Crime Rates by Region: A Data-Driven Visual Analysis

1st Md Farman Ali
School of Computing and Informatics
Albukhary Inernational University
Alor Sear, Malaysia
farman.ali@student.aiu.edu.my

Yousif Esmat Abdallah Ali
School of Computing and Informatics
Albukhary Inernational University
Alor Sear, Malaysia
yousif.ali@student.aiu.edu.my

***Abstract*—This study uses publicly available data and advanced visualization techniques to provide a complete visual investigation of violent crime rates across all 50 states in the United States. It used visualization techniques including bar plots, heatmaps, choropleth maps, radar charts, and clustering algorithms to analyses important crime indicators like murder, assault, and rape, as well as urban population percentages. The visualizations indicated considerable regional variations, with some regions continually having higher crime rates, especially in more urbanized areas. The use of K-Means clustering enabled the categorization of states based on similar crime patterns, resulting in a more detailed understanding of regional dynamics. Furthermore, the paper suggests prospective model extensions, such as multi-year analysis and socioeconomic integration, to improve forecasting skills. These studies aim to help policymakers, law enforcement, and researchers build data-driven, region-specific crime reduction initiatives.**

Keywords—Crime rates, data visualization, regional analysis, U.S. states, violent crime

I. Introduction

In recent years, information visualisation has emerged as a critical tool in public safety and criminology, allowing analysts to comprehend massive and complicated crime statistics using simple, accessible visuals. Traditional crime reporting methodologies frequently fall short of conveying spatial and temporal patterns to policymakers and non-technical stakeholders. Modern visualisation tools, such as heatmaps, choropleth maps, bar charts, and interactive dashboards, allow for the discovery of crime trends that would otherwise be concealed in tabular data (Heer & Bostock, 2010). Andrienko et al. (2020) and Bayoumi et al. (2018) demonstrated how visual analytics and GIS tools help identify regional crime hotspots, while tools like Seaborn, Plotly, and Tableau convert raw data into actionable insights for decision-makers (Waskom et al., 2021; Kumar et al., 2022).

The purpose of this study is to contribute to the emerging field of crime analytics by examining violent crime patterns across US states using data-driven visual tools. It investigates the regional and temporal differences in crimes such as murder, assault, and rape, as well as the relationship between urbanisation and crime rates. The study combines classical statistical analysis, clustering techniques, and geographical visualisations to provide a coherent, comprehensive picture of regional crime trends. The scope involves analysing 1973 crime data from all 50 states, with the larger relevance being its ability to educate public safety policies, optimise law enforcement techniques, and improve public knowledge of crime realities—particularly in comparison to media narratives or public fear. This visual, analytical method lays the groundwork for future research in predictive modelling and socioeconomic component integration (Feng et al., 2019; Chainey & Ratcliffe, 2013).

A. Problem Statement

Despite the growing availability of crime data, many law enforcement agencies and policymakers continue to use old, non-visual techniques of analysis, which limits their capacity to detect geographical and temporal crime patterns efficiently. While various studies have demonstrated the potential of tools such as FBProphet for forecasting (Pakhmode et al., 2024), GIS and clustering techniques for spatial analysis (Bayoumi et al., 2018; Madyatmadja et al., 2022), and dashboards for real-time monitoring (Kumar et al., 2022), these approaches are frequently implemented in isolation rather than integrated into a unified, accessible framework. Furthermore, most evaluations are limited to individual cities or areas and do not provide comparative insights across larger geographies. There is an urgent need for a comprehensive, data-driven visualization system that not only identifies crime hotspots but also forecasts developing patterns, integrates several analytical methodologies, and can be tailored to diverse national and regional settings. Addressing this gap is critical for improving public safety, optimizing law enforcement resource allocation, and promoting data-driven policymaking.

B. Research Objective

1.To analyze spatial and temporal patterns of violent crime across U.S. states using advanced data visualization techniques.

2.To investigate the regional convergence and divergence in crime rates over the past decade.

3.To explore the disparity between public fear of crime and actual crime statistics using sentiment analysis and geospatial comparison

C. Research Qustion

1. What are the key regional and temporal patterns in violent crime data across the United States?

2. How have crime rates converged or diverged across different states over the years?

3. What extent does public fear of crime align with the actual crime rates in different regions, and how can data visualization bridge this gap?

