Mapping & Predicting Arrest Patterns in New York City

Group 1: Carolina Forero, Daniel Gutierrez, Kristen Kane, Natalia Rigol

FLORIDA ATLANTIC UNIVERSITY

PROBLEM STATEMENT

This project aims to identify and predict the probability of the most prominent offense types for each precinct within New York City for each day of the week, providing actionable insights for local law enforcement agencies and communities.

BACKGROUND & INFORMATION

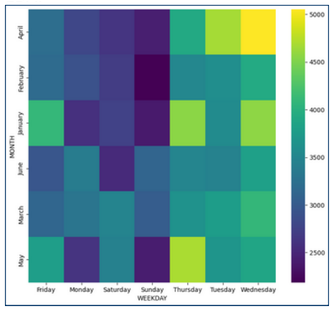
NYC has faced many challenges when trying to reduce crime with limited resources. Many existing predictive models for crime have relied heavily on personal demographic details, which can create bias in policing. This project takes a different approach to predictive methods by focusing on spatial and temporal patterns within NYPD arrest data to identify crime patterns. This project aims to enhance law enforcement and safety planning strategies to reduce demographic bias by using only location and time-related features.

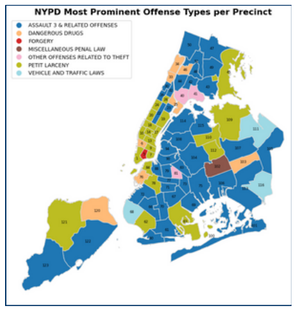
DATA PROCESSING : EXPLORATORY DATA ANALYSIS

Data Details:

- Our primary dataset contains arrests made from January 2025 – June 2025 including variables such as offense type, arrest date, & arrest precinct.
- Our second dataset contains historical NYC arrest data (2006-2024) that will be used for linear regression to predict the number of arrests in the following years

EDA:





Analysis:

- The heat map shows arrest counts by month and day of week displaying a frequency peak made in the month of April and days Tuesday through Thursday
- The hotspot map highlights the most prominent offenses within each precinct showing that the most prominent offenses throughout New York City are “Assault 3 & related offenses” and “Petit larceny”

METHODS & ANALYSIS

Figure 1:

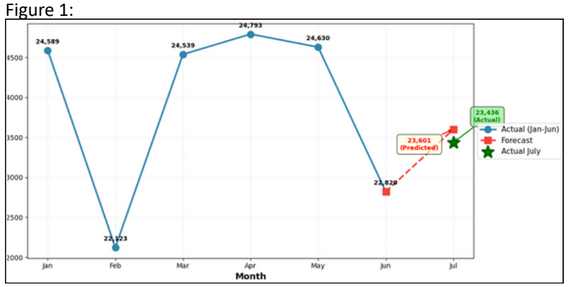


Figure 1 shows a Random Forest Regression model which forecasted July 2025 arrest counts with great accuracy, R^2= 0.974 (explaining 97.4% of arrest count variation) and MAE = 64 arrests (average prediction error of only 67 arrests), demonstrating accurate forecasting for resource planning. Comparing predicted arrests for the month of July with actual arrest counts.

Figure 3:

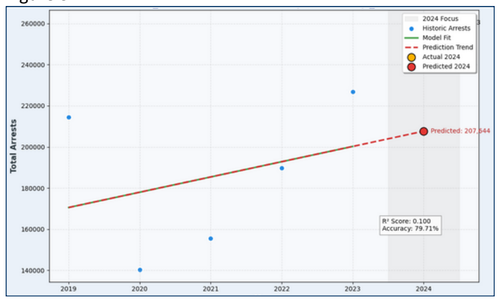


Figure 3 is showing a linear regression model, that forecasts the total amount of arrests for the rest of 2025 and 2026, using data from 2021 to 2025. With an accuracy score of 80%, we can reliably utilize its prediction.

