International Journal of Engineering Applied Sciences and Technology, 2025
Vol. 10, Issue 01, ISSN No. 2455-2143, Pages 43-54
Published Online May 2025 in IJEAST (http://www.ijeast.com)

CRIME DATA ANALYSIS

Potharaju Srihitha, Jillapalli Nikitha, Gaikwad Sneha, Dr Diana Moses
UG Scholar, Department of Artificial Intelligence and Data Science,
Methodist College of Engineering and Technology, Hyderabad, India

Dr. P Lavanya, Dr. Diana Moses
Professor, Department of Computer Science and Engineering,
Methodist College of Engineering and Technology, Hyderabad, India

*Abstract—***This dataset holds crime rates of many U.S. cities and states along with precise information on violent and property crimes and demographic data including population counts. The dataset has a date-stamped record for every entry so that one can analyze across different time frames, with the time attribute starting from the year 2022 and continuing through at least 2025. The central emphasis lies with the frequency of certain categories of crimes—violent crimes (murder, rape, robbery, aggravated assault) and property crimes (burglary, larceny, motor vehicle theft)—and the respective population figures, allowing for per capita crime rate calculation. The dataset consists of 305 rows and 16 columns, with all rows containing usable data, although some values may require formatting or standardization. From the data, we incurred two key points: first, that crime entries span across a meaningful recent timeframe (2022–2025), enabling temporal trend analysis; and second, that the dataset is relatively clean in structure but still benefits from light preprocessing. This dataset is highly beneficial to various stakeholders, such as local law enforcement agencies, government offices, criminologists, and the press, all of whom can leverage the information to monitor crime trends, manage resources optimally, and frame educated policies for public safety. The dataset also has the potential for researching the socio-geographic dynamics of crime, urban development strategies, and aiding community action or awareness programs.**

I. INTRODUCTION

Literature survey:
Nwankwo et al., This project analyzes crime data from Chicago using tools like Excel, Tableau, and Python to uncover patterns and correlations with socioeconomic factors, aiming to aid law enforcement in crime prevention. Findings show that theft is the most common crime, with peak activity occurring on Fridays and between 6 PM and midnight, suggesting targeted police allocation [1].
Garg et al., compares R and Tableau for big data analytics, highlighting Tableau's efficiency, user-friendliness, and speed in handling large datasets. The study concludes that Tableau outperforms R in big data analytics, though additional features were not explored due to time constraints [2].
Gaur conducted a survey on public priorities reveals a strong concern for reducing crime, despite the FBI data showing low crime rates. This discrepancy highlights the difference between public perception and actual crime statistics, suggesting that people's concerns may not always align with reality [3].
Silva Atencio et al., uses statistical models, including linear regression, to predict crime trends in Costa Rica, highlighting the Gran Area Metropolitana's increasing crime risk. It emphasizes the need for government transparency, regular data publication, and proactive community and police efforts to prevent future crimes [4].
Ajagbe et al., uses the Folium data visualization tool to track and display Fulani herdsmen attacks in Nigeria, providing detailed crime data on a Chloropleth map. It demonstrates the advantages of Folium in visualizing crime patterns and offering more precise insights compared to traditional methods [5].
Kane, Robert et al., analyzes the impact of weather conditions on crime in Chicago between 2008 and 2017, using data analysis and forecasting methods like Holt-Winters exponential smoothing. The study confirms a strong positive correlation between weather and crime, with findings suggesting that weather-based forecasts could aid law enforcement in resource allocation and prevention strategies [6].
Nair, Lekha et al., compares the effectiveness of D3.js and Tableau for visualizing large datasets like the Chicago crime data, highlighting how visualization uncovers hidden patterns. It concludes that no single tool is perfect for all Big Data visualization needs, and selection depends on factors like cost, interactivity, and dataset features [7].
Sharma, Sharad et al., compares crime data from Chicago, Baltimore, Dallas, and Denton, analyzing crime rates in relation to population, unemployment, and poverty. It reveals that crime patterns are influenced by factors like population density and commercial areas, helping law enforcement better allocate resources and understand crime dynamics [8].
Liu et al., develops a geospatial crime dashboard prototype to aid crime analysis and prediction, helping policymakers

visualize crime patterns and relationships. It evaluates existing crime dashboards, revealing a gap in utilizing GIS analysis for deeper crime event understanding [9].

Akhtar et al., reviews Tableau's role in analyzing and visualizing COVID-19 data, emphasizing its ability to create intuitive dashboards and predictive models for decision-making. Tableau's features like data blending, real-time reporting, and mapping help users explore and understand data, especially in the context of the ongoing pandemic [10].

Feng, Mingchen, et al., applies big data analytics and visualization techniques to crime data from San Francisco, Chicago, and Philadelphia, uncovering patterns and predicting trends. Advanced models like Prophet and stateful LSTM outperformed traditional neural networks, with three years of training data identified as optimal for accurate forecasting. The findings offer valuable insights for law enforcement to enhance resource allocation, crime prevention, and decision-making. Future work aims to develop a scalable analytics platform with multivariate visualization, graph mining, and fine-grained spatial analysis capabilities [11].

Agarwal, Jyoti, et al., presents a crime analysis approach using k-means clustering on crime data with the RapidMiner tool, aiming to support law enforcement through data-driven insights. By focusing on homicide rates from 1990 to 2011, the analysis reveals a decreasing trend in such crimes over the years. The clustered results effectively highlight crime patterns, making it easier to understand historical trends and develop precautionary strategies for the future [12].

