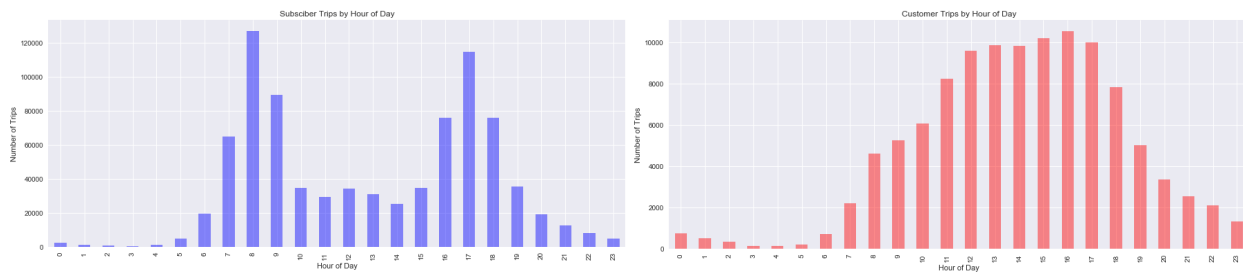


# Bay Area Bike Share Inferential Statistics

## Identify Key User Groups

The Bay Area Bike Share users have taken just over 980,000 trips between August 29, 2013 and August 31st, 2016. With more than 97% of trips taking no more than one hour to complete, this number is reduced to 955,000 total trips.

There are two key user groups in the data. Customers, who purchases rides in 24 hour or 3 day increments, and Subscribers, who purchase an annual membership. Subscriber Trips account for 844,570 trips which is 88.39% of all trips and Customer trips account for 110,987 trips which is 11.61% of all trips. Customers are busy during daylight hours (8a-7p) accounting for 95,953 trips, which is 10.04% of all trips, and 86.45% of all Customer trips.



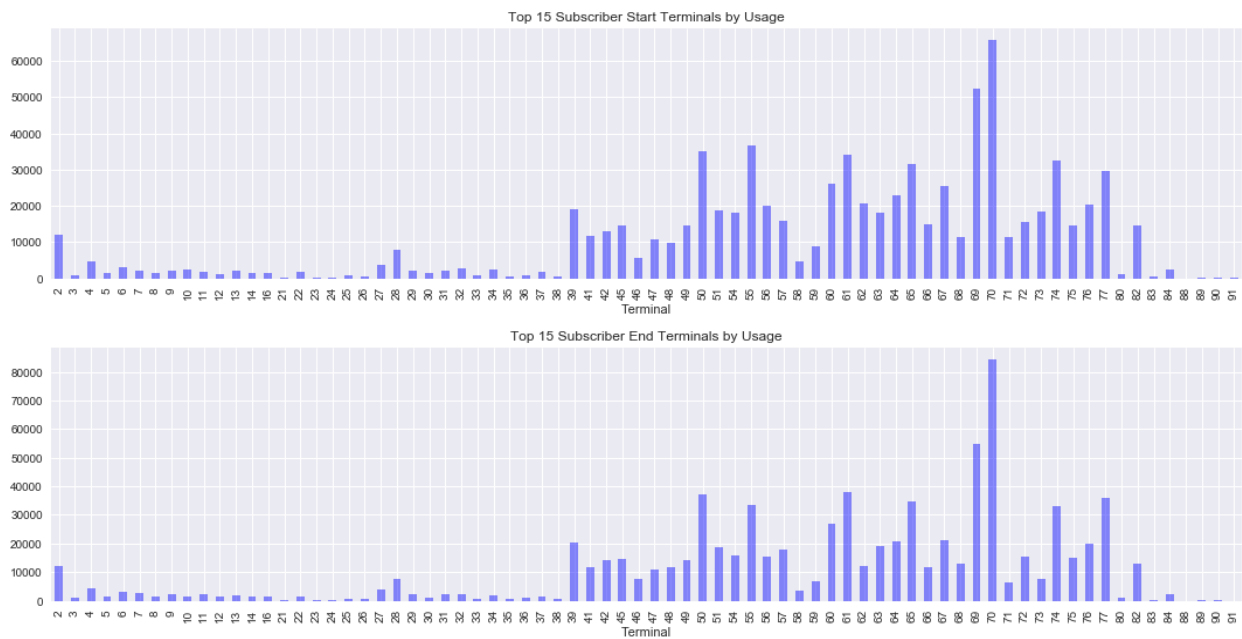
## Identify Key Start and End Stations

The Bay Area Bike Share has stations in 5 geographical areas, San Francisco, San Jose, Redwood City, Mountain View, and Palo Alto.



