

A low-angle, upward-looking perspective of several tall skyscrapers in a city, likely New York City, with the Empire State Building prominently in the center. The image is overlaid with a semi-transparent blue and purple gradient.

# EDA - MTA REPORT

Leveraging Foot-Traffic For Marketing  
Purposes



# INTRODUCTION

1

## MTA DATASET

(MTA) publishes data on a weekly basis. This data does not represent value, Unless it is analyzed and used correctly to serve certain business domain to take advantage of this foot-traffic.

2

## OBJECTIVES

- Identification of the subway stations that have the most crowded.
- Helping an Ads agency to target crowds from busy stations.
- Improve the ads cost based on the busiest stations and day hours.

# Overview



Data Scraping



Data Preparing



Data  
Visualisation

# Data Scrapping

- Data gathered from MTA official web-sit in the period from August to October of 2021

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	07/31/2021	00:00:00	REGULAR	7611181	2603110
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	07/31/2021	04:00:00	REGULAR	7611192	2603113
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	07/31/2021	08:00:00	REGULAR	7611197	2603126
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	07/31/2021	12:00:00	REGULAR	7611235	2603178
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	07/31/2021	16:00:00	REGULAR	7611357	2603211

# Data Preparing

- Replace the outliers with mean values of the turnstile before and after the error data datetime
- Total traffic at time period = sum of entry difference and exit difference

STATION		DATETIME	DAY_OF_WEEK	ENTRY_DIFF	EXIT_DIFF	TOTAL_TRAFFIC
0	59 ST	2021-07-31 04:00:00	5	11.0	3.0	14.0
1	59 ST	2021-07-31 08:00:00	5	5.0	13.0	18.0
2	59 ST	2021-07-31 12:00:00	5	38.0	52.0	90.0
3	59 ST	2021-07-31 16:00:00	5	122.0	33.0	155.0
4	59 ST	2021-07-31 20:00:00	5	155.0	30.0	185.0

# Data Visualisation

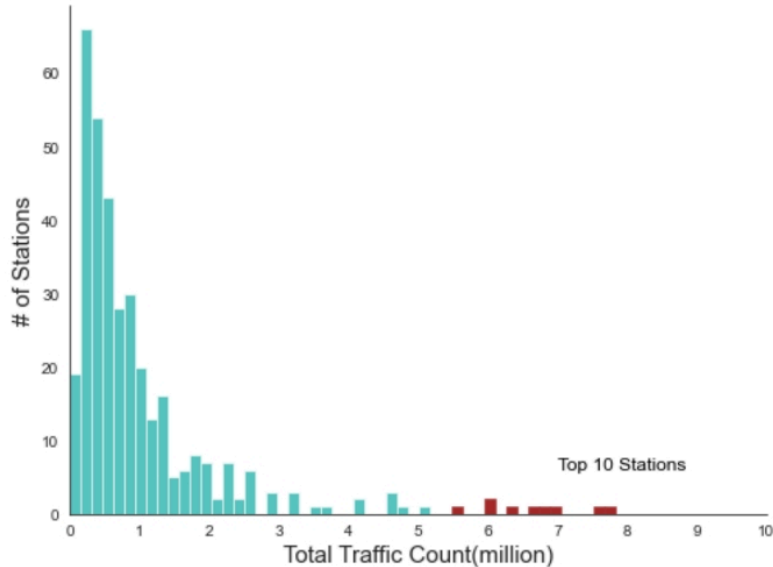
- Distribution of traffic across station
- Top 10 Busiest Subway Stations In NYC



