NY Crime Detective

A comprehensive analysis of crime patterns across New York State

By: Hana Alyahya | Advanced Policy Analysis

Syracuse University

1. Project overview

Crime impacts not just public safety, but also housing markets, urban development, local economies, and overall public trust. However, decision-makers often lack timely insights into where, when, and what types of crimes are happening—until it's too late.

This project, NY Crime Detective, aims to address that gap by building a predictive model to classify crime types and uncover meaningful behavioral patterns across New York State. The objective of this project is to analyze crime patterns in New York State using open data sources. It aims to uncover trends, identify high-crime areas, and explore the relationship between different types of crimes, locations, and times. The analysis focuses on historical crime data from 2021 to 2023 to build a foundation for future predictive modeling efforts. By leveraging big data analytics techniques, this project provides insights to support community safety efforts and policy recommendations.

2. Project Goals:

Predict Crime Type

Develop a machine learning model that classifies reported crimes into broad categories using structured incident data.

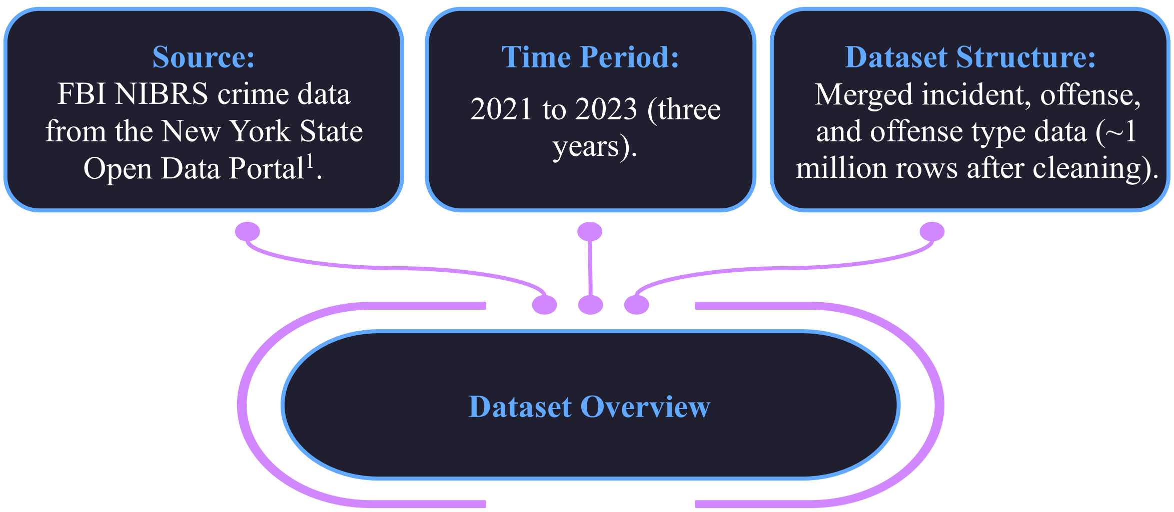
Learn Crime Behavior

Extract meaningful features from time, location, and past trends to reveal how different crimes behave over hours, days, and seasons.

Support Decision-Making

Support proactive public safety planning with predictive analytics.

3. Dataset Overview



4. Model Overview

4.1 Model Selection

An **XGBoost Classifier** was selected for this project. This model was chosen because:

⇒ It performs well with **large datasets** and **imbalanced classes**, which are typical in crime datasets[2].

⇒ It supports **regularization techniques** to prevent overfitting.

⇒ It provides **feature importance scores**, enhancing interpretability for public safety planning.

4.2 Feature Engineering

To enhance model performance and better capture crime behavior patterns, several types of features were created:

Temporal Features

Contextual features

Smart features

Temporal Features:

- **Hour of day**: to capture time-of-day effects.
- **Day of week**: to identify weekday vs weekend patterns.
- **Season**: to detect seasonal crime trends.

Contextual Features:

- **Location ID**: representing where the crime occurred.
- **Crime Against Category**: whether the crime was against person, property, or society.
- **Completion Status**: whether the crime attempt was completed or attempted.

Smart Features:

- **Past Crime Counts:**
  - Number of crimes recorded in the past 7 days at the same location.
  - Number of crimes recorded in the past 30 days at the same location and time window.

