**Visualizing and Predicting Crime in Los Angeles: A Data-Driven Dashboard Approach*.***

*Submitted b*y

Group F

Rajesh Anne – 11701768

Amrutha Vamshi Goud Anantha - 11674521

*In partial fulfillment of the credit requirements in*

**ADTA 5940.004**

*Under the guidance of*

**Denise R. Philpot, PhD, Clinical Associate Professor**



**MARCH 2025**

Table of Contents

1. Acknowledgments .........................................................................2

2. Abstract ........................................................................................3

3. Introduction ..................................................................................4

4. Problem Statement ........................................................................5

5. Research Questions .......................................................................6

6. Literature Review ............................................................................6

7. Data Preparation and Overview .......................................................11

8. Data Cleaning and Preprocessing .....................................................13

9. Exploratory Data Analysis ................................................................14-20

10. Modeling ........................................................................................21

- Logistic Regression Model ............................................................ 23

- XGBoost Classifier ....................................................................... 25

- Random Forest Classifier ............................................................. 27

11. Results and Interpretation ..............................................................28

12. Discussion, Summary and Limitations ............................................ 29

13. Conclusion and Recommendations .................................................30

14. Appendix ........................................................................................ 31

15. References ......................................................................................32

### **Acknowledgments**

We would like to express our sincere gratitude to Dr. **Denise R. Philpot, PhD, Clinical Associate** for herinvaluable mentorship and continuous guidance throughout the duration of our research. Her insights and support during the ADTA Analytics Capstone Experience course played a pivotal role in shaping our understanding and execution of the project.

Our heartfelt thanks also go to God Almighty, whose blessings and grace provided us with the strength, clarity, and perseverance to carry out this work to completion.

We extend our appreciation to every member of Group F for their dedicated contributions and collaborative efforts toward the successful completion of our project, “Visualizing and Predicting Crime in Los Angeles: A Data-Driven Dashboard Approach.”

We sincerely hope that this research meets its intended objectives and adds meaningful value to the field of Data Analytics, particularly in the application of machine learning for enhancing safety and informing regulatory decision-making.

# **Abstract**

This project, Visualizing and Predicting Crime in Los Angeles, is based on the analysis and forecasting of crime patterns from historical data of the Los Angeles Police Department (LAPD) between 2020 and 2024. The study makes use of data-driven techniques to identify where and when crimes occur most frequently. Exploratory Data Analysis (EDA) showed that the rates of crime are highest over weekends, especially Saturdays and Fridays, and are concentrated in zones such as 77th Street, Central, and Hollywood. Battery, burglary, and vehicle theft were among the most frequently reported crimes. Tableau dashboards were built to plot these trends by region and over different time periods so that even technical stakeholders could obtain useful insights and make sound decisions concerning safety.

To further elaborate on the analysis, machine learning models were developed to predict types of crime from time and location characteristics. A baseline Logistic Regression model was 61% accurate, while the better XGBoost classifier was 65% accurate. The most influential predictors were hour of day, weekday, and area name. These prediction findings present a real-world foundation for proactive police work, enabling strategic patrol planning and optimized resource allocation. Whereas the project shows the potential of crime prediction, it also recognizes the need for socioeconomic factors in the effort to further enhance model accuracy and applicability to the community in future work.

**Keywords:**  
Crime Data Analysis, Predictive Policing, Exploratory Data Analysis, Machine Learning, Tableau, XGBoost, Logistic Regression, LAPD, Crime Forecasting.

**1.Introduction**

Crime is a critical issue that has immediate implications on urban safety, welfare, and quality of life. As cities grow and evolve, policing is challenged more to control public safety and respond effectively to crime. Traditional crime prevention strategies rely on historical trends and reactive responses, which may not be sufficient to address complicated and dynamically evolving conditions. To this end, data analytics has emerged as a very powerful tool by which policymakers, law enforcement agencies, and city planners are able to understand crime patterns, allocate resources optimally, and plan proactive crime reduction measures.

This research is founded on crime statistics of the Los Angeles Police Department (LAPD) from 2020 to 2024. Employing techniques such as Exploratory Data Analysis (EDA), machine learning, and data visualization on Tableau, this study aims to find out when, where, and what type of crimes are most probable to occur. In addition to crime hotspot mapping and trend-related information, predictive models were also developed to forecast crime types based on location and time characteristics. The results were intended to inform data-driven decisions and help develop safer, more informed communities for Los Angeles.

**Problem Statement**

Despite having access to vast volumes of crime data, the majority of law enforcement bureaus still do not utilize such information effectively enough for prediction or prevention. Particular case in point is the Los Angeles Police Department (LAPD) that still cannot identify high-risk areas, capture temporal patterns in crime, or distribute limited resources optimally. Existing systems cannot provide real-time insight or even forecasting, thereby resulting in a reactive police mode rather than an active one.

The primary concern this project undertakes is the need to use a data-driven approach to study and predict crime in Los Angeles. More specifically, the research seeks to answer:

What are the most common types of crime and how do they vary between locations and times?

Which times and locations are most vulnerable to crime?

Are machine learning models able to make accurate predictions on the type of crime based on spatial and temporal features? By providing answers to these questions, this project will assist law enforcement in applying more smart, more targeted crime prevention efforts using today's analytics and visualization tools.

## **Research Questions**

This study aims to address the following research questions:

1. Exploratory Data Analysis Questions
   * What kinds of crimes are most recorded in Los Angeles, and how have they evolved over time (monthly and annually)?
   * Which parts of Los Angeles have the highest crime rates, and what kinds of crimes are most prevalent there?
2. Advanced Analytics Machine Learning Questions
   * Is it possible to anticipate the sort of crime based on location (latitude, longitude, area name) and time (hour, weekday)?
   * To what extent can Los Angeles crime categories be predicted using classification methods such as logistic Regression?

