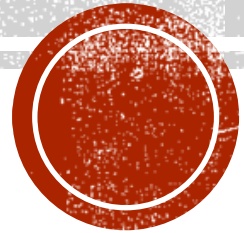
CHICAGO CRIME CLASSIFIER

By Ian Korir



INTRODUCTION

- The City of Chicago has been facing persistent crime-related issues, impacting its residents' safety and well-being. With over 1.5 million crime reports recorded between 2001 to present, the sheer volume and complexity of the data make it challenging for law enforcement agencies to analyze and respond effectively to criminal activities.
- Utilizing advanced data analytics and machine learning techniques presents an opportunity to transform raw crime data into actionable insights, enhancing the city's ability to prevent and manage crime.
- Law Enforcement Agencies: Primary users of the crime classifier, benefiting from data-driven insights to improve resource allocation, crime prevention strategies, and operational efficiency.
- City Government and Public Safety Officials: Utilize insights from the classifier to inform policy decisions, budget allocations, and community safety initiatives.

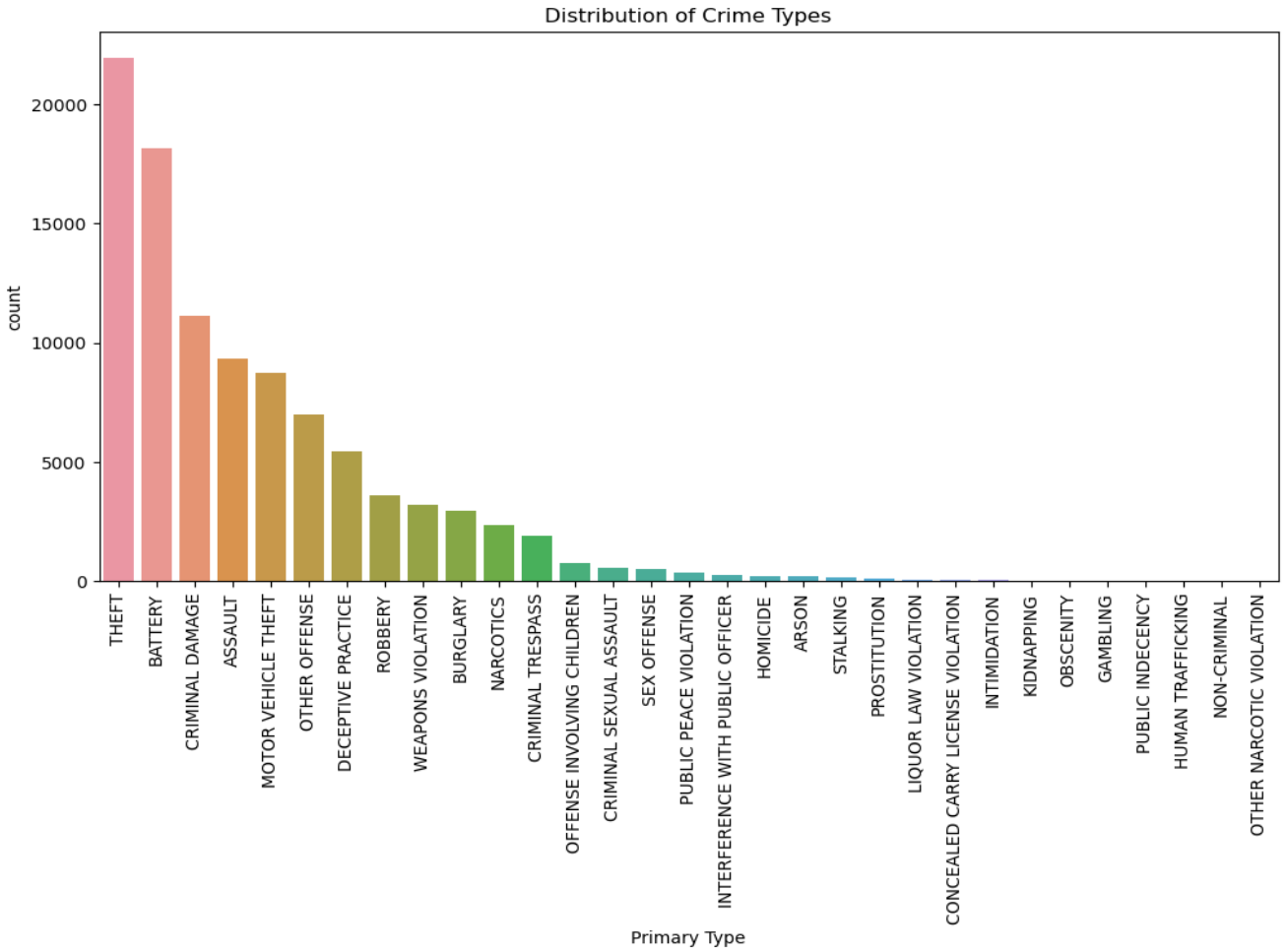
OBJECTIVES

- **Main Objective:**
  - Develop a Crime Classifier: Build and train machine learning models to classify and predict criminal activities based on historical data.
