

EDA Project



METHODOLOGY



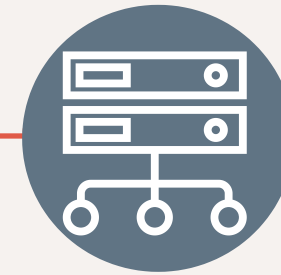
**Problem
understanding**



**Data
integration**



**Data
preprocessing**



**Feature
engineering**



**Exploratory data
analysis (EDA)**

PROBLEM UNDERSTANDING



Overview: tss is an advertising company responsible for the billboard in train stations. Its main problem is to determine the price of advertisement at each train station to increase the income of the advertising company and the advertiser. In this analysis we used the New York subway MTA turnstile data .

DATA INTEGRATION



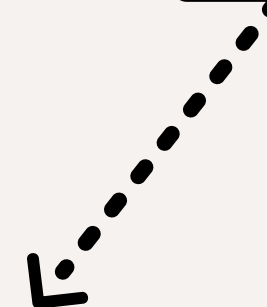
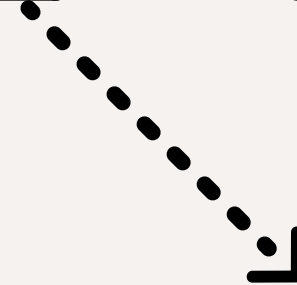
four months of
2021

March

April

june

Augast



3347283 rows,
11 columns

DATA PREPROCESSING



- Drop duplicates
- Check nulls
- Check outliers
- Replace some special character such as '/' with '-'

