MTA Turnstile Data Analysis

By Wesley Sheh

Introduction

The objective of this project was to find out optimal days and time to schedule maintenance work in the MTA stations.

NYC historically has been a busy place with tons of traffic going through their underground transportation system. To minimize congestion, we take a look at MTA turnstile data and look for patterns within the week and day to find four hour blocks for maintenance to occur within the week.

Methodology

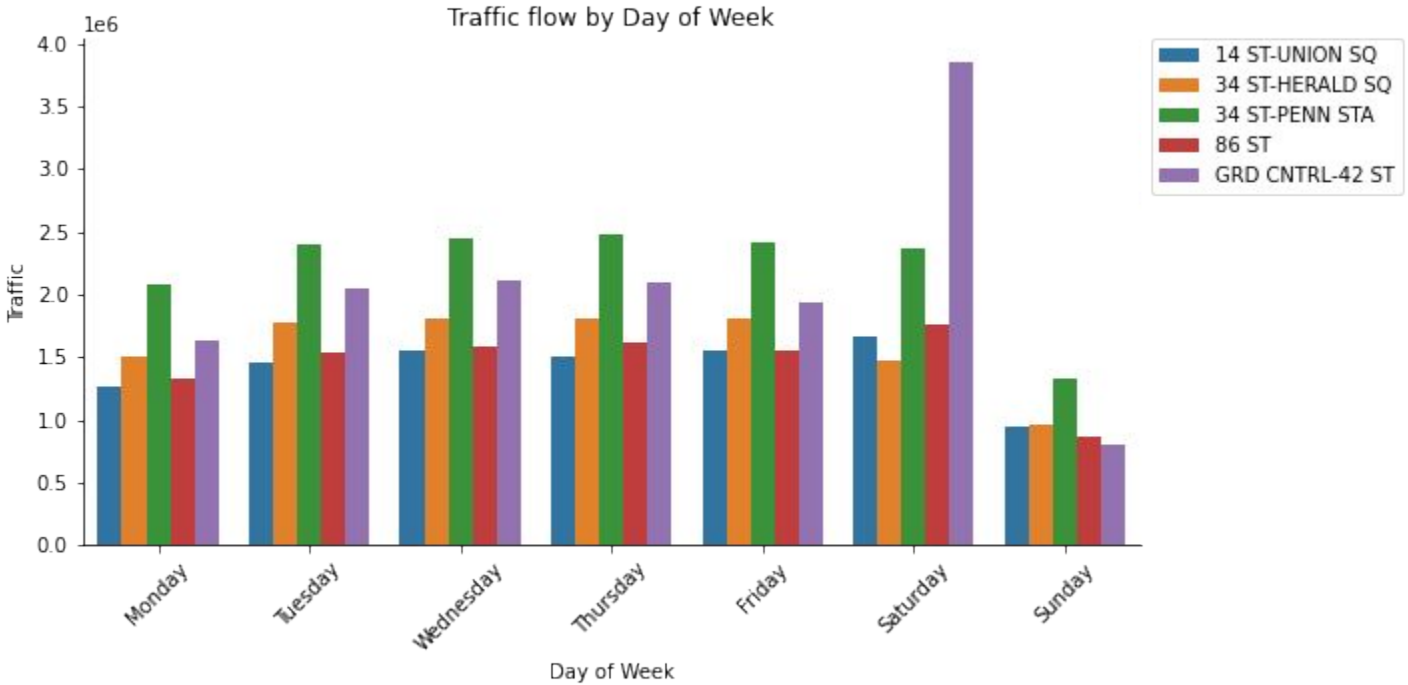
Data Sources:

MTA Data

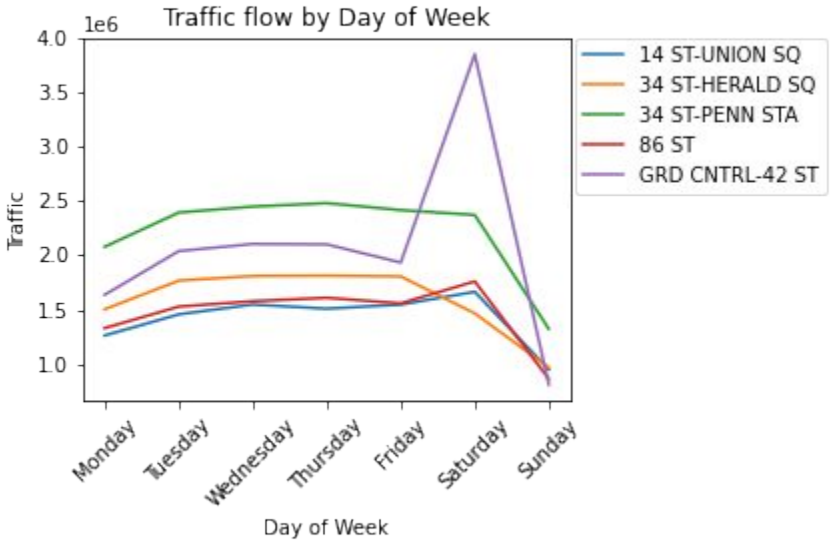
Tools used:

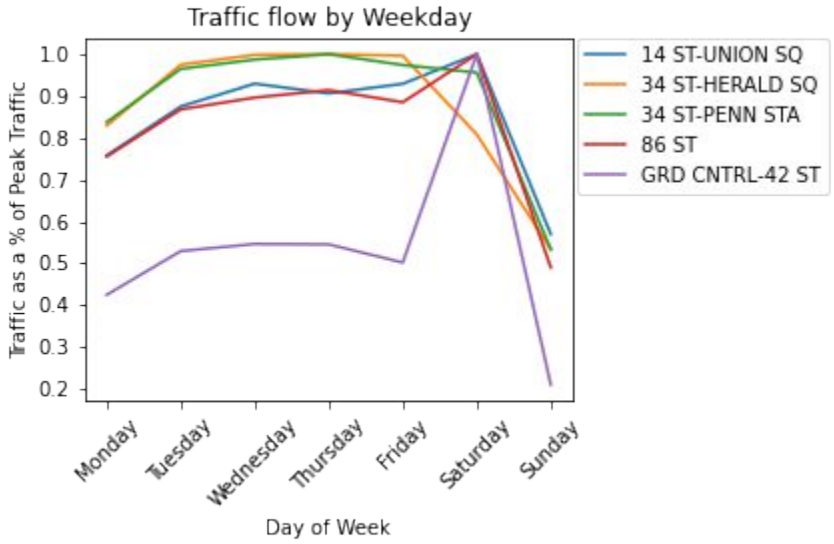
Pandas, Numpy, Matplotlib, Seaborn, Datetime, Dateutil