These features helped capture recent trends and potential crime hotspots dynamically.

4.3 Crime Type Grouping

To improve model performance and simplify classification, the original dataset's 24+ crime types were consolidated into six broader, more manageable categories:

- Theft/Fraud
- Assault/Sex Crimes
- Property Damage
- Drug/Weapon

- Violent Crimes
- Other

This grouping was essential to address class imbalance issues and to make the predictive insights more practical and actionable for public safety applications.

4.4 Model Training and Performance

The model was trained on the structured dataset with the engineered features.

- **80% overall accuracy** was achieved on the test set.
- The model performed especially well in predicting frequent crime categories such as:
  - **Theft/Fraud** crimes
  - **Assault/Sex** crimes
- Performance on the "Other" crime category was lower, likely due to the heterogeneous nature of that group.

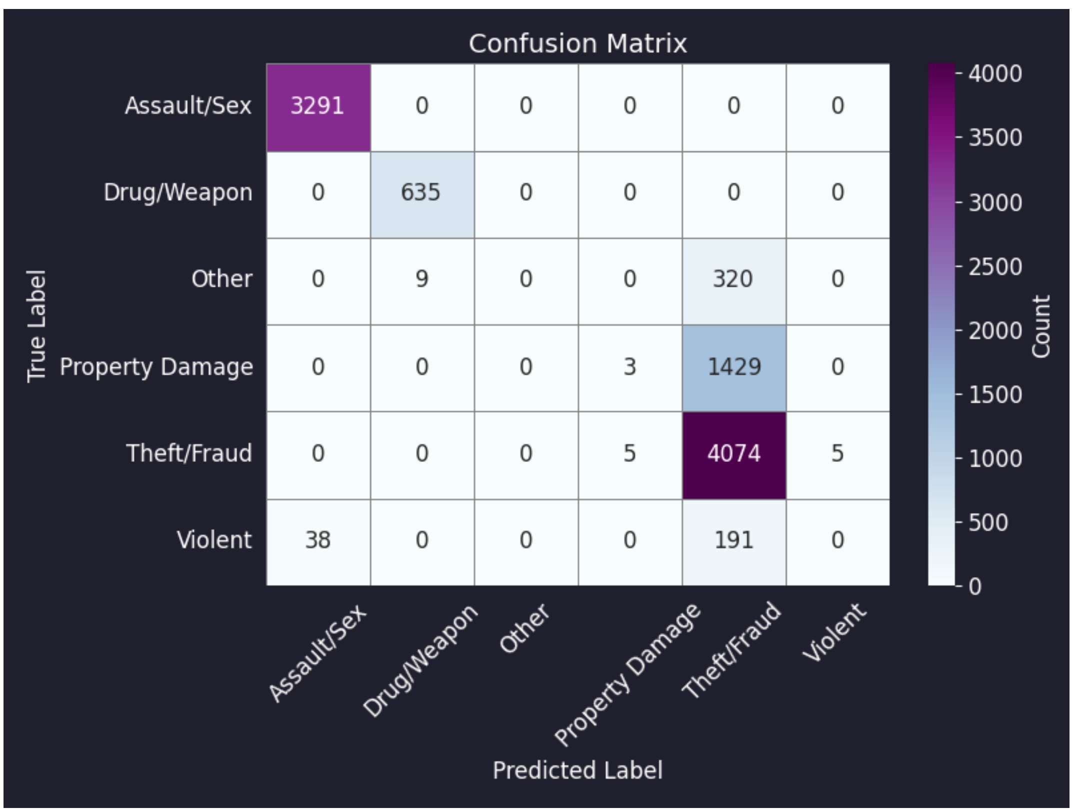


Figure 1. Confusion Matrix

4.5 Key Predictive Features Identified

Based on feature importance analysis:

- **Crime Against Person** category contributed **27%** to the model's predictive power.
- **Crime Against Society** contributed **15%**.
- **Crime Against Property** contributed **10%**.
- While crime type and location features were the most influential predictors, time-related features such as day of the week also contributed modestly to the model's performance.

