

Los Angeles Police Department



## FOREST FLUX

Project Plan

# CRIME ANALYSIS

by

Victoire Migashane, Evan Ghulbet, and Sarah Mohammed  
DAT 490 - Spring A, 2026  
Laurence Schneider  
Jan 26, 2026

## Organization Description

The Los Angeles Police Department (LAPD) is the municipal law enforcement agency for the City of Los Angeles, California. As the third-largest police department in the United States, the LAPD serves a population of approximately 4 million residents across 502 square miles. The department operates 21 community police stations (called "Areas") throughout the city, employing over 9,000 sworn officers and approximately 3,000 civilian personnel. The LAPD is organized into four main bureaus: Office of Operations (patrol and specialized units), Office of Special Operations (metro, K-9, SWAT), Detective Bureau (investigative services), and Office of Administrative Services (support functions). The department is governed by the Los Angeles Board of Police Commissioners and reports to the Mayor of Los Angeles.

The LAPD's mission is to safeguard the lives and property of the people it serves, reduce the incidence and fear of crime, and enhance public safety while working with diverse communities to improve their quality of life. The department is guided by six core values: Service to Our Communities, Reverence for the Law, Commitment to Leadership, Integrity in All We Say and Do, Respect for People, and Quality Through Continuous Improvement. These values emphasize dedication to public safety, upholding constitutional principles, leading by example, maintaining the highest ethical standards, treating all people with dignity and respect, and embracing innovation and continual change.

Current leadership includes Chief of Police Jim McDonnell, who was sworn in as the 59th Chief on November 8, 2024. Chief McDonnell is a 29-year LAPD veteran and the first person to serve in senior executive leadership positions in the three largest policing agencies in Los Angeles County (LAPD, LASD, LBPD). He succeeded Dominic H. Choi, who served as Interim Chief from March to November 2024 and was the first Asian American to lead the LAPD. Previous Chief Michel Moore led the department from 2018 to 2024.

The LAPD's Strategic Plan 2021-2023 outlines six primary goals that guide departmental operations and improvements. These include protecting Los Angeles through crime reduction and community partnerships, serving Los Angeles by building trust through engagement initiatives, improving organizational accountability by reducing uses of force and promoting fairness, modernizing technology to improve field efficiencies, enriching training quality with cultural perspectives, and maximizing workforce potential through diverse recruitment and employee retention. The plan emphasizes constitutional policing, transparency, and continuous improvement in service delivery across all neighborhoods.

Headquarters: Los Angeles California (100 West 1st Street, Los Angeles, CA 90012)

Contacts: 911 for emergencies and (877) 275-5273 for non-emergencies

Website: <https://www.lapdonline.org/>

Learn more: <https://www.lapdonline.org/lapd-organization/>

## History of Organization

The Los Angeles Police Department was formed in 1869 when the city hired six officers to serve under City Marshal William C. Warren. Before this, Los Angeles relied on volunteer forces like the Los Angeles Rangers, founded in 1853, to maintain order in what was then a violent frontier town known for gambling, vice, and reportedly the highest murder rate in the United States during the Gold Rush era. The early LAPD grew slowly, and in 1889 Chief John Glass organized the city into four police districts for better supervision. The department made history by hiring Alice Stebbins Wells in 1910, the first female police officer with full arrest powers in the United States, and Georgia Ann Robinson in 1916, the first African American female officer. However, corruption plagued the department through the early 1900s until Mayor Fletcher Bowron forced major reforms in 1938, removing dozens of corrupt officers and commissioners.

From 1950 to 1966, Chief William H. Parker transformed the LAPD into one of the most professional police departments in the country, introducing modern management systems and the famous motto "To Protect and To Serve" in 1955. Despite these improvements, the department faced serious challenges including the 1965 Watts Riots and the 1991 Rodney King incident, which exposed problems with police brutality and community relations. These events led to major reforms through the Christopher Commission, which increased diversity in hiring and improved accountability. In 2001, the U.S. Department of Justice entered into a consent decree with the LAPD to address civil rights violations, and after significant reforms, the decree was lifted in 2013. Today, the LAPD continues to focus on constitutional policing, community partnerships, and transparency.