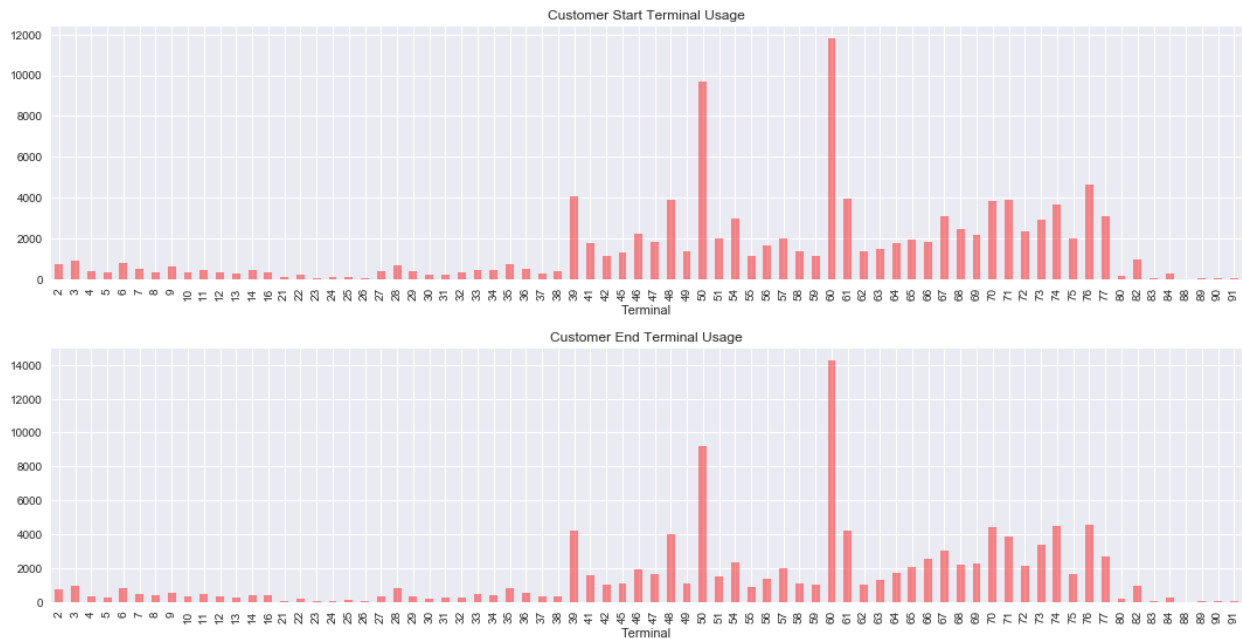
# Identify Key Stations for Subscribers

Subscribers heavily favor a distinct group of stations to start and to end their trips.



# Identify Key Stations for Customers

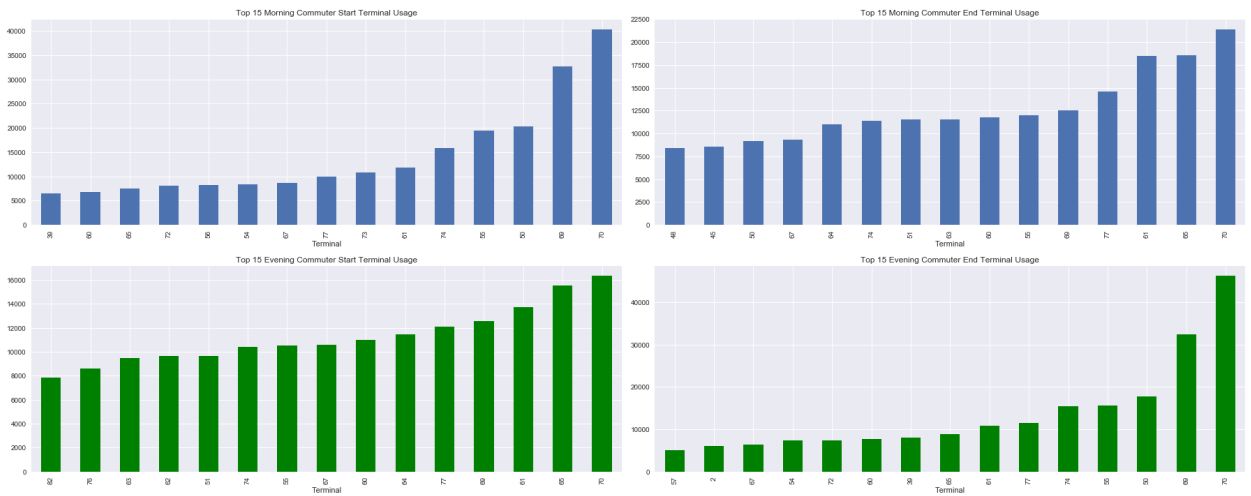
Customers heavily favor a distinct group of stations to start and to end their trips, and the data indicates that there are a large number of trips that start and end at the same station.



