TIME SERIES ANALYSIS OF CRIMES IN LA

TREVOR GARRITY, DRAKE BROWN, DANIEL PERKINS, DAVIS HUNTER

Abstract. Crime is a complex societal issue that impacts communities and individuals, requiring thoughtful analysis to inform effective policymaking. In this project, we analyze crime data in Los Angeles to identify trends, seasonal patterns, and factors influencing crime volume. Using a dataset containing millions of crime records from 2020 onward, we clean and structure the data for time series modeling. We decompose the time series into trend, seasonal, and residual components to better understand long-term patterns. An Autoregressive Integrated Moving Average (ARIMA) model is used to predict future crime trends. Additionally, we explore demographic factors associated with crime volume and type, highlighting correlations without implying causation. Our analysis reveals significant seasonal patterns, demographic disparities, and age-related disparities, providing insights that could guide policymaking and resource allocation by law enforcement and public safety officials.

1. Problem Statement and Motivation

Crime is a pervasive global issue that impacts societies at every level, undermining economic stability and social well-being. Analyzing how different types of crime evolve over time is crucial for enhancing public safety, as it equips policymakers with the insights needed to develop effective policies and interventions.

In this study, we analyze crime data through the lens of time series analysis, identifying trends, seasonal patterns, and anomalies. Our findings contribute to the growing field of data-driven criminology, offering a quantitative perspective on crime dynamics. We analyze the weekly and yearly seasonal components of the total number of crimes over time. We employ an ARIMA model to forecast future crime trends, providing insight into potential shifts in overall crime rates. Finally, we investigate how factors such as victim age, sex, race, day of the week, and day of the year correlate with the types and frequency of crimes committed.

There have been a number of works analyzing the the temporal nature of crime (this list is in no way comprehensive due to the magnitude of literature). Andresen et al. found that crime rates typically increased over the weekend and that different days of the week correlated to crime occurring in different locations ([AM15]). Wyatt Lam found that major holidays increased violent crime rates, but decreased property crimes ([CR03]). Some

Date: 12 April 2025.

have attempted to use modern machine learning models to predict crime hotspots for certain times of the year([ID25]). Lastly, Cohen et al. found that a likely reason for crimes rates increasing over the years may be due to "the dispersion of activities away from households and families" ([CF79]).

2. Data

Our dataset comes directly from the Los Angeles Police Department [Dep25] and includes 1,005,104 records of crime incidents in Los Angeles from 2020 to the present. Since the data was transcribed from original paper-based crime reports, some inaccuracies may exist. However, the source is reliable and contains limited bias, making it well-suited for our analysis. Each record includes a timestamp, inherently structuring the dataset as time-series data. Key features utilized in our study include the type of crime (e.g., burglary, theft, arson, assault), as well as the victim's gender, age, and race. However, the dataset lacks the perpetrator's demographic information and information on the economic class of both victims and perpetrators, which could have provided additional context for our analysis.

2.1. **Data Cleaning.** To prepare the dataset for analysis, we first removed the 142 crime incidents with an unknown case status or a negative victim age and eliminated redundant columns. Given the large size of the dataset and the significantly small proportion of missing data, this approach allowed us to maintain a robust dataset without compromising its integrity. We also found a large number of crimes with a victim age of 0. After manually checking the corresponding crime descriptions, we found that these crimes were victimless. We therefore replaced every instance of a victim age of 0 with NA. When conditioning on victim age, we ignore these values. We understand that this could inadvertently remove the age of crimes where the victim actually was infant, but we presume that the number of such crimes is quite small (which our later analysis verifies). Furthermore, we applied one-hot encoding to the most common categorical features and removed columns with an excessive number of unique values to streamline the dataset and improve its analytical efficiency.

Next, we restructured the data into a time series format, where each entry represents the total number of crimes that occurred within a single day. This transformation aligns with our understanding of crime patterns: Although individual crimes may not directly influence one another, aggregating crimes by day allows us to capture potential temporal trends and dependencies in criminal activity over time. The dataset also recorded which crimes were "Part 1", which generally corresponded to violent or property crimes. We therefore created a separate time series that counts the number of violent or property crimes each day. For brevity, we sometimes refer to this quantity as the number of violent crimes. For both time series, the last few months of data had significantly lower numbers. This is likely due to crimes not yet being reported, so we removed the last 100 days from the time series.

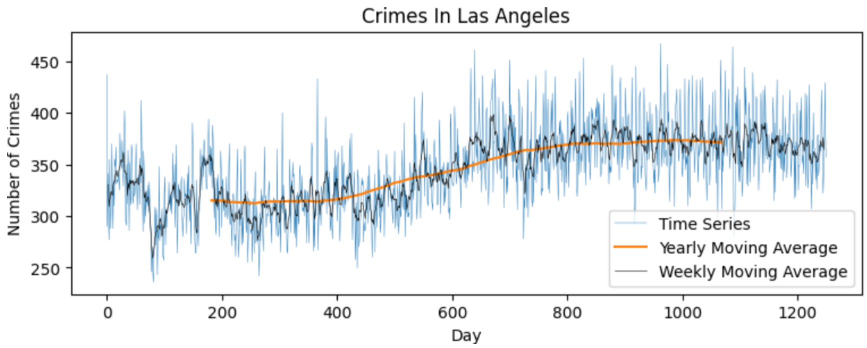

FIGURE 1. This plot shows the daily crime count over time. To better visualize the overall trend, we computed both yearly and weekly moving averages, highlighting long-term and short-term fluctuations in crime patterns.

2.2. **Train-Test Split.** Since the dataset is inherently time-based, randomly splitting it for training and testing would disrupt its temporal structure. Instead, we preserved the chronological order by assigning the first 70% of days to the training set, the next 10% to the validation set, and the final 20% to the test set.

