Crime Data from 2020 through March 1, 2023, Analysis and Recommendations

Authors: Jose Aviles, Desarie Gonzales, Javier Gutierrez, Patricia Santana, Abby Ruiz and Marko Vuksanovic
California State University, Los Angeles
FEMBA
BUS 5100-93 Intro to Business Analytics

Abstract: By analyzing crime data, type of offense, time of day, area concentration/region, and victim demographics, we will provide the police department with information on how to better serve the citizens and where to concentrate resources. Based on the findings, we will analyze the crime descriptions and suggest new training/strategies for the police department. We believe this data is essential to allocate resources appropriately based on need. Police departments should focus on identifying high-crime areas and allocate resources accordingly. Police departments can use predictive analytics to forecast future crime trends and adjust their resources accordingly. In addition, analysis of this data is conducted using Excel and SAC, depicting visuals such as charts, timelines, and maps of crime in the City of Los Angeles from 2020 through March 1, 2023.

1. Introduction

This paper uses SAP Cloud to process and visualize the current and historical data of crime in the City of Los Angeles, CA. The dataset was retrieved from Data.gov and it consists of information surrounding the details of crimes committed within the 21 police station locations within the City of Los Angeles from January 2020 through March 1, 2023. We selected this dataset to analyze the crime trends in the City of Los Angeles to provide the Los Angeles Police Department (LAPD) recommendations on how to utilize the information for planning and preventative measures. The City of Los Angeles is not only the home of 3.8 million residents (2021 Census Data), it also is a place where people come to work and to visit. Data analytics can help the LAPD keep the City of Los Angeles as safe as possible. The recommendations we provide to the LAPD will strengthen their strategies in reducing/preventing criminal activity.

2. Related Work

LAPD crime data has been analyzed by numerous profit- and non-profit organizations due to the high-level of importance and concerns over safety in the second most populated city in the United States. In the article, "Crime in Los Angeles jumped 11.6% in 2022" by Data Driven Stories for Los Angeles, a non-profit organization who collects data to better inform citizens and hold law enforcement agencies accountable, the author Khan used the same LAPD crime data to illustrate how several types of crimes have changed year over year since 2010. According to their 10-year data, assaults and vehicle theft have hit the highest levels in 2022. Our analysis only reveals data from 2020 to March 2023, which can be more relevant to a person concerned with crime during the existing and post Covid 19 pandemic. Khan's article also revealed stolen vehicles, burglaries, homicides, and assaults for the period. Similarly, our data reveals the top crimes from 2020 to March 2023 showed assaults to be the number one crime committed within the period. Since the article extracted data for 10 years, they can analyze crime data descriptions and count to assess if L.A is a much more dangerous place, and according to the data, there was 1.1% more crime from 2021 to 2022.

Additionally, the LAPD released its yearly report using the data in our analysis to inform all stakeholders of the crime rates in the city for 2021. The report is called, "LAPD Crime & Initiatives", and states the rates of homicides, Part 1, violent, and property crimes. The data in the report goes back to 1992 and up to 2021. Similarly, our crime data includes "Part 1" crimes which means they are more series crime classifications, but in comparison, we also have "Part 2" which are less severe crimes. Overall, LAPD states its goal is: "Improving organizational accountability, leveraging technology, and training, and maximizing our workforce potential" in which our whole data analysis is about.

3. Specifications

The dataset was retrieved from Data.gov (https://catalog.data.gov/dataset/crime-data-from-2020-to-present) which is an official website of the United States government, where data sets can be published and examined by the public. The data is updated daily, with data for the City of Los Angeles to be inclusive of 2020 to 2023 year to date, although this analysis has a firm cut of date of March 1, 2023.

This data set is transcribed from original crime reports with location information reported as the nearest hundred block for privacy purposes. In addition, due to transcription of typewritten reports, there may be inaccuracies or missing information within the data, denoted with (0°, 0°) in the impacted field. The dataset is constantly updated and therefore, an analysis can be conducted of crimes reported within the data period. The dataset consists of all reported criminal activity for the City of Los Angeles at its 21 stations: 77th Street, Central, Southeast, Hollywood, Newton, Southwest, Rampart, Olympic, Pacific, Harbor, West Valley, North Hollywood, Hollenbeck, Mission, Wilshire, Topanga, Northeast, Van Nuys, West L.A., Devonshire and Foothill. The size of the data set is 167.9 MB.

