



The Relationship between MTA Traffic and Crime Rate

Presented by: Amena Alsamah

The Steps To Find the Relationship

Introduction:

- The problem.
- The Goal

Methods:

- Data cleaning and preparation.
- Analyzing and finding patterns

Result:

- Data interpretation.
- Future Recommendation



How all this started

- **The sudden increase of crime rate**
After the pandemic.
- **People worried about their safety**
That affect their decision of using public transportation.
- **NYPD and MTA have to take action**
Finding the busiest station from MTA subway data and the most dangerous station from NYPD data.

Methods:

1-Choose the period of time that the analysis will take place which is Oct 03, 2020 to Dec 26, 2020

2- Upload both datasets.

3- Convert the daily report to weekly in the NYPD.

4-Cleaning the data, dropping duplicates and imputing any missing value.



| | C/A | UNIT | SCP | STATION | LINENAME | DIVISION | DATE | TIME | DESC | ENTRIES | EXITS |
|---|------|------|----------|---------|----------|----------|------------|----------|---------|---------|---------|
| 0 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 09/26/2020 | 00:00:00 | REGULAR | 7463495 | 2538694 |
| 1 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 09/26/2020 | 04:00:00 | REGULAR | 7463500 | 2538697 |
| 2 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 09/26/2020 | 08:00:00 | REGULAR | 7463516 | 2538712 |
| 3 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 09/26/2020 | 12:00:00 | REGULAR | 7463552 | 2538767 |
| 4 | A002 | R051 | 02-00-00 | 59 ST | NQR456W | BMT | 09/26/2020 | 16:00:00 | REGULAR | 7463664 | 2538795 |

Methods:

- The sum of entries were taken for each station, taking into account the uniqueness of each of the " C/A, UNIT, and SCP" for each station.



Note:

Since the Entries value were cumulative, an extra step was needed to find the difference between each entry for each hour. To find the actual number

Methods:

-Finding the busiest station was calculated from the sum of all the entries for each day of the period between Oct to Dec in each and every station.

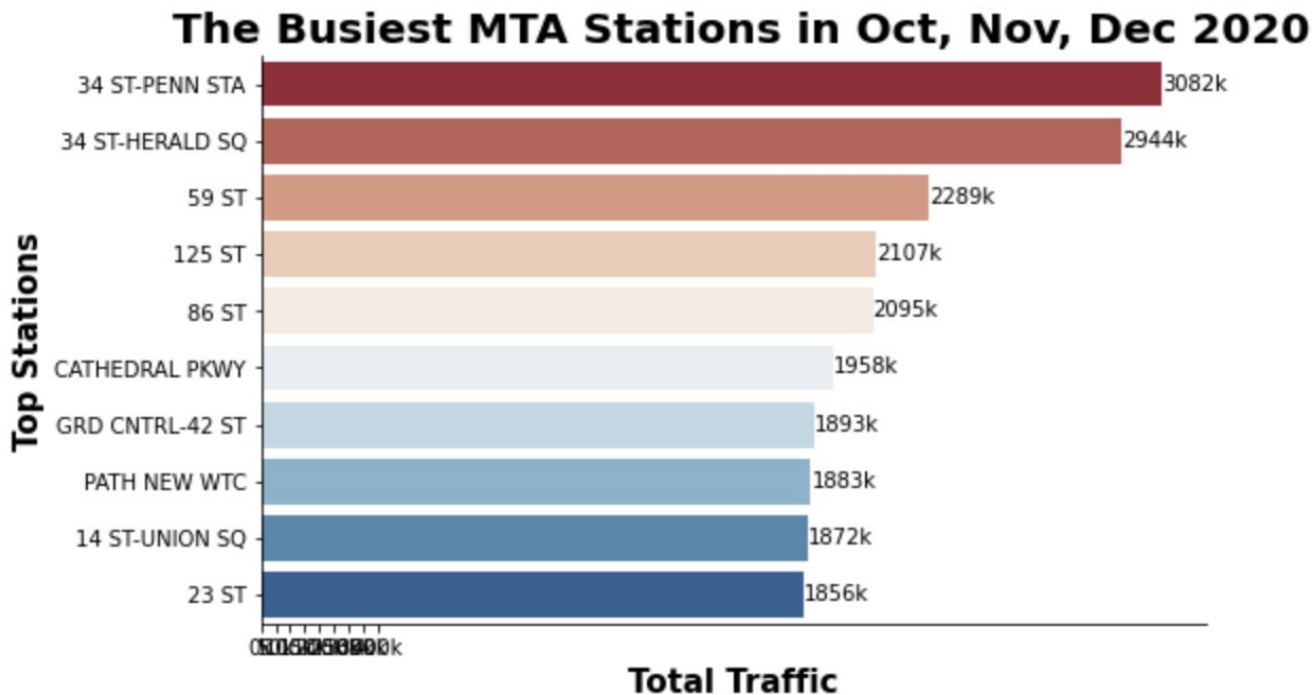
-At the end Sorting them and taking the largest 10 station in regard of traffic.

```
station_totals.head(10)
```