3. Methods

3.1. **Detrending The Time Series.** The first step of our analysis was to determine how crimes change over large periods of times. Given that crimes tend to increase on holidays and decrease during regular workweeks, a yearly seasonal component is likely present in the data. Additionally, since crime rates are generally higher on weekends than on weekdays, our time series also exhibits a weekly seasonal pattern. Unfortunately, we could not include a monthly seasonal component, as the varying number of days in each month would have introduced inconsistencies in the analysis. Therefore, we split the dataset into a trend, two seasonal components, and a residual. In other words, we have

$$Z_t = T_t + S_t^{\text{yearly}} + S_t^{\text{weekly}} + R_t \tag{1}$$

As expected, the original time series (Figure 1) is quite noisy due to fluctuations caused by yearly and weekly seasonal components. To decompose the time series into the four components, we first found the trend by taking the length-365 moving average of Z_t via the formula

$$T_t \approx \hat{T}_t^{(365)} = \frac{1}{365} \sum_{j=-182}^{182} Z_{t+j} \tag{2}$$

To find the yearly seasonal components, we subtracted the yearly moving average from the original time series to get $y_t = Z_t - T_t$. Then we averaged

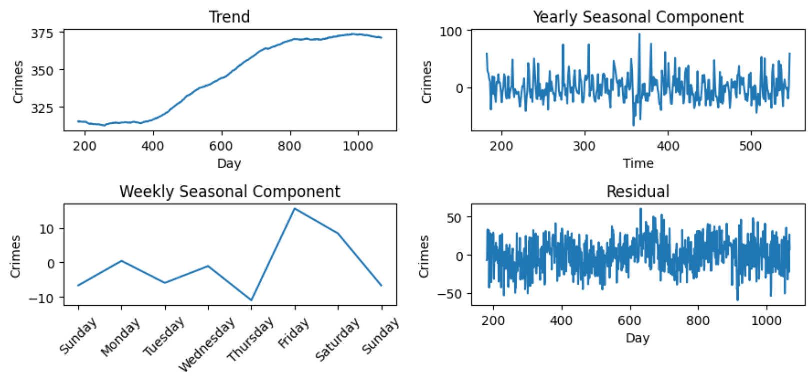

FIGURE 2. These plots show how Z_t decomposes into $T_t + S_t^{\text{yearly}} + S_t^{\text{weekly}} + R_t$. After removing the seasonal components and residuals, the trend reveals a long-term increase in crime, which eventually tapers off. While the overall pattern appears relatively linear, it's shape more closely resembles an affine transformation of the sigmoid function.

over all values of one period

$$S_k^{\text{yearly}} \approx \frac{1}{\lfloor (N-k)/365 \rfloor} \sum_{t \equiv k \text{ mod } 365} y_t \quad \forall k \in \{0, 1, ..., 364\} \tag{3}$$

Next, after subtracting S_k^{yearly} from y_t, we calculated S_k^{weekly} using the same equation 3.1 (where the period is changed from 365 to 7). Finally, we computed the residual by subtracting the trend and both seasonal components from the original time series. The results are shown in Figure 2

3.2. **Factors influencing Volume of Crime.** By splitting up the time series according to various aspects of the data we are able to identify trends in location, weapon used, type of crime, age of violator, and other demographic factors of violator.

We note that although our model may associate certain attributes like sex or ethnicity with a higher volume of crime on a specific day, this is not a causal model in that other factors likely contribute to any amount of crime. For our main results refer to section 4.1.

3.3. **ARIMA Model and Crime Volume Prediction.** To model and predict future crime volumes, we utilized an AutoRegressive Moving Average (ARMA) model. ARMA models are well-suited for time series data with stationary behavior, making them ideal for capturing the temporal dependencies in crime patterns.

Time series models like ARMA require stationarity, where the mean and variance remain constant over time. However, the crime data is non-stationary. To account for this, we applied first-order differencing to both

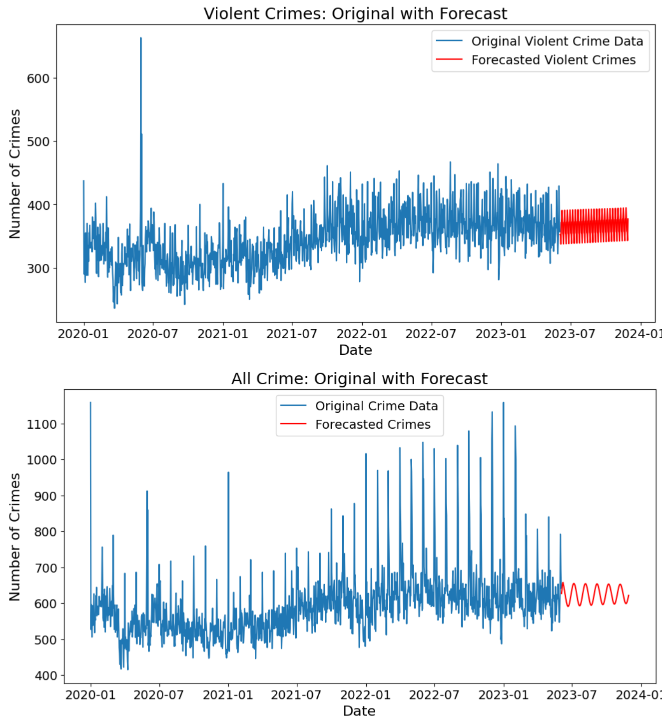

FIGURE 3. Time series forecasts of violent and total crime volumes based on ARIMA modeling.

the violent crime and overall crime time series to stabilize the mean and remove any trends.

To find the optimal parameters of the model, we used the Akaike Information Criterion (AIC) to compare model performance. After doing a grid search over a range of values for p and q, we found the model that minimized the AIC, which was an ARMA(5,5) model for the violent crime data and an ARMA(6,4) model for overall crime data.

We used the best-fitting ARMA model to forecast future crime volume. These predictions provide an estimate of expected crime patterns based on historical data and can offer insights for policymakers and law enforcement agencies. Initial forecasts for violent and overall crime volumes were generated, with plots of the predicted values shown in Figure 3.

To evaluate the model's performance, we compared the ARMA predictions to the actual recorded values. Figure 4 shows this comparison for both violent crimes and all crimes.

