Analyzing Crime Trends in Chicago (2001 – 2025)

Patterns, Predictions, and Policy Implications

By

Saikiran Reddy, Salama

Table of Contents

Abstract

The research uses extensive crime data from Chicago spanning from 2001 to 2025 to detect major crime patterns which guide public safety approaches. The exploratory data analysis (EDA) results show theft as the leading crime type which mostly happens outside residential areas while violent offenses mainly occur in residential neighborhoods. The seasonal examination reveals violent offenses increase most significantly during summer months. The analysis of arrest data reveals that weapons violations and domestic incidents result in immediate arrests, but theft and burglary cases rarely lead to arrests. The research implements predictive analytics through XGBoost for arrest predictions and employs both LSTM and Facebook Prophet models for time-series crime forecasting. The XGBoost model shows moderate success because it identifies crime type and location and temporal factors as crucial elements for predicting arrests. The forecasting models successfully track seasonal patterns while enabling basic predictions about upcoming criminal activities. The research explores limitations and ethical concerns about bias and transparency while stressing the need for fairness in predictive policing. The research findings offer practical insights which help law enforcement agencies make better decisions about resource distribution and implement proactive policing strategies.

Introduction

Modern law enforcement relies heavily on crime analytics to transform resource allocation and prevention strategy development (Perry, McInnis, Price, Smith, & Hollywood, 2013; Ferguson, 2017). Crime analysts who use historical crime data to identify patterns can make predictions about areas and times where criminal activity tends to occur (Ratcliffe, 2016). The use of statistical information through predictive policing enables law enforcement to move from traditional reaction-based methods toward proactive prevention measures (Ferguson, 2017; Perry et al.,

2013). The data allows police departments to raise patrol presence along with community programs in areas with elevated crime rates known as hotspots (Weisburd & Telep, 2014). The purpose extends beyond efficient crime response to include crime prevention through analysis of past trends (Ratcliffe, 2016). Researchers alongside community supporters question the equity and operational success of predictive policing because historical prejudice in data sources might generate unbalanced forecasts (Lum & Isaac, 2016; Richardson, Schultz, & Crawford, 2019).

Crime trend analysis finds a valuable application through the Chicago case study. The United States' third-largest city (U.S. Census Bureau, 2020), Chicago has faced extensive documented problems regarding violent and property crime rates. Through its open data portal Chicago provides access to an extensive crime dataset that includes more than 8 million reported incidents from 2001 until 2025. The study investigates whether violent crimes increase during summer months while property crimes reach their peak during winter months according to past criminological theories. We must understand both the frequency of arrested crimes particularly identifying the types of crimes that most frequently result in an arrest-and the potential to forecast arrest outcomes through machine learning analysis of incident circumstances. This research project seeks solutions to these inquiries by using combined methods. We start by determining which crimes occur most frequently in Chicago while analyzing their distribution between residential and non-residential (commercial/public) areas. We analyze the distribution of violent and property crimes across different times and seasons to determine their separate patterns. The third section evaluates arrest results by determining which offenses generate arrests most frequently along with identifying the most effective variables for predicting arrests from given circumstances. We accomplish these goals through a combination of Exploratory Data Analysis (EDA) and

sophisticated modeling techniques which include gradient-boosted trees neural networks and time-series forecasting tools.

Our research adds both operational value for police strategies and knowledge about how urban crimes behave. The research results enable better resource management by pinpointing specific crime areas where theft and domestic violence occur which need targeted intervention. The time-based discoveries about crime patterns including summer violence increases provide direction for officer and community outreach deployment during specific seasons. The study demonstrates machine learning applications in crime analysis by developing predictive models yet emphasizes the importance of ethical considerations like preventing biased enforcement in practical implementation.

This study investigates how different types of crimes spread out in Chicago and focuses on figuring out the common offenses as well, as how they relate to residential and non-residential areas. This research also delves into trends and time-based patterns to distinguish between crimes and property related offenses effectively. Furthermore, the report delves into how arrests made for each type of crime and uses advanced technology like machine learning models to identify the factors that impact arrest outcomes. The subsequent analysis then uses explanations and predictive tools to cover these aspects.

Literature Review

Crime analytics informs proactive policing (Ourania, Alina, Araujo, & Leitner, 2020). Analysts use statistical analysis together with GIS and machine learning algorithms to detect crime patterns while predicting crime probabilities across different spatial locations and time periods. Predictive policing represents a new approach to law enforcement which uses data analytics to shift from traditional reactive methods toward preventive measures that optimize resource deployment and

enhance public safety (Mugari & Emeka, 2021). Big data analytics in policing has generated substantial controversy about its effectiveness and fairness during its development. This review evaluates present-day crime data analytics research by focusing on recent predictive modeling techniques and Spatio-temporal analysis methods as well as current law enforcement practices and existing challenges with prejudice and transparency issues and potential transformational changes. Research studies have shown that developers have created new crime analysis tools which have been tested, and Spatio-temporal modeling has proven essential because analysts understand that criminal activity clusters in both space and time (Yingjie & Ning, 2023). The authors of Butt et al. (2020) conducted a systematic review of crime hotspot detection and prediction techniques which showed that current methods use various strategies. Researchers typically use clustering algorithms to identify high-crime areas together with time series analysis for tracking temporal patterns and deep learning models for complex data pattern analysis. Research has demonstrated that neural networks including CNNs and LSTMs show better performance than traditional methods when predicting crime rates or crime locations in analysis tasks. The research indicates that future crime forecasting will progress beyond basic regression analysis and hotspot mapping by adopting AI-based prediction systems.

The implementation of various data types represents an essential progress in the field. New crime prediction models combine traditional historical crime incident data with additional elements such as demographic information and land use data and social media post content to improve their predictive accuracy (Mohler et al., 2015). The analysis of geotagged Twitter data together with crime records showed Matthew (2014) that the prediction of specific crimes in Chicago such as thefts and assaults and stalking became slightly more accurate through the identification of typical area activities that correlated with criminal activity.

The implementation of big data analytics (including socio-economic indicators and real-time feeds) aligns with the overall goals of smart city crime analytics initiatives. The multi-scale prediction models described by Yingjie and Ning (2023) enable analysts to make precise forecasts through their examination of short-term, medium-term and long-term horizons and micro-level street predictions and meso-level district predictions and macro-level city predictions. These models enable analysts to make accurate forecasts that range from identifying neighborhood burglary hotspots for the upcoming week to predicting annual crime patterns throughout the entire city by selecting appropriate temporal and geographical scales.

The development of predictive policing applications has occurred through the implementation of these analytical enhancements in police departments throughout the world. The most common application of predictive policing involves utilizing algorithms to determine crime-prone locations so police can allocate their presence to those specific areas. Mugari and Emeka (2021) report that numerous cities have implemented systems such as PredPol (which operates under the name Geolitica) and Risk Terrain Modeling and near-repeat hotspot theories. The real-world application of crime data analytics manifests through these tools because model outputs display predicted crime locations on maps that guide operational decisions about officer deployment. Multiple research studies have been performed to evaluate their success rates. The field trial conducted by Mohler et al. (2015) in Los Angeles employed an epidemic-inspired point process predictive model which led to aa 7.4 reduction in crime volume and outperformed traditional hotspot policing methods that relied on human analysts. The results of predictive policing implementations vary because some areas experience measurable benefits, but other locations fail to achieve noticeable outcomes thus proving that context and execution matter. Meijer and Wessels (2019) highlight that predictive policing tools have widespread adoption, yet the actual results do not match the expected

outcomes because studies demonstrate both minimal crime reduction and no noticeable effect and the available evidence lacks substantial proof of extensive benefits. The systematic review conducted by Lee et al. (2024) discovered that data-driven predictive policing had strong supporting evidence in only six out of 161 examined studies. Predictive policing momentum exists primarily through strong examples instead of solid evidence so independent assessments must be conducted to establish genuine evidence.