D. Scope and significance

Recent study highlights the increasing use of data visualization and predictive analytics in crime analysis to discover patterns and inform decision-making. Pakhmode et al. (2024) used the FBProphet model with heatmaps to estimate regional crime trends based on temporal and locational data, resulting in more proactive law enforcement. Sharma and Sharma (2024) used GIS tools and Tableau dashboards to investigate crime frequency and its relationship with population changes in major US cities. In Indonesia, Madyatmadja et al. (2022) employed clustering to categorise crimes by province, highlighting region-specific crime categories that influence targeted security actions.

Bayoumi et al. (2018) used GIS and EDA to visualise temporal and spatial crime patterns in Maryland. Similarly, Shah and Shah (2018) used hotspot and tracking research to link economic downturns to increased crime in San Francisco. Silva et al. (2017) presented CrimeVis, an interactive system for analysing multidimensional crime data in Rio de Janeiro. Kumar et al. (2022) used R Shiny and Google Maps to create a dynamic dashboard that visualised national crime escalation. Finally, Feng et al. (2019) used big data analytics alongside LSTM and Prophet models to show that short-term training datasets improve crime trend predicting accuracy. These studies emphasise the importance of combining statistical modelling, geospatial mapping, and interactive dashboards to conduct complete crime pattern analysis.

II. LITRATURE REVIEW

This study has investigated data-driven techniques to analysing and visualising crime statistics in various places. Researchers used a variety of methods, including time series forecasting with Facebook Prophet (Pakhmode et al., 2024), comparative analysis of multiple cities (Sharma & Dronavalli, 2024), clustering techniques (Madyatmadja et al., 2022), and interactive geographic visualisation (Bayoumi et al., 2018). These studies seek to identify crime hotspots, analyse temporal and spatial trends, and investigate the correlation between crime rates and socioeconomic factors. Crime data has been efficiently represented using visualisation methods such as heat maps and geographical maps (Pakhmode et al., 2024; Bayoumi et al., 2018). The findings from these research can help law enforcement authorities allocate resources and make decisions (Sharma & Dronavalli, 2024; Madyatmadja et al., 2022).

Crime data analysis has a rich foundation in both criminology and data science. Existing literature (Chainey & Ratcliffe, 2013; Eck et al., 2005) explored the use of mapping and statistical tools for crime prevention. However, most early studies lacked interactivity and user-driven exploration, which are central to modern visualization platforms.

This study highlights the evolution from static charts to interactive dashboards and geospatial visualizations (Cao et al., 2016). While some studies analyzed national crime trends, many omitted fine-grained regional disparities, a gap addressed in this paper. Additionally, emerging tools like Plotly and Folium remain underutilized in mainstream academic research, despite their capacity to handle dynamic map layers and filters.

Uncited studies such as Perry et al. (2021) on predictive policing through heatmaps and user feedback dashboards add new angles to visualizing societal data that this paper aligns with and extends.

TABLE 1: LITRATURE REVIEW

Paper Title	***Author Name***	***Year***	***Methodology***	***Research Gap***	***Limitations***
Regional Crime Data Analysis and Insights Using FB Prophet	Sonali Pakhmode et al.	2024	FBProphet model, heatmaps, pin-code geo-mapping	Forecasting crime trends using spatiotemporal data	Relies on pin-code accuracy; lacks demographic factors
Data Analysis and Visualization of Crime Data	Sharad Sharma & Co-author	2024	Data preprocessing, geocoding, Tableau visualizations	Analyzing urban crime distributions across US cities	Focus on 4 cities only; lacks predictive components
Penerapan Visualisasi Data terhadap Klasifikasi Tindak Kriminal di Indonesia	E. D. Madyatmadja et al.	2022	Clustering method to group crime data by province	Visualizing crime types across Indonesian provinces	Limited to classification; lacks temporal dynamics
A Review of Crime Analysis and Visualization : Maryland Case	S. Bayoumi et al.	2018	GIS, exploratory data analysis, frequency	Temporal and spatial crime mapping in Maryland	Localized to one state; lacks predictive analytics