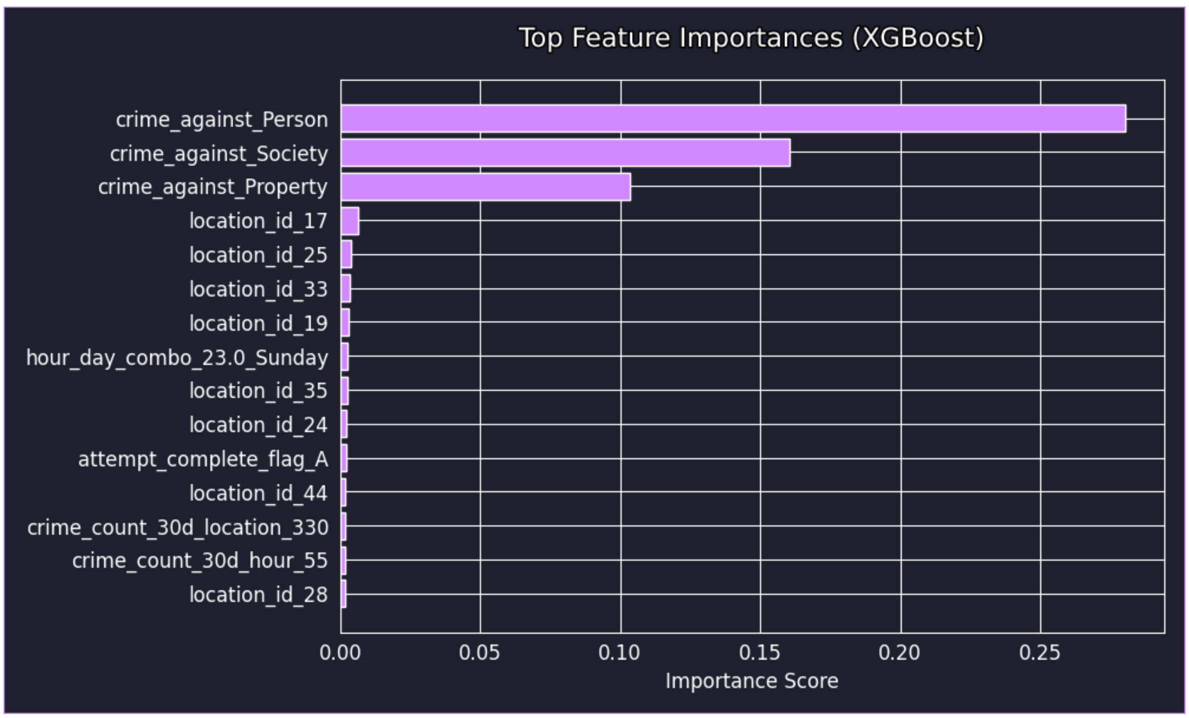


Figure 2. Top Features

5. Results summary

The analysis revealed several important patterns related to crime trends across New York State between 2021 and 2023:

As shown in Figure 3, Theft/Fraud crimes were the most frequent, followed by Assault/Sex crimes and Property Damage incidents. Other categories like Drug/Weapon and Violent crimes were less common.

