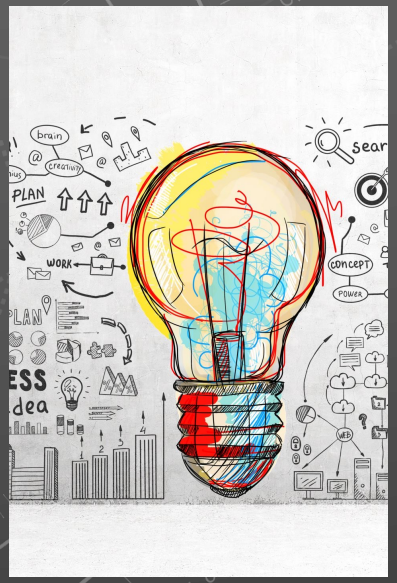


OBINWANNE ALISIGWE

DATA ANALYST PORTFOLIO

TABLE OF CONTENT

MY PROFILE

I am a data analyst with backgrounds in oil & gas, sales, project management, and strong analytical and organizational skills. My technical expertise includes Microsoft excel, python, SQL, tableau, and various web platforms. I have led quality assurance and safety initiatives in automotive, bioplastics and oil & gas industries. Fluent in English and proficient in German, I possess strong leadership and communication abilities, with a passion for data analytics, teamwork, and innovation. My approach is guided by reliability, flexibility, and a commitment to continuous improvement.

Tools

MY PROJECTS

(GUN VIOLENCE DATA)

- **Context**

Gun violence is a pressing public health and safety issue in the United States. Understanding the patterns, trends, and underlying factors of gun violence are crucial for developing effective prevention and intervention strategies.

- **Tools**

Python, jupyter notebook, pandas, seaborn and matplotlib libraries

- **Key questions**

What are the patterns and factors contributing to gun violence incidents in the U.S., and how can they be categorized or predicted over time?

- **Data Set**

The gun violence dataset from Kaggle provides detailed information on gun-related incidents in the United States from 2013 to 2018. This dataset supports comprehensive analysis and helps inform strategies to combat gun violence.

- **Skills**

Geographical analysis

Exploratory analysis

Cluster