

# Data-Driven Policing at the MTA

## Abstract

With increased vaccination rates throughout NYC<sup>1</sup> and with more workers returning to the in-person workforce, subway crime has increased by 93% compared to May of 2020<sup>2</sup>. As a response to this, additional police officers will be deployed throughout the 3 borough command areas<sup>3</sup>.

The goal of this project is to use exploratory data analysis of historical turnstile and crime data to determine which of the 472 stations could benefit the most from additional deployment of law enforcement officers.

## Design

One approach to answering the question involves analysing MTA turnstile data on the station level based on entries per day. The current ratio of Police to Residents in New York City is 436:100,000<sup>4</sup> so higher concentrations of people might require additional deployments.

A second approach is to look at historical crime data to identify concentrations of reported criminal activity. The NYPD Transit Bureau divides NYC into 12 districts with specific precincts covering the MTA rail system across 4 of the 5 boroughs<sup>5</sup>.

Both perspectives will hopefully aid in planning and budgeting officer deployment for the NYPD.

---

<sup>1</sup> [NYC.gov](https://www.nyc.gov) website data

<sup>2</sup> As reported by [NY1 News](https://www.ny1.com) June 05 2021.

<sup>3</sup> Manhattan, Brooklyn and Bronx/Queens

<sup>4</sup> As reported on [Bloomberg](https://www.bloomberg.com)

<sup>5</sup> Staten Island is not considered part of the NYPD Transit coverage

## The Data

The extract of MTA turnstile data I used contains 2.7 million rows featuring subway station usage broken down by individual turnstiles measured by counters every 4 hours. These turnstiles can be grouped by station and each station is covered by a specific NYPD transit district. Reported crimes are broken down by date, time, precinct and district. These were the sources used:

- MTA Turnstile Data - <http://web.mta.info/developers/turnstile.html>
- MTA Station location Data - <http://web.mta.info/developers/data/nyct/subway/Stations.csv>
- NYC Crime Data - <https://data.cityofnewyork.us/Public-Safety/NYC-crime/qb7u-rbmr>
- NYPD Transit Data - <https://www1.nyc.gov/site/nypd/stats/reports-analysis/transit-bus.page>

## Exploratory Data Analysis

- Individual turnstile entries data was aggregated by day and grouped by station name.
- Station locations were determined by merging turnstile data with station data.
- Reported crimes were grouped by day and station name.
- Stations were ranked by the highest traffic and reported crimes
- Ranked lists were overlaid onto a NYC map

## Tools

- SQLite3 to read the MTA data into a Pandas dataframe
- Pandas and NumPy for manipulation and analysis of the data
- Matplotlib, Seaborn and/or Altair for data visualization

EDA Project by: Mike Bernardo

# Communication

Slides and visuals are currently on my [github](#).