Dashboards

US Crime Trends and Analysis

Abstract

This report is based on the interactive dashboard that analyses crime trends across the United States, providing insights into crime patterns by location, demographics, weapon usage, and time. The design emphasizes clarity, user engagement, and dynamic filtering to facilitate data-driven decision-making for law enforcement and policymakers

Group 4: Aitana Hierro Garzon, Khushboo Surana, Tanya Jain

Introduction

Crime rates have long been a social concern, and understanding the underlying patterns can help authorities allocate resources efficiently. This project aims to identify trends and patterns in crime occurrences across cities and identify insights based on demographic factors. Our project covers the Crime Dataset, which contains approx. 638,454 records (more than 10,000 crime records), covering various crime categories, locations, and dates.

Objective

Crime rates have long been a critical societal issue, and understanding patterns and trends in crime occurrences can significantly assist law enforcement agencies in allocating resources more effectively. This project focuses on analyzing a dataset containing crime records for murder to uncover insights related to crime categories, locations, timelines, and demographic factors.

Part 1: Data Cleaning

For the purpose of this project, we utilized a dataset sourced from Kaggle link here and performed data cleaning using Python in Jupyter Notebook. This section outlines the data cleaning techniques used to ensure consistency, handle missing values, and standardize the dataset for effective analysis.

The original dataset contained 638,454 rows with no explicit null values, but several columns had entries labeled as “Unknown.” These were initially converted to ‘Nan’ for cleaning. We then dropped columns such as 'Victim Ethnicity,' 'Perpetrator Sex,' 'Perpetrator Race,' and 'Perpetrator Ethnicity' due to the high volume of missing values. Rows with null values in critical columns like 'Agency Name,' 'Victim Race,' 'Victim Sex,' and 'Weapon' were also removed. Afterward, remaining nulls in other columns were reverted to "Unknown" where they held meaningful information.

The final cleaned dataset contains 599,679 rows with several columns that provide specific information about each crime record. Record ID serves as the unique identifier, while Agency Code, Agency Name, and Agency Type specify the reporting agency. The City and State columns indicate the location of the incident, with Year and Month marking the time it occurred. The

Incident column details the incident itself, and Crime Type categorizes the type of crime, with Crime Solved indicating whether it was resolved. Victim details include Victim Sex, Victim Age, and Victim Race, while the Relationship column defines the victim-perpetrator connection. The Weapon column lists the weapon used, with Victim Count and Perpetrator Count indicating the number of victims and perpetrators involved, respectively. Record Source identifies the data source, and Perpetrator Age records the age of the perpetrator.

Part 2: Key Data Fields and Structure

After completing the data cleaning process, we refined our focus to specific areas of analysis that would yield the most valuable insights. Below are the key areas and the corresponding fields used for each:

- **Geographical Trends:**
  To analyze crime distribution across locations, we used fields like *City*, *State*, and *Record ID* (count), allowing us to map and quantify crime occurrences at the state and city level.
- **Demographic Trends (Victim and Perpetrator):**
  We explored demographic patterns for both victims and perpetrators:

i. **Perpetrator by Age and Relationship:** Using *Perpetrator Count*, *Perpetrator Age (bin)*, and *Relationship (group)*, we examined the age distribution of perpetrators and their relationship to victims (e.g., domestic or stranger).

ii. **Victim by Age, Race, and Sex:** For victims, we focused on *Victim Sex*, *Victim Age (bin)*, *Victim Race*, and *Victim Count*, to understand how demographics vary across different types of crime.

- **Weapons Trends:**
  The fields *Weapon*, *Crime Solved*, and *Record ID* (count) helped us analyze the types of weapons used in crimes and the solvability of those crimes.
- **Time-Series Trends:**
  To track how crime trends evolved over time, we used *Year*, *Record ID* (count), and *Victim Sex* to observe changes in crime rates and gender distribution across different years.

Note: The field *Crime Type* was unused in this analysis, as the dataset primarily focused on two types of crime: involuntary manslaughter and murder, limiting the variability in crime categories.