Crime data analytics serves multiple purposes because it extends beyond spatial prediction since some programs concentrate on analyzing individual behaviors including offender recidivism prediction and victim vulnerability assessment. However, this review focuses on spatial analytics. Person-based predictive systems including offender risk scores and potential shooter "heat lists" have received attention from researchers who discuss both positive and negative aspects of these systems (Mugari & Emeka, 2021). The status of crime analytics demonstrates advancement because law enforcement agencies implement these tools, but their effectiveness remains unclear thus requiring additional research. The literature consensus supports that data-assisted policing offers potential benefits to enhance public safety when properly implemented. The obstacles that hinder their success such as bias and transparency issues need proper resolution.

Methods

Data Source and Preparation

Our analysis used the City of Chicago Open Data Portal crime dataset which contains the "Crimes from 2001 to 2025" collection. The dataset includes all recorded crime incidents that occurred within Chicago from January 1, 2001, until early 2025. The dataset contains about 8.14 million

records which represent individual crime reports and each record contains 22 attributes. The analysis makes use of the following key variables.

Variable	Description
Primary Type	Broad classification of the crime (e.g., Theft, Assault, Homicide).
Location Description	Specifies the crime scene (e.g., Street, Residence, Bank).
Date/Time	The date when the crime was reported or estimated to have occurred.
Arrest	Indicates whether an arrest was made (True/False).
Domestic	Whether the incident is classified as domestic-related (True/False).
Community Area	One of 77 designated community areas within Chicago.

The study did not examine additional fields (like IUCR codes, beat, ward, etc.) even though they were present. We cleaned the data before analysis to guarantee its reliability by taking care of missing values records with missing location descriptions or community area were omitted and by properly formatting the date-time (we split the date and time into separate columns and created Month, DayOfWeek, Hour columns as needed). We also adapted data types to suitable levels for memory efficiency because of the extensive dataset. We performed geospatial consistency checks including tests for valid Chicago latitude/longitude coordinates although we used the provided categorical location fields instead of raw coordinates.

We assigned every crime incident to either Residential or Non-Residential categories based on the Location Description. The "Residential" category included all descriptions that referred to houses and dwelling places (e.g. Residence, Apartment, House, Yard, Porch, Domestic residence). All remaining areas including streets and businesses along with transit and parking lots received the simplified classification of "Non-Residential (Commercial/Public)." A two-category system

permitted us to evaluate the number of incidents that took place in residential areas compared to non-residential areas. The classification method provides straightforward results even though it merges various public place categories into one group because it directly answers our research question about general location context. The creation of a "Crime Category" variable helped us distinguish between violent crime and property crime for seasonal analysis. Violent crimes included offenses that involved physical force or threats against people (homicide, assault/battery, robbery, criminal sexual assault, etc.) whereas property crimes involved theft or damage to property which did not result in personal injury to victims (burglary, larceny/theft, motor vehicle theft, arson). Narcotics or Weapon violations failed to align with either violent or property crime categories thus we analyzed only the main FBI index crime categories (violent vs property) while omitting ambiguous categories to ensure a clear distinction.

Exploratory Data Analytics (EDA)

We analyzed the complete 2001–2025 period by counting incidents according to Primary Type to determine the most frequent crime types. The analysis produced a list of crime categories based on their total number of occurrences. We analyzed the leading crimes through location-specific analysis by determining the percentage of each major crime type that took place in residential versus non-residential settings. Bar charts display overall frequencies and side-by-side or stacked bar graphs to present location-specific data.
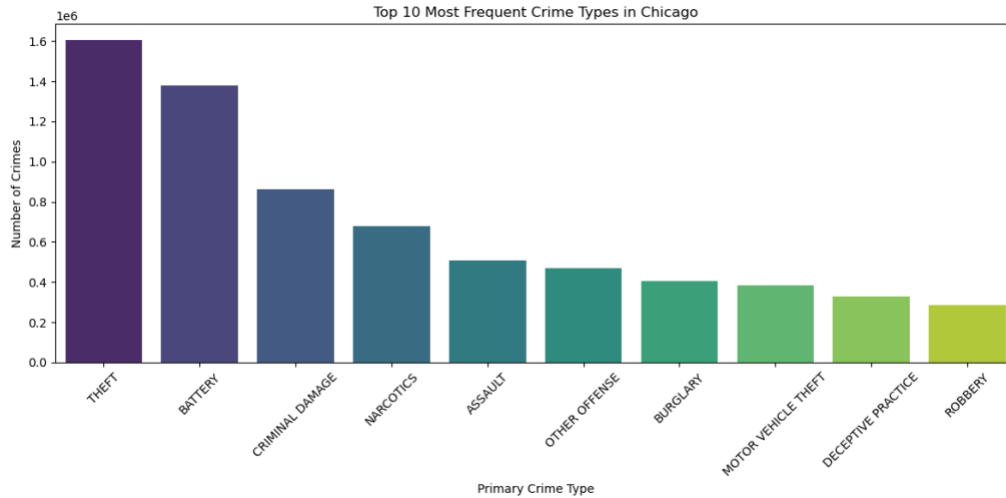
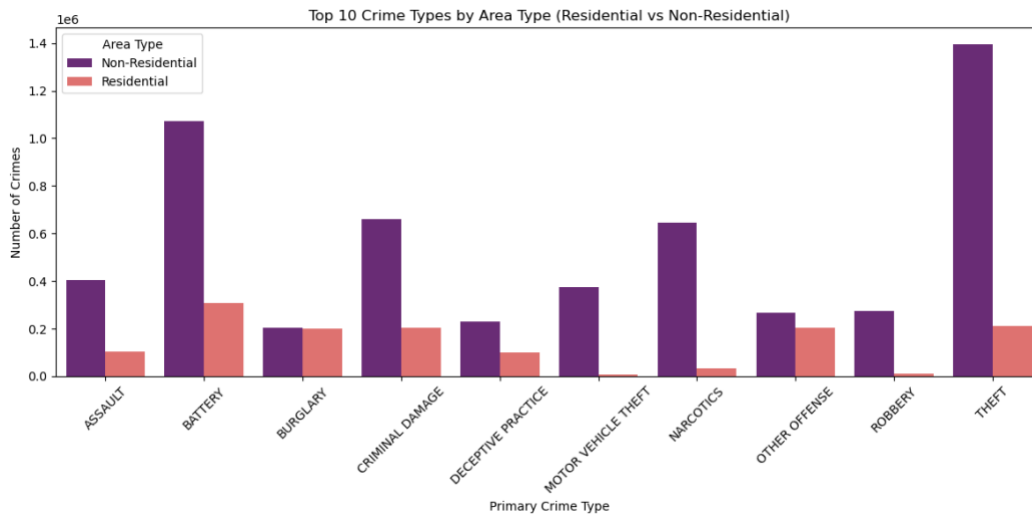*Figure 1 Shows the top ten crime types and their overall frequencies*

Figure 2 Illustrates the percentage of each top crime occurring in residential vs. non-residential locations.

