Crime Data Analysis for Control and Prevention

Introduction

This project focuses on analyzing crime data from 2020 to the present to uncover patterns and trends that can aid in reducing crime rates in specific areas. By examining critical factors such as crime types, locations, and timing, the analysis aims to provide actionable insights. These insights will support law enforcement agencies and policymakers in developing and implementing targeted strategies for effective crime prevention and public safety enhancement.

Source of the data

Data set was obtained from Data.gov (https://catalog.data.gov/dataset/crime-data-from-2020-to-present).

Streamlining the data

The crime_data DataFrame was structured with numerous columns, some of which were deemed unnecessary for the current analysis. To streamline the dataset and focus on the most relevant information, specific columns were removed, simplifying the data and enhancing its usability. The removed columns included general identifiers such as AREA, crime-related codes like Part 1-2, Crm_Cd, and Mocodes, and demographic details such as Vict Descent. Additionally, geographical data (LOCATION, Cross Street, LAT, LON) and case-specific details (TIME OCC, Premis Cd, Weapon Used_Cd, Status) were excluded as they did not contribute directly to the analysis objectives. The resulting cleaned dataset, stored as crime_data_cleaned, retains only the essential columns, ensuring efficiency and clarity for subsequent analysis.

Addressing missing values in the dataset

To ensure data completeness, the crime_data_cleaned.isnull().sum() function is employed to detect missing values across all columns in the crime_data_cleaned DataFrame. This step identifies gaps in the dataset by counting the total null entries in each column, enabling a targeted approach to handling incomplete data.

Insights

The analysis revealed missing values in the following columns:

- **Vict Sex**: Contains 71 missing entries.
- **Weapon Desc**: Has a substantial 861 missing entries.

Replace Missing Values in Vict Sex:

Missing values (NaN) in the Vict Sex column are filled with 'Unidentified' to label cases where the victim's gender is not recorded.

Replace Missing Values in Weapon Desc:

Missing values (NaN) in the Weapon Desc column are filled with 'None' to indicate no weapon description is provided or involved.

Addressing these gaps is critical for maintaining the integrity of the analysis. Depending on the project's objectives, appropriate strategies such as imputation with relevant values, replacing with default placeholders, or excluding rows with missing data can be applied to mitigate the impact on results and ensure reliable insights.

Standardizing the date column

The Date Rptd column in the dataset contains dates in varying formats, which can hinder consistent analysis. To address this, a function was defined to clean and standardize the dates. The function resolves inconsistencies such as different separators (slashes vs. hyphens), removes extraneous details like "AM/PM" or trailing "00:00," and converts each date into a uniform format. Once the function was applied, a new column, Date Rptd Standardized, was created, ensuring consistency across all date entries. This standardization simplifies subsequent analysis and ensures compatibility with time-based queries. A preview of the original and standardized columns was displayed to verify the success of the conversion.

Define the Refined Function:

- A function (standardize_date_refined) is defined to handle various date formats, removing unwanted parts (like 00:00 or AM/PM) and trying multiple formats for parsing.

Apply the Function:

- The apply method is used to apply this function to the Date Rptd column in the crime_data DataFrame.

Create a New Column:

- The standardized dates are stored in a new column, Date Rptd Standardized, within the crime_data_cleaned DataFrame.

The original Date Rptd and DATE OCC columns is removed from the crime_data_cleaned DataFrame after the dates were standardized into a new column. This step ensures consistency and clarity in the dataset while streamlining it for analysis. The updated DataFrame is previewed to confirm the changes.

Creating Separate Columns for Month, Day, and Year

The Date Rptd Standardized column is refined by splitting it into separate **Month**, **Day**, and **Year** columns, enabling more detailed time-based analysis. The column was first converted to a datetime format to facilitate the extraction of these components using .dt.month, .dt.day, and .dt.year. Additionally, a Month_Str column was created to display month names (e.g., January, February) for improved readability in visualizations and reports. Afterward, the original Date Rptd Standardized column was dropped to maintain a clean and focused dataset. The updated DataFrame was reviewed to ensure the transformation was successful.