Learn more: <https://www.lapdonline.org/history-of-the-lapd/>

## Community Background

Los Angeles is one of the most diverse cities in the United States, with residents representing over 140 countries and speaking 224 different languages across approximately 214 distinct neighborhoods. The city includes wealthy areas like Bel Air and Pacific Palisades as well as working-class communities in South LA and the San Fernando Valley. This diversity creates unique policing challenges that require cultural understanding and tailored approaches for different communities. The LAPD operates 21 community police stations throughout the city, each serving its local area with over 9,000 sworn officers and 3,000 civilian personnel.

To address community needs and rebuild trust after historical tensions, the LAPD has adopted a Community Policing approach where police and residents work together as partners. Each of the 21 Areas has a Community-Police Advisory Board (CPAB) that gives community members a voice in local policing decisions. The department uses a problem-solving model called SARA (Scanning, Analysis, Response, and Assessment) to identify and address crime issues with input from residents. The LAPD has also increased transparency by publicly releasing crime data.

Despite these efforts, Los Angeles continues to face challenges including homelessness, property crime, gang activity, and ensuring fair policing across all neighborhoods regardless of income level or racial background.

Learn more: <https://www.lapdonline.org/community-policing-unit/>

## Competitors and Finances

The Los Angeles Police Department operates with a budget of \$2.14 billion for fiscal year 2025-26, representing an 8.1% increase from the previous year, with combined federal, state, and other funding bringing the overall budget to approximately \$3.3 billion. The department's budget priorities include salary adjustments and recruitment efforts, vehicle and helicopter replacements, and technology upgrades including the Real-Time Crime Center and cybersecurity operations. The city has set a goal to increase the number of officers to 9,084, with long-term plans to hire 9,500 officers by 2028 or earlier. As of 2024-2025, the LAPD employs over 9,000 sworn officers and approximately 3,000 civilian personnel across 21 community police stations serving 4 million residents across 502 square miles, making it the third-largest municipal police department in the United States.

The LAPD's major peer agencies include the New York City Police Department (NYPD) and the Chicago Police Department (CPD). The NYPD is the largest municipal police department in the country with a fiscal 2025 budget of \$5.8 billion, supporting 48,844 full-time positions including 35,001 uniformed officers and plans to grow to 40,000 officers by 2029. The Chicago Police Department operates with a fiscal year 2024 budget of nearly \$2.0 billion but faces significant staffing challenges, operating at only 87% capacity with 12,329 active employees as of January 2024, down 1,400 officers since 2019. All three departments face similar pressures including recruitment difficulties, overtime costs, technology modernization needs, and balancing effective crime reduction with accountability reforms. The LAPD distinguishes itself through its community policing model with 21 Area-based Community-Police Advisory Boards and strategic emphasis on diversity, equity, and inclusion initiatives.

Learn more:

<https://nenc-la.org/2024/11/lapd-budget-approved-2-14-billion-spending-plan-for-2025-26/>

## Project Description

### Research Questions

#### **Research Question 1: Predicting Crime Reporting Delays in Los Angeles**

Our research will investigate the factors that cause the gap between when a crime occurs and when it is reported to the Los Angeles Police Department. Specifically, we want to understand

which factors influence crime type, victim demographics, location, and timing predict whether someone reports a crime right away or waits days or even weeks to file a report, and whether these patterns are different for violent crime versus property crimes. We think this is an important question because when someone experiences a crime, their decision to report isn't random. It could be driven by fear, how serious they think the crime is, whether they trust the police, or just practical stuff like not knowing how to report or being too busy. Using regression analysis on LAPD crime data from 2020 to 2025 (about 1,000,000 rows and 28 columns), we can identify which factors most strongly predict reporting delays. Understanding these patterns helps the LAPD identify which communities or types of crime face barriers to timely reporting, which has real implications for how police build trust with different neighborhoods, allocate resources, and ensure everyone has equal access to police services regardless of where they live or who they are.

### **Research Question 2: Predicting Crime Hotspot Intensity Across Los Angeles**

Our research will investigate why some Los Angeles neighborhoods feel unsafe while others don't. We want to know what factors predict where and when crime concentrates most intensely across the city. We want to understand how time patterns (rush hour vs. late night, weekdays vs. weekends, summer vs. winter), the mix of different crime types, the places where crimes happen (streets, parking lots, homes), and past crime history predict whether an area becomes a crime hotspot, or if it's safer than surrounding neighborhoods. We think this is crucial because every Angeleno deserves to feel safe in their own neighborhood, yet some communities live with constant anxiety about crime, while others rarely think about it. This inequality isn't fair, and it's not inevitable. Using regression analysis on LAPD crime data from 2020 to 2025, we will group the 1,000,000 rows and 28 columns by area and time. Doing this will allow us to pinpoint exactly what makes certain places and times so vulnerable to crime. Understanding these patterns means the LAPD can actually protect the communities that need it most, putting officers where they'll make the biggest difference, working with neighborhoods to fix dangerous conditions, and giving families in high-crime areas the same peace of mind that people in safer neighborhoods take for granted.