FEATURE ENGINEERING



Month

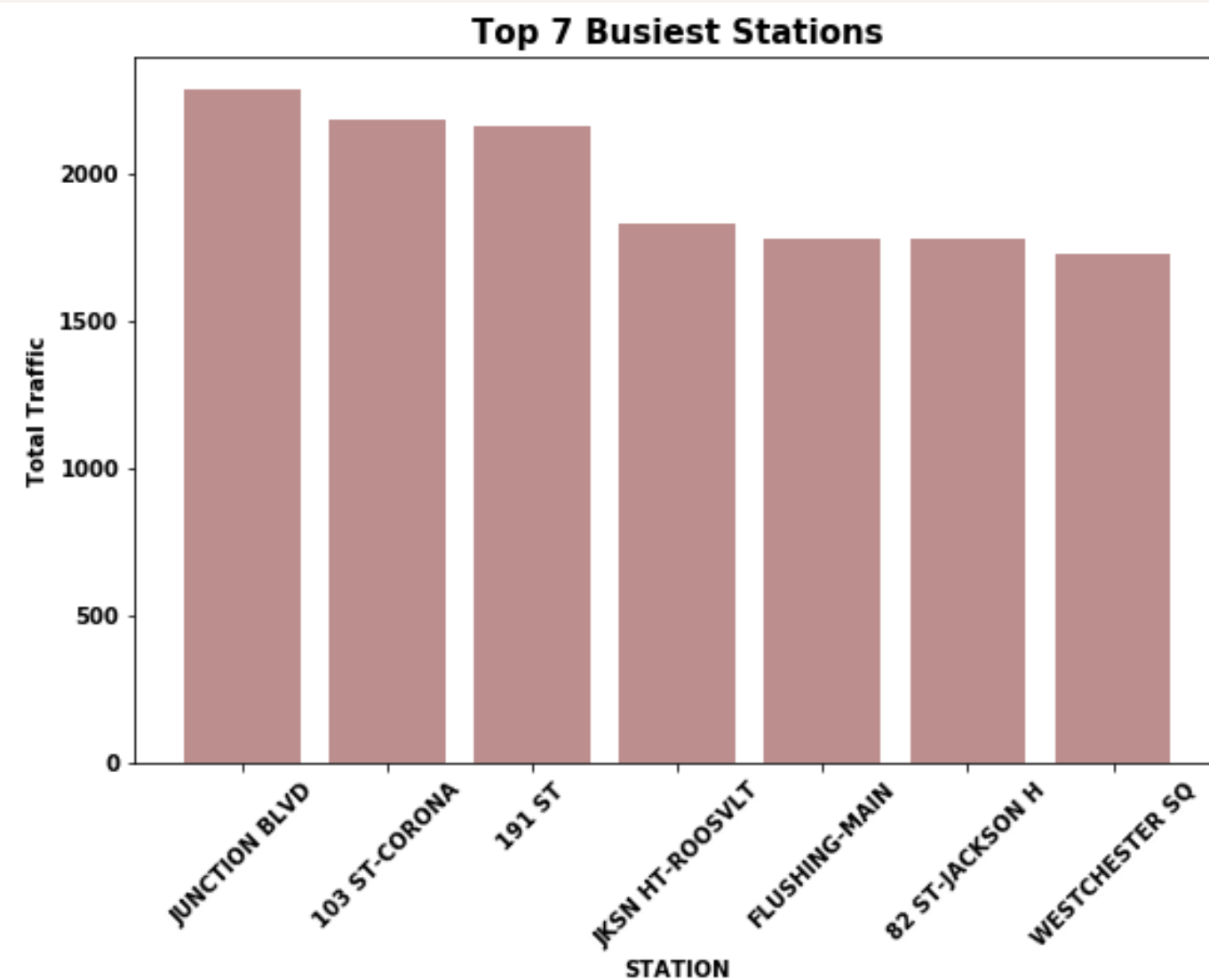
Turnstile

DateTime