Identify the most common types of crime in each area

This analysis is to identify the most common crime types in each area to uncover specific crime patterns. This information is vital for authorities and stakeholders to allocate resources effectively and implement targeted measures addressing high-risk areas. To achieve this, we employed a multi-step process that includes data grouping, ranking, and visualization. First, we grouped the data by AREA NAME and Crm Cd Desc (crime description) to count the frequency of each crime type per area. This granular breakdown highlights not only the overall crime trends but also the most prevalent crime types in specific locations. Next, we sorted the data by AREA NAME and frequency in descending order, selecting the highest-frequency crime for each area. This ranking simplifies the findings, allowing us to focus on actionable insights rather than overwhelming details. The data was then prepared for visualization by creating a pivot table, ensuring it is organized for clear presentation. Finally, we visualized the results using a bar chart, which highlights the most common crime type and its count for each area. This approach provides an intuitive and accessible representation of the data, enabling stakeholders to quickly interpret trends and prioritize interventions. By focusing on the most frequent crimes, the analysis reduces noise and enhances clarity, supporting targeted, data-driven decision-making.

Visualizing the Most Common Crimes by Area

This bar chart showcases the **most common crime type in each area**, with the height of each bar representing its frequency. A gradient color scale is applied to the bars, where **lighter colors** indicate areas with lower crime counts and **darker colors** signify areas with higher crime counts. A color bar alongside the chart provides a reference for interpreting the intensity of crime counts across areas. This visualization offers a clear and concise comparison of crime patterns, helping to quickly identify regions with the highest frequency of specific crimes and enabling targeted decision-making.

Group and Count Crimes:

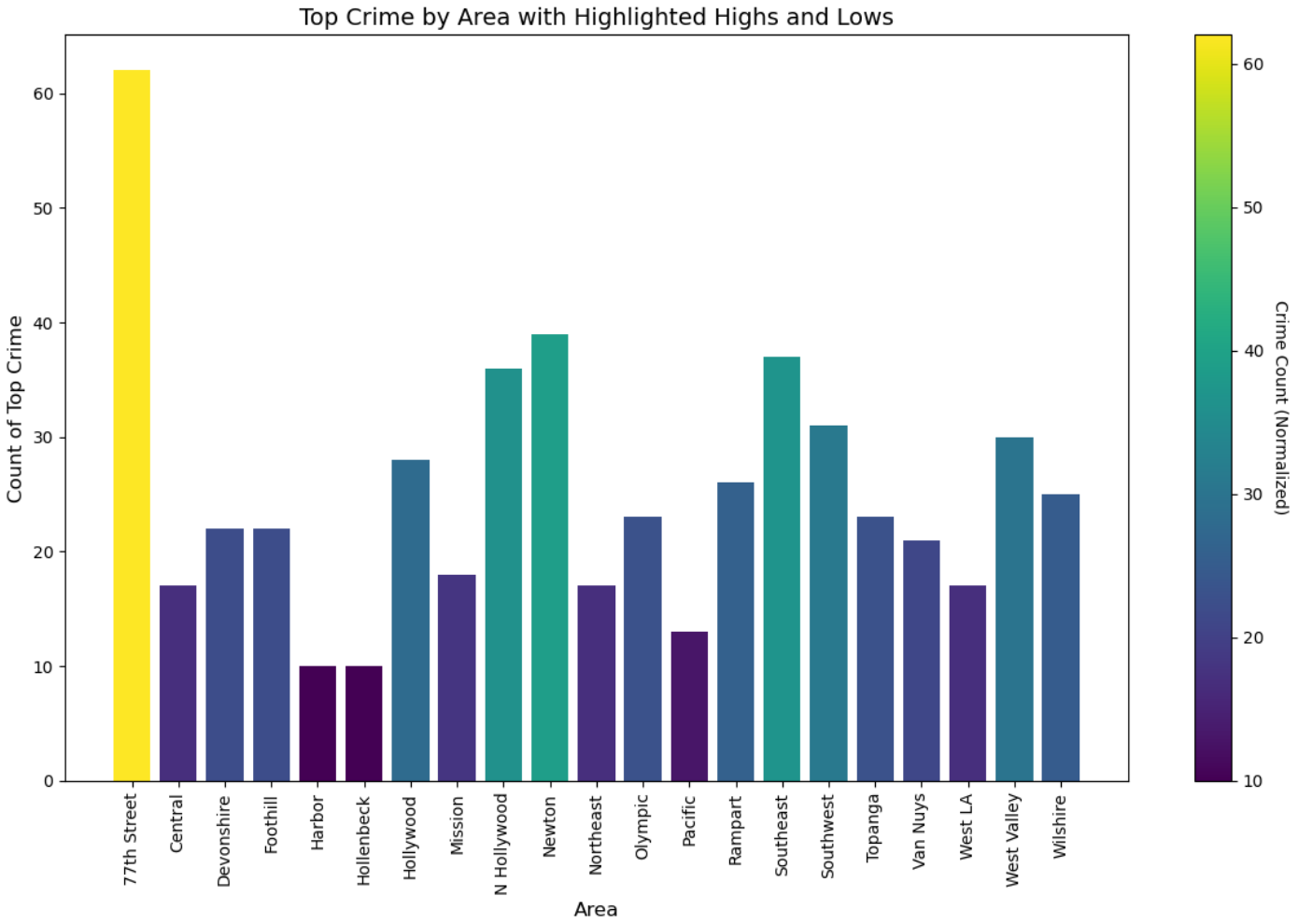
- Group the dataset by area name and crime description.
- Count the number of occurrences for each crime type in each area.

