Crime Forecasting Analysis

IE 6700 Data Management for Analytics

Final Report

Group Number 21

Sathvik Ramappa (002847460)

Saiteja Reddy Gajula (002872000)

Pooja Arumugam (002872003)

Bhakti Paithankar (002833722)

Nihal Mallikarjun(0036010381)

Part 1: Introduction

With the advent of cutting-edge technology and the previously unheard-of level of data availability, the field of crime prediction has entered a new era. This change has created a rare chance to investigate and create cutting-edge approaches for detecting crimes by utilizing the capabilities of deep learning and machine learning techniques. This paper explores current trends in the field and clarifies how crime prediction and prevention are being redefined by this cutting-edge technology.

As a branch of artificial intelligence, machine learning is a potent toolkit of statistical models and algorithms for data analysis and prediction-making. Simultaneously, deep learning, a different branch of machine learning, uses multiple-layer artificial neural networks to simulate intricate correlations between inputs and outputs. The science of crime prediction is about to undergo a paradigm revolution because to the combination of machine learning and deep learning. Algorithms for machine learning have become indispensable resources for crime prediction. A number of algorithms, including support vector machines, random forests, and decision trees, have been used to analyze crime data in an effort to predict crime patterns remarkably accurately. These algorithms offer deeper insights into the changing trends and patterns of criminal activity, beyond just their prediction ability. The allocation of resources and tactics that can successfully combat crime is made easier by this understanding. Moreover, machine learning algorithms are particularly good at identifying complex relationships between a variety of environmental and demographic factors and criminal episodes. Among the variables that are carefully considered are the location, the weather, and the time of day. The data obtained from this study serves as the basis for the creation of specially designed crime prediction and prevention plans that are well matched to the needs of particular communities. In conclusion, this study article takes a tour around the modern terrain of crime prediction. It emphasizes how important machine learning and deep learning are to improving our understanding of crime trends and strengthening our group's efforts to deter criminal activity more successfully. This study also highlights the ways in which law enforcement organizations could use these technologies, offering a more proactive and data-driven approach to fighting crime in our changing society.

Part 2: Summary of Results

Our research has established a significant relationship between the presence of external factors and the occurrence of various types of crimes, based on a thorough analysis of historical crime data and external factors." This finding demonstrated the consistent impact of outside variables on crime patterns in numerous cities and regions.Our results highlight how crucial it is to take into account outside factors when predicting crime trends, including social events, meteorological circumstances, economic indicators, and law enforcement activities. These outside variables were regularly incorporated into our prediction models to increase the accuracy of crime forecasts, which allowed for more efficient resource allocation and proactive law enforcement tactics.As a result, this study highlights the critical role that outside variables play in crime predicting and offers insightful information that law enforcement organizations, local governments, and legislators may use to improve public safety and crime prevention initiatives in a variety of metropolitan contexts. In conclusion, a number of important conclusions emerged from the examination of crime data spanning 2020 to 2023. Crime rates rose gradually between 2020 and 2022 before showing a notable decline in 2023, which may have been caused by incomplete data. For the most part of 2023, there was a discernible decline in crime rates; nevertheless, January and February stood out as anomalies, maybe as a result of altered policy.The majority of crimes, according to seasonal study, happened in the midst of the year. During the COVID-19 shutdown, identity theft increased dramatically; in 2022, crimes like credit card scams and loan fraud peaked.After analyzing the correlation between victim age and crime, it was found that, with variations among various age categories, those between the ages of 0 and 20 experienced the highest number of crimes. Crime rates were higher on Sundays.Geographically, the most violent neighborhoods were Central, Hollywood, Pacific, Southwest, and 77th Street. Males and females committed crimes at the same rates, with lower rates for other gender categories.Comprehending the age and gender-based aspects of criminal activity is crucial for both law enforcement and policy formulation. The incidence of crimes follows a diurnal pattern, which highlights the necessity for prompt, data-driven solutions to public safety issues.