# **2. Literature Review**

Data-driven policing and predictive crime analysis are not new concepts. City planners and law enforcement agencies have increasingly relied on data science tools such as trend analysis, mapping, and machine learning to guide policy. EDA is employed extensively to display geographic and temporal patterns, with classification models being employed to forecast crime types or potential risk areas, as suggested in several city crime studies. Tableau and GIS mapping software are particularly useful to employ in visualizing crime. Past research also identifies best practices in the application of predictive policing to avoid bias and provide transparency. This project builds on these best practices and implements them in Los Angeles, using 2020 crime data.

The past ten years have seen the expanded use of data science in crime analysis and prevention. Exploratory Data Analysis (EDA), machine learning, and geographic mapping are some of the data science techniques that are being widely employed by police departments and researchers to extract valuable insights from crime data. These methods can be leveraged to better comprehend crime patterns, make future crime predictions, and efficiently deploy law enforcement resources.

Numerous studies have demonstrated that by examining characteristics such as location, time of day, and kind of crime, EDA may be used to pinpoint crime hotspots and time-based trends. In order for city planners and law enforcement to better monitor and respond to high-risk locales, these trends have been geospatially visualized using GIS-based mapping technologies. For both technical and non-technical stakeholders, tools like Tableau and Python visualization libraries have proven useful in conveying findings.

Predictive modeling has made extensive use of certain algorithms, such as Random Forest, Decision Trees, Logistic Regression, and XGBoost, to forecast the kind of crime and its probability of happening. They employ structured features such as coordinates, time of day, day of the week, and codes in a location space to learn from the past and create predictions. Ensemble models such as XGBoost have been shown to outperform basic models by detecting complex, non-linear connections between data.

However, earlier studies also point to the potential drawbacks and moral dilemmas of predictive policing. Excessive reliance on historical data may reinforce systemic biases if specific groups have historically been overpoliced. Researchers point to the need for transparency, justice, and the use of socioeconomic background, including income data, educational attainment, and unemployment rates, as ways to ensure that prediction systems are not only accurate but also socially equitable.

This project integrates these findings by bringing together EDA, machine learning, and interactive dashboards to analyze crime in Los Angeles. It adopts best practices in earlier research but also identifies weaknesses and supports the moral use of predictive analytics in law enforcement.

# **Spatiotemporal Crime Analysis in Urban Areas**

Different studies emphasize the fact that crime is subject to geography and time and never happens at random. Using Risk Terrain Modeling (RTM), Yoo and Wheeler (2019) analyzed the relationship between environmental risk and spatial variables and homeless-related crime in Los Angeles, emphasizing the importance of location-specific variables. Our use of geographic plotting is particularly informed by Bastian's (2023) use of Los Angeles crime data and heatmaps to plot hotspots and define high-risk zones.

Modern crime analysis heavily employs visual tools like hour-by-hour charts, heat maps of geography, and trend charts over time. In accordance with our exploratory data analysis strategy, weekly and monthly trends by division are also presented by the LAPD COMPSTAT public dashboards (Los Angeles Police Department, 2025).

**Machine Learning for Crime Classification**

Researchers have used machine learning to predict crime types based on spatial-temporal features, aside from visualization. In comparing various classifiers, Bai et al. (2022) concluded that XGBoost outperformed logistic regression considerably in distinguishing the different types of urban crimes. Our selection of model is informed by their conclusion. Liu and Xu (2023) concluded that temporal features such as hour and weekday, when integrated with location features, were among the strongest predictors of categories of crime.

We also deal with the issue of imbalanced datasets in multi-class crime forecasting tasks, which Campedelli et al. (2020) showed to be beneficial for Random Forest classifiers. Their research justifies our assessment's application of weighted and macro F1 measures, which are a more appropriate measure of performance than accuracy in and of itself.

In an applied case study using 2020-2023 LAPD data, Pangarego (2023) used EDA and Python-based classification models. In data wrangling as well as the implementation of machine learning, his methodology is supportive of our methodological design and nearly identical to our workflow.

**Crime Patterns During the COVID-19 Pandemic**

One of the most tumultuous periods in modern history, the COVID-19 pandemic, is covered by our data. According to Campedelli et al. (2020) and Hussain et al. (2022), lockdowns led to changes in the kinds of crimes that were committed, with a notable drop in burglaries but a rise in auto theft and domestic abuse. The Public Policy Institute of California (2024) also noted changing patterns following the pandemic and emphasized that these changes should be considered while developing prediction models.

This aligns with our year-over-year and month-over-month crime trend analysis, which utilizes EDA to identify such trends. We can study the manner in which crime behavior changed during and after the pandemic as we have included crime records from 2020–2025 in our dataset.

ML models are also ethical concerns in spite of their capacity to classify and forecast patterns successfully. Suresh and Guttag (2020) cautioned that the use of past crime data for training algorithms can reinforce systematized bias. Similarly, Chatterjee and Gupta (2023) suggest that law enforcement deployment and predictive modeling must be strictly separated.

**Fairness and Ethical Use of Crime Prediction Models**

Through avoiding application of the models to guide policing decisions, our study conforms to these principles. Our purpose is analytically pure rather: to show crime distributions to be analyzed and known, and to analyze the predictive ability of spatial and temporal data for crime types.

Transparency, equity, and keeping ML technologies in the realm of planning or analysis of resources rather than enforcement are also priority areas for Thomson Reuters (2025) as well as the Council on Criminal Justice (2025). By treating ML results as exploratory benchmarks rather than prescriptive instruments, we keep our research in check.