Sort and Select Top Crimes:

- Sort the grouped data by area and the count of crimes in descending order.
- Select the top crime(s) for each area.

Reshape for Visualization:

- Pivot the data so that areas are rows, crime types are columns, and their counts are the values.



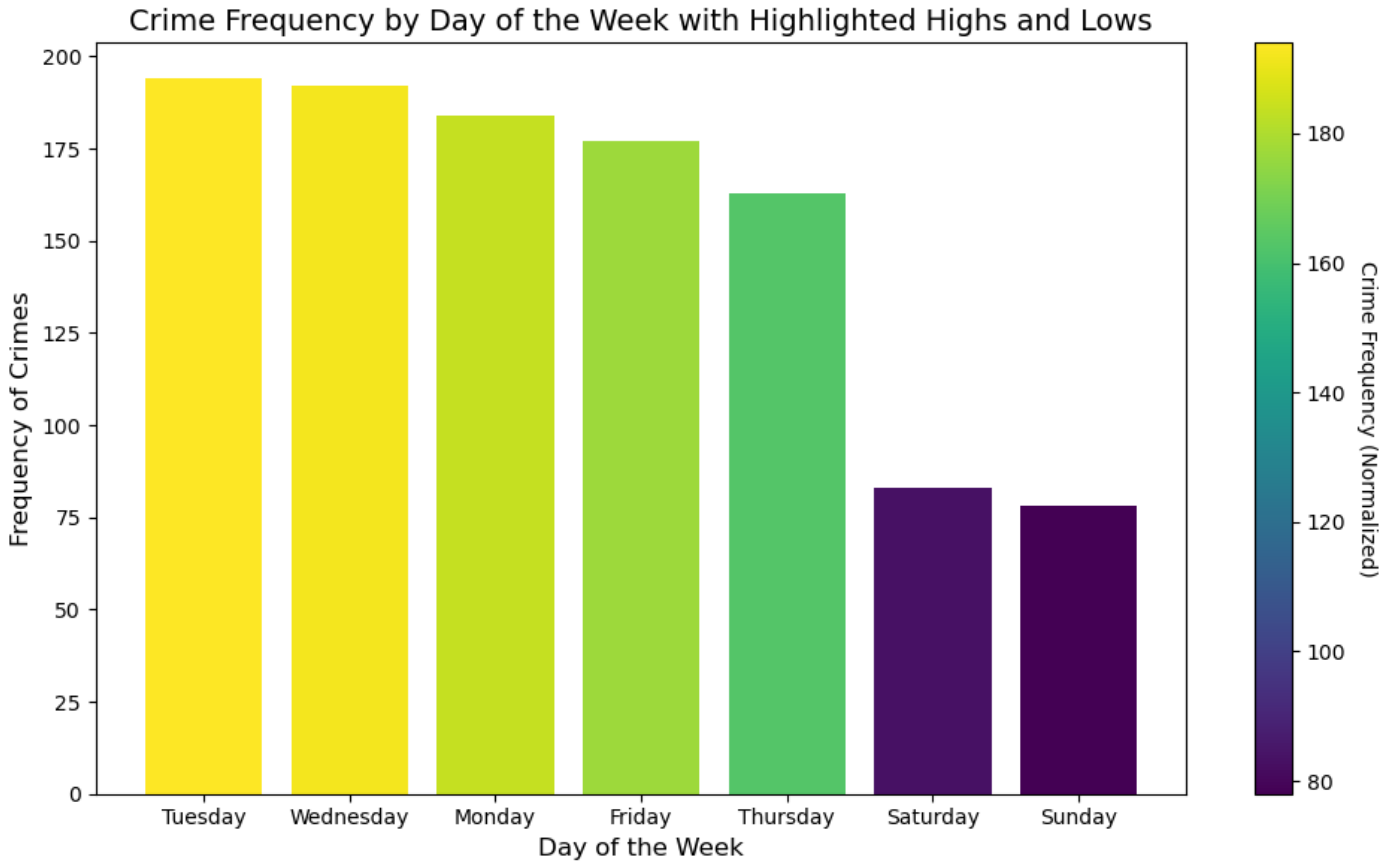
When crimes happen most frequently

This analysis is to uncover patterns in crime occurrences by examining day of the week, providing crucial insights for planning targeted measures and effectively allocating resources during high-crime periods. Analyzing crime timing allows law enforcement and community stakeholders to identify peak hours and days for heightened vigilance and intervention. For instance, recognizing specific time frames such as late nights or