| | STATION | DAILY_ENTRIES |
|---|-----------------|---------------|
| 0 | 34 ST-PENN STA | 3082301.0 |
| 1 | 34 ST-HERALD SQ | 2944189.0 |
| 2 | 59 ST | 2289213.0 |
| 3 | 125 ST | 2107892.0 |
| 4 | 86 ST | 2095087.0 |
| 5 | CATHEDRAL PKWY | 1958771.0 |
| 6 | GRD CNTRL-42 ST | 1893006.0 |
| 7 | PATH NEW WTC | 1883282.0 |
| 8 | 14 ST-UNION SQ | 1872849.0 |
| 9 | 23 ST | 1856797.0 |

— Findings:

From MTA database, 34-ST-PENN and 34 ST-HERALD SQ stations were the busiest with a significant difference from the other stations.



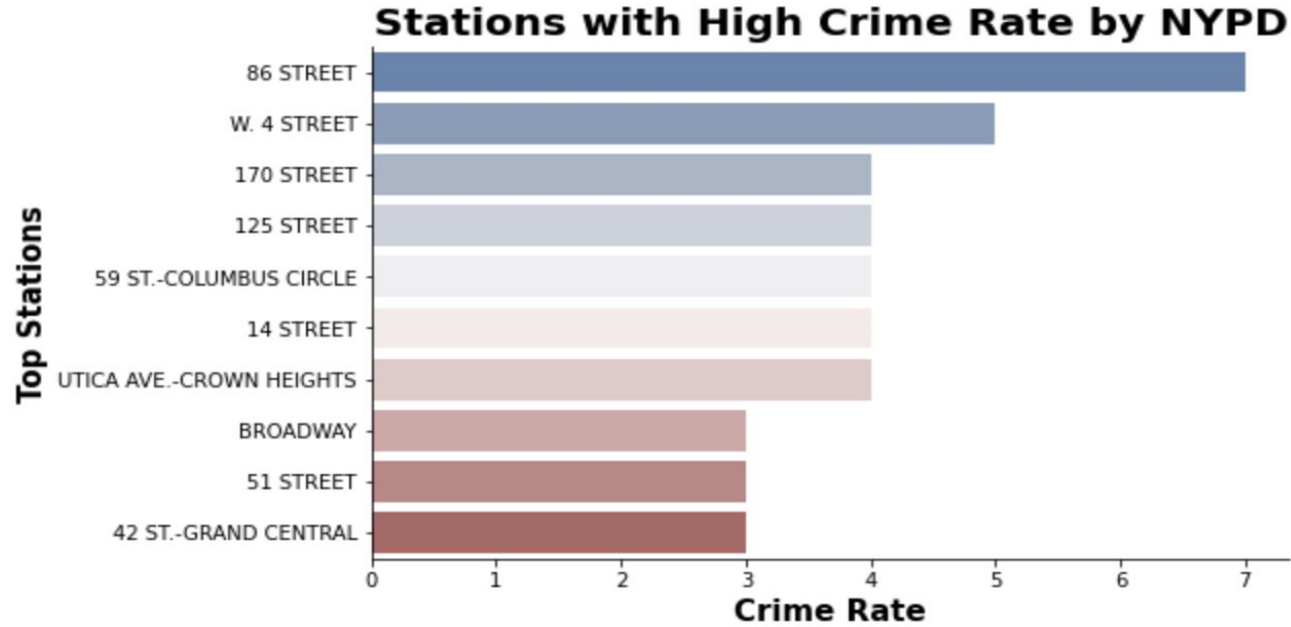
Methods:

From the NYPD report
taking the sum of complaints
for each station for everyday
in a period of 3 months.

crime_rate_daily

| | KY_CD | STATION_NAME | PD_CD | CMPLNT_NUM |
|-----|-------|-----------------|-------|------------|
| 0 | 578 | W. 4 STREET | 638.0 | 102802571 |
| 1 | 361 | CHURCH AVENUE | 661.0 | 102939554 |
| 2 | 109 | 59 STREET | 406.0 | 108244259 |
| 3 | 344 | ATLANTIC AVENUE | 101.0 | 113475424 |
| 4 | 344 | 86TH STREET | 101.0 | 114122384 |
| ... | ... | ... | ... | ... |
| 151 | 113 | 170 STREET | 729.0 | 952342725 |
| 152 | 344 | 145 STREET | 101.0 | 962935226 |
| 153 | 361 | 33 STREET | 661.0 | 967400024 |
| 154 | 678 | 42 STREET | 633.0 | 984437897 |
| 155 | 678 | 42 STREET | 633.0 | 986928304 |

156 rows x 4 columns



Findings:

86 Street had the highest crime with a noticeable difference among other stations.

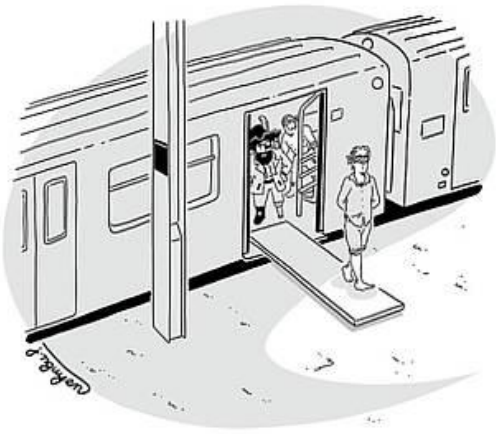
Conclusion

The hypothesis was that there is a correlation between the rate of crime and how busy that station was at that day.

The data did not support the hypothesis and that is due to:

- Huge number of missing values.
- The period of time was not long enough to generate bigger data to get better result.

| | ENTRIES | EXITS | CMPLNT_NUM | ADDR_PCT_CD | KY_CD | PD_CD | JURISDICTION_CODE |
|-------------------|----------|----------|------------|-------------|-----------|-----------|-------------------|
| ENTRIES | 1.000000 | 0.782164 | NaN | NaN | NaN | NaN | NaN |
| EXITS | 0.782164 | 1.000000 | NaN | NaN | NaN | NaN | NaN |
| CMPLNT_NUM | NaN | NaN | 1.000000 | 0.014358 | 0.002088 | -0.005157 | -0.012437 |
| ADDR_PCT_CD | NaN | NaN | 0.014358 | 1.000000 | 0.040324 | 0.020254 | -0.015939 |
| KY_CD | NaN | NaN | 0.002088 | 0.040324 | 1.000000 | 0.401678 | 0.000645 |
| PD_CD | NaN | NaN | -0.005157 | 0.020254 | 0.401678 | 1.000000 | -0.017984 |
| JURISDICTION_CODE | NaN | NaN | -0.012437 | -0.015939 | 0.000645 | -0.017984 | 1.000000 |
| HOUSING_PSA | NaN | NaN | -0.020950 | 0.080609 | -0.032568 | -0.057611 | -0.061489 |
| X_COORD_CD | NaN | NaN | 0.012113 | 0.239656 | -0.012341 | -0.017921 | -0.021532 |
| Y_COORD_CD | NaN | NaN | -0.007989 | -0.476645 | -0.021951 | -0.011118 | -0.016547 |
| TRANSIT_DISTRICT | NaN | NaN | 0.007699 | 0.851740 | -0.150349 | -0.103737 | NaN |
| Latitude | NaN | NaN | -0.008408 | -0.477040 | -0.022053 | -0.010761 | -0.016603 |
| Longitude | NaN | NaN | 0.012033 | 0.239473 | -0.012303 | -0.017810 | -0.021561 |



After finding out those results, an action toward this issue is still required, training the regular commuters at those stations and hire more cops.

Also, further investigation will help solve the root of the issue



**Future
Recommendations**