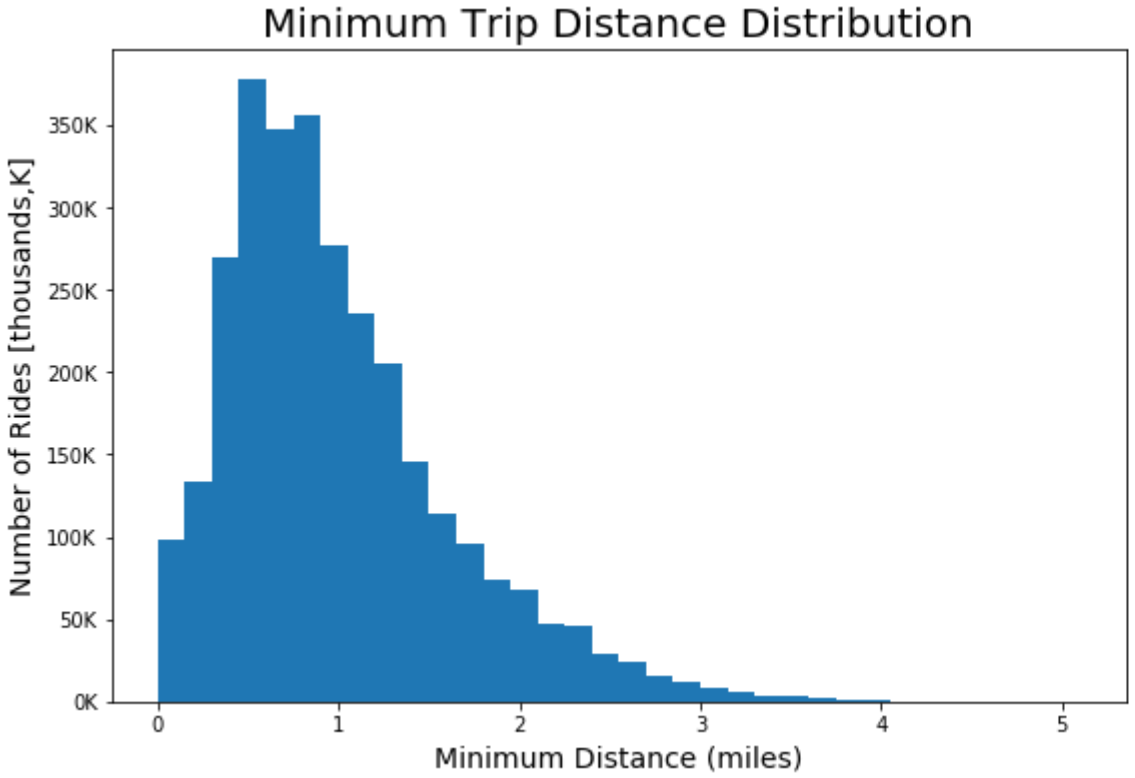
It is also interesting to investigate the trip duration and its relationship with *`user_Type`* (Subscriber or Customer – “Subscriber” = Member or “Customer” = Casual), *`member_year_of_birth`* and *`member_gender`*.

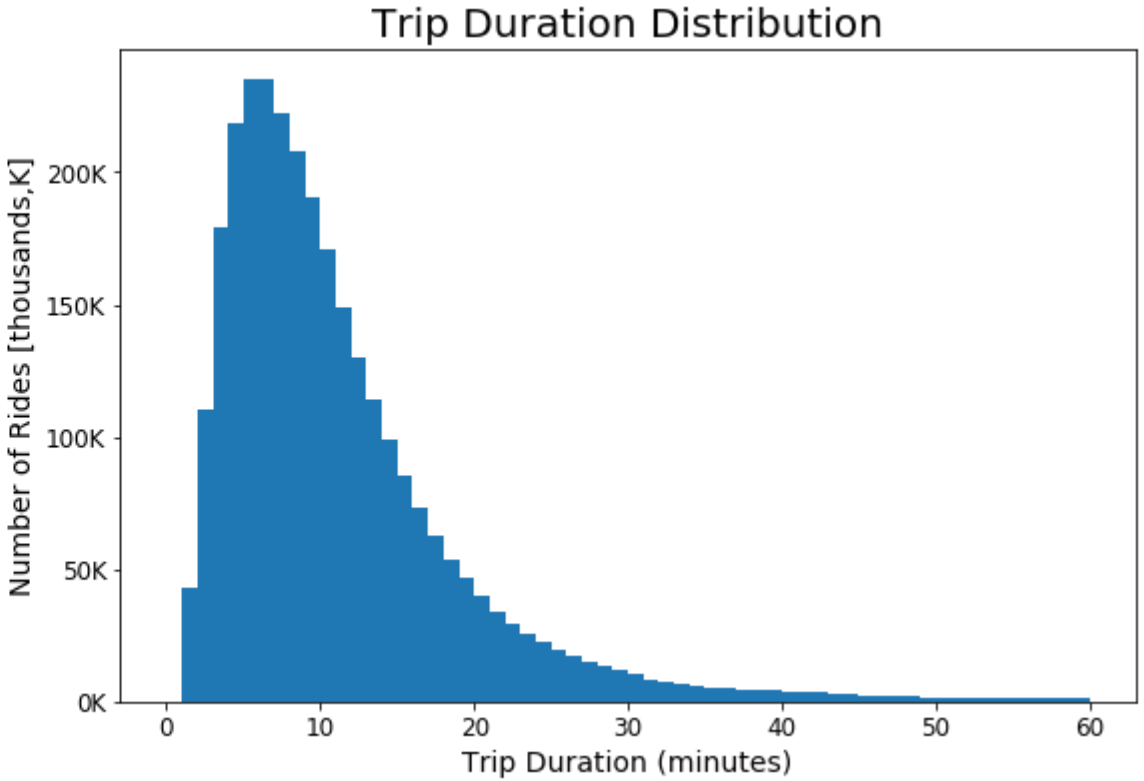
What new features are introduced?

We use geopy module for spatial data manuplation, it can calculate geodesic distance between two points. With time and distance then we calculated the average speed, so new feature called *`minimum_distance_miles`* and *`average_speed_mph`* were created.

Based on latitude and longitude filtering we create new feature called *`city`* and group as `San Francisco`,`Oakland`,`San Jose` and `Others`.

We created new feature called *`member_age_group`* which shows grouping based on the member age range. From `member_birth_year` we use 2019 as reference to find the `member_age`, and then we use bining to divide in groups of meaningful size for visualization.