Name of day

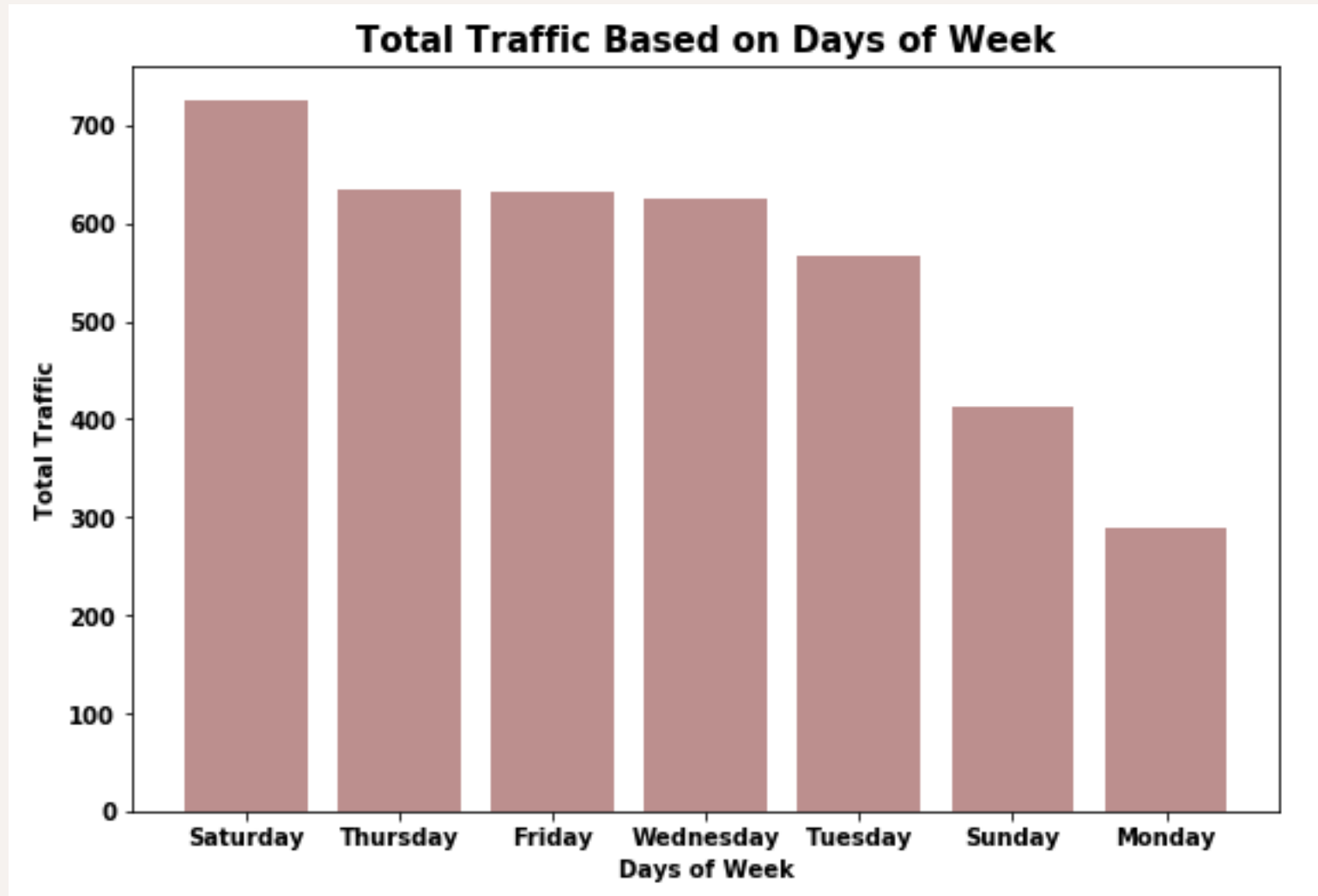
Traffic

EXPLORATORY DATA ANALYSIS (EDA)