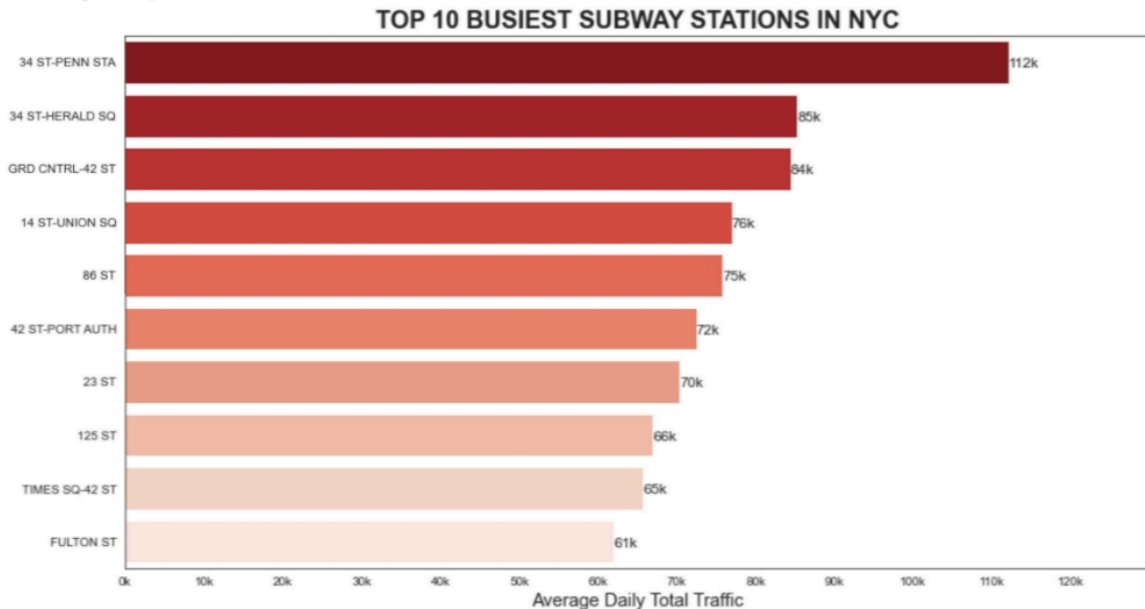
# Data Visualisation

- Distribution of traffic across station

**Distribution of Traffic Across Stations**



# Data Visualisation



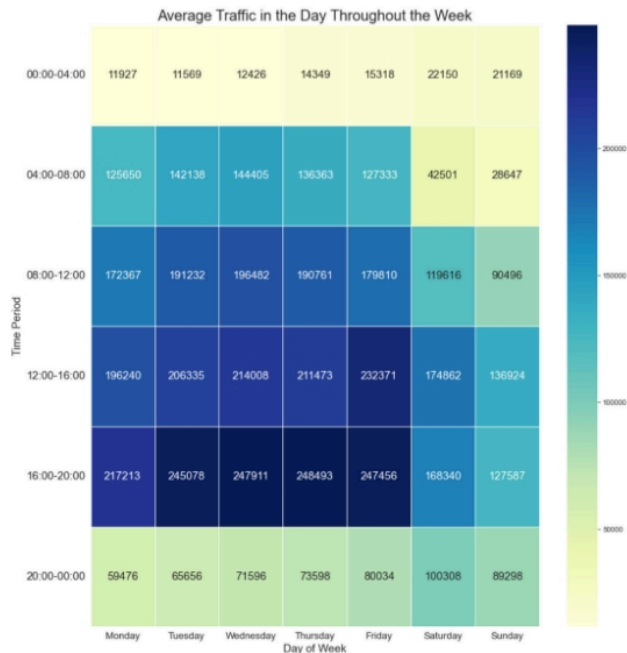
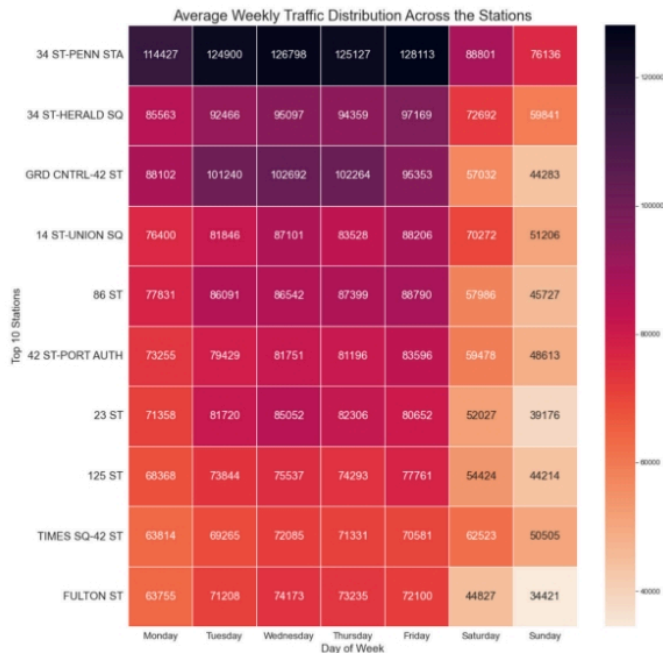


- Distribution of traffic across the week
- Traffic on weekdays > weekends

# Result

## Top 10 Stations Traffic pattern

- Most crowded days: Tues, Wed, Thu, Fri.
- Traffic is heavier in afternoon





## Future Work

- Interactive Dashboard
- More EDA on each stations



# Thank You

Presented By: Hatim

SDAIA Bootcamp T5