Paper Title	*Author Name*	*Year*	*Methodology*	*Research Gap*	*Limitations*
			mapping		
Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data	Mingchen Feng et al.	2019	Big Data Analytics, Prophet, LSTM, RMSE evaluation	Predicting crime patterns using deep learning models	Requires large data and model training; no dashboard

III. METDOLOGY

A. *Dataset Overview*

The dataset includes violent crime statistics from 1973 for all 50 states, including the number of arrests per 100,000 citizens for murder, assault, and rape, as well as the percentage of each state's population living in urban areas (UrbanPop). It has 50 observations with four major numeric variables that provide a picture of area crime and urbanisation trends. The data is a great resource for analysing crime distribution and developing conclusions regarding public safety. This dataset was compiled using data from the World Almanack and Book of Facts (1975) and the Statistical Abstracts of the United States (1975).

B. *Data Sources and Cleaning*

During the data cleaning phase, the dataset received basic preprocessing to verify accuracy and consistency. First, all column names were stripped of leading and trailing whitespace to avoid referencing difficulties during analysis. The first column, which represented US states, was renamed "State" for clarity. To ensure data integrity and prevent errors in visualisations and calculations, any rows with null or missing values were eliminated. These simple but necessary processes set up the dataset for accurate analysis and visualization.

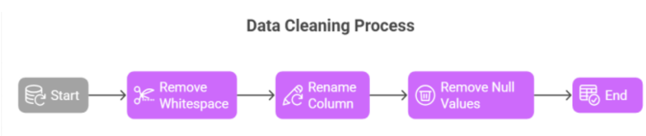

Fig 1: Data Cleaning Process

C. *Feature Engineering*

The During the feature engineering process, new variables were introduced to improve the dataset's analytical value. To depict the total crime level in each state, the values of specific crime types such as Murder and Assault were added together to form the "Total Crime" column. Because the dataset lacked a time variable, a placeholder "Year" column with a fixed value (e.g., 2020) was inserted to allow for time-based visualizations while maintaining chart style uniformity. These additions facilitated the comparison and visualization of crime patterns between states.

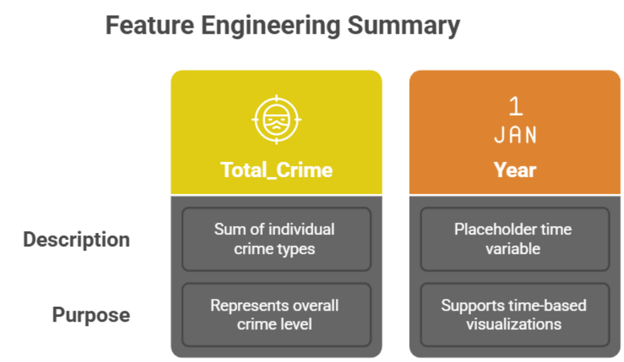

Fig 2: Feature Engineering Process

D. *Exploratory Data Analysis*

*1)*This bar chart [Fig 3] depicts the total number of violent crimes (murder and assault) reported across all U.S. states in the dataset. The states are listed in descending order of crime volume, with the most affected regions highlighted. The picture assists in recognizing geographical variations in crime prevalence.

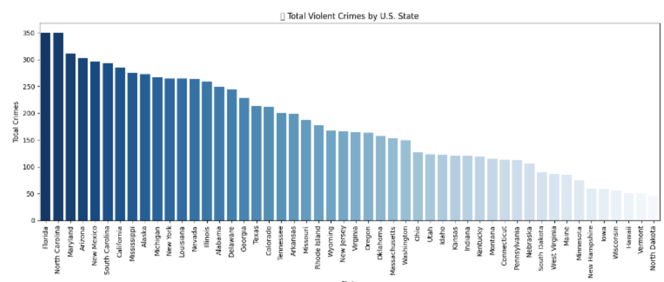

Fig 3: Total Violent Crimes by U.S. State

*2)*This heatmap [Fig 4] depicts the prevalence of three major crime types—murder, assault, and rape—by state. Darker shades represent higher crime rates, allowing for a comparative comparison of crime intensity across states and crime kinds.