Part 3: Data Sources

The main source of data used in this analysis was a government website that provided incident reports of crimes committed in the City of Los Angeles going back to 2020. It is significant to remember that the data in this dataset is a transcription of the original crime reports that were first recorded on paper. As a result, it's probable that the manual transcription method resulted in some errors in the data. When precise location information was lacking, these situations are denoted by the notation (0°, 0°). Address fields are purposefully limited to the closest hundred block in order to protect people's privacy; precise addresses are not shared. It's critical to understand that the quality of the original records determines how accurate the data is the record. Any issues or complaints about particular data points or the quality of the data are duly acknowledged and can be addressed with comments or inquiries. In addition, a number of crucial procedures were involved in the compilation of the crime dataset, including data collection, inspection, cleaning, and exploratory data analysis (EDA). This procedure included finding and removing unnecessary or redundant columns from the dataset, such as date records and URL references. In addition, the feature extraction technique was used to retrieve pertinent information from the given data. The rigorous preparation of data is of utmost importance in guaranteeing the trustworthiness and relevance of the dataset for the extensive analysis carried out in this study. It provides the fundamental framework around which all other discoveries and understandings are based.

Part 4: Results and Methods

The crime database was collected from the referenced source, and the built-in pandas and numpy methods were used to explain its analytics. Missing values were addressed by taking into account a number of parameters throughout the data cleaning process. In particular, all missing values for Vict Sex were imputed as 'X,' in line with the details on the landing page. The Vict Descent data was filled in with the same 'X' value. Duplicate rows were also eliminated during the data cleaning procedure.
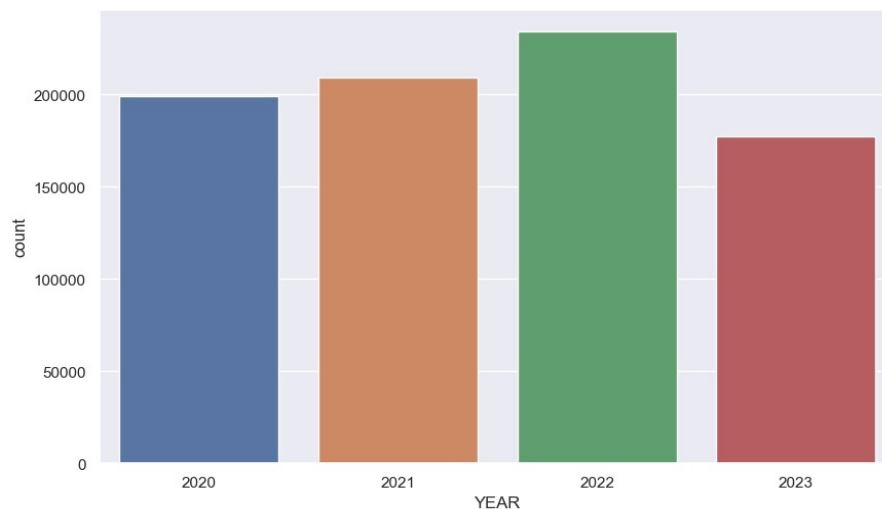

