



MTA TRAFFIC ANALYSIS FOR SUBWAY STATION ACCESSIBILITY UPGRADE

Presented by Crystal Huang
Metis Data Science Bootcamp
Project I

INTRODUCTION

- MTA and Department of City Planning are inviting private businesses:
 - subway stations accessibility upgrades
 - Increase “Americans with Disabilities Act-compliant” (ADA) stations
- Lift Inc. (fictional elevator construction company)
Stations for upgrades

Prioritize the higher traffic stations for the upgrades

Employee schedule

lower traffic days and times at those stations to minimize exposure and risk during pandemic

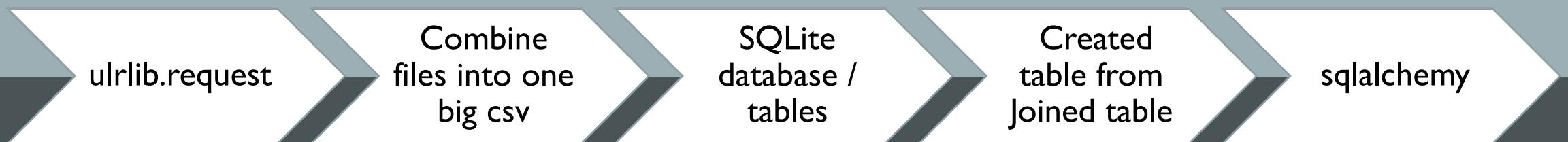
MTA Looking for Private Business Help With Subway Accessibility Upgrades

BY JOSE MARTINEZ | MAR 17, 2021, 7:17PM EDT



METHODOLOGY – DATA COLLECTION

- Turnstile data during pandemic when NYC WFH order started – from 03/21/2020 to 03/20/2021
- MTA Stations info data – contains ADA info of each MTA station



	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/21/2020	00:00:00	REGULAR	0007411940	0002515962 ...
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/21/2020	04:00:00	REGULAR	0007411942	0002515966 ...
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/21/2020	08:00:00	REGULAR	0007411945	0002515979 ...
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/21/2020	12:00:00	REGULAR	0007411969	0002516000 ...
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/21/2020	16:00:00	REGULAR	0007412028	0002516024 ...

