NYC Crime with MTA Station Traffic

The Problem :

Fear of crime and harassment in New York City subways is keeping many New Yorkers out of the system,The subway ridership in NYC looking for secure station to use away from crime.

Our goal in this work to capture NYC subway station entry/exit data along with crime statistics around the stations, then determine the correlation between them.

Data collection:

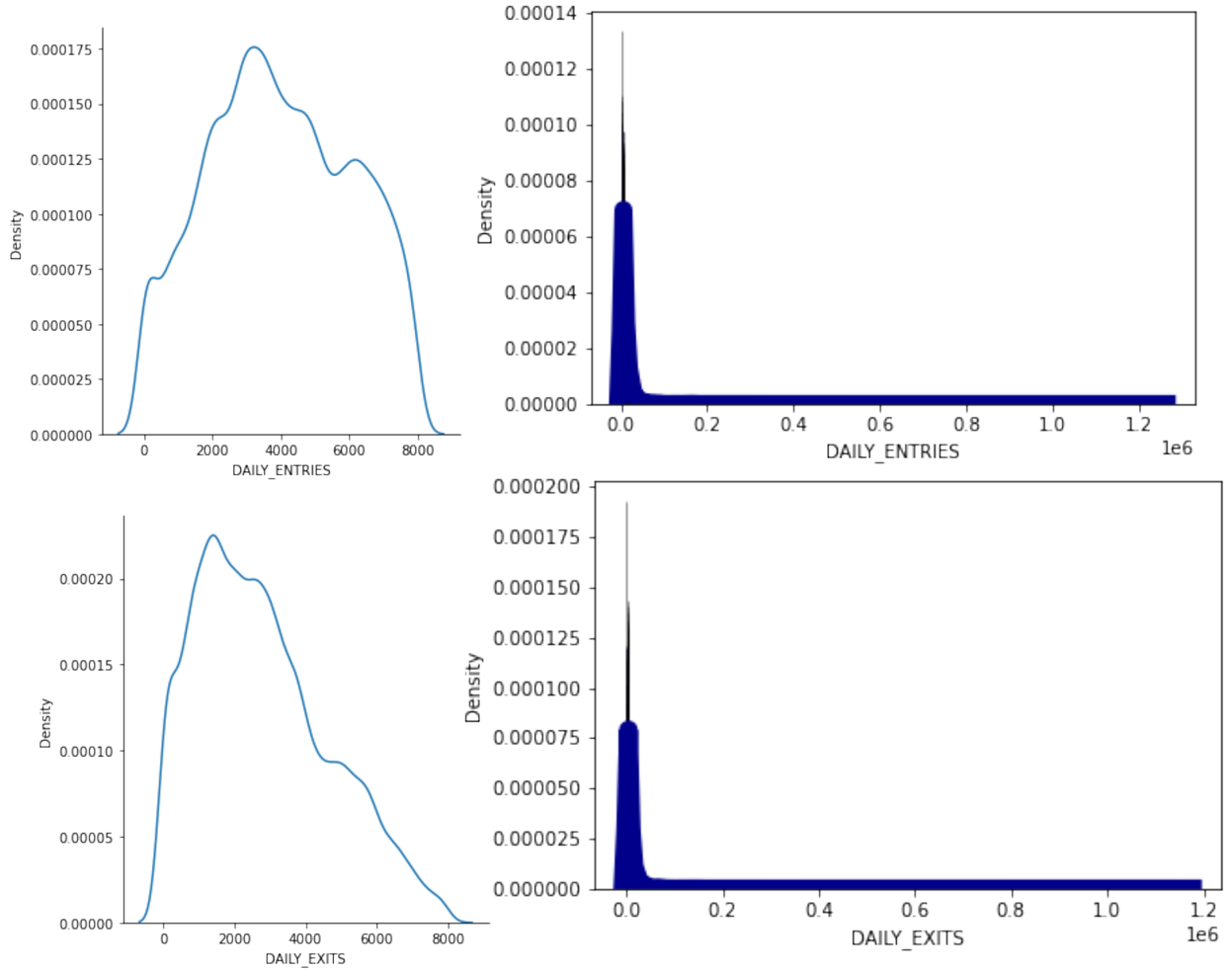
1-MTA DATA use data from January to may,2019

2-LOCATION DATA was used to fill in each station’s location coordinates (latitude/longitude) using station name

3-CRIME DATA Merge the above for sampling target, then further append longitude and latitude columns.