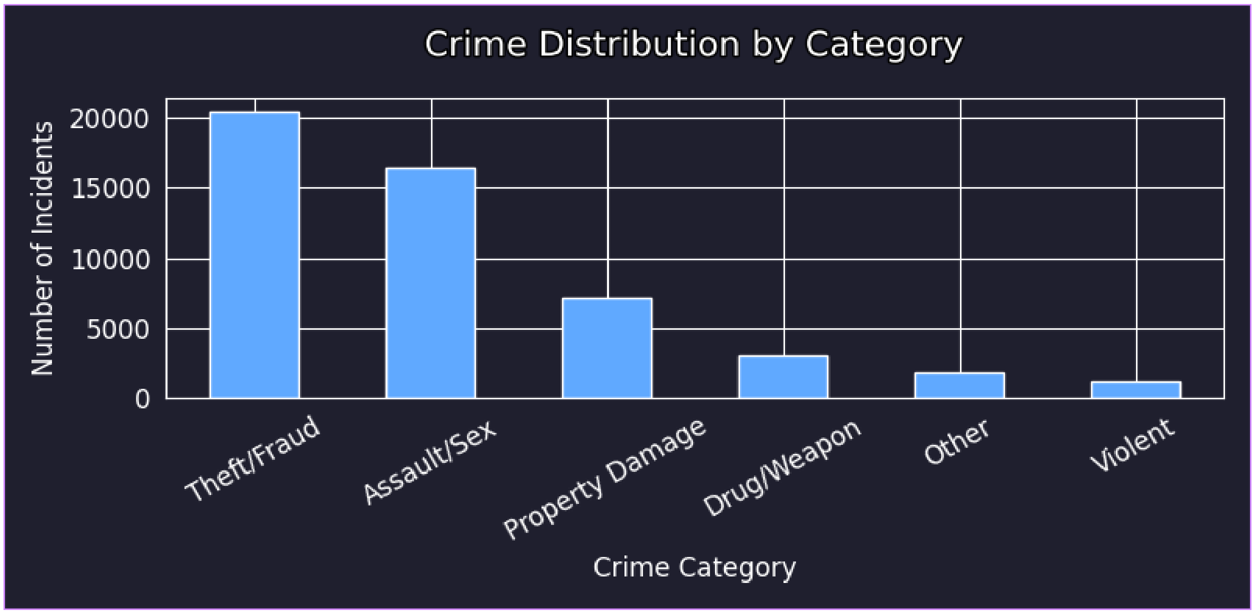


Figure 3. Crime Distribution by Category

As illustrated in Figure 4, crime activity was highest on Wednesdays and Fridays, while Sundays consistently recorded the lowest number of crimes.

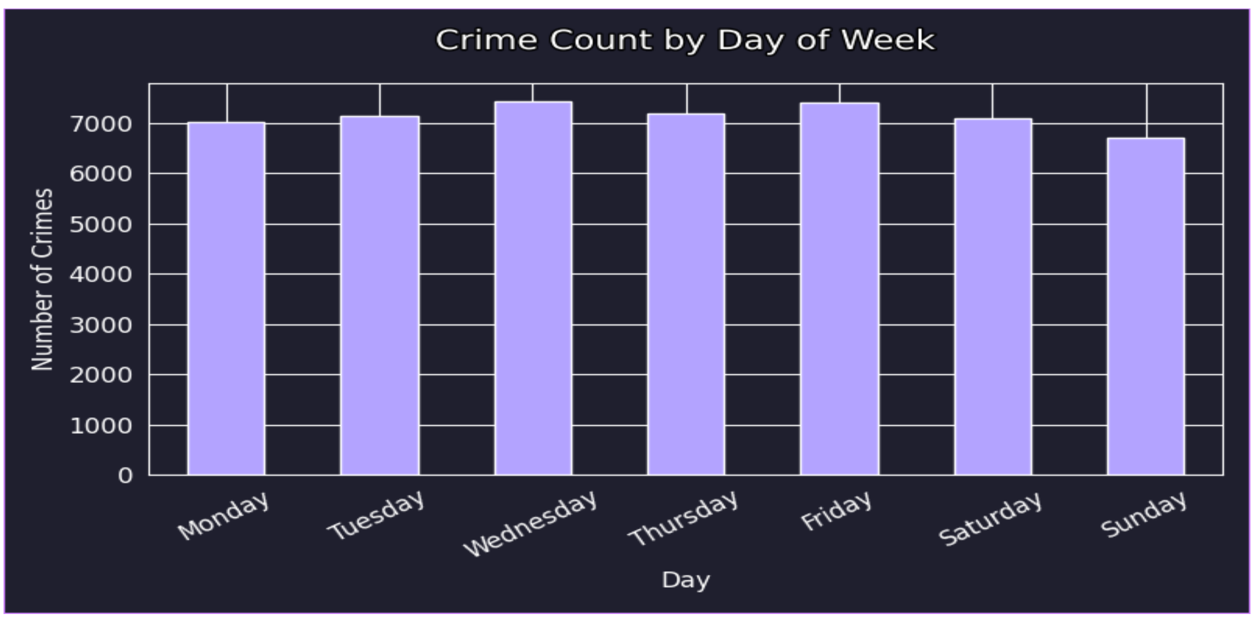


Figure 4. Crime Count by Day of the Week

Since crime patterns changed across the days of the week, it was also important to examine how crime levels shift during different times of the day. This helped identify when crime is most likely to occur within a 24-hour period. As shown in Figure 5, the afternoon period (12 PM to 5 PM) experienced the highest concentration of crimes, whereas early morning hours (midnight to 5 AM) had the lowest.

