

March 2020 Ford GoBike Data Exploration

-by Neha Kale

Introduction to the Dataset:

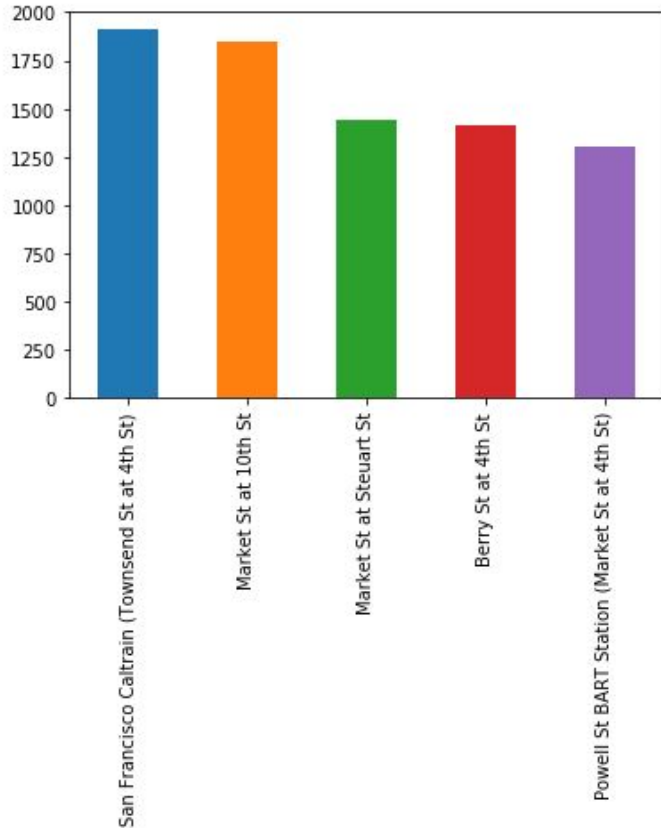
It is a public bicycle sharing platform in San Francisco. The bicycles are available for the people at any point of time for easy and reasonable transportation. This data consists of the information of each and every ride taken by the citizens such as duration of the ride, start point, end point, date, time, etc.

The dataset contains information for all the rides in March 2020. The dataset has 176799 rows and 14 columns.

The main questions for analysis:

1. Most common start stations
2. Most common end stations
3. The average duration of riding
4. The average distance travelled
5. Maximum distance travelled
6. Minimum distance travelled

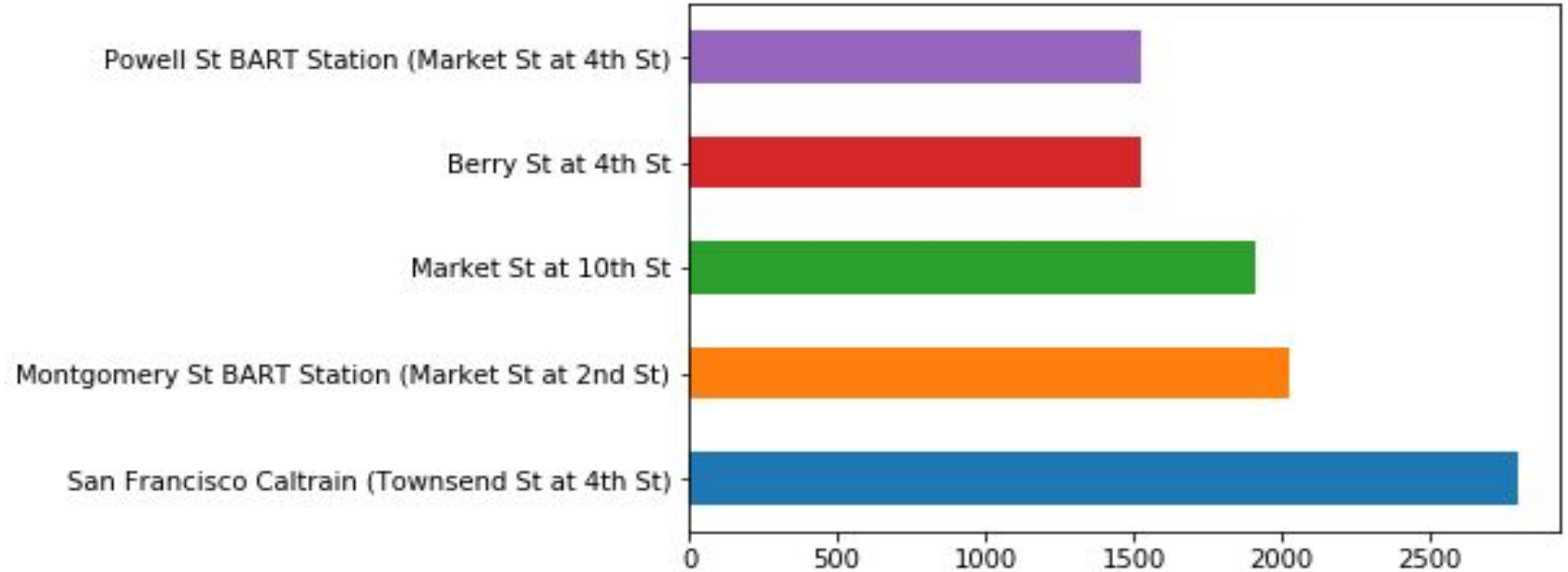
Which are the most common start stations?