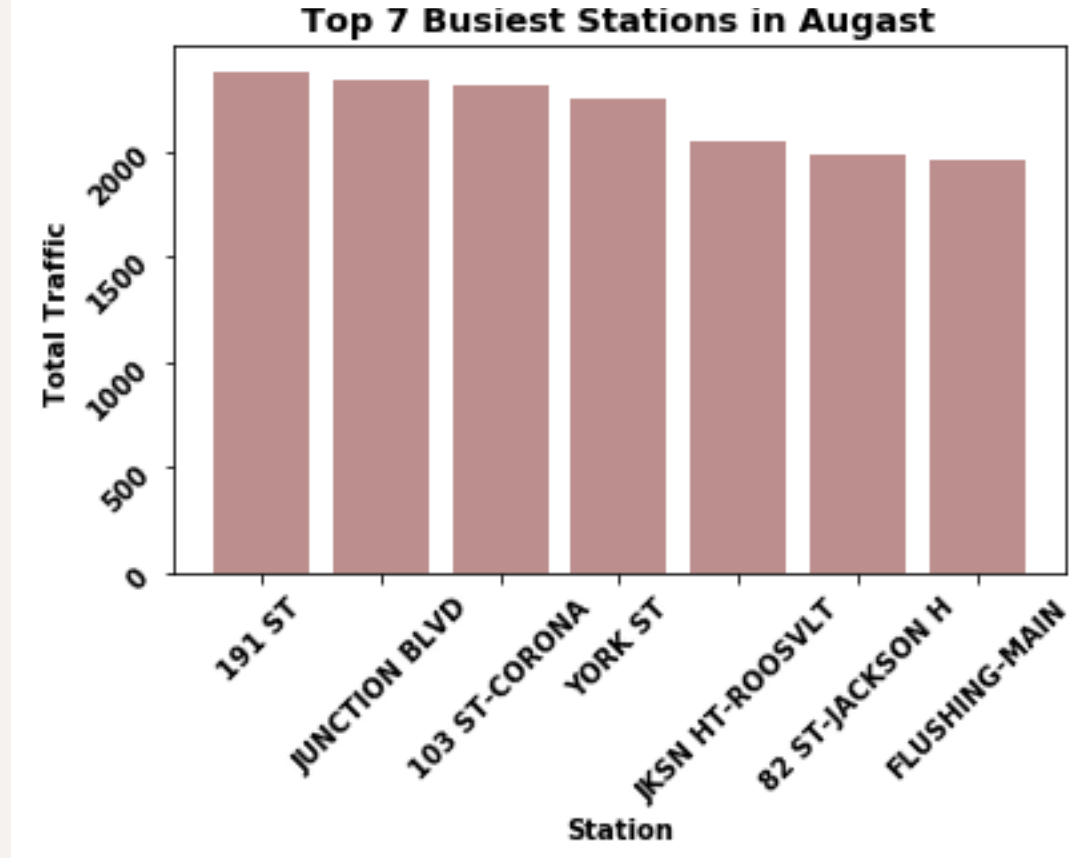
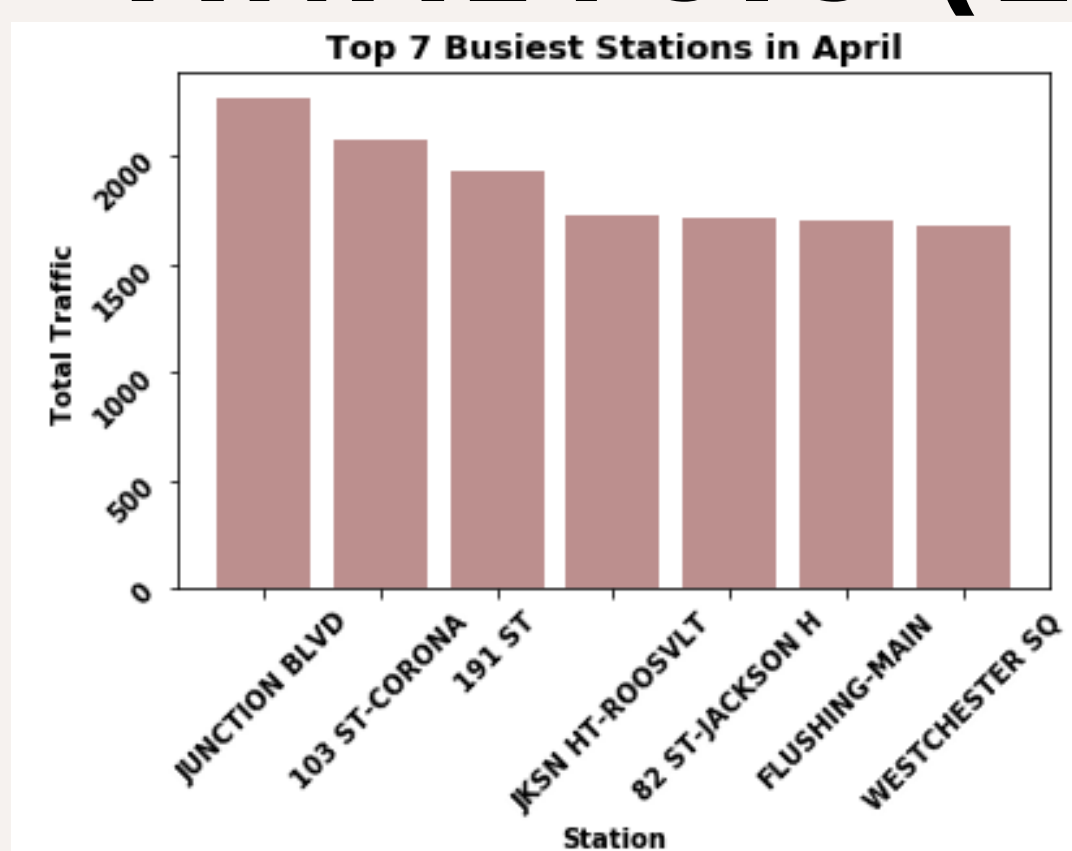
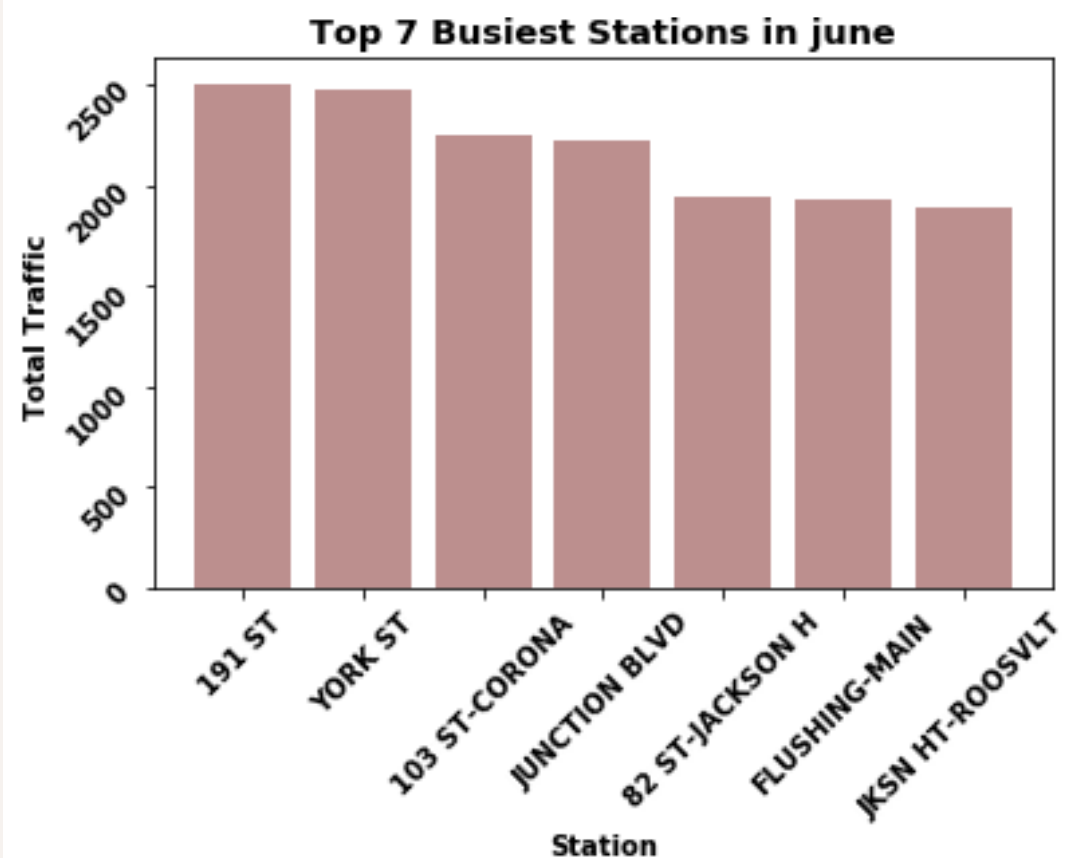