- **Other Objectives:**
  - Analyze Crime Data: Conduct a comprehensive analysis of historical crime data to identify patterns and trends.
  - Evaluate Model Performance: Assess the accuracy and effectiveness of different models to determine the best approach for crime prediction.
  - Enhance Resource Allocation: Use the insights gained from the classifier to optimize the allocation of police resources and improve response strategies.
  - Improve Crime Prevention: Develop proactive measures and targeted strategies to prevent crime and enhance public safety in Chicago.

DATA OVERVIEW AND UNDERSTANDING

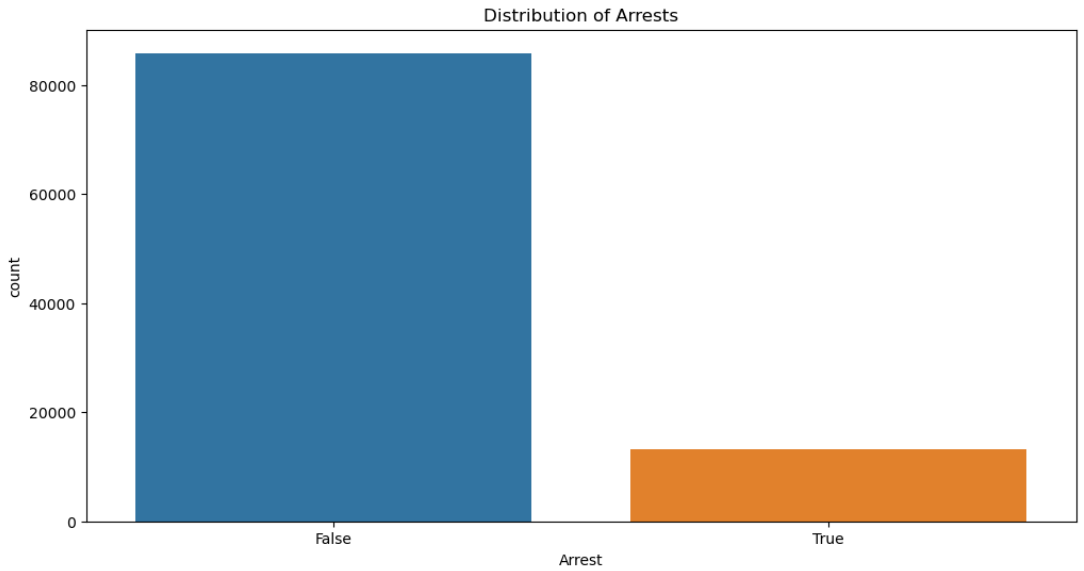
- The dataset comprises records of crimes in Chicago.
- Key features used
  - Date and time of the crime
  - Location description
  - Type of crime
  - Arrest made or not
  - Domestic crime or not
  - Beat, district, ward, community area, etc.
- Removed irrelevant or redundant features: 'ID', 'Case Number', 'Block', 'IUCR', 'Description', 'FBI Code', 'X Coordinate', 'Y Coordinate', 'Updated On', 'Location', 'Year'

EXPLORATORY DATA ANALYSIS (EDA)



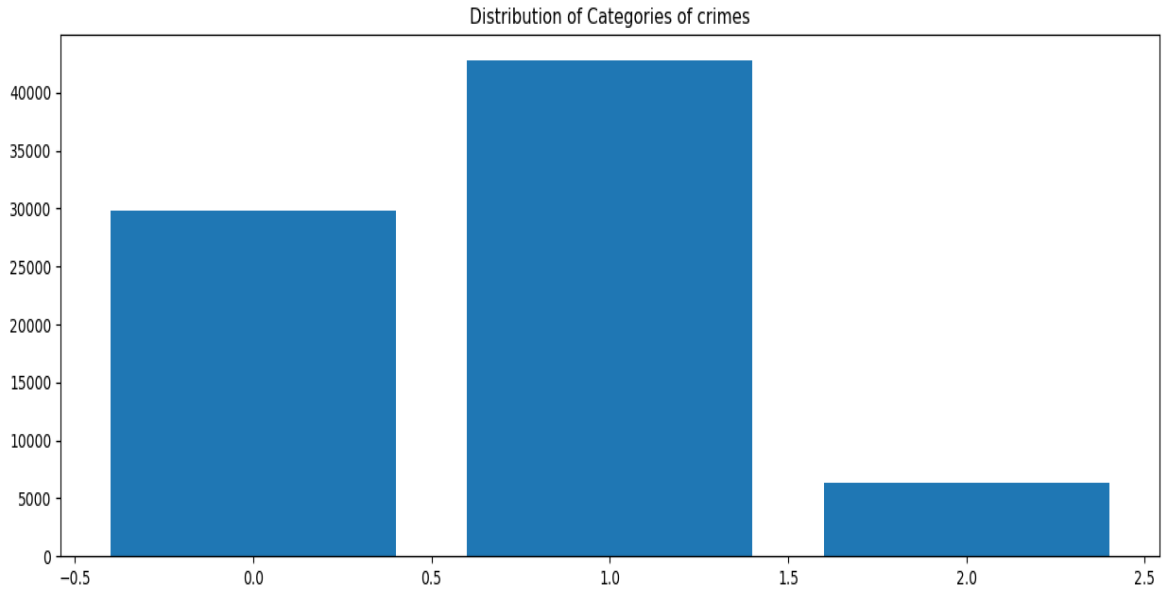
- From the table we can conclude that Theft and Battery are the leading crimes.