# Identify Key Stations for Commuters

Commuters heavily favor a distinct group of stations to start and to end their trips. The Morning Start Stations and Evening End Stations align closely indicating ‘Home’ Stations’, while the Evening Start Stations and Morning End Stations align closely indicating ‘Work’ Stations.

The locations of these terminals all support the assumptions of commuters making up the largest collection of users as they are at key Caltrain and Bart Stations, the local Commuter Rail and Subway systems.



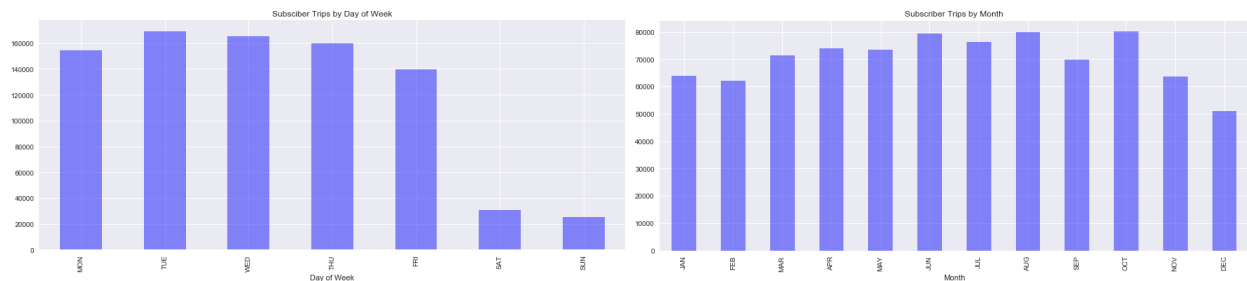
Top ‘Home’ Stations		Top ‘Work’ Stations	
70	San Francisco Caltrain (Townsend at 4th)	70	San Francisco Caltrain (Townsend at 4th)
69	San Francisco Caltrain 2 (330 Townsend)	65	Townsend at 7th
50	Harry Bridges Plaza (Ferry Building)	61	2nd at Townsend
55	Temporary Transbay Terminal (Howard at Beale)	77	Market at Sansome
74	Steuart at Market	69	San Francisco Caltrain 2 (330 Townsend)

---

## Identify Subscriber Usage Patterns

Subscriber usage clearly has visible peaks during morning (7a-11a) and evening (4p-8p) commute hours. These Morning Commuters account for 340,345 trips which is 36.61% of all trips, while Evening Commuters account for 318,936 trips which is 33.38% of all trips.

This assumption that most Subscriber trips are commutes is supported when we look at the number of trips by day of the week. Weekday usage outnumbers weekend day usage nearly five to one. Also, Subscriber usage by month is stable as well, with only a slight dip during November and December, likely associated with users being out of town for Holidays.



---

## Identify Customer Usage Patterns

Customers usage is a smaller number of the overall usage, but has distinct trends of their own. It is clear to see that the bulk of Customer trips are taken during daylight hours, but looking at Usage by day of the week, and by month a pattern of weekend usage and summer month usage being key time periods is evident.

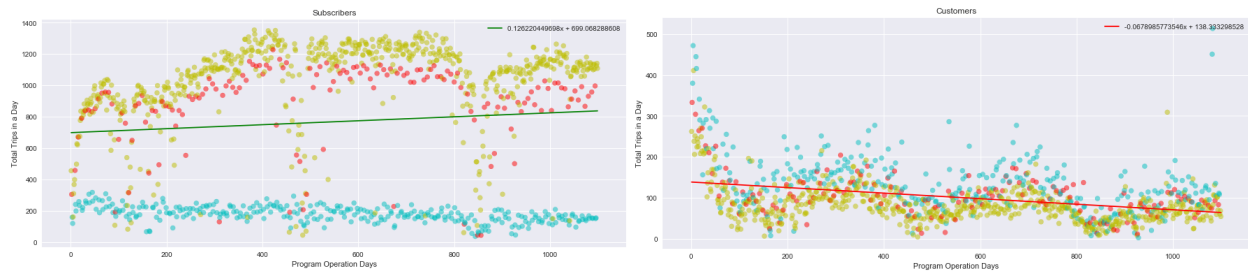


---

## Statistical Analysis

Plotting the number of daily trips for user group overtime has an interesting trend early on. In the first two months of the Bike Share Program's operation, there is a sharp decline in the number of Customer trips and a reciprocal boost in the number of Subscriber trips.

