SDAIA BOOTCAMP PROJECT 1 -EDA

By Taef alkhales

Database

Step 1

Create a server on AWS RDS

Step 2

Connect it with Postgres

Step 4

Create an connection with SQLAlchemy in python

Step 3

Import my dataset CSV file as a table in the database

Problem

MEMO-RICH a company that sells merch across the U.S wants to place five new trucks that sell souvenirs for the tourists in summer in front of the MTA subway stations, the company wants to optimize the profits by placing the five trucks in the stations that are on high demand in the Summer.

Dataset : extract the MTA turnstiles dataset from the date 19 June 2021 until 22 September 2021

Cleaning

Step 1 ------------------------- **Step 2**

Renaming columns   Removing duplicates rows

Step 4 ------------------------- **Step 3**

Removing outliers   Calculating the daily exits and entries
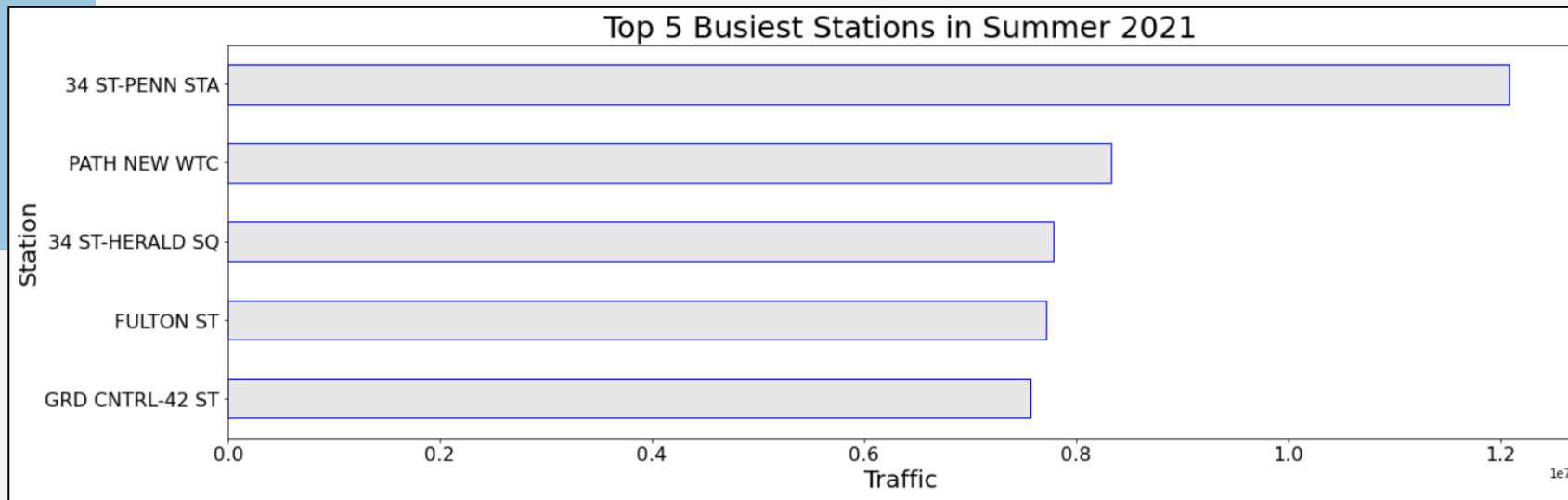Using Boxplot