Wang, Hongjian, et al., study addresses neighborhood-level crime rate inference by integrating modern urban data, specifically Point-Of-Interest (POI) and taxi flow data, in Chicago. Compared to traditional demographic and geographic features, these new features significantly improve prediction accuracy, with taxi flows modeling social interaction and POIs reflecting regional functionality. Using a negative binomial regression model, the approach achieves a 17.6% reduction in prediction error and proves consistent across multiple years [13].

Researchers have applied data mining and big data techniques to uncover crime patterns and predict trends. Feng et al. used LSTM and Prophet models on U.S. city crime data, achieving accurate forecasting. Nwankwo et al. analyzed Chicago crimes, linking them to socioeconomic factors for better police resource allocation.[14]

Researchers use data mining techniques like Naïve Bayes, Decision Trees, and Apriori to detect crime patterns. The "Series Finder" project by Cambridge Police showed over 80% accuracy in burglary pattern recognition.NER and Coreference Resolution improve crime data extraction.[15]

II. MATERIALS & METHODS

DATASETS: The dataset contains **crime statistics** for various U.S. cities, detailing offenses like murder, robbery, and property crimes along with population, date, and time. It aids in analyzing crime trends, guiding law enforcement, and supporting public safety planning. The dataset here includes crime statistics from different U.S. cities and states. Every entry reflects a particular location with a timestamp, including date and time, which provides background information on when the data was taken. The main emphasis of the dataset is on crime rates, including both violent and property offenses. Violent offenses are calculated into subtypes like murder, rape, robbery, and aggravated assault, whereas property offenses are burglary, larceny, and theft of vehicles. Property crimes are also present along with the population data for each city, calculated per capita presumably as a benchmark for calculating per-capita crime rates.

Structurally, the dataset has 16 columns and 1130 rows. Nonetheless, only roughly 305 rows are filled with genuine data, and the remaining ones are either empty or placeholders. There are geographic identifiers (`states`, `cities`), demographic data (`population`), some crime measures, and metadata such as `date`, `time`, and a numeric `area` code. One column (`Unnamed: 12`) is entirely empty and can be safely deleted for data cleaning. Notably, several of the numeric columns are string stored and potentially need to be converted to proper formats for analysis.

The stakeholders of this dataset can be quite varied. Police departments at local levels or even the FBI can utilize this data to track crime trends and decide how to allocate resources. Local governments and city planners can utilize the statistics to inform urban development and enhance public safety. Researchers and criminologists could undertake an analysis of the dataset to identify patterns, causes or correlations in crime across different geographic or temporal contexts. Policy makers can utilize such information to develop regulations and crime prevention policy. Journalists and media sources would also be able to utilize the dataset in reporting on issues of public safety. The general public or activist groups can also use information from the dataset to inform community action or awareness campaigns.

This dataset holds crime statistics across many U.S. cities and states, with detailed records on violent as well as property crimes, along with demographic information like population counts. The dataset holds a time-stamped record for every entry such that the data can be analyzed over different time segments. The key emphasis is placed on the rates of certain crime types—violent crimes (such as murder, rape, robbery, and aggravated assault) and property crimes (burglary, larceny, and motor vehicle theft)—along with the respective population figures, which allow per capita crime rate estimations. In spite of the 16 columns and 1130 rows, valid data appear in only around 305 rows, the others being either empty or dummy data, making it necessary to clean the data before analysis.

Software: Tableau

Tableau is a powerful data visualization and business intelligence tool designed to help users analyze, interpret, and visualize data. Its primary purpose is to transform raw data into interactive dashboards, charts, and graphs that are easy to understand and interpret. Tableau's key features include a drag-and-drop interface, real-time data analysis, forecasting, and advanced analytics such as trend identification and clustering. It also supports data blending, allowing users to combine data from various sources for more comprehensive analysis. With its ability to map geographic data, Tableau is highly effective for geospatial analysis, providing location-based insights.

The application of Tableau spans across a wide range of industries, aiding businesses in decision-making by providing insights into past and present performance, as well as customer behavior. It enables automated reporting, ensuring that stakeholders have timely access to critical data. Tableau also facilitates collaboration through Tableau Server and Tableau Online, allowing teams to share and discuss insights efficiently. Additionally, Tableau Prep helps clean and structure data before analysis, and calculated fields allow users to create customized metrics to suit specific analytical needs, enhancing the overall data analysis process