A thorough examination of crime statistics from 2020 to 2023 provides important new insights into the dynamics of criminal activity. First, it shows a rising tendency in crime rates, with 2022 showing the greatest incidence and 2021 and 2020 following closely after. Interestingly, 2020 saw a significant increase in crime at the same time as the COVID-19 pandemic's first emergence. Though it should be noted that the decline in crime rates in 2023 raises some concerns, it's possible that this decline is the result of incomplete data for that year. The fact that crime trends have changed throughout the years is a sign of changing socioeconomic conditions and calls for more research.
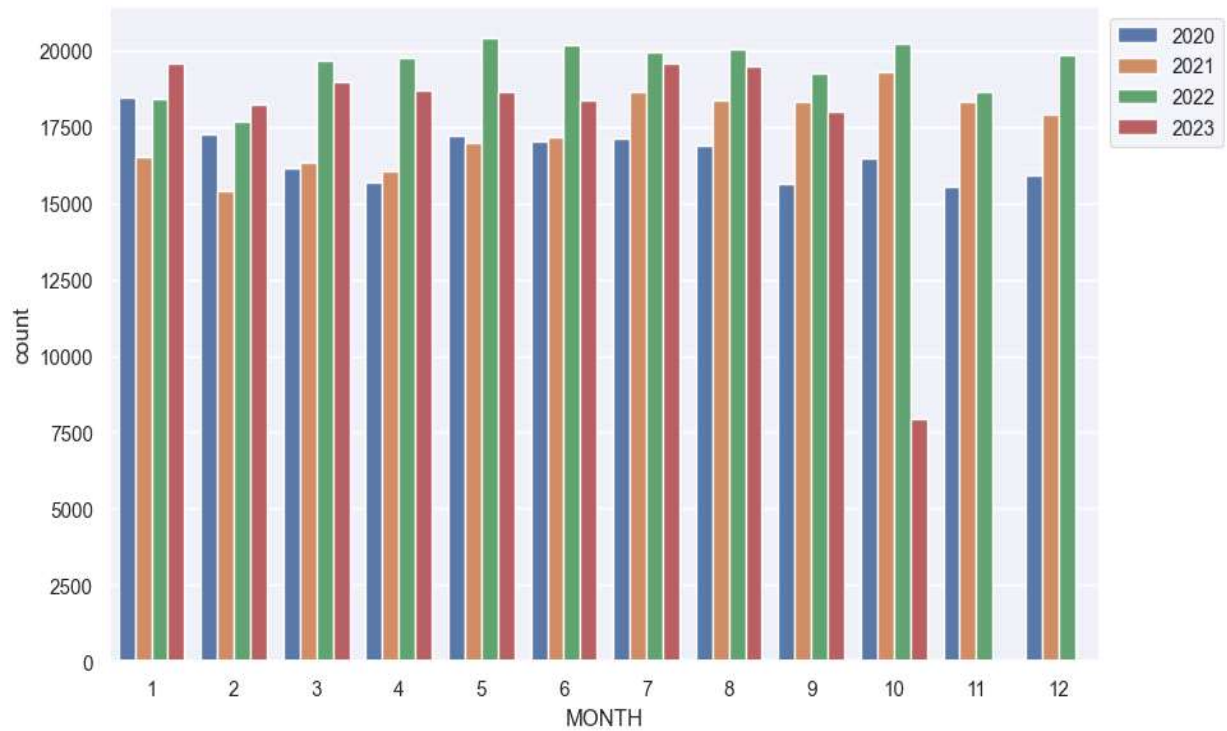
Secondly, a fascinating pattern emerges from the examination of year-by-year crime statistics within months.

With the exception of 1 and 2 of 2023, there is a discernible decline in crime rates for the months for which data for that year is available. This anomaly points to the possibility of a big policy shift taking place in 2023, which would significantly lower global crime rates. To fully comprehend the nature and significance of this policy change and how it affects law enforcement and crime prevention initiatives, more research is necessary.
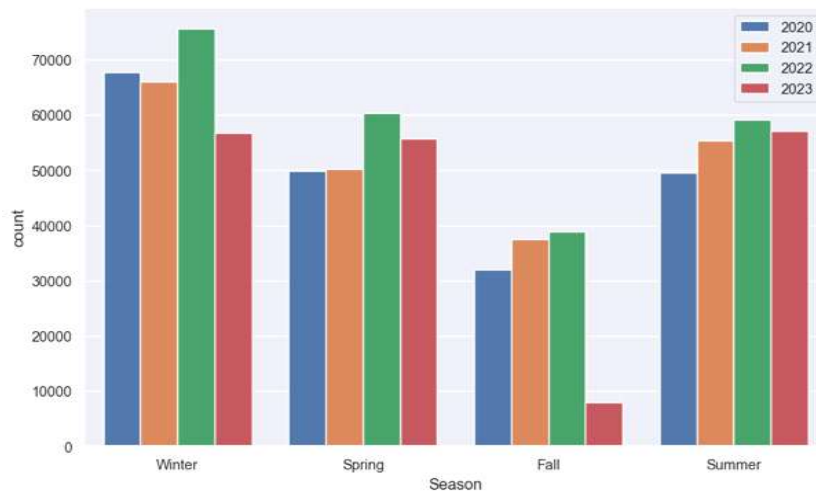
Thirdly, the seasonal differences in crime are emphasized, with the bulk of criminal activity evidently taking place in winter with the midpoint of each season serving as the focal point. The fact that these seasonal tendencies may be influenced by elements like weather, social dynamics, or cultural events highlights the need for more thorough research to identify the underlying causes of these patterns.
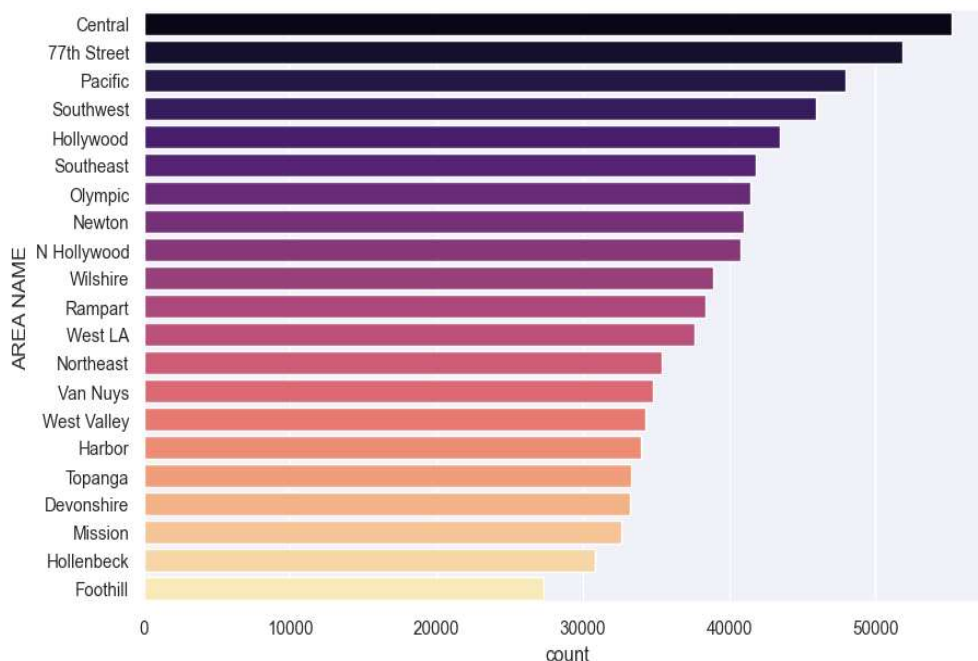
The allocation of crimes by gender and geographic location is covered in the fourth point. Every kind of crime is counted, and the percentages of M, F, and O genders involved in committing these crimes are reported for the various areas appraised. The ability to clearly illustrate the distribution of crimes across different parameters makes these charts indispensable for data visualization of crime statistics. Law enforcement and policymakers can make well-informed judgments by visualizing crime trends and using the reliable statistics comparison they provide.