### **Research Question 3: Predicting Victim Age Patterns in Los Angeles Crime**

Our research will examine who is harmed by crime in Los Angeles. We will investigate factors that predict whether victims are children, young adults, middle-aged people, or seniors, and whether certain crimes deliberately target the most vulnerable among us. We want to understand how the type of crime, where it happens, what time it occurs, and whether weapons are involved, predict the age of victims, and whether some age groups face dangers that others don't even think about. We think this matters deeply because a grandmother getting robbed, a teenager getting assaulted, and a young parent getting their car stolen are all tragedies, but they're different kinds of tragedies that need different solutions. Elderly people might live in fear of scams or being attacked because they can't fight back. Young people might not realize they're walking into

danger in certain neighborhoods or at certain times. Kids depend on adults to keep them safe, but sometimes fall through the cracks. Using regression analysis on LAPD crime data from 2020 to 2025 (about 1,000,000 rows and 28 columns, including victim ages), we can identify which groups are most at risk of which crimes and why. Understanding these patterns means the LAPD can actually protect people based on their real vulnerabilities, warning seniors about the specific threats they face, teaching young adults how to stay safe, and making sure kids have safe routes to school.

## Initial Project Hypotheses

### **Hypothesis 1: Crime type predicts reporting delay**

Violent crimes (crimes such as assault with deadly weapons) will have significantly shorter reporting delays compared to non-violent crimes (such as theft), with violent crimes being reported within 24-48 hours of occurrence while theft and fraud related crimes will show reporting delays exceeding 7 days.

### **Hypothesis 2: Temporal patterns and premise type predict crime hotspot intensity**

Crime concentration in specific LAPD areas will be significantly predicted by nighttime hours (10 PM - 4 AM) and commercial premise types (stores, parking lots, gas stations), with areas experiencing the highest crime intensity during late-night weekend hours in commercial districts showing 3-5 times more incidents compared to residential areas during daytime weekday hours. Additionally, areas with historically high crime rates will continue to demonstrate elevated crime intensity regardless of temporal factors, suggesting persistent geographic vulnerability.

### **Hypothesis 3: Crime type and time of occurrence predict victim age patterns**

Violent crimes involving weapons will disproportionately affect younger victims (ages 18-35), with over 60% of weapon-related assaults targeting this age group, while property crimes like theft and fraud will show a bimodal distribution targeting both young adults (18-35) and seniors (60+). And, crimes occurring during late-night hours (10 PM - 4 AM) will predominantly victimize younger individuals, while crimes during daytime business hours (9 AM - 5 PM) will more frequently target middle-aged and senior victims, reflecting different daily activity patterns and vulnerabilities across age groups.

## About the Data

The data for this project comes from the Los Angeles Police Department (LAPD) and is publicly available through the U.S. Government Open Data Catalog at data.gov. The dataset is provided as a single CSV file containing both structured and unstructured variables that support crime records for the greater Los Angeles metropolitan area. The dataset currently includes 1,004,991 reported crime records spanning from January 1, 2020 to June 4, 2025 and comprises 28

variables. The data is organized into five groups: Incident Identification, Temporal Information, Geographic Information, Victim Demographics, and Crime Characteristics. These groups collectively support the investigation of crime reporting delays, which is the primary focus of this project.

### *Incident Identification*

Each of the 1,004,991 crime records in the dataset is uniquely identified through a set of core identification and classification variables. The division record number (DR\_NO) serves as the unique identifier for each incident. The reporting district number (Rpt Dist No) and the LAPD area code and area name indicate which precinct handled the case. Each record contains up to four crime code fields (Crm Cd, Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4) along with their corresponding descriptions (Crm Cd Desc), which allows for both primary and secondary crime classifications when an incident involves more than one type of crime. The method of operation (mocodes) provides additional detail about how the crime was carried out. The current status of each case is tracked through the status code (status) and its corresponding description (Status Desc).

