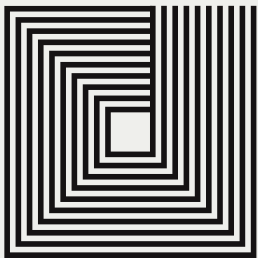# MAPPING & PREDICTING ARREST PATTERNS IN NEW YORK CITY

Carolina Forero
Daniel Gutierrez
Kristen Kane
Natalia Rigol

PROBLEM STATEMENT

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

BACKGROUND

- NYC faces challenges reducing crime while managing limited resources.
- Crime often forms clusters, making trends identifiable, but policing is still mostly reactive.
- Many predictive models are biased when relying on personal demographics.
- This project uses neutral features (location and time of arrest) to identify trends more transparently.
- The insights can help law enforcement, emergency services, and policymakers improve planning and public safety.

METHODS - EDA
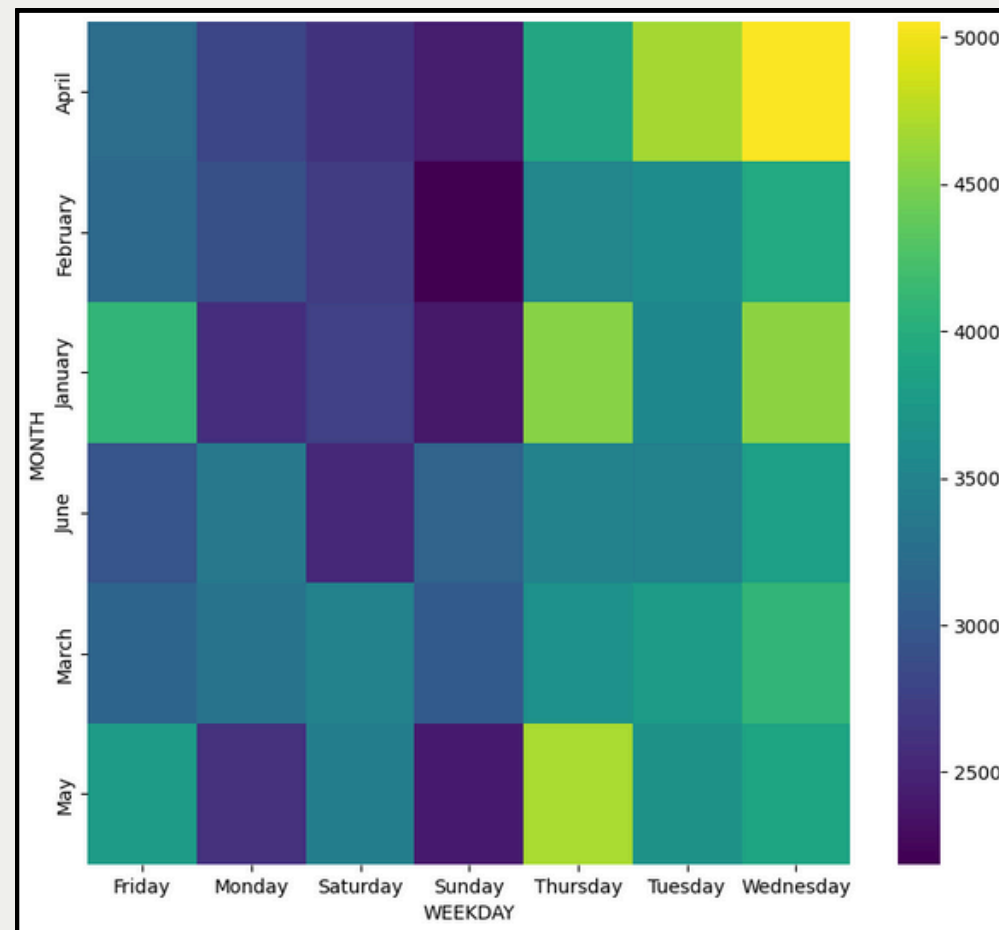
Heat map: Arrest counts by month and day of the week
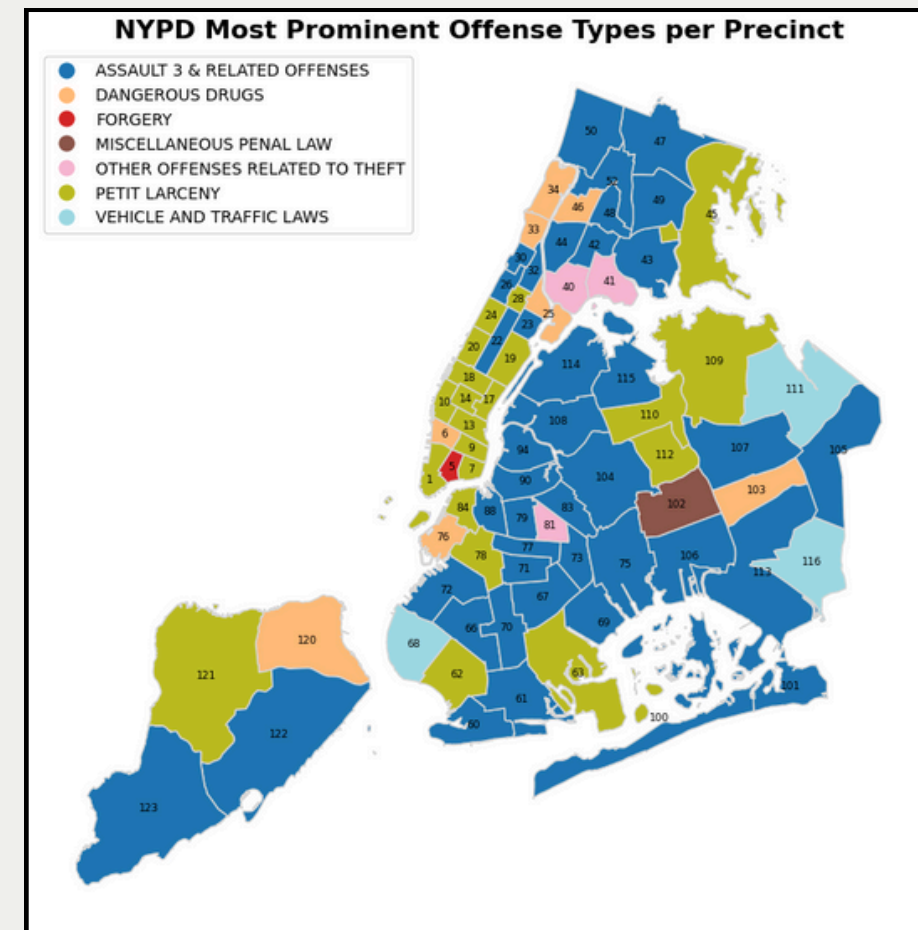
Hotspot: Top offenses by precinct

