

# Assignment – Case Study

## Bay Area Bike Share Data Analysis

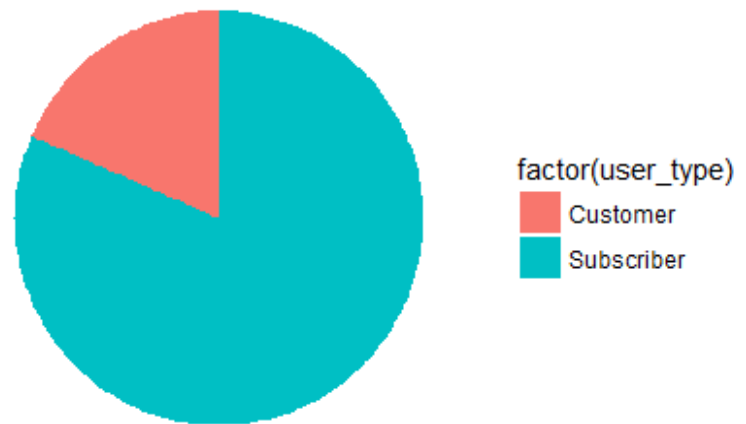
(Vinod Varma)

I have downloaded the datasets as CSV files on BigQuery from 2017 to March-2018 and preprocessed accordingly for the analysis. And I used R Programming for the same, please find the code in appendix.

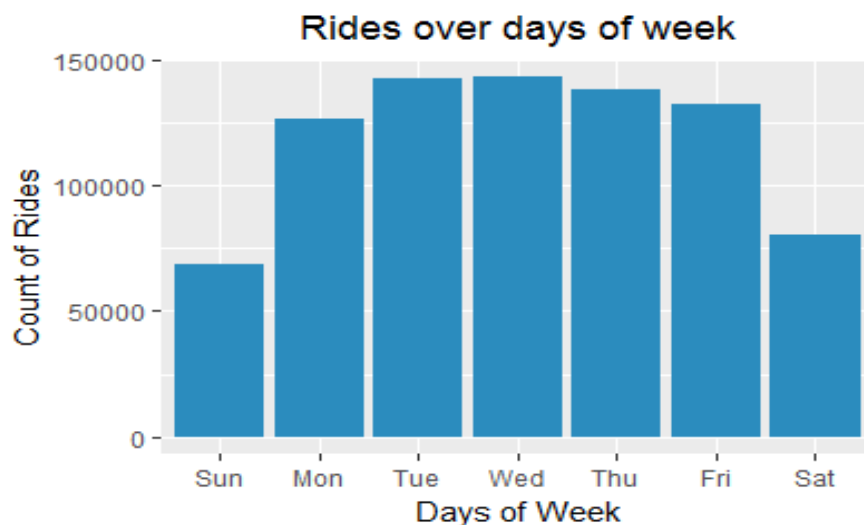
### Who uses Bay Area bike share?

#### Riders by Membership

Almost **80%** of riders have an annual subscription. They are called as Subscribers. The remaining **20%** of riders were Customers who purchased a 24-hour or 3-day pass.



