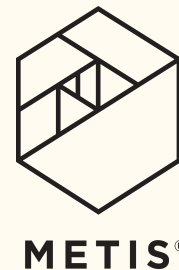# MTA Traffic Analysis for WTWY
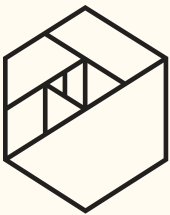
METIS®

Presented by: Alvin, Daniel, Xinni

As Project 1 of
Metis Data Science Bootcamp

Introduction

METIS®

- WomenTechWomenYes (WTWY) is holding an annual gala, and hopes to invite as many interested individuals as possible.

- WTWY is intending to place street teams at the entrance to NYC subway stations to collect email addresses and give out free tickets.

- **Agenda**: Apply data analytics on the MTA subway data on traffic flow to optimize placement of street teams to maximize engagements.
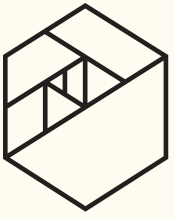
Methodology

- Wrote a script to webscape weekly MTA turnstile data to analyze traffic.

- Analyze turnstile data from the latest week 22/8/20 to 28/8/20.

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-08-22	00:00:00	REGULAR	7447810	2532191
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-08-22	04:00:00	REGULAR	7447812	2532197
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-08-22	08:00:00	REGULAR	7447824	2532208
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-08-22	12:00:00	REGULAR	7447852	2532248
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	2020-08-22	16:00:00	REGULAR	7447937	2532276

217832 rows

Methodology

- Each turnstile has a unique combination of 'UNIT' and 'SCP' number.

- Traffic at each time period is given by the difference between consecutive entry/exit serial numbers.

|	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS	ENTRY_DELTA	EXIT_DELTA
10969	A060	R001	00-00-00	WHITEHALL S-FRY	R1W	BMT	2020-08-15	01:00:00	REGULAR	5106326	2946514	NaN	NaN
10970	A060	R001	00-00-00	WHITEHALL S-FRY	R1W	BMT	2020-08-15	05:00:00	REGULAR	5106329	2946516	3.0	2.0
10971	A060	R001	00-00-00	WHITEHALL S-FRY	R1W	BMT	2020-08-15	09:00:00	REGULAR	5106475	2946572	146.0	56.0
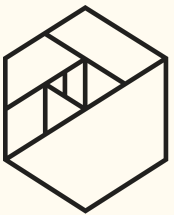
METIS®

Data Preparation

- Sort and look out for **error datas** those duplicates or outliers -
- Dropped duplicate rows in the dataset.

METIS®

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
726085	N400A	R359	2/6/01	COURT SQ	EMG	IND	22/8/20	3:59	REGULAR	4624908	934503
726086	N400A	R359	2/6/01	COURT SQ	EMG	IND	22/8/20	4:00	REGULAR	4624908	934503
726087	N400A	R359	2/6/01	COURT SQ	EMG	IND	22/8/20	4:00	REGULAR	4624908	934503
726088	N400A	R359	2/6/01	COURT SQ	EMG	IND	22/8/20	4:02	REGULAR	4624908	934503
726089	N400A	R359	2/6/01	COURT SQ	EMG	IND	22/8/20	4:03	REGULAR	4624908	934503
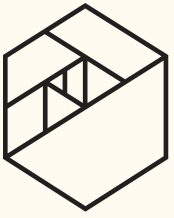
Data Preparation

-             Negative   or   More than 10,000.
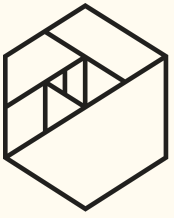
```
252044     -8291685.0
264128          -91.0
264129           -4.0
264130          -34.0
264131          -25.0
               ...
341647       -17768.0
339394      -210087.0
```

```
271427     1.057536e+09
271465     1.549702e+09
300430     8.368614e+07
297544     1.443527e+06
297546     1.443555e+06
```

- Replace **errors datas** with mean values of the station before and after the error data timing.

