Ford Bike Rental- Data Analysis

By- Utkarsh Upadhyay

Investigation Overview

———

In this investigation, I wanted to look at the characteristics of trips that could be used to predict their prices, most traffic bearing station and high traffic hours. The main focus was on the attributes such as : duration_sec, tart_time, end_time, user_type, start_station_name and end_station_namefour.

Dataset Overview

———

The data consisted of prices and attributes of approximately 183,412 trips.

The attributes included the characterstics such as : duration_sec, tart_time, end_time, user_type, start_station_name and end_station_namefour.

Eight thousand four hundred and sixty two data points were removed from the analysis due to inconsistencies or missing information.
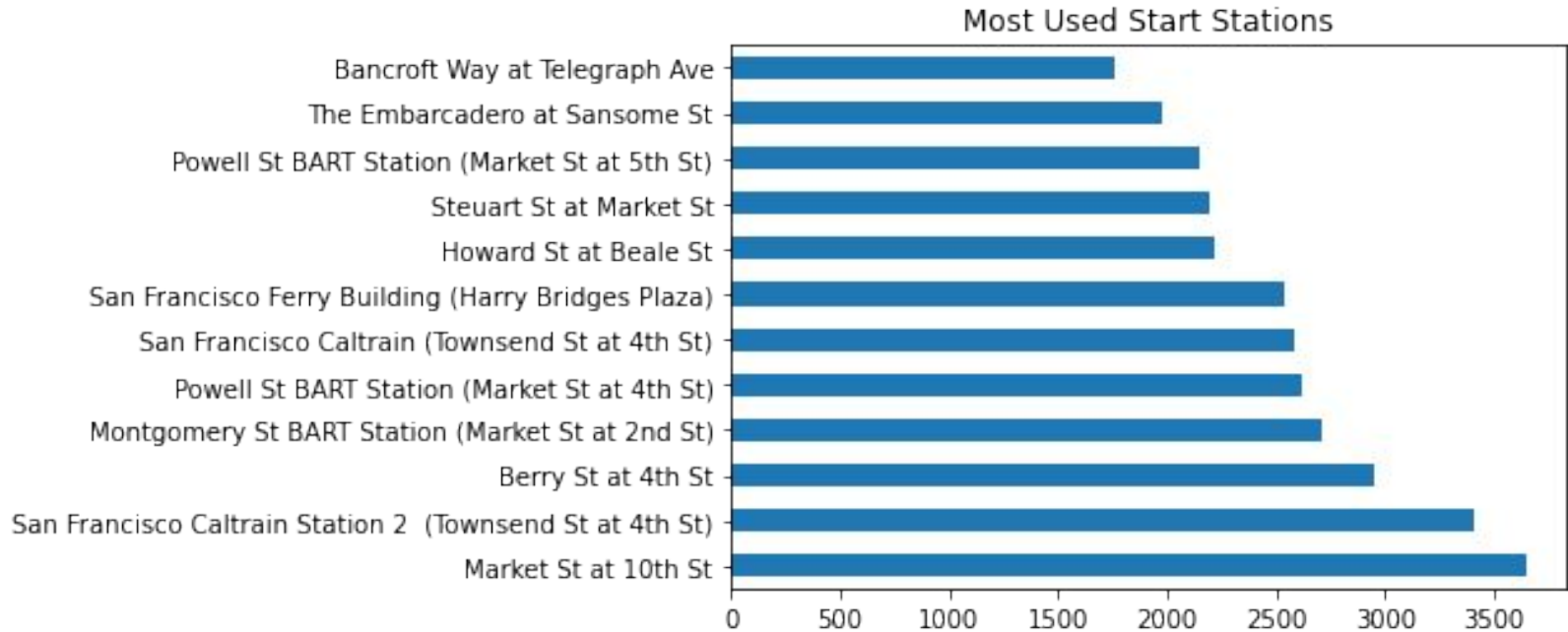
The Engineered dataset can be found in the repository : https://github.com/Arshofficial/FordBikes_Udacity
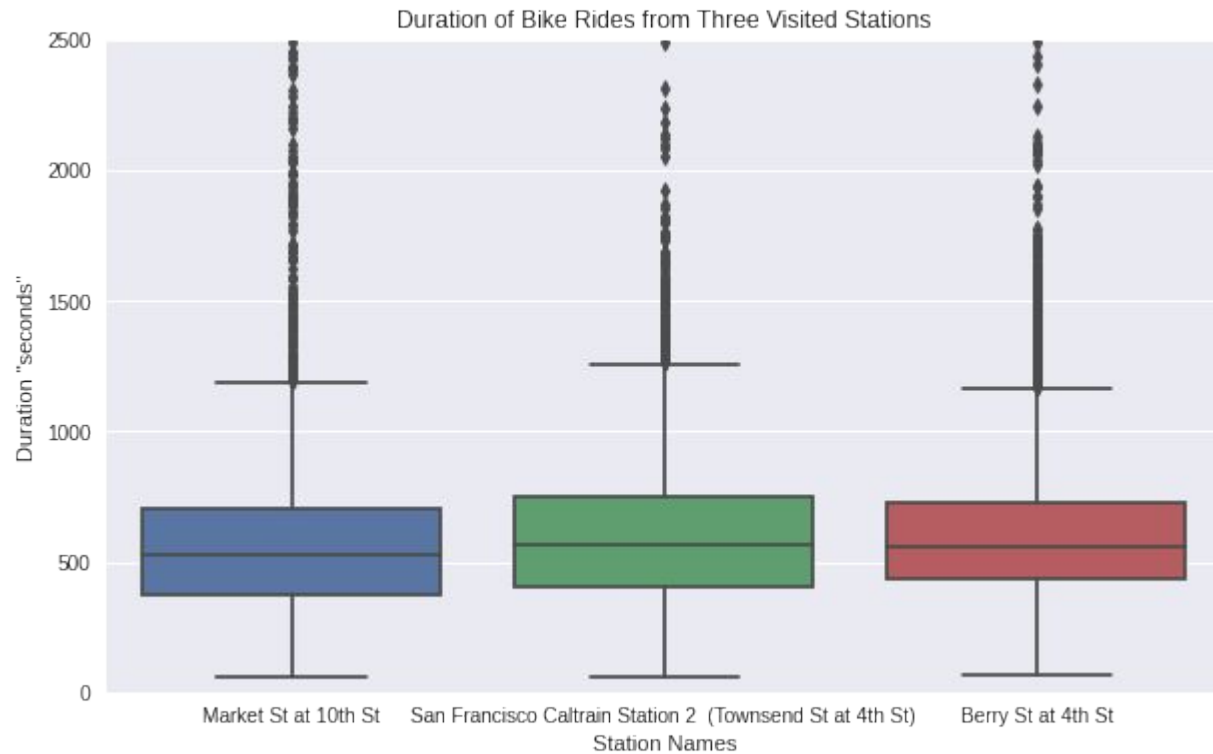
Distribution Of users : Customers/Subscribers
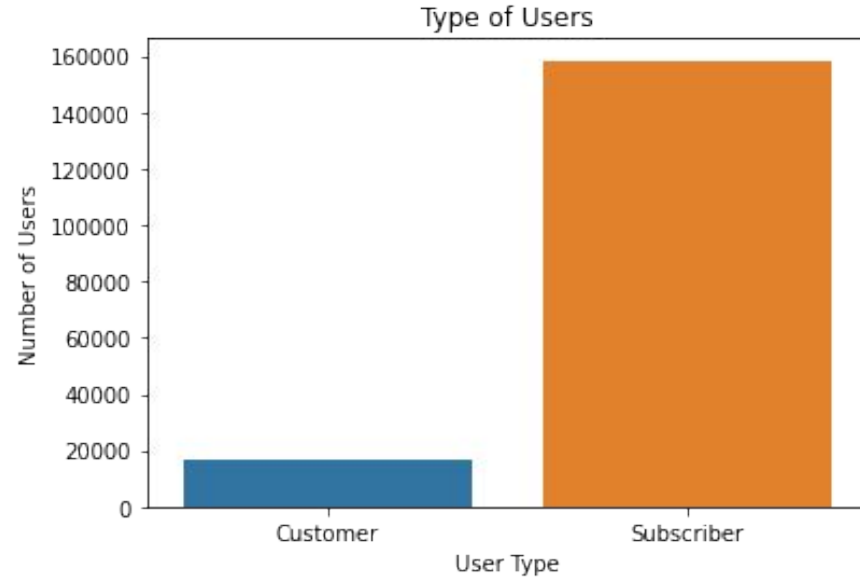
Bike Rides in the dataset take on a very large range of trips, from about 183,412 trips. Plotted on a boxplot, the distribution of diamond prices takes on ratio of users between the two categories. This illustrates that significant number of users are Subscribers rather than normal customers.