### Distribution of rides by days of week:



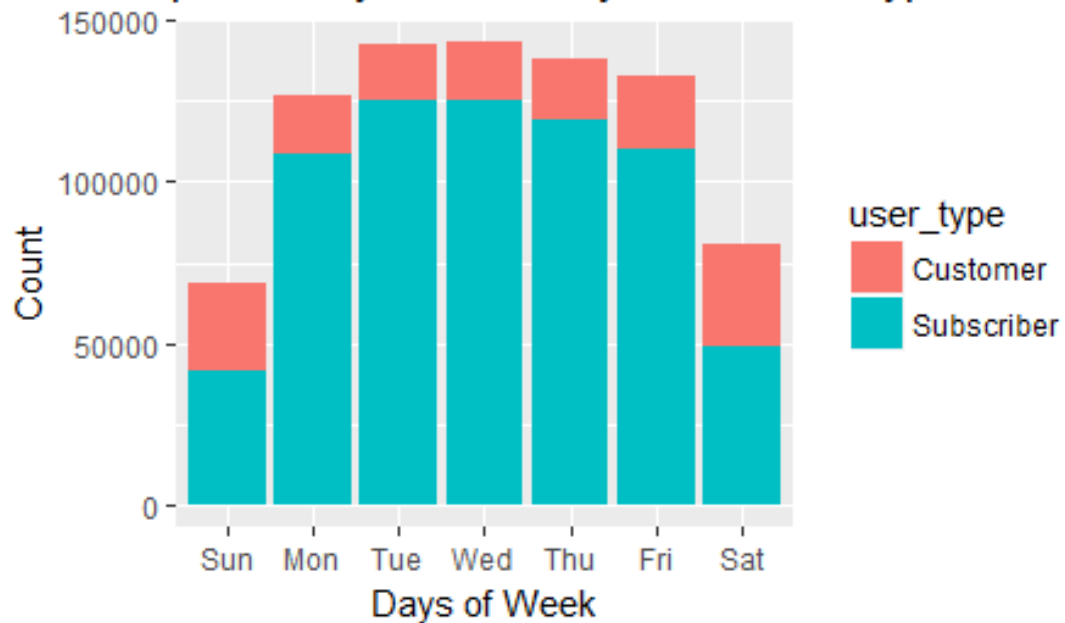
From the result, the number of rides occurred mainly during weekdays and fell to a lower number during weekends. Hence, we would expect the availability of bikes problem in the bike stations/docs to occur more during weekdays.

# Assignment – Case Study

## Bay Area Bike Share Data Analysis

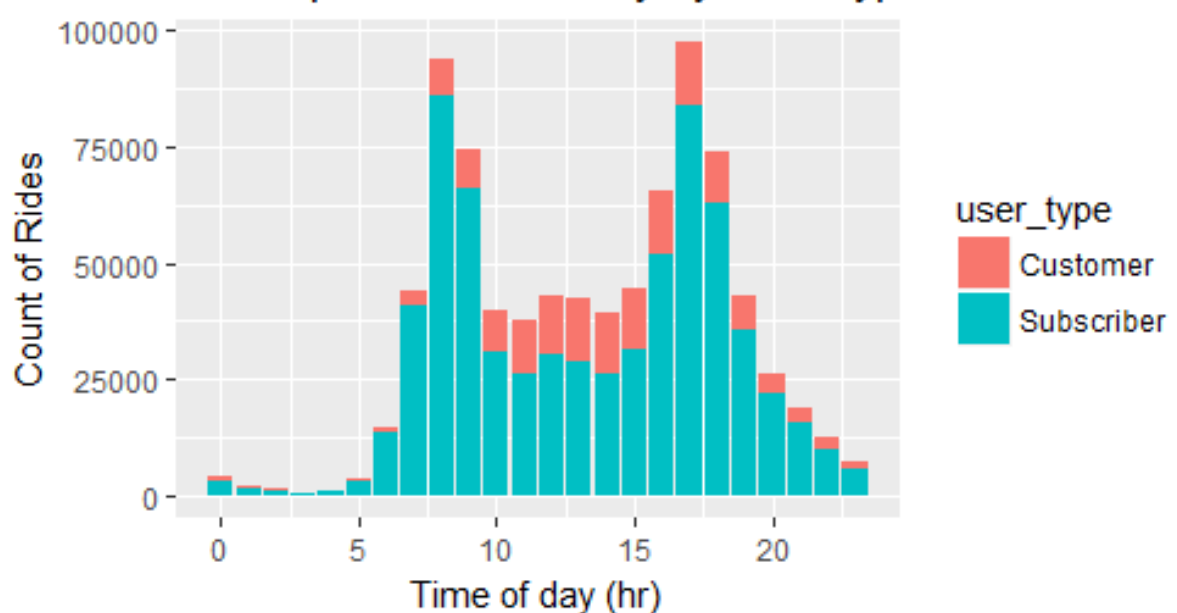
(Vinod Varma)

### Ridership over days of week by Subscriber Type



This brings me to the point, could the number of non-subscribers (customers) drive up the usage on weekdays? As much as we want to cater to both groups of customers, however, if there is a large number of non-subscribers using the bikes, this would cause much displeasure to the subscribers. Hence, we may want to increase the prices for customers' usage such that we can allow more subscribers to use instead.