We analyzed seasonal patterns by combining crime data with monthly frequencies and violent versus property crime categories. We calculated yearly and yearly averaged crime statistics for violent offenses and property offenses to minimize annual statistical fluctuations. The analysis resulted in separate seasonal patterns for violent crimes and property crimes.
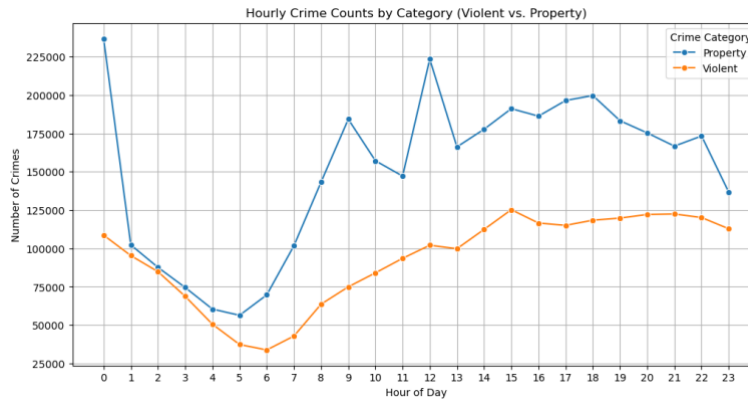
Figure 3 Shows the average hourly crime rate for violent vs. property crimes.

The visual aids show temporal differences by revealing that violent crime rises more steeply during evenings than property crime and that violent crime shows a larger summer increase than property crime. We ran an interaction model as a statistical test (two-way ANOVA with factors for crime category and month) to determine if the seasonal pattern varies between crime categories. The test reveals whether violent crime shows a larger summer increase than property crime.

Most Frequent Crime Types:

The evaluation of 8.14 million crime records shows theft stands as the most frequent crime type in Chicago from 2001 through 2025. Most reported incidents during this period consisted of theft offenses which included pickpocketing and shoplifting and theft from buildings. The larceny-theft rate in Chicago during 2021 reached 481.5 per 100,000 inhabitants making theft the most widespread crime category. The analysis of our complete dataset showed theft as the most common offense followed by battery and criminal damage and narcotics and assault. These top five categories made up most recorded crimes. *Figure 1 Shows the top ten crime types and their overall frequencies.* Homicides and other well-known serious crimes remained infrequent throughout the 23-year period because they occurred fewer than 300 times per year resulting in only a few thousand cases while theft incidents reached into the millions. Property crimes together with minor

assaults form the main drivers of Chicago's crime statistics even though they do not diminish the importance of other crimes.

Predictive Modeling

Which crime types lead to arrests most frequently, and what features are most predictive of arrest outcomes?

To answer this question, we implemented multiple modeling techniques like XGBoost, FB Prophet, and LSTM.

Arrest Outcome Prediction (XGBoost Classifier)

We trained a classification model to predict the Arrest outcome (True/False) for individual crime incidents. We selected the XGBoost model due to its strong performance with structured data.

The prediction model leveraged factors to boost its accuracy; in making predictions it considered the main type of crime committed (Primary Type) where the crime took place (Location Description. Whether in residential or non-residential areas) and temporal aspects like the time of day (Hour) month of occurrence (Month) and the day of the week it happened (Day of Week). To enrich the understanding of each event furtherly; markers indicating Domestic Incidents and geographical details such, as Community Area and Police District were also considered. The prediction model used Primary Type of crime as an input feature along with Location Description which included both residential/non-residential flags and detailed categories through one-hot encoding of popular locations. The prediction model included temporal features such as Hour of day Month and Day of week alongside Domestic incident markers and Community Area or police district information. Our goal was to determine which elements play the most significant role in determining arrest outcomes. The model received training data from a stratified random sample of several hundred thousand cases because 8M was too large (the reduced dataset maintained a fair

representation of the distribution). The evaluation of the model took place on a separate dataset which was set aside for testing purposes. We conducted 5-fold cross-validation on the training data to establish robust performance metrics which included accuracy precision recall and ROC-AUC. The model needed special handling for class imbalance because arrests occur less frequently than other incidents, so we applied either scale_pos_weight in XGBoost or resampling methods to prevent constant No Arrest predictions. The model generated feature importance scores which determined the most influential inputs for prediction.

We anticipated that crime type and perhaps location or domestic flag would be top predictors, which indeed was the case.

Feature	Importance
Primary_Type_enc	0.709241
Area_Type_enc	0.211334
Domestic	0.043503
Hour	0.020545
District	0.010725

Figure 4 Summarizes the top 5 features and their relative importance values.
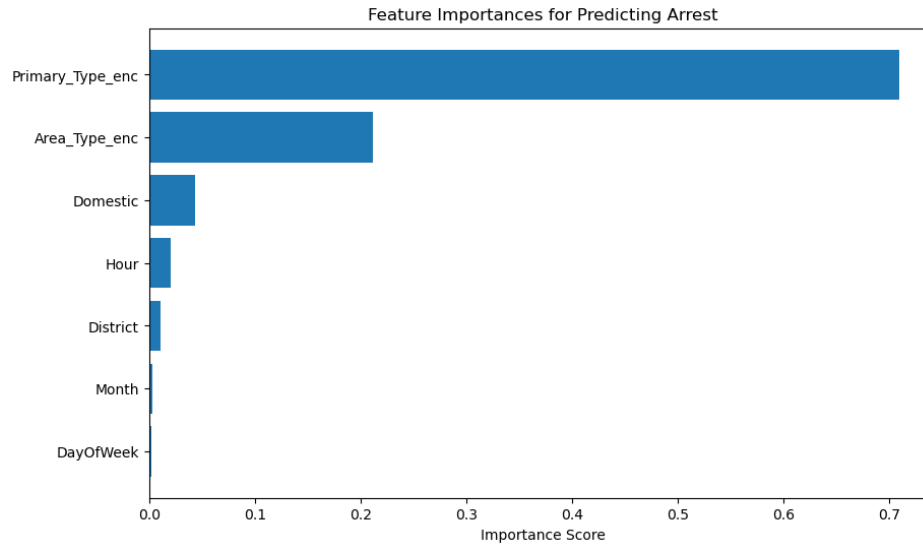
Figure 5 Displays a feature importance plot from XGBoost Model.

Interpretation of Feature Importance

The analysis of features, in the XGBoost model shows that the type of crime being the factor influencing arrest predictions with indicators, like 'Area Type' and 'Domestic Incident' following closely behind in importance rankings. It's important to approach this finding as the model's predictions are not solely based crime attributes but also consider wider law enforcement procedures and practices in play. Crimes categorized as serious or violent tend to have chances of leading to arrests since they receive greater attention and priority from law enforcement agencies. Property crimes may sometimes not result in arrests because of evidence or when law enforcement places a priority on such cases. Moreover, the chances of an arrest are influenced by how resources allocated the discretion of officers and procedural aspects, like deciding on pressing charges or carrying out investigations. It's important to recognize that the model doesn't just focus on crime traits but also considers the interactions, within policing and legal systems.