4. Data Cleaning

The CSV file was downloaded for the Catalog database. The data represented crime in Los Angeles city from 2020 to March 1, 2023. The dataset was cleaned by removing data that didn't apply or added bulk, making the file larger. We removed the following columns Crm Cd2, Crm Cd3 and Crm Cd4. Additionally, removed the space between the address in location and fused Time OCC and Date OCC to provide us with additional ways to formulate the data. Lastly, we froze the first row making it the header of the data and uploaded it into SAP Analytics Cloud.

5. Analysis and Visualization

5.1 Top 10 Police Department w/ the Highest Crime. 2020-2023

The first chart (Figure 1), a donut chart, was created in SAP Analytics Cloud and shows the highest crimes produced per Police Station. This visual representation was created with the intent which police stations tend to deal with the most crime during 2020-2023. Looking at the data we determined that Central (**12.08**), 77th street (**10.89**), Pacific (**10.79**) were the top police stations that dealt with the highest crimes.

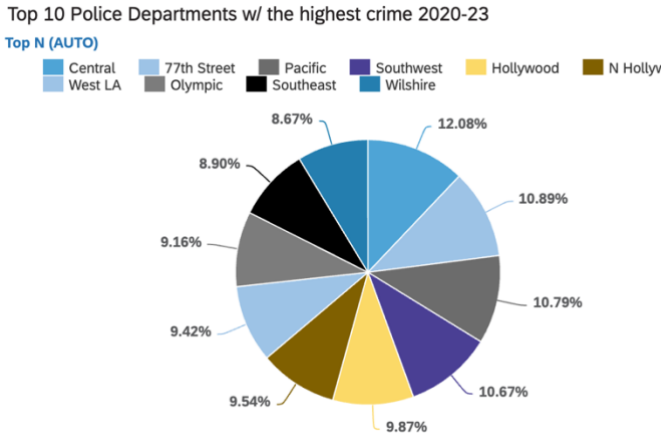

Figure 1. Donut Chart "Top Police Departments w. The Highest Crime 2020-2023"

5.2 Top 5 Weapons Used From 2020-2023

When analyzing the weapons used in crimes, we used a bar graph. The graph was created with SAP Analytics Cloud and formatted to rank the top weapons used between the time period of 2020-23. This visual representation displayed a strong arm (**127,306**) as the highest weapon used in crimes in Los Angeles. Figure 2 ranged weapons, with a semi-automatic pistol (**5,472**) and Handguns (**14,681**) being at the bottom.

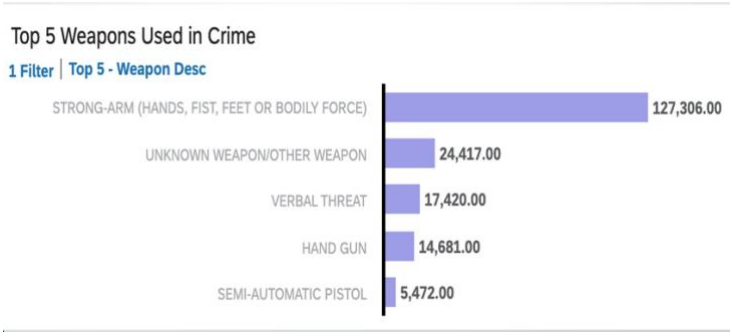

Figure 2. Bar Chart "Top 5 Weapons Used in a Crime 2020-2023"

5.3 Top Crimes Committed by Gender 2020-2023

Figure 3 analyzes the top 2 crimes committed per victim gender. This chart was created with SAP Analytics Cloud and formatted by ranking the top 2. This visual representation was created to give the police awareness of the vulnerable populations that can become victims of these specific crimes. When observing the data, we found that Battery Assault (**28,121**) was top crime for males and Assault by an intimate partner (**28,121**) was the top crime from Female victims. For the unknown gender the top two crimes were robbery and burglary.

