Fatimah Alghamdi – RTA

Abstract

The goal of this project is to conduct the exploratory data analysis (EDA) on MTA turnstile data to assist RTA company in finding the stations with smooth traffic around New York State popular tourists areas. The results will help the company to minimize the cost of hiring drivers and rental cars. However, The Empire State Building, Central Park, Little Italy, Times Square, New York Public Library, The Statue of Liberty and Fifth Avenue are examples of popular areas.

Design

To achieve the aim of this project, a list of places that RTA's consumers visit was provided. Then 3 places were selected randomly to find the proper transit lines. These places are Empire State Building, Central Park and Times Square. Based on Google Maps, the stations that have routes and pass near them were determined.

Data

The MTA turnstile data was scraped from the last week of April till the end of July. It's in shape (2930792, 12).

Algorithms

1- Conduct exploratory data analysis.

2- Create new columns that will assist in achieving the aim of the project such as [EX Diff] column and [Name_Day].

3- Study the MTA stations first, then focus on 5 specific stations which are:

['49 ST','34 ST-HERALD SQ','86 ST' , 'TIMES SQ-42 ST', '42 ST-BRYANT PK', '34 ST-HERALD SQ'].
The subway lines that pass these stations are:
[1, 2, 3, 4,5, 6, 7, A, B, C, D, E, F, M, N,Q, R, W,X].

4- Present the analyzed data in graphs.