Two Issues :
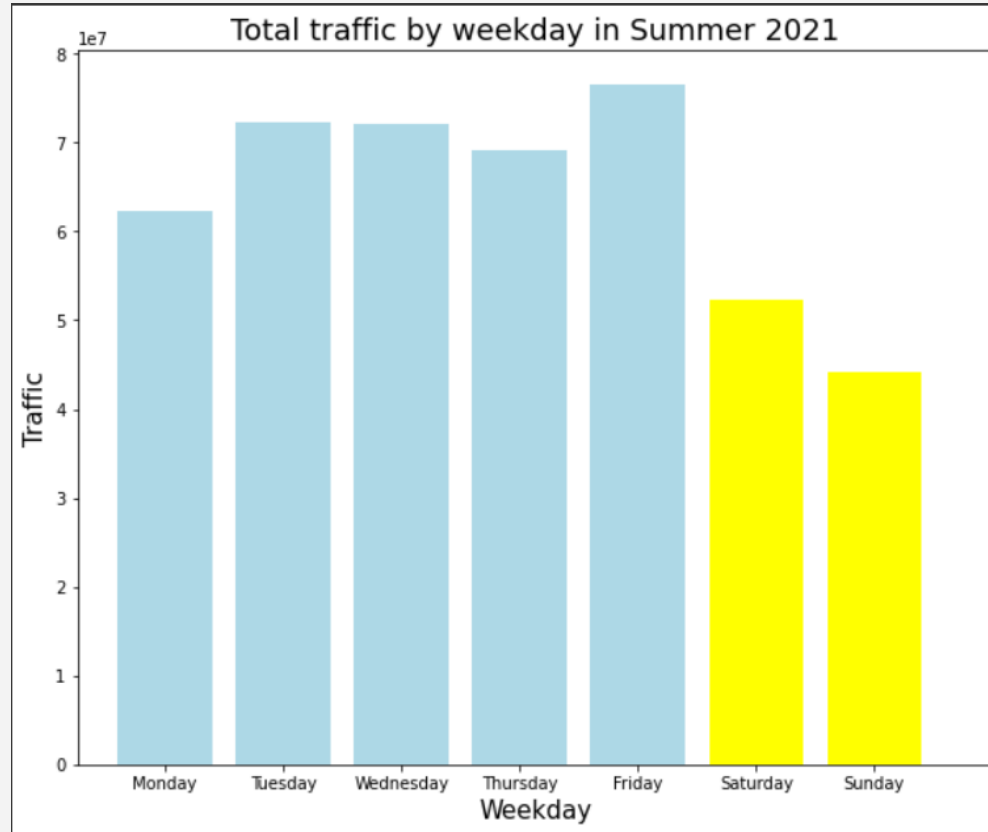
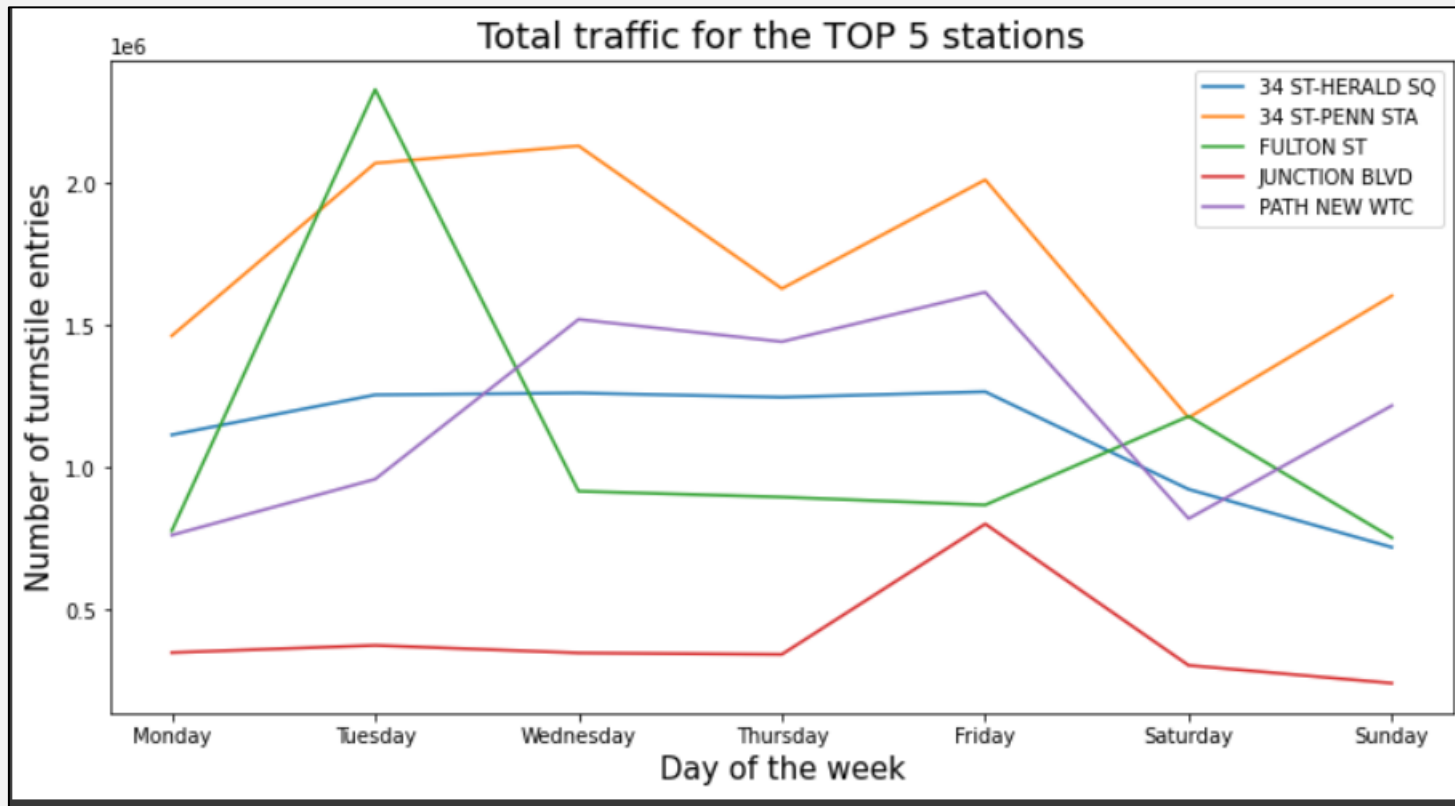1- Counter