Distribution of the main feature Trip Duration and Distance

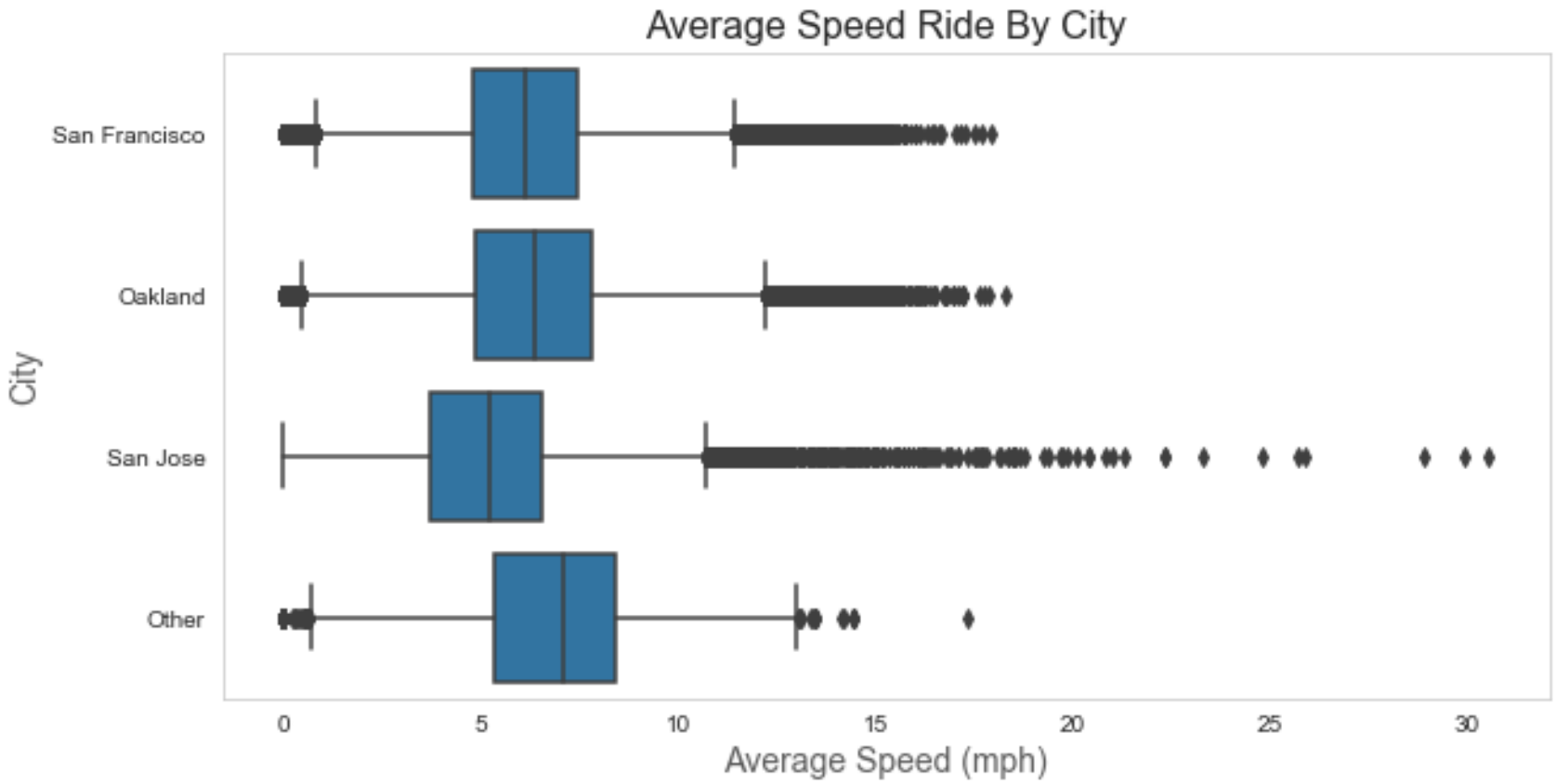
Both the trip distance and trip duration shows long-tailed distribution with a lot of data on the left, around 1miles and 10minutes.

What are the features statistics tells us ?

- ✓Average trip ride duration is 12.35 minutes and ride distance is 1 mile.
- ✓75% of the rides took less than 15 minutes and travel less than 1.4 miles.
- ✓99% of the rides took less than 102 minutes and travel less than 3.1 miles.
- ✓Subscriber or member riders are 84.7 % while casual customer are 15.3 %.
- ✓Male riders are 69.3 % while Female riders are 22.6 %.
- ✓Bike Share for all trips are only 7.3% while the rest are 75.6%
- ✓San Francisco, Oakland, San Jose riders are 74.4%, 20.7%, and 4.8% respectively

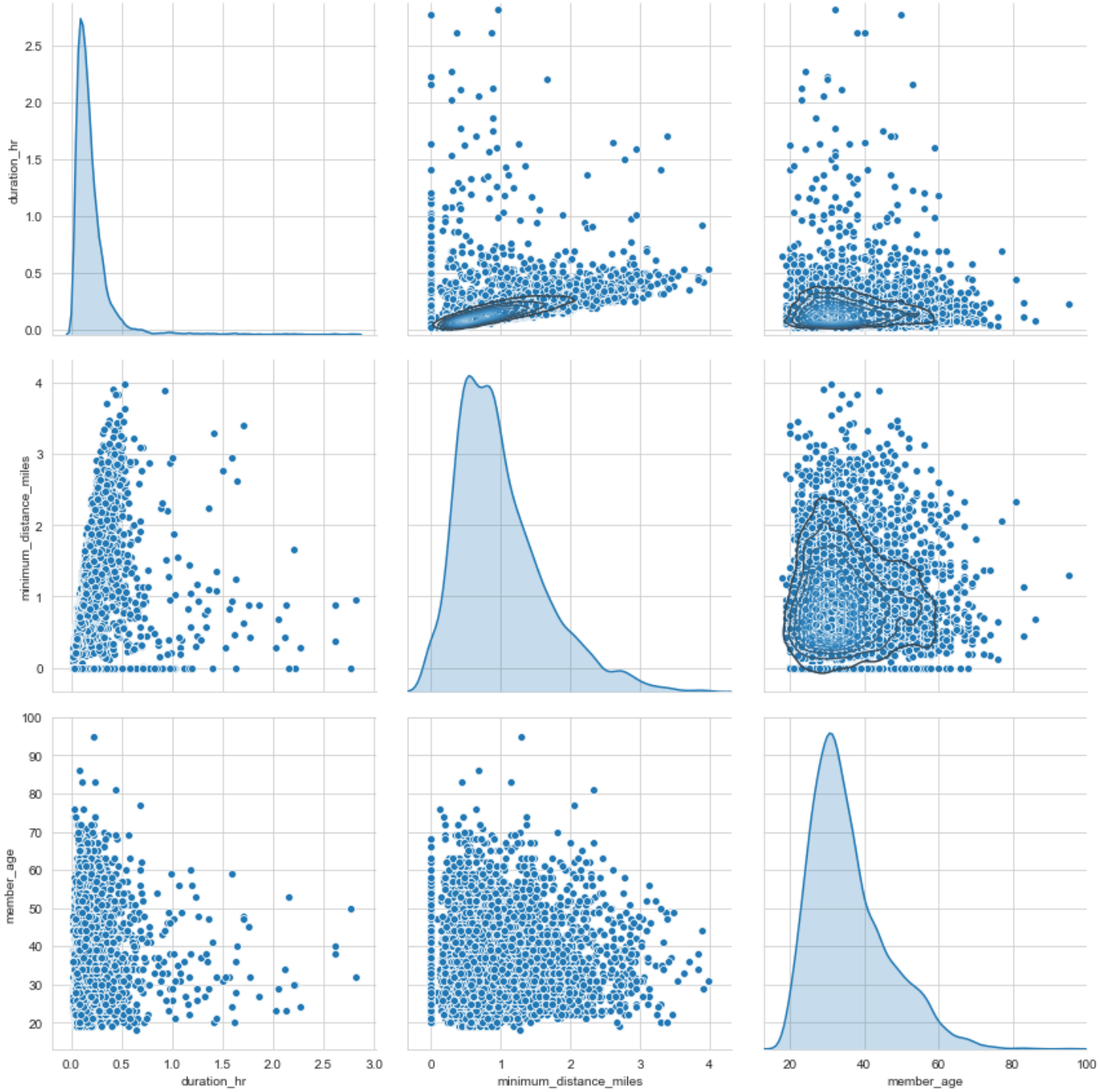
What are the features statistics tells us …?