### Ridership over time of day by user Type



# Assignment – Case Study

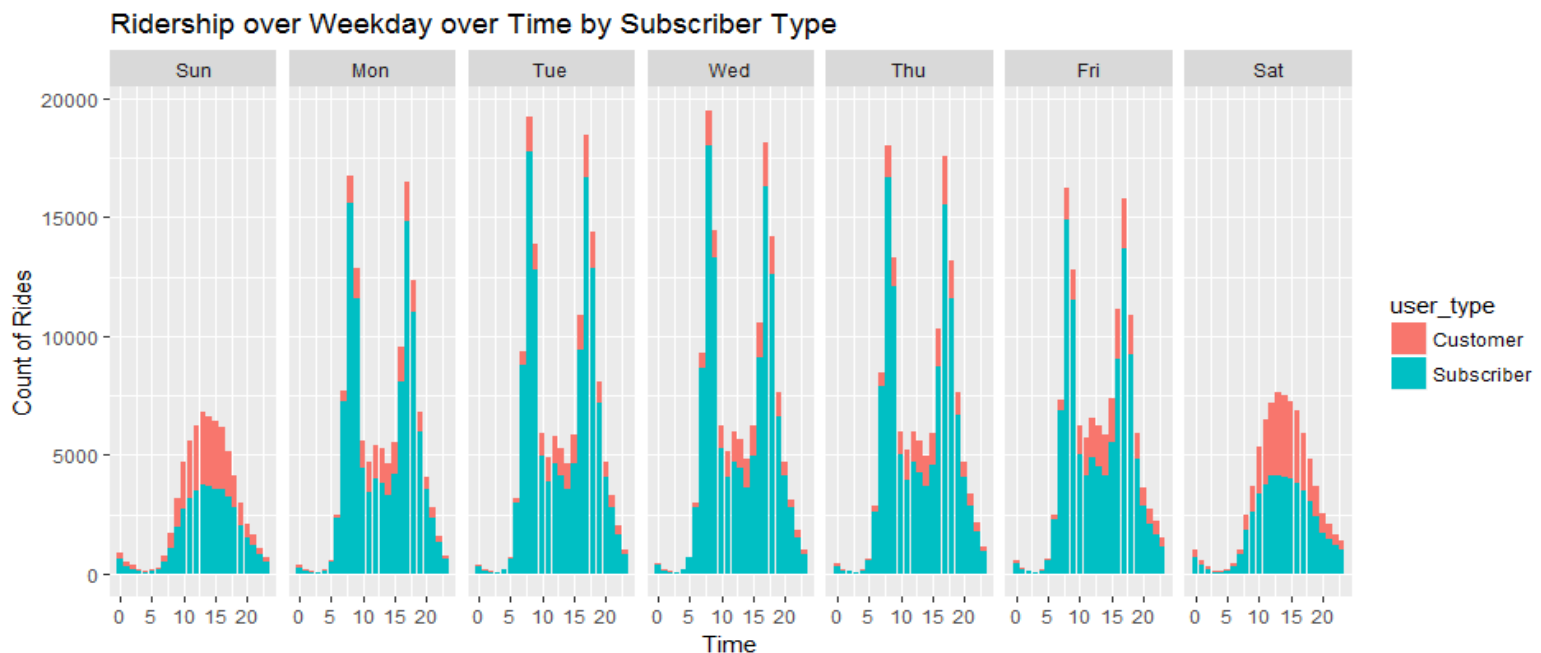
## Bay Area Bike Share Data Analysis

(Vinod Varma)

Among subscribers we see spikes in use at 8am and 5pm with another small bump at 12 noon. These users must be riding a bike to get to work, to go to lunch, and to head home.

And customers' hourly usage seems to fall along a bell-shaped distribution peaking at two in the afternoon. There doesn't seem to be a lunchtime rise among customer use. These users must be riding around throughout the daytime at their leisure.

Now let's combine all these factors together and see how does ridership varies over the days of week over hour of the day, broken down by subscriber type.