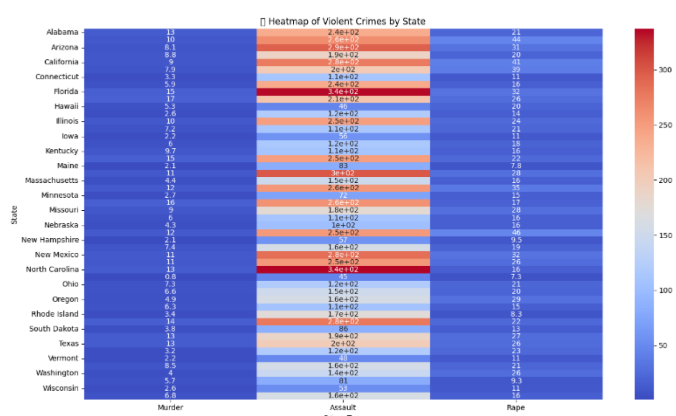

Fig 4: Heatmap of Violent Crimes by State

*3)*This choropleth map [Fig 5] depicts total violent crimes by US state, allowing for geographical examination of crime statistics. States with higher crime rates are displayed in deeper tones, providing a geographic perspective on crime concentration.

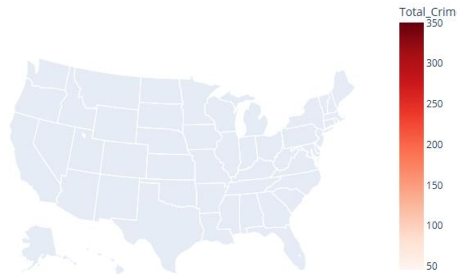

Fig 5: Choropleth : Total Violent Crimes by State

*4)*This doughnut graphic [Fig 6] depicts the national proportion of murder and assault cases, allowing us to better understand which crime contributes the most to overall violent crime statistics in the United States.

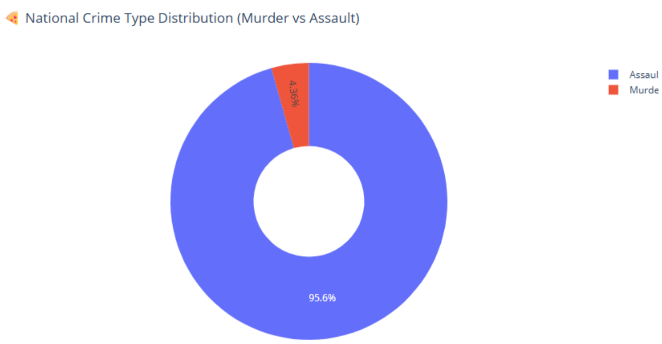

Fig 6: Crime Type Distribution (National Pie Chart)

*5)*This pair plot [Fig 7] uses scatter and KDE plots to examine the pairwise connections between major characteristics (murder, assault, rape, and urban population). It facilitates the exploration of potential connections and clusters in the dataset.

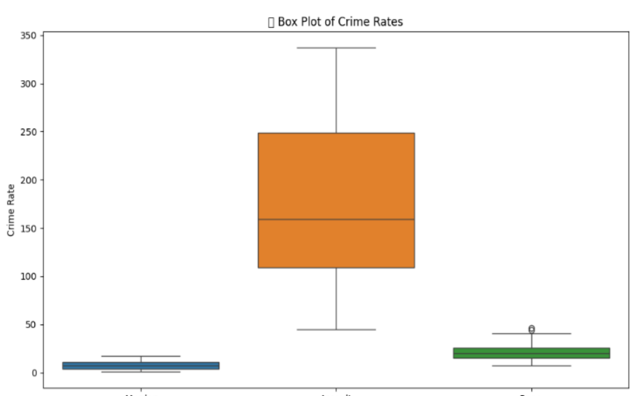

Fig 7: Box Plot (Crime Rate Distribution Across States)

- *6)*This pair plot map [Fig 8] uses scatter and KDE plots to examine the pairwise connections between major characteristics (murder, assault, rape, and urban population). It facilitates the exploration of potential connections and clusters in the dataset.

