

# Comprehensive Crime Data Analysis and Forecasting (2020–2025)

Authors: Aarya Gadekar, Pooja Patil, Andrew Roth

Scope: Los Angeles Crime Data (2020 – 2025)

Tools: Python (Pandas, NumPy, Matplotlib), ARIMA, Prophet, FRED Economic Data

## Abstract

The purpose of this project is to analyze crime trends from the Los Angeles area between 2020 and 2025. Analysis will include, but not be limited to, yearly and seasonal crime trends, and crime patterns based on region, economic factors, major events, and demographics. This project will also provide models for future crime trends.

Data for this project is derived from the given Crime\_Data\_from\_2020\_to\_Present data set from data.gov. Outside data was obtained from FRED Federal Reserve Bank of St. Louis to gain insight into real-time economic patterns.

## Data Cleaning

Data Cleaning was done in three steps: the removal of columns, filling null values, and clarifying columns names and descriptions. In total, 12 columns were dropped from the original data set. Many of these columns, such as ‘Crm Cd 4’ and ‘Cross Street’, had over 100,000 null values, and simply dropping them made sense. Other columns, such as ‘Crm Cd’ or ‘Mocodes’, were dropped simply because its contents were vague and unable to be analyzed due to incomplete descriptions provided by the original data set. Lastly, ‘AREA’ was dropped because it was vague; as well as redundant with ‘Area Desc’, which was much more insightful for analysis.

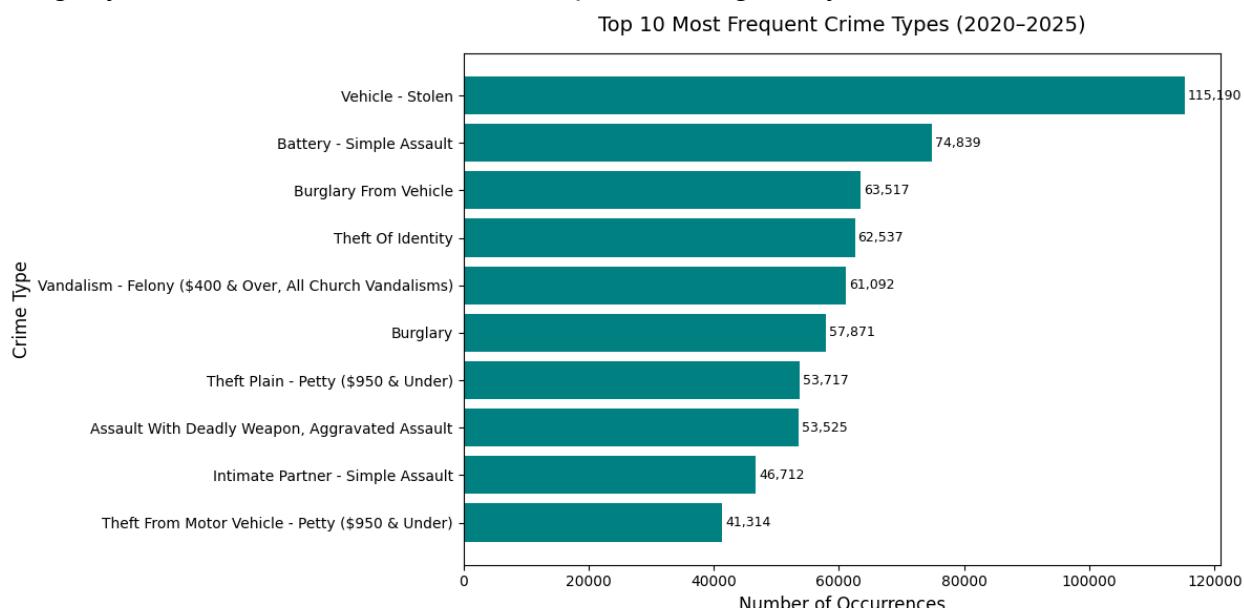
After dropping columns, there were very few that remained with missing values; these were filled with “Unknown” in accordance with its column specification. Before proceeding to analysis, some column entries were better labeled for easier analysis. For example, ‘Vict Descent’ and ‘Vict Sex’ entries were defined with their own dictionaries. Additionally, ‘Part 1-2’ was renamed to ‘Crime\_type’ and a new ‘Age\_category’ was created. The data

frame was then renamed as cleaned\_df to emphasize its completeness and readiness for analysis.