From the above plot few points to take away are:

- The distribution of trips over time during weekdays is approximately the same across all 5 days, having peaks at 7 - 9 a.m. and 4 - 6 p.m.
- The distribution of trips over the weekend is, different from weekdays but similar across both days.
- The number of trips taken by customers are roughly more than the number of trips taken by subscribers on weekends

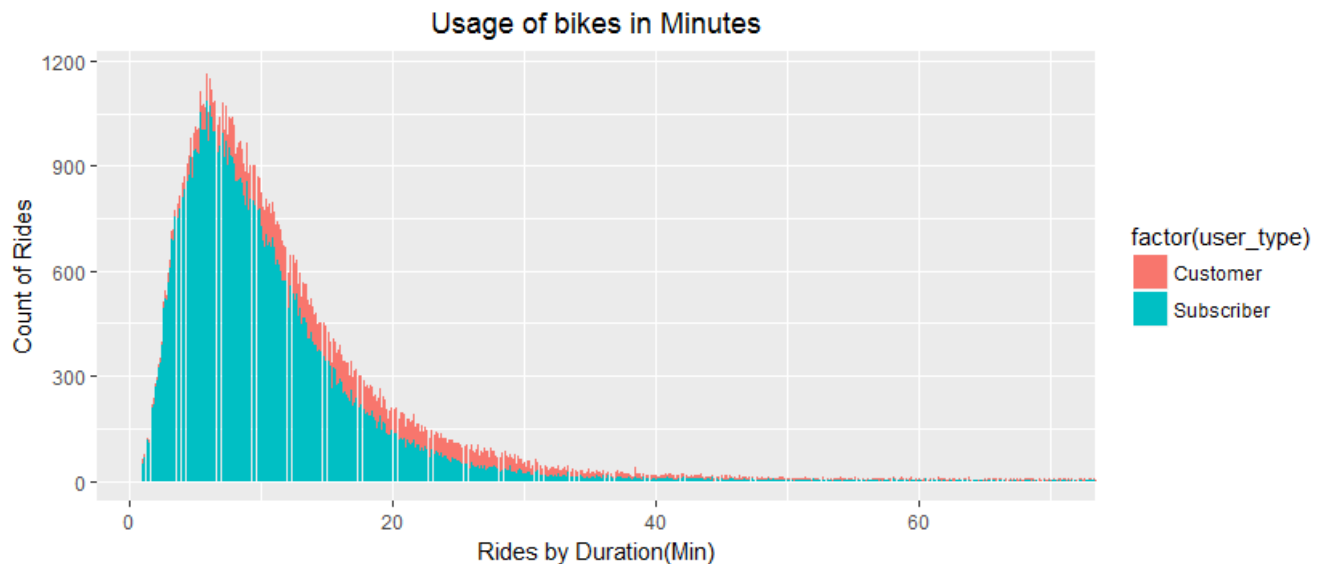
# Assignment – Case Study

## Bay Area Bike Share Data Analysis

(Vinod Varma)

### When is Bay Area Bike Share System is most used?

In the Bike Share company website, it has been mentioned that the share system is intended to be used for short rides: i.e trips under a half hour do not incur any additional charges. So here I wanted to find out whether riders are using the system in the intended way!



And Yes. The chart above shows that most common ride duration is in between 5 to 10 Minutes. Ride Share System Subscribers are clearly savvy about the pricing structure. A very small number of rides are taken for more than half an hour. And for the most part customers are also savvy as well.

### Most Popular Start & End stations of Rides by numbers:

	start_station_name	n
1	San Francisco Ferry Building (Harry Bridges Plaza)	22357
2	The Embarcadero at Sansome St	19668
3	San Francisco Caltrain (Townsend St at 4th St)	19558
4	Market St at 10th St	18392
5	San Francisco Caltrain Station 2 (Townsend St at 4th St)	17810
6	Berry St at 4th St	17506

	end_station_name	n
1	San Francisco Caltrain (Townsend St at 4th St)	26787
2	San Francisco Ferry Building (Harry Bridges Plaza)	25302
3	The Embarcadero at Sansome St	23903
4	San Francisco Caltrain Station 2 (Townsend St at 4th St)	21293
5	Montgomery St BART Station (Market St at 2nd St)	20228
6	Market St at 10th St	17145