### *Temporal Information*

The temporal variables are central to this project, particularly for Research Question 1, which investigates crime reporting delays. The dataset contains both the date of occurrence (Date OCC) and the date reported (Date Rptd), each accompanied by pre-derived year, month, and day fields (Date OCC Year, Date OCC Month, Date OCC Day, Date Rptd Year, Date Rptd Month, Date Rptd Day). The time of occurrence (Time OCC) is also recorded, which provides hour-level granularity. The reporting delay can be calculated at both the day and hour levels by computing the difference between the date and time of occurrence and the date and time reported. This delay measurement will serve as the primary classification target for the predictive model.

### *Geographic Information*

The geographic variables provide spatial context for each crime record. The LAPD area code (Area) and area name (Area Name) identify the police district in which the crime occurred. More precise location data is captured through the full address (LOCATION), the nearest cross street (Cross Street), and the geographic coordinates represented by latitude (LAT) and longitude (LON). These variables support geographic-based analysis, particularly for Research Question 2, which examines crime hotspot intensity across Los Angeles neighborhoods. Geographic location also serves as a predictor variable for Research Question 1, as the district in which a crime occurs may influence the speed at which it is reported.

### *Victim Demographics*

The dataset includes three victim demographic variables: victim age (Vict Age), victim sex (Vict Sex), and victim descent (Vict Descent). These variables are relevant across all three research questions. For Research Question 1, victim demographics, such as age and descent, may

influence the likelihood or speed at which a crime is reported. For Research Question 3, victim age is the primary outcome variable used to classify and predict which age groups are most affected by different crime types. Together, these demographic variables provide important context for understanding the human factors behind crime reporting behavior.

### *Crime Characteristics*

The crime characteristics variables capture the nature and context of each incident. The premise code (Premis Cd) and premise description (Premis Desc) indicate where the crime took place, such as a residence, street, or commercial establishment. The weapon used code (Weapon Used Cd) and weapon description (Weapon Desc) detail whether a weapon was involved and, if so, what type. The crime code description (Crm Cd Desc) is an unstructured text field that requires processing to establish consistent crime-type categorizations, such as grouping crimes into violent versus non-violent categories. These classifications are particularly important for testing Hypothesis 1, which predicts that violent crimes will have significantly shorter reporting delays compared to non-violent crimes such as theft or fraud.

### *Measurements*

A critical component of any analytical project, especially one involving predictive analytics, is a clear understanding of what is being measured. The primary focus of this project centers on classification accuracy and reporting delay patterns. The primary outcome variable for Research Question 1 is the reporting delay itself, measured as the time difference between the date and time a crime occurred and the date it was reported to the LAPD. This delay will be calculated at both the day-level and hour-level to capture broad and fine-grained patterns. The delay will then be converted into a classification target, categorizing incidents as reported quickly (within 24 hours) versus reported with a significant delay (exceeding 7 days). This classification serves as the output that the predictive model is trained to predict.

The LAPD captures all relevant information at the moment a crime is logged into their system, including timestamps, location details, victim information, and crime characteristics. This creates a comprehensive dataset that directly supports the measurement objectives of this project. The key predictors for the reporting delay classification model include crime type (violent vs. non-violent), victim demographics (age, sex, and descent), geographic location (LAPD area), time of occurrence, and premise type. These variables each represent a factor that may influence how quickly a crime is reported, and together they provide the predictive model with the information needed to identify patterns in crime reporting behavior across Los Angeles.

### *Methodology*

Our project will use statistical and predictive modeling to analyze crime patterns in LA by using the public crime data specifically between 2020-2025. Our overall goal is to understand how crime reporting behavior, crime concentration, and victim age patterns vary across time, location, and crime characteristics.

Before completing an analysis, we will clean and prepare the data set by removing records with invalid/missing dates, handling missing demographic values and creating new variables as needed. The most important variable would be ‘crime reporting delay’ as it calculates the difference between the date and time a crime occurred and when it was actually reported to LAPD.

### **Research Question 1: Predicting Crime Reporting Delays in Los Angeles**

The research on the reporting delay will be transformed into a classification problem. We will have the crimes labeled as either reported within 24 hours or reported with a significant delay (7+ days). Logistic regression will be used to determine which factors such as crime type, victim demographics, location, and time to help us predict delayed reporting.

### **Research Question 2: Predicting Crime Hotspot Intensity Across Los Angeles**

The research on crime hotspot intensity will be grouped by areas within LAPD and intensity. Regression models for count data such as a negative binomial regression will be used to identify how time patterns, premise types, and historical crime levels predict where crime concentrates most heavily around the city.