Maps Created:
- Hotspot: Arrest count by precinct
- Hotspot: Top offenses by precinct
- Hotspot: Top offenses by precinct per day of the week
- Heat map: Arrest counts by month a day of the week

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 shows that the most prominent offenses through the city are Assault 3 & related offenses and Petit larceny

METHODS - MODELS


```
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
```
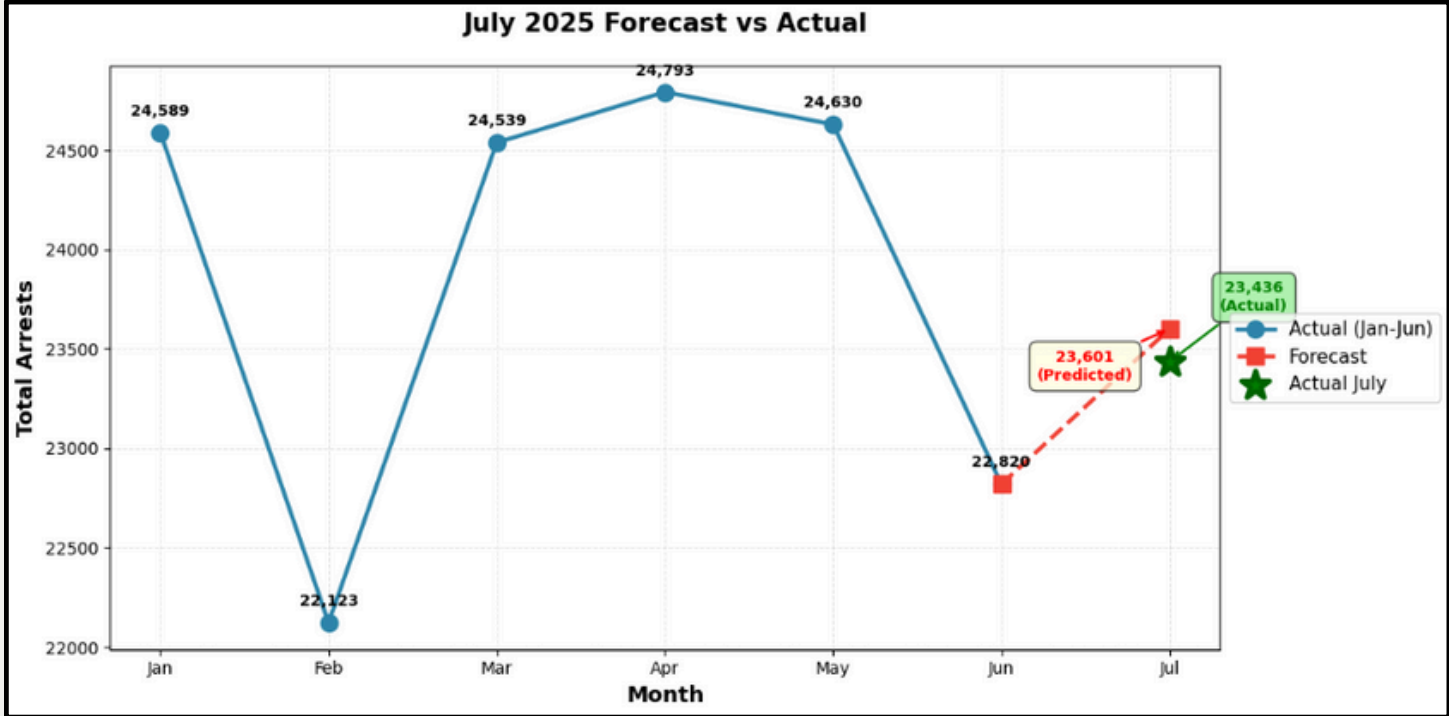
Random Forest Classifier

Regression Model

Random Forest Classifier:
- Predicting which offense may occur when a person is arrested
- Focuses on the top 10 most common offenses
- 24% accuracy
- Has the ability to predict which offense may occur based on the precinct and day of the week

Random Forest Regression:
- Predicting the number of arrests that may have been made in July 2025 with 97% accuracy
- Graph shows the arrest counts that have already been recorded and compared it to the predicted arrest count for July 2025
- Predictions can be confirmed once the NYPD arrest dataset is updated

