Analyzing Complaints to the NYPD With a Focus on Crime in the Subway System

Danielle Shelton

Seidenberg School of Computer Science & Information Systems

INTRODUCTION

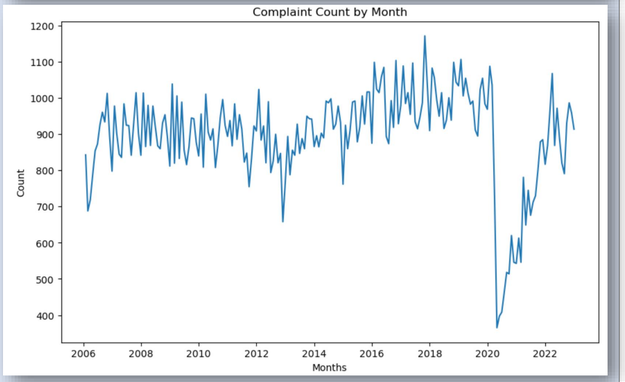
New York City is the most populated city in the United States:

- 8.34 million residents, April 2022.
- Appr. 1 million commuters daily.
- Largest police department: 36000 officers, 19000 civilian employees.

From 2006 to 2022, there were 8.36 million complaints filed with the NYPD. These complaints include felonies, misdemeanors, and violations that take place across all 5 boroughs in homes, places of worship, on city streets, and in the subways. NYC Open Data has detailed documentation of these complaints allowing anyone to gain a better understanding of crime patterns over time, location, and demographics. For this project, I chose to focus on complaints of crime committed exclusively within subway stations. Due to this, the borough of Staten Island is not included in this research project. Querying the data by Station Name and Borough the initial dataset of 8.36 million complaints was paired down to 182,959 complaints.

NYPD Complaint Data Historic

	Bx	Bkln	Manh	Qns	SI	totals
Complaints	27113	47655	87868	20323		182959
Percent	14.819%	26.047%	48.026%	11.108%		
Stations	70	170	121	81		442
Percent	15.837%	38.462%	27.376%	18.326%		
Population	1,356,476	2,561,225	1,597,451	2,252,196	490,687	8,258,035
Percent	16.426%	31.015%	19.344%	27.273%	5.942%	



OBJECTIVES

1: Utilize time series analysis to determine patterns in complaints over time before, during, and after COVID-19.

2: Review geographic data to determine complaint patterns and their level of offense in relation to location.

3: Review the demographics of alleged suspects and victims in the complaints compared to the demographics of the city and individual boroughs.

LITERATURE REVIEW

Before 2020, the NYC population had been on the rise:

- April 2020, population 8.8 million

Since COVID-19, the population has fallen off:

- July 2022, population 8.335 million

As of 2017:

- 53.9% of residents lived and worked within the same borough.
- 39.3% lived and worked in different boroughs within city limits.
- The most common form of travel is the subway for residents.

With the start of COVID-19, ridership in the NYC subway system and crime initially plummeted however their recovery rates have been different. As of January of 2023, “ridership stagnated at 60% of 2019 levels.” At the same time, it was noted that “subway crime rose by 30% in 2022 from a year ago, outpacing the 22% jump in major crimes across the city during the same period” (Akinnibi, Korte, 2023).

Routine Activity Theory (RAT):

- “Predatory crimes against persons or property require three elements – (1) a motivated offender, (2) presence of a suitable target, and (3) the absence of a capable guardian” (Koppel, Capellan, Sharp, 2023).
- Data collected since the implementation of SAH restrictions in New York City support this.

METHODS

1: time series analysis using ARIMA (autoregressive integrated moving average) modeling and XGBoost.