10,858,718 rows/ data points

METHODOLOGY – DATA PREPARATION

Problem – counter resets or counts backward

- Traffic count at each time period may contain error data

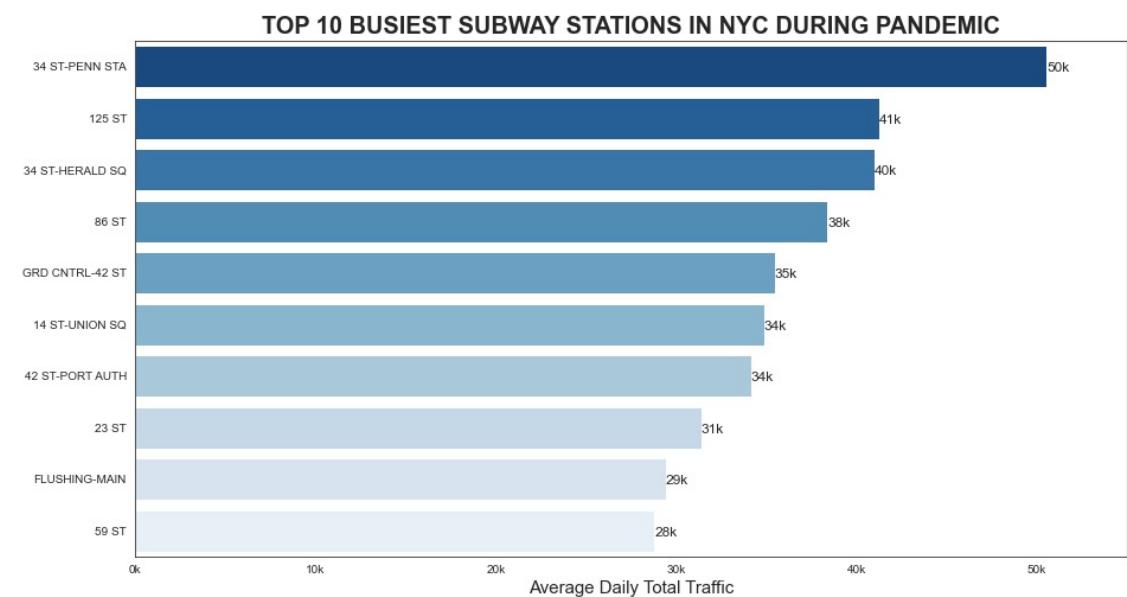
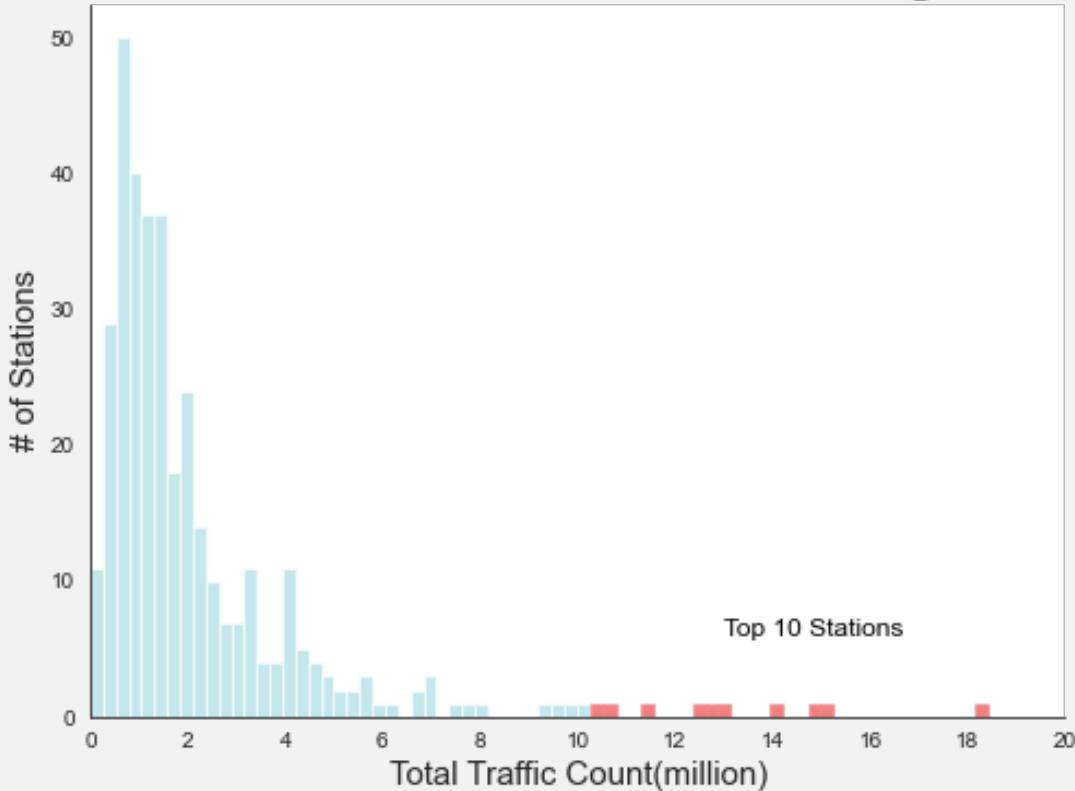
Error Data	Negative counts	Extremely high counts (outliers)
Method	$ (\text{Entry/Exit}) - (\text{previous Entry/Exit}) $	Traffic count > 14,400 (total seconds in 4-hour interval) Replace the outliers with mean values of the turnstile before and after the error data datetime