reversed (Solved)
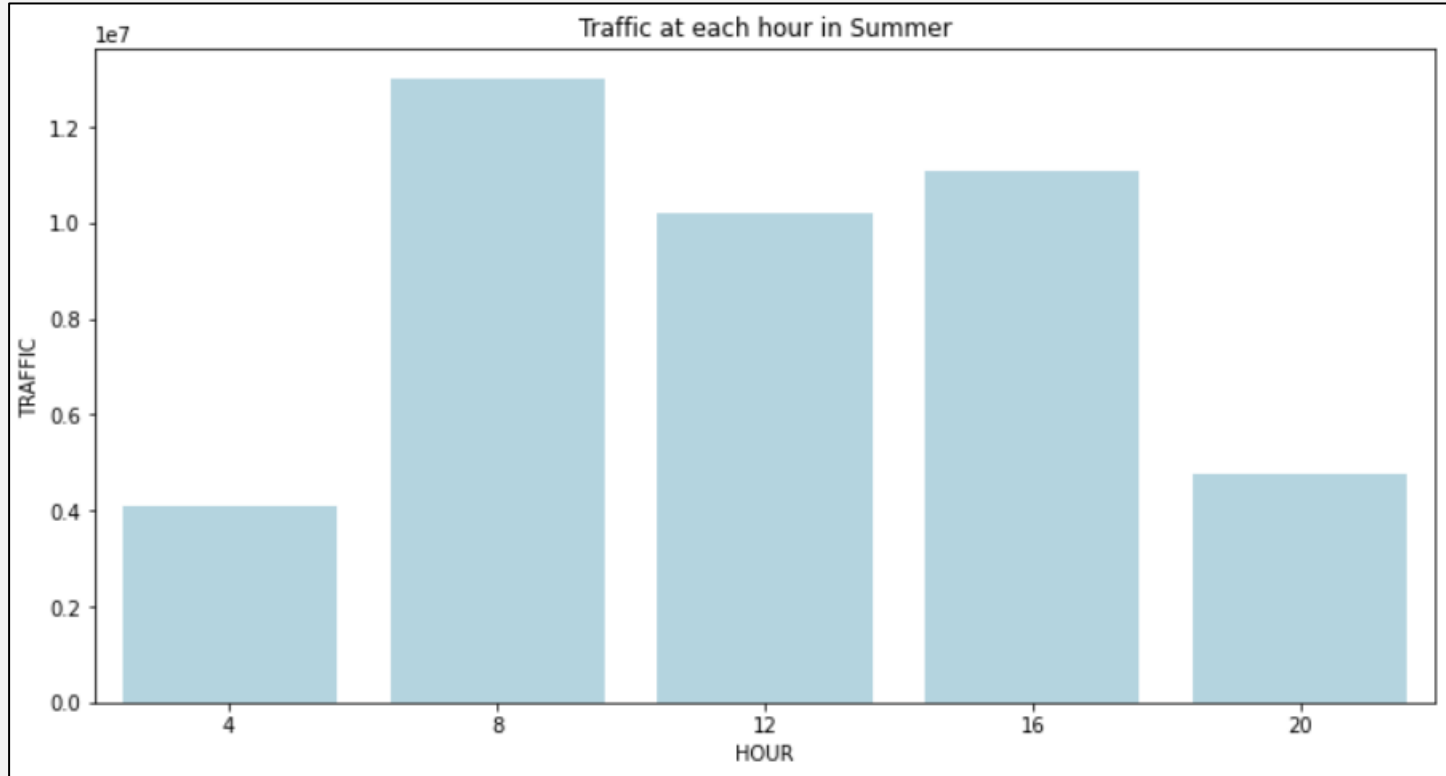
2- Counter reset (Solved)