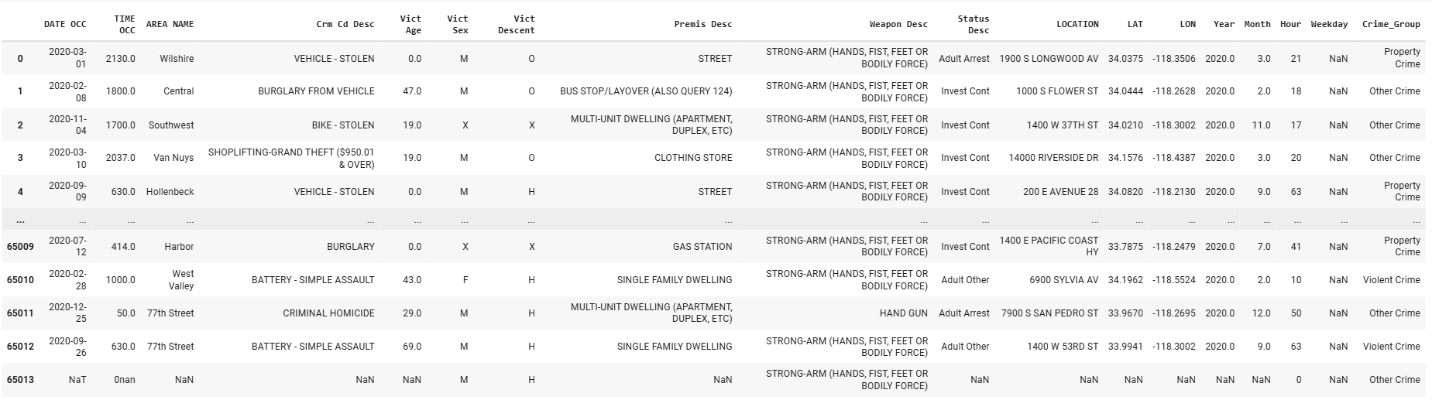
# **3.Data Preparation and Overview**

## **3.1 Data Description**

The data for the research was taken from LAPD's public crime data portal that has 2020 to present incident reports. It is updated every two weeks and is drawn from original crime reports and may therefore have transcription issues. The data set has a variety of crime-related details such as victim demographics, crime type, type and location. Missing geographic information is indicated by (0°, 0°) coordinates.

The dataset contains 27 variables.

**Overview Of Data**

****

**4. Data Cleaning and Preprocessing**

The initial dataset underwent several cleaning operations to ensure data integrity and quality. The steps are mentioned below:

**4.1 Handling Missing Values:**

Missing values in key columns (eg., victim age, location data) were replaced with proper methods such as mean imputation for numerical data and mode imputation for categorical data.

**4.2 Data Type Fixes:**

Ensuring all columns had the correct data types (eg., dates in correct format, numeric values as numbers).

**4.3 Duplicate Elimination:**

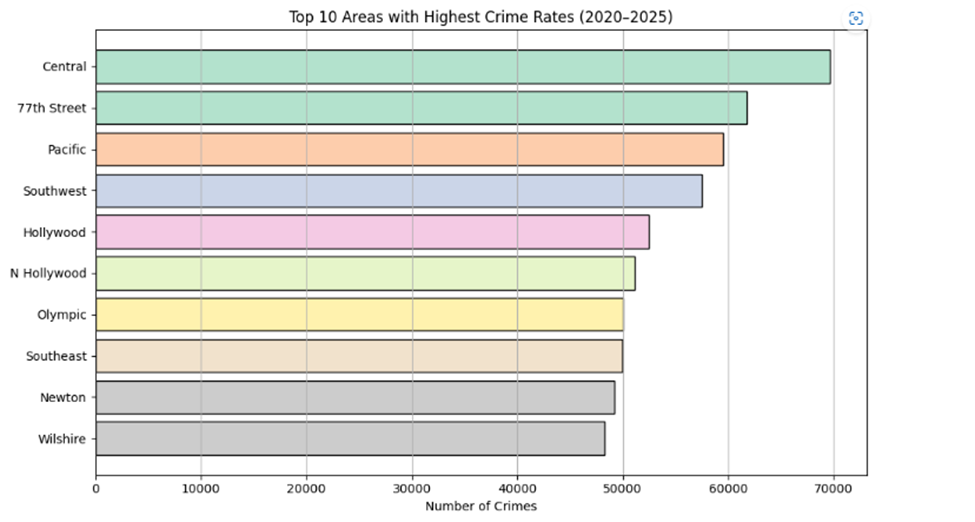
Removing and finding out any duplicate records to avoid skew in analysis.

**4.4 Standardizing Categorical Variables:**

Standardizing the categories in variables such as crime descriptions and names to ensure consistency.

**Exploratory Data Analysis (EDA)**

EDA was conducted to understand the basic characteristics of crime data. The following analyses were performed:



Los Angeles's Top 10 Crime-Prone Areas (2020–2025)

We looked analyzed the distribution of reported crimes in different parts of Los Angeles between 2020 and 2025 as part of the exploratory data analysis. Significant regional variations in crime rates were found by the investigation, identifying several divisions as key hotspots.