Results

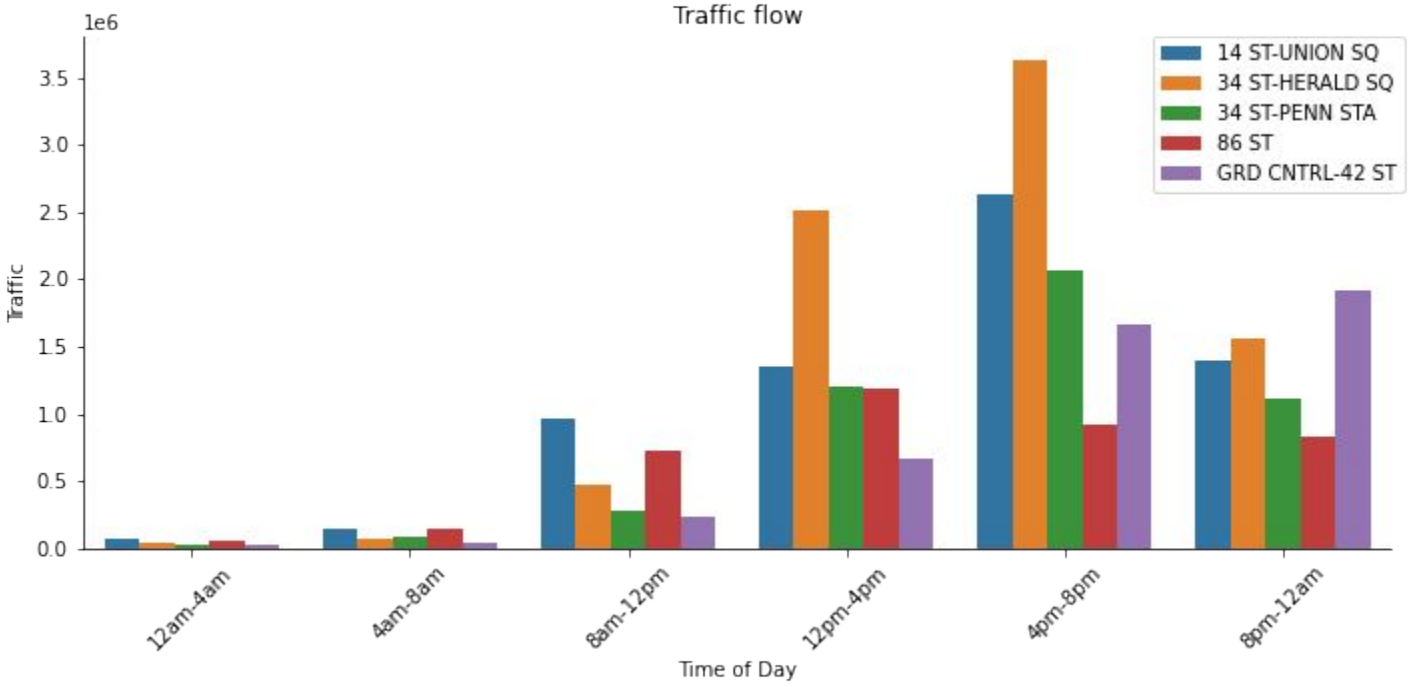

Traffic flow by Day of Week
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Traffic
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week
14 ST-UNION SQ
34 ST-HERALD SQ
34 ST-PENN STA
86 ST
GRD CNTRL-42 ST