- ✓The largest number of riders 67.2% are in the age range of 20 to 40 years old.
- ✓San Francisco members age from 30 to 40 are the largest number of riders of all age group. Thursday is the busiest day of all weekdays, Tuesday is the second busiest.
- ✓We observed that there are period where number of bikes in use decrease from 3.3K peak in March 2018 down to 2.8k, End of year 2018. And then increased at the high rate towards end of March 2019 with 6.8K unique bikes.



Average Speed Ride by Cities

San Francisco has IQR smaller than all but the median average speed, ~6mph, is about the same as Oakland.
San Jose shows a lot of data points beyond the outliers and the median average speed is the smallest, ~5mph.

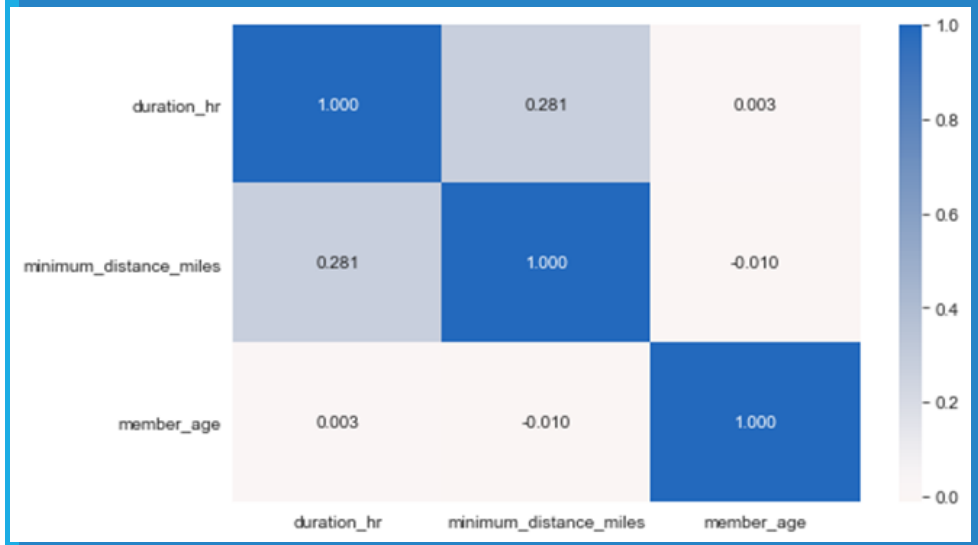


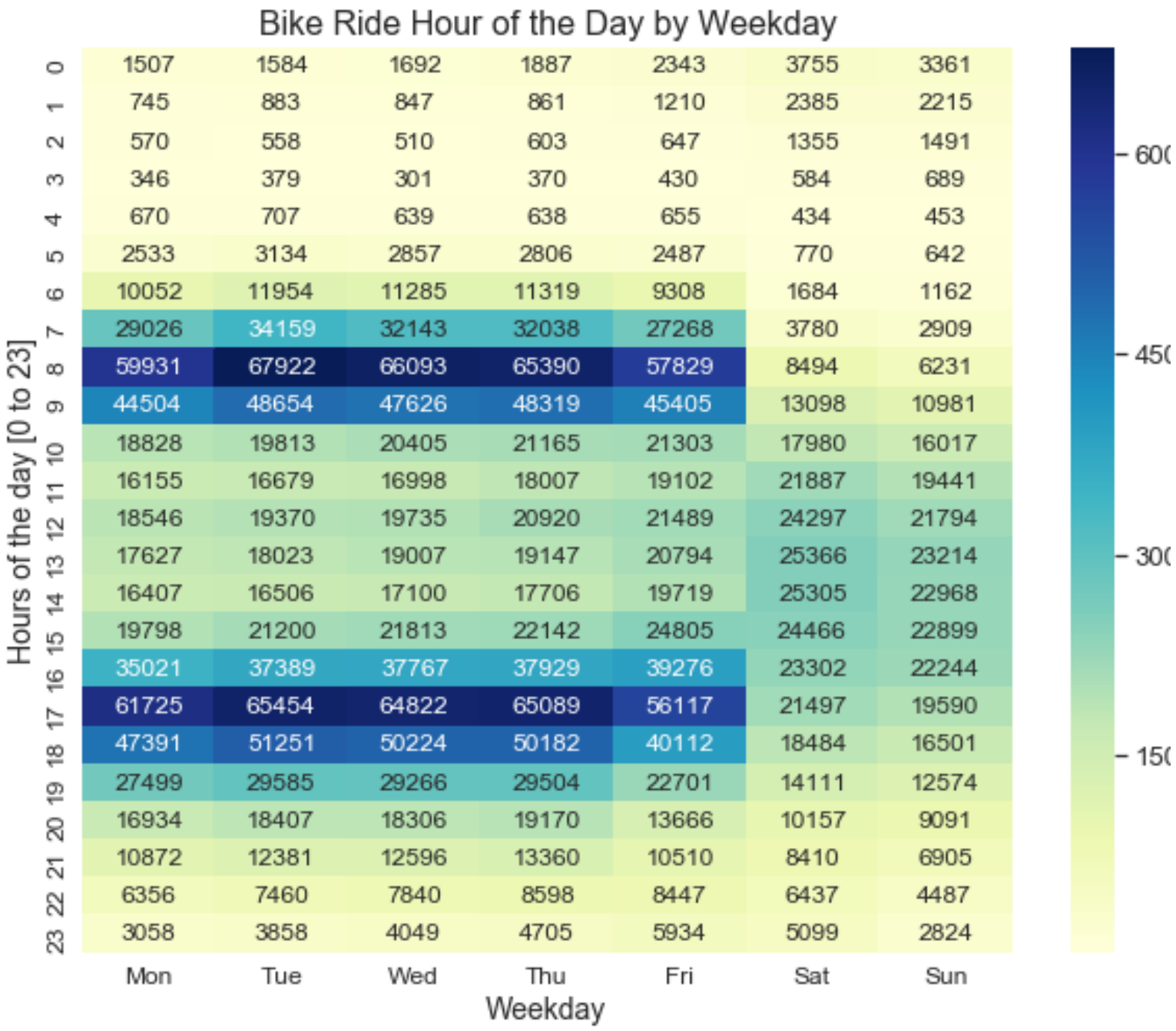
Feature Relationship: Duration, Distance, and Member Age