We then calculated how the accuracy changes as we predict further into the future (Figure 5). For both violent and all crime data, the ARMA model captures periodic trends effectively, even when daily predictions deviate from actual values. As expected, prediction accuracy decreases as the forecast horizon increases, but the error appears to stabilize over time.

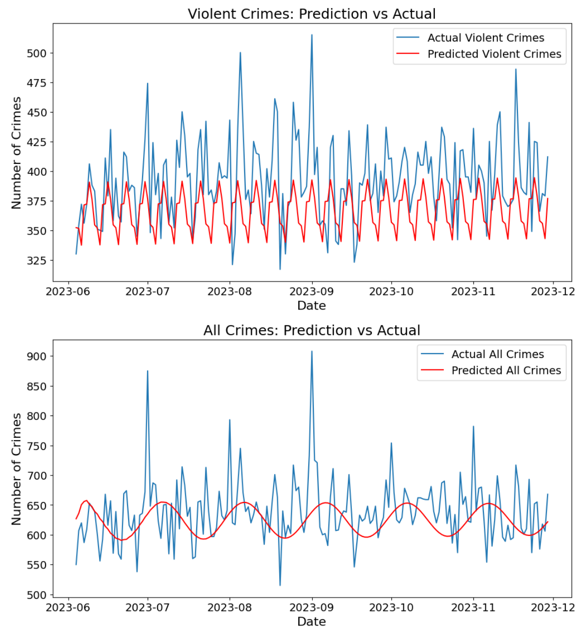

FIGURE 4. Predicted vs. actual crime volumes for violent and all crimes. The ARMA model captures seasonal patterns despite deviations in day-to-day accuracy.

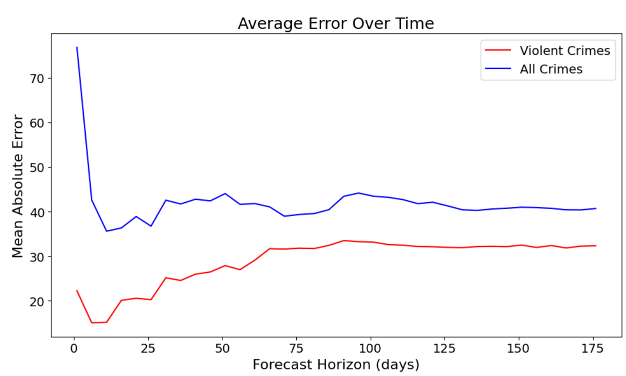

FIGURE 5. Mean absolute error (MAE) over time for violent and all crimes. As the number of days increase, the error gradually increases and then plateaus.

In the first week of predictions, the model's average daily error is approximately 14.57 for violent crimes and 36.65 for all crimes. After six months, the average daily error increases to 32.31 for violent crimes and 40.35 for all crimes. This shows that short-term forecasts are more precise, but the model still provides insightful season trends in the long term.

3.4. **Time-Invariant Analysis.** We calculated general correlations in the data and with the demographic data of the city of LA and used various statistical tests to measure the significance of any deviations. The data was obtained from two different sources within the U.S. Census Bureau. First, data on race demographics was obtained from the 2020 Decennial Census Redistricting Data for the city of LA ([Bur20b]). There is a separate census from that same year called "Demographic and Housing Characteristics" that contains racial demographics ([Bur20a]). However, this second census does not differentiate Hispanic/Latino as its own race while the redistricting data does. Since the crime data separately records Hispanic/Latino as its own racial category, we decided to use the redistricting data. Second, we used the 2023 American Community Survey ([Bur23]) for the city to get demographic data of age and sex.

We used the Pearson Chi-squared test and the G-test to calculate p-values for the distributions of victim age, sex, and race compared to city demographics. We also calculated the proportion of crimes divided by the proportion of people within each demographic. Finally, we used the Chi-square test of independence to calculate the p-values of different features being independent, and we calculated the proportion of each feature along varying demographics.

4. Results and Analysis

4.1. **Trend and Seasonal Analysis of various crime factors.** Due to the large number of factors influencing this time series, we cannot include all the graphics illustrating our results. Instead, this section summarizes the most interesting patterns (in our view) from the analysis, supported by a selection of key graphs. A full list of all generated plots can be found in the appendix.

Key takeaways include:

- Crimes in single-unit and multi-unit housing spiked sharply on the first of the month and around the Fourth of July. General spikes in crime were also observed at start of each month.
- The most frequently used weapon in crimes was a "strong arm" or the perpetrator's own body. These crimes peaked on Saturdays, with a notable spike around the Fourth of July.
- Identity theft saw a significant rise in reported incidents from 2020 to 2024. These crimes were most common at the beginning of the month and on Thursdays or Sundays.
- Burglaries—particularly in single-unit houses—were most likely to occur on Thursdays, whereas domestic assaults peaked on Saturdays.
- Most crimes occurred on Fridays or during the first few days of a new month.
- The total number of crimes committed against individuals aged 20–40 increased overall. While Fridays saw the highest crime volume

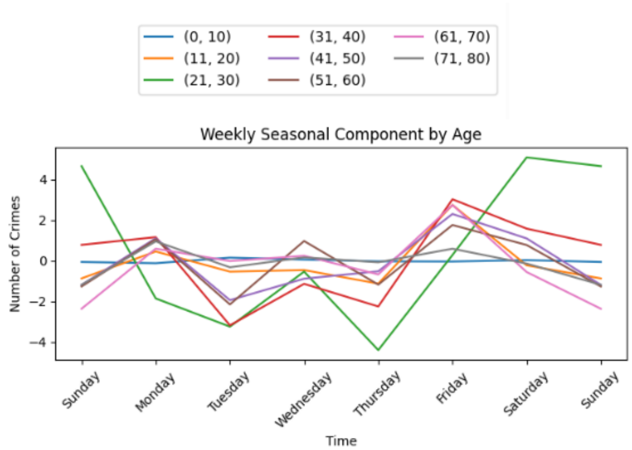

FIGURE 6. Depiction of Weekly Seasonal Component of Crimes according to Age. Note that the crimes committed by 20-30 year olds peak at a different time than the rest.

across all ages, 20–30-year-olds exhibited a much larger proportion of offenses on Saturdays and Sundays (see Figure 6).

