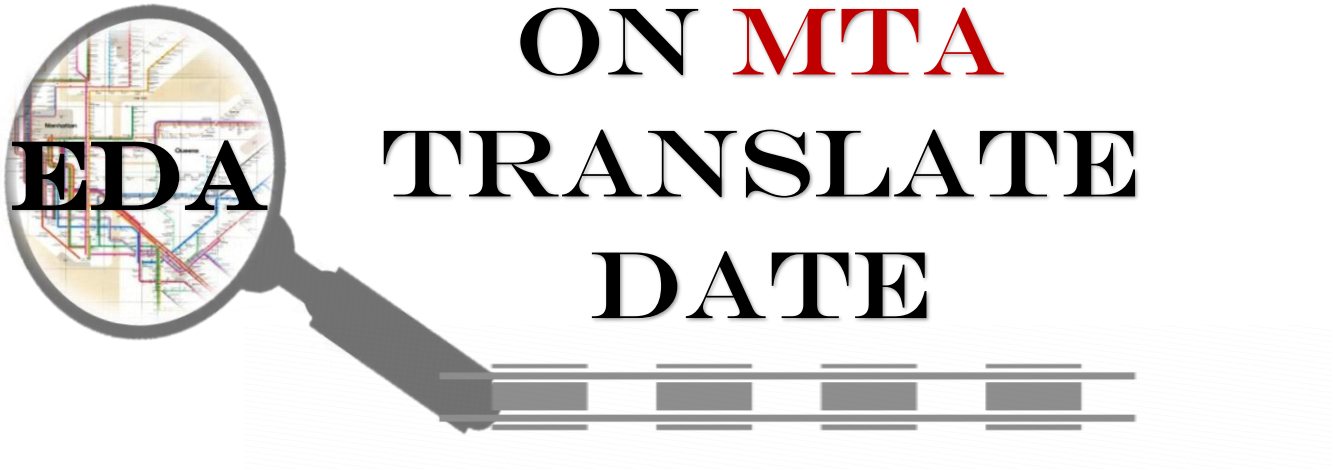
SDAIA T5_bootcamp
Raghad Rashed Al-Saqer
Oct.09.2021



Abstract

The objective of this project was exploratory data analysis (EDA) of the publicly available MTA dataset. I aim to provide insight into the fictional entity (RSTABUS), a fictional bus company covering several regions in the USA. Interested in transporting passengers from several specific hubs to destination stations throughout the day, wants to cover more high demand areas and there are no buses available for customers to connect them to the destination station, they need to know the bus stops(locations) with the mentioned specifications, worked with MTA dataset, Using EDA to provide the required results. Finally, a model with the required specifications was built after several EDA operations.

Data

The MTA turnstile data is scraped for the week 10 JULY 2021 to 25 SEPTEMBER 2021. And I have

(2724418rows,11columns)Taking the first few rows of the data, we observed the following data frame:

```
rstabus_df.head()
```

Out[289]:

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	09/18/2021	00:00:00	REGULAR	7637026	2613455
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	09/18/2021	04:00:00	REGULAR	7637036	2613457
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	09/18/2021	08:00:00	REGULAR	7637047	2613478
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	09/18/2021	12:00:00	REGULAR	7637104	2613524
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	09/18/2021	16:00:00	REGULAR	7637236	2613561

The descriptions of the column features are given here
web.mta.info/developers/resources/nyct/turnstile/ts_Field_Description.txt.