We see some correlation betwen duration and distances as shown in both pairplot and heatmap

*Duration_hr ~0.281*Minimum_distance_miles*

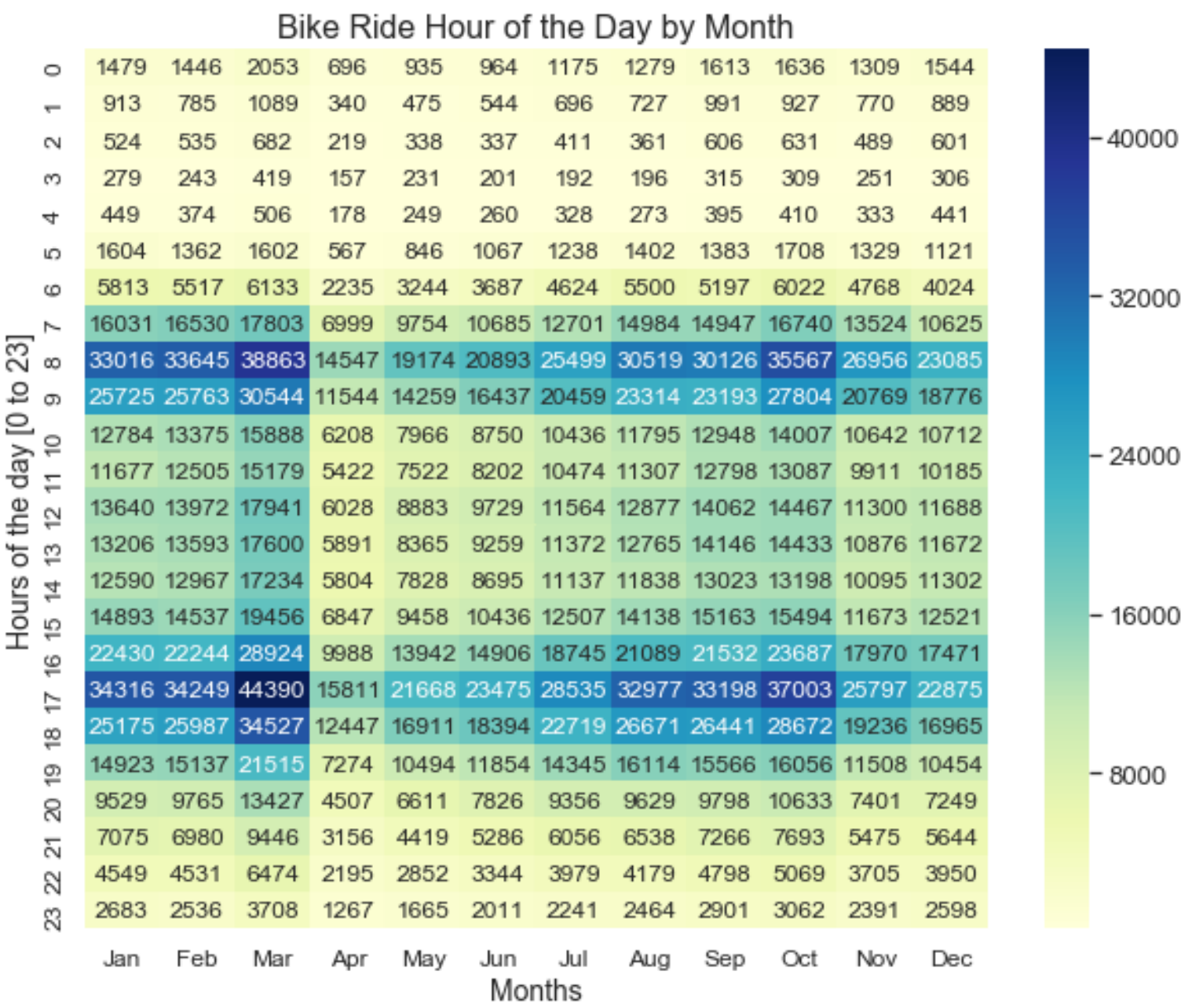
Maximum riders miles are most around 1 miles
Riders duration are most around 0.25hr or 15min.
Member age most riders are from 20 and 40 years





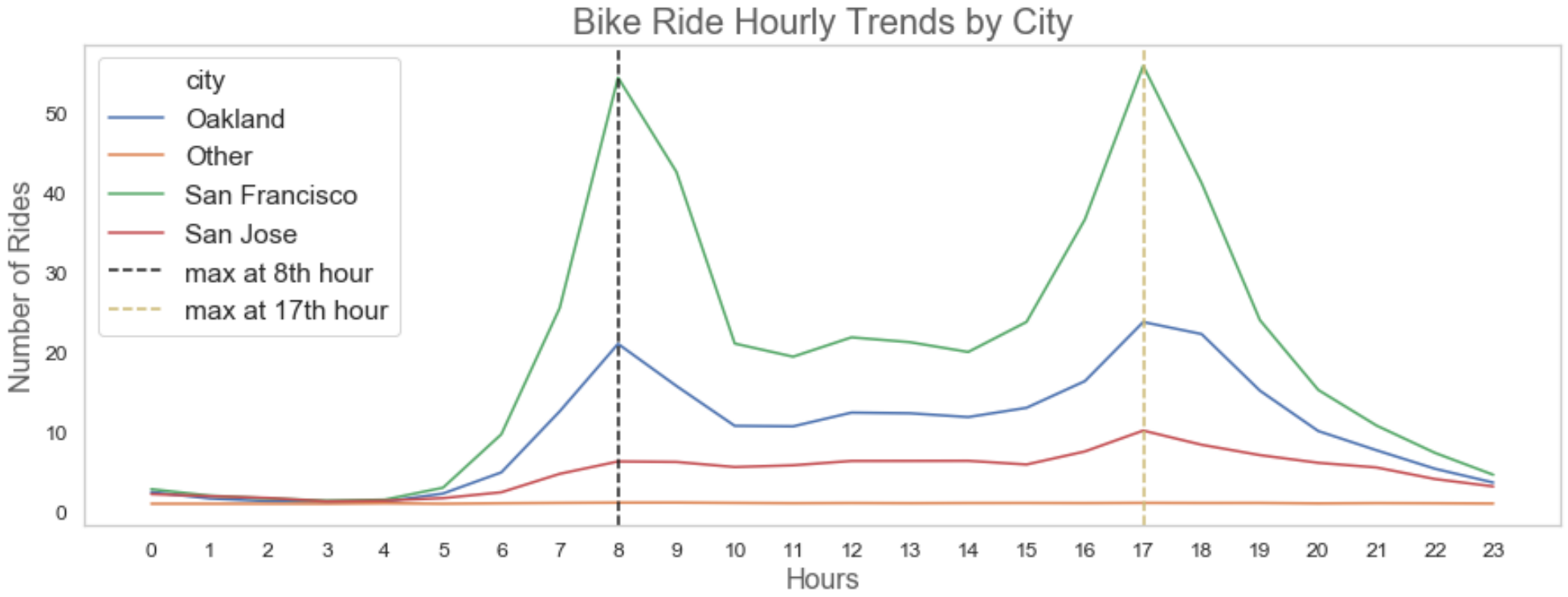
Bike Ride Hour of the Day relationship with Weekday

We see 8th and 17th hours are the busiest hours for each work week as expected, typical work hour start and end times. 2nd busiest are 9th and 18th hours, followed by 7th and 16th hours all around the work start and end periods. Tuesday seems the most used rides.