- Geographically, Central LA experienced a sharp increase in crime between 2020 and 2024 (see Figure 7).
- Crime rates among the white population were highest on Thursdays, while the Hispanic population saw a higher volume of offenses on Saturdays. (Crimes committed against said demographic)
- Although no major trend shift was observed, there was a small peak in nonviolent crimes committed against males around July 4th.

For the time-invariant analysis, the p-values were all so small that their logarithms were returned as negative infinity. Figure 8 displays a couple of plots of the proportion of crimes divided by the proportion of people in a demographic. The corresponding test statistics are also displayed rather than the p-values, as the returned p-values were not quantitatively meaningful. The first plot shows that there the number of Black victims is disproportionately high and the number of Asian victims is disproportionately low. The second plot shows that the number of victims who are children (under 20) is disproportionately very low while the number of victims who are over 60 is also disproportionately low. Figure 9 displays the proportion of different features conditioned on race, along with the chi-square test statistic. The first plot shows that Native Americans, Asians, and Pacific Islanders are proportionately much more likely to be victims of a violent or property crime. The second plot shows that Black women are more likely to be the victim of a crime than Black men, while White and Native American men are more likely to be the victim of a crime than White or Native American women.

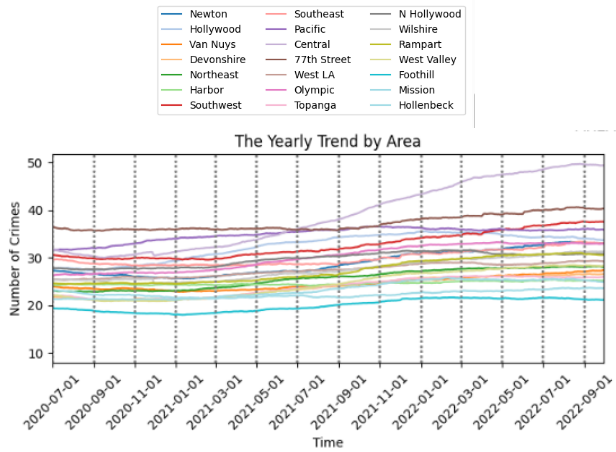

FIGURE 7. Depiction of increase in crimes in area Central of LA. Note this starts later in the year because of the time averaging window.

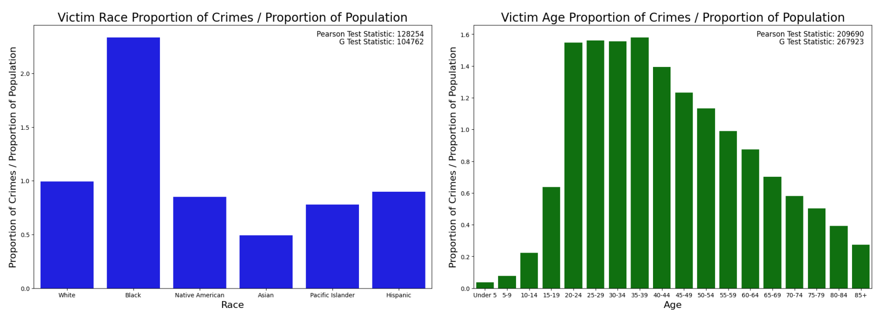

FIGURE 8. The proportion of crimes divided by the proportion of the population of varying demographics along with a couple test statistics.

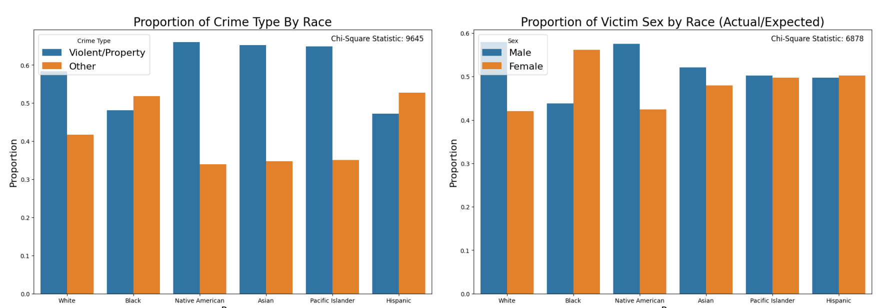

FIGURE 9. The proportion of crimes according to crime type or victim sex and conditioned on victim race, along with a corresponding test statistic.

5. Ethical Implications and Conclusions

When analyzing crime data, ethical considerations must be carefully addressed to prevent unintended consequences. We took deliberate steps to ensure an objective analysis, minimizing potential biases in our approach. While variables such as race, age, and gender can introduce concerns regarding fairness, we handled them with caution to avoid reinforcing harmful stereotypes or contributing to biased outcomes in law enforcement practices. Our methodology prioritized transparency and neutrality, focusing on data-driven insights while being mindful of ethical implications.

Misinterpretation or misuse of the dataset or our analysis could inadvertently reinforce harmful stereotypes or contribute to biased outcomes in law enforcement practices. We urge policymakers to interpret the data with caution to avoid disproportionately focusing on certain areas, which could create a cycle where increased attention leads to heightened enforcement, further amplifying crime statistics in those regions. A balanced approach is essential to ensure that interventions are data-informed without unintentionally reinforcing disparities.

We also recognize the possibility of inherent biases in police reports and data collection processes. Social and systemic factors may lead to disproportionate reporting of crimes involving certain demographic groups, which can skew the data and affect the validity of our analysis.

Furthermore, protecting individual privacy is paramount. The data used in this analysis is aggregated and anonymous, with address fields restricted to the nearest hundred block to prevent any potential re-identification. These privacy safeguards help mitigate the risk of exposing sensitive information while preserving the integrity of the analysis.