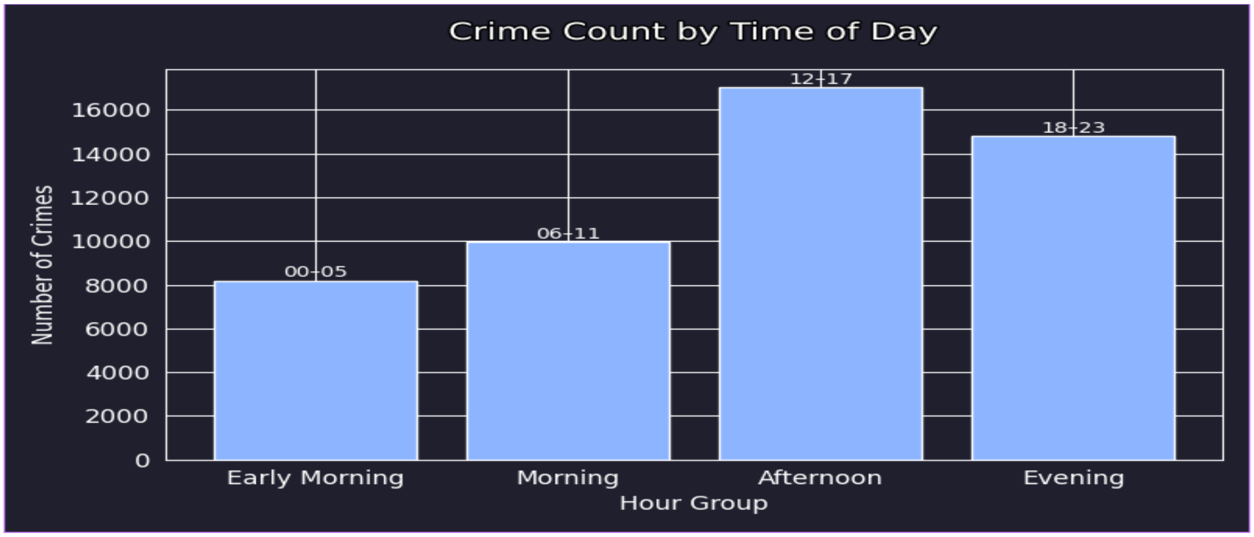


Figure 5. Crime Count by Time of Day

Understanding daily time patterns also raised the question of whether crime levels vary across different seasons, as external factors like weather and public activity levels change throughout the year. Figure 6 shows that summer recorded the highest number of crimes, followed by spring and fall, with winter showing the lowest crime rates.