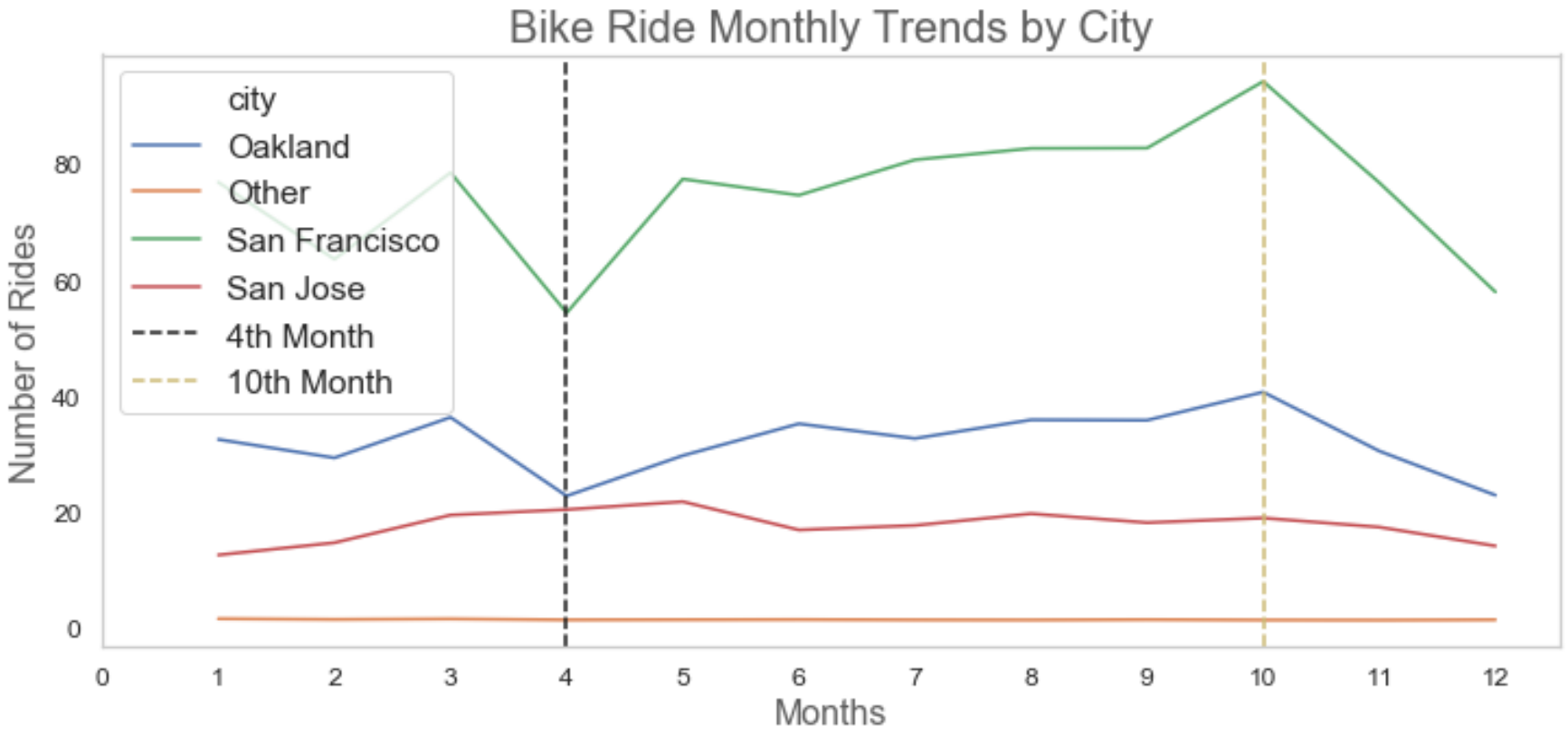
Bike Ride Hour of the Day relationship with Month

Busiest hours of the day generally follows the same as previously observed. March is the busiest and April is the slowest month of the year.