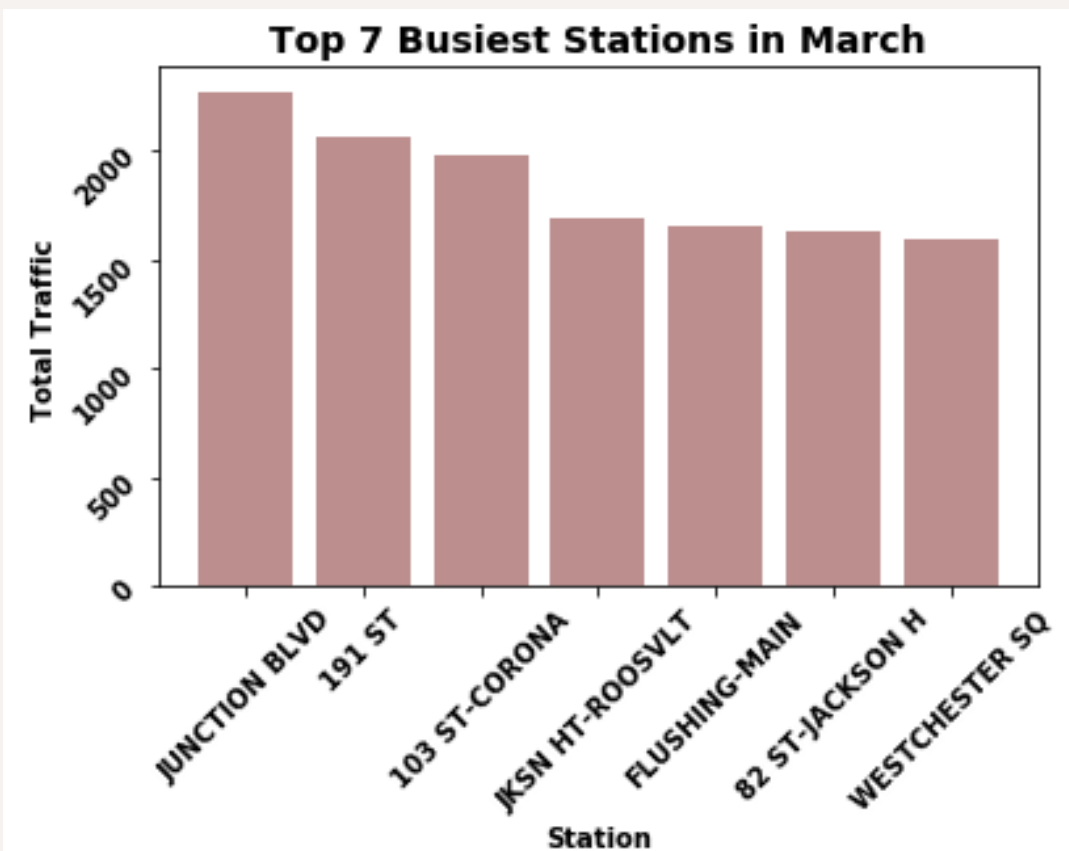
Observations: Through this Bar plot, it becomes clear to us that the (Junction Blvd) station is the most crowded station during the four months .