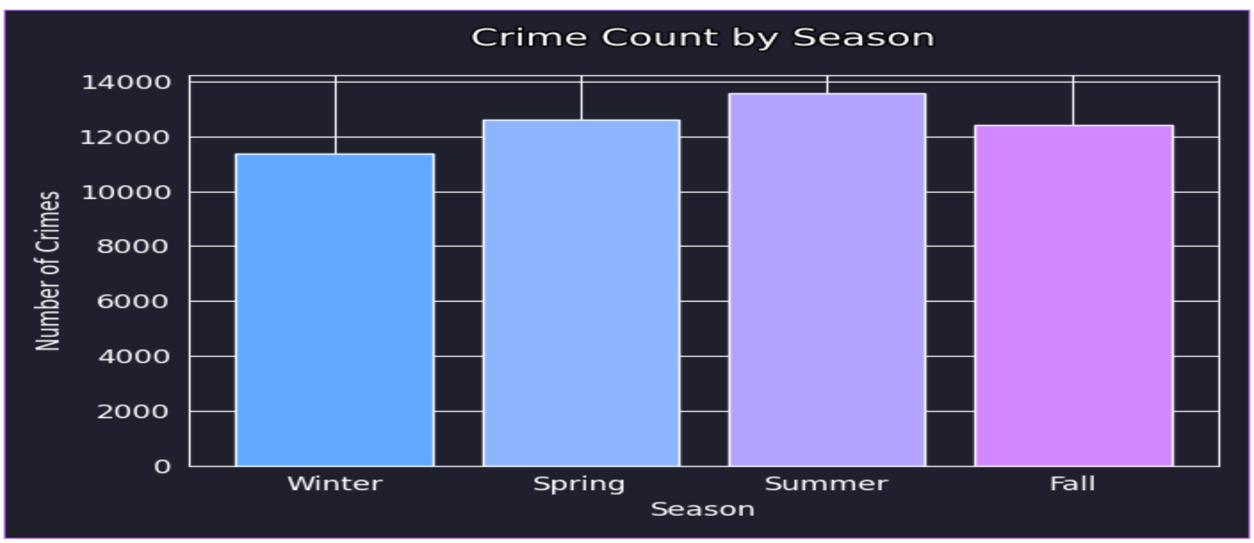


Figure 6. Crime Count by Season

Given the strong seasonal variations, it became important to further explore how different crime types behave throughout the day. This deeper analysis would reveal whether specific types of crimes peak at certain hours. As seen in Figure 7, the heatmap highlights that Theft/Fraud and

Assault/Sex crimes peak during afternoon hours, while property damage incidents are more evenly distributed throughout the day.

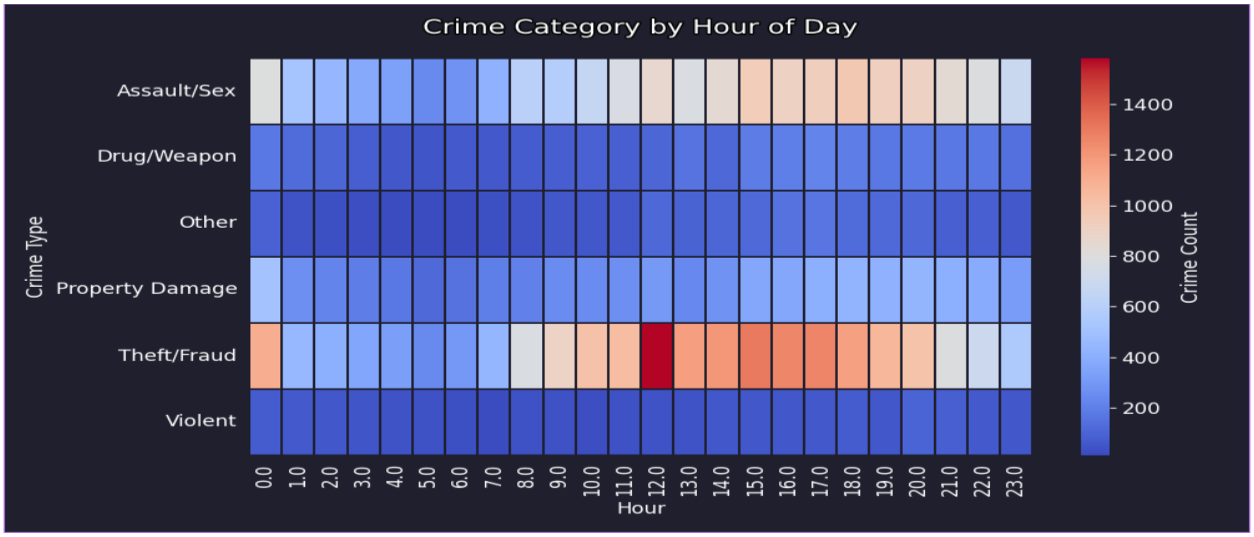


Figure 7. Crime Patterns by Hour and Category

After understanding hourly patterns for different crime types, the next step was to compare crime frequency between weekdays and weekends, as lifestyle changes between these periods could influence crime behavior. Figure 8 demonstrates that crimes, especially Assault/Sex and Theft/Fraud, were more frequent on weekdays compared to weekends.