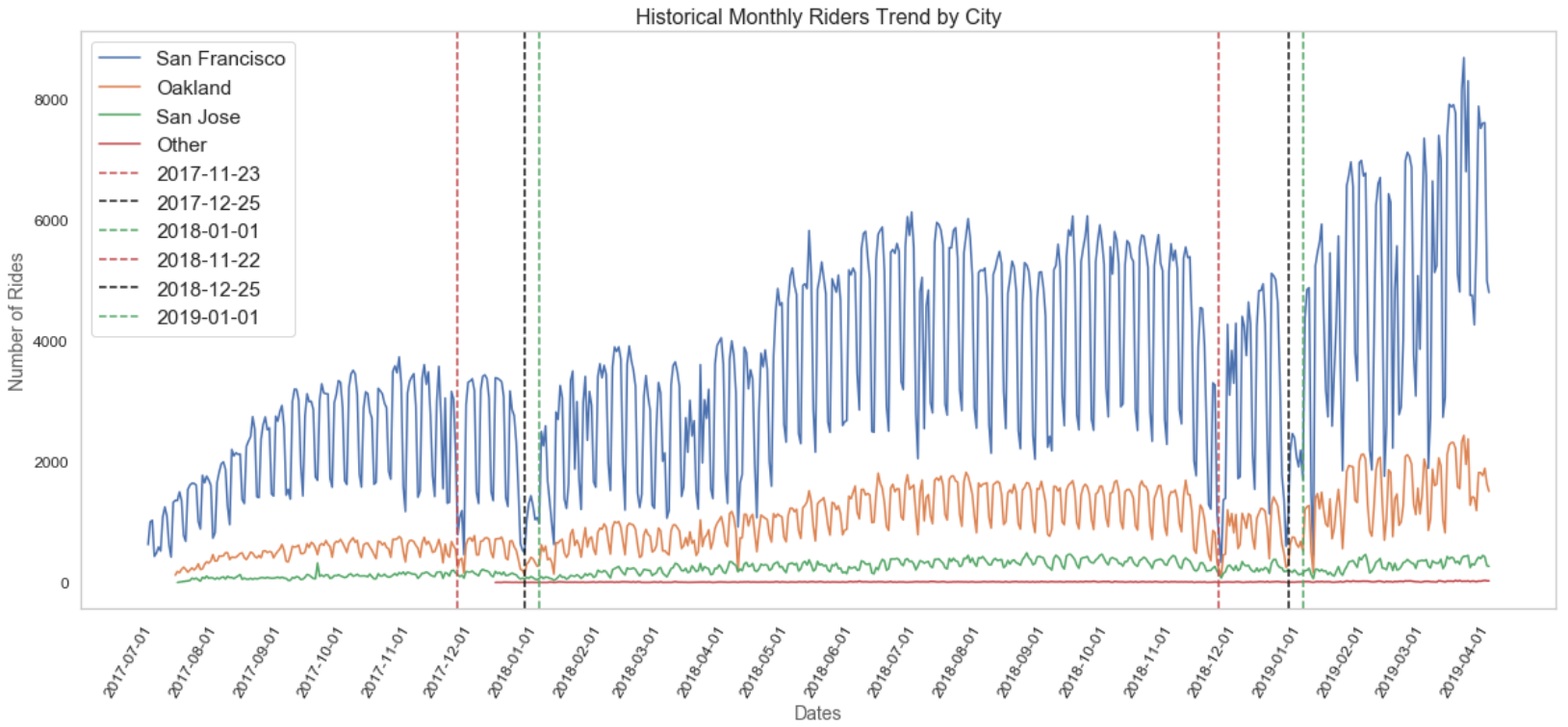
Bike Ride Hourly Trends by City

As previously observed this lineplot shows us the steep rise and fall around peak hours, 8th and 17th hours especially for San Francisco. Oakland follows the same as San Francisco for 8th hour but with lower magnitude. For 17th hour its rate of decrease is slower until 18th hour. San Jose, we can assume more consistency in the number of rides between 8th hours until 15th hours, then slowly rise at 17th hour.



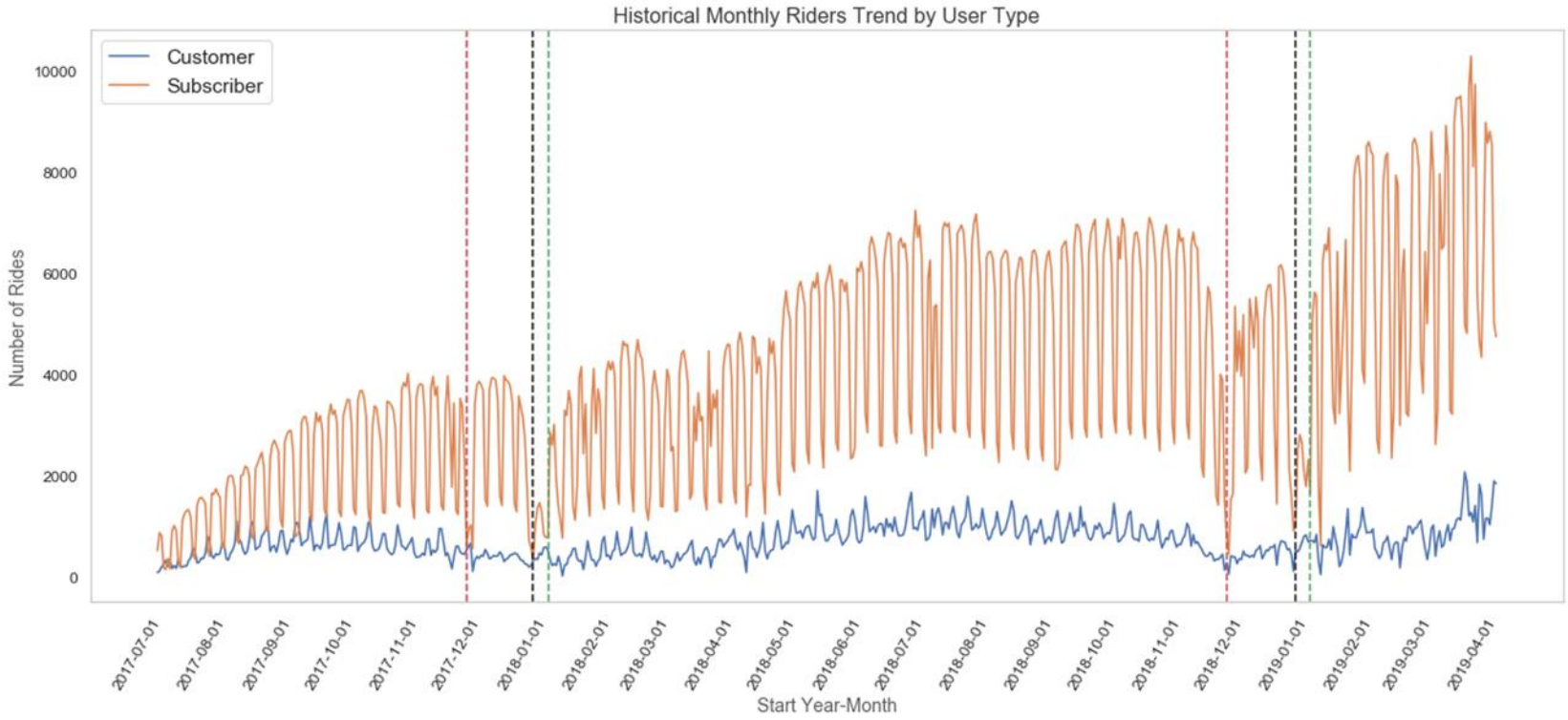
Bike Ride Monthly Trends by City

For both San Francisco and Oakland area we show decrease of riders in the month of April while it is peak in the month of October. San Jose shows relatively consistent number of rides across all months of the year.