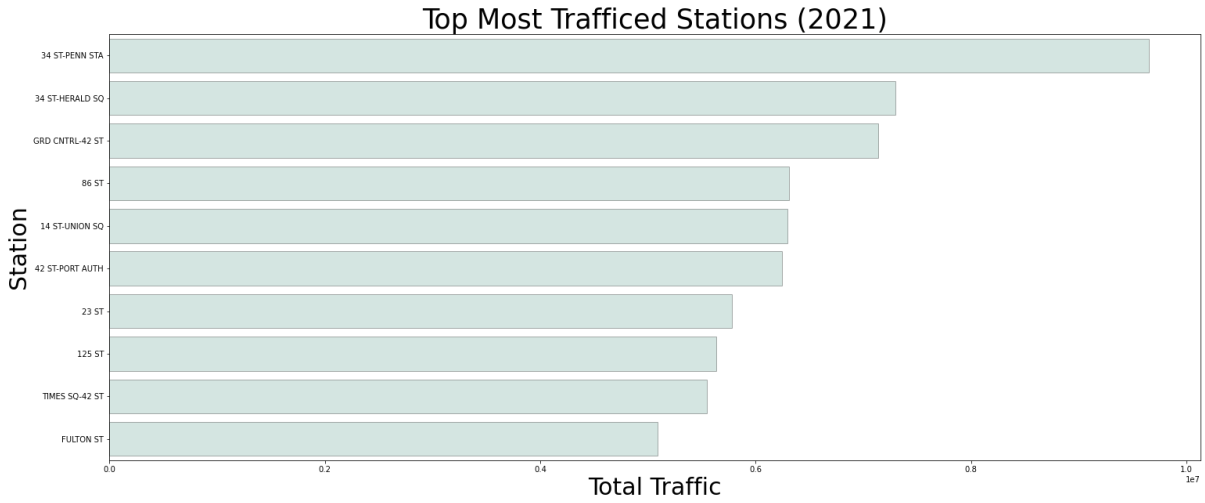
Workflow

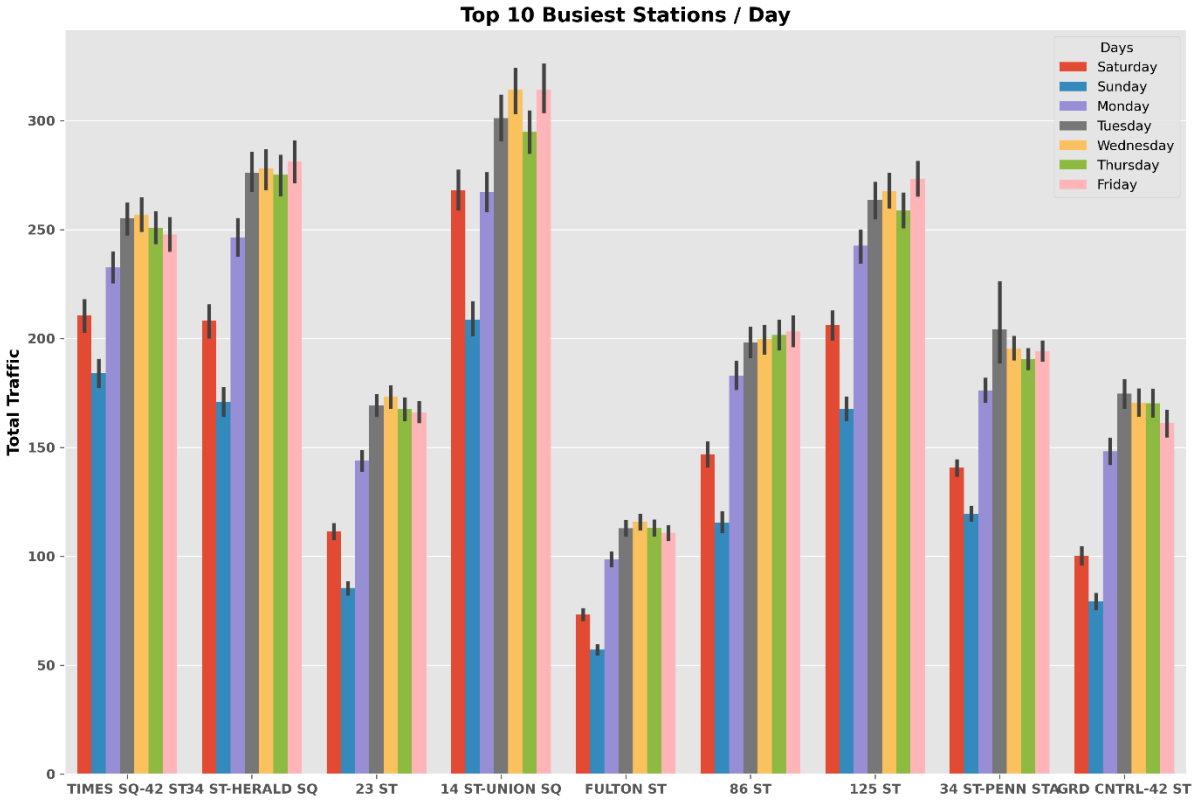
- Data scraping
- Data cleaning and preparation
- EDA and visualization
- Future work

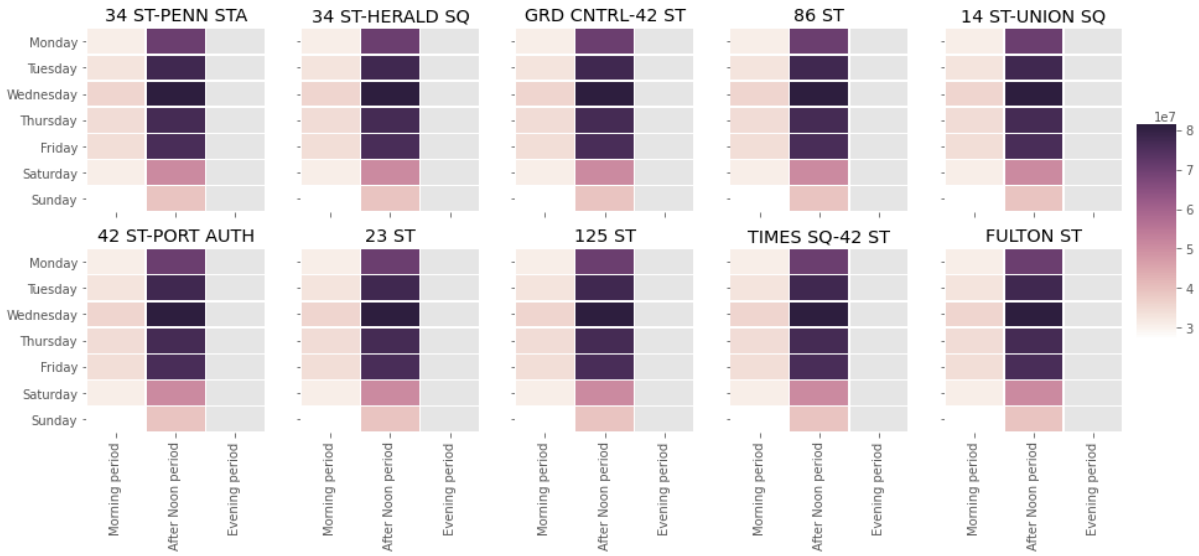
Tools

- SQLite
- Python (sqlalchemy, Numpy and Pandas, etc…)
- Google Map to ensure that the stations serve a good geographic area and a bus centre may serve a remote line station.

Communication






34 ST-PENN STA
34 ST-HERALD SQ
GRD CNTRL-42 ST
86 ST
14 ST-UNION SQ
42 ST-PORT AUTH
23 ST
125 ST
TIMES SQ-42 ST
FULTON ST
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Morning period
After Noon period
Evening period
1e7
8
7
6
5
4
3