EXPLORATORY DATA ANALYSIS (EDA)



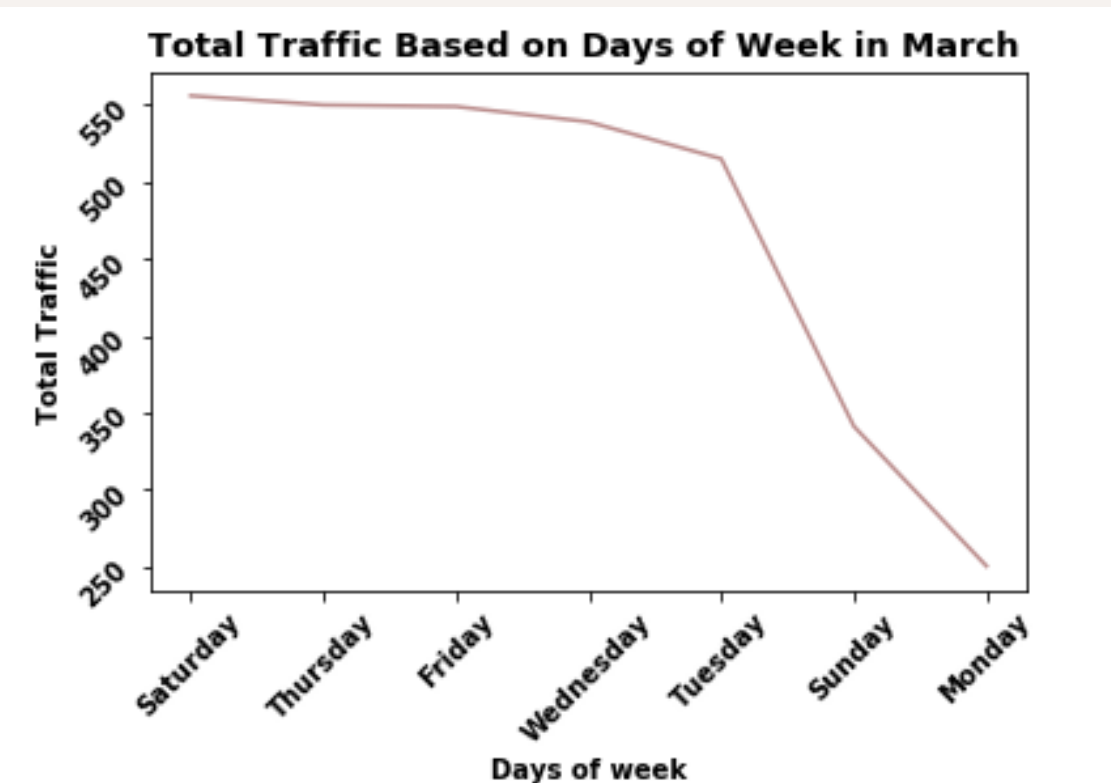
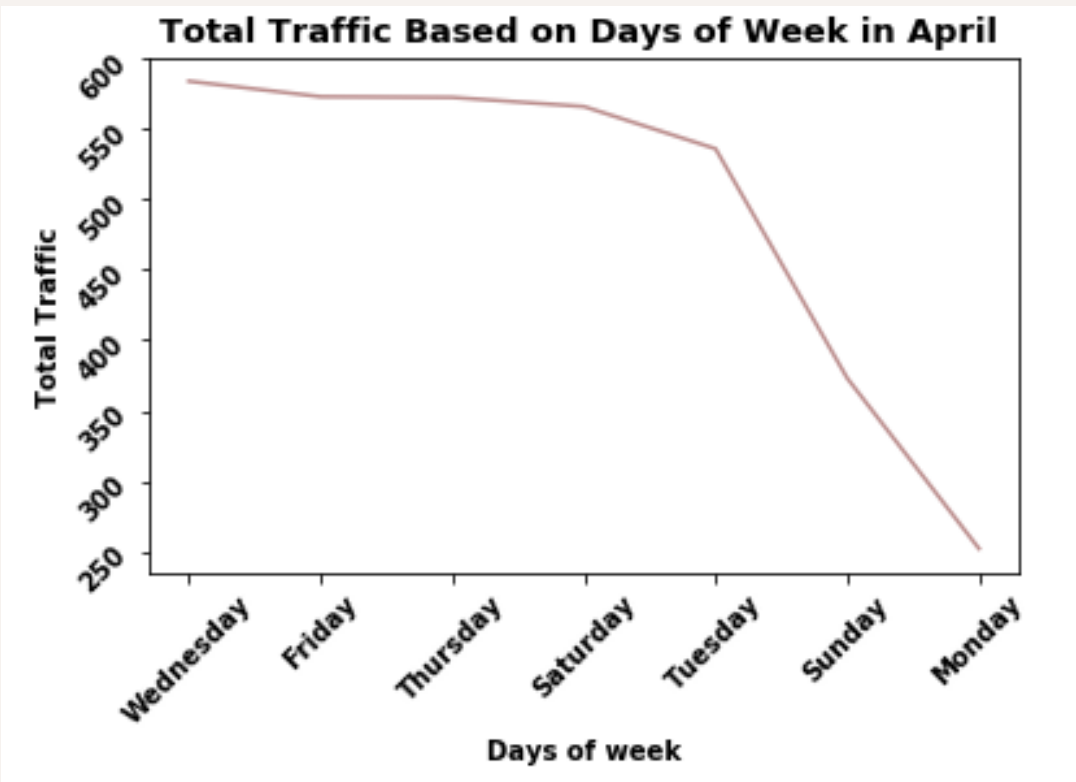
Observations: The bar blot shows the traffic according to the days of the week, and it turns out that Saturday is the most traffic day of the week and Monday is the least.