III. DATA VISUALIZATION

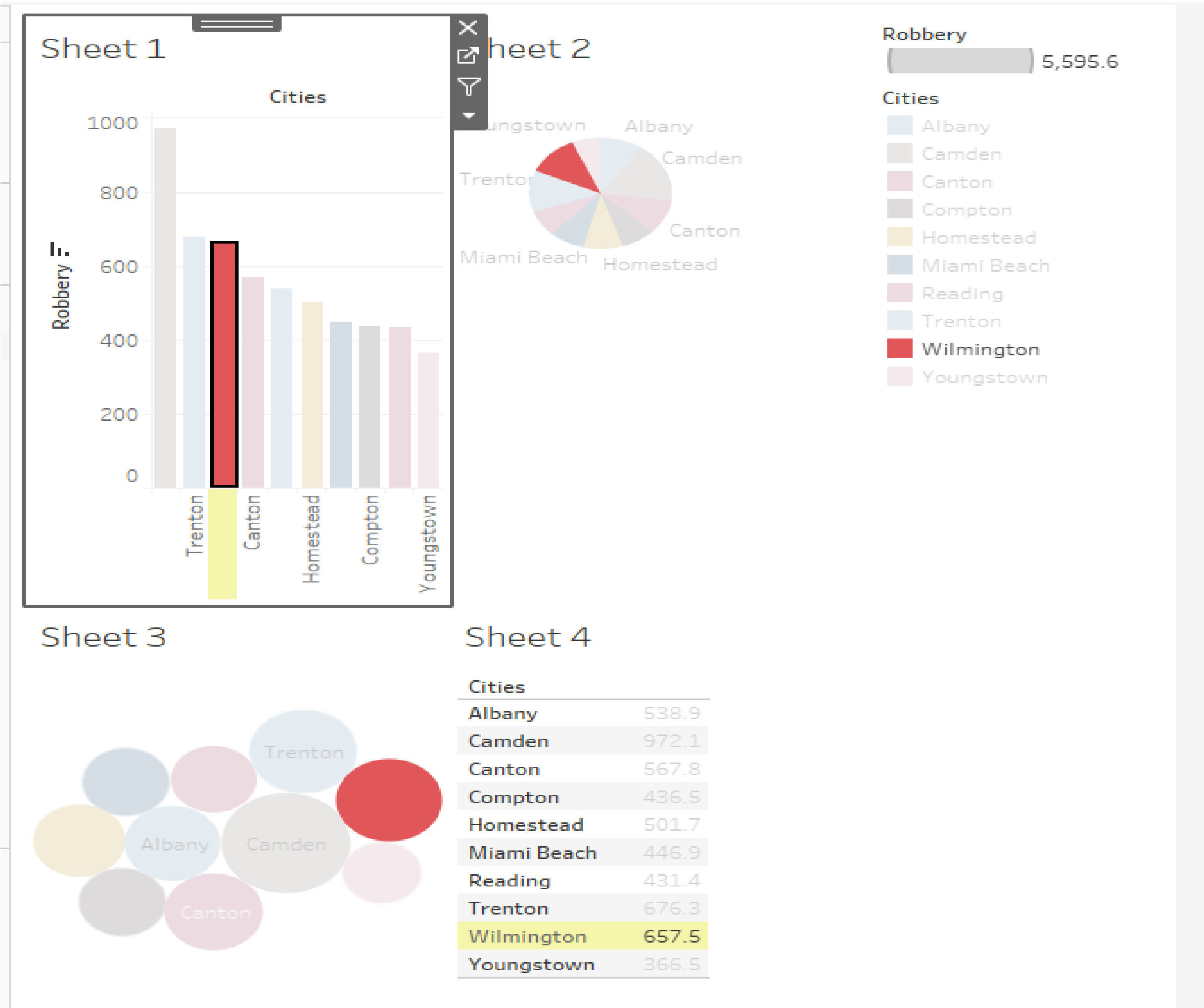

Figure 1:

The above image contains multiple sheets visualizing crime data for different cities. It includes a bar chart, a pie chart, a bubble chart, and a table displaying robbery statistics. Specific cities like Wilmington and Trenton are highlighted in different charts.

Figure 2:

The above graph is used to represent crime trends, in this case, robbery and car theft over various time intervals. It is made up of three sheets with data at varying granularities: quarterly (Sheet 1), monthly (Sheet 2), and daily (Sheet 3). The line graphs depict a general drop in both crimes across quarters, whereas monthly and daily trends fluctuate with observable peaks and dips. The blue and orange lines signify robbery and car theft, with identical patterns of movement, implying a connection between the two offenses.