Total traffic at time period = sum of entry difference and exit difference

RESULTS

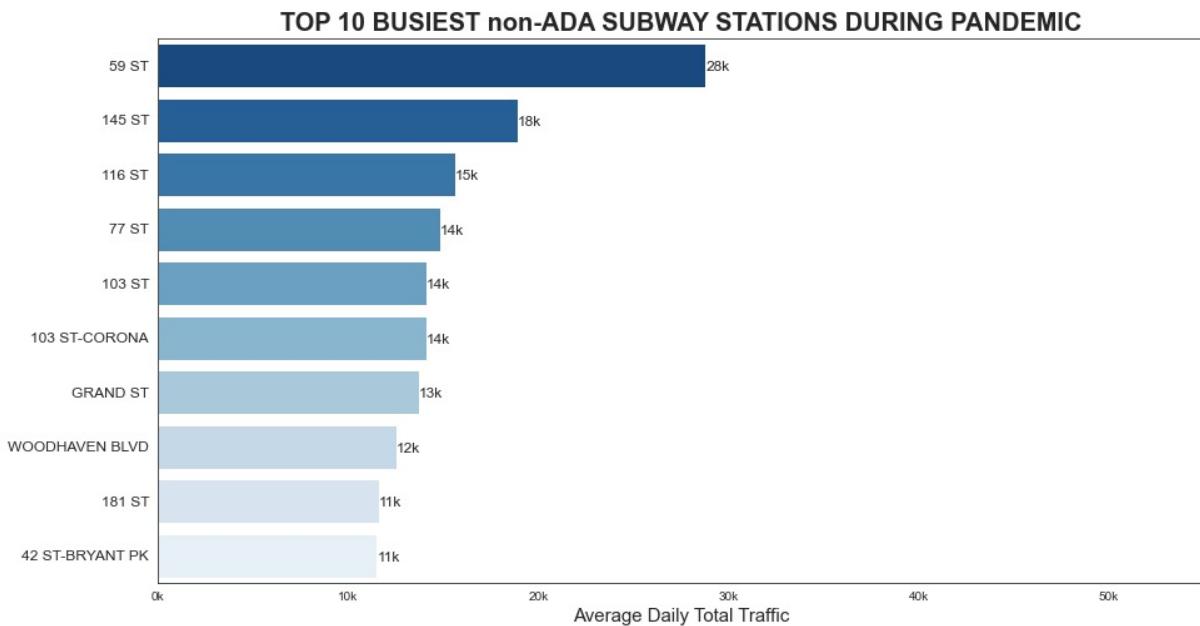
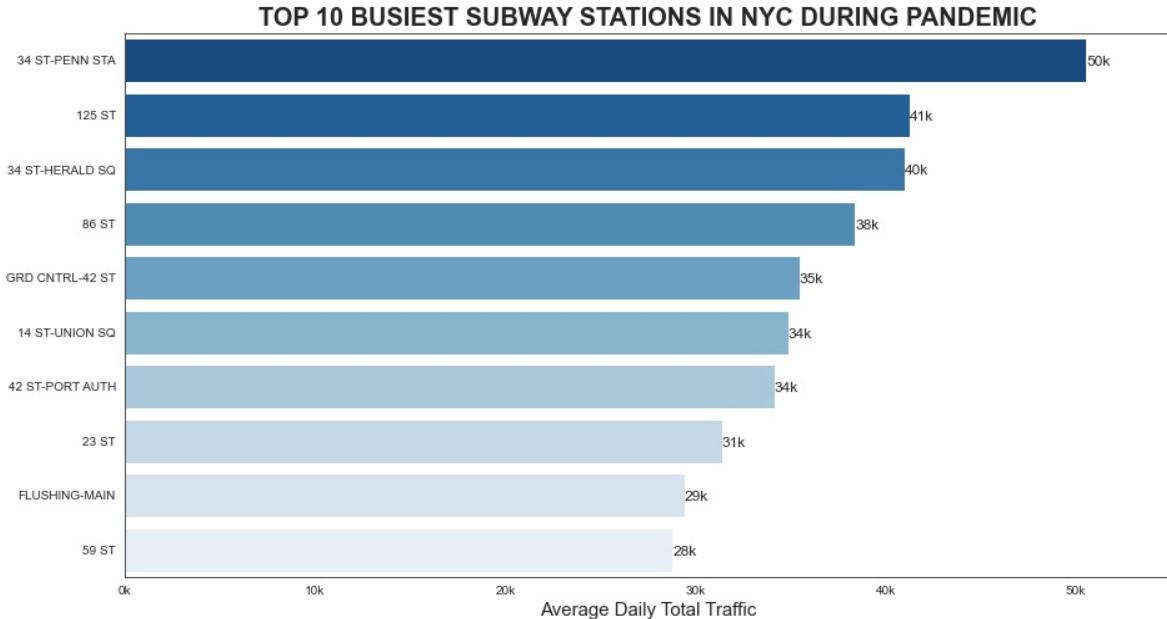
- Distribution of traffic across station is heavily right-skewed
- Top 10 stations – outliers in distributions
- 34 St – Penn Station has notably more traffic than the rest