Crime Rates for LYNCHING - ATTEMPTED During Major Events Over the Years
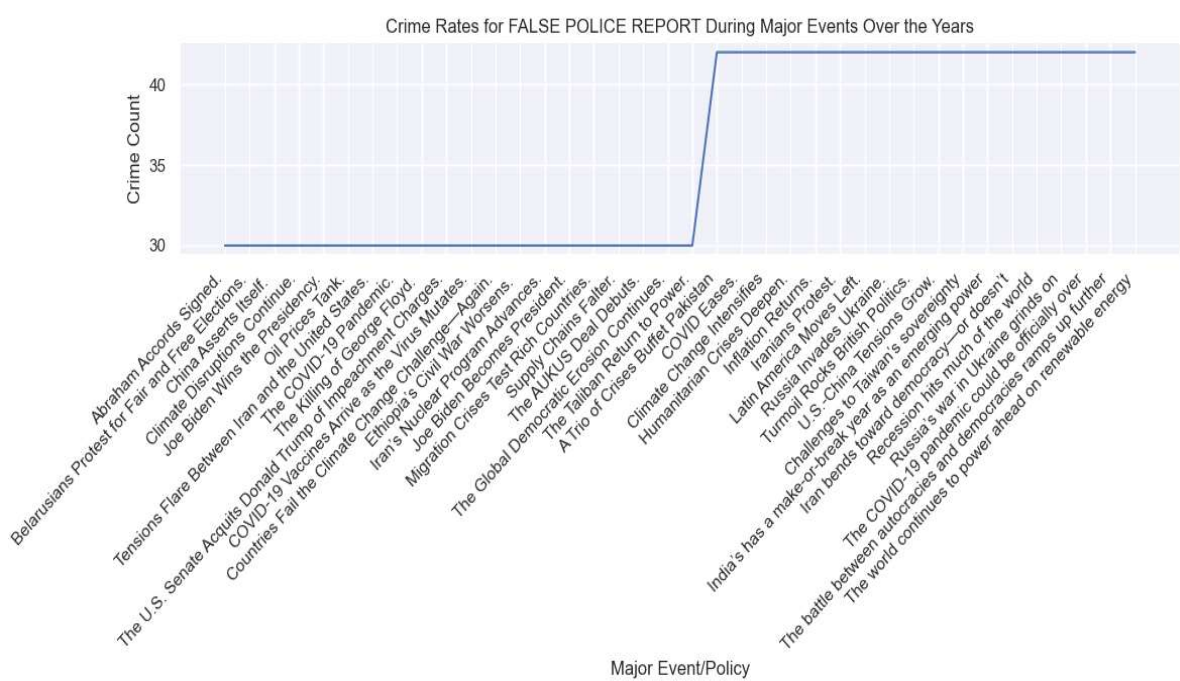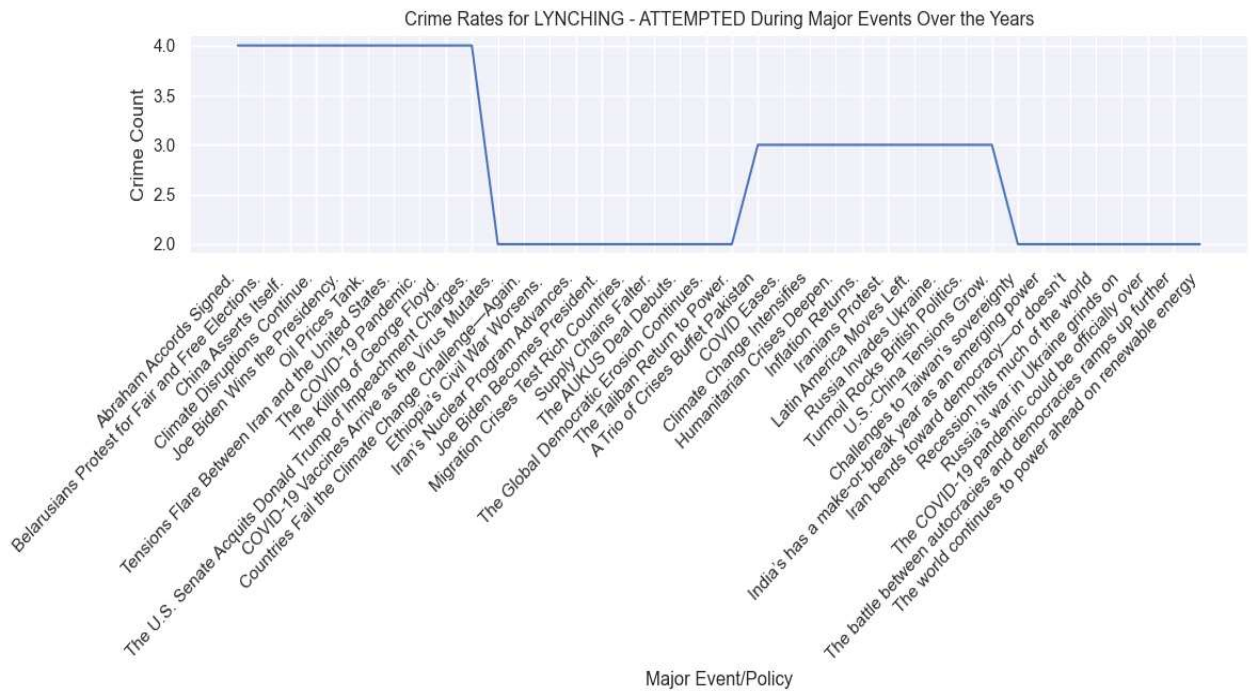
On the fifth, the gender-related mean age of the victims is examined. The analysis indicates that the mean age of males is between 38 and 40, that of females is between 34 and 37, and that of those in the "other" gender category is between 21 and 23. These insights offer useful information regarding the demographic traits of victims across various gender categories.