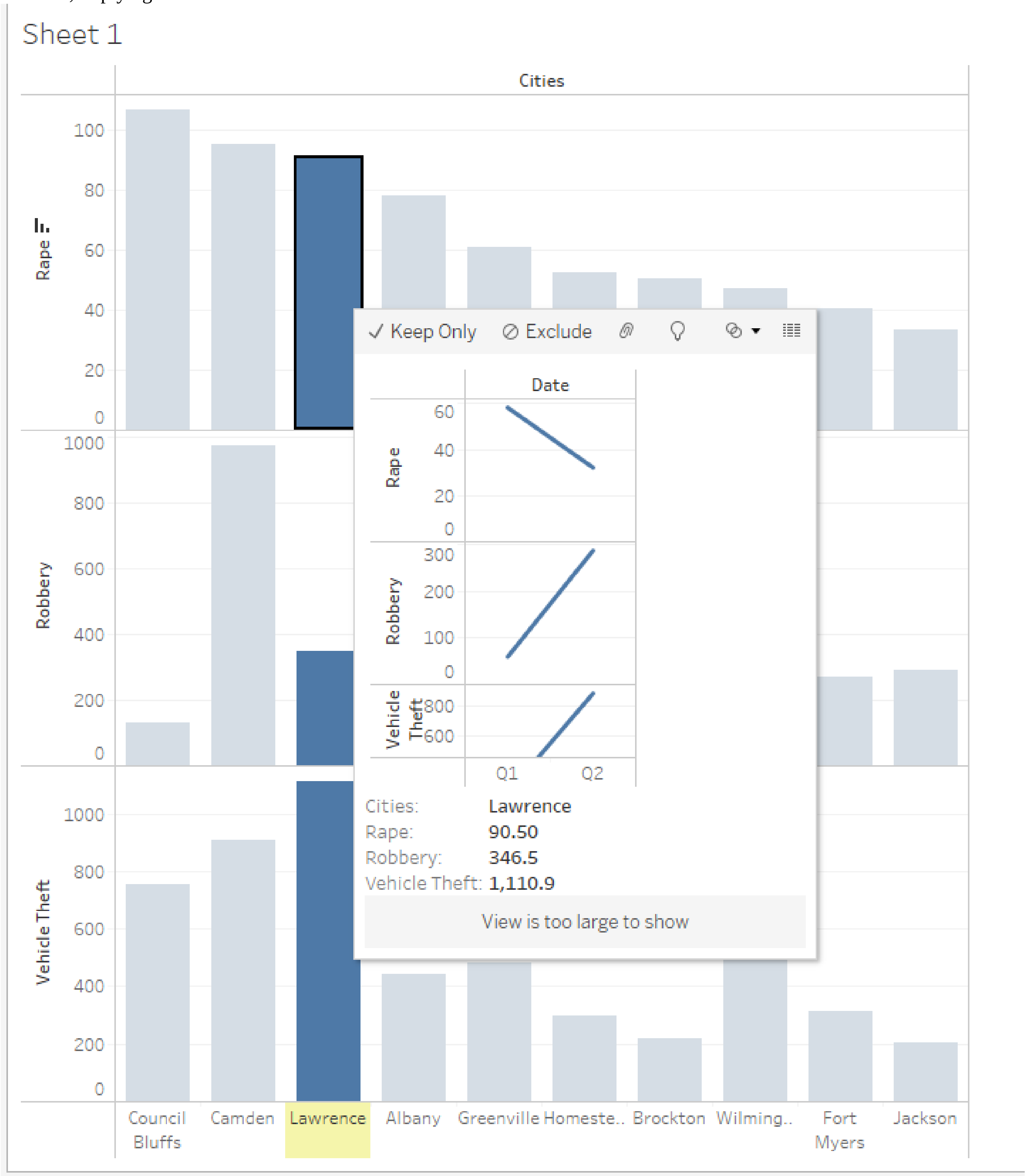

Figure 3:

The graph illustrates crime rates in cities, in terms of rape, robbery, and car theft. Lawrence is highlighted with most car theft (1,110.9) and medium-sized robbery (346.5). A popup for trend analysis illustrates crime variations for two quarters. City filters facilitate interactive exploration of crime trends. On the whole, car theft is the most common crime in Lawrence.

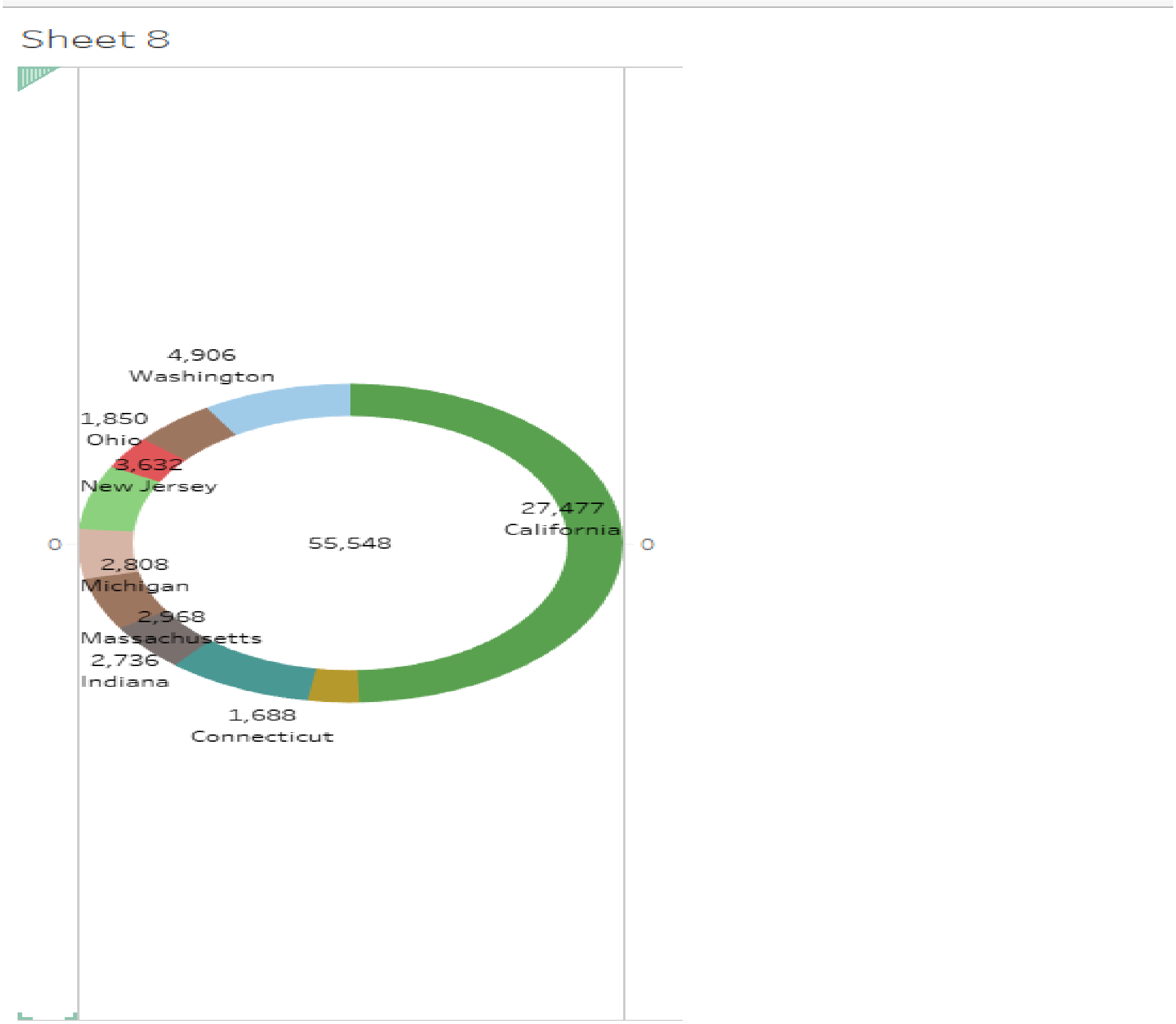

Figure 4

The donut chart shows data distribution across various U.S. states, with **California (27,477)** and **New Jersey (6,632)** having the highest values. The total sum of all states shown is **55,548**, indicating California alone accounts for nearly half of the total.