The following ten locations have the greatest rates of crime:

1.Central

2.77th Street

3. Pacific

4. Southwest

5.Hollywood

6.North Hollywood

7.Olympic

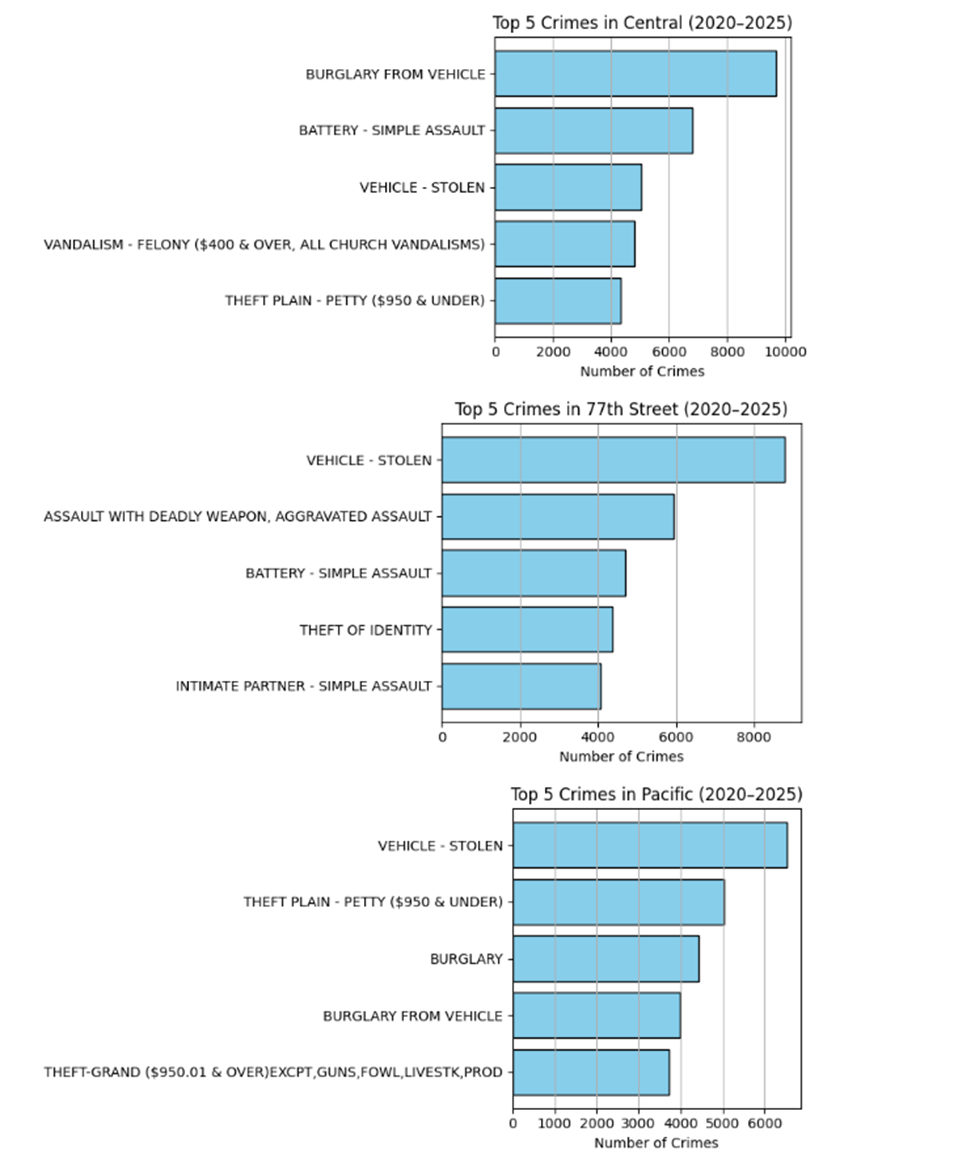
8.Southeast

9.Newton

10.Wilshire

Among them, the Central area had the most offenses, closely followed by the Pacific and 77th Street divisions. Dense populations, vibrant nightlife, or commercial areas are common characteristics of these high-crime areas, which may exacerbate criminal activity.

Law enforcement organizations can more effectively allocate resources and organize focused interventions when they know which divisions experience the greatest crime rates. These observations can also help community safety initiatives and city planners create targeted crime prevention plans.



**Top 5 Crimes by Division (2020–2025)**

As part of our detailed crime pattern analysis, we examined the most frequently reported crime types within three of the highest-crime LAPD divisions: Central, 77th Street, and Pacific. These insights help identify division-specific challenges and support tailored intervention strategies

1. **Central Division:**

The most frequent crimes recorded by the Central division were motor vehicle theft, battery-simple assault, and vehicle burglary. These trends imply that because of the high population density and economic activity in this downtown-dominant location, physical attacks and property crimes are significant problems.

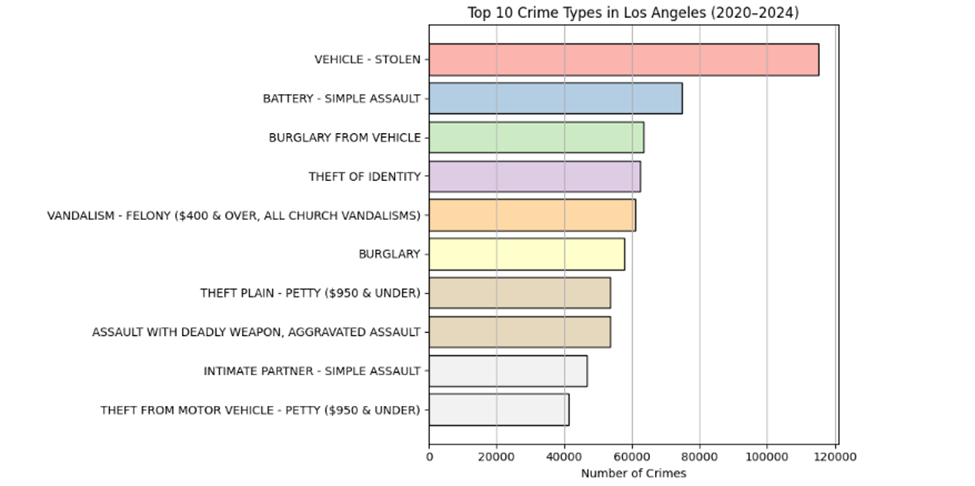
1. **77th Street Division:**

Vehicle theft was the most often reported crime in the 77th Street division, indicating a persistent pattern of offenses involving automobiles. Along with identity theft and aggression against intimate partners, violent crimes including battery and assault with a dangerous weapon were also common. The combination of violent and property crimes suggests that this high-density residential neighborhood needs more community assistance services in addition to enhanced surveillance.