Distribution of Traffic Across Stations during Pandemic



RESULTS

- Goal 2: To increase # of ADA stations
 - Focus on stations that are not ADA stations
- Distribution of traffic on non-ADA stations
 - Identify top 10 non-ADA stations
- Original top 10 vs. non-ADA top 10
 - Non-ADA stations have notably less traffic

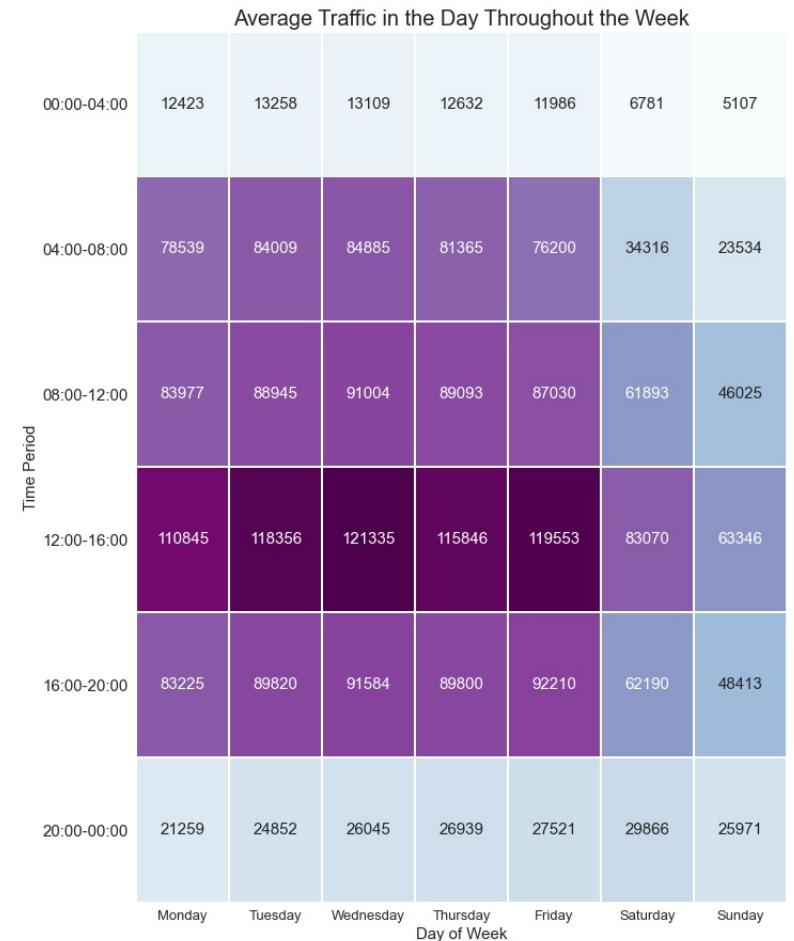
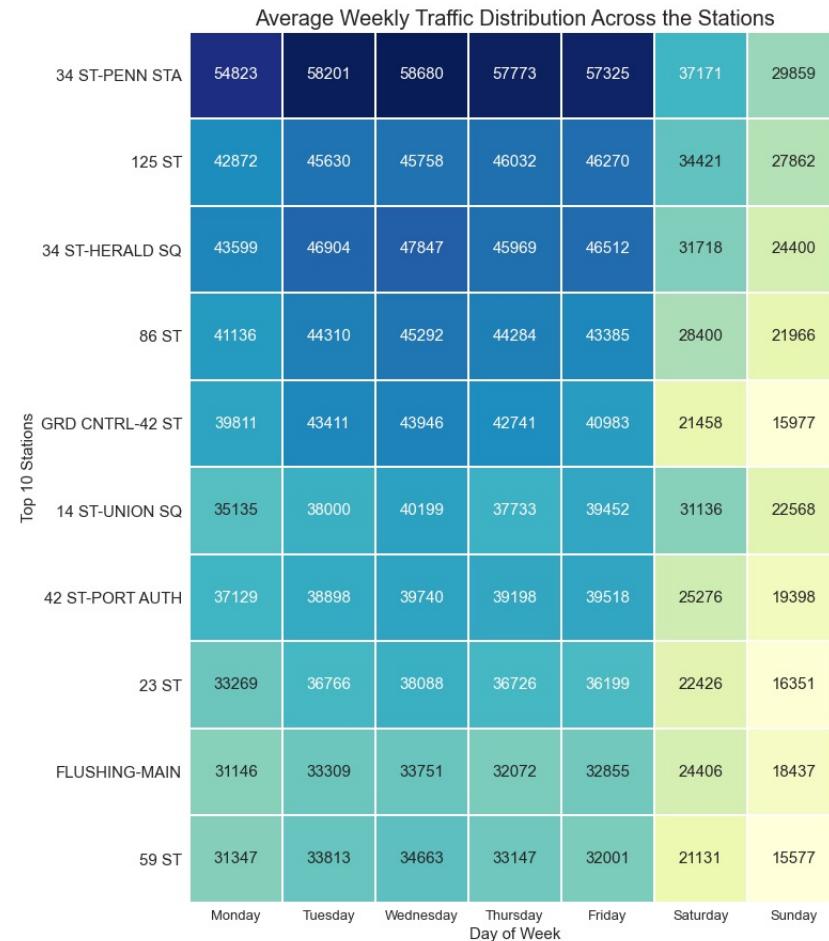