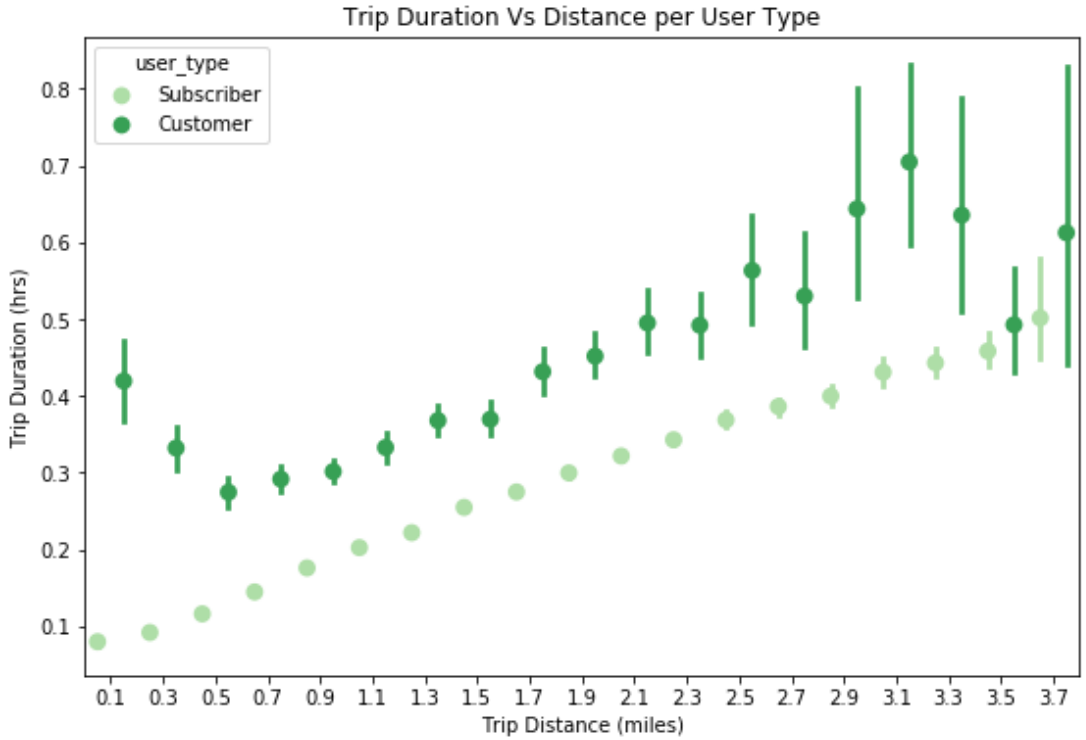
Historical Monthly Trend by City

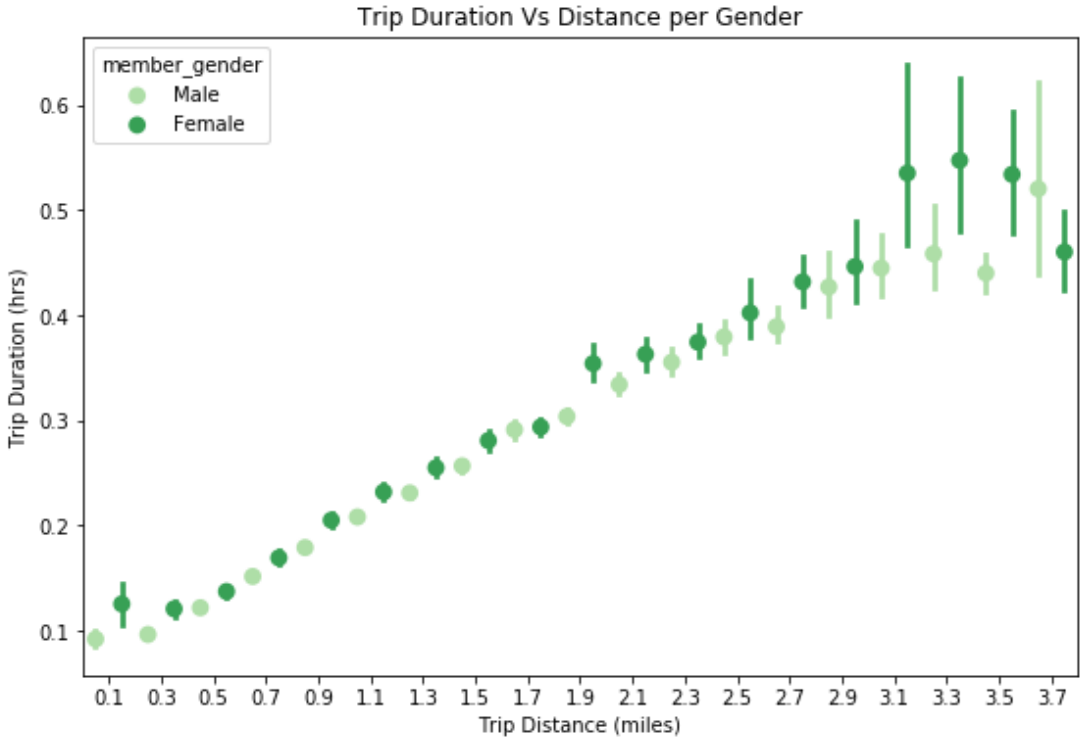
From previous analysis we show busy hours are aligned with the commute hours and from the lineplot what we show is the big drops on the number of rides are around Thanksgiving, X-Max, and New Year. The rate of increase after new year in 2019 is considerably very high comparing with the increase after January 1 of 2018. This is mainly due to the high number of bikes brought in service, about 2000 just in the first two months of 2019. Note that San Francisco riders rate of increase is huge compared to other cities.



Historical Monthly Riders Trend by User Type

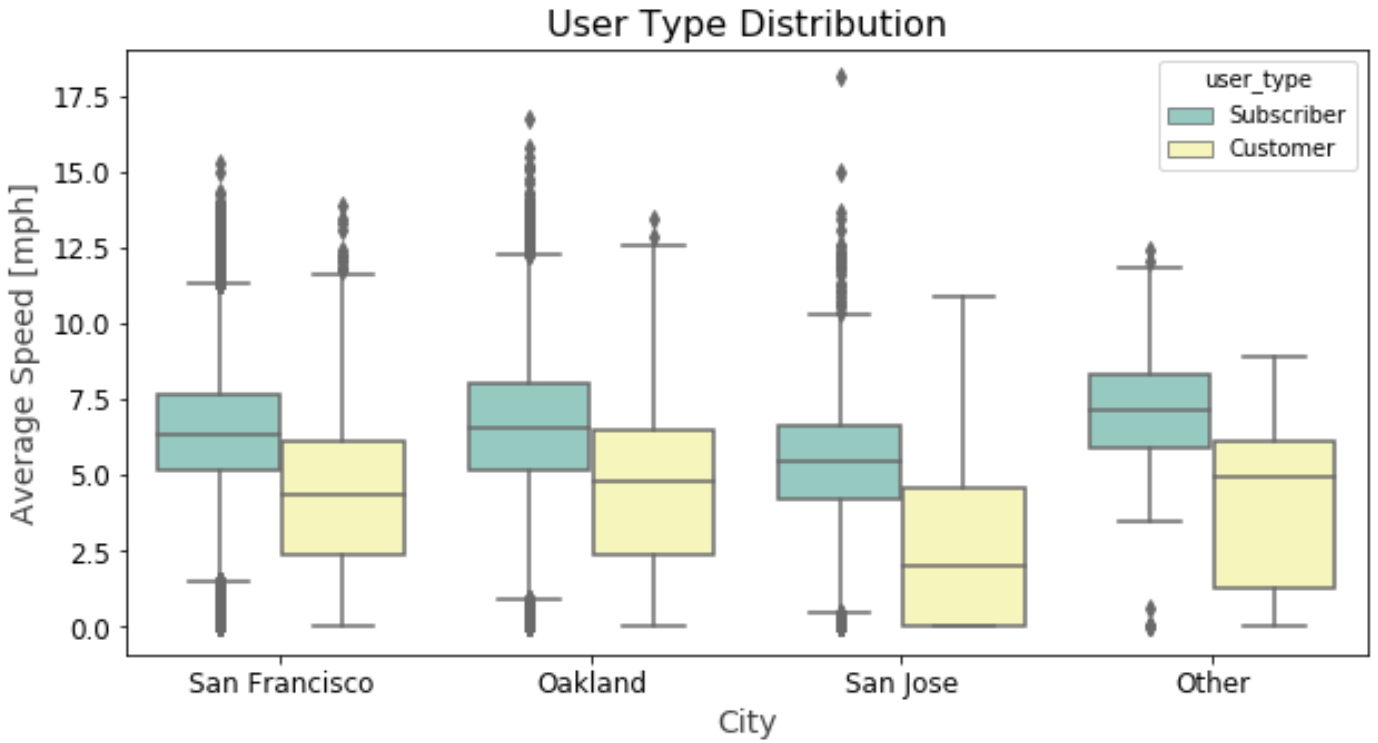
Increase of Subscribers riders are generally at higher rate than customer, this rate is more visible during 2019.