EDA graphs

Top 5 Stations

Top 5 Busiest Stations in Summer 2021

Busiest weekdays

Total traffic by weekday in Summer 2021

Top 5 stations traffic by weekdays

Total traffic for the TOP 5 stations

Legend:
- 34 ST-HERALD SQ
- 34 ST-PENN STA
- FULTON ST
- JUNCTION BLVD
- PATH NEW WTC

Y-axis: Number of turnstile entries (1e6)

X-axis: Day of the week — Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday

Traffic per hour

Traffic at each hour in Summer

Traffic per week

Top 5 stations by the traffic for each week of Summer

Legend:
- Top1: 134 ST-PENN STA
- Top2: PATH NEW WTC
- Top3: 34 ST-HERALD SQ
- Top4: FULTON ST
- Top5: GRD CNTRL-42 ST

y-axis: traffic (1e6)
x-axis: week

Weeks:
June 3 - June 17, June 14 - June 20, June 21 - June 27, June 28 - July 4, July 5 - July 11, July 12 - July 18, July 19 - July 25, July 26 - Aug 1, Aug 2 - Aug 8, Aug 9 - Aug 15, Aug 16 - Aug 22, Aug 23 - Aug 29, Aug 30 - Sep 5, Sep 6 - Sep 12, Sep 13 - Sep 19
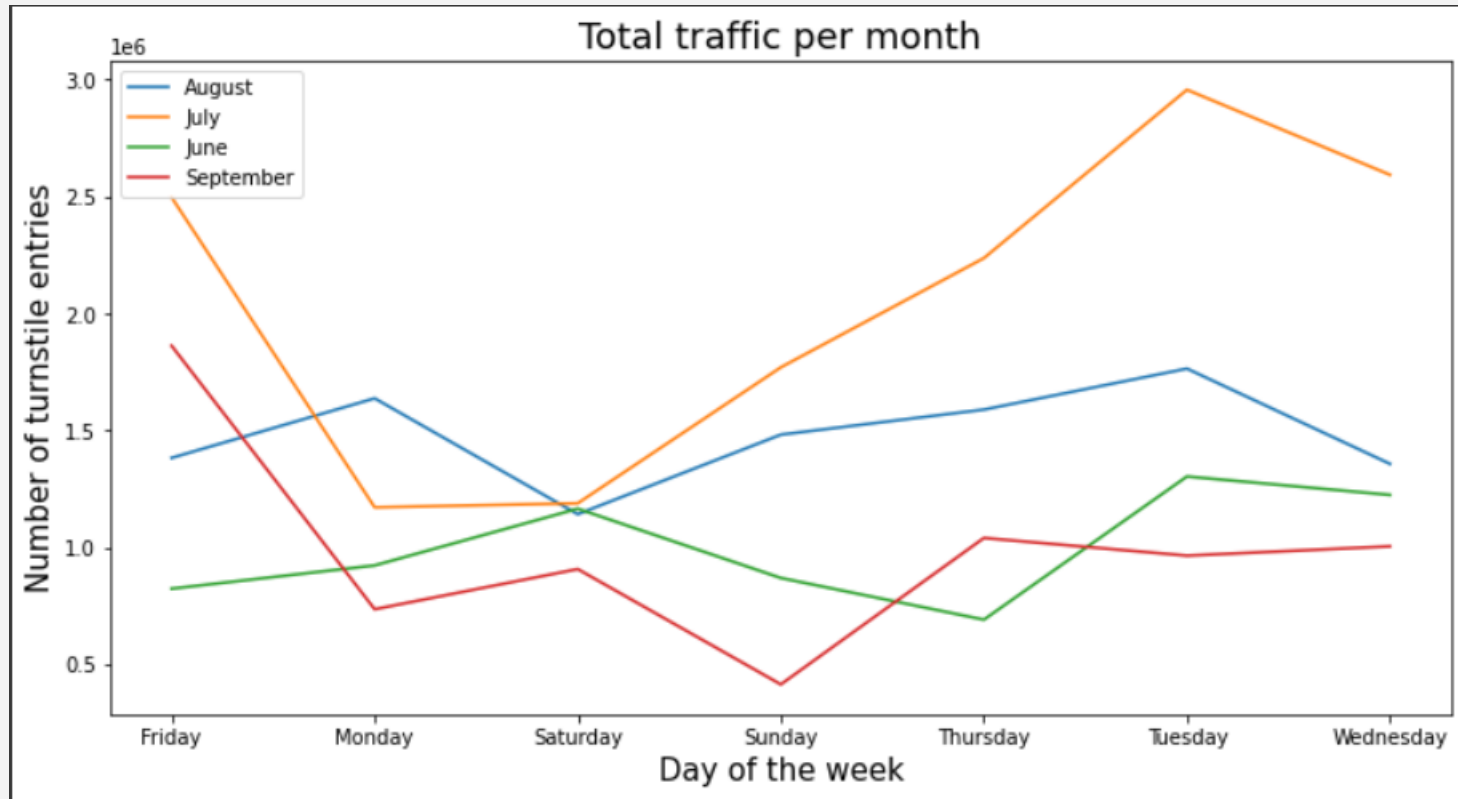
Traffic per month

Conclusion and Recommendation

- Focus on weekdays.

- Place the 5 trucks Infront of the Top 5 stations'

- Best opening times are 8 AM -4 PM

Thanks