- Distribution of traffic across the week
- traffic on weekdays > weekends

RESULTS

- No morning rush hour crowd
- Traffic is heavier in afternoon

Top 10 Stations Traffic pattern



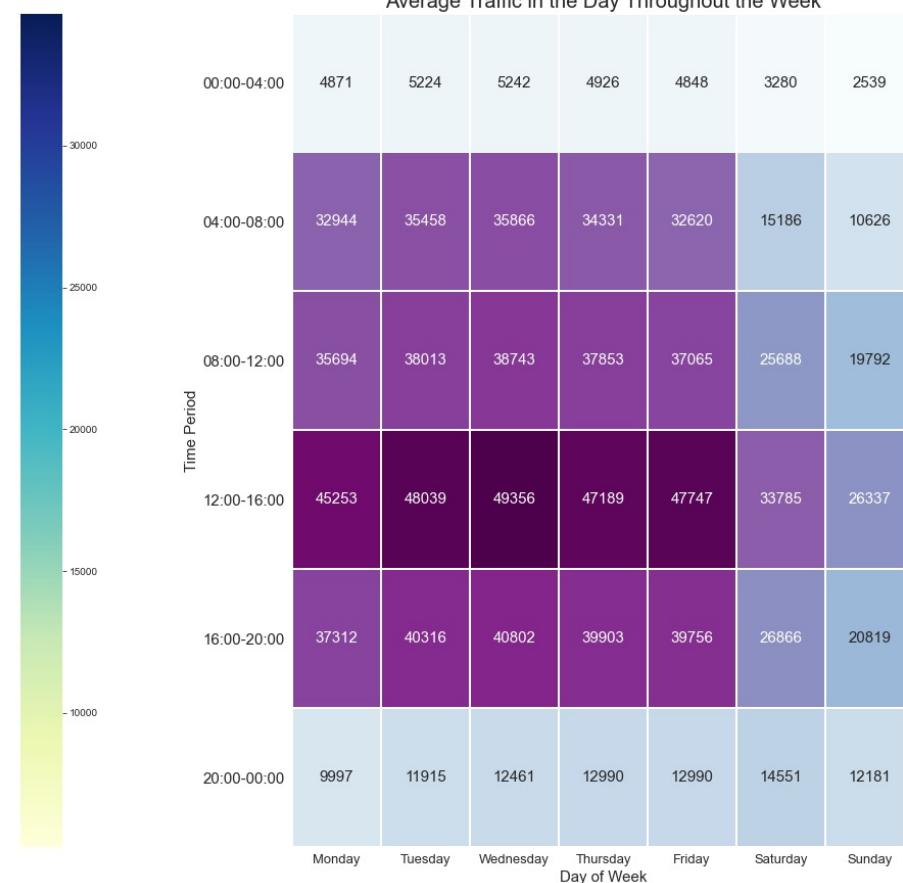
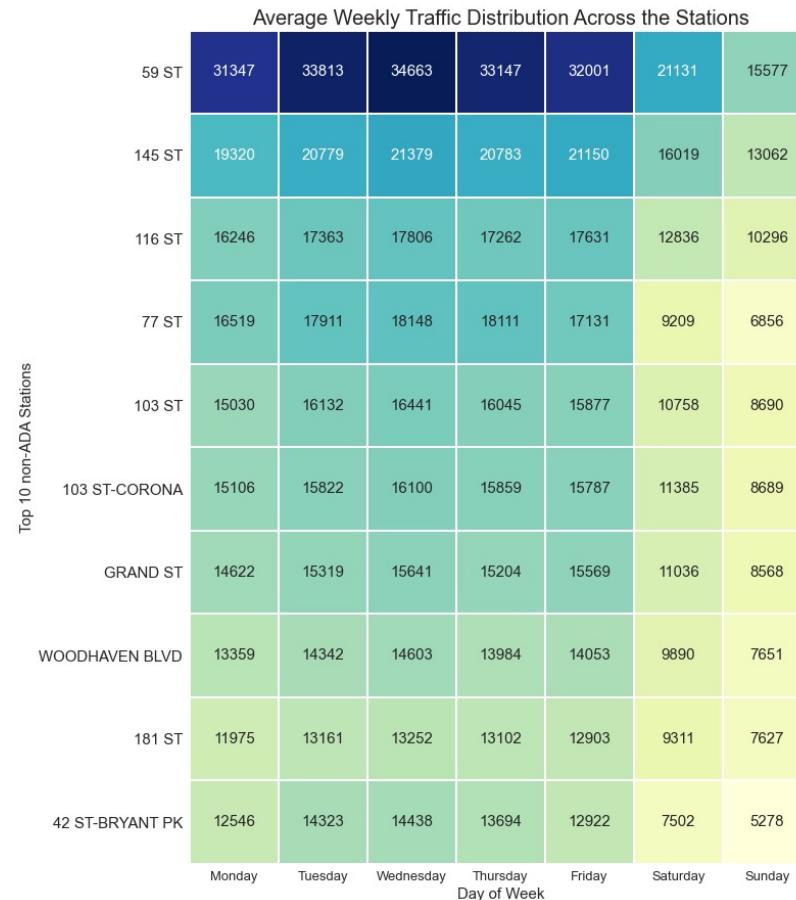
RESULTS