Time Series Data and ARIMA Predictions
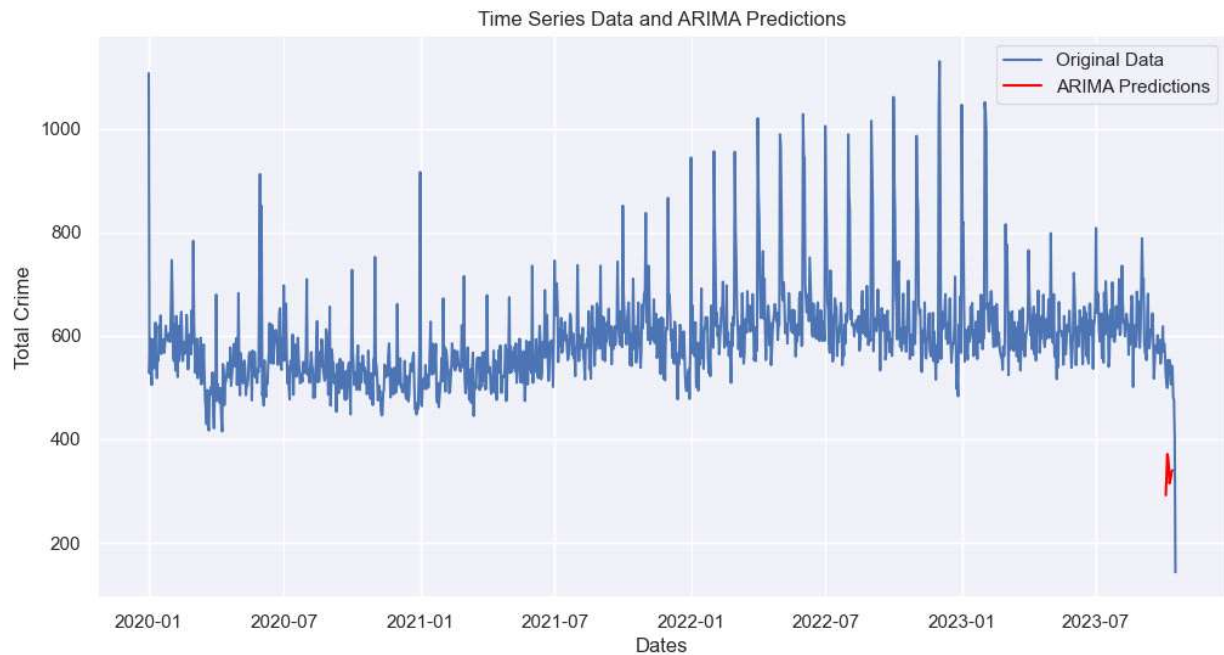
Finally, the sixth point looks at the correlation between the number of crimes and the time of day, expressed in hours. The data indicates that there is a tendency for crime rates to climb in the morning, peak in the afternoon, then progressively decline until midnight before rising once more in the morning. The need of matching law enforcement resources to these temporal trends for efficient crime prevention is highlighted by the diurnal pattern of crime frequency. Because they can influence law enforcement tactics, public policies, and crime prevention initiatives, these findings highlight how crucial it is to comprehend crime trends, demographics, and patterns.