- Sum the change of entry and exit to get the total traffic for each timing

Data Preparation

- Add weekday columns to the date and change the list to matrix by station & weekday

WEEKDAY	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
STATION							
34 ST-PENN STA	63.9k	64.3k	66.2k	65.9k	65.3k	42.4k	33.4k
34 ST-HERALD SQ	54.3k	56.0k	57.7k	55.9k	55.3k	39.6k	31.2k

- Change time to time format %H:%M:%S3 then group timings into <u>6 intervals of 4 hours</u> each. (Eg. 12am-4-8-12pm-4-8-12am)

Results
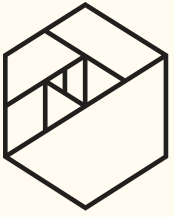
Distribution of Traffic Across Stations
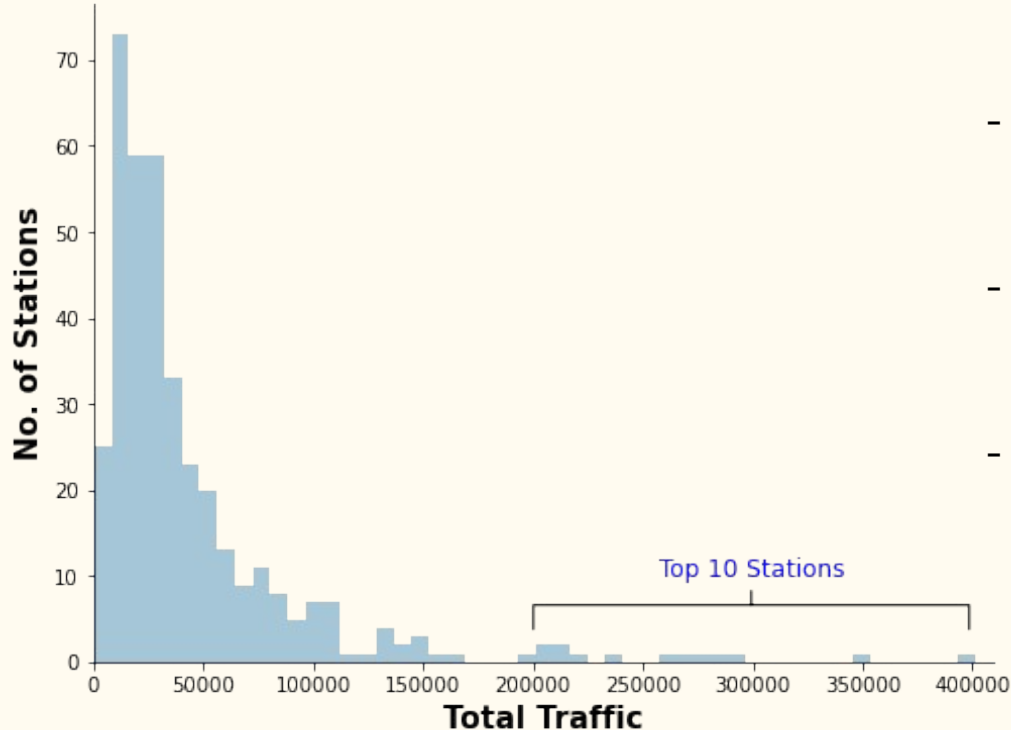
- Distribution of traffic is heavily right-skewed.

- Top 10 stations in traffic are outliers in distributions.