Time-Series Forecasting (Prophet and LSTM)

We applied two forecasting techniques which included Facebook Prophet as a decomposable time-series model (additive model with trend, seasonality, and holiday components) and an LSTM (Long Short-Term Memory) neural network which demonstrates sequence learning capabilities. The forecasting target focused on measuring the quantity of criminal incidents per period. The analysis started with daily crime statistics, yet we shifted to monthly citywide totals because daily counts showed excessive fluctuations. The change in approach did not require us to abandon yearly pattern analysis. Prophet maintained the entire date structure to detect yearly seasonal input for annual pattern detection and monthly time series data for trend analysis. The LSTM model received monthly total data for training while maintaining the complete date index to preserve time order and enable the model to learn seasonality and long-term patterns.

Our models received training data through 2022, with predictions generated for the period from 2023 to 2024. The established framework allowed us to check some 2023 predictions by using the expanded dataset up to 2024. The development of distinct predictions for violent crimes and property crimes allowed us to analyze their separate trend patterns. Model accuracy was evaluated using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) on a validation

dataset. The Prophet model implemented time-based cross-validation (rolling origin forecast evaluation) as a built-in feature to detect potential overfitting.
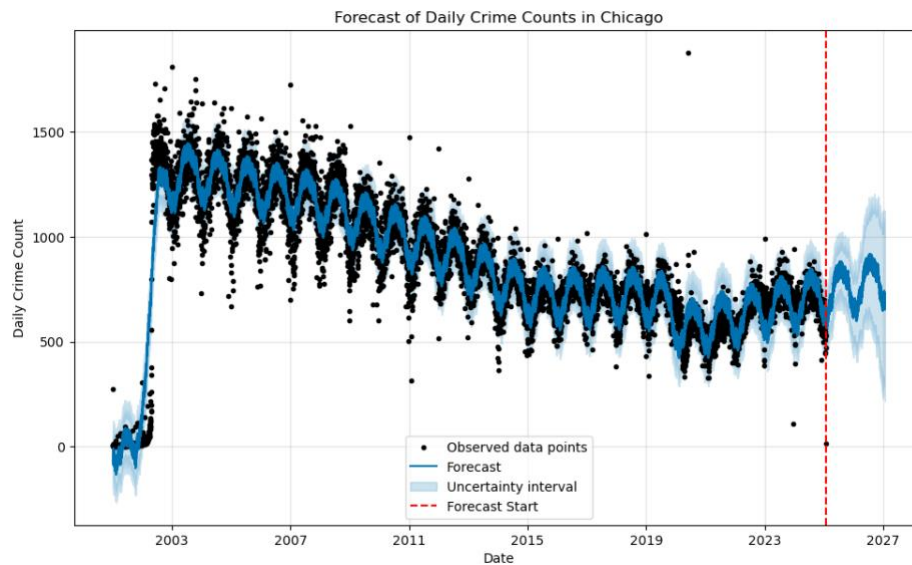

Figure 6 Illustrates the comparison between the Actual vs. Forecasted crime counts for 2027
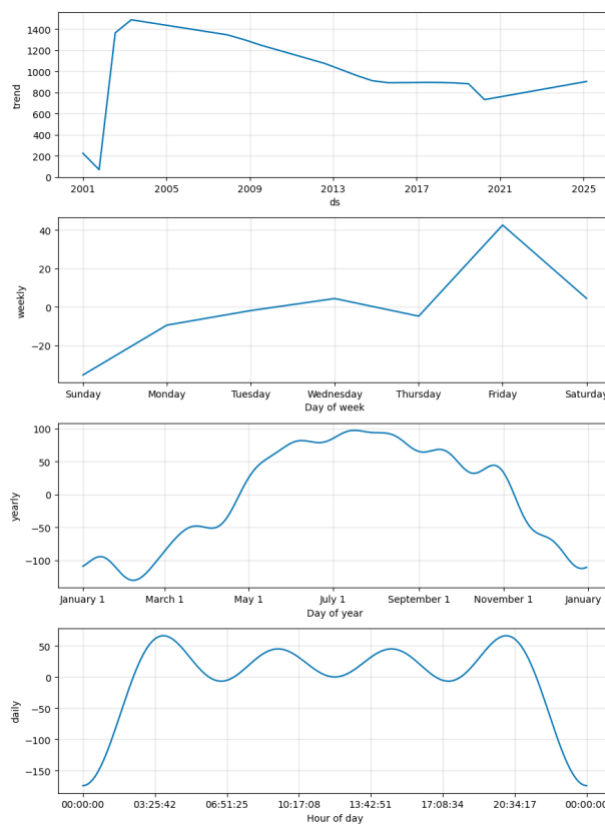

Figure 7 Shows the Seasonal Components for Overall Crime.

We were aware of model limitations during modeling: for example, the crime forecasting does not incorporate external variables (like economic indicators, weather, or policy changes) which certainly affect crime but are beyond the scope of this dataset. Similarly, the arrest prediction model does not include suspect-specific information (like prior records) that might influence an officer's decision to arrest or not. it purely relies on incident characteristics. We treated our models as exploratory tools to glean insights (e.g., "what factors matter most for arrests?" or "does the data show a consistent seasonal wave we can exploit for prediction?") rather than as deployable systems. The following section shows the results of these analyses in detail.

Results

Crime Frequency and Location Analysis

Residential vs. Non-Residential Patterns: The analysis reveals that particular types of crimes tend to appear in specific geographical settings. The analysis of crime incidents into Residential and Non-Residential areas revealed separate patterns. Among the most frequently reported crimes theft and narcotics offenses primarily take place outside residential areas. Most theft incidents occur in public spaces along with commercial areas. The high number of thefts along with narcotics offenses makes sense because these crimes often occur during busy times in public areas such as streets and commercial spaces. Retail thefts along with vehicle break-ins which fall under theft category primarily occur outside residential areas while drug arrests take place in public drug markets and vehicles. The majority of specific crime categories show a strong preference for residential locations. Domestic battery as a battery subset always occurs in homes because it involves conflicts between family or household members thus these incidents naturally appear in residential areas. Burglary (unlawful entry to steal) is another crime strongly associated with residential locations – burglaries often target houses or apartments (though there are also burglaries

of businesses, many involved entries into homes). Private residences serve as common locations for criminal sexual assaults to take place. Most robberies which involve theft through force or threats take place on streets and commercial areas including muggings and convenience store hold-ups thus showing alignment with non-residential areas despite their violent nature. *Figure 2 shows the proportion of incidents for selected frequent crime types that took place in residential versus non-residential locations.*

Our data showed that about 60% of burglaries took place in residential areas (home invasions) while narcotics incidents occurred mostly in non-residential/public locations at more than 85% (drug transactions or police stings take place outside). Most thefts occurred outside homes because retail theft and vehicle thefts took place on streets while residential thefts included only a few cases such as guest residence thefts and porch package thefts. The assault and battery cases divided into two groups because simple assaults frequently take place in homes and private residences, yet many incidents happen in public locations such as bars and streets. The crime prevention strategies need to be location-specific because residential neighborhoods should concentrate on burglary prevention through community watch and better locks and lighting and domestic violence interventions, yet downtown and commercial districts require strategies for theft prevention through improved surveillance and anti-pickpocketing measures and drug-related crime control through narcotics enforcement or treatment programs.