This selection of fields allowed us to perform a focused and meaningful analysis of the dataset, yielding insights into geographical, demographic, weapon usage, and time-related crime trends.

Part 3: Worksheet Creation

To extract meaningful insights from the dataset, the following focused worksheets were developed:

1. **Map Visualization**: A geographical representation displaying crime distribution across states, with tooltip details on specific cities to enhance data exploration.
2. **Crime by City**: A detailed view of crime occurrences by city, which integrates with the dashboard map for interactive analysis.
3. **Perpetrator Relationship**: This worksheet examines the relationship between perpetrators and victims, providing insights into crime dynamics.
4. **Victim Demographics**: An analysis of victim demographics, including age, race, and gender, to better understand the population affected by crime.
5. **Over Years**: A trend analysis of crime rates over time, highlighting changes and patterns in criminal activity.
6. **Race Distribution**: A pie chart illustrating the racial distribution of crime victims, with a breakdown of percentages.
7. **Weapons Used**: A bar chart displaying the most frequently used weapons in crimes, with handguns leading the list, followed by other firearms and knives.
8. **Crime Solvability**: A toggle option that indicates whether crimes in various regions and times were solved, aiding in the assessment of law enforcement efficiency.

Part 4: Dashboard Creation

The dashboard provides a comprehensive analysis of crime trends and patterns across the United States, offering valuable insights for law enforcement, policymakers, and researchers. This report highlights its key features, design strategies, and steps taken to create a highly interactive and informative visualization tool.

With a focus on visualizing crime data across dimensions like time, location, demographics, and weapon usage, the dashboard covers multiple years of data, including details on victims, perpetrators, relationships, and crime outcomes.

Design Considerations:

1. **Consistent Color Schemes**

   Colors were chosen carefully to distinguish different categories and data points clearly, enhancing readability. The color palette was selected to align with data context.

2. **Interactive Filters and Parameters**

   Filters allow users to dynamically refine visualizations based on criteria like gender, weapon type, crime resolution, year, and age group. Parameters such as the "Top N Weapons" selector enable customization of chart details, adapting the dashboard to different analysis needs. Additionally, interconnected filters ensure that selections in one chart automatically update others, providing a seamless and synchronized analysis experience.

3. **Tooltips and Annotations**

   Interactive tooltips provide additional contextual information when users hover over data points, including exact values, percentages, and related details. This helps to minimize chart clutter while still providing in-depth information.

4. **Information Prioritization**

   Efforts were focused on presenting data clearly by prioritizing clarity over clutter, ensuring charts and tables convey key metrics without overwhelming users. A hierarchical organization

was implemented, placing essential charts at the top and supplementary details below, allowing users to quickly grasp critical insights.

Part 5: Story and Insights

1. **High-Crime States:** California, New York, Texas, and Florida have the highest crime rates, likely due to their large populations and urban centers. These states may require targeted crime prevention strategies.
2. **Weapon Usage:** Handguns are the most frequently used weapon in crimes, followed by other firearms. This emphasizes the ongoing issue of gun violence in the U.S.
3. **Age Distribution:** Both perpetrators and victims are predominantly aged between 10-40 years, with a peak in the 20s. This suggests that young adults are at the highest risk of being involved in violent crimes, either as victims or perpetrators.
4. **Unsolved Crimes:** The 'Unknown' field on the perpetrator -victim relationship often indicates unsolved crimes, highlighting inefficiencies in crime-solving and the need for improved law enforcement resources.
5. **Domestic Violence and Female Victims:** Filtering for female victims reveals a higher proportion of domestic-related crimes, underlining the seriousness of domestic violence, particularly against women.
6. **Racial Disparities:** Black victims are more common in southern states, while the northwestern region shows fewer Black victims, indicating regional racial disparities in crime victimization.
7. **Gender Differences:** Men are more often victims of murder and involuntary manslaughter in general, while women are more frequently victims of domestic violence.

After analyzing the dataset, several critical insights emerge from the crime trends across the U.S., specifically related to the geographic distribution, demographic factors, weapon usage, and victim-perpetrator relationships. These findings highlight various patterns that can inform crime prevention strategies and policy-making decisions.