- Top 10 stations are probably in dense residential areas or close to city centre with a lot of human activities.
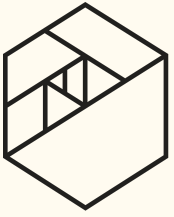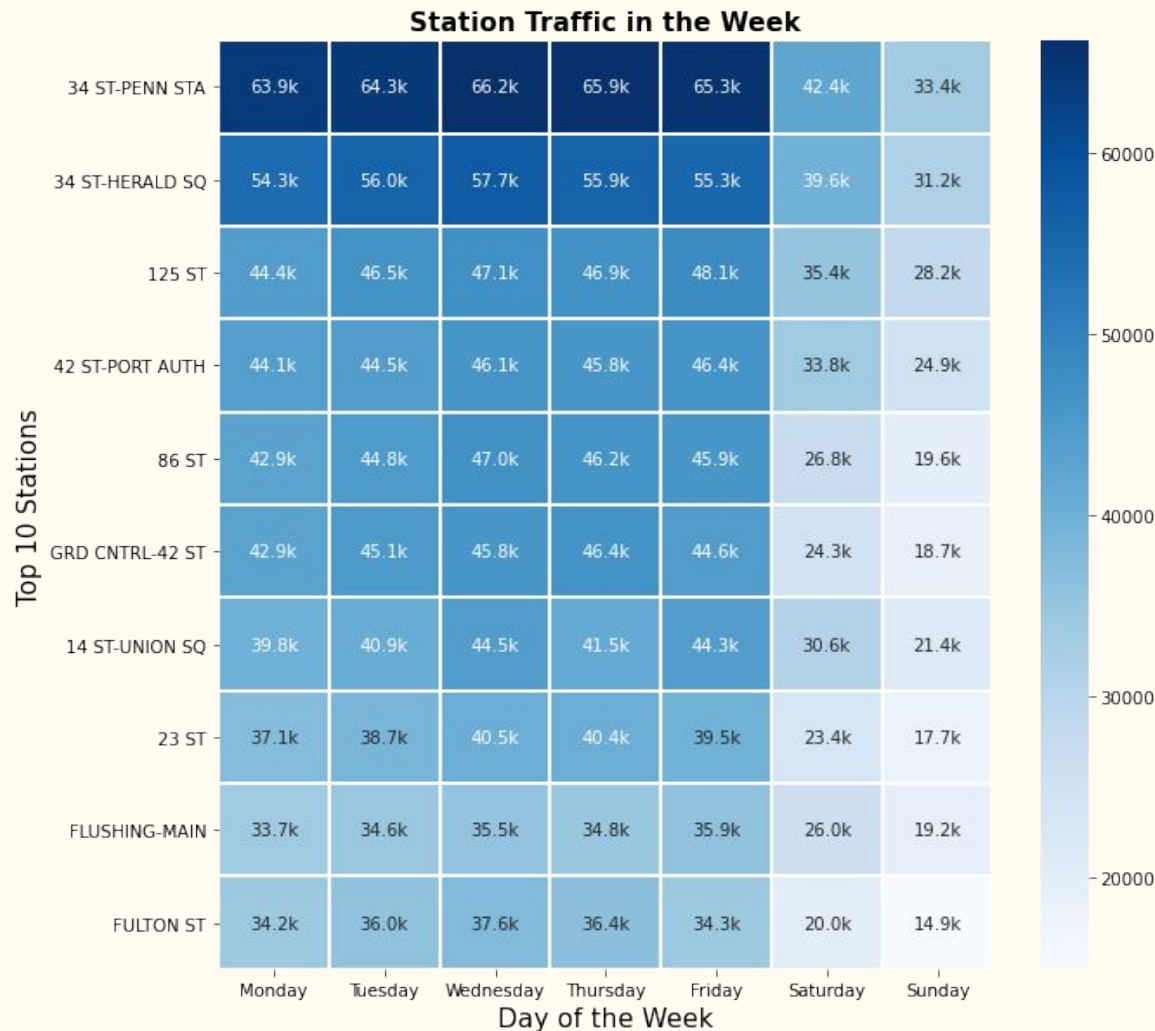
Results

Busiest MTA Stations from 22/8 to 28/8
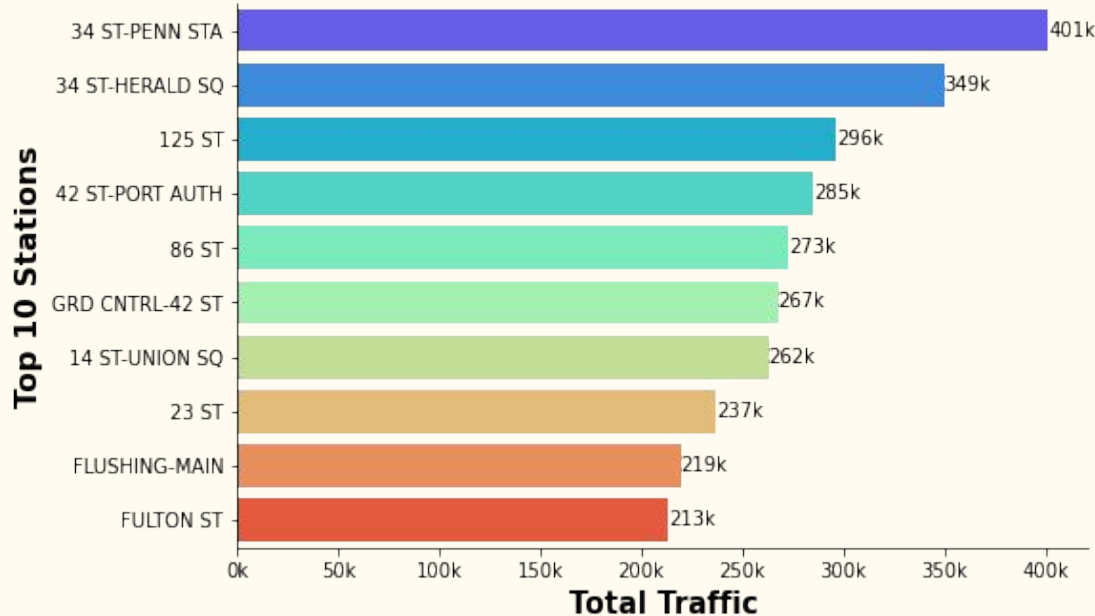
- Distribution of traffic across the top 10 stations.