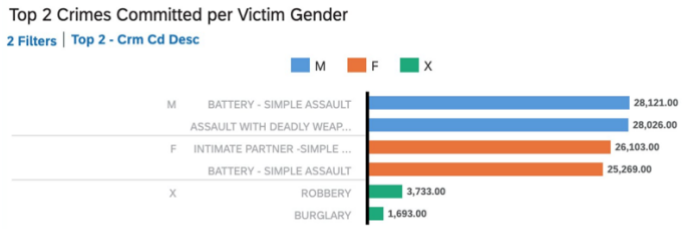

Figure 3. "Top 2 Crimes Committed per Victim Gender"

5.4 Geo-Heat Map and Tempo-Spatial Analysis

Geo-Heat Map

The LAPD consists of 21 divisions which are community police stations spread throughout the city of Los Angeles. For our Geo-map, we constructed a heat map because it was a good visualization to show where the hot spots are within each station's districts. Figure 1 below shows the "Central Station" located at 251 E 6th St, Los Angeles, CA 90014, and it gives us a count of crimes committed in that location. The outlined blue area falls out of the latitude and longitude of the reported crime committed. The yellow area can be considered safer than the red area since it is a hot spot for crimes within our data set and indicates where previous crimes have occurred by the count and where the LAPD should patrol.

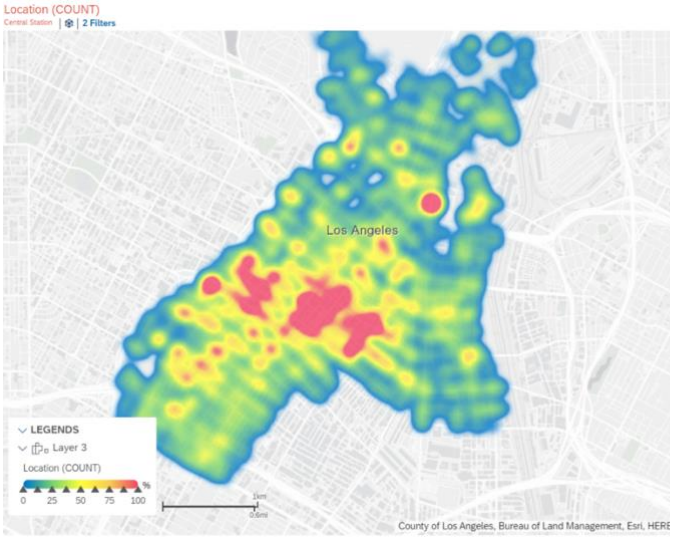

Figure 2. Geo-Heat Map

Furthermore, location and police presence are not enough to prevent more crimes. So, we looked at the time series within the data set and concluded there was a time and location correlation between the crimes committed. In figure 2, we found that the most crimes committed were at 12 pm and 6 pm. Both data sets can provide guidance to better serve each station and allocate resources efficiently.

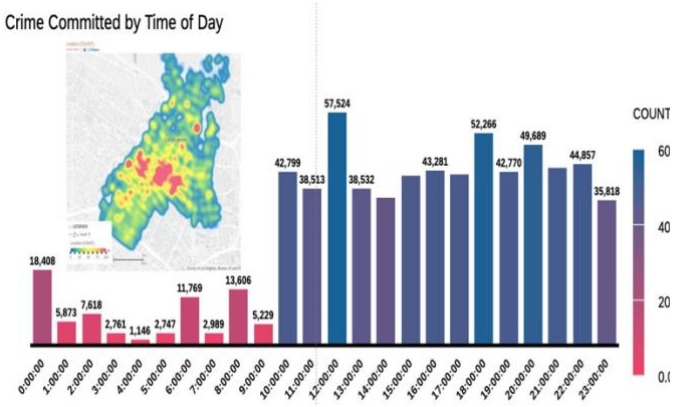

Figure 3. Bar chart "Crime Committed by Time of Day."

5.5 Regression Analysis

The 2020 Crime Dataset was aggregated to ensure that our margin of error was not high as there were categorized left blank for several columns.