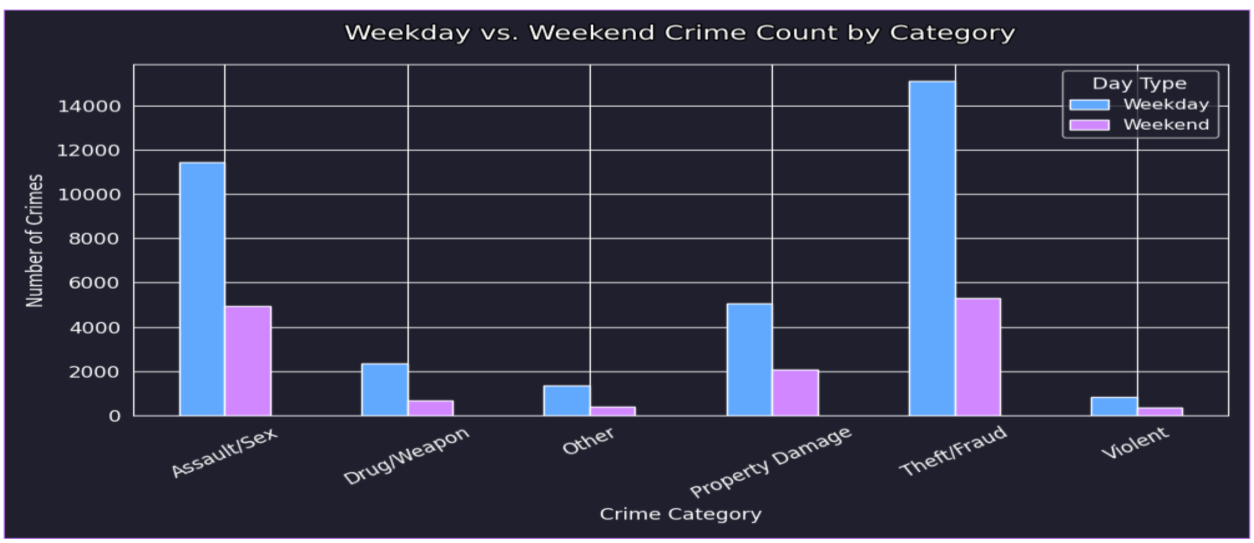


Figure 8. Weekday vs. Weekend Crime Counts

Overall, the analysis revealed that crime incidents are strongly influenced by time and seasonality. Nearly half of all crimes occur during the afternoon hours, highlighting important time-based trends that can guide resource allocation strategies. In addition, crime activity tends to peak midweek, particularly on Wednesdays and Fridays, rather than on weekends, challenging common assumptions about when crimes are most likely to occur. Seasonal effects were also significant, with substantially higher crime rates recorded during the summer months compared to winter, underscoring the role of environmental and social factors in crime patterns.

6. Challenges

1

The original dataset contained over 24 crime types, requiring consolidation into six broader categories to make modeling feasible.

Data Complexity

2

Some crime categories had significantly fewer samples, which made achieving balanced accuracy across classes difficult.

Class Imbalance

3

Limited contextual data (e.g., lack of demographics, socioeconomic factors) constrained the predictive power for some crime types.

Feature Gaps

7. Conclusion and Future Work

The project successfully developed a strong predictive model capable of classifying crimes into broader categories with good accuracy. Throughout the analysis, clear temporal and seasonal patterns were identified, supporting the goal of understanding crime behavior across different timeframes. Overall, the results provide actionable insights that can assist public safety decision-makers in developing more proactive and targeted strategies, aligning closely with the original objective of using predictive analytics to enhance community safety efforts.

Looking ahead, future work could focus on several areas for improvement. Incorporating additional contextual data such as demographic, socioeconomic, and environmental factors could enhance model accuracy and deepen the understanding of crime drivers. Improving classification within the "Other" crime category could also lead to more precise predictions across all crime types. Additionally, applying advanced modeling techniques, such as deep learning or spatial-temporal models, could further improve the ability to forecast crime patterns at finer geographic and temporal resolutions.

Policy Implications

The predictive model developed in this project offers valuable insights that can directly support public safety decision-making. Policymakers and city officials can use the findings to allocate police resources more efficiently, focusing on peak hours and high-crime days identified in the analysis. Seasonal trends can guide adjustments in staffing levels during summer months when crime rates are higher. Additionally, insights into weekday versus weekend crime patterns can inform targeted prevention strategies and public awareness campaigns. Overall, the model helps translate big data analytics into actionable, proactive measures that enhance community safety.

References

1. Federal Bureau of Investigation. *National Incident-Based Reporting System (NIBRS) Data*. New York State Open Data Portal, 2021–2023.
2. Chen, Tianqi, and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.