



MTA

Souvenir shop

By Samaher AL-Hadi

AGENDA

Backstory

Goals

Tools

Preprocessing

Result

Future Work





Backstory

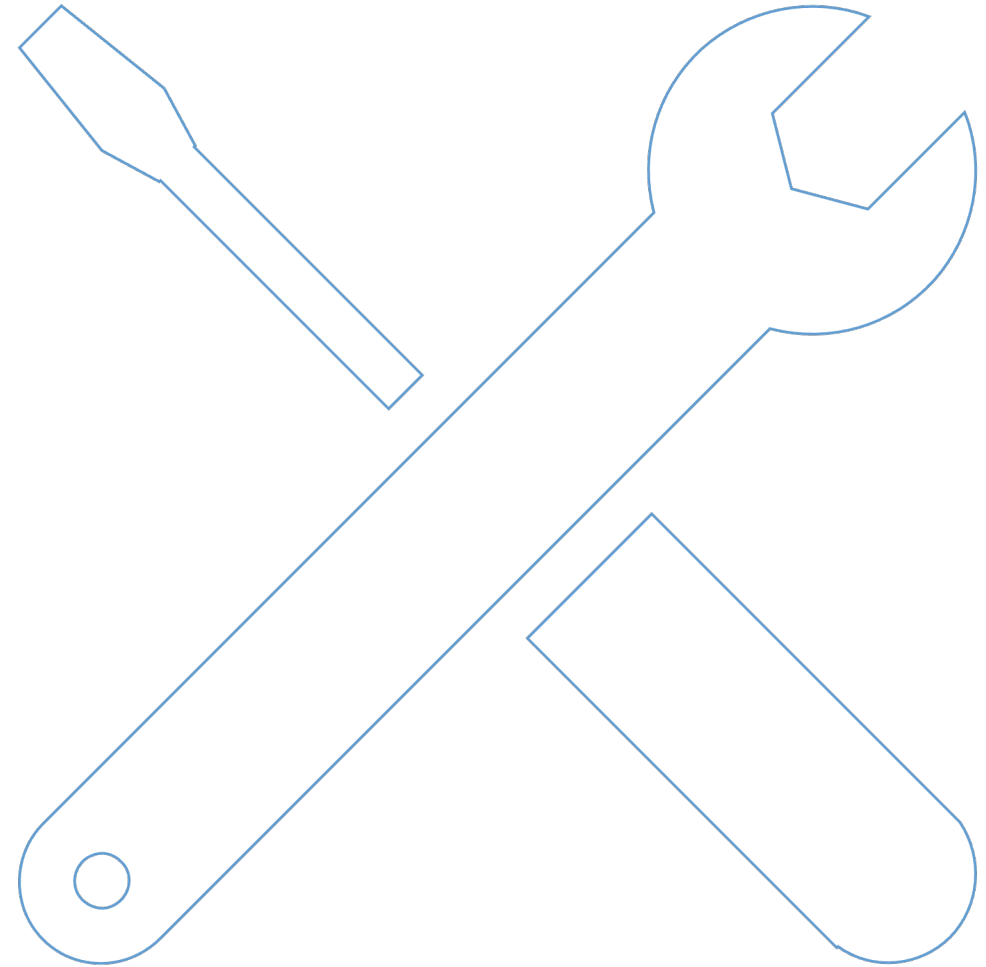
- There is a Souvenir Company intends to open a booth in a metro stations located in New York to sell its products, aiming to increase its profits during the summer season from June 2019 to September 2019.

GOALS

- Choose the best metro station at New York city by:
 - a. Analysis all stations (Entries & Exits) .
 - b. Specify the best station by the high number of entries from June 2019 to September 2019.

Tools

- Jupyter
- Python
- SQLAlchemy
- Numpy
- Pandas
- Seaborn
- Matplotlib





Preprocessing

Data Gathering

- reading the data for the date from June 2019 to September 2019, 11 columns. 3,714,064 rows.