Data cleaning:

- Subway data was cleaned using the following steps:
- renaming columns, adding calculated fields, removing erroneous data and duplicate data
- Aggregate entry and exit data per station per date.
- After plotting daily enters and exits we removed all values of entries and exits greater than 8,000(outliers cut off point >8000)



Data Analysis

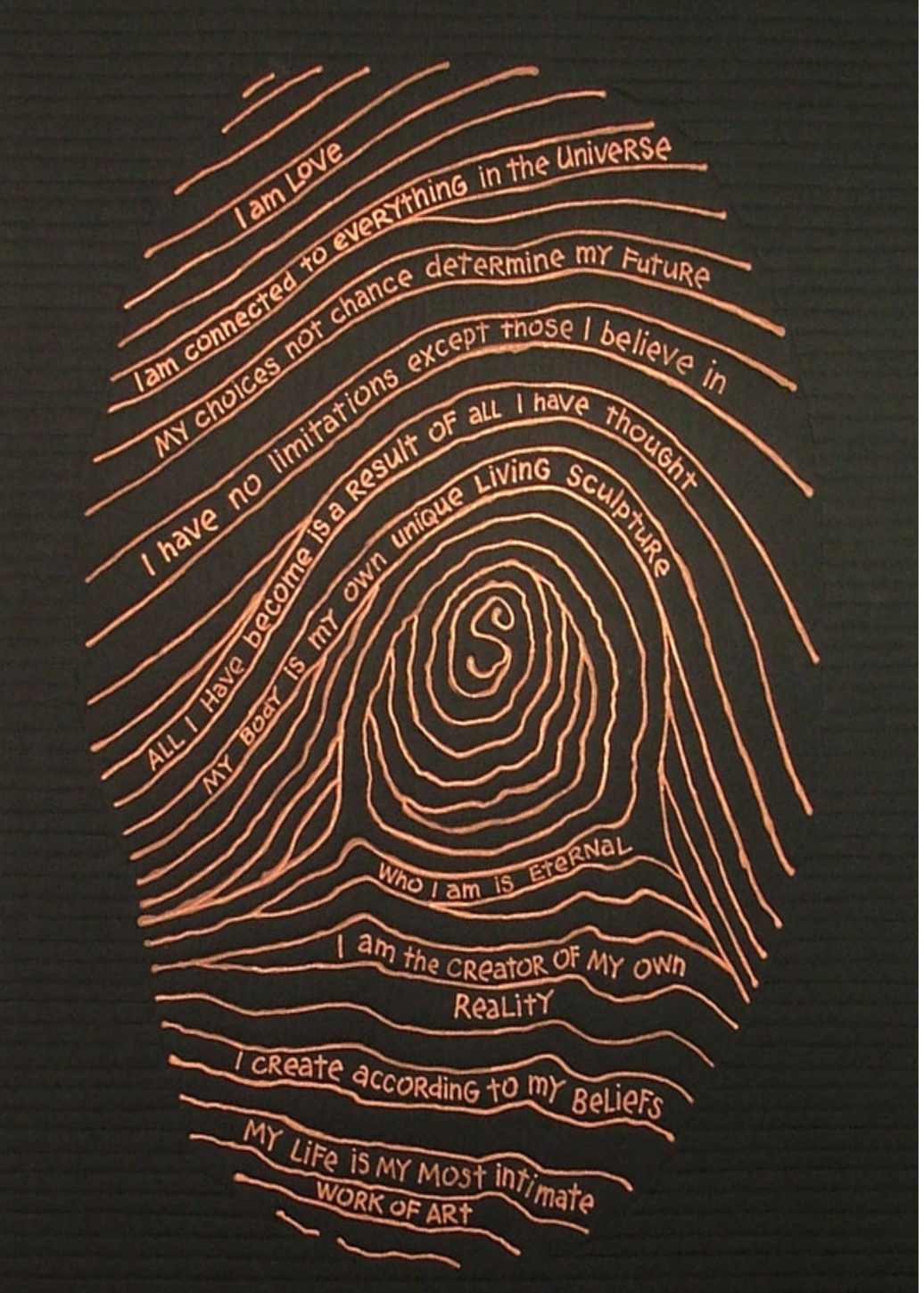
Data analysis

Motivation:

Explore correlation between MTA station traffic and adjacent crime statistics.