The noted fluctuations in crime rates over time demonstrate the dynamic character of criminal activity and the requirement for flexible, data-driven strategies to address and tackle new issues pertaining to public safety.

Part 5: Limitations and Future Work

Limitations

Our research's biggest limitation is its limited geographic scope, as it is solely based on data from Los Angeles. Compared to other major American cities, Los Angeles may have quite distinct crime dynamics, socioeconomic features, and law enforcement strategies. Therefore, it is possible that our findings might not fully represent the metropolitan environment.Our analysis relies heavily on Los Angeles-specific characteristics, which makes it challenging to extrapolate our results to other locations. It's probable that other cities don't have Los Angeles' distinct economic, cultural, and demographic traits.Biases or anomalies in the Los Angeles dataset could have a significant impact on the reliability and accuracy of our results. It is necessary to thoroughly assess the representativeness and quality of the dataset.

Assurance of Data Quality: Careful steps need to be taken to ensure that the data collected is accurate, consistent, and reliable. Working with law enforcement agencies and subject matter experts may be necessary to handle data.

Comparative Analysis: To identify patterns and distinctions in crime trends, socioeconomic attributes, and law enforcement strategies, comparative analysis must be performed following the collection of data from other cities. This will provide a more comprehensive understanding of the dynamics of urban crime.

Future Work

Expansion of Data: To address the limitation of limited geographic coverage, future research should mainly concentrate on gathering data from numerous significant cities around the United States. Obtaining large datasets for cities such as Chicago, Atlanta, Miami, New York, and others is necessary in order to present a more thorough picture of crime patterns.

Data Diversity: It is imperative to ensure that the dataset is diverse and includes cities with a range of characteristics, including differences in crime rates, population densities, and demographics. This diversity will make the results more reliable and widely applicable.