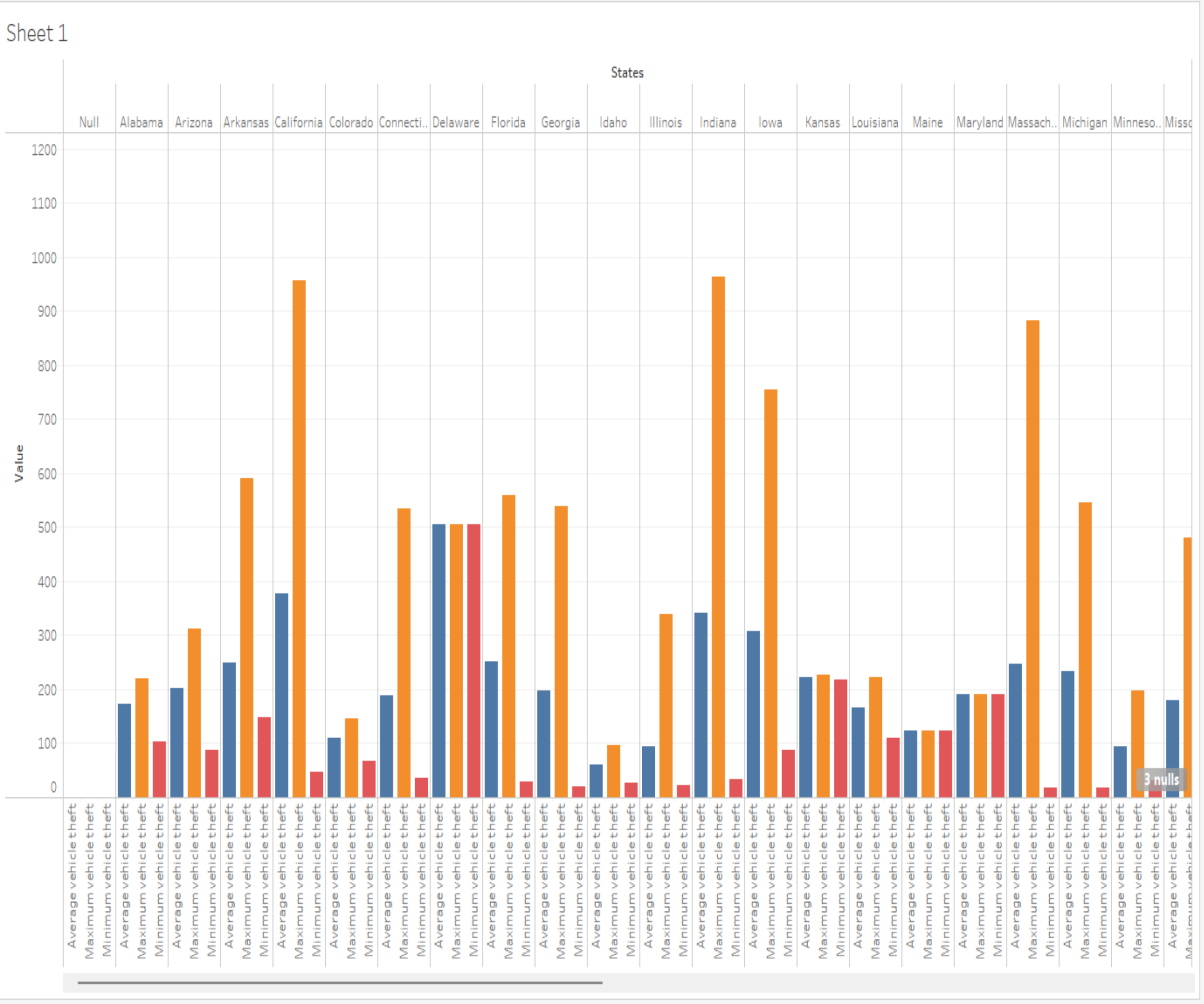

Figure 5:

The bar chart displays sales across various states, segmented by product categories (shown in different colors).
California, Texas, and New York have the highest overall sales.
Each state shows a different distribution of sales across categories.
California leads in almost all product categories.
Texas and New York follow, with strong but varied category performance.
Lower sales are observed in smaller states across most categories.