Result

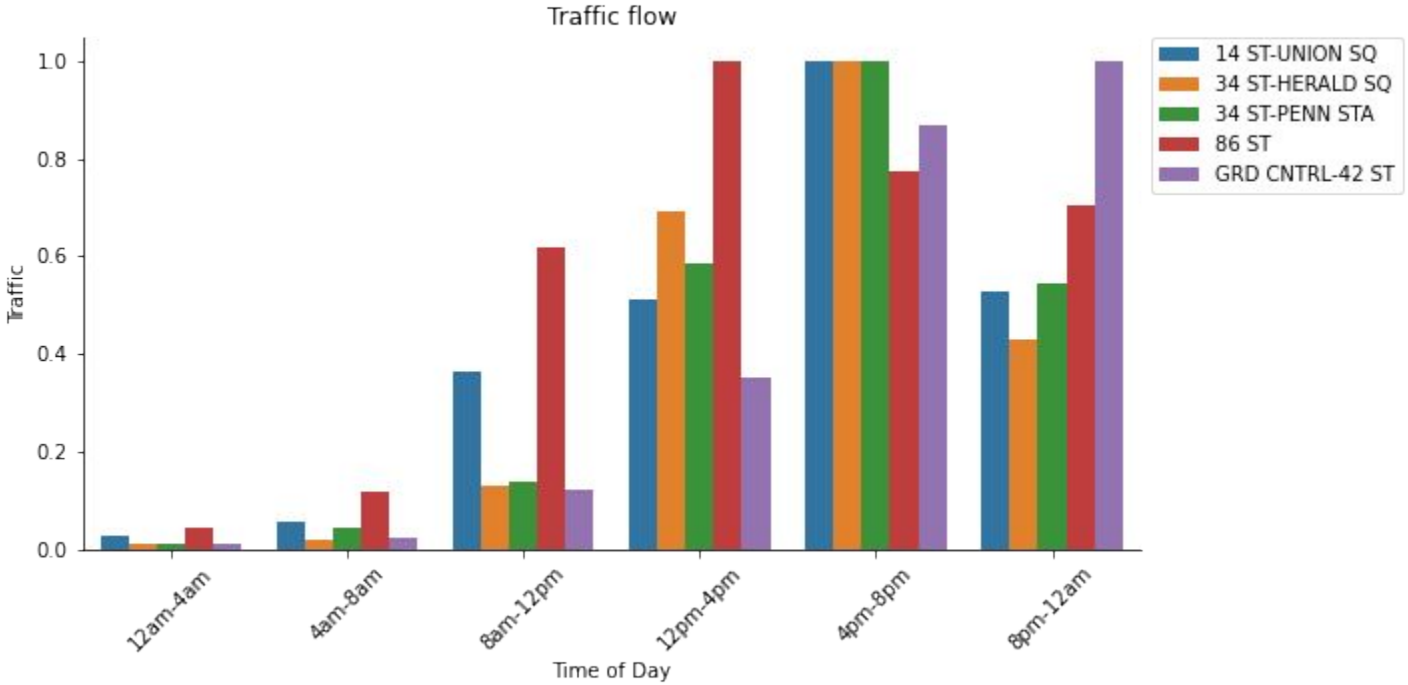

Traffic flow by Day of Week
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Traffic
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week
14 ST-UNION SQ
34 ST-HERALD SQ
34 ST-PENN STA
86 ST
GRD CNTRL-42 ST


Traffic flow by Weekday
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Traffic as a % of Peak Traffic
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week
14 ST-UNION SQ
34 ST-HERALD SQ
34 ST-PENN STA
86 ST
GRD CNTRL-42 ST

Results


Traffic flow
1e6
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Traffic
14 ST-UNION SQ
34 ST-HERALD SQ
34 ST-PENN STA
86 ST
GRD CNTRL-42 ST
12am-4am
4am-8am
8am-12pm
12pm-4pm
4pm-8pm
8pm-12am
Time of Day

Results


Traffic flow
Traffic
1.0
0.8
0.6
0.4
0.2
0.0
12am-4am
4am-8am
8am-12pm
12pm-4pm
4pm-8pm
8pm-12am
Time of Day
14 ST-UNION SQ
34 ST-HERALD SQ
34 ST-PENN STA
86 ST
GRD CNTRL-42 ST

Conclusion

In this sample of data, we see that the lowest foot traffic is universally on sunday which makes the a lot of sense. However we also saw a breakdown of time which concludes that between 12am and 8am is the lowest amount of people. Outside of these times we see different patterns emerge from different stations, some being more busy at certain hours. Thus I would schedule construction to fit where it is lowest in comparison to the peak hours times to reduce the possibility of congestion.

One of these examples is the Grand Central Station, where foot traffic is moderate in the morning to afternoon, however quickly picks up during evening hours. We would tailor the maintenance to each individual station