6. Conclusion

Our analysis of crime data in Los Angeles uncovered several meaningful insights that can assist policymakers and law enforcement in making data-driven decisions. By decomposing the time series, we identified both yearly and weekly seasonal patterns, with crime rates spiking during holidays and weekends. We also observed long-term trends indicating an overall increase in certain types of crimes over the study period.

The ARIMA model we used demonstrated a reasonable ability to predict future crime trends. However, the unpredictability of human behavior and other factors imply that we need to proceed with caution when interpreting these forecasts. Furthermore, although our analysis reveals correlations between certain demographic factors—such as race and age—and crime patterns, these associations should not be interpreted as evidence of causality

This project provides a foundation for further research into crime prediction and prevention. Our use of time series modeling underscores the value of data-driven approaches in uncovering crime dynamics and informing more effective intervention strategies.

References

[AM15] Martin Andresen and Nick Malleson. Andresen, m.a., & malleson, n. (2015). intra-week spatial-temporal patterns of crime. crime science, 4, article 12. *Crime Science Journal*, 4:Article 12, 07 2015.

[Bur20a] U.S. Census Bureau. Dec demographic and housing characteristics. U.S. Department of Commerce, 2020. Available at `https://data.census.gov/table/DECENNIALDHC2020.P8?g=160XX00US0644000`.

[Bur20b] U.S. Census Bureau. Dec redistricting data (pl 94-171). U.S. Department of Commerce, 2020. Available at `https://data.census.gov/table/DECENNIALPL2020.P2?g=160XX00US0644000`.

[Bur23] U.S. Census Bureau. Acs 5-year estimates subject tables. U.S. Department of Commerce, 2023. Available at `https://data.census.gov/table/ACSST5Y2023.S0101?g=160XX00US0644000`.

[CF79] Lawrence Cohen and Marcus Felson. Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44, 08 1979.

[CR03] Ellen G Cohn and James Rotton. Even criminals take a holiday: Instrumental and expressive crimes on major and minor holidays. *Journal of Criminal Justice*, 31(4):351–360, 2003.

[Dep25] Los Angeles Police Department. Crime data from 2020 to present, 2025. Accessed: 2025-03-31.

[ID25] Esen Ilgun and Murat Dener. Exploratory data analysis, time series analysis, crime type prediction, and trend forecasting in crime data using machine learning, deep learning, and statistical methods. *Neural Computing and Applications*, pages 1–26, 03 2025.

Appendix A. Time Series Split along different Keys Graphics

Here we present the comprehensive set of graphs that partition the time series up according to different factors of the crime. Note that when splitting alogn factors it does not mean that this factor is the sole cause of crime. Rather it just seems to be correlated with it in the data we were given.