Seasonal and Temporal Patterns (Violent vs. Property Crimes)

Seasonal Trends

Our analysis reveals an intense yearly pattern of crime in Chicago especially for violent crimes. The analysis of crime statistics from 2001–2025 revealed that violent and property crime rates reach their highest points between June and August and their lowest points between December and

February. The seasonal pattern of violent crimes and property crimes becomes apparent when looking at *Figure 2* which displays the monthly average of each type of crime.

During summer the number of violent crimes shows significant growth with July standing as the month with the highest number of violent crimes followed by August and June. The lowest number of violent crimes occurred in February which is typically the coldest month of the year in Chicago. Summer saw an increase in property crimes but the variation between summer and winter months was less dramatic than with violent crimes. The number of thefts and burglaries rose in summer possibly because more people were outside creating more targets and opportunities while seasonal factors including vacationing residents and unattended vehicles during outings contributed to the increases. The frequency of arson crimes shows two notable spikes during July associated with Fourth of July fireworks and civil disturbances as well as in October linked to Halloween events. The ANOVA test demonstrated significant monthly differences ($p < 0.001$) in violent crime and property crime data indicating that monthly variations exceed random chance. The percentage increase between winter and summer reveals that violent crime reacts more strongly to seasonal variations than property crime does. The relationship between warm weather and increased aggression alongside social interaction supports the observed seasonal patterns of violence. The data supports that violent incidents become more common during summer, yet property crimes show different contributing elements while particular property crime types might occur during winter months thus moderating the peak of property crimes.

Statistically, we performed a two-way ANOVA analysis to examine Crime Category (violent/property) and Month as independent variables. The Category × Month interaction term reached statistical significance at ($p < 0.05$) which showed that violent and property crimes

exhibited different monthly patterns. Specifically, violent crimes showed higher rates during summer months, while property crimes were more evenly distributed throughout the year.

ANOVA analysis has its limitations because it fails to consider yearly changes and other variables including location type and time of day and specific crime types. Our interpretation remains limited to general seasonal patterns because we cannot determine exact temporal patterns. The initial exploratory analysis used ANOVA to detect seasonal patterns, but we confirmed these results through time-series models including Prophet and LSTM which handled temporal trends and seasonal patterns and yearly variations. The models delivered a more detailed understanding of seasonal crime patterns through analysis of multiple years of data.

We conducted post-hoc tests to compare month-by-month values for each crime category. The results showed that every summer month produced higher violent crime rates than winter months and most of these differences reached statistical significance at the 0.01 level. The property crime pattern showed similar characteristics although some month-to-month comparisons were not as pronounced.

Research suggests that weather conditions together with changes in daylight hours drive seasonal patterns in crime rates. The cold winters of Chicago probably reduce criminal activity because people stay home and there are fewer chances to meet others on the streets, yet summer heat and outdoor activities create more criminal opportunities and stressors. The observed results match both the existing theoretical explanations and previous studies conducted in different cities. The data shows that theft rates (property crime) reached their highest point during summer months even though traditional research suggested winter would see increased property crime. The modern urban environment produces such high summer activity levels that all types of crime increase together with violent crimes. Some particular property crimes such as residential burglary might

have different patterns than other thefts, but the combined effect of all property crimes showed a summer trend.

Temporal (Hourly and Daily) Patterns

The study investigated the differences, in patterns of property crimes throughout the day and week. Both categories exhibited levels of activity from 4 to 6 AM. Saw an increase in incidents during the morning hours. Peak times for both types of crimes were observed between 6 PM and 11 PM which coincides with evening rush hours and social gatherings. Violent crimes such as assaults and robberies showed a spike during late night hours (from 8 PM to midnight) whereas property crimes like theft and burglary continued into hours due, to shopping times and vacant residences. Property crimes also saw an increase during midday hours due to incidents like shoplifting and vehicle thefts as depicted in Figure 3 which showcases the balance, between violent and property crimes.

According to the report findings revealed an increase, in crime rates during weekends specifically with violent crimes reaching its peak on Saturdays possibly influenced by social gatherings and gang related incidents while property crimes spiked on Fridays due to heightened shopping activities, around the area overall Chicago's daily crime rates remained relatively steady with Mondays recording the lowest figures.

During seasons of the year violent crimes tend to spike in the summer while property crimes stay relatively steady throughout. Violent crimes also tend to peak in the day reflecting the effects of social gatherings. On the hand property crimes, like burglary follow timing patterns as they often happen during daylight hours when houses are unoccupied. Summer nights pose a risk, for offenses while property crimes are more spread out affected by shopping and work schedules.

Arrest Outcomes and Predictive Factors

Arrest Rates by Crime Type

One of the discoveries, from our exploratory data analysis was the different arrest rates associated with various types of crimes. We calculated the proportion of incidents in each crime category where an arrest was made – when Arrest = True. This generally indicates arrests made on site. After but doesn't encompass arrests made later due to continuing investigations. This differentiation is crucial as it influences our understanding of what "arrest" signifies, in our dataset. Crimes involving violence, like weapon violations and domestic violence incidents had the rates of arrest at times exceeding 50%. Cases of weapons violations often stem from policing measures such as stops or raids leading to arrests in illegal firearm possession situations. Public indecency cases show arrest rates well despite being uncommon due to the evident nature of the offense happening in front of law enforcement officers. Incidents of violence frequently result in individuals being apprehended because of regulations and guidelines established to safeguard victims. Requiring law enforcement to detain the perpetrator in cases of domestic disagreements. Property crimes such, as theft and burglary have rates of immediate arrests at the other end of the spectrum in contrast to violent crimes like assault or robbery which often lead to prompt apprehensions by law enforcement authorities according to our data analysis results Theft incidents typically end in apprehensions only in a small fraction of cases (less, than 10%) reflecting the common scenario where thefts are reported post incident making it challenging for law enforcement agencies to nab the culprit red handed Similarly burglary cases exhibit a similar trend where the crime is typically unearthed after its occurrence reducing the likelihood of quick arrests Burglary was also low – a homeowner comes home to find a break-in, but an arrest at the scene is unlikely unless there is an eyewitness or alarm trigger which leads to a quick police response. Arson also had low on-scene arrest rates because fires are responded to by the fire departments