METHODS - MODELS
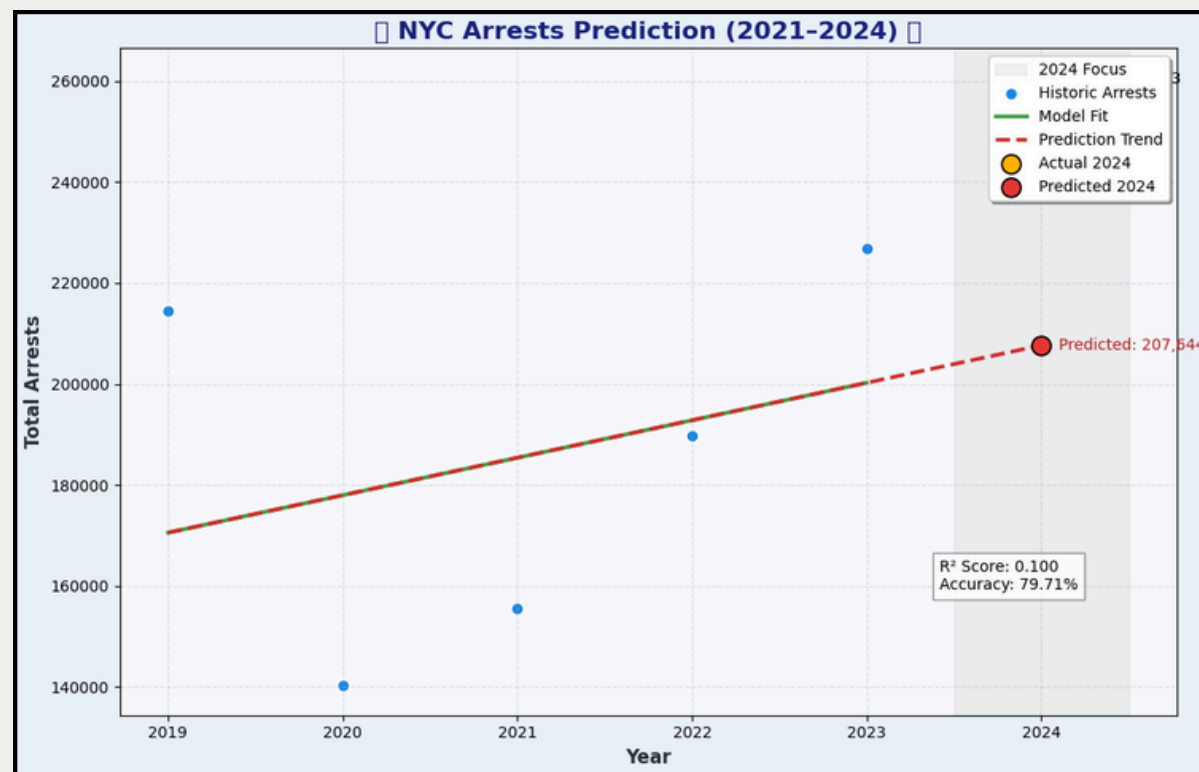

Linear Regression
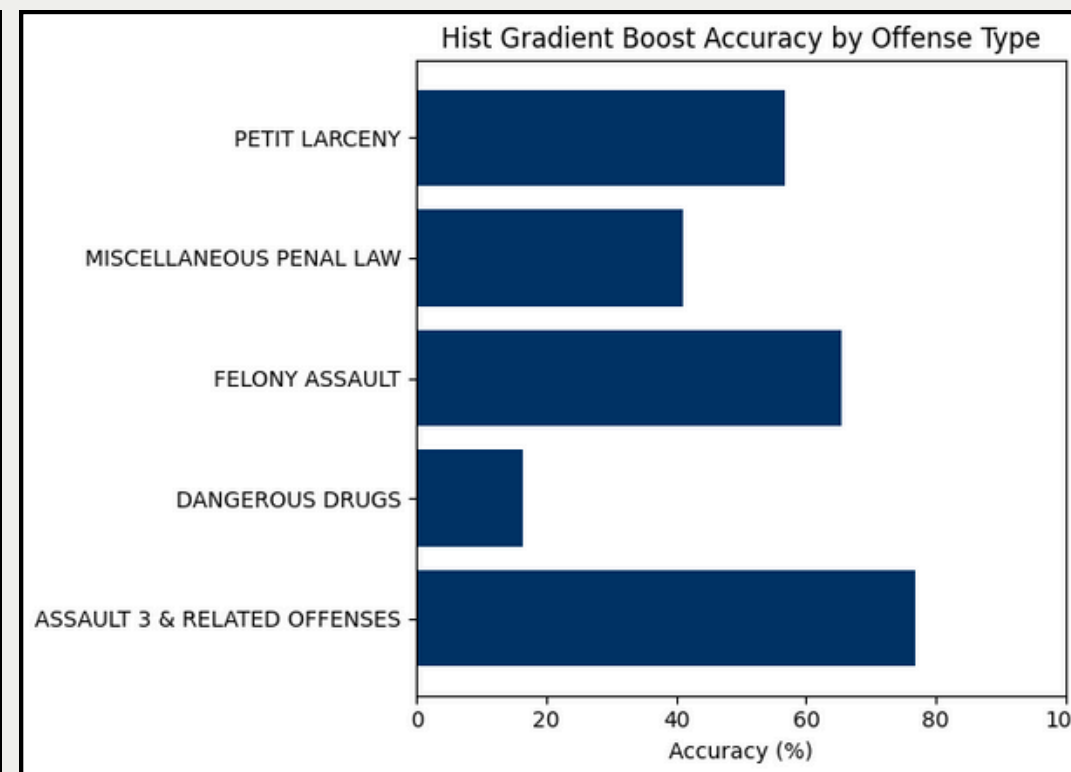
Hist Gradient Boosting