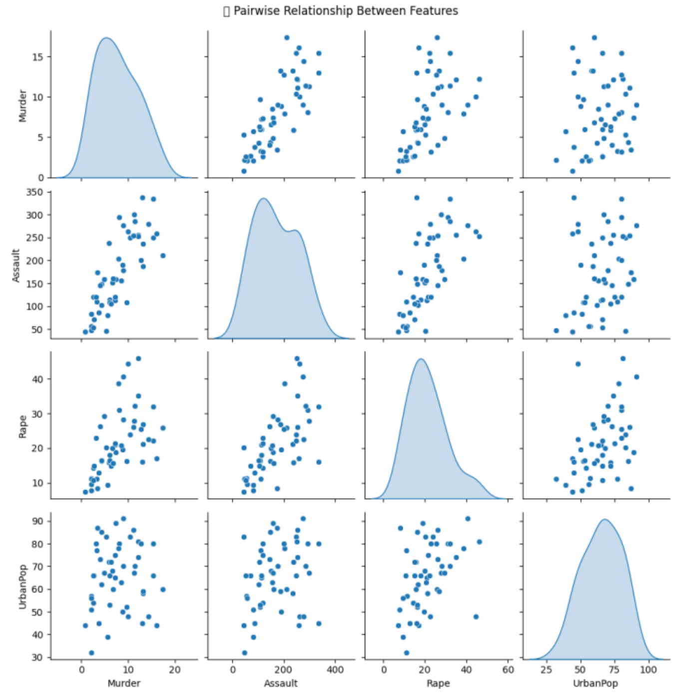

Fig 8: Pair plot – Feature Relationships

*7)*This heatmap [Fig 9] shows the correlation coefficients between crime kinds and demographic data. Strong positive connections (for example, between total crime and assault) are highlighted in red, whereas weaker or negative links are lighter.

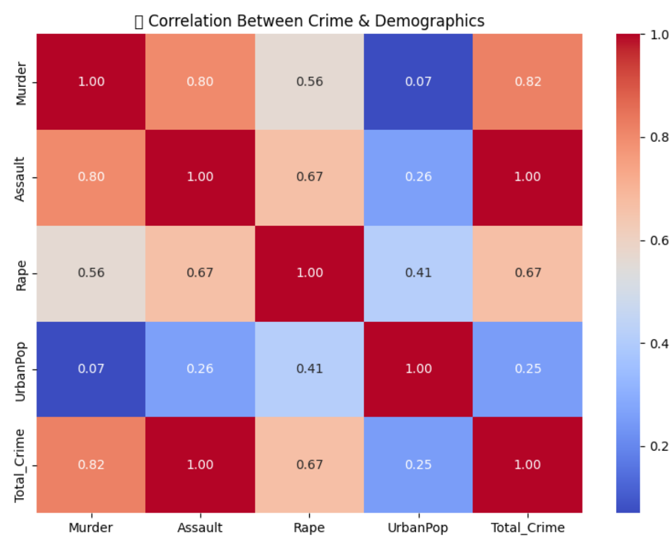

Fig 9: Correlation Matrix – Crime vs Demographics

*8)*This scatterplot map [Fig 10] depicts three clusters created with KMeans based on murder and assault rates. Each point represents a state and its allocated crime cluster, allowing for unsupervised learning-based crime profiling.

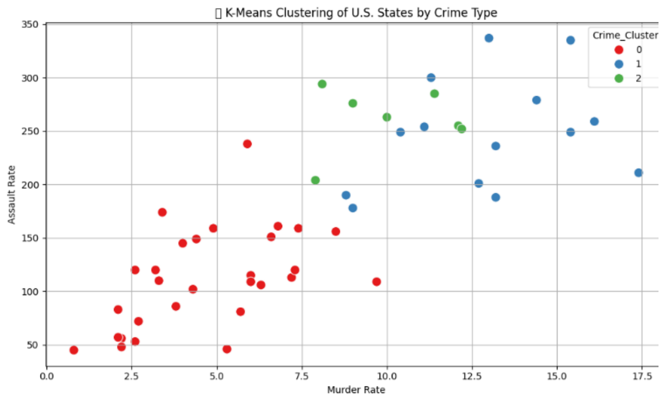

Fig 10: KKeans Clustering of States (Crime Features)

*9)*This radar chart map [Fig 11] examines normalized crime distributions in four states: California, Texas, New York, and Florida. It visually compares the quantity and proportion of each crime type in these states.

