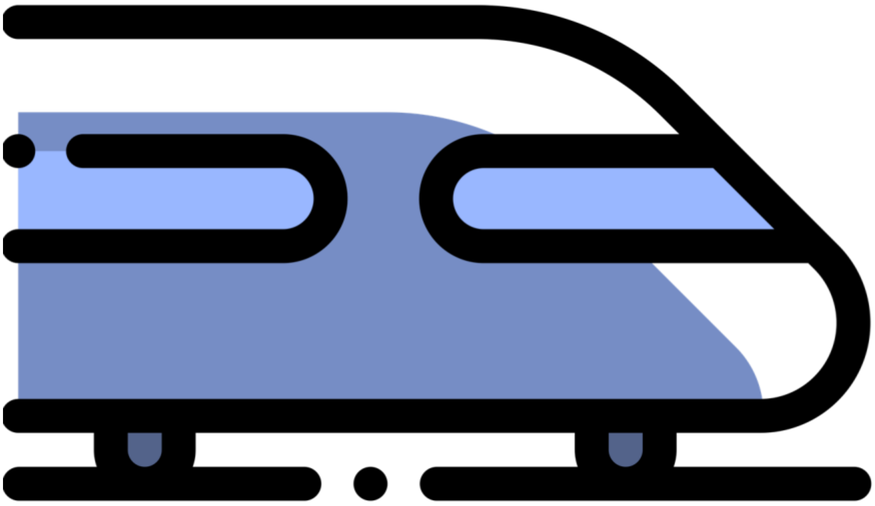
أكاديمية سدايا
SDAIA Academy

MTA DATA UTLIZATION

8 SEP 2021



NOUF ALSAEED

Table of Contents

Abstract

The aim of this project is to explore and preprocess the Metropolitan Transportation Authority (MTA) data, which is North America's largest transportation network. And to explore and prepresses that data I did it via the use of some techniques and showed the best location for opening a mall near to one of the stations in New York City.

Data description

The data I used in this project is provided by the MTA with some columns that I used and drive new columns that are useful for my project.
Here are the columns that was provided by the MTA:

Column	**Description**
C/A	Control Area (A002)
UNIT	Remote Unit for a station (R051)
SCP	Subunit Channel Position represents an specific address for a device (02-00-00)
STATION	Represents the station name the device is located at
LINENAME	Represents all train lines that can be boarded at this station Normally lines are represented by one character. LINENAME 456NQR repersents train server for 4, 5, 6, N, Q, and R trains.
DIVISION	Represents the Line originally the station belonged to BMT, IRT, or IND
DATE	Represents the date (MM-DD-YY)
TIME	Represents the time (hh:mm:ss) for a scheduled audit event

DESc	Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours) 1. Audits may occur more that 4 hours due to planning, or troubleshooting activities. 2. Additionally, there may be a "RECOVR AUD" entry: This refers to a missed audit that was recovered.
ENTRIES	The comulative entry register value for a device
EXIST	The cumulative exit register value for a device

And I needed to drive some columns which are:

Column	Description
WEEK_DAY	The weekday name for each date
PREV_TIME	Shifted time per date and time
PREV_ENTRIES	Shifted entries per date and time
PREV_EXITS	Shifted exits per date and time
TOTAL_ENTRIES	Total entries per date and time
TOTAL_EXITS	Total exits per date and time
TOTAL_TRAFFIC	Total traffic per date and time

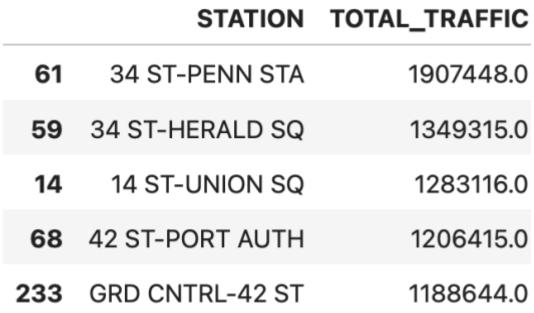
Tools

The tools that I used for the MTA project are:

- Python programming language
- Jupyter lab as programming environment
- Numpy and Pandas for data manipulation
- scipy and math for mathematical operations
- Matplotlib and Seaborn for plotting

Algorithms and results

I grouped the data by the weekends which are Saturdays and Sundays in New York and calculated the total traffic for each station, I used group by method, and sort to sort data ascendingly lastly, I chose the top 5 stations The below table shows the top 5 stations:



	STATION	TOTAL_TRAFFIC
61	34 ST-PENN STA	1907448.0
59	34 ST-HERALD SQ	1349315.0
14	14 ST-UNION SQ	1283116.0
68	42 ST-PORT AUTH	1206415.0
233	GRD CNTRL-42 ST	1188644.0

Table 1 top 5 stations

Top 5 station showed in plot:

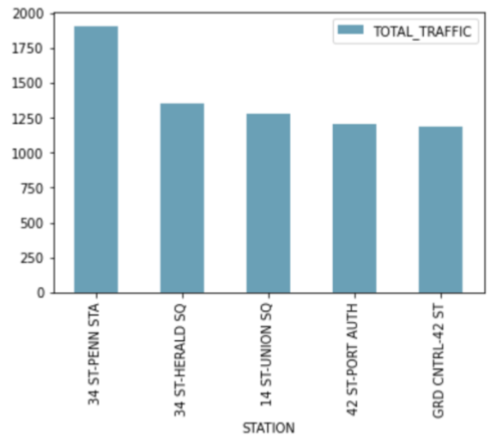


Figure 1 top 5 stations

Here I wanted to see the different between Sundays and Saturdays:

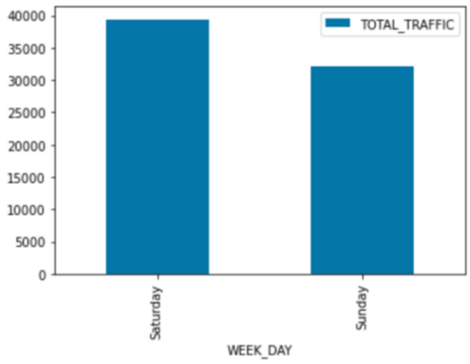


Figure 2 total traffic in Sundays and Saturdays

Conclusion

Finally, after all the preprocessing and data exploration, I have found that station ("34 ST-PENN STA") is the best area to open a mall based on data results as I have checked the station with the highest amount of ridership and exists and entries and saw the weekdays and weekends differences.

Reference

1. MTA. 2021. About Us. [online] Available at: <https://new.mta.info/about-us> [Accessed 30 August 2021].