### **Research Question 3: Predicting Victim Age Patterns in Los Angeles Crime**

The research on predicting the victim age patterns will be grouped into categories (children, young adults, middle age, and seniors). We will use multinomial logistic regression to examine the crime type, weapon involvement, time of occurrence, and location to help us predict which age group is most likely to be victimized.

#### *Computation method and output*

All data processing and analysis will be conducted using Python and R. Regression models will be developed using statistical modeling tools and conventional data analysis libraries will be used for data cleaning and feature engineering.

For classification models such as research questions 1 and 3, we will evaluate performance using accuracy, precision, recall, F1 score, and ROC-AUC. Additionally, confusion matrices will be used to understand classification errors, specifically when the cases of delayed reporting are incorrectly predicted as timely.

The hotspot analysis for research question 2 will assess the model fit and predictive strength using goodness of fit measures and comparisons between observed and predicted crime counts. Spatial visualizations, including heat maps and area level crime plots will be used to clearly communicate hotspot patterns.

Our outputs for this project will include cleaned and processed datasets, regression results and coefficient tables, model performance metrics, and visualizations of reporting delays, crime

hotspots, and victim age distributions. These outputs are designed to be easy to interpret while still providing us with strong analytical evidence.

#### *Output Summary*

#### **Research Question 1: Predicting Crime Reporting Delays in Los Angeles**

The output will show which crimes, locations, and victim characteristics are associated the most with delayed reporting. Our results will highlight differences between violent and non-violent crimes as well as identifying communities where barriers to timely reporting exist.

#### **Research Question 2: Predicting Crime Hotspot Intensity Across Los Angeles**

The output will reveal which areas within LA experience the highest crime intensity and when the hotspots are most active. The maps and time based summaries provided will show how crime concentration changes across neighborhoods which will assist with explaining why some areas feel consistently less safe compared to others.

#### **Research Question 3: Predicting Victim Age Patterns in Los Angeles Crime**

The output will identify which age groups are most vulnerable to specific crime types and conditions. Our results will show how age patterns differ by crime type, weapon involvement, and time of day which will provide insight into age specific risks across LA.

#### *Campaign Implementation*

It is known that Los Angeles is one of the largest and most diverse cities in the United States. It houses over four million residents across neighborhoods that experience a very different level of safety, resources and trust with law enforcement. Since there is a big difference in how each neighborhood experiences crime, the LAPD must be strategic with how they allocate resources, communicate with the residents, and implement prevention efforts.

#### **Research Question 1: Predicting Crime Reporting Delays in Los Angeles**

For our first research question, identifying the patterns in delayed crime reporting will allow the LAPD to focus outreach efforts on crimes and communities where reporting barriers are highest. If non-violent crimes such as theft or fraud are consistently reported late, then the department could implement targeted public awareness campaigns explaining how and why to report these crimes quickly, even if they seem minor. Initiatives such as online reporting tools, multilingual reporting guides, and community workshops in affected neighborhoods could possibly help reduce those barriers and improve data accuracy for LAPD.

#### **Research Question 2: Predicting Crime Hotspot Intensity Across Los Angeles**

For our second research question, understanding crime hotspot intensity allows for LAPD to deploy officers and resources more efficiently. The department could start focusing on staffing

during high risk hours and in consistently vulnerable areas identified within our research, rather than spreading officers evenly across the city. We believe this data driven approach could lower crime, while also reducing the unnecessary policing in low risk neighborhoods.

### **Research Question 3: Predicting Victim Age Patterns in Los Angeles Crime**

For our third research question, identifying age patterns allows the LAPD to tailor prevention strategies to different populations. Depending on the population affected, initiatives can be taken through social media posts, support centers, and programs to provide insight and awareness.

## **References**

1. Los Angeles Police Department. *LAPD organization*.  
<https://www.lapdonline.org/lapd-organization/>
2. Los Angeles Police Department. *History of the LAPD*.  
<https://www.lapdonline.org/history-of-the-lapd/>
3. Los Angeles Police Department. *Community policing unit*.  
<https://www.lapdonline.org/community-policing-unit/>
4. North East Neighborhood Council. (2024, November). *LAPD budget approved: \$2.14 billion spending plan for 2025–26*.  
<https://nenc-la.org/2024/11/lapd-budget-approved-2-14-billion-spending-plan-for-2025-26/>