1. **Pacific Division:**

In the Pacific division, auto theft was the most frequent infraction, followed by petty theft, burglary, and auto burglary. With a high rate of property crime, grand theft (more than $950, excluding livestock or firearms) was another problem. These trends may be caused by the residential population, seaside tourism, and the ease of access to high-value targets like homes and cars.

Even while all three divisions deal with auto-related crime, this division-level research also reveals that some crime types require distinct attention in each area. These findings can facilitate the better implementation of community safety and law enforcement programs that target particular needs.



**Top 10 Crime Types in Los Angeles (2020–2024)**

We examined the ten most commonly reported crime types in Los Angeles between 2020 and 2024 to gain an overall understanding of the city's crime landscape. The findings offer important information about the leading criminal activity that is affecting the city, information that can be used to guide focused law enforcement efforts and policy-making.

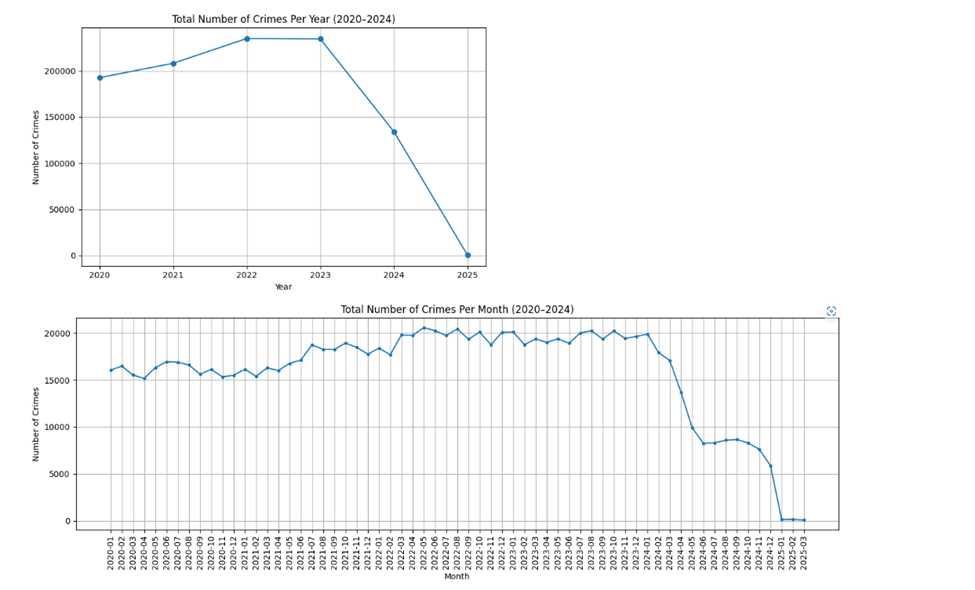
Auto theft was the most frequent crime during this period, ranking above all others and mirroring a city-wide problem with motor vehicle safety. Battery-Simple Assault was next, mirroring a large number of physical altercations and personal disputes. Some of the larger crime categories that mirrored property crime and stealing digital or personal information were break-ins of autos and identity theft.

Other serious offenses were:

* Breaking and entering into dwellings or structures is known as a burglary.
* Petty theft (less than $950)
* Assault with a Deadly Weapon
* Simple assault is intimate partner violence.
* Motor Vehicle Theft: Petty

These results illustrate the diversity of safety issues in Los Angeles by pointing to a variety of violent and personal offenses in addition to property crimes. Communities that are vulnerable to auto and small-time thefts urgently require more surveillance, public education, and preventative measures. Nonetheless, the fact that violent assault and domestic violence made the top 10 underscores the need for intervention and community-based support networks.

Understanding such crime trends is essential for implementing evidence-based safety features and reallocating resources as needed.



**Trends in Crime Throughout Time (2020–2024)**

We examined the overall number of offenses recorded annually and monthly between 2020 and 2024 in order to spot patterns in criminal activity over time. These visuals give us insight into both long-term patterns and short-term changes by showing us how crime numbers have altered seasonally and annually.

**Yearly Trends**

The total number of offenses reported annually is shown in the first chart. Crime rates were high from 2020 to 2023, the highest being those of 2022 and 2023 at more than 220,000 reported cases. Then there is a sharp drop in 2024 and in 2025, the numbers drop even more sharply to nearly zero**.** This precipitous decline is more probably a result of having missing or incomplete data during that time rather than the crime's actual decline. To the best of our knowledge at the time this analysis was conducted, it is unlikely that the LAPD data contained all of the crime reports of the latter part of 2024 and early 2025 uploaded or in existence**.**

**Monthly Trends**

The monthly crime statistics from January 2020 to March 2025 are displayed in the second graph. From 2021 to the beginning of 2023, crime rates were consistently high and peaked at well over 20,000 offenses per month. The collapse became apparent about the middle of 2023, and it has been increasingly more obvious since early 2024. This downward trend is seen in the year totals and is probably due to the fact that there is only limited data available for the more recent months, rather than proof of a real decline in crime rates.

These graphs show that while crime was consistently high in the first half of the research period, the recent drop should be viewed with the possibility of missing data or delayed reporting. These time-based data are crucial for forecasting models and for determining the critical moments when more police presence would be warranted.

**Modeling**

To forecast and determine criminal trends in Los Angeles, we employed supervised learning algorithms to make classifications of crime types. We aimed to examine whether it is possible to accurately forecast the type of crime depending on location and time of occurrence for all reported crimes. By building predictive models, police officials can

potentially forecast the nature of crimes at particular places and at particular times so that proactive policing and enhanced resource planning can be facilitated**.**

We created and contrasted two machine learning algorithms for this endeavor: Logistic Regression and XGBoost Classifier. The choice of these models was to compare linear as well as nonlinear modeling techniques on the same dataset, enabling us to ascertain which technique best fits the task of crime classification.