and investigating a perpetrator takes time and any eventual arrest comes later and may not be linked in the dataset if the case is not immediately cleared. Assault and battery cases were moderate as for arrest rates (higher if the assault is domestic or if the police arrive during the incident, lower if the offender fled). Robberies (a violent crime) also had moderate arrest rates at best – many street robberies do not result in the immediate capture of the suspect unless the police are very nearby. These disparities suggest that the nature of the crime (how it is typically discovered and handled) strongly influences whether an arrest is made quickly.
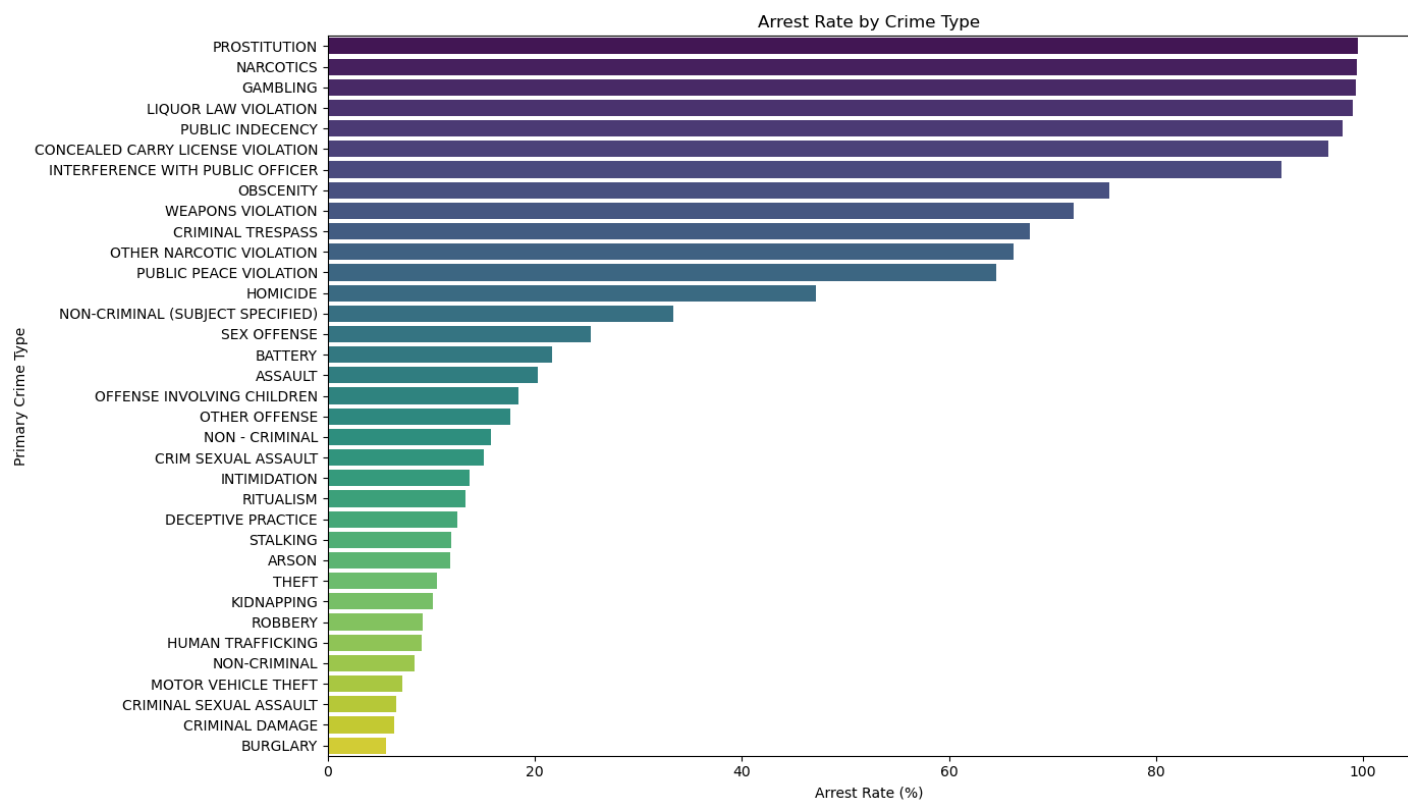

Figure 8 Presents examples of arrest rates for select crimes.

Features Predictive of Arrest (XGBoost Model Results)

We further analyzed this by training the XGBoost classification model to predict the Arrest outcome using various incident features. The model got a cross-validated accuracy of about 75% distinguishing arrest vs. no-arrest cases. For context, simply guessing "No Arrest" for every case

(the naive baseline) would be right about, say, 70% of the time if 30% of incidents have arrests – so the model does offer an improvement over baseline, and more importantly, provides insight into which factors drive those predictions. The feature importance rankings from the trained model confirmed several intuitive drivers:

The most influential feature was Primary Type of the crime. The model has a strong bias towards one direction or the other depending on the type of crime as discussed (for example, if the incident is coded as a narcotics violation, the model will heavily lean towards predicting an arrest, whereas if it is a theft, it will lean towards no arrest). The model learned these associations well.

The second important feature was the Location Description (or our residential vs non-residential flag). This may serve as a proxy for context: for example, crimes that occur at a "Residence" may correlate with domestic incidents which have higher arrests, whereas "Street" may correlate with crimes where the offender can flee more easily. In the detailed importance, specific location categories like "Gas Station" or "CTA Train" had their own weights – perhaps reflecting that crimes in certain public transit areas often involve quick police response (there are police units for transit), etc.

The Domestic indicator was another key feature. Domestic incidents (regardless of type) have a higher probability of arrest because many domestic violence cases have mandatory arrest policies. The model found this flag informative – any domestic-related incident boosted the likelihood of an arrest prediction.

Time of day and Day of week had some influence too. The model noticed patterns such as: Crimes committed at night may be less likely to result in immediate arrests because there are fewer witnesses or police patrols (except in nightlife areas), whereas crimes committed during certain peak patrol hours or shifts may result in quicker arrests. For instance, the model may have

identified that incidents that occurred in the very early morning hours (3-5 AM) had fewer arrests (perhaps because there were not many police around or the crimes were discovered later), whereas early evening incidents might have more. Day of week had a smaller effect, but possibly certain days (like weekends) could influence arrest likelihood if police deployment differs or if weekend incidents are more chaotic.

The Community Area/District feature also became important, but we need to be careful about how to interpret this, it could be due to differences in police staffing or efficiency across districts, or differences in crime mix. For example, one district may have a higher arrest rate overall because of proactive policing. The model can use area as a proxy for these unobserved factors. We did see that downtown district (where many thefts happen) had low overall arrest probabilities, whereas some more residential districts or those with targeted police initiatives had higher ones.

The model's predictions showed a high level of accuracy in detecting no-arrest cases through its true negatives but failed to recognize some arrest cases as false negatives. The model received adjustments to boost its ability to detect arrest cases because law enforcement agencies find it more valuable to identify arrest situations or to determine scenarios where arrests are unlikely for resource allocation purposes after incidents. We made a threshold modification to achieve proper precision-recall equilibrium. The final model achieved precision at 0.6 and recall at 0.5 for the arrest class because individual prediction accuracy was not expected to be high due to random factors and unmeasured elements in officer suspect apprehensions, but it provided sufficient results to determine relevant features.

	Predicted: No Arrest (0)	Predicted: Arrest (1)	Total
Actual: No Arrest (0)	1,106,148 (TN)	22,575 (FP)	1,128,723
Actual Arrest (1)	1,70,809 (FN)	208,766 (TP)	379,575
Total	1,276,957	231,341	1,508,298

Figure 9 Confusion Matrix.