Figure 5:

	Precinct	Day_of_Week	Top_Offense	count	total_arrests	Offense_Probability
0	1	Tuesday	PETIT LARCENY	163	405	40.20%
1	1	Wednesday	PETIT LARCENY	130	388	33.50%
2	1	Monday	PETIT LARCENY	89	273	32.60%
3	1	Thursday	PETIT LARCENY	117	376	31.10%
4	1	Saturday	PETIT LARCENY	98	315	31.10%
5	1	Friday	PETIT LARCENY	102	371	27.50%
6	1	Sunday	PETIT LARCENY	57	221	25.80%
7	5	Saturday	FORGERY	55	190	28.90%
8	5	Monday	DANGEROUS DRUGS	27	195	13.80%
9	5	Thursday	PETIT LARCENY	47	344	13.70%
10	5	Friday	PETIT LARCENY	28	211	13.30%
11	5	Sunday	ASSAULT 3 & RELATED OFFENSES	16	125	12.80%
12	5	Wednesday	GRAND LARCENY	35	291	12.00%
13	5	Tuesday	OFFENSES AGAINST PUBLIC ADMINI	44	380	11.60%
14	6	Tuesday	DANGEROUS DRUGS	68	209	32.50%
15	6	Friday	DANGEROUS DRUGS	59	216	27.30%

Figure 5 is a conditional probability model that shows the likelihood of the most common offense type to occur within each precinct on every day of the week. Displays how prominent offense types may fluctuate throughout the week. A seemingly low probability can still be deemed significant when considering the amount of offense types that may occur (58).

Figure 2:

```
Original dataset: 142797 arrests
Filtered dataset: 101082 arrests
Offense types reduced from 56 to 10

Top 10 Offenses:
OFNS_DESC
ASSAULT 3 & RELATED OFFENSES        19590
PETIT LARCENY                       14461
DANGEROUS DRUGS                     11966
FELONY ASSAULT                      11479
MISCELLANEOUS PENAL LAW              9462
OTHER OFFENSES RELATED TO THEFT      8736
VEHICLE AND TRAFFIC LAWS             8325
CRIMINAL MISCHIEF & RELATED OF       6182
ROBBERY                              5519
DANGEROUS WEAPONS                    5362
Name: count, dtype: int64
Accuracy: 0.243
Number of offense types: 10

Feature Importance:
           Feature  Importance
1  ARREST_PRECINCT    0.494151
3            MONTH    0.250904
2          WEEKDAY    0.227639
0      ARREST_BORO    0.027306
```

Figure 2 shows a Random Forest Classifier using top 10 offense types (70% of data), to achieve a better accuracy (24.3%), successfully predicting property crimes (46% of thefts) but struggling with random violent crimes (2% of weapons cases).

Figure 4:

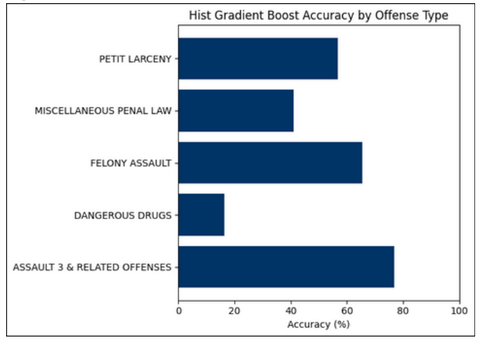


Figure 4 shows the accuracy of the Hist Gradient Boosting model for the top 5 most dominant offense types. The model received an overall accuracy of 55%, best predicting assault-related offenses, while drug and property-related offenses were harder to predict due to having more widely distributed data across precincts and days. Results indicate that spatial and temporal patterns allow consistency without including personal demographics.

CONCLUSION & NEXT STEPS

Conclusion:

The results show that NYC exhibits clear spatial and temporal arrest patterns by place and day of the week. Arrests seem to spike in April and mainly occur Tuesday through Thursday with the dominant offense in most precincts being “Assault 3 & related offenses” and “Petit larceny”. Our models that only utilize location and time features were able to capture this variation, showing that these patterns can be used to predict arrest activity.

Next steps:

Additional features could be utilized, if included in the data, allowing for a more precise recognition of arrest patterns. This could include the time of day, the socioeconomic factors of the area, or weather conditions. These results could be used for real-time decision making, such as specifc resource allocation, patrol scheduling, or community planning. These findings may greatly improve safety planning in many areas throughout New York City.

REFERENCES

Background: https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained

GIS data for hotspot mapping: https://www.nyc.gov/content/planning/pages/resources/datasets/police-precincts

Historical Arrest Data: https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u/about_data

NYPD Arrest Data (Main Dataset): https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc/about_data

POLICE DEPARTMENT CITY OF NEW YORK