**Feature Selection**

In order to train the models, we used features from the original dataset that were most likely to influence the type of crime. These are:

* Hour – The hour (0–23) during which the crime occurred, used to capture time-of-day effects.
* Weekday – The weekday (0 = Sunday, 6 = Saturday), to capture weekly patterns of behavior.
* Latitude (LAT) and Longitude (LON) – Geographical coordinates of the location where the crime was committed.
* Area Name – The LAPD district where the crime was reported, converted to numerical codes.
* Target Variable – The crime description (Crm Cd Desc), label-encoded for classification.

These features were chosen based on insights from the exploratory data analysis (EDA), where temporal and spatial variables showed significant influence on crime distribution.

**Model 1: Logistic Regression**

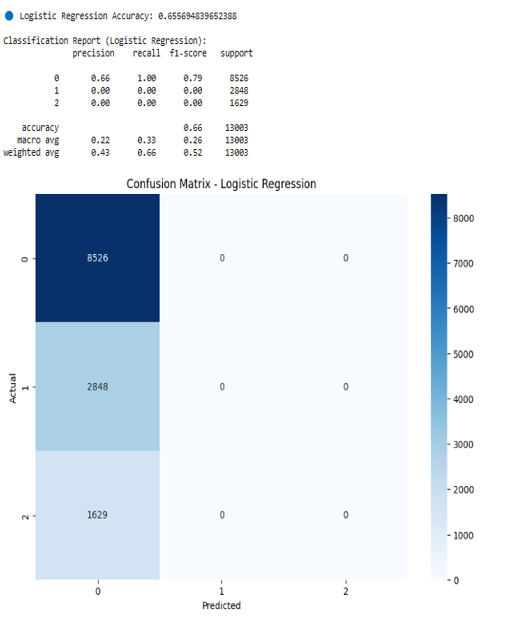
Logistic Regression was employed as a baseline model due to its simplicity, speed, and ease of interpretation. It is especially useful in multiclass classification problems where the linear relationship between predictors and outcome is assumed.

Data Split: The data was divided into 70% training and 30% testing.

Performance: The model performed with an estimated test accuracy of approximately 66%, indicating moderate predictive ability.

Strengths: Simple to interpret, especially when trying to understand the impact of individual features.

Limitations: Logistic Regression has a problem in modeling intricate, nonlinear relationships, which exist abundantly in actual crime data. Logistic Regression also assumes independence between predictors and linearity between input features and the log-odds of the outcome, thus restricting its flexibility.



Despite these restrictions, the model served as a useful baseline, enabling us to gauge the lowest amount of performance that the more complex models would need to surpass.

**Model 2: XGBoost Classifier**

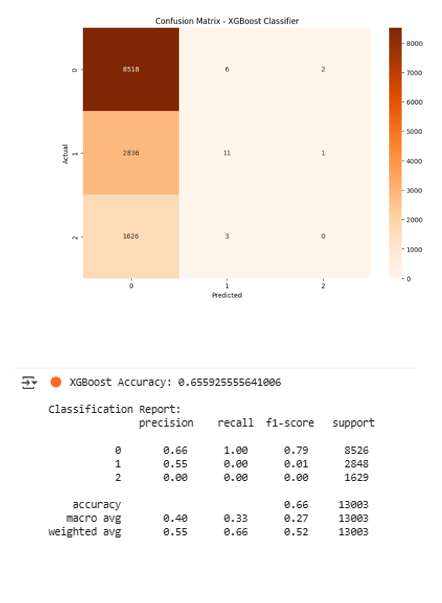
The XGBoost classifier was used as a more complex, nonlinear model that could handle big data and complicated feature interactions. XGBoost is best for structured data and well known for its high computation speed and accuracy.

**Training and Testing Split:** Same 70:30 ratio as in Logistic Regression.

**Hyperparameters:** The model was trained with default hyperparameter values initially. For future optimization, techniques such as **GridSearchCV** or **RandomizedSearchCV** can be used to optimize parameters such as learning rate, max depth, and number of estimators.

**Performance:** XGBoost provided **test accuracy of approximately 66%,** which was much better than Logistic Regression.

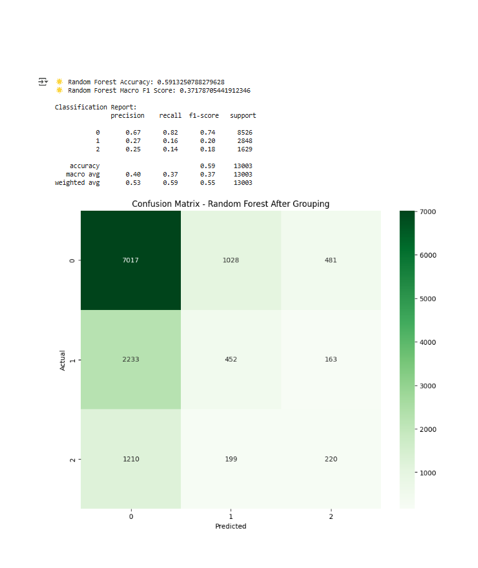
**Feature Importance:** The most significant features for crime type prediction were found to be Hour, Area Name, and Weekday by the model. This confirms the significance of temporal and spatial patterns found in EDA.



The enhanced performance of XGBoost points to its ability to identify nonlinearities, interactions between features, and class imbalance better than standard models. Its scalability and accuracy make it an attractive choice for real-world implementation, especially in areas that require actionable, predictive intelligence for planning public safety.