weekends with higher crime incidents enables strategic planning and optimized resource deployment to address crime proactively.

To achieve this, the analysis is structured into logical steps to ensure clarity and accuracy. First, crimes are grouped by Day of the Week, and their occurrences are counted to highlight the days with the highest frequency. This step identifies high-risk days, helping to prioritize when to intensify patrols or community initiatives. Next, a bar chart is used to visualize the frequency of crimes across the week, offering a

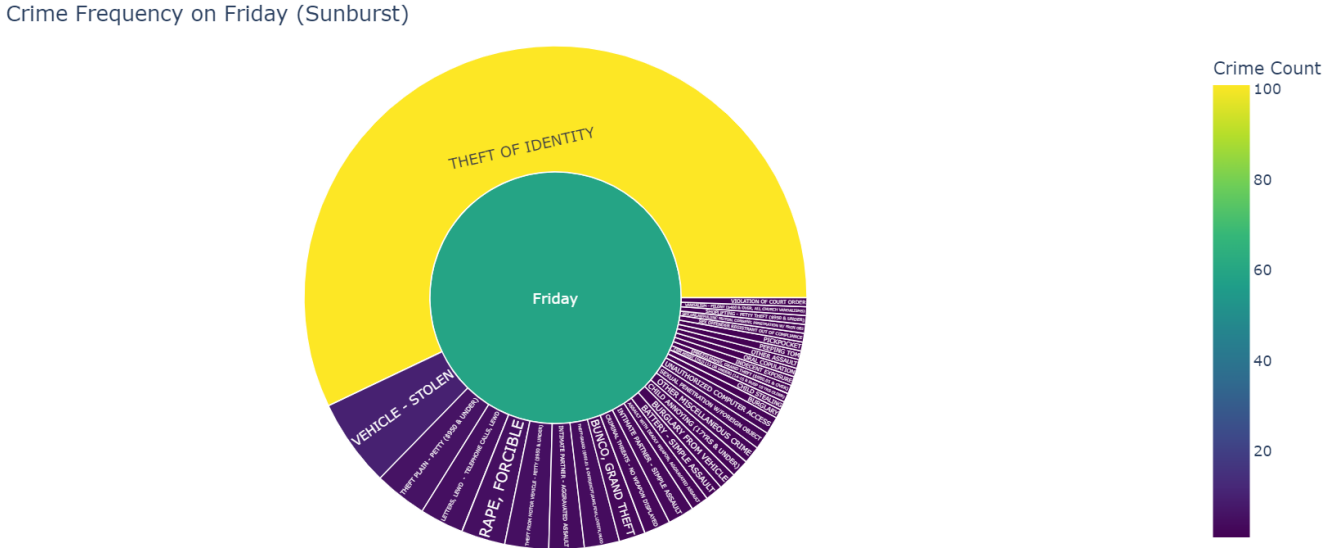
straightforward comparison of crime rates. The bar chart's intuitive format makes it accessible to both technical and non-technical audiences, facilitating informed decision-making.