# Assignment – Case Study

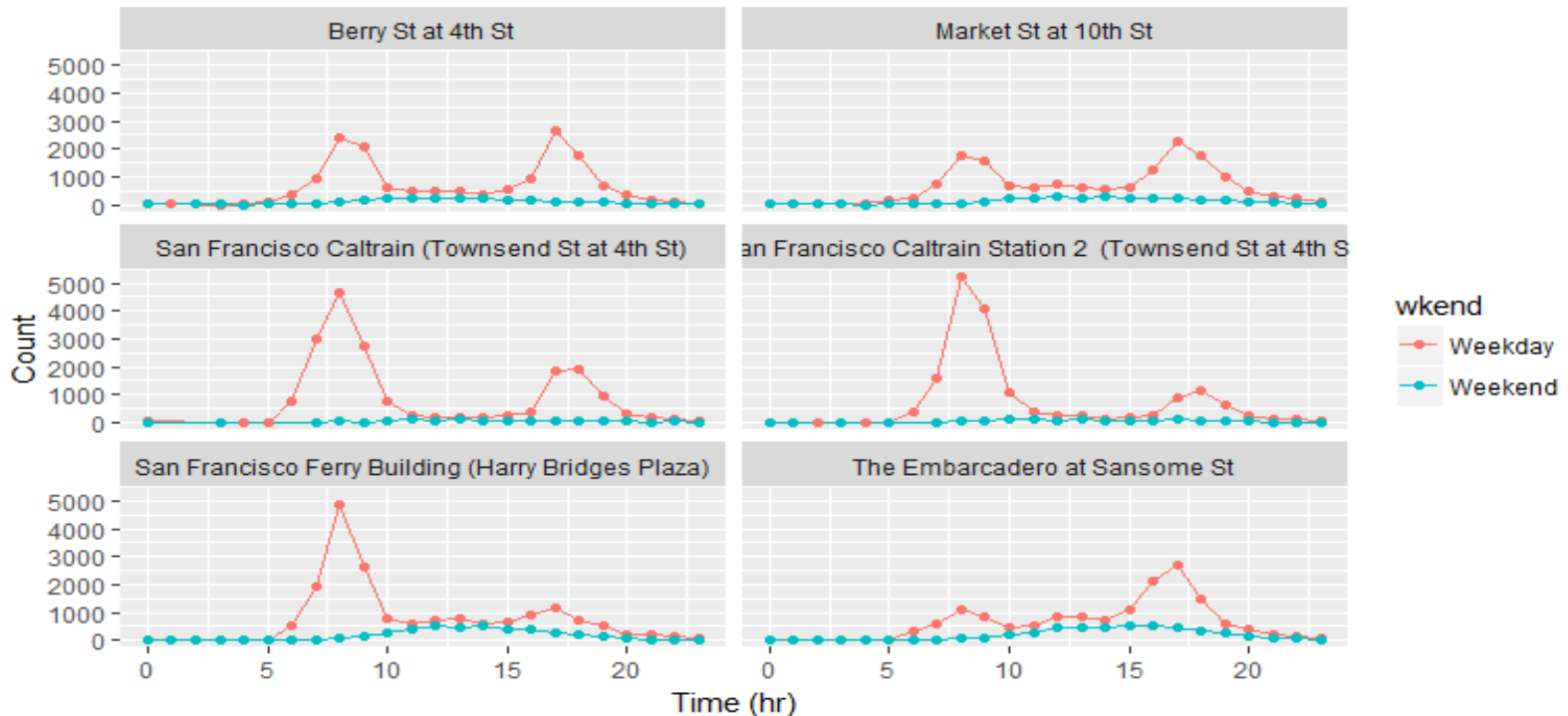
## Bay Area Bike Share Data Analysis

(Vinod Varma)

From the above popular stations info, it appears that top five start stations are same as top end stations. After a bit of browsing all of those are in side San Francisco city, which is a good sign.

To go deep in this popular stations rides condition, I did plot the distribution of rides over the time in these six stations as shown below.

**Distribution of trips starting from each station across time by weekday/weekend**



With the above flow of data points we can notice that each of these stations have very different distribution of rides across time and day. From here, we could in fact see that San Francisco Caltrain (Townsend at 4th), San Francisco Caltrain 2 (330 Townsend) and Harry Bridges Plaza are places where people either go home to or to take on another form of transport (most likely the commuter train) in the evening to some-where else (high number of departures in the evening)! On the other end, the other 3 stations are places where many people are most likely to work at, hence the high number of arrivals in the morning and high number of departures in the evening.

Hence, we can then see how several factors such as time of the day, day of the week, date, and subscriber type can then affect the general demand and supply across the stations.