- Lower traffic on weekends vs. weekdays

Similar traffic pattern as original top 10 stations

- Lower traffic in
 - night shift throughout the week
 - early morning weekends

Top 10 non-ADA Stations Traffic pattern



CONCLUSION

Which stations have higher traffic?

Overall accessibility upgrades

- Top 10 busiest Stations

	STATION	TOTAL_TRAFFIC
0	34 ST-PENN STA	50547.0
1	125 ST	41264.0
2	34 ST-HERALD SQ	40993.0
3	86 ST	38396.0
4	GRD CNTRL-42 ST	35475.0
5	14 ST-UNION SQ	34889.0
6	42 ST-PORT AUTH	34165.0
7	23 ST	31403.0
8	FLUSHING-MAIN	29425.0
9	59 ST	28811.0

Increase # of ADA Stations

- Top 10 busiest non-ADA Stations

	STATION	TOTAL_TRAFFIC
0	59 ST	28811.0
1	145 ST	18928.0
2	116 ST	15634.0
3	77 ST	14841.0
4	103 ST	14139.0
5	103 ST-CORONA	14107.0
6	GRAND ST	13708.0
7	WOODHAVEN BLVD	12554.0
8	181 ST	11619.0
9	42 ST-BRYANT PK	11529.0

CONCLUSION

What are the lower traffic days and times at those stations?

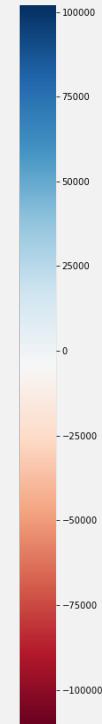
- Recommended employee schedule:
 - Days – weekends
 - Times – early morning or night

FUTURE WORK

- Residential vs Commercial area
 - Net Entry/Exit at each station during each time period
- GeoMapping –
 - Stations vs. Nursing facilities/hospitals
 - Disability Population
- Other visualizations such as interactive plots
 - Plotly, [Bokeh](#)

APPENDIX

Net Entry/Exit of Commuters



APPENDIX

```
# Functions for entry, exit counts, and traffic counts
def get_entry_counts(row, max_counter):
    counter = abs(row['ENTRIES'] - row['PREV_ENTRIES'])
    # Set anomaly values due to reset of counters to the uniform NaN values
    if counter > max_counter:
        counter = np.nan
    return counter

def get_exit_counts(row, max_counter):
    counter = abs(row['EXITS'] - row['PREV_EXITS'])
    # Set anomaly values due to reset of counters to the uniform NaN values
    if counter > max_counter:
        counter = np.nan
    return counter

def get_counts(df):
    # Set max_counter = 14400, assuming only 1/person/sec/turnstile at a time in a 4HR interval
    # Entry count
    df['entry_count'] = df.apply(get_entry_counts, axis=1, max_counter=14400)

    # Exit count
    df['exit_count'] = df.apply(get_exit_counts, axis=1, max_counter=14400)

return df
```

APPENDIX

```
# Fill NaN values with the mean of values before and after NaN value
# AKA take the mean of the daily count

entry_list = list(df['entry_count'])
ind = 0
for i in entry_list:
    if np.isnan(i) == 1:
        entry_list[ind] = np.nanmean([entry_list[ind-2],entry_list[ind-1],entry_list[ind+1],entry_list[ind+2]])
    ind += 1

df['ENTRY_DIFF'] = entry_list

df['ENTRY_DIFF'].isna().sum()

0
```



THANK YOU