Linear Regression Model:
- Predicting the amount of arrests made in 2025 and 2026 with 80% accuracy
- Forecasted to have 205k -210k arrests over the next 2 years
- Most arrests occur on Wednesdays in October, with assault-related offenses being the most common
- Precinct 14 records the highest arrest activity, indicating strong police presence in Manhattan

Hist Gradient Boosting Accuracy:
- Received an overall accuracy of 55% when considering the top 5 most dominant offense types
- Utilized features such as precinct, day, month, borough, weekend flag, and law category
- Assault-related offenses showed greater accuracy, while drug and property-related offenses were harder to determine

METHODS - MODELS
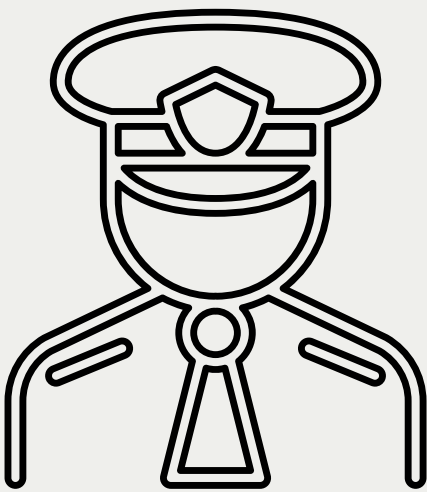
Conditional Probability Model:
- Shows which offenses are most likely to occur in each precinct of NYC on any day of the week
- Groups arrests by precinct, day of the week, and offense description
- Allows the ability to identify arrest patterns for the most prominant offense types without relying on predictive features

Analysis:
- Some precincts have offense types that occur consistently throughout the week
- Other precincts do not show the same consistency
- In Precinct 1, the most dominant offense type every day of the week is Petit Larceny (25%-40%)
- In Precinct 5, offense types vary from Petit Larceny to Dangerous Drugs
- Different precincts should use different safety planning strategies based on their specific prominent offense type

	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%

Conditional Probability Model

CONCLUSION

Analysis

- NYPD arrest data shows clear spatial and temporal patterns by location and day of week.
- Most common offenses: "Assault 3 & related offenses" and "Petit larceny".
- Even with only location and time features, the model captured most data variation.
- These patterns give law enforcement useful insights for safety planning and resource allocation.

Next Steps

- Add more features (time-of-day, weather, seasons, socioeconomic factors).
- Share insights with law enforcement and community groups.
- Create dashboards for precinct-level forecasting.
- Include a confusion matrix for model evaluation.

THANK YOU
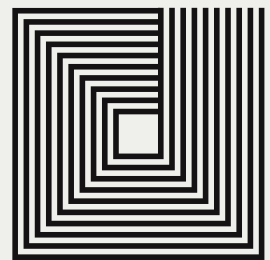
Carolina Forero
cforero2024@fau.edu

Daniel Gutierrez
gutierrezd2023@fau.edu

Kristen Kane
kanek2022@fau.edu

Natalia Rigol
nceleirorigo2023@fau.edu

Website: https://kristennkane.github.io/NYPD-Capstone-FAU-2025/

QUESTIONS?

Florida Atlantic University