## Exploratory Data Analysis (EDA)

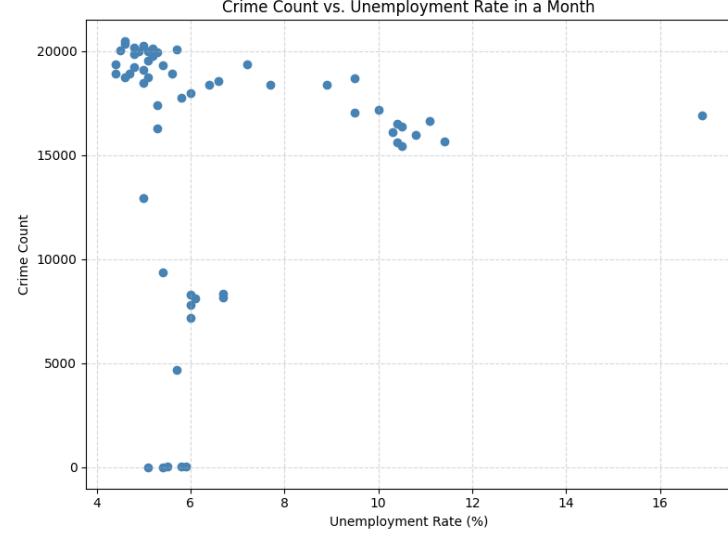
EDA was structured around the key questions, but included extended graphical analysis. First, we explored the overall crime trend. Crime activity seemed to peak in 2022, and has seemingly declined since then. It is important to note that 2025 has not yet been completed, so 2025 may display incomplete data; therefore, there are time we elect to remove 2025 data from graphs, or disregard its information completely if graphed. According to the data itself, crimes may be reported months after they occur, so this may also be a contributing factor the steep decline in crimes in 2025. Alternatively, as the data suggests, Los Angeles may have done better to limit the number of crimes. Crimes also could be reported at a less regular rate, and as mentioned in the data set description, LAPD is in the process of both transferring to a new system to report crimes and is facing issues posting the data. A combination of these factors is most likely responsible for the low number of crimes thus far in 2025.

Regardless, this should have little effect on the count of crimes per month in LA. January has seen the most crimes over the last five years, with an overall decreasing crime count until December, which has seen the least number of crimes. Crimes per month was also analyzed on a monthly basis since 2020, and saw very small increases in crime until 2024, when crime dropped steadily each month. Of all crimes committed, a Stolen Vehicle was most popular, representing over 10% of all reported crimes. Simple Assault and Burglary from Vehicle rounded out the top three. Regionally, Central, 77<sup>th</sup> Street, and



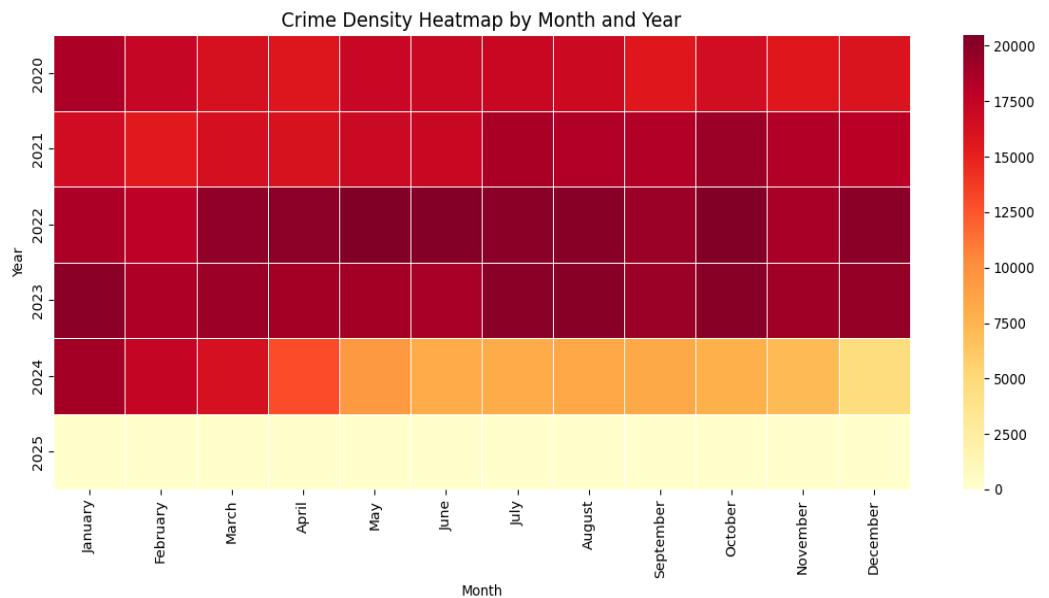
Pacific had three most crime since 2020, while Foothill, Hollenbeck, and Mission reported the least. The number of crimes was also highest on Friday and Saturday, while lowest on Tuesday. This is likely due to people traveling on weekends and being away from personal property, while they are almost certain to be there on a Tuesday.