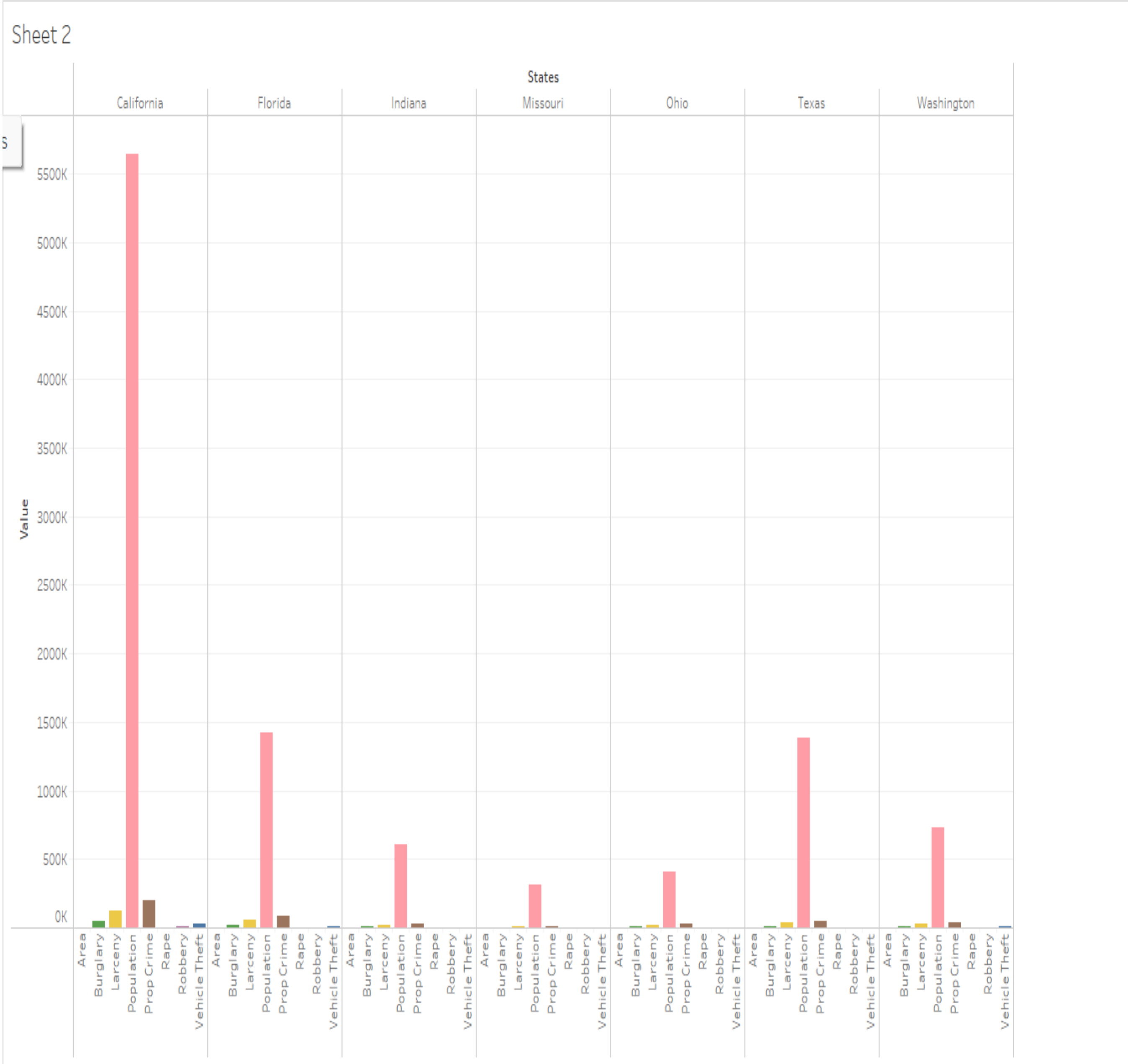

Figure 6:

This bar chart displays **values by state and product category**, with **California** showing a significantly higher value—over 5.5 million—mainly from the "Producers" category. Other states like **Florida**, **Texas**, and **Washington** also show notable contributions, but much lower in comparison