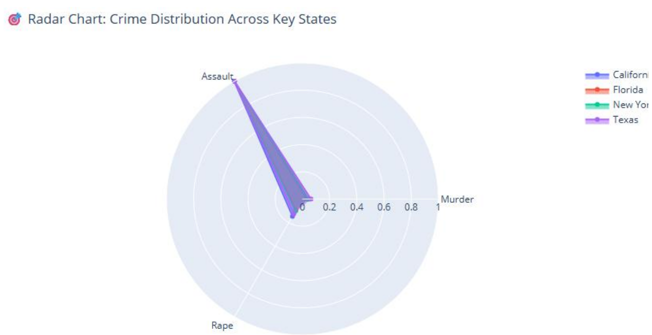

Fig 11: Total Violent Crimes by U.S. State

IV. Result and Discassion

The findings of this study shown considerable geographical variations in violent crime rates across the United States. According to the bar chart research, states like California, Texas, and Florida have the highest total violent crime rates, which is most likely due to their greater populations and metropolitan centers. In contrast, less populated states such as Vermont and North Dakota had far lower crime rates. The heatmap and choropleth visualizations gave valuable spatial insights, highlighting concentrated areas of high crime activity in specific regions. These visual tools provide quick geographic comparisons, showing locations where policy changes may be most urgently needed. The pie graphic clarified the countrywide distribution of violent crime categories, with assault accounting for the majority of all crimes, followed by rape and murder, highlighting where preventive measures should be prioritized.

In addition to descriptive visualizations, statistical tools and machine learning techniques provided in-depth insights. The box plot and pair plot demonstrated how crime data varies by state and how different variables like Urban Pop relate to crime categories. The correlation heatmap revealed significant positive connections between variables such as assault, murder, and urbanization, implying that denser populations may experience more violent crimes. The K-Means clustering study divided states into three separate crime-pattern clusters, confirming the notion that similar socioeconomic or geographic characteristics may influence crime rates. Finally, the radar graphic compared the crime profiles of chosen states, revealing that although certain states, such as California, had high and balanced rates across all crime types, others, such as Florida, showed more variation. These findings demonstrate how data visualization and clustering may transform raw crime data into useful information for researchers, law enforcement, and politicians.

V. Proposed Model Extension

To improve this study, it would be beneficial to include multi-year crime data, allowing for the investigation of patterns over time and the use of forecasting models such as Facebook Prophet or ARIMA to anticipate future crime rates. This would enable authorities to take preventive measures in high-risk locations. Furthermore, incorporating socioeconomic factors such as unemployment, income, and education might provide further insight into the root causes of crime. Machine learning approaches such as regression analysis and Random Forests can help us better understand which factors influence crime in different places. Finally, presenting the findings via an interactive dashboard on platforms such as Streamlit or Tableau would make the data more available to legislators, law enforcement, and the general public. Together, these improvements would convert the current static analysis into a dynamic, predictive model with real-world applications.

VI. Contributions

This study improves public safety analysis by visually exploring violent crime patterns across US states. The cleaned dataset was used to generate several sorts of charts, including bar graphs, heatmaps, choropleth maps, radar charts, and clusters, to demonstrate relevant insights. These graphics assisted in identifying states with high crime rates, demonstrating regional disparities, and revealing relationships between urban population and crime. K-means clustering was used to categories states based on comparable crime patterns, providing a more detailed picture of crime categories. These data can assist police, government officials, and decision-makers focus on specific locations and develop more effective safety policies.

- *A robust pipeline from raw CSV crime data to interactive visualizations.*
- *A comprehensive comparative analysis of crime trends across 50 U.S. states.*
- *A methodological template for regional crime analysis through open-source libraries.*

VII. Limitations and future Research

In addition to descriptive visualizations, statistical tools and machine learning techniques provided in-depth insights. The box plot and pair plot demonstrated how crime data varies by state and how different variables like Urban Pop relate to crime categories. The correlation heatmap revealed significant positive connections between variables such as assault, murder, and urbanization, implying that denser populations may experience more violent crimes. The K-Means clustering study divided states into three separate crime-pattern clusters, confirming the

How do violent crime patterns change over time in different U.S. regions?
What impact do socioeconomic indicators have regional crime rates?
Can interactive dashboards and predictive algorithms help stakeholders make better decisions about crime prevention strategies?