Tools:

1) Python to ingest and aggregate the raw data into a SQL database
2) SQLAlchemy to query resulting database into Python
3) Pandas for exploratory data analysis and cleaning
4) MatPlotLib to plot data for correlation identification
5) Seaborn, Plotly., and MatPlotLib to visualize station and crime locations using geo coordinates



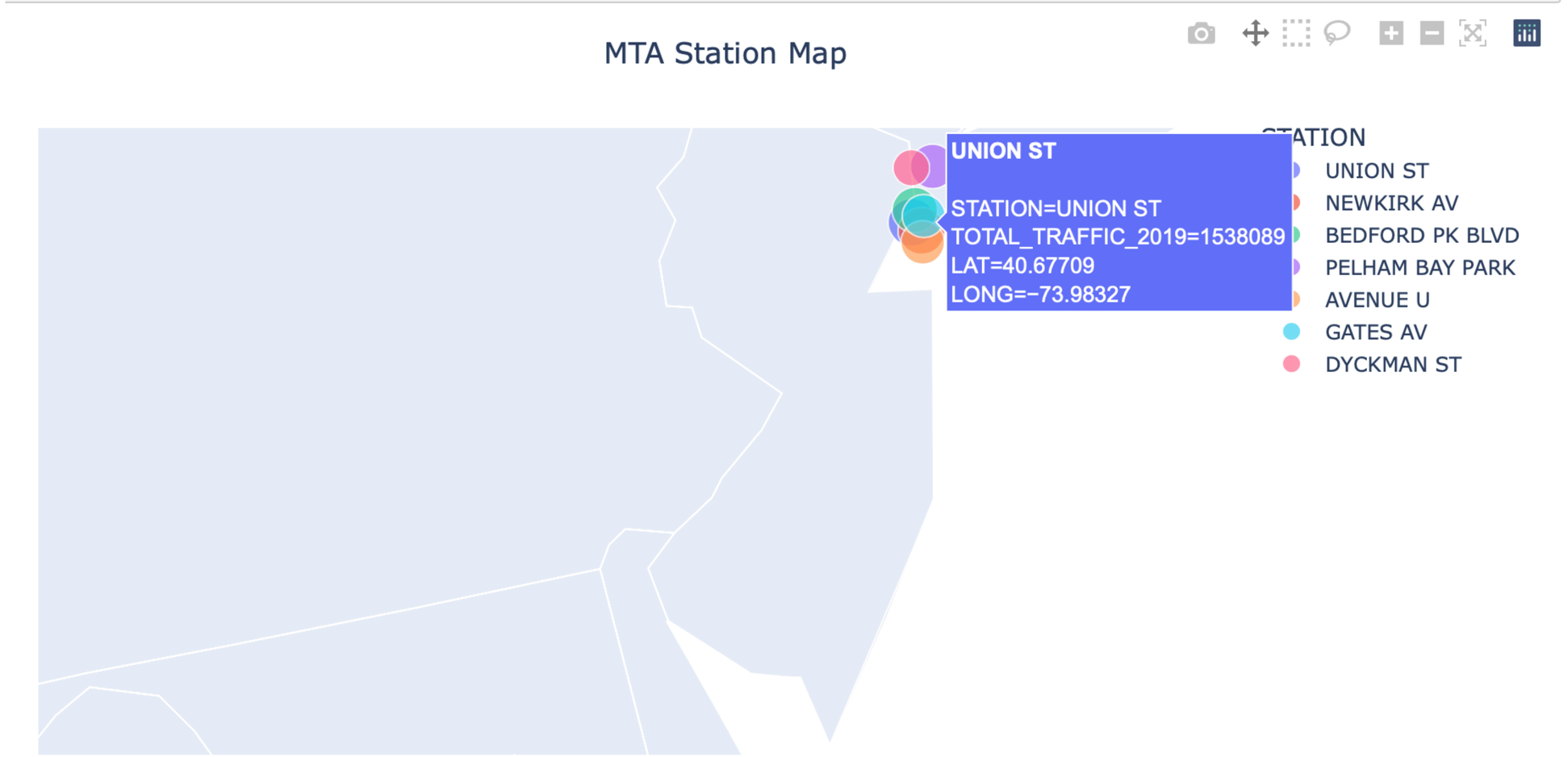
Data analysis

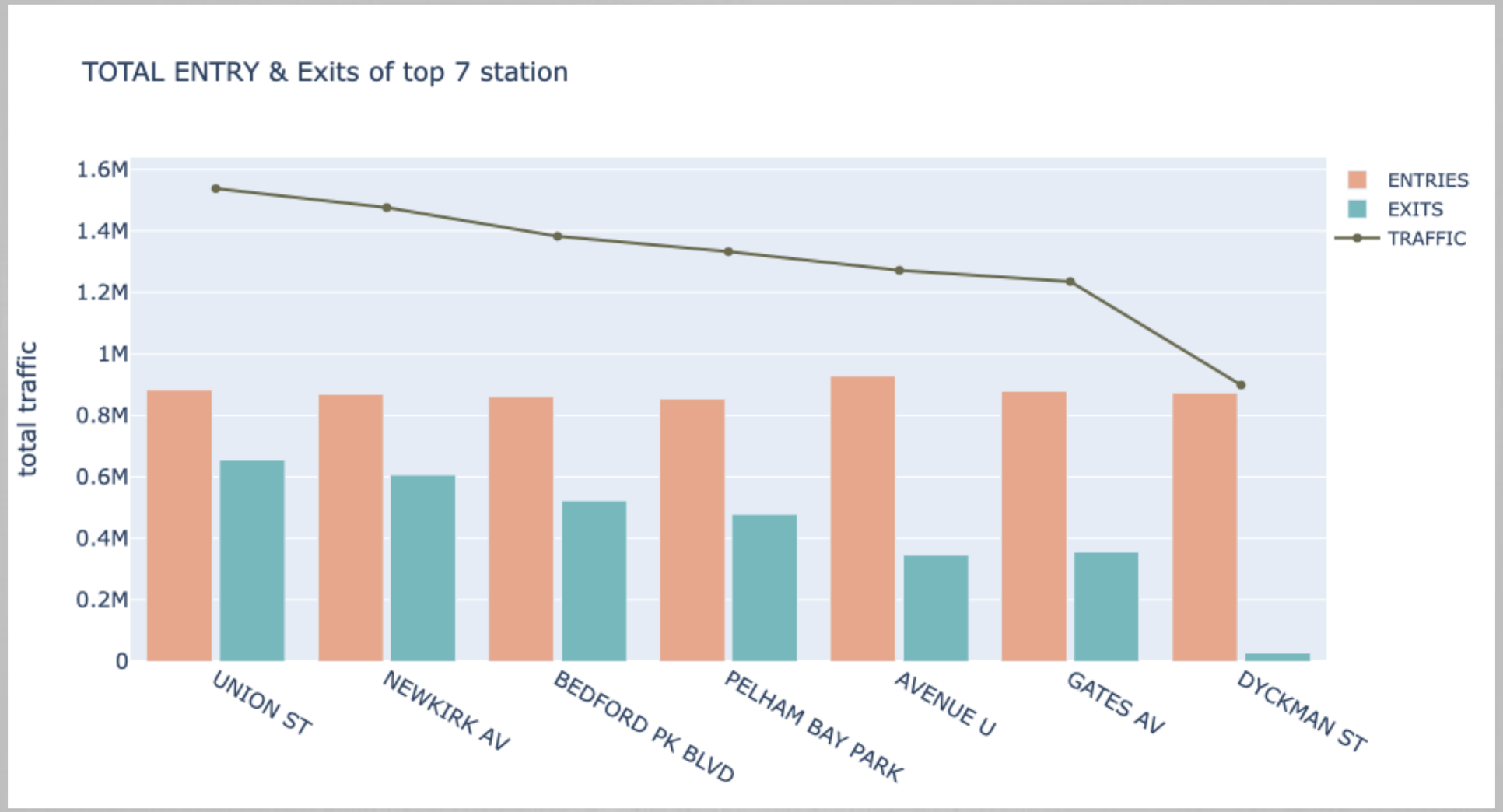
data frames for (2019) MTA ridership And crime data:

1) Concatenate the top entry and exit data frames to obtain Top 7 stations by total traffic.