The model processed information according to its internal logic which matched the EDA by determining that weapon offenses led to arrest probabilities while thefts did not. The model produced an interesting discovery about feature interactions when it assigned higher arrest probabilities to domestic thefts compared to regular thefts because the domestic label indicates the offender's presence at the scene. The model assigned higher arrest probabilities to cases with "School" locations because of dedicated school officers and clear suspect identities but lower probabilities to "Alley" locations because the offender had probably escaped. The model incorporated numerous specific patterns which were too complex to list but it stored them within its intricate decision trees.

The predictive model confirmed that crime type stands as the leading predictor of arrest outcomes followed by location and domestic involvement as contextual factors. The model's performance demonstrates that arrest probabilities have some degree of predictability, yet they also contain substantial unpredictable elements. The dataset does not show the complete picture because crimes without immediate arrests may get solved in the future and various elements such as suspect identity and victim cooperation or police response time remain outside the data range.

Crime Count Forecasting (Prophet & LSTM)

The inclusion of forecasting results as a research element emerged during our exploration of data predictability although it was not listed as one of the original research questions. The Facebook

Prophet model analyzed monthly total crime counts from 2001–2025 by separating the time series into both general trends and seasonal patterns. The overall crime pattern in Chicago exhibited decreasing rates during the initial ten-year span before rising briefly from 2016 through 2017 because of increased violence in 2016 then showing inconsistent patterns before dropping in 2020 because of pandemic restrictions and later rising back up. The Prophet model identified a minor negative trend while adapting to both the increased and decreased crime rates. The seasonal pattern revealed summer as the peak season while winters represented the lowest points which supported our previous findings. The Prophet model predicted future seasonal patterns along with a slow decrease in total numbers by maintaining a trend towards the mean value of recent years. Prophet made a prediction for 2023 which included a summer crime rate increase that exceeded 2020 numbers but remained below the 2017 peak levels. We compared these forecasts to actual 2023 numbers (which we have through 2023 in the data): The model correctly showed the summer increase after winter but its predictions for crime levels in this period were lower than actual numbers by several percentage points. The MAPE (Mean Absolute Percentage Error) for the 2023 monthly predictions was around 8 to 10%, which is fairly good for this context. The LSTM model learned the seasonal pattern but training an LSTM with 270 data points (12 months × 22.5 years) proved difficult because of limited data even after performing input sequence windowing. Prophet showed comparable performance to the LSTM model when forecasting one year ahead but required precise tuning to achieve better results. Prophet provided more straightforward interpretation because it generated seasonal results directly.

The main outcome from this forecasting exercise shows that Chicago crime rates follow a regular yearly pattern, but short-term fluctuations occur because of social and political events and policy changes. The models would show better performance when additional covariates such as

unemployment rates and police staffing levels and weather extremes are included to capture external influences. A simple time-series model would provide officials with advance notice about summer increases so they can make necessary preparations for potential violence based on projected high peaks from recent patterns. The decline in forecasts should be interpreted as successful interventions or pattern changes but one should remain cautious since forecasts do not determine future outcomes.

Conclusion

The research aimed to analyze Chicago crime data between 2001 and 2025 while implementing predictive modeling approaches to discover additional information. The findings validate expected patterns while revealing unexpected information which can guide policy decisions. We analyze the findings for each research question together with their relation to existing literature and operational factors.

Frequent Crimes and Location Context

In our study findings show that theft is the type of crime, in Chicago according to crime statistics. This aligns with data which highlights theft as the reported crime category, in the United States (FBI 2023). The significant number of theft incidents, alongside offenses, like assaults and drug related issues highlights the existence of opportunistic crimes that flourish in city settings characterized by social unrest (Ourania et al., 2020).

The way these crimes are spread out shows a trend; frequent offenses tend to happen in public places and commercial areas with a high concentration of people and valuable assets (Butt et al., 2020). The clustering of incidents, in places than areas indicate a significant requirement for specific crime prevention tactics like improving surveillance, in downtown areas and introducing theft prevention measures in retail zones. The research backs the use of small-scale interventions

that highlight the importance of targeted monitoring and raising awareness in commercial zones with high risks (Butt et al., 2020).

In contrast, to that viewpoint when we look at neighborhoods where people live. We can see a trend in criminal behavior unfolding there as well. Thefts like burglaries and incidents of violence tend to happen often in residential areas compared to other places. This suggests that being proactive with policing efforts and having neighborhood watch programs could really make a difference in curbing crimes. The variation in crime trends seen in locations is consistent with what previous research has found. Emphasizing the need for tailored crime prevention approaches to each area (Ourania et al., 2020). Understanding how different types of crimes unfold in locations can provide insights, for police departments aiming to use their resources effectively and implement targeted strategies.

Seasonal Difference between Violent and Property Crimes

Research has consistently shown that violent crimes tend to increase during the summer months due to higher temperatures, which lead to more social interactions and conflicts (Anderson, 2001). The combination of warm weather and longer daylight hours creates opportunities for outdoor gatherings, which sometimes escalate into violent encounters. Our analysis aligns with this pattern, revealing those violent crimes in Chicago peak during summer, placing a greater burden on police and emergency services.

Interestingly, while property crimes generally exhibit less seasonal fluctuation, our study shows a slight summer increase, which contrasts with some academic literature that suggests higher property crime rates during winter. In Chicago, summer property crimes primarily consist of bicycle thefts and thefts targeting tourists, reflecting the city's active tourism season. Our findings

support the need for targeted summer crime prevention strategies, including increased patrols in high-risk areas and youth diversion programs to reduce potential conflicts.

The temporal analysis of crime data further reveals that violent crimes tend to occur at night, particularly between 6 PM and 11 PM, which aligns with social activities and nightlife. In contrast, property crimes show a more consistent daily distribution, with peaks during business hours and early evening, reflecting opportunities for theft in commercial areas. These temporal patterns underscore the importance of strategic resource allocation for law enforcement, such as increased night patrols for violent crime hotspots and daytime surveillance for property crimes.

Arrest Likelihood and Predictors

The differences in arrest rates between various crimes create significant consequences for the criminal justice system. The low clearance rates for minor thefts indicate that numerous victims will not experience prompt justice. The public confidence in the criminal justice system may deteriorate unless these issues receive proper attention. The high number of arrests for particular offenses such as weapons indicates successful policing efforts in these specific domains. Police leadership should use knowledge about rare arrest rates to launch programs that enhance investigation methods and community relations. The low rate of burglary arrests suggests that investing in improved forensic burglary units and community watch programs could increase the chances of catching burglars during their crimes.

The model confirms existing practices because weapon crime investigations produce arrestable cases, and domestic incidents lead to arrests because of established policies. The high arrest numbers require attention because they might indicate either strong enforcement of minor offenses or aggressive policing of small crimes. The aggressive arrest policies of a department can result in nearly all drug incidents ending in arrests because of their low tolerance enforcement approach.

The practice of over-criminalization emerges as a potential drawback from this approach. The historical data from Chicago shows that narcotics offenses produced numerous arrests during the "war on drugs" period but the city started favoring drug treatment instead of imprisonment for minor drug offenses which might lead to decreased narcotics arrests in our 2025 dataset (a yearly arrest rate analysis would show this policy shift).