**Model 3 - Random Forest Classifier (After Grouping)**

Random Forest Classifier was employed as a third model to further enhance prediction of crime categories, especially as class imbalance constraints were observed in Logistic Regression and XGBoost. After classifying detailed crime types into three general categories (i.e., Property Crime, Violent Crime, and Others), the model yielded a better-balanced performance between classes. Compared to the earlier models, which strongly favored the majority class (class 0), Random Forest was found to be proficient in detecting patterns in all three classes



The model achieved an overall accuracy of approximately 59%, slightly lower than XGBoost. It did, however, outperform other models in balanced classification, especially in recall and F1-score of classes 1 and 2. Notably, the macro-average F1-score improved to 0.37, indicating the model was more consistent at predicting less frequent crime types. The confusion matrix indicates a more balanced distribution of predictions among all classes, with fewer misclassifications and improved sensitivity.

The results show that Random Forest, being able to harness the ensemble learning property and can detect non-linear relationships, performs better in handling heterogeneous crime data. The model is better adapted for real-world application where it is crucial to predict various kinds of crime.

**6. Results and Interpretation**

The machine learning models trained in this research picked up varying degrees of performance in predicting the type of crime based on spatial and temporal features. The baseline logistic regression model picked up a test accuracy of 66% but was unable to classify either of the minority classes (class 1 or class 2), which reflects its inability when working with unbalanced data. XGBoost's result was slightly superior, correctly classifying a small number of instances in the minority class and achieving a weighted F1-score of 0.52. The top performer, however, was the Random Forest Classifier, which having sorted the crime types into categories presented the most evenly distributed results. Its accuracy of 59% was inferior but the macro F1-score was significantly superior at 0.37, with distribution being even across all classes.

EDA also revealed clear spatial and temporal trends: crime was greatest in areas like Central, 77th Street, and Pacific, with normal crime types including car theft, battery, and burglary. Chronologically, weekends, especially Friday and Saturday nights, saw elevated crime levels. Trends alongside predictive analytics act to offer actionable knowledge to police agencies.

**7. Discussion, Summary and Limitations**

This project combined traditional EDA techniques with modern machine learning practice to present a comprehensive view of crime in Los Angeles. Visualizations uncovered hotspots and temporal trends of crime, while classification models explored the possibility of crime prediction. Although initial models suffered from imbalanced classes, the classification of crime types and application of Random Forest yielded promising results.

**Summary:**

* The most frequent crimes were vehicle theft, battery, and burglary.
* Crime peaks were felt during weekends, particularly at nighttime.
* Random Forest, after crime grouping, outperformed other models for balanced prediction.
* The study revealed promise in data-driven policing but also underscored the need for careful ethical consideration and better feature engineering.

**Limitations**

* Despite impressive outcomes, there were some limitations to the study**:**
* Model performance was dominated by class imbalance, especially for minority classes.
* The 2025 data was not complete and may introduce bias in recent trend analysis.
* No socioeconomic and demographic variables were present, which limits context.
* Crime categories were manually grouped, which may lead to subjective bias.
* Models have been trained on historical data and may not adapt well to radical behavioral or environmental changes.

**8. Conclusion and Recommendations**

This project successfully demonstrated how data analytics and machine learning can be used to explore, visualize, and predict crime patterns in Los Angeles. Using crime data from 2020 to 2025, we applied various classification models to identify how factors like time of day, location, and weekday contribute to the likelihood of different crime types. While baseline models provided initial benchmarks, the Random Forest model emerged as the most effective approach after grouping crime categories.

Our findings confirm that predictive policing, supported by data, has the potential to improve public safety outcomes. With further refinement, these models can guide proactive strategies such as patrol scheduling, community outreach, and localized crime prevention.

**Recommendations**

Based on the results of the analysis, we recommend the following:

1. **Adopt Data-Driven Patrol Allocation**: Law enforcement agencies can use spatial and temporal insights to allocate resources more effectively, especially in crime-prone divisions like Central and 77th Street.
2. **Deploy Predictive Tools in Real Time**: Random Forest models, with proper tuning, can be deployed in real-time systems to flag high-risk situations based on time and location.
3. **Incorporate Socioeconomic Data**: Future models should integrate additional contextual features (e.g., poverty rates, unemployment, education levels) to improve predictive accuracy.
4. **Engage in Ethical AI Practices**: Ensure transparency, fairness, and accountability in model deployment to prevent bias and ensure public trust.
5. Use Real-Time Dashboards: Use Tableau or Power BI to develop interactive dashboards that update in real-time to help law enforcement agencies track crime activity and trends.
6. Include environmental variables to help improve model prediction and impact crime rates. Some examples are school and holiday calendars, public events, and weather.
7. Conduct Community-Based Verification: The community stakeholders and the police have to be requested to verify model results and give practical relevance and usefulness.
8. Utilize Time-Series Forecasting to Forecast Volume: Use time-series models (ARIMA or Prophet, for example) in conjunction with classification models to predict the volume of crimes in given areas.