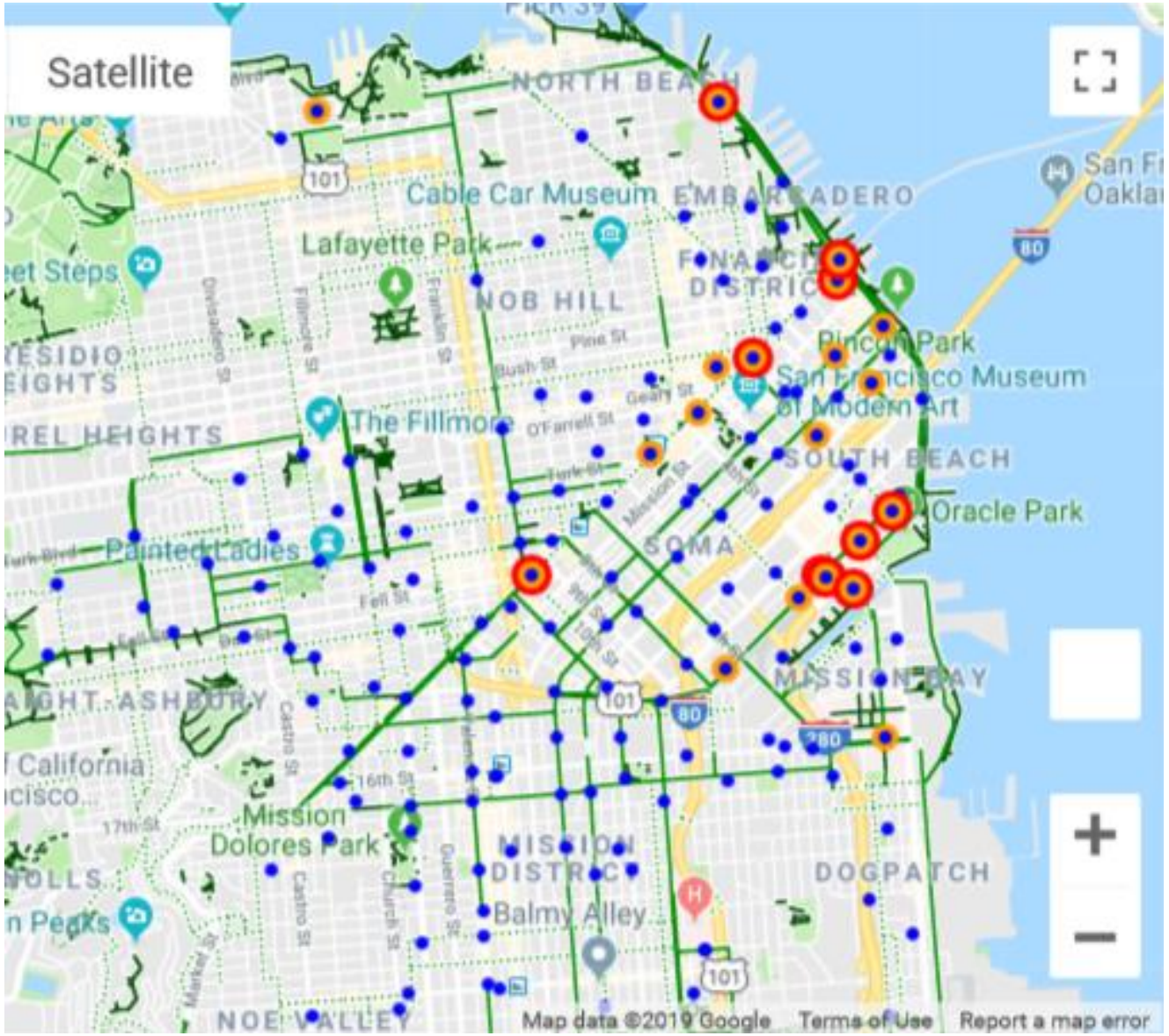
Trip duration Vs Distance per User Type or per Gender

Subscriber and Customer trip duration increases as trip distance increases. Subscribers trip duration are generally longer than Customers. Male and Female riders follow linear relationships between trip duration and distance.



User Type Interaction with Average Speed by City

It seems that the subscriber average speed is consistently higher than customers in all city, which is expected as casual customers could be new to the area and ride slower. Worth nothing that San Jose is the slowest for Customer but a bit higher for Subscriber. Need to investigate why this difference occurring assuming the traffic congestion will be the same for all riders.



San Francisco Area showing Top 10 and Top 25 Stations

What we show in Google Maps are: the Top 10 (red outer circle), Top 25 (orange outer circle) and blue dot for Ford GoBike stations.

Key Insights

The analysis of the dataset can help to tell story Ford GoBike how their services are effective and efficient. Understanding the historical growth of the number of bike rides of Ford GoBike network will help policy makers in order to facilitate for future growth of the infrastructure. And also helps to further study the benefit of environmental friendly nature of the service around the Bay area in general.

We show big gender gap as almost 68% of riders are men, this possibly indicate that there is a need for Ford GoBike to research possible cause of such gap. Is business dresses, high hill shoes etc. discourage women to use bike rides, or any other reason? Will introducing different style of bike will help to improve gender gap?

Key Insights

Increase of Subscribers riders at higher rate than customer is a good indication that more awareness by advertisement or environment friendly group joining the club. Bike Share for All are available to Bay Area residents ages 18 and older who qualify as Low Income member, another reason for increase of subscribers.

We identify the top busiest stations by number of riders hourly or daily. We show station id of 30, San Francisco Caltrain (Townsend St at 4th St) has served 71 riders at a maximum Busy Hour which occur on March 26 of 2019 at 8A.M.

Key Insights

And knowing the peak number of riders and maximum duration of bike rides per station used to determine if expanding for another nearby location is needed. It also helps to understand the most efficient use of resource allocation based on area, as San Francisco is the most utilized compared to other cities it is worth focusing for more growth.

Knowing the slowest periods as we show around Thanksgiving, X-Max, and New Year can be used to service bikes of high mileages in addition to malfunction bikes. Ford GoBike can prepare by introducing new bikes, new stations, etc. for the high demand periods immediately after new year as we show in 2019. Which was considerably very high comparing with the previous year especially in San Francisco area. Or re-focus on the potential growth opportunities in the city that has more demands or new areas with similar behaviour.