This chart provides a comprehensive view of crime frequencies across the days of the week, offering actionable insights for decision-making. Each bar represents the number of crimes reported on a given day, with a color gradient emphasizing variations—darker bars indicate high-crime days, while lighter bars represent lower activity. Key annotations highlight the day with the highest crime frequency, marking peak activity, and the day with the lowest frequency, signaling the least active day for crime. The chart reveals distinct patterns, such as peaks on specific days potentially linked to work schedules, holidays, or social trends, and dips on other days, like weekends, reflecting reduced public engagement. These insights enable law enforcement to allocate resources strategically, increasing patrols on high-crime days and optimizing efforts during low-crime periods. Additionally, the trends may correlate with broader social behaviors or events, offering opportunities for further analysis. This visualization simplifies complex data, making it accessible for non-technical audiences and empowering stakeholders to implement targeted interventions effectively.



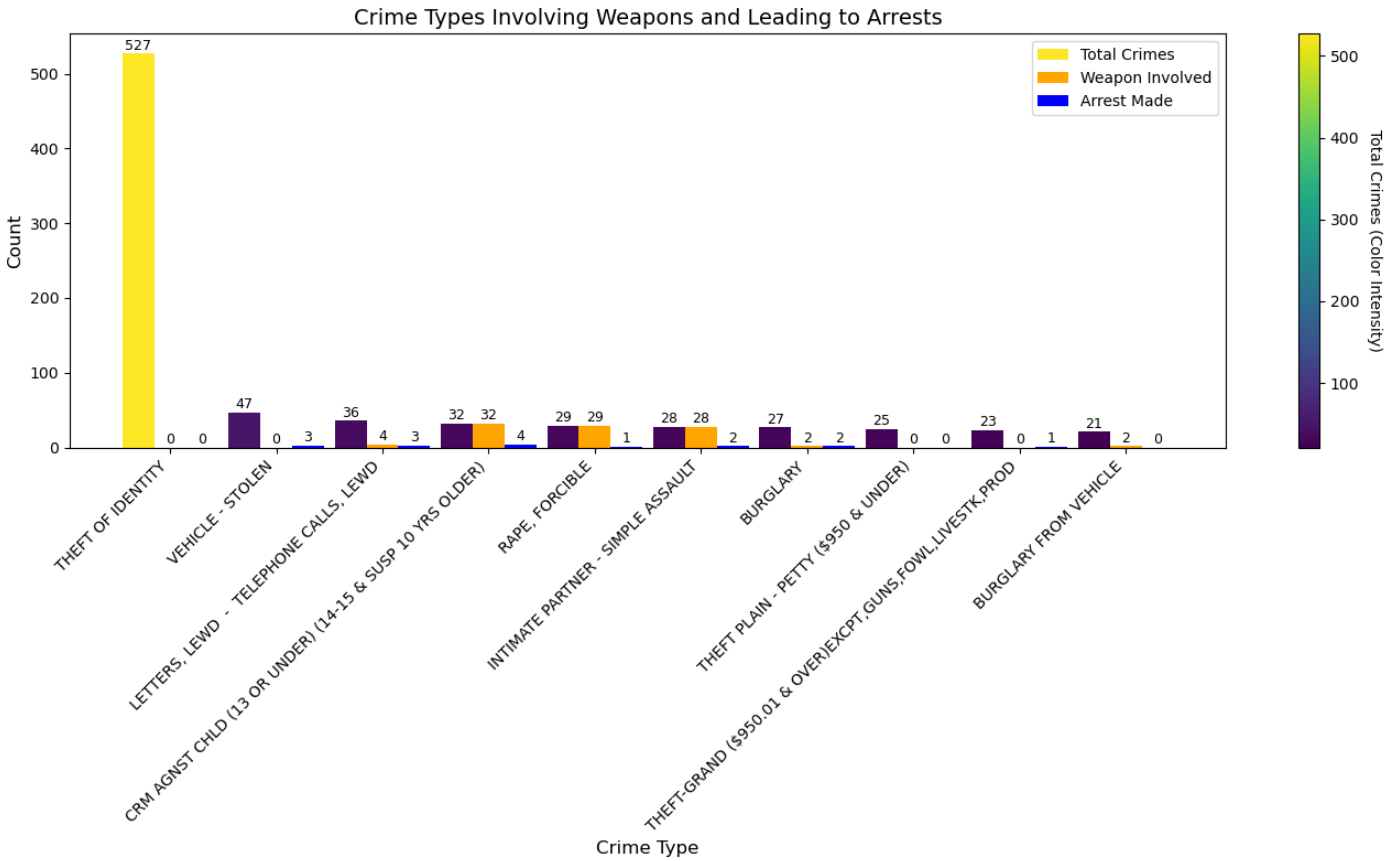
Crime Frequency by Day and Type

The **sunburst chart** provides a detailed breakdown of crime data, illustrating patterns of criminal activity across days of the week and types of crimes. The central segments represent the days, with the outer layers showing the frequency of specific crime types on each day. This visualization highlights dominant crime days, with larger central segments indicating the days with the highest overall crime rates. It also reveals patterns in crime types, such as certain crimes dominating specific days, while others are more evenly distributed throughout the week. Additionally, the chart supports resource allocation by identifying days with varied crime types, suggesting a need for broader coverage, and days with fewer or more concentrated crimes, allowing for targeted interventions. This comprehensive view not only uncovers overarching trends but also provides actionable insights to guide strategic resource deployment and crime prevention efforts.



Crime Types Involving Weapons and Arrests

This bar chart highlights the top 10 crime types, detailing their frequency and providing insights into the involvement of weapons and arrests. The chart reveals that **Theft of Identity** is the most frequent crime, with over 500 reported cases, yet it involves neither weapons nor arrests, indicating its non-violent nature. In contrast, violent crimes such as **Rape (Forcible)** and **Intimate Partner - Simple Assault** show significant weapon involvement, emphasizing their potential danger. Crimes like **Vehicle Theft** result in higher arrest rates, suggesting effective law enforcement, while others, such as **Burglary**, have fewer arrests, which could point to challenges in investigation or apprehension. These patterns have policy implications, with high-frequency but low-arrest crimes like **Theft of Identity** requiring preventative measures, and high-arrest crimes showcasing enforcement strategies that could be adapted to other areas. This visualization provides a comprehensive perspective, enabling policymakers and law enforcement to prioritize resources and refine strategies for crime prevention and control.