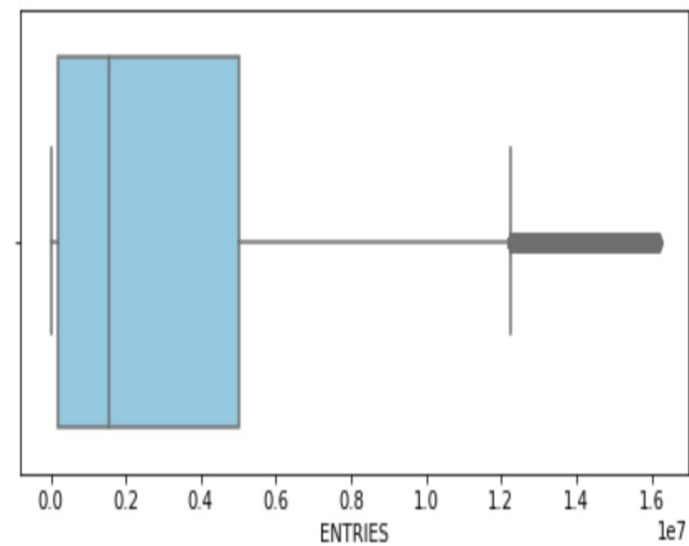
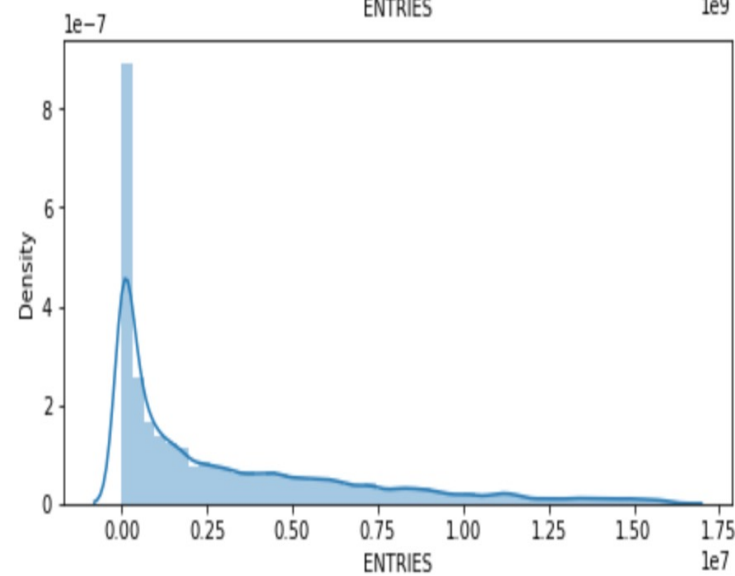
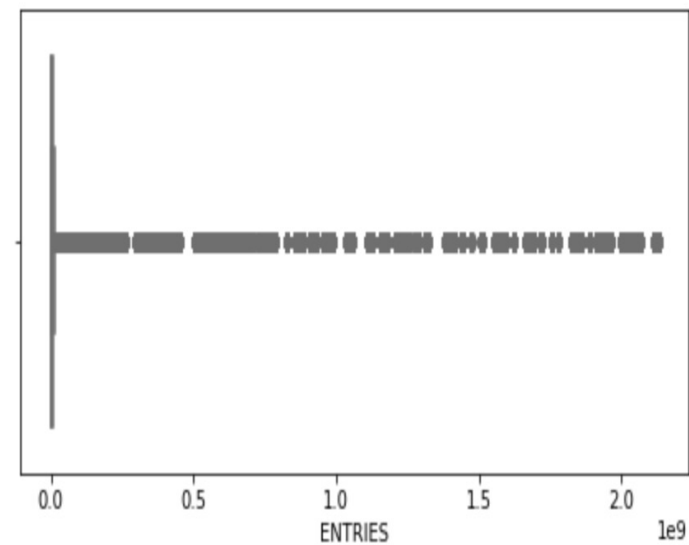
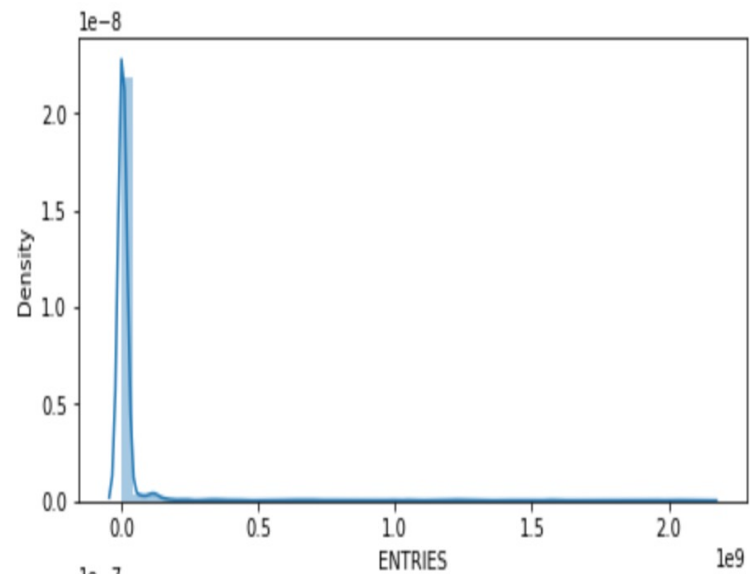
SQL Data

- Connecting data to the database

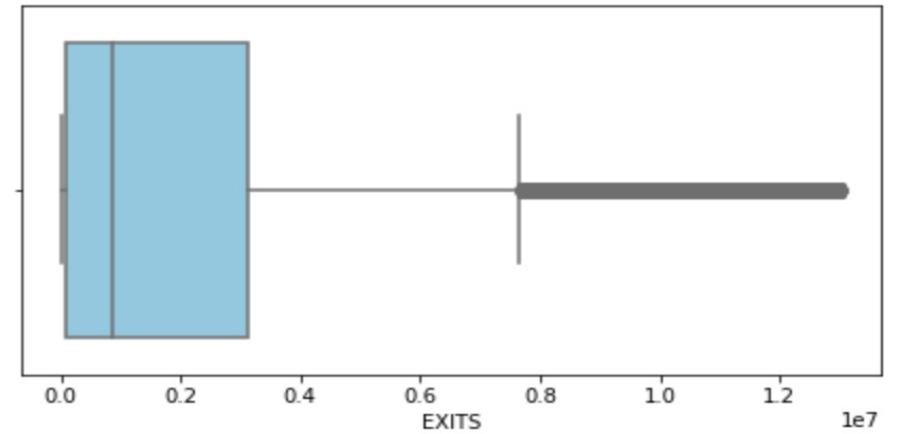
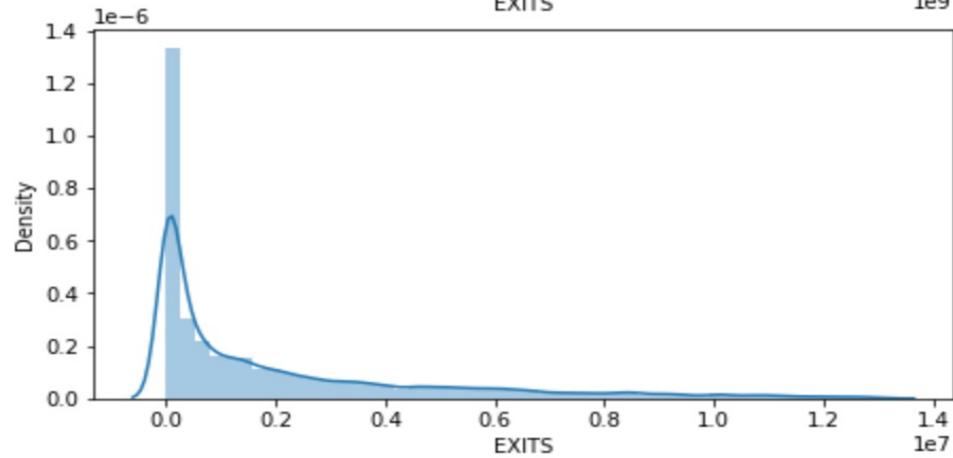
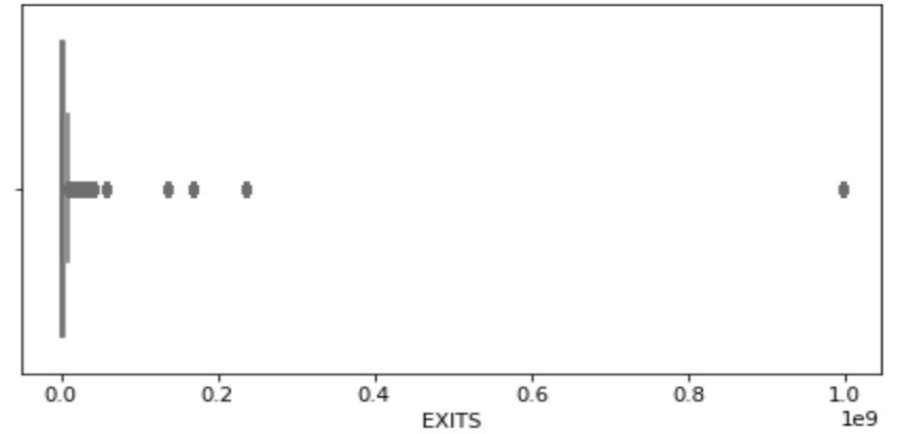
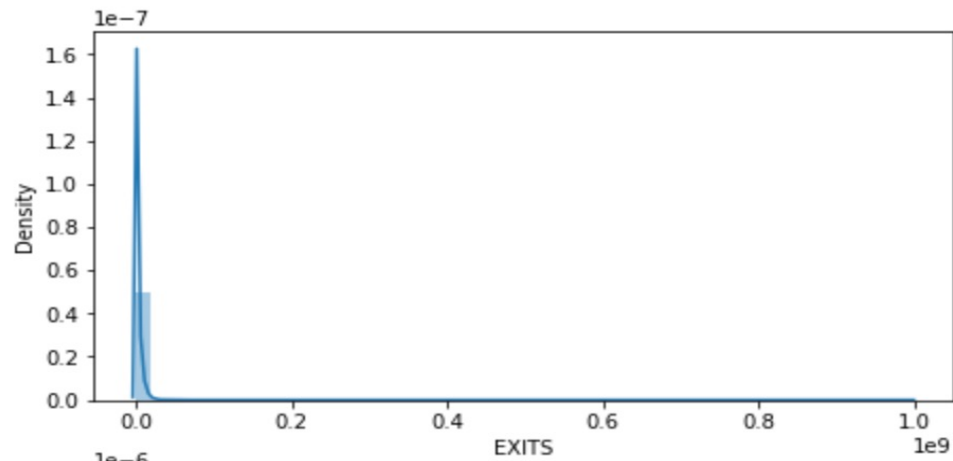
Data Cleaning

- Check Nulls
- Strip Column Name
- Check Duplicate
- Detecting outliers

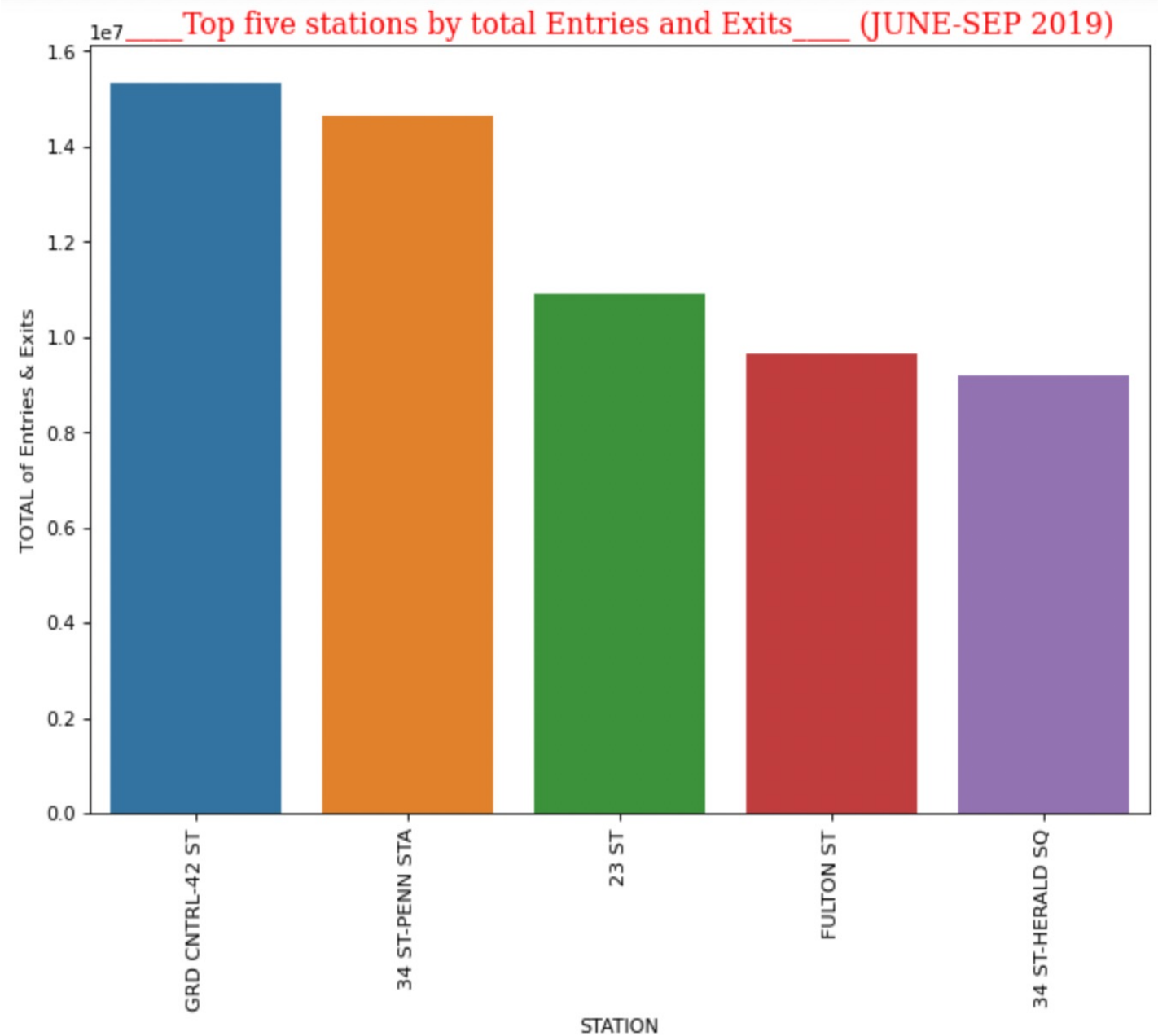
Outliers for entries



Outliers for exits



Top 5 Stations are more crowded



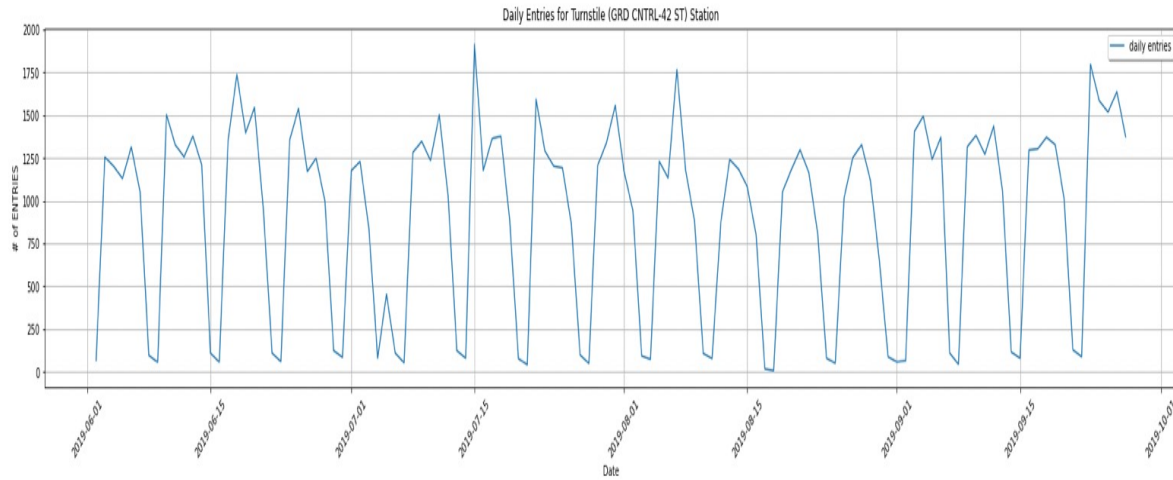


Figure (1)

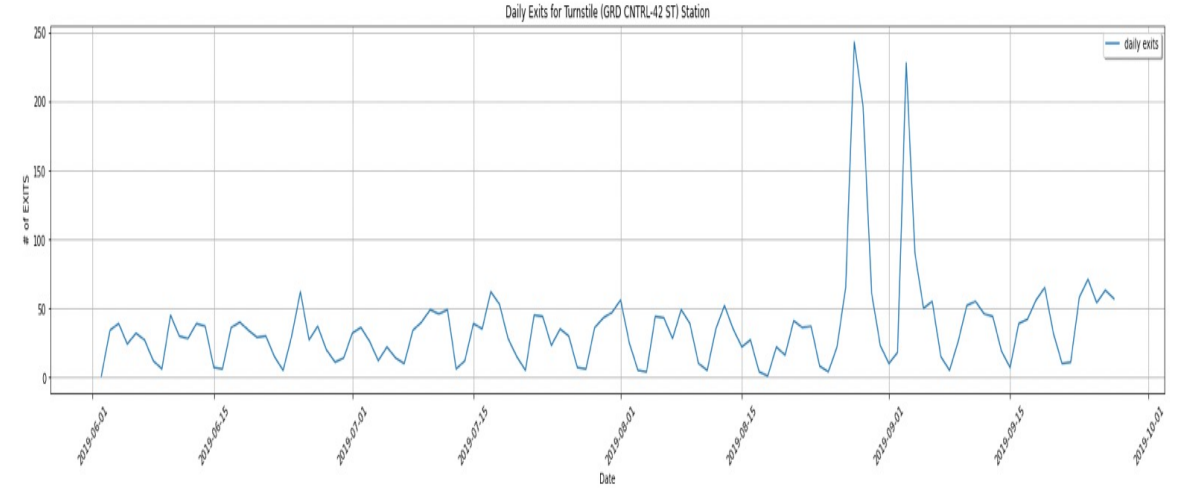


Figure (2)

Result

After analysis we found out the (GRD CNTRL-42 ST) is the busiest stations at NY in Midtown Manhattan

Figure (1) - Showing the indicators of the daily entry to the station in the summer season and the most peak day.

Figure (2) - Showing the indicators of the daily exit from the station during the same time.

Future Work

Providing booths in other stations



Thank You