2) Append the crime count to the original station data frame to obtain Top 7 stations by number of crime

Data analysis

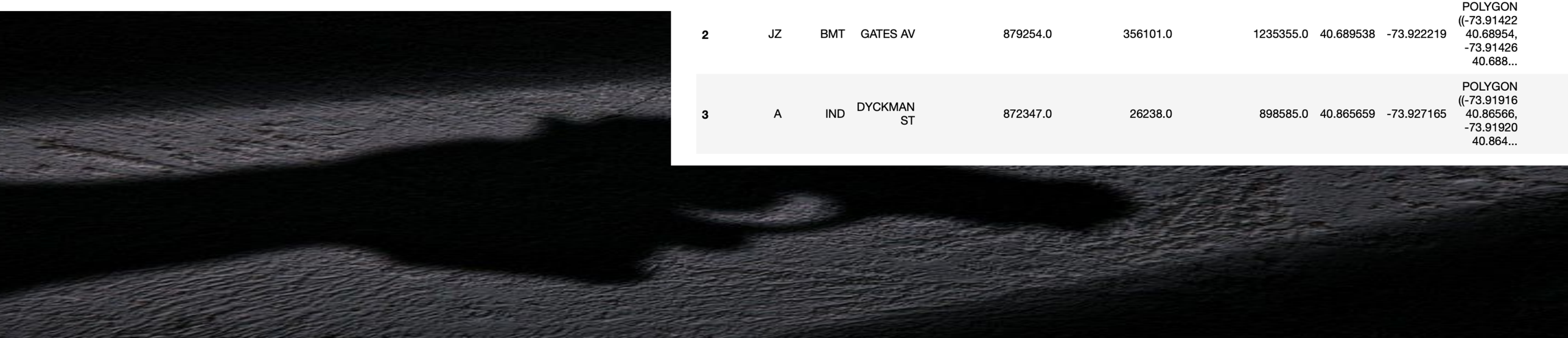


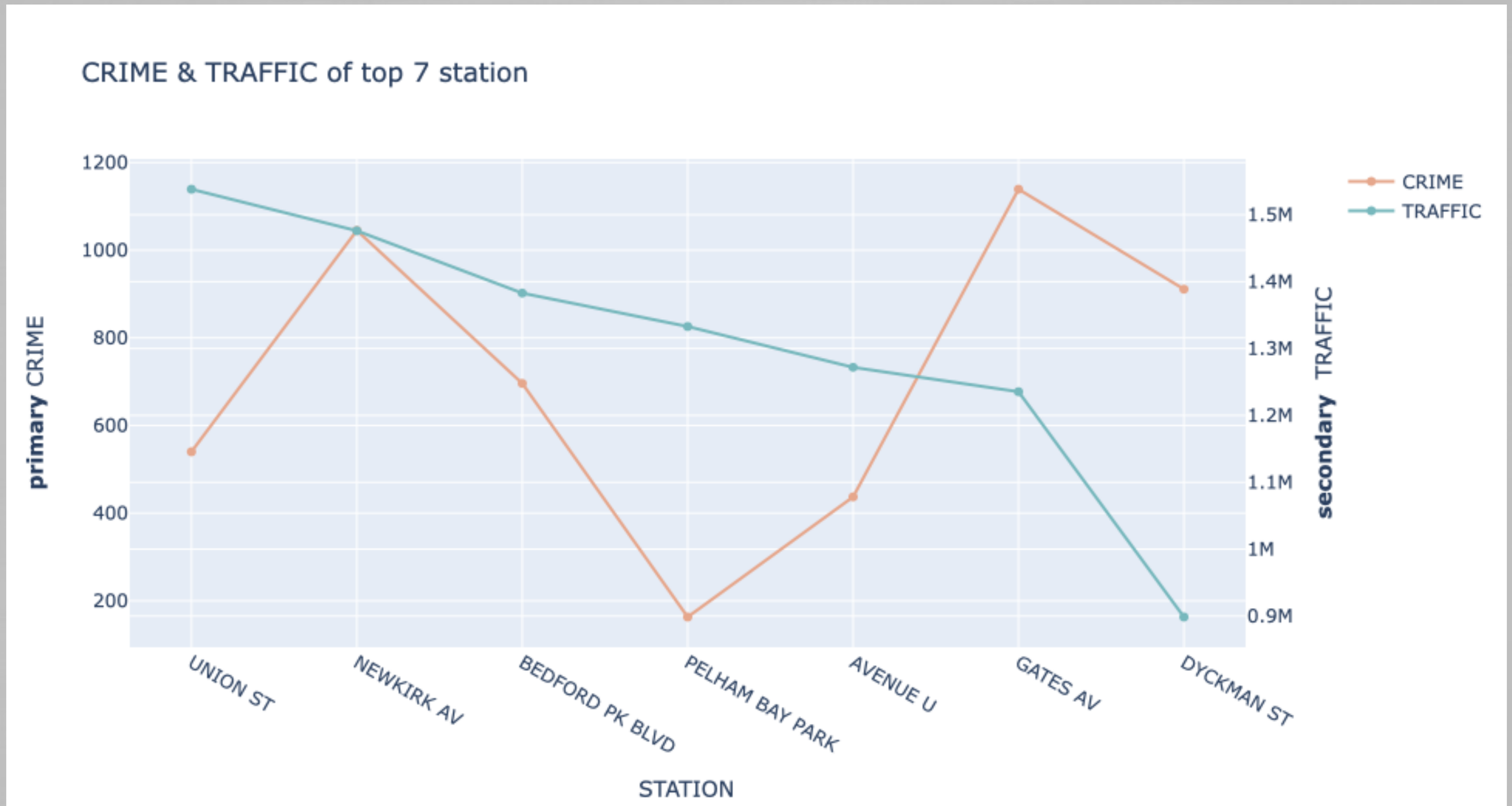

TOTAL ENTRY & Exits of top 7 station
total traffic
1.6M
1.4M
1.2M
1M
0.8M
0.6M
0.4M
0.2M
0
ENTRIES
EXITS
TRAFFIC
UNION ST
NEWKIRK AV
BEDFORD PK BLVD
PELHAM BAY PARK
AVENUE U
GATES AV
DYCKMAN ST

METHODOLOGY

Create a polygon protruding 0.5 mi in all directions from the selected stations and count number of crimes reported within the circular buffer

	LINENAME	DIVISION	STATION	TOTAL_ENTRIES_2019	TOTAL_EXITS_2019	TOTAL_TRAFFIC_2019	LAT	LONG	POLYGON	CRIME
1	R	BMT	UNION ST	881965.0	656124.0	1538089.0	40.677087	-73.983273	POLYGON ((-73.97527 40.67709, -73.97531 40.676...	
4	25	IRT	NEWKIRK AV	869584.0	606435.0	1476019.0	40.639932	-73.948465	POLYGON ((-73.94046 40.63993, -73.94050 40.639...	
5	BD	IND	BEDFORD PK BLVD	860879.0	521926.0	1382805.0	40.717353	-73.956897	POLYGON ((-73.94890 40.71735, -73.94894 40.716...	
6	6	IRT	PELHAM BAY PARK	855026.0	477842.0	1332868.0	40.852990	-73.826868	POLYGON ((-73.81887 40.85299, -73.81891 40.852...	
0	BQ	BMT	AVENUE U	927216.0	344624.0	1271840.0	40.599032	-73.955924	POLYGON ((-73.94792 40.59903, -73.94796 40.598...	
2	JZ	BMT	GATES AV	879254.0	356101.0	1235355.0	40.689538	-73.922219	POLYGON ((-73.91422 40.68954, -73.91426 40.688...	
3	A	IND	DYCKMAN ST	872347.0	26238.0	898585.0	40.865659	-73.927165	POLYGON ((-73.91916 40.86566, -73.91920 40.864...	