From the visualization it is clear that the most common start station is San Francisco Caltrain.

It has around 2000 pickup record.

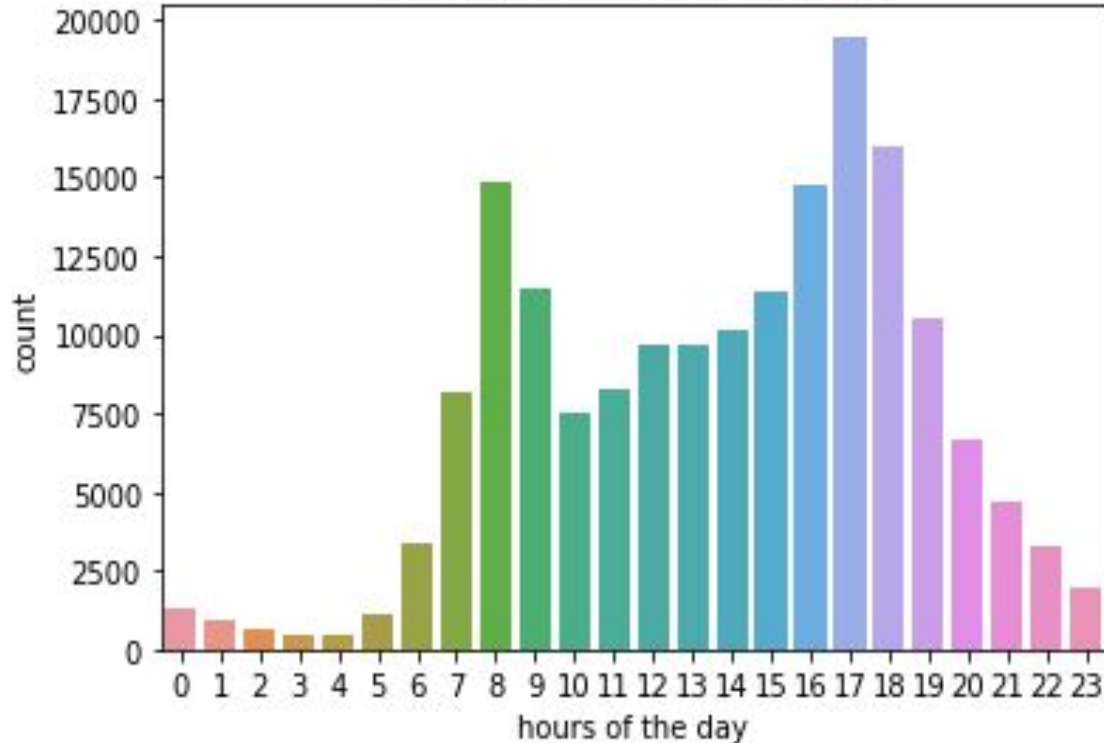
Which are the most common end stations?



Also, the most common end station is San Francisco Caltrain. It has the count above 2500.

Which are the most busy hours of the day?

Most active hours of the day



Looking at the countplot, we can see that the most busy/active hours of the day are 5pm and 8am as well.

Is the subscriber count more or the customer count?

For this question, as we have only 2 types of users, I am using a pie chart to get an idea about the same.

Here, from the pie chart we can see that the number of Subscribers is more than the number of Customers.