- 34 St - Penn Station and 34 St - Herald Square station has notably more traffic than the rest.

METIS®

Results

- Distribution of traffic across the week.

- Trend shows that traffic on weekdays is generally more than weekends.

Station Traffic in the Week

Top 10 Stations	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
34 ST-PENN STA	63.9k	64.3k	66.2k	65.9k	65.3k	42.4k	33.4k
34 ST-HERALD SQ	54.3k	56.0k	57.7k	55.9k	55.3k	39.6k	31.2k
125 ST	44.4k	46.5k	47.1k	46.9k	48.1k	35.4k	28.2k
42 ST-PORT AUTH	44.1k	44.5k	46.1k	45.8k	46.4k	33.8k	24.9k
86 ST	42.9k	44.8k	47.0k	46.2k	45.9k	26.8k	19.6k
GRD CNTRL-42 ST	42.9k	45.1k	45.8k	46.4k	44.6k	24.3k	18.7k
14 ST-UNION SQ	39.8k	40.9k	44.5k	41.5k	44.3k	30.6k	21.4k
23 ST	37.1k	38.7k	40.5k	40.4k	39.5k	23.4k	17.7k
FLUSHING-MAIN	33.7k	34.6k	35.5k	34.8k	35.9k	26.0k	19.2k
FULTON ST	34.2k	36.0k	37.6k	36.4k	34.3k	20.0k	14.9k

Day of the Week

METIS®

Results

- Trend shows that traffic is heavier in the late afternoon and evening periods, even during weekdays.
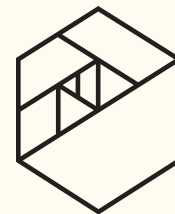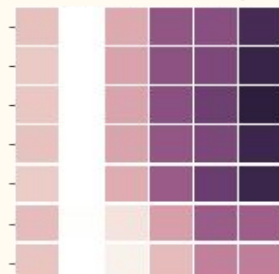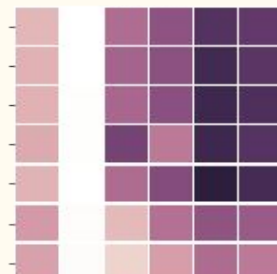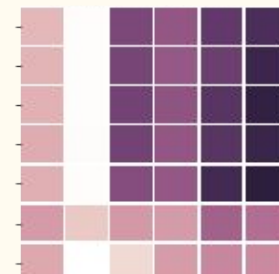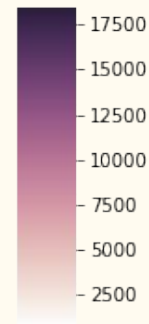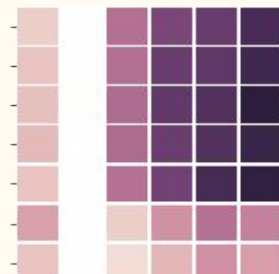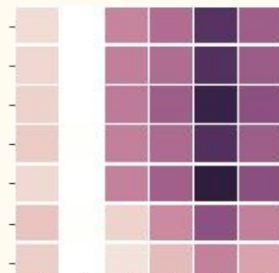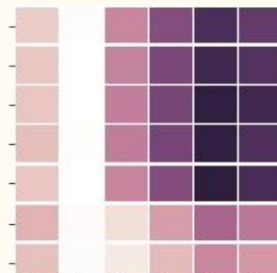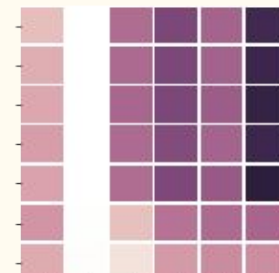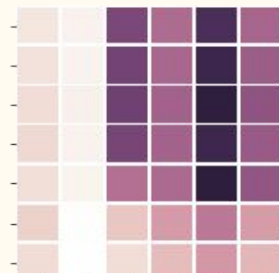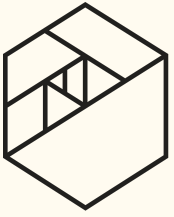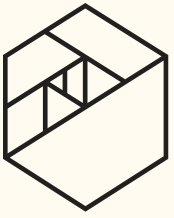
Conclusion

- Given limited manpower resources, WTWY should focus on sending street teams to the top 10 busiest MTA stations.