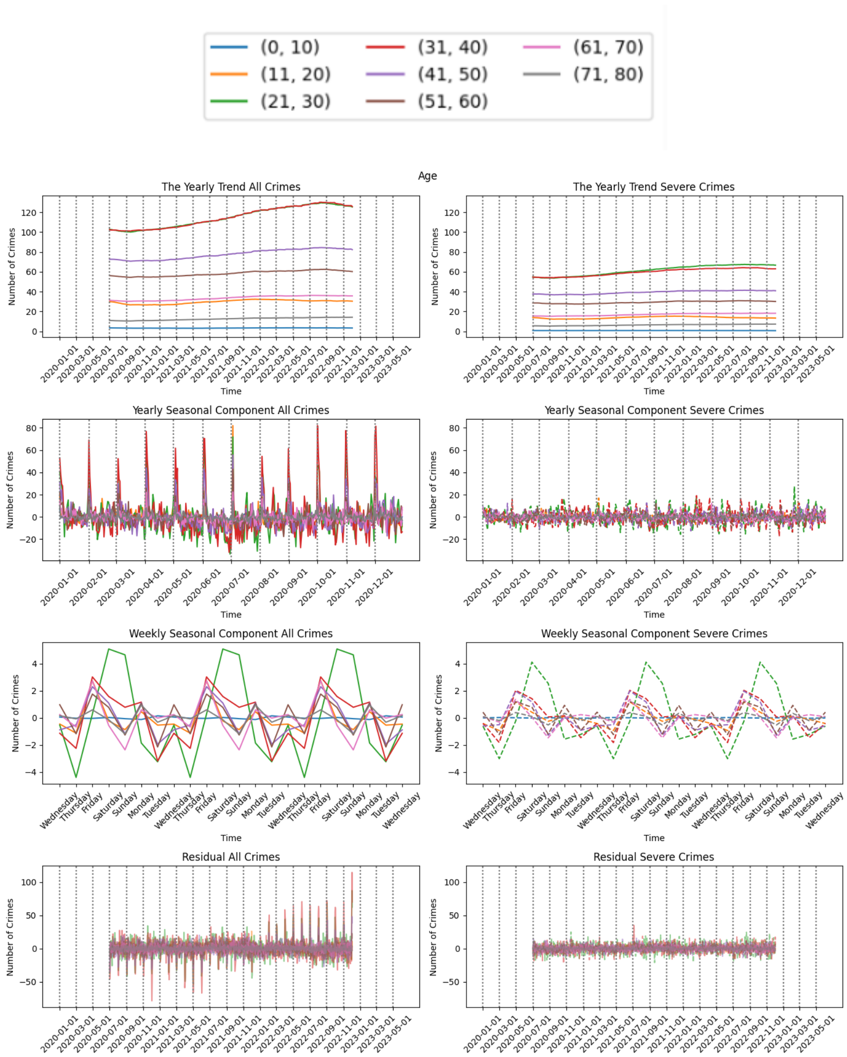

FIGURE 10. Decomposition of Crime Time series divided up by age

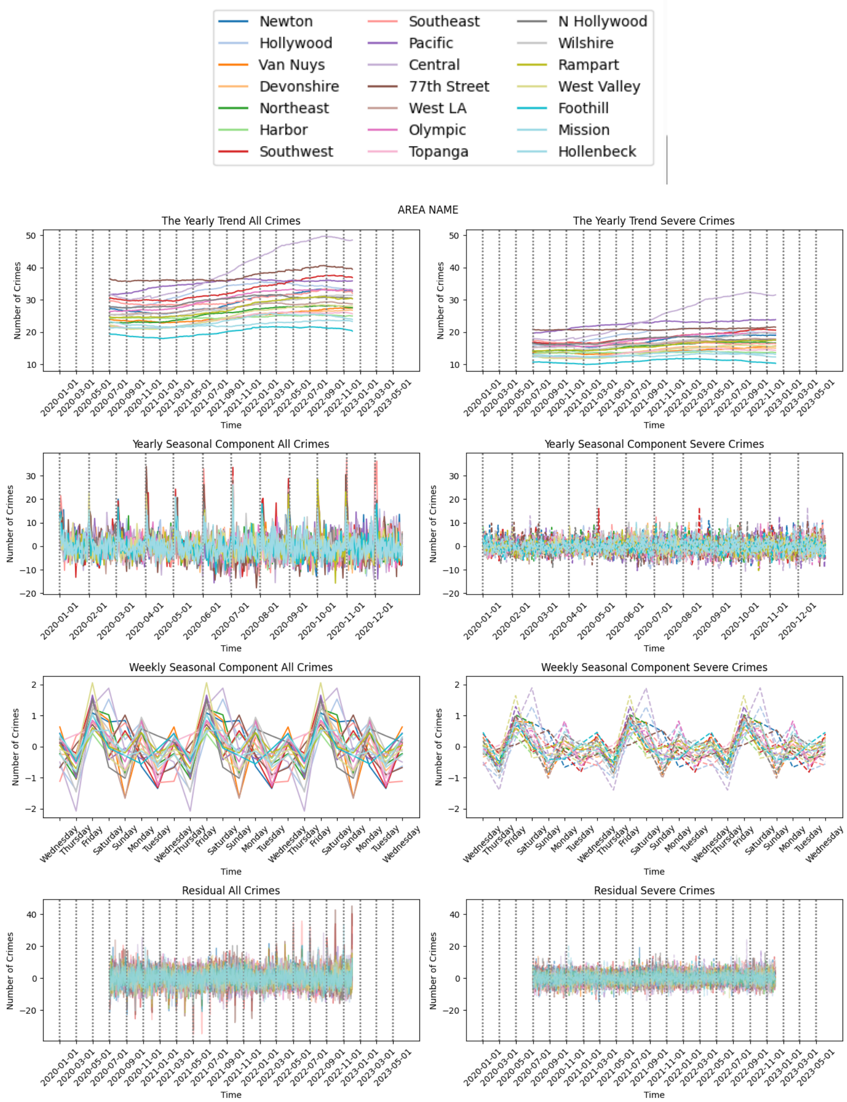

FIGURE 11. Decomposition of Crime Time series divided up by Area Name

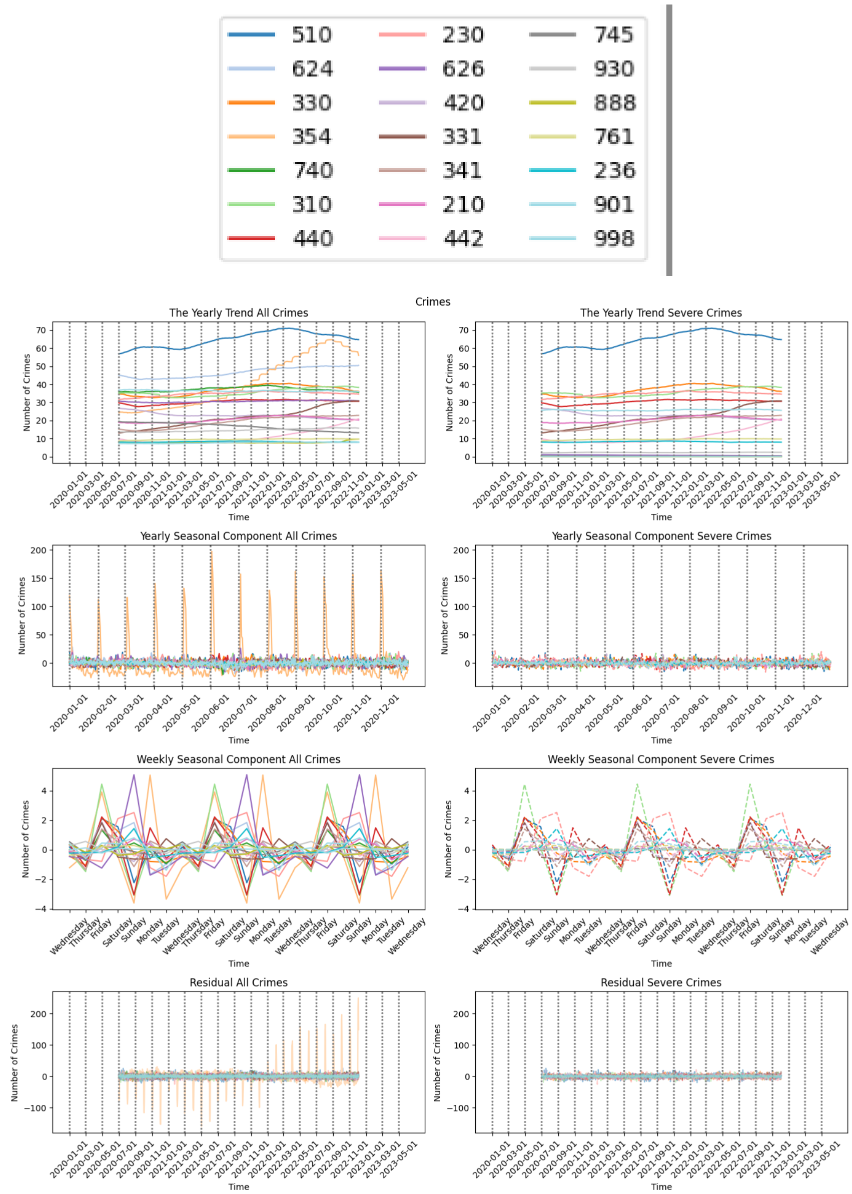

FIGURE 12. Decomposition of Crime Time series divided up by Crime Committed