Distribution Of Journey Origination Frequency of All stations.

Most Used Start Stations

Station	Frequency (approx.)
Bancroft Way at Telegraph Ave	~1750
The Embarcadero at Sansome St	~1950
Powell St BART Station (Market St at 5th St)	~2150
Steuart St at Market St	~2200
Howard St at Beale St	~2250
San Francisco Ferry Building (Harry Bridges Plaza)	~2500
San Francisco Caltrain (Townsend St at 4th St)	~2550
Powell St BART Station (Market St at 4th St)	~2600
Montgomery St BART Station (Market St at 2nd St)	~2700
Berry St at 4th St	~2950
San Francisco Caltrain Station 2 (Townsend St at 4th St)	~3400
Market St at 10th St	~3650

Duration of Bike Rides from Three Visited Stations

Major Distribution Of Trip in the given dataset originate at 'Market St. at 10th St.' followed by: 'San francisco Caltrain Station 2' and 'Berry St. at 4th St.'
Below bargraph ilustrates the above conclusion

Distribution Of Journey Origination Frequency of All stations.

Most Used End Stations

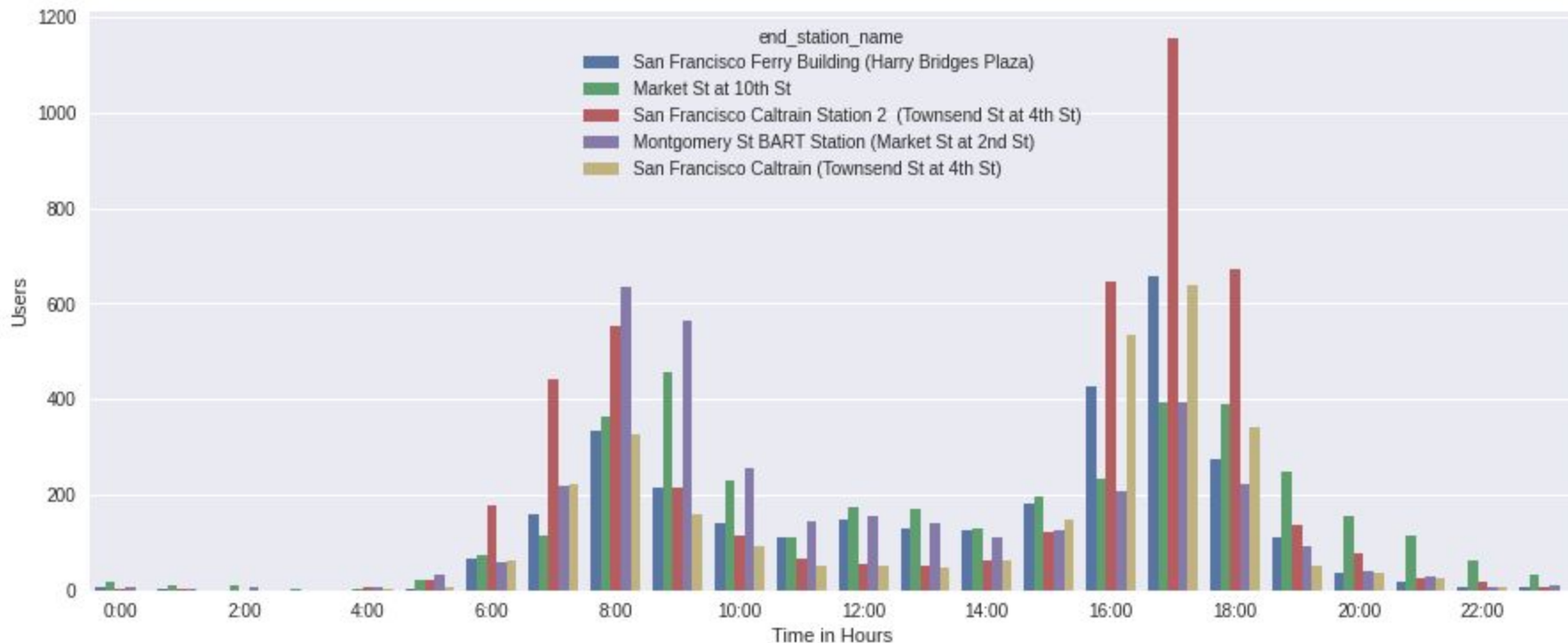Duration of Bike Rides from Three Most Concluded Stations
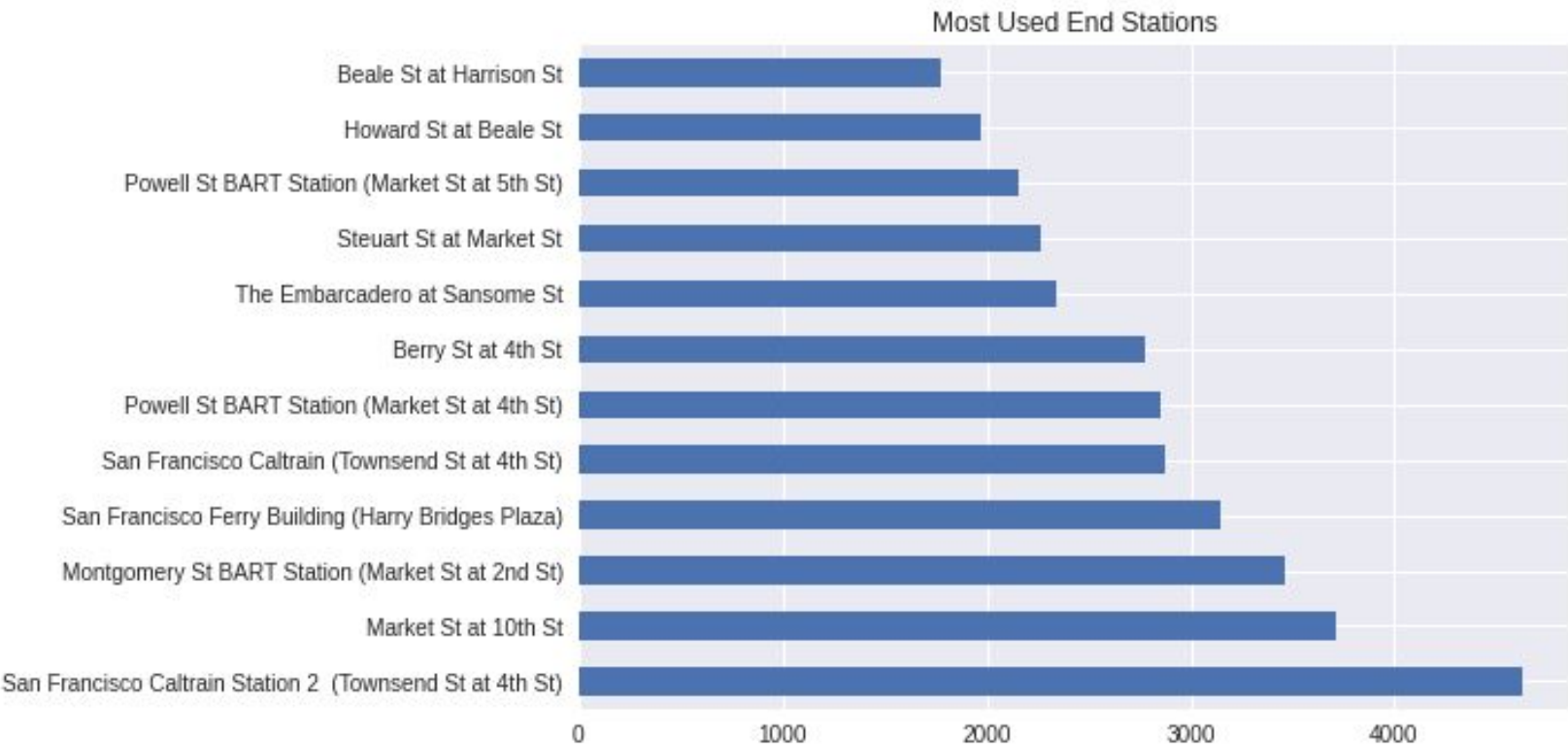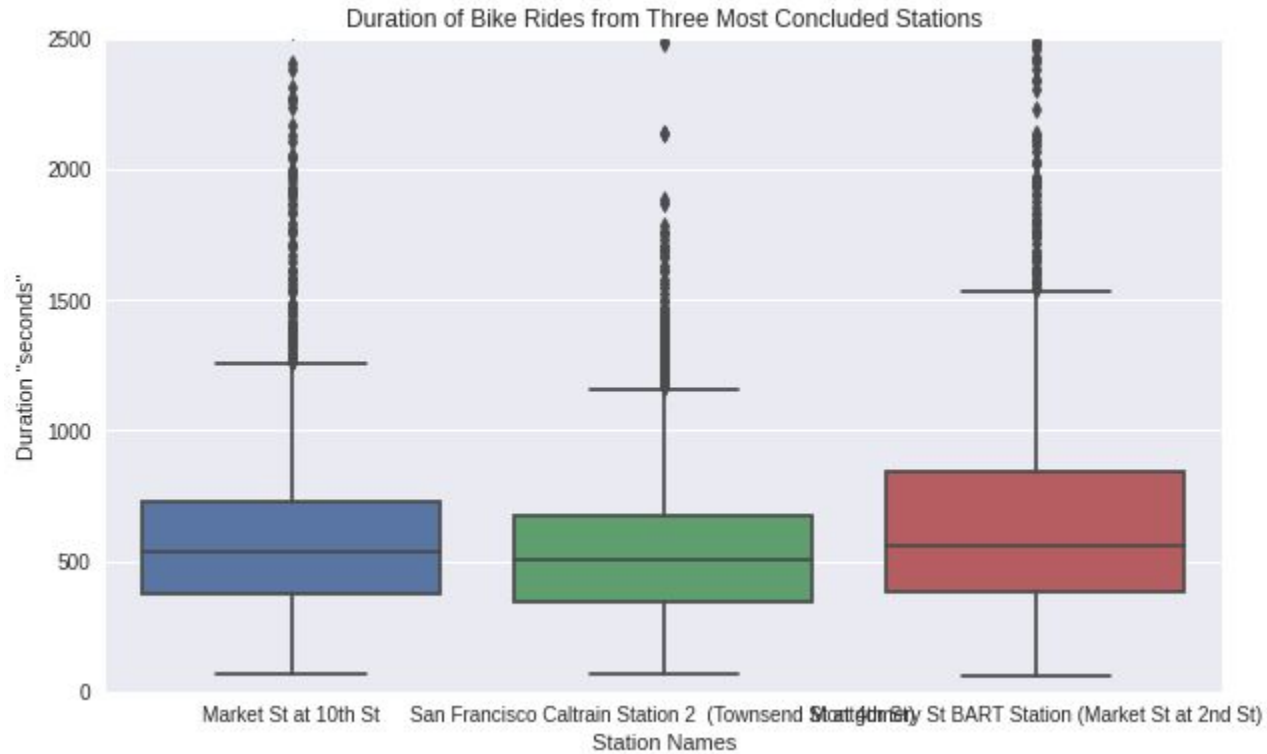
Majority of Trip in the given dataset conclude at 'San francisco Caltrain Station 2'
followed by: 'Market St. at 10th St.' and 'Montgomery St. BART St.'.
Bargraph and Boxplot illustrates the above conclusion.