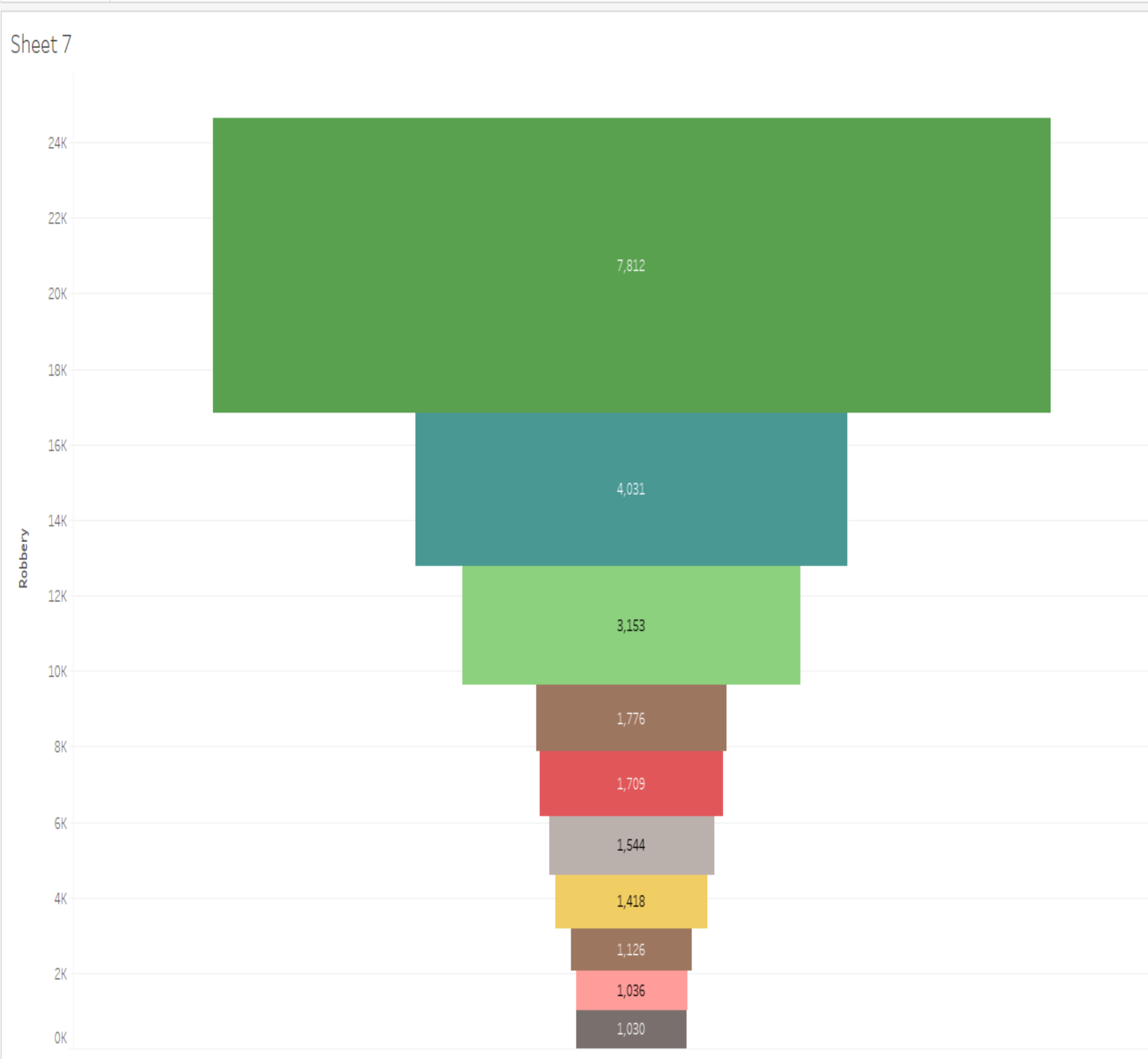

Figure 7:

This image is a **funnel chart** representing a sequential drop in values across different stages or categories. The top segment shows the highest value (7,612), and each subsequent layer decreases in size, ending with the smallest value of 1,089. The chart visually emphasizes the **progressive reduction** through each stage, likely indicating a process like customer conversion, filtering, or data refinement

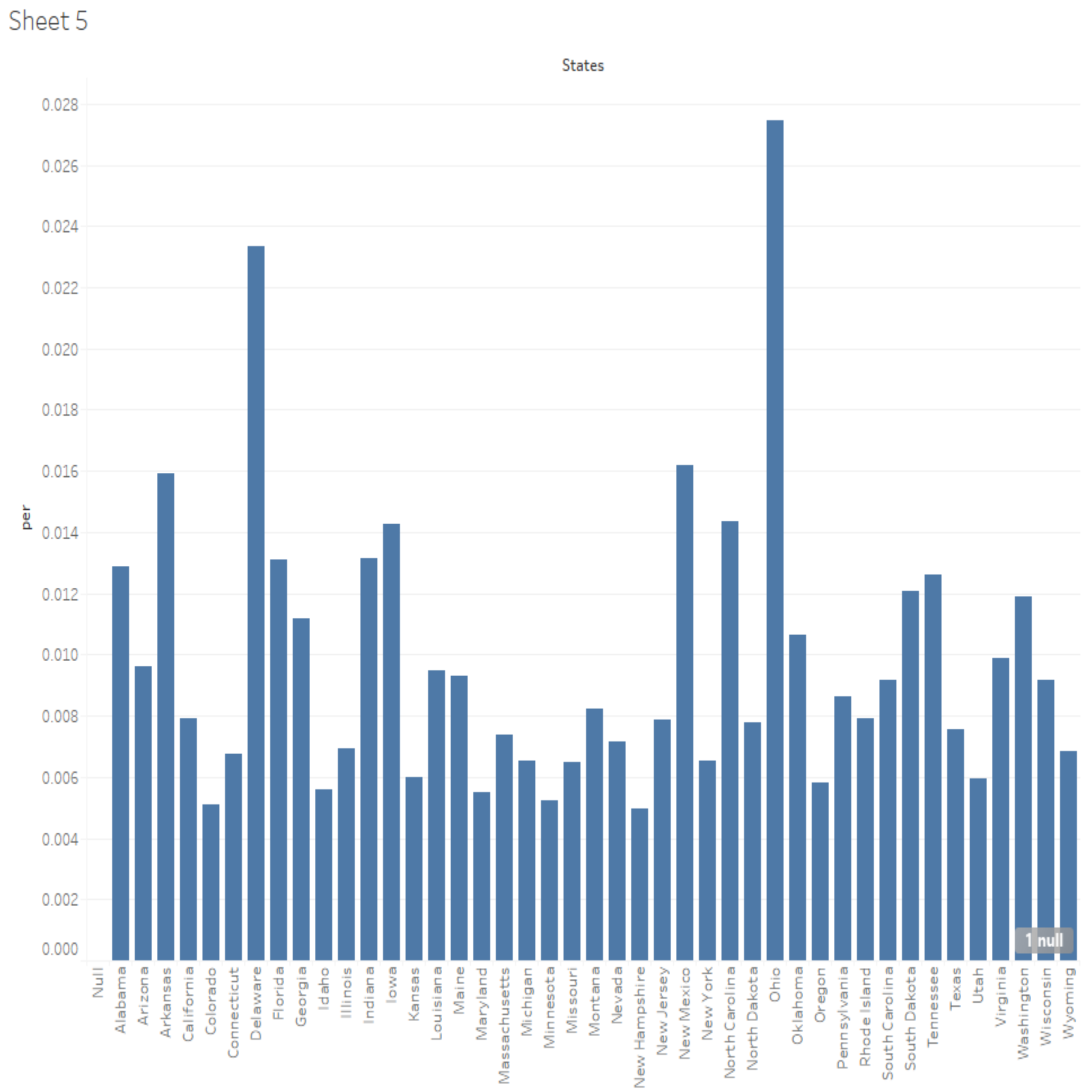

Figure 8:

This bar chart shows the **percentage (per) distribution across U.S. states**, with **North Dakota** having the highest value (around 0.028). Other states like **Colorado** and **New Hampshire** also stand out, while most states fall below the 0.015 mark.