ARREST DISTRIBUTION



- Many crimes go unpunished since few arrests are made.
- This shows how unsafe Chicago neighbourhoods are.

DISTRIBUTION OF CATEGORIES OF CRIMES



- 0 - Offenses directed against the person' : 29296
- 1 - Number of crimes in the category 'Offenses directed against property' : 42834
- 2 - Number of crimes in the category 'Offenses affecting public health, safety and decency/affecting governmental functions' : 6373

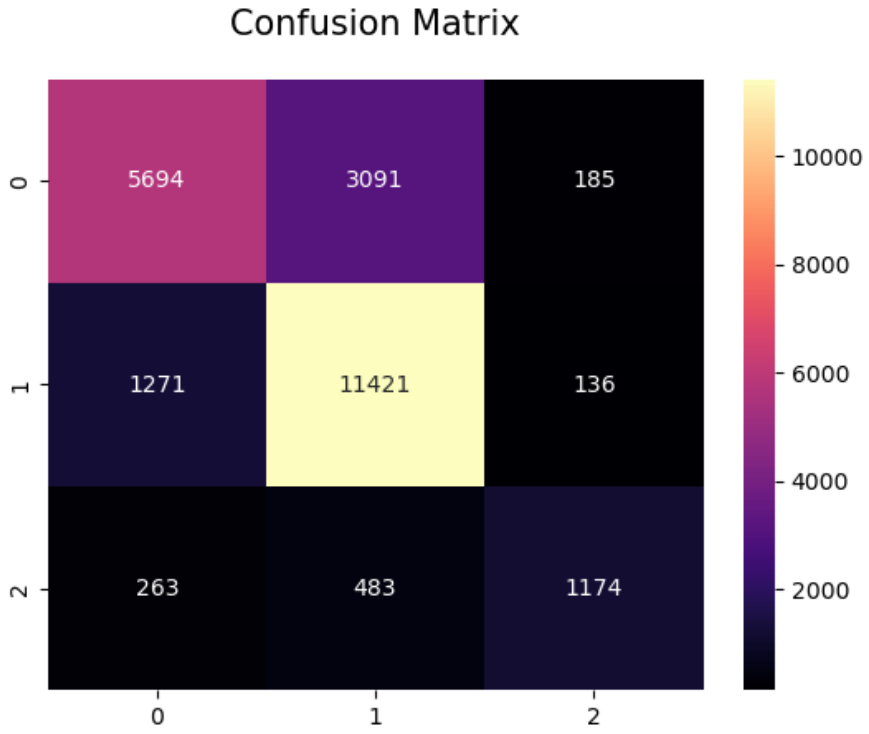
MODELING

- Tested various models: K-Nearest Neighbour(KNN), Random Forest, XGBoost.
- Evaluated models using cross-validation and test set performance.
- Models and Performance:

Model	Accuracy
KNN	71%
Random Forest	76%
XGBoost	77%

MODEL EVALUATION

- The selected XGBoost model was evaluated using a classification report and a confusion matrix. Observations indicated that the model had difficulty differentiating between certain categories.
- precision recall f1-score support
- ODATP 0.79 0.63 0.70 8970
- ODAP 0.76 0.89 0.82 12828
- OAPHSD 0.79 0.61 0.69 1920
- accuracy 0.77 23718
- macro avg 0.78 0.71 0.74 23718
- weighted avg 0.77 0.77 0.77 23718



RECOMMENDATIONS

- Further tuning of hyperparameters could improve model performance.
- Consider collecting more data to improve model accuracy.
- Regularly update the model with new crime data for sustained accuracy

NEXT STEPS

- Implement the classifier in the real-time crime reporting system.
- Monitor the model's performance and retrain periodically.
- Explore additional features that could improve classification accuracy, such as weather data, social events, etc.

CONCLUSION

- The XGBoost model provides a reliable starting point for classifying crimes in Chicago.