The ability to predict arrests through predictive policing enables departments to direct their resources after crimes happen by understanding arrest-causing factors. When the model shows that an incident has low chances of immediate arrest such as non-domestic theft in busy areas police will dedicate additional follow-up investigation resources to enhance the chances of solving the case. The predicted arrest outcomes in weapons violations cases usually mean police have already resolved these incidents at the scene thus requiring minimal additional resources. The analysis of arrest patterns allows departments to direct their training resources by providing better investigative methods and community engagement training to officers working in areas with low arrest rates.

Predictive Modeling Efficacy

Our use of XGBoost, Prophet and LSTM demonstrates how contemporary analytical tools operate on crime data. The XGBoost model demonstrates a moderate level of success in predicting individual incident outcomes yet the results depend heavily on random and case-specific elements that influence arrests. The exercise provides valuable benefits for incident prioritization because it helps detectives determine which unsolvable cases require their limited attention. Time-series forecasts demonstrate that general patterns in data are more predictable than specific individual occurrences. The observation makes sense because aggregate crime patterns tend to follow gradual patterns or cyclic trends, but individual crimes and specific results exhibit chaotic behavior. Police

departments should expect recurring seasonal patterns in their workload which enables them to develop budget plans and staffing schedules, but they will never eliminate the unpredictability of crime occurrences or arrest outcomes.

Comparing to Literature and Other Cities

Our research results match previous observations made in other urban crime studies. Most cities identify theft as their primary offense while summer represents their most active crime period and police officers arrest different numbers of offenders depending on the specific offense. Every community maintains its own unique characteristics. The violent crime peak in Chicago stands out because of the city's high levels of gang violence and gun-related incidents which could explain this difference from other cities. The data revealed a major increase in gun-related incidents during the mid-2010s which represents a development specific to Chicago (the 2016 surge). The models recognized historical patterns but exceeded their knowledge limits when attempting to predict unexpected future spikes. Research on crime forecasting demonstrates inconsistent results between studies that use self-exciting point processes for crime analysis (Mohler et al.) and ARIMA models (others) and our approach with Prophet/LSTM which demonstrates acceptable results but might not reach maximum potential with advanced methods. The COVID-19 pandemic presented an unprecedented situation that any forecasting model would have found challenging to predict. The situation requires human experts to use present-day information and intelligence when evaluating model-generated results.

Ethical and Policy Implications

Predictive policing tools, like our arrest prediction model pose a challenge due to biases when trained with data that may reflect unequal law enforcement practices in certain areas. Our study recognizes that past under policing in neighborhoods could influence the model to predict arrest

rates in those regions. This could lead to focus on these areas. Possibly perpetuate existing inequalities. On the hand larger predictions may be made for regions with historical arrest rates supporting increased policing efforts, in those communities.

The moral consequences of these biases have been extensively discussed in studies (referencing Fergusons work in 2017). A case, in point is the heat list" initiative in Chicago's policing program that drew scrutiny for unfairly singling out minority communities and sparking community distrust and legal disputes. Our study emphasizes the significance of utilizing models as aids for making decisions than as standalone systems. Suggesting their use for guidance rather than absolute control over policing strategies while also emphasizing the need, for openness in how forecasts are created and put into practice.

We want to highlight the importance of conducting fairness assessments even though we haven't developed a predictive system in our study yet. In the model's development phase to detect any biases present, in them and guarantee fairness across neighborhoods must include evaluating positive and false negative rates. Furthermore, no model outcomes should unequally affect groups. This is where community feedback becomes crucial in shaping systems to maintain transparency and accountability.

In terms of policy considerations utilizing data driven analyses is crucial, for establishing policing methods. Law enforcement agencies have the opportunity to shift their attention from surveilling infractions in minority neighborhoods to concentrating efforts on combating major property crimes or violent acts in areas where they are most prevalent. For instance, understanding the correlation between temperatures during summer months and a rise in criminal activities could guide the allocation of resources towards proactive initiatives, like summer youth programs and conflict resolution teams.

Limitations

This research has a limitation to consider. The Chicago crime data used in the study is quite extensive; however, it had some details that necessitated the exclusion of entries, with incomplete information (such as absent location descriptions). This exclusion could potentially introduce bias since the missing data might not be distributed randomly. Moreover, the dataset captures crimes that were reported leaving out incidents that went unreported, to authorities, which could skew the crime statistics presented. Furthermore, the Arrest variable oversimplifies a process by indicating solely whether an arrest occurred without distinguishing between arrests and those resulting from prolonged investigations. Moreover, the data covers the years 2001 to 2025 a time frame characterized by shifts, in demographics, policies and technology in Chicago. These alterations could impact crime trends and rates of arrests. Our study concentrated on patterns rather than effects of a given period. It's important to view our results as a summary than an exhaustive scrutiny of all changes, over time.

Reference

Alikhademi, K., Drobina, E., Prioleau, D. et al. Correction to: A review of predictive policing from the perspective of fairness. Artif Intell Law 30, 19–20 (2022). https://doi.org/10.1007/s10506-021-09299-z

Anderson, C. A. (2001). Heat and violence. Current Directions in Psychological Science, 10(1), 33-38. https://doi.org/10.1111/1467-8721.00109

Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial. Statistics and Public Policy, 5(1), 1–6. https://doi.org/10.1080/2330443X.2018.1438940

Butt, Umair Muneer, Letchmunan, Sukumar, Hassan, Fadratul Hafinaz, Ali, Mubashir, Baqir, Anees and Sherazi, Hafiz Husnain Raza (2020). Spatiotemporal crime HotSpot detection and prediction: a systematic literature review. IEEE Access, 8. https://doi.org/10.1109/access.2020.3022808

Cohn, E. G., & Rotton, J. (2000). Weather, Seasonal Trends, and Property Crimes: A Time Series Analysis of Miami, Florida. Journal of Quantitative Criminology, 16(1), 51–74. https://doi.org/10.1023/A:1007540309470

Du, Y., & Ding, N. (2023). A Systematic Review of Multi-Scale Spatio-Temporal Crime Prediction Methods. ISPRS International Journal of Geo-Information, 12(6), 209. https://doi.org/10.3390/ijgi12060209

Ferguson, A. G. (2017). The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. NYU Press.

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, 115–125. https://doi.org/10.1016/j.dss.2014.02.003

Google, NotebookLLM was used for reasoning tasks on the research papers to get ideas about methodologies used in paper.

Kristian Lum, William Isaac. (2016). To Predict and Serve?. Significance, 13(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Lee, Y., Bradford, B., & Posch, K. (2024). The Effectiveness of Big Data-Driven Predictive Policing: Systematic Review. Justice Evaluation Journal, 7(2), 127–160. https://doi.org/10.1080/24751979.2024.2371781

Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. International Journal of Public Administration, 42(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. Journal of the American Statistical Association, 110(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710

OpenAI, ChatGPT 4 was used to brief some of the research papers to write this literature review and formatting the structure of the Research Paper creating Sections to compile the research paper.

Perry, Walter L., Brian McInnis, Carter C. Price, Susan Smith, and John S. Hollywood. (2013). Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR233.html

Ratcliffe, J.H. (2016). Intelligence-Led Policing (2nd ed.). Routledge. https://doi.org/10.4324/9781315717579

Weisburd, D., & Telep, C. W. (2014). Hot Spots Policing: What We Know and What We Need to Know. Journal of Contemporary Criminal Justice, 30(2), 200-220. https://doi.org/10.1177/1043986214525083