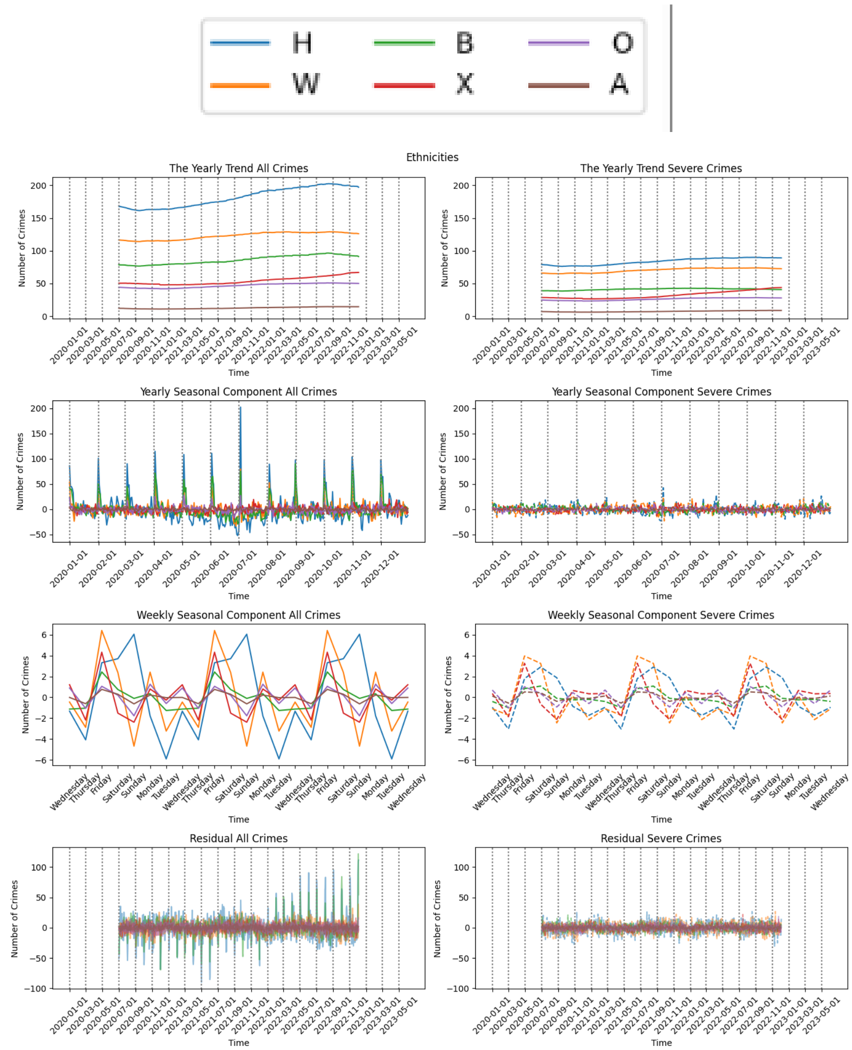

FIGURE 13. Decomposition of Crime Time series divided up by Ethnicity of individual who committed the crime.

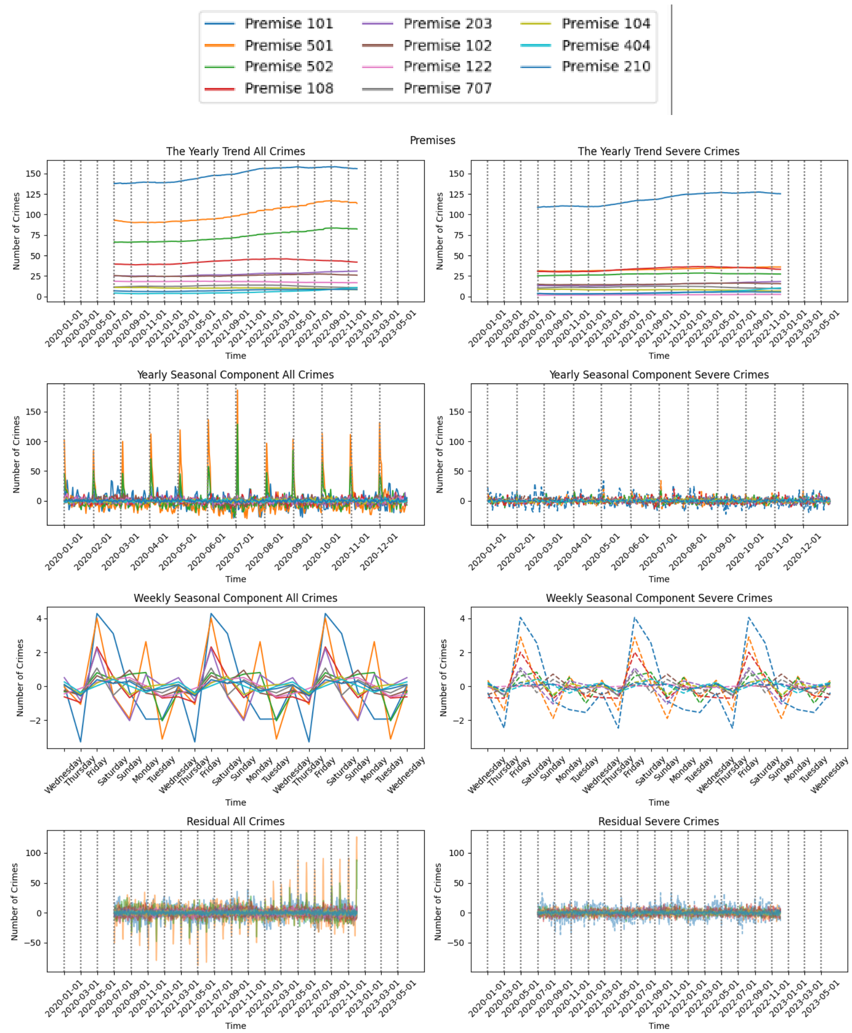

FIGURE 14. Decomposition of Crime Time series divided up by Premise (like street or house) of the crime.