Using the Crime dataset, we created a predictive Crime count model using regression analysis. The average crime count was used as the predictive goal. The root means square error (RMSE) came to 2.2, where the closer to zero the RMSE was, the better the model would be due to fewer errors (Figure 4) The low score indicates how small the margin of error is in the model when comparing the predicted vs. actual data. At the same time, Prediction Confidence is 99.08% (Figure 4). Prediction Confidence measures if the predictive model can make the predictions with the same reliability and capacity when using a new dataset, the closer that percentage is to 1 the better.

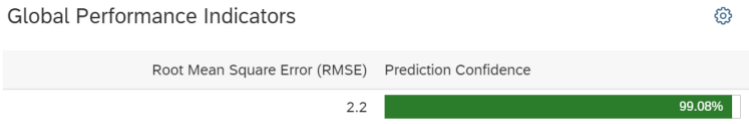

Figure 4. RMSE 2.2 & Prediction Confidence 99.08%

Figure 5 represents Predicted vs. Actual chart data while showing the crime count per Month and per Station. The regression analysis shows a Crime Count Validation - Error Max of 80 and a Validation - Error Min of 1. This helps indicate that our predictive model is close.

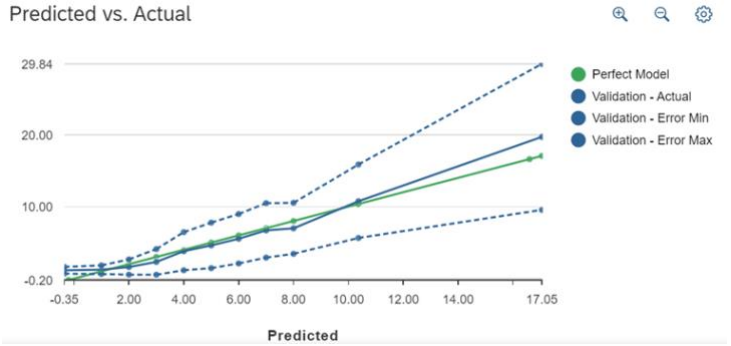

Figure 5. Predicted Crime Count vs. Actual Crime Count

Figure 6 shows the biggest influencers that contributed to the analysis. From highest to lowest are: Premis Description (43.49%) , Crm Cd Description (41.50%) , and Area Name (15.00%)

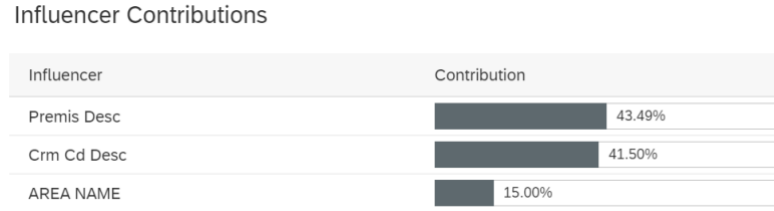

Figure 6. Influencer Contribution

Figure 7 shows a bar chart with the Predicted vs Actual Crime Count by Month. Comparing the monthly trends of predicted vs actual crime count shows how close our predictions are to actual crime count in Los Angeles City.

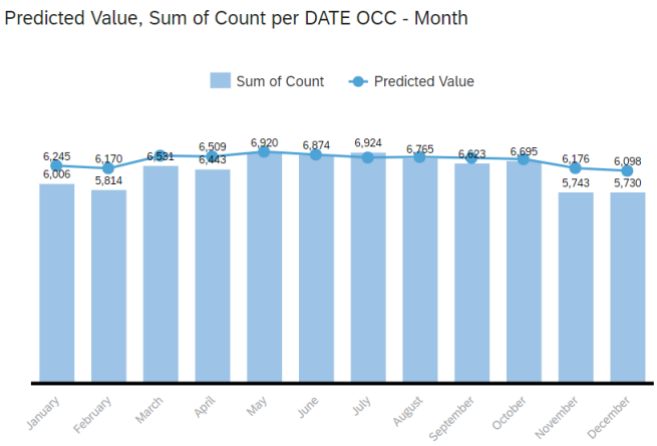

Figure 7. Predicted Crime Count Value Bar Chart by Month in 2022

Figure 8 shows a bar chart with the Predicted vs Actual Crime Count by Area Name. Comparing the trends of predicted vs actual crime count shows how close our predictions are to actual crime count by Area in Los Angeles City. When we sorted the data, we looked at the Two areas in Los Angeles city with the highest crime count: 77th Street (6,673) and Central (6,499) and the two lowest areas in Los Angeles city with the lowest crime count: Devonshire (2,402) and Foothill (2,291).