2: Logistic Regression, Bernoulli Naïve-Bayes, and Gaussian Naïve-Bayes

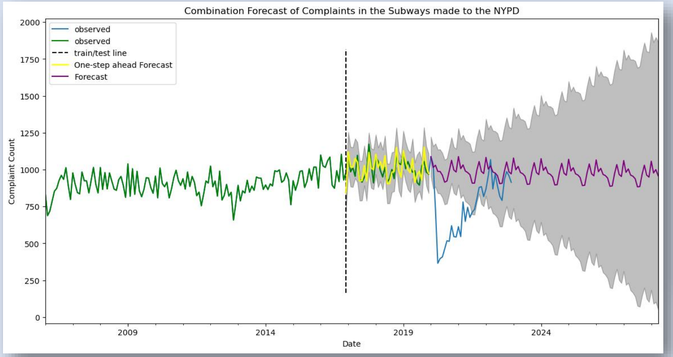
3: Exploratory Data Analysis and compilation of data found online

	RMSE score
ARIMA - monthly	
including COVID	101.3
excluding COVID	63.35
ARIMA - daily	
including COVID	9.96
excluding COVID	11.02
XGBoost - daily	
including COVID	9.27
excluding COVID	6.86
XGBoost - Cross Val.	
excluding COVID	6.966

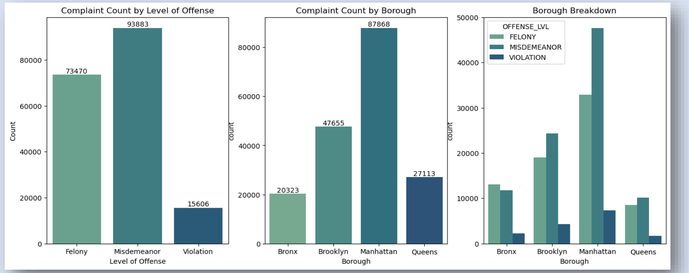
RESULTS

Of the 442 stations in the New York City subway system, 9 of the top 10 stations for most complaints are in Manhattan.

1: Working with Monthly and Daily totals using both Autoregressive Integrated Moving Average (ARIMA) modeling and XGBoost ss Validation show the effects of COVID-19 on predicting future crime complaints. Working with data that ends on 12/31/2019, effectively removing COVID-19 from the train/test model, shows a forecast of complaints trending slightly down. By utilizing the entire dataset, through the end of 2022, the forecast changes to an uphill trend in complaints.



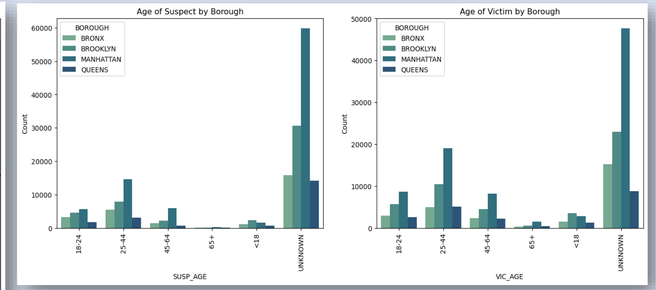
2: A breakdown by Offense level and Borough shows that where crime complaints are made does not align with population and station statistics. Manhattan ranks 3rd for population, 2nd for number of stations, and 1st for number of complaints. Additionally, the Bronx, which ranks last for population and number of stations is the only borough where the number of Felony complaints is higher than Misdemeanor complaints



RESULTS (cont.)

2. Attempts to predict each borough using Logistic Regression with features ['YEAR', 'MONTH', 'DAY', 'FELONY', 'MISDEMEANOR', 'VIOLATION'] proved futile. The only results to reach 50% were those for Manhattan. Predicting the type of crime, Felony with features ['BRONX', 'BROOKLYN', 'MANHATTAN', 'QUEENS', 'YEAR'] was only slightly better with accuracy on both the train and test sets of 60% but precision on the test set was only 43%.

3. Lack of data for age, gender, and race for both the alleged suspect and victim leads to the conclusion that this information is not being collected when crime complaints in the subway system are being made to the NYPD.



CONCLUSIONS & FUTURE WORK

Complaints of crime do not correlate with the residential population or the number of stations in a given borough. Missing demographic information in this dataset meant a better understanding of victims and who was alleged to have committed such crimes was not available to us here. There is another dataset, NYPD Arrests Data, that includes the demographics of alleged suspects but has no information on victims. Based on Routine Activity Theory (RAT), combined with the SAH orders of 2020 and the current popularity of WFH and Hybrid models, we are only beginning to see and understand the effects of COVID-19 on crime over time. Future areas of work should include ridership models along with crime studies. Additional areas of interest are a deeper look into data by borough and demographics (if/when found) looking for patterns and clusters of complaints made over time. Create a series of interactive maps and a dashboard that updates yearly.