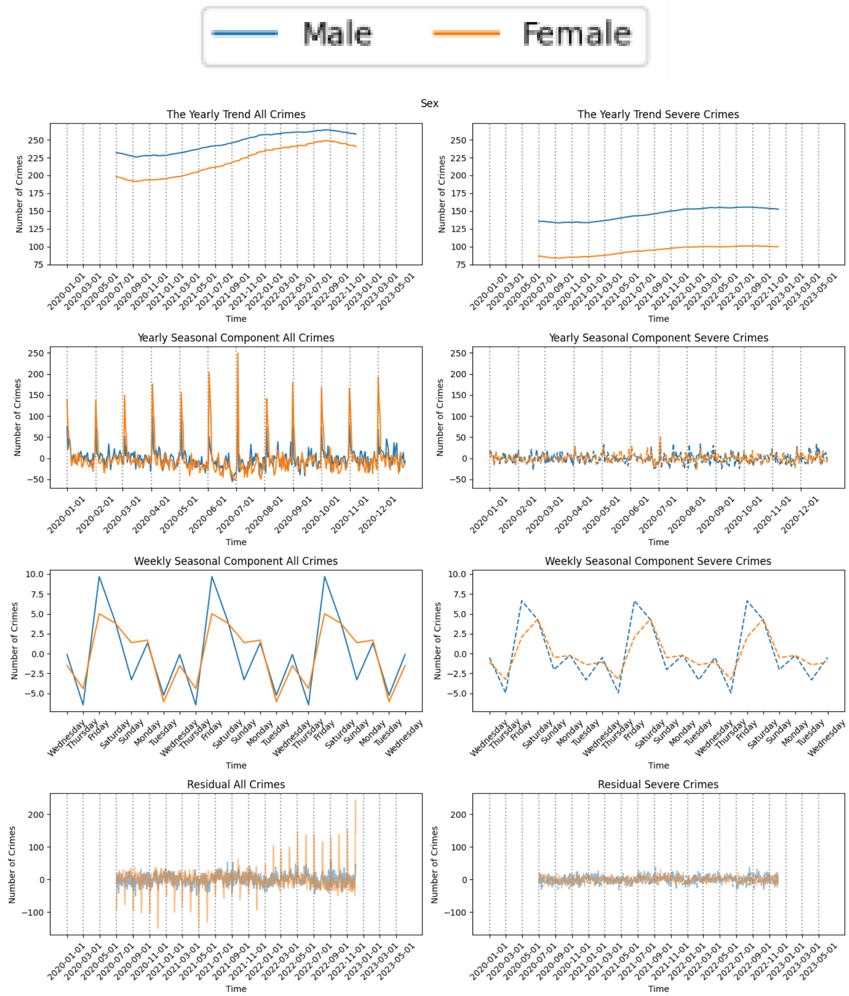

FIGURE 15. Decomposition of Crime Time series divided up by Sex of the individual who committed the crime.

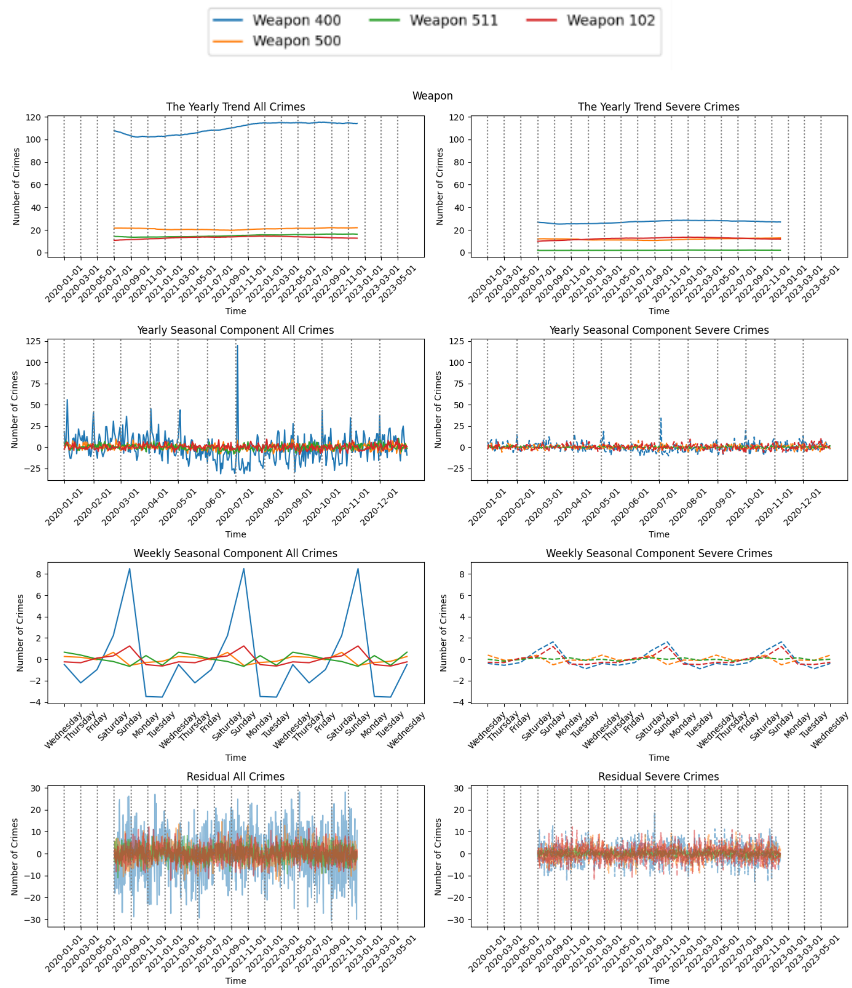

FIGURE 16. Decomposition of Crime Time series divided up by Weapon used to commit the crime.