- If time is a constraint, WTWY should focus on weekdays in the period of late afternoons to late evenings.

- Morning period on weekdays can be avoided due to has less traffic (due to WFH), and commuters are busy reporting to work.
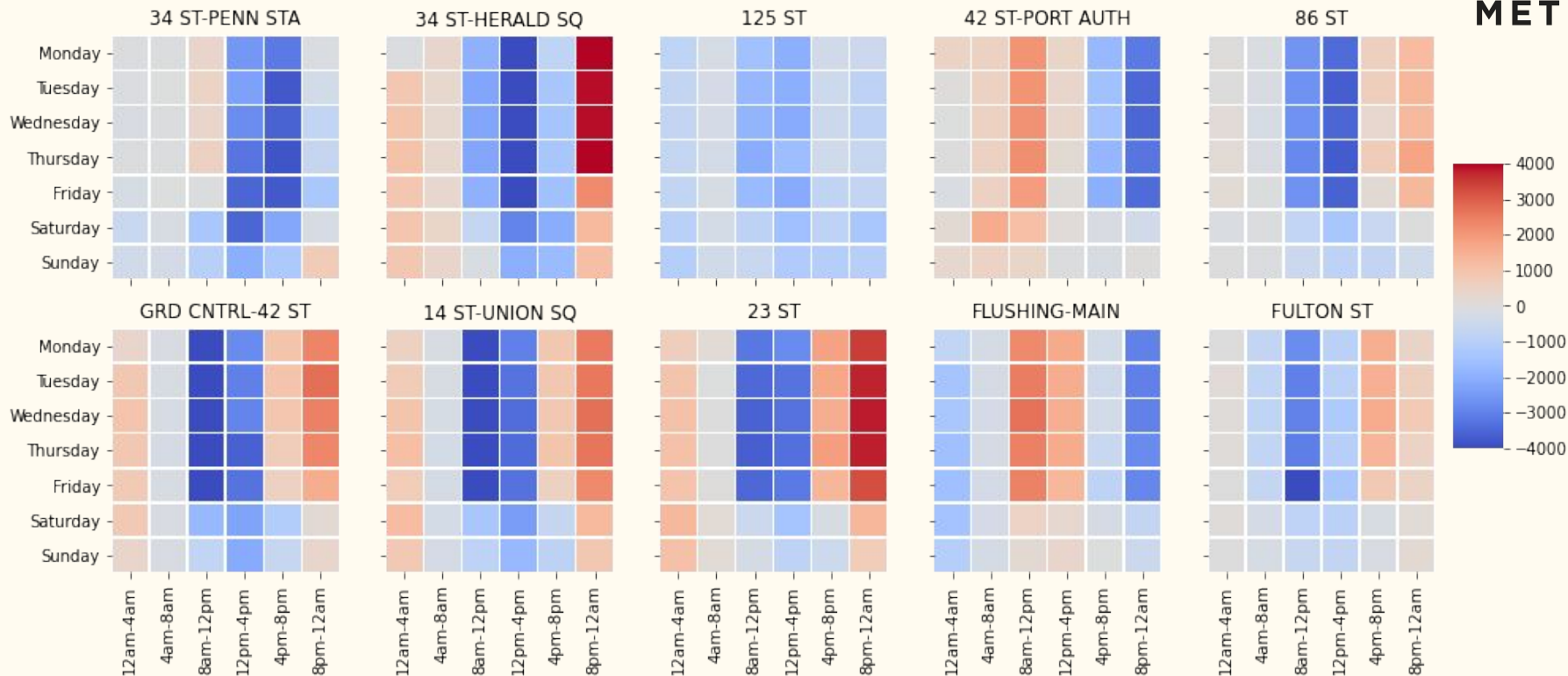
Future Work

- Using our automated web scraping script on the MTA website, our team can quickly produce similar data insights for other date periods.

- Generate other insights by analyzing trends based on total entries and total exit respectively.

- Generate other visualizations such as using a choropleth map for the locations of MTA stations in NYC.

- Possibility of including areas outside the MTA transit where there are more commercial and leisure activities taking place.
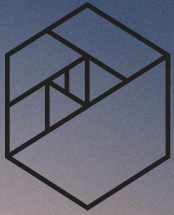
METIS®

Appendix
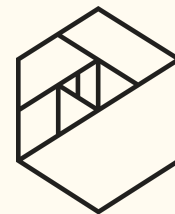
Net Entry/Exit of Commuters

Thank You!

METIS®