Distribution Of Passengers over 24Hrs.
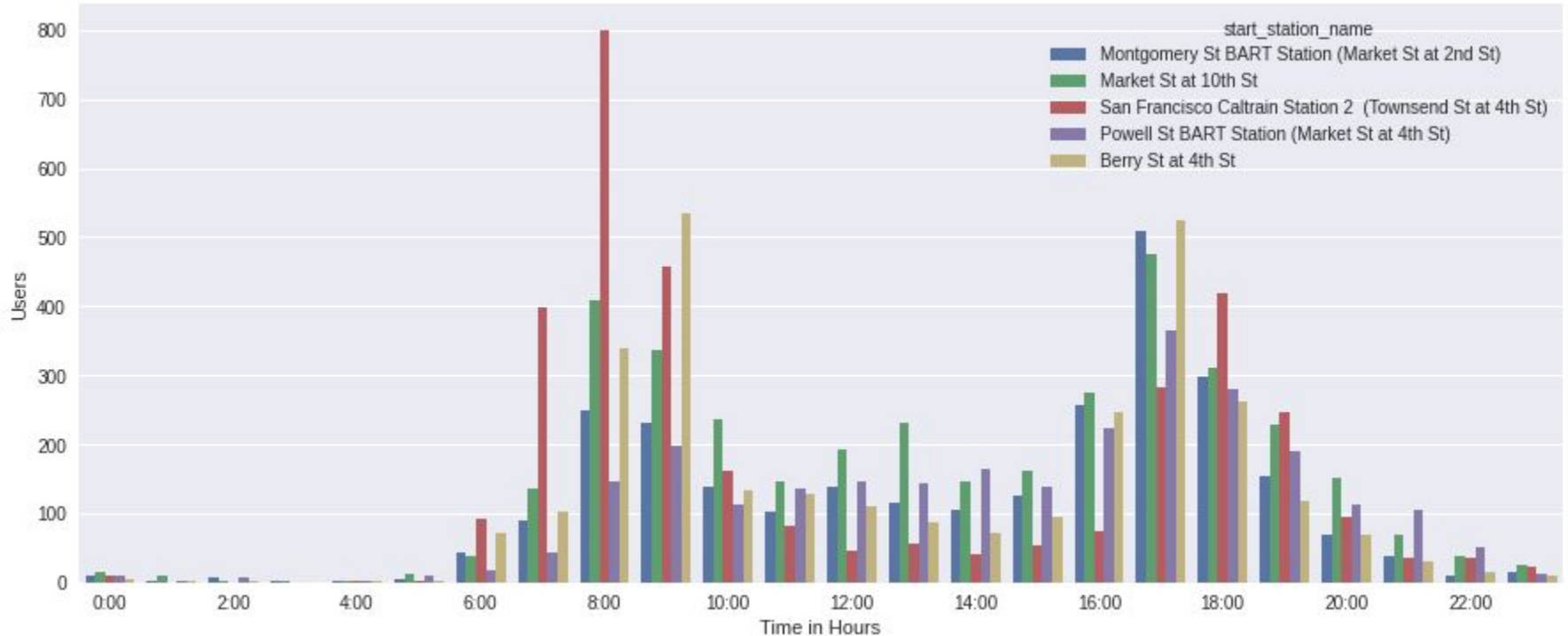
Graphs suggest, bigger number of users prefer ride bike in the morning at 8. On the other hand, the users end the trip usually at 5 pm.