Treemap of Weapon-Related Crimes and Arrests

Treemap visualization offers a detailed view of crimes involving weapons, highlighting their frequency and corresponding arrest rates. Each block represents a specific crime type, with the block size proportional to the number of weapon-related incidents. A color gradient is used to indicate the number of arrests, where **darker blocks** represent crime types with higher arrest rates. Hovering over a block reveals additional details, such as the total number of crimes, weapon-related incidents, and arrests, providing deeper context.

The largest blocks represent the most prevalent weapon-related crimes, such as **Rape** or **Aggravated Assault**, reflecting the violent nature of these offenses. Darker-colored blocks indicate crime types with effective law enforcement responses, while lighter blocks highlight areas with fewer arrests, suggesting enforcement challenges or investigative difficulties. This visualization underscores the need for targeted interventions in crimes with high weapon involvement but low arrest rates and highlights successful strategies that could be applied to other crime categories. The treemap provides a comprehensive perspective, linking weapon-related crimes to enforcement outcomes, enabling policymakers and law enforcement to refine strategies and allocate resources effectively.

Future Steps

To further enhance the Crime Data Analysis project, several advanced steps can be undertaken. Implementing predictive modeling using machine learning techniques, such as logistic regression or decision trees, could help forecast high-risk areas and times based on historical data. Geospatial analysis using GIS tools can identify crime hotspots and their correlation with environmental factors, while integrating socio-economic indicators like unemployment rates and education levels could uncover underlying causes of crime. Developing an interactive dashboard would enable real-time monitoring and visualization of crime trends for law enforcement and policymakers. Additionally, analyzing demographic patterns and the profiles of victims and perpetrators can provide targeted insights for crime prevention. Ethical considerations, such as bias evaluation and mitigation, should be integral to ensure fairness and transparency in the analysis. Finally, sharing simplified visualizations and actionable strategies with community stakeholders would promote transparency and engagement, ultimately enhancing the effectiveness of crime prevention measures. These steps collectively pave the way for a comprehensive, data-driven approach to crime control and public safety.