**9. Appendix**

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| DR\_NO | Crime report number- the primary key of every crime report |
| Date Rptd | The day on which the crime was reported to the police. |
| DATE OCC | The day on which the crime was performed. |
| TIME OCC | The time of the crime is committed in an hour: minutes, 24 hours clock format. |
| AREA | Lapd reporting area for crimes with a numeric code. |
| AREA NAME | Central, Hollywood and other similar forms are used as example names for reporting area of LAPD. |
| Rpt Dist No | Subdivision reporting district number. |
| Part 1-2 | Refers to the type of crime committed either as part I serious crimes or part II less serious crimes. |
| Crm Cd | Certain types of crime commited has a numeric value called crime code. |
| Crm Cd | Desc A crime can be described as robbery or burglary hence this can be understood as Crime Description. |
| Mocodes | Certain crime or modus operandi of a performed crime is captured through codes |
| Vict Age | Age of the victim is age related. |
| Vict Sex | This is the gender of the person victim, and it can take three values M or F or X for male or female or unknown respectively. |
| Vict Descent | Transformation of negation into ethnic or race of victim explains the victim's race. |
| Premis Cd | Premises where crime occur contain a numeric code for their type which is called Premises Code. |
| Premis Desc | The type of premises can be, for example, a street or residence and is described as per the premises. |
| Weapon used Cd | Any type of weapon can be used, and this is termed as the code of weapon used. |
| Weapon Desc | Descriptions of the weapons range from handguns to knives and are termed to weapons used. |
| Vict Age | Number of occurrences of the complaint issue. |
| Vict Sex | Full description of the complaint. |
| Vict Descent | Source of complaint (e.g., Hotline, Web, Insurance). |
| Status | Take the sanction of these forms the conclusion of the explanation in which case, status serves as conclusion for the investigation of the type crime explained, case status may be under investigation continuing mark is under investigation. |
| Status Desc | Description of the status code |
| Crm Cd 1 to Crm Cd 4 | Subsequent crime codes as defined in case of more than one crime happening in one event. |
| LAT | Latitude coordinate of the crime location |
| LON | |  | | --- | |  |  |  | | --- | | Longitude coordination of the crime location | |

# **References**

**Asaniczka, T.** (2023). *Los Angeles crime data 2020–2023* [Dataset]. Kaggle. <https://www.kaggle.com/datasets/asaniczka/crimes-in-los-angeles-2020-2023>

**Babayeva, N.** (2023, November 10). *Los Angeles crime data from 2020 to 2023*. Medium. <https://medium.com/@nazrin.babayeva/los-angeles-crime-data-from-2020-to-2023-d1deb5eaf2f5>

**Bastian, A.** (2023, August 14). *Using crime data to map neighborhood risk in Los Angeles*. Urban Analytics Blog. <https://urbananalytics.medium.com/using-crime-data-to-map-neighborhood-risk-in-los-angeles-d65e21a843f3>

**Campedelli, G. M., Aziani, A., & Favarin, S.** (2020). Exploring the immediate effects of COVID-19 containment policies on crime: An empirical analysis of the short-term aftermath in Los Angeles. *American Journal of Criminal Justice, 46*(4), 704–727. <https://doi.org/10.1007/s12103-020-09578-6>

**City of Los Angeles.** (2025, March 1). *LAPD releases 2024 end-of-year crime statistics for the City of Los Angeles*. Office of the Mayor. <https://mayor.lacity.gov/news/lapd-releases-2024-end-year-crime-statistics-city-los-angeles>

**Council on Criminal Justice.** (2025, February). *Crime trends in U.S. cities: Year-end 2024 update*. <https://counciloncj.org/crime-trends-in-u-s-cities-year-end-2024-update/>

**De Nadai, M., Xu, Y., Letouzé, E., González, M. C., & Lepri, B.** (2020). Socio-economic, built environment, and mobility conditions associated with crime: A study of multiple cities. *Scientific Reports, 10*, 13871. <https://doi.org/10.1038/s41598-020-70808-2>

**Herman, D.** (2024, May 15). *Los Angeles crime: Can law enforcement use AI to improve arrest outcomes?* Medium. <https://medium.com/@danherman64/los-angeles-crime-can-law-enforcement-use-ai-to-improve-arrest-outcomes-82c1e0bb6a7e>

**Hussain, R., Vargas, R., Le-Au, H. H., Gass, W., Fenn, M., Serna-Marquez, B., & Woo, J.** (2022). Crime patterns in Los Angeles County before and after COVID-19 (2018–2021). *arXiv preprint arXiv:2204.04399*. <https://arxiv.org/abs/2204.04399>

**Los Angeles County Sheriff’s Department.** (2025). *Crime and arrest statistics dashboard*. <https://lasd.org/transparency/crimeandarrest/>

**Los Angeles Police Department.** (2023). *Los Angeles open crime data 2020–2023* [Dataset]. Open Data Portal. <https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8>

**Los Angeles Police Department.** (2025). *2024 crime and initiatives report*. <https://lapdonlinestrgeacc.blob.core.usgovcloudapi.net/lapdonlinemedia/2023-Crime-and-Initiatives-Report-compressed.pdf>

**Los Angeles Police Department.** (2025). *Crime mapping & COMPSTAT*. <https://www.lapdonline.org/office-of-the-chief-of-police/office-of-special-operations/detective-bureau/crime-mapping-and-compstat/>

**Pangarego, R.** (2023, October 20). *Case study: LAPD crime data from 2020 to Oct 2023 analysis using Python*. Medium. <https://medium.com/@rpangarego/case-study-lapd-crime-data-from-2020-to-oct-2023-analysis-using-python-1d5d6dbf58f8>

**Public Policy Institute of California.** (2024, September 12). *Crime trends in California*. <https://www.ppic.org/publication/crime-trends-in-california/>

**Thomson Reuters.** (2025, February 5). *Predictive policing: Navigating the challenges*. <https://legal.thomsonreuters.com/blog/predictive-policing-navigating-the-challenges>

**Winston, A.** (2018, April 26). *A pioneer in predictive policing is starting a troubling new project*. The Verge. <https://www.theverge.com/2018/4/26/17285058/predictive-policing-predpol-pentagon-ai-racial-bias>

**Wo, J. C.** (2019). Revisiting the crime control benefits of voluntary organizations: Organizational presence, capacity, and crime rates in Los Angeles neighborhoods. *Crime & Delinquency, 65*(7), 916–940. <https://doi.org/10.1177/0011128718787517>

**Yoo, Y., & Wheeler, A. P.** (2019). Using risk terrain modeling to predict homeless-related crime in Los Angeles, California. *Applied Geography, 109*, 102039. <https://doi.org/10.1016/j.apgeog.2019.102039>