Future Work

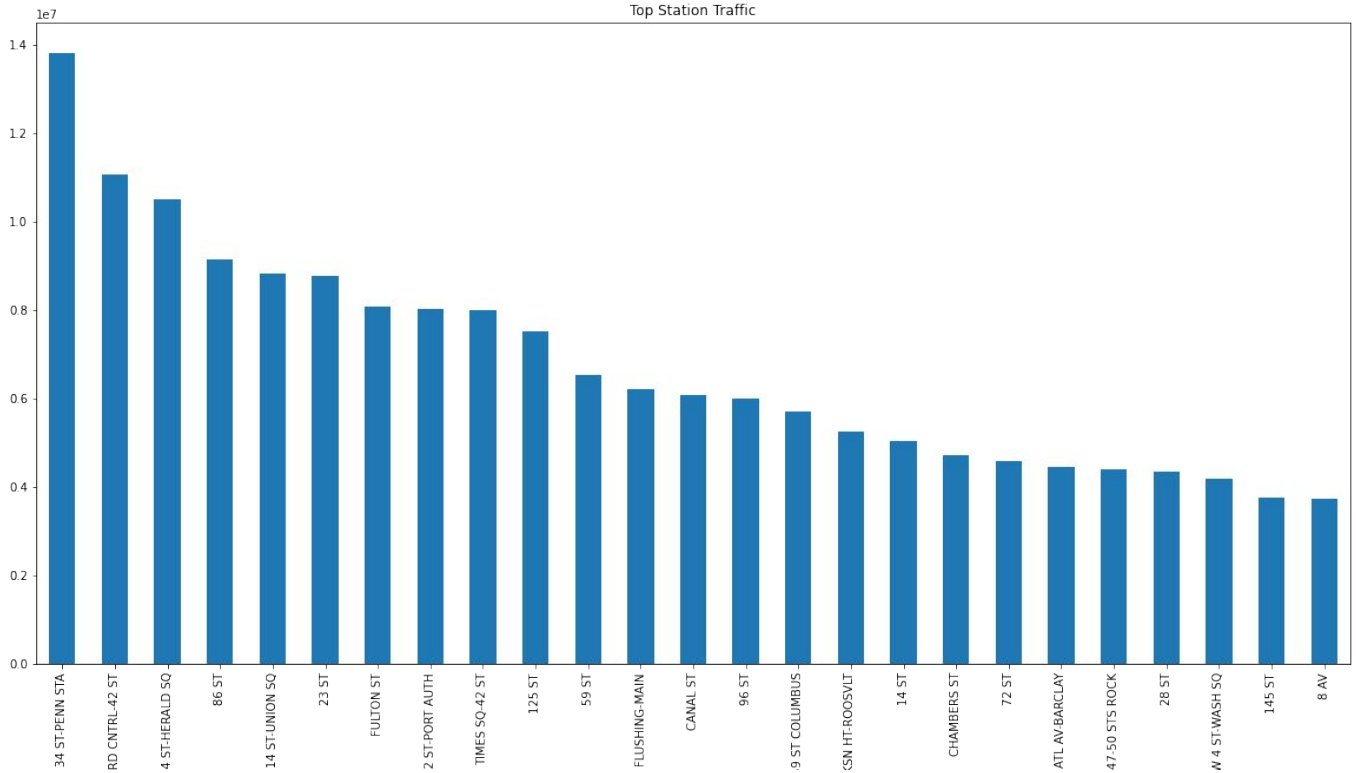
This approach can applied to each station individually to optimize the traffic within the MTA stations. We can also dig deeper into themes for certain months. Our approach assumes a relatively same traffic throughout each month. However, there might be changes due to tourism, university schedules, and a number of other confounding variables.

Another question we would have to address is how much is our budget, and weigh the pros and cons of hiring individuals at night (assuming a hire base pay). Otherwise what is a profit maximizing solution for MTA?

Questions?

Appendix

Time Series Plots - Traffic Totals



Percentage of Peak Calculations

In [45]:
```
top_stations["PEAK_TRAFF"] = top_stations.groupby("STATION")["TOTAL_TRAFF"].transform('max')
top_stations["% of PEAK"] = top_stations['TOTAL_TRAFF']/top_stations['PEAK_TRAFF']
top_stations
```

Out[45]:

	STATION	PAR_TIME	TOTAL_TRAFF	PEAK_TRAFF	% of PEAK
0	14 ST-UNION SQ	2021-09-14 00:00:00	1.389922e+06	2639485.0	0.526588
1	14 ST-UNION SQ	2021-09-14 04:00:00	6.614850e+04	2639485.0	0.025061
2	14 ST-UNION SQ	2021-09-14 08:00:00	1.436126e+05	2639485.0	0.054409
3	14 ST-UNION SQ	2021-09-14 12:00:00	9.565810e+05	2639485.0	0.362412
4	14 ST-UNION SQ	2021-09-14 16:00:00	1.354362e+06	2639485.0	0.513116
5	14 ST-UNION SQ	2021-09-14 20:00:00	2.639485e+06	2639485.0	1.000000
6	34 ST-HERALD SQ	2021-09-14 00:00:00	1.552935e+06	3631797.0	0.427594
7	34 ST-HERALD SQ	2021-09-14 04:00:00	3.207929e+04	3631797.0	0.008833
8	34 ST-HERALD SQ	2021-09-14 08:00:00	6.865500e+04	3631797.0	0.018904
9	34 ST-HERALD SQ	2021-09-14 12:00:00	4.766176e+05	3631797.0	0.131235
10	34 ST-HERALD SQ	2021-09-14 16:00:00	2.514948e+06	3631797.0	0.692480