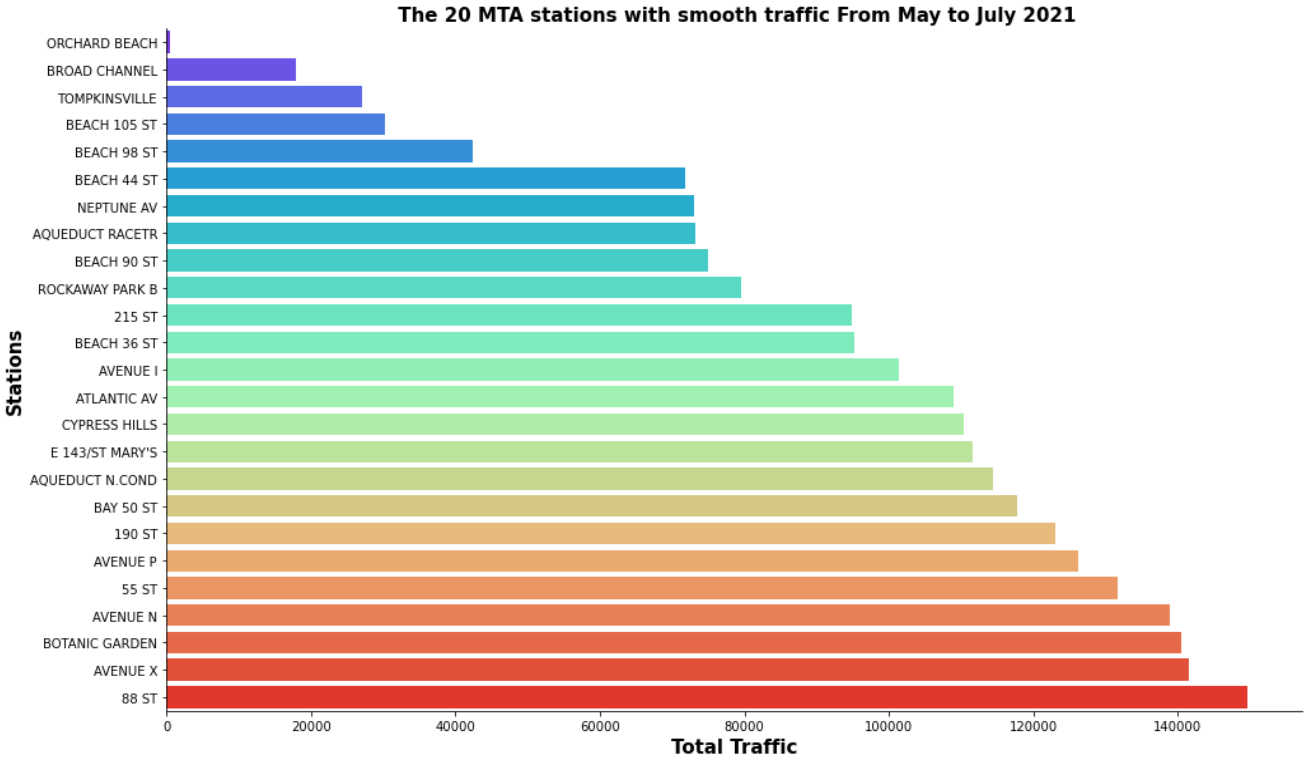
Tools

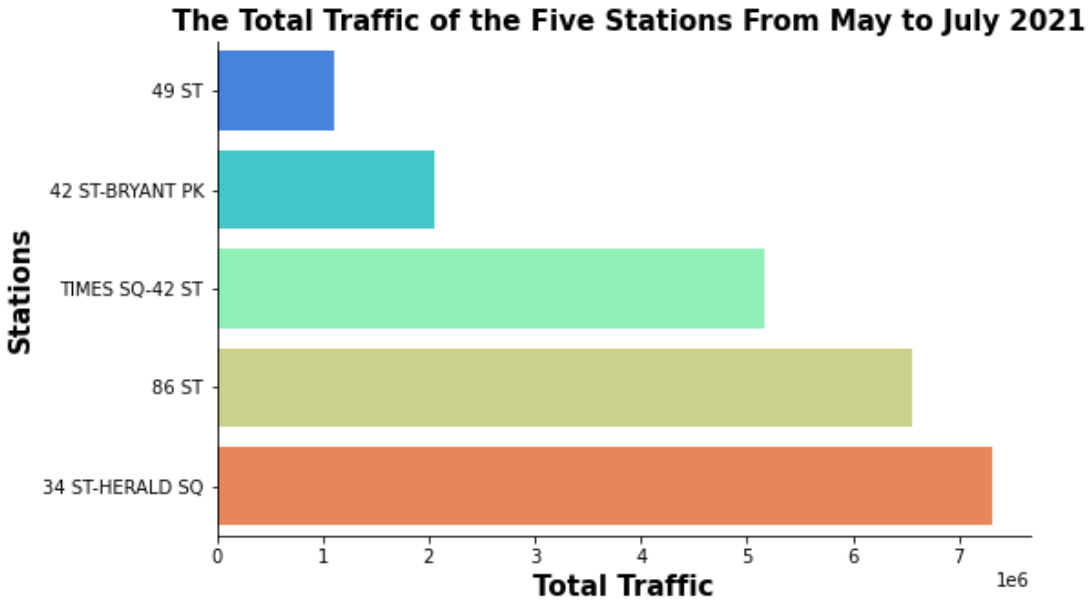
1- Numpy and Pandas for data manipulation

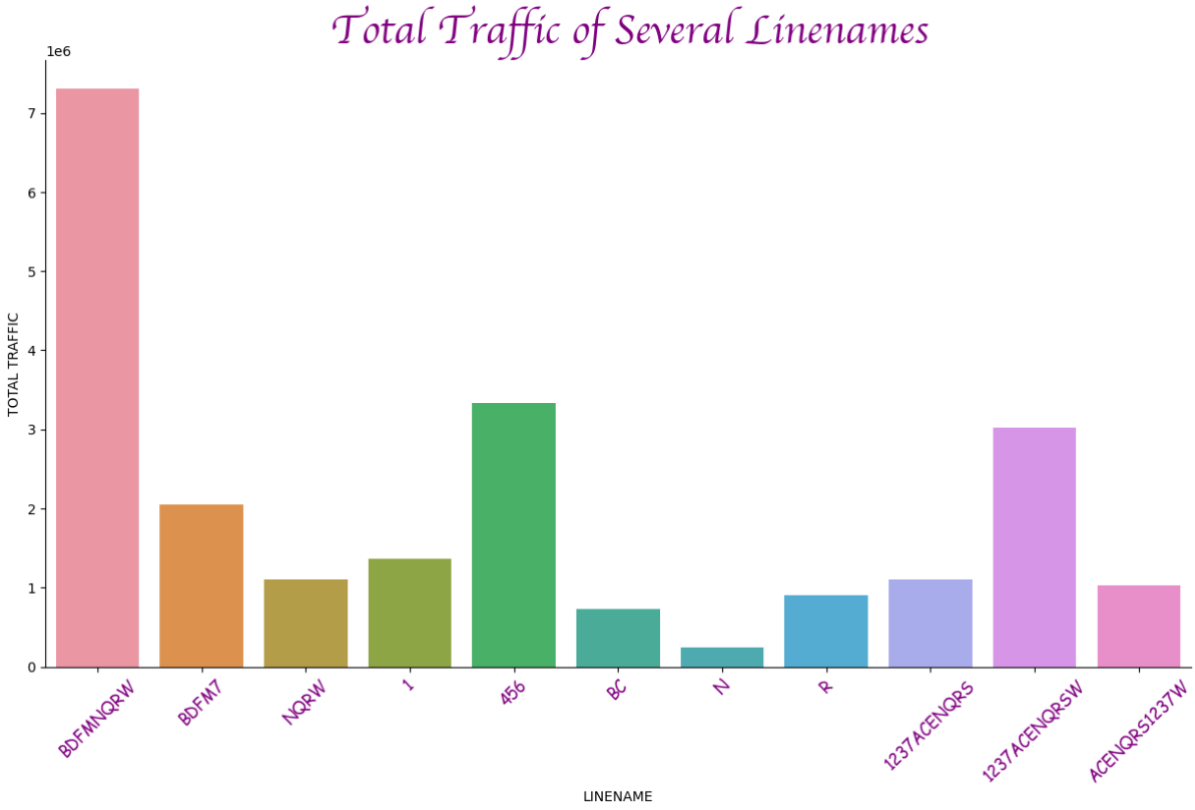
2- Matplotlib and Seaborn for plotting.