Next, we obtained data from the Federal Reserve Bank of St. Louis (FRED) for unemployment rates in LA over the last five years. After synchronizing time frames, we saw a negative correlation between the number of crimes and unemployment rate, as seen below. This included omitting the late 2024 and 2025 plots in the bottom left of the graph. This negative relationship may revolve around one of the greatest social



events of the 2020s: Covid-19. From 2020 to 2022, crime was lowest in 2020 when most stores were shut down or not running at full capacity. This occurred simultaneously with the highest unemployment rates of the 2020s. As nationwide fear subsided and the stay-at-home order was lifted, unemployment rates began to lower. However, crime rates began to rise as stores were more open, and people had no restrictions on going outside.

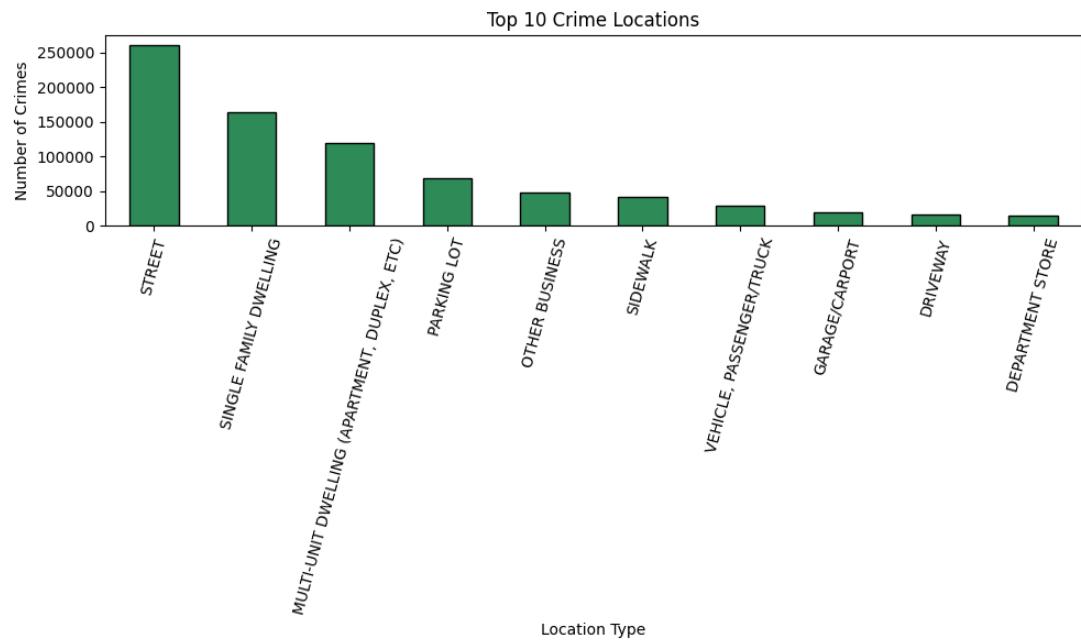
As mentioned before, all 2025 data was found to be an outlier. Many of the months from 2024 were outliers as well. This is likely due to the LAPD prompt that the crime reporting system would be changing in March 2024. However, a January 2025 report suggests that most 2024 data is available, but that the LAPD is working through issues gathering 2025 crime reports and making them available. The crime density heatmap, shown to the right, displays this phenomenon perfectly. There were easily 15,000



or more crimes per month until April 2024, the first full month after the reporting system change. The rest of 2024 shows the decline in crimes per month, and it is unknown if this is a result of the reporting change or an actual decline in crime via some measure of police enforcement. The issues noted in January 2025 remain present for the rest of 2025.