Calculations conclude that the growth pattern in the first 60 days of each user group are not equal to the growth rate in the remainder of the data, and as such we can safely disregard it when calculating long term growth trends.



For Subscribers, we first create a null hypothesis that the growth rate of the first 60 days is equal to the growth rate for all remaining days the program has been operating.

$H_0$  : Growth Rate for Subscribers in first 60 Days = Growth Rate for Subscribers after first 60 Days

$H_a$  : Growth Rate for Subscribers in first 60 Days  $\neq$  Growth Rate for Subscribers after first 60 Days

Conducting a Two Sample T Test, we arrive on a T Statistic of **-3.178** with a P Value of **0.0015**, so we can safely reject the null hypothesis. Concluding that the growth rate during the first 60 days of the program is not equal to the growth rate in the rest of the data. As such, we will can safely disregard this time period when analyzing the data for long term trends in program growth.

For Customers, we first create a null hypothesis that the growth rate of the first 60 days is equal to the growth rate for all remaining days the program has been operating.

$H_0$  : Growth Rate for Customers in first 60 Days = Growth Rate for Customers after first 60 Days

$H_a$  : Growth Rate for Customers in first 60 Days  $\neq$  Growth Rate for Customers after first 60 Days

Conducting a Two Sample T Test, we arrive on a T Statistic of **19.141** with a P Value of **1.081e-17**, so we can safely reject the null hypothesis. Concluding that the growth rate during the first 60 days of the program is not equal to the growth rate in the rest of the data. As such, we will can safely disregard this time period when analyzing the data for long term trends in program growth.

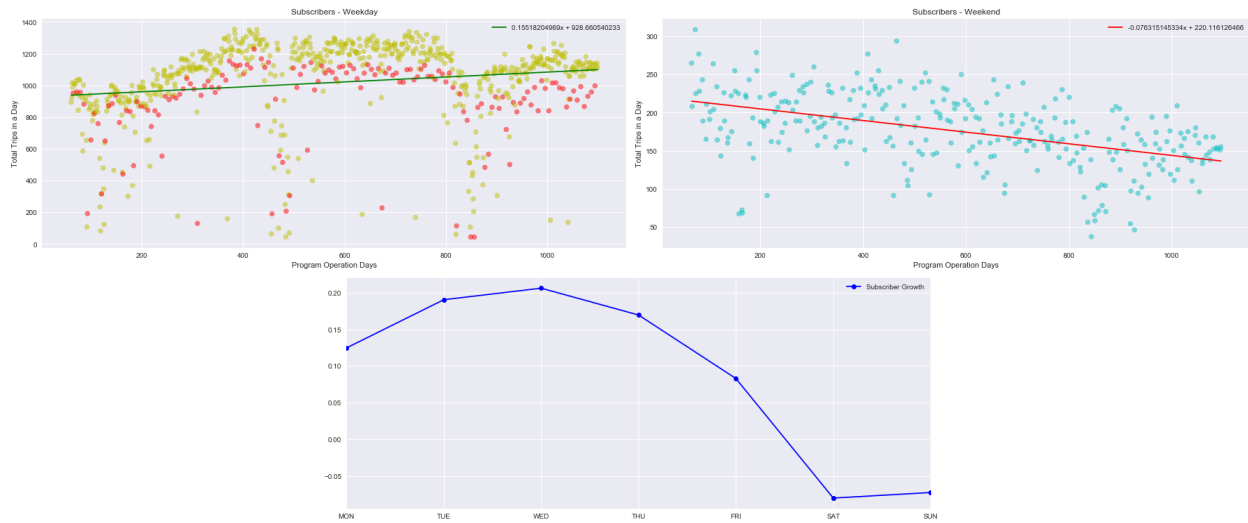
---

## Subscriber Usage is growing on weekdays, shrinking on weekends

Overall daily Subscriber usage is growing by 0.087 trips per day

Weekend daily Subscriber usage is growing by 0.115 trips per day

Weekend daily Subscriber usage is shrinking by 0.076 trips per day



---

## Customer Usage is Shrinking, but not too much

Overall daily Customer usage is shrinking by 0.035 trips per day

Weekend daily Customer usage is shrinking by 0.029 trips per day

Weekend daily Customer usage is shrinking by 0.050 trips per day



A workbook of all these calculations and further investigation can be found in this [Jupyter Notebook](#)