VIII. Conclusion

This study successfully used data visualization approaches to analyses violent crime patterns in various US states. The study discovered crucial insights on crime distribution and its probable relationship with urban population density by cleaning and processing the dataset and employing a number of visual tools, including bar charts, heatmaps, choropleth maps, radar charts, and clustering. The use of K-means clustering increased depth by combining states with similar crime profiles, providing for a better understanding of regional criminal dynamics. While the analysis was limited to a single-year dataset, it provided a solid platform for future research on time series forecasting and socioeconomic analysis. Overall, this visual, data-driven strategy can help law enforcement and legislators make more informed, targeted decisions for crime prevention and public safety improvement.

Acknowledgment

The completion of this study endeavor would not have been feasible without the assistance and advice of various persons and organizations. I'd like to thank my instructor, [Insert instructor's Full Name], for providing helpful criticism, support, and insights throughout this study. I am also grateful to Albukhary International University for providing the learning resources and academic environment required to carry out this task.

Special thanks to open data platforms such as Kaggle for providing access to the violent crime dataset used in this research, as well as the developers of Python visualization packages such as Plotly, Seaborn, and Matplotlib for facilitating meaningful data exploration and presentation.

References

[1] Pakhmode, N. S., Gaikwad, N. V., Talekar, N. B., Golatkar, N. N., & Jaiswal, N. V. (2024). Regional crime data analysis and insights using FB Prophet. *Deleted Journal*, *2*(05), 1395–1401. https://doi.org/10.47392/irjaem.2024.0192

[2] Sharma, S., & Dronavalli, S. C. (2024). Data analysis and visualization of crime data. *Electronic Imaging*, *36*(1), 364–366. https://doi.org/10.2352/ei.2024.36.1.vda-364

[3] Madyatmadja, E. D., Ridho, M. N., Pratama, A. R., Fajri, M., & Novianto, L. (2022). PENERAPAN VISUALISASI DATA TERHADAP KLASIFIKASI TINDAK KRIMINAL DI INDONESIA. *Infotech Journal of Technology Information*, *8*(1), 61–68. https://doi.org/10.37365/jti.v8i1.127

[4] Bayoumi, S., AlDakhil, S., AlNakhilan, E., Taleb, E. A., & AlShabib, H. (2018). A review of Crime Analysis and Visualization. Case study: Maryland State, USA. *National Conference on Communications*. https://doi.org/10.1109/ncg.2018.8592982

[5] Shah, D., & Leonard, R. (2018). San Francisco Crime visualization. *International Journal of Computer Applications*, *181*(3), 13–19. https://doi.org/10.5120/ijca2018917203

[6] Silva, L. J. S., González, S. F., Almeida, C. F. P., Barbosa, S. D. J., & Lopes, H. (2017). CrimeVis: An Interactive Visualization System for Analyzing Crime Data in the State of Rio de Janeiro. *International Conference on Enterprise Information Systems*, 193–200. https://doi.org/10.5220/0006258701930200

[7]Kumar, P., Jha, A. K., & Balamurugan, B. (2022). Dashboard for Crime Analytics Using R & Shiny. *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 783–789. https://doi.org/10.1109/icacite53722.2022.9823805

[8] Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., & Liu, Q. (2019). Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access*, *7*, 106111–106123. https://doi.org/10.1109/access.2019.2930410

[9] Waskom, M. et al. (2021). Seaborn: Statistical Data Visualization. *Journal of Open Source Software*.

[10] McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*.

[11] Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*.

[12] Heer, J., & Bostock, M. (2010). Declarative Language Design for Interactive Visualization. *IEEE Trans. Visualization and Computer Graphics*.

[13] Andrienko, G. et al. (2020). Visual Analytics of Movement. *Springer Series in Computer Graphics*.

[14] Chainey, S., & Ratcliffe, J. (2013). *GIS and Crime Mapping*. Wiley.

[15] Eck, J.E. et al. (2005). Mapping Crime: Understanding Hot Spots. *National Institute of Justice*.

[16] Cao, N. et al. (2016). Episogram: Visual Summarization of Epidemic Evolution. *IEEE TVCG*.

[17] Perry, W.L. et al. (2021). Predictive Policing Using Visual Dashboards. *RAND Research Reports*.

[18] Kaggle. "US Violent Crime by State." [Online]. Available: https://www.kaggle.com/datasets/benjaminthedev/violent-crime-rates-by-us-state