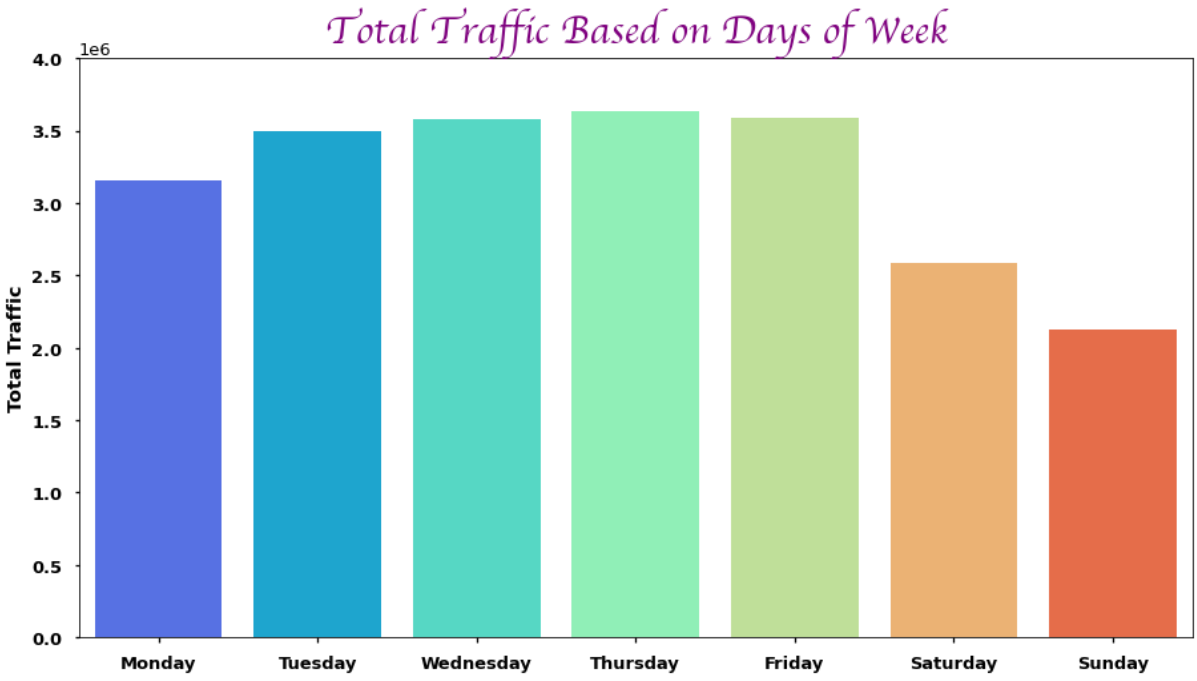
Communication

This graph shows the 20 stations of MTA with minimum total traffic from May to July 2021.






Total Traffic of Several Linenames
1e6
TOTAL TRAFFIC
LINENAME
BDFMNQRW
BDFM7
NQRW
1
456
BC
N
R
1237ACENQRS
1237ACENQRSW
ACENQRS1237W


Total Traffic Based on Days of Week
1e6
Total Traffic
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday