RTS – MTA Exploratory Data Analysis

Abdulrahman Ali Almegren

Background & project scope

- **RTS (Road TO Success) is A company that own several vending machines that sells drinks , food and electronics.**
- **RTS focus is on where to put their vending machines on stations in the summer.**
- **By analyzing the given data I want to find top stations , best entries for station ,and daily traffic per week to do maintenance and refill vending machines.**

Datasets:

MTA Turnstile data

Dataset that count number of entries and exits of turnstile in NYC metro stations.

Selection Criteria

- Three months per year (Summer).
- Three years 17/18/19.
- 11 columns.
- 7,809,430 rows.
- Focus in both entries and exits.

Tools:

- SQLite, DB browser, SQLAlchemy.
- Python, Pandas, NumPy.
- Seaborn and Matplotlib.

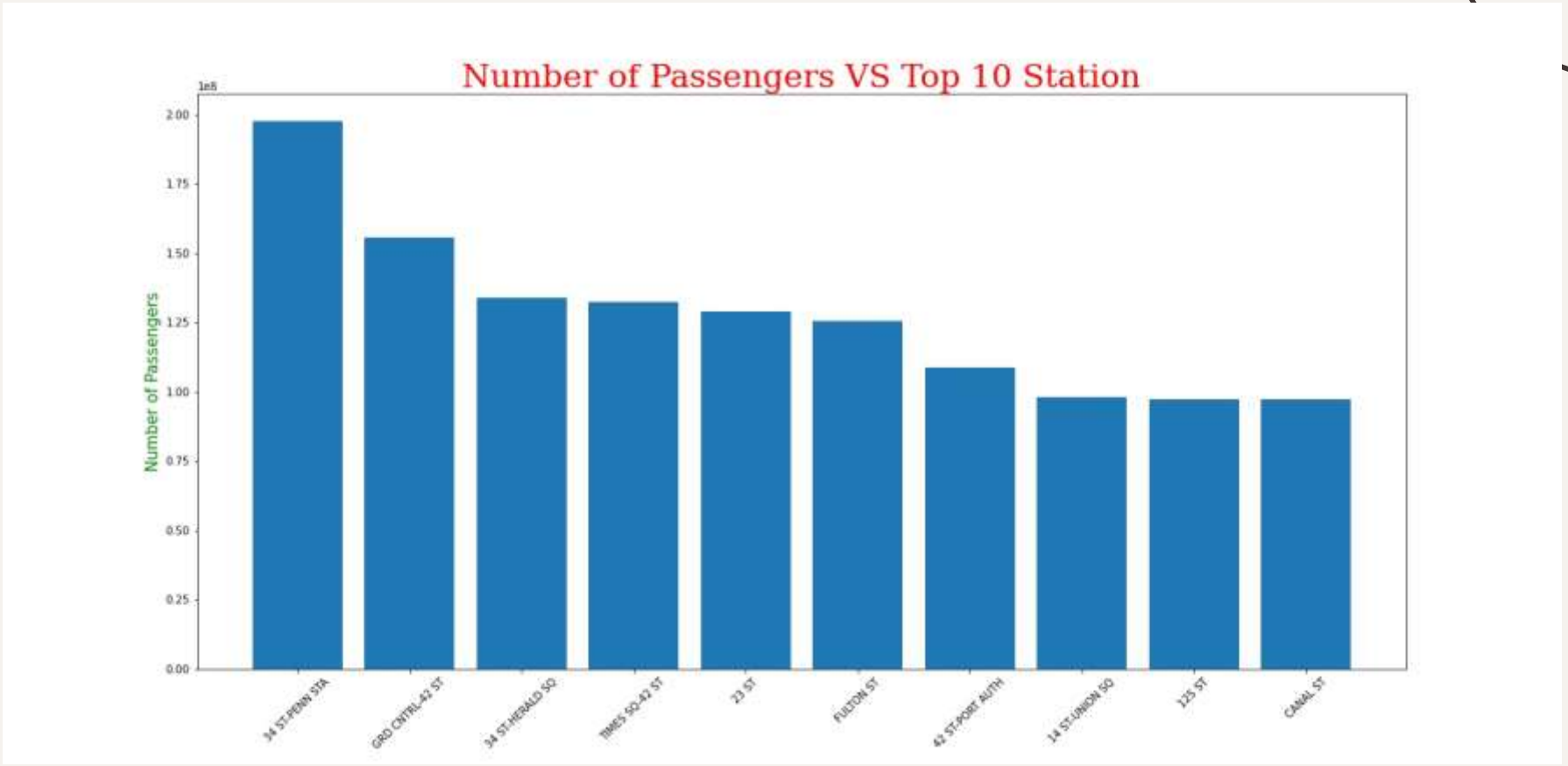
Data Cleaning

- Null values
- Fix column
- Duplicates Rows
- Outliers

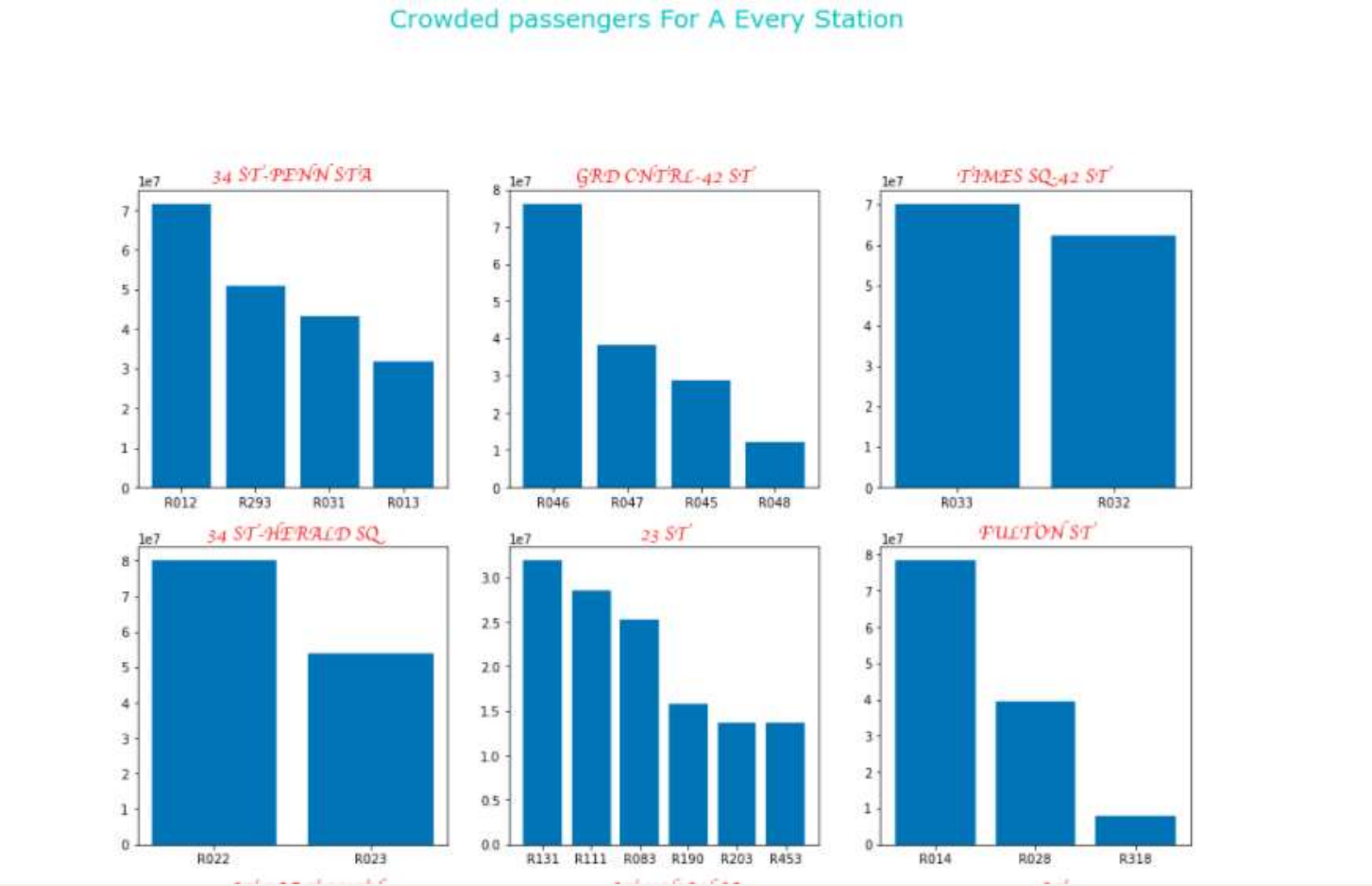
- ✓ NO Null values
- ✓ Rename , add columns , change types of columns
- ✓ Delete duplicates, fix cumulative entries and exits , find daily traffic

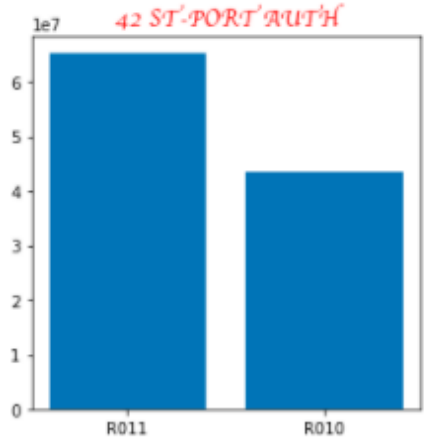
Analysis

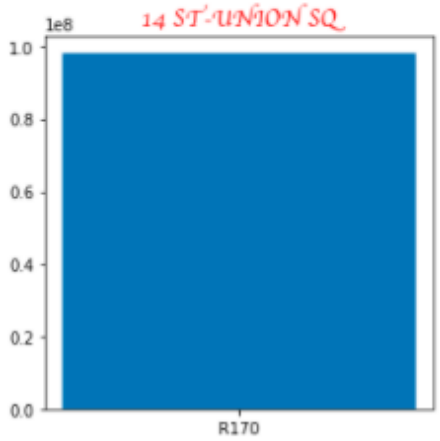
Top Ten Traffic Stations

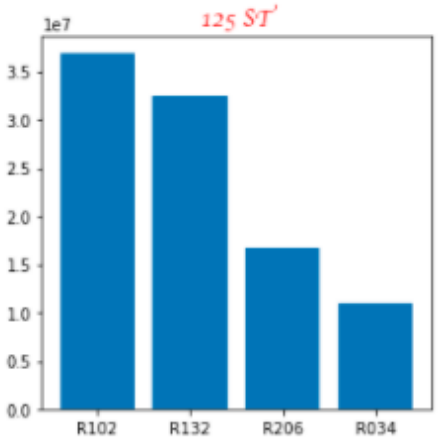


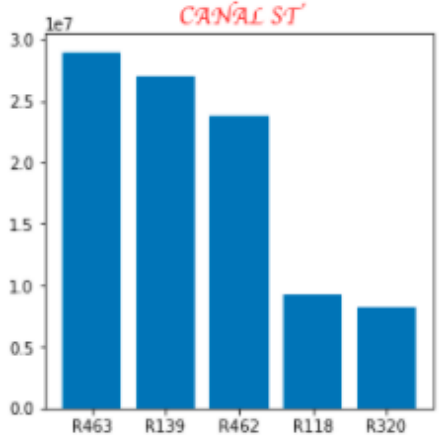
Best entry in single station



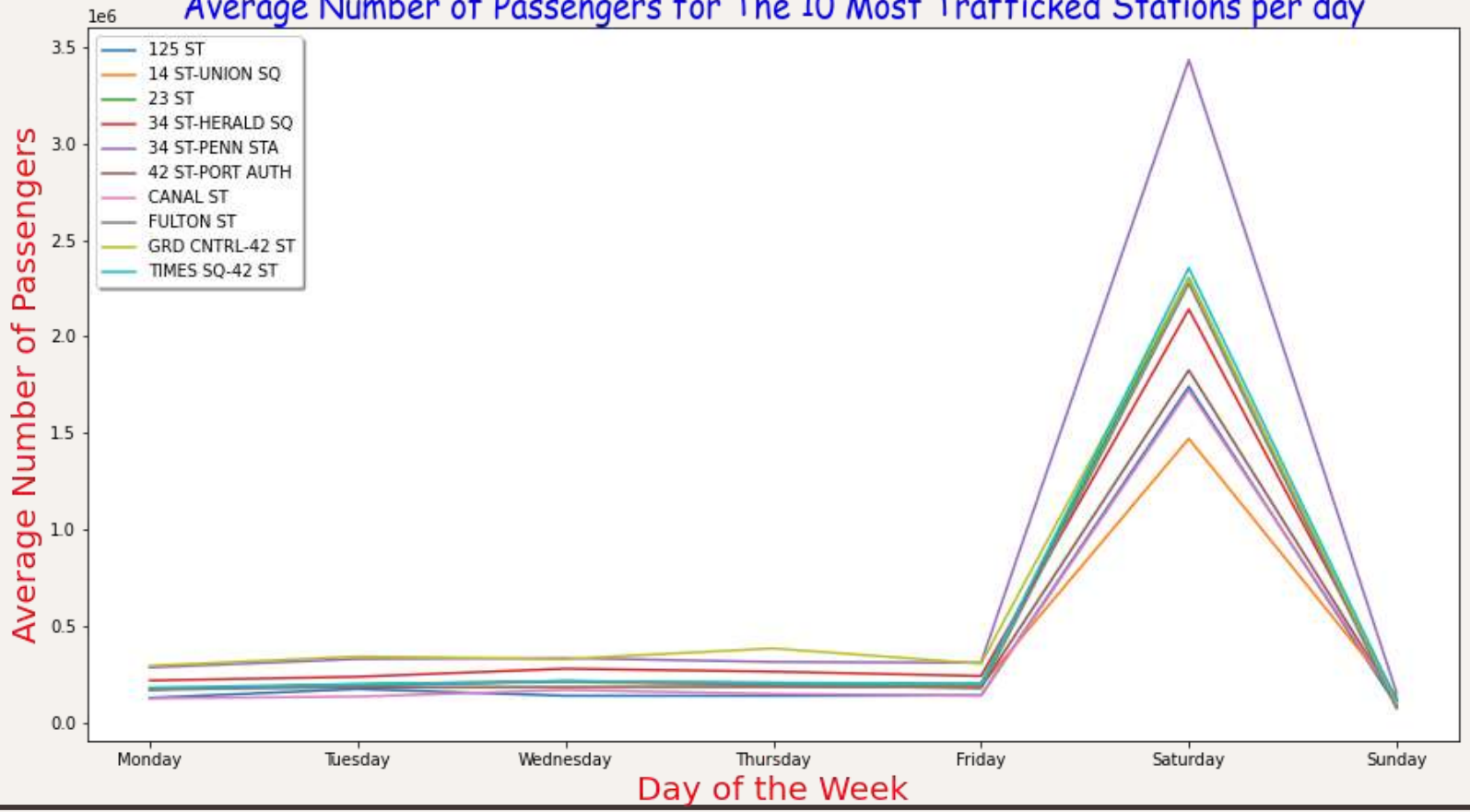

42 ST-PORT AUTH
1e7
6
5
4
3
2
1
0
R011
R010


14 ST-UNION SQ
1e8
1.0
0.8
0.6
0.4
0.2
0.0
R170


125 ST
1e7
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
R102
R132
R206
R034


CANAL ST
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
R463
R139
R462
R118
R320

Avg daily traffic in a week:



Conclusion & Recommendation:

- The top stations are :
  - ✓ 34 ST-PENN STA
  - ✓ GRD CNTRL-42 ST
  - ✓ 34 ST-HERALD SQ
  - ✓ TIMES SQ-42 ST
  - ✓ 23 ST
  - ✓ FULTON ST
  - ✓ 42 ST-PORT AUTH
  - ✓ 14 ST-UNION SQ
  - ✓ 125 ST
  - ✓ CANAL ST.
- Each station have multiple entries so I choose best entry.
- Weekdays is the best days for maintenance and refill vending machines.

Future Work:

- Find rush hours per day.
- Adding dataset for crime percentage to find most safe stations.

Thanks

Do you have any questions?