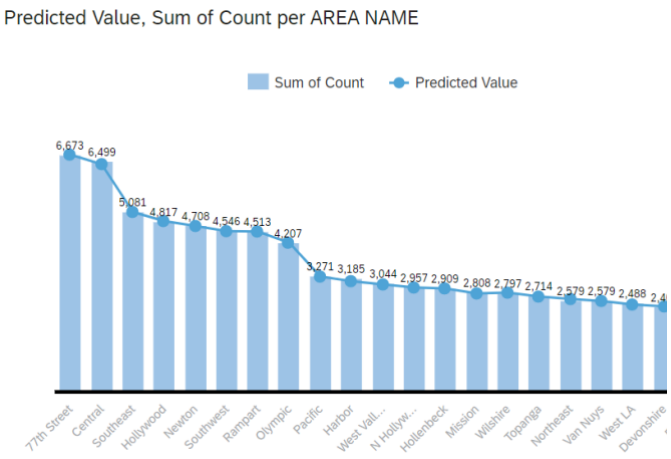

Figure 8. Predicted Crime Count Value Bar Chart by Area Name in 2022

5.6 Time Series Analysis and Forecast

The graph below (Figure 9) shows the time series analysis and forecast for crime incidents in the City of Los Angeles. The graph clearly shows a yearly seasonal trend with high crime counts in summer months and conversely lower crime counts in the cold season. We can also spot spikes due to holidays and events/protests that involve gatherings of large crowds.

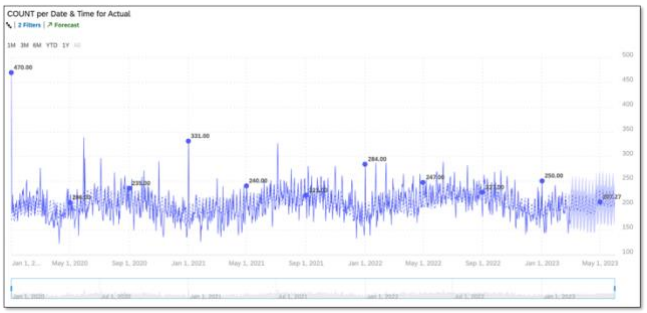

Figure 9. Time Series Analysis and Forecast of Crime Incidents

6. Conclusion

The basic & continued (https://post.ca.gov/Peace-Officer-Training) training standards for LAPD peace officers were reviewed and compared against the top crimes reported. After review, we concluded that although the LAPD does have a well-documented training program for their officers, there are some areas where additional training is needed as it pertains to the highest volumes of crimes. Finally, summing up all the above work and research, we conclude that the LAPD implement the following recommendations:

i. Increase patrolling at 12 and 6 pm in the hot spot areas identified on the heat map.
ii. All officers receive training in Verbal Judo.
iii. Domestic Violence training frequency updated from bi-annually to annually.
iv. Use of Force training frequency updated from bi-annually to annually given the top two crimes committed against both men and women are related to assault and/or battery.

References

[1] Crime Data from 2020-Present (https://catalog.data.gov/dataset/crime-data-from-2020-to-present)
[2] Census Data for City of Los Angeles https://www.census.gov/quickfacts/losangelescitycalifornia
[3] Commission on Peace Officer Standards and Training (https://post.ca.gov/Peace-Officer-Training)
[4] Los Angeles Police Department 2021 Crime and Initiatives (https://lapdonlinestrgeacc.blob.core.usgovcloudapi.net/lapdonlinemedia/2022/01/2021_crime_initiatives.pdf)
[5] Crime in Los Angeles jumped 11.6% by 2022 (https://xtown.la/2023/01/23/crime-rate-los-angeles-2022/)