EXPLORATORY DATA ANALYSIS (EDA)

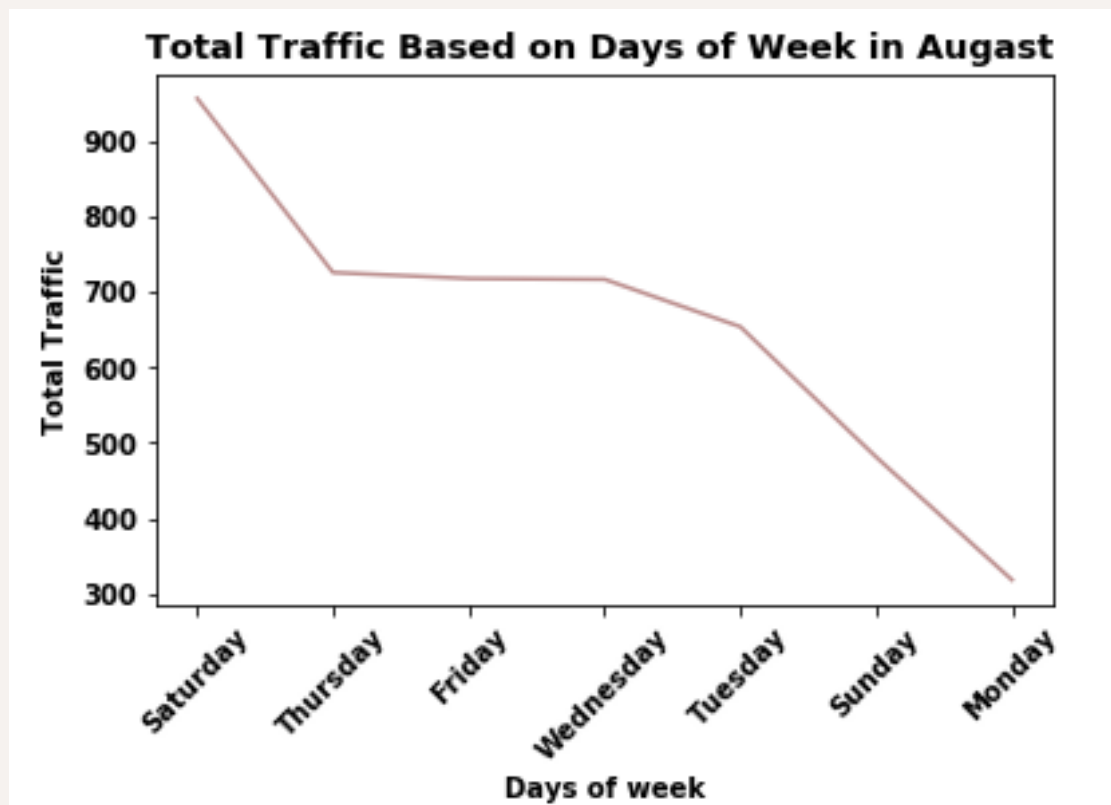


Observations: All four months share the names stations of the most traffic.

EXPLORATORY DATA ANALYSIS (EDA)




The line chart represents the relationship between traffic and days.



Observations: The most of the months agreed on the same results that Monday is the least traffic and Saturday is the highest with different numbers



CONCLUSION

- Monday is the least traffic and Saturday is the highest
 - Junction Blvd) station is the most crowded station during the four months .
- 



THANK YOU