IV. RESULT & DISCUSSION

The dataset shows that **property crimes like larceny and burglary are the most common**, echoing national trends—FBI reports say larceny-theft makes up about **70% of all property crimes** in the U.S. Violent crimes like robbery and assault appear less often but still impact certain cities more heavily, similar to patterns seen in cities like Detroit or St. Louis.

Crimes often occur in the **evening hours**, matching national studies that show peaks between **8 PM and midnight**. This timing insight helps police better plan patrols. Overall, the analysis highlights where and when crimes happen, making it valuable for **resource planning, public safety strategies**, and understanding **urban crime dynamics** more effectively.

Looking into crime stats from different U.S. cities, some patterns really stood out. Property crimes—especially things like theft and burglary—happen way more often than violent crimes like murder or assault. That lines up with national trends, too. It's also noticed that a lot of these crimes take place in the evening, which could help police figure out when and where to focus their efforts. This kind of data is super useful—not just for understanding crime, but for helping cities make smarter, safer decisions for the people who live there.

According to the FBI, larceny-theft makes up about 70% of all property crimes in the U.S.

Violent crimes like robbery and assault are less frequent but still heavily affect certain urban areas, such as Detroit and St. Louis.

Evening hours—particularly between 8 PM and midnight—see the highest crime rates. This pattern mirrors national studies and provides key insights for law enforcement planning. Understanding when and where crimes happen allows police to better target patrols and prevention strategies. The analysis not only reveals urban crime patterns but also supports more effective public safety initiatives. By using this data, cities can allocate resources wisely and improve safety for their communities. Ultimately, data-driven approaches can lead to smarter decision-making and more secure neighborhoods.

V. CONCLUSION

This dataset gives a useful snapshot of what crime looks like in different U.S. cities. It shows us that property crime is the bigger issue in most places, and that there are noticeable differences in crime rates from one city to another. Because it includes when crimes happened, there's also a chance to explore patterns over time—maybe even predict or prevent some of them. The dataset spans a recent and relevant period from 2022 to 2025, offering a solid foundation for analyzing short-term crime trends. While it contains 305 fully populated records, some values may still benefit from formatting adjustments or standardization. To get even more out of it, a bit of cleanup would help—like refining number formats and checking consistency. All in all, it's a solid starting point for anyone looking to understand or reduce crime through better data.

VI. REFERENCE

[1]. **Nwankwo, C.S., Raji, M.K., and Oghogho, E.S. (2018).** Application of Data Analytics Techniques in Analyzing Crimes.

[2]. **Rajeswari, C., Basu, D., and Maurya, N. (2017).** Comparative Study of Big Data Analytics Tools: R and Tableau. IOP Conference Series: Materials Science and Engineering, Vol. 263, No. 4, p.042052.

[3]. **Gaur, A., Kammadanam, A., and Krishna, S.G. (Year N/A).** Visualization of Crime Survey Data.

[4]. **Silva Atencio, G., and Umaña Ramírez, M. (2023).** Predictive Models in Pandemic Times and Their Impact on the Analysis of Crime. Journal of Applied Research and Technology, 21(3), (pp. 484–495).

[5]. **Ajagbe, S.A., Oladipupo, M.A., and Balogun, E.O. (2020).** Crime Belt Monitoring via Data Visualization: A Case Study of Folium. International Journal of Information Security, Privacy and Digital Forensic, 4(2), (pp. 35–44).

[6]. **Kane, R. (2018).** A Statistical Study on the Impact of Weather on Crime: Technical Report. National College of Ireland (Doctoral Dissertation).

[7]. **Nair, L., Shetty, S., and Shetty, S. (2016).** Interactive Visual Analytics on Big Data: Tableau vs D3.js. Journal of e-Learning and Knowledge Society, 12(4), (pp. **–**).

[8]. **Sharma, S., and Dronavalli, S.C. (2024).** Data Analysis and Visualization of Crime Data. Electronic Imaging, 36, (pp. 1–6).

[9]. **Liu, S. (2021).** Design and Implementation of a Geospatial Dashboard for Crime Analysis and Prediction. Toronto Metropolitan University, (pp. **–**).

[10]. **Akhtar, N., Tabassum, N., Perwej, A., and Perwej, Y. (2020).** Data Analytics and Visualization Using Tableau Utilitarian for COVID-19 (Coronavirus). Global Journal of Engineering and Technology Advances, (pp. **–**).

[11]. **Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., and Liu, Q. (2019).** Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data. IEEE Access, 7, (pp. 106111–106123).

Agarwal, J., Nagpal, R., and Sehgal, R. (2013). Crime Analysis Using K-means Clustering. International Journal of Computer Applications, 83(4), (pp. **–**).

[12]. **Wang, H., Kifer, D., Graif, C., and Li, Z. (2016).** Crime Rate Inference with Big Data. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (pp. 635–644).

[13]. **Feng, M., Zheng, J., Han, Y., Ren, J., and Liu, Q. (2018).** Big Data Analytics and Mining for Crime Data Analysis, Visualization and Prediction. In Advances in Brain Inspired Cognitive Systems: BICS 2018, (pp. 605–614). Springer.

[14]. **Sathyadevan, S., Devan, M.S., and Gangadharan, S.S. (2014).** Crime Analysis and Prediction Using Data Mining. In First International Conference on Networks & Soft Computing (ICNSC), (pp. 406–412). IEEE.