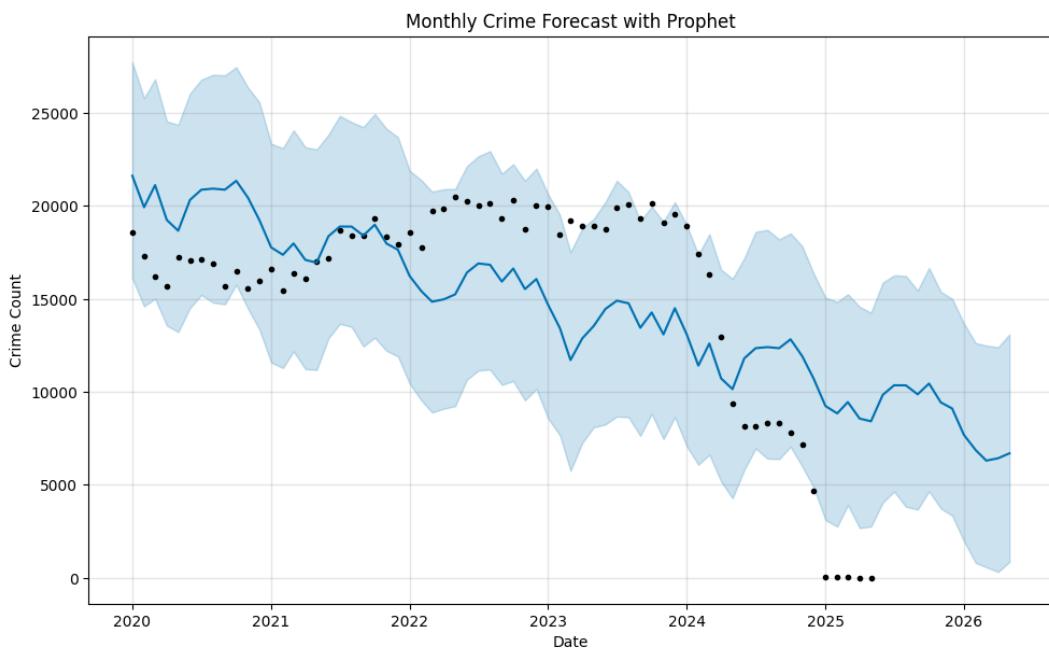
Analysis now shifts to crime as it relates to demographic factors provided by the data. The number of crimes is greatest at night, which is between 9 PM and 5 AM, and least in the morning, between 5 AM and noon. This makes sense, as most criminals prefer darkness to avoid detection. In terms preying upon older members of the population, the most common crimes are stealing from an inebriated person, blocking the door, and train wrecks. Purse snatching and purse snatching attempts are also high on the list. The average victim age across all crimes is people in their late thirties, with women having a lightly higher victim age than men. It is likely that those with an unknown sex are babies or children (where victim gender may be less important), which is why that average age is so low. Lastly, a majority of crimes occur in the street, a family dwelling (single or multi-unit), or in a parking lot. This makes sense since the majority of all crimes were vehicle related and

would occur in a street or parking lot.



# Predictions for the Future

We used both PRIMA and Prophet to make predictions about future crime in LA. We excluded 2025 from our ARIMA model and found that crime should decrease to around 120,000 yearly crimes and stabilize there for the next three years. The Prophet prediction is modeled monthly, as opposed to yearly for ARIMA, and does include 2025 data. This causes Prophet to predict monthly crimes reaching as few as 7500 crimes per month by 2026, decreasing at an average rate of 2500 crimes per month since 2020.



## Conclusion

Even with a decreased crime count in 2024, it is okay to use 2024 data until an alternative reason for potential missing data is stated. However, 2025 data is likely missing or incomplete. Therefore, it is okay, for now, to assume the visualization of a decreasing crime rate is correct, but only to an area near 2024 monthly crime rates (10,000-15,000 a month). Data regarding demographics should not be greatly affected by this. Due to an overall negative correlation between unemployment rate and crime (and a relatively low unemployment rate now as compared to 2020), it is possible that the decrease in crime rate in 2024 can be explained by LA enforcement of crime.