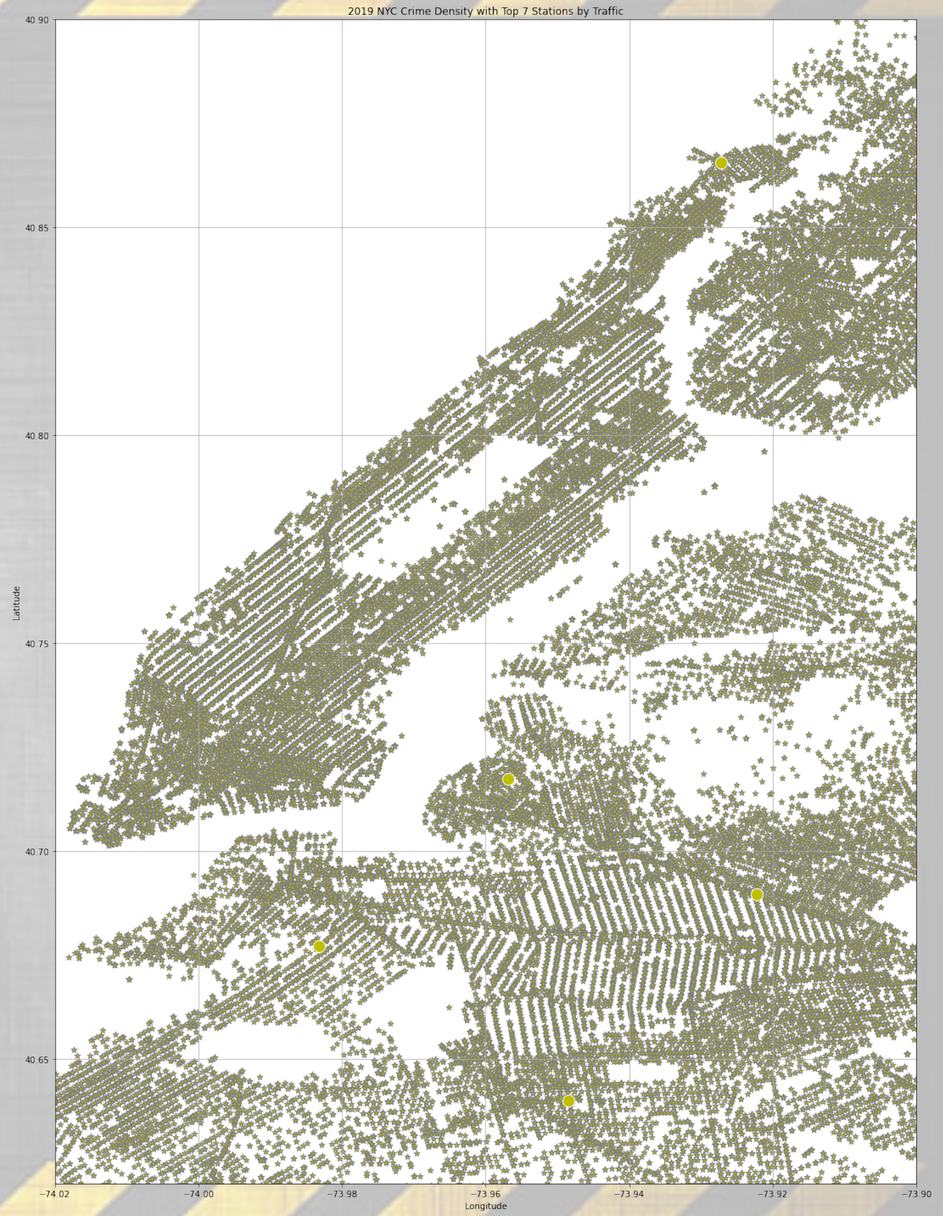



CRIME & TRAFFIC of top 7 station
primary CRIME
1200
1000
800
600
400
200
secondary TRAFFIC
1.5M
1.4M
1.3M
1.2M
1.1M
1M
0.9M
CRIME
TRAFFIC
UNION ST
NEWKIRK AV
BEDFORD PK BLVD
PELHAM BAY PARK
AVENUE U
GATES AV
DYCKMAN ST
STATION

Questions:

-what impact between NYC Crime Statistics with MTA Station Traffic?

-where will be more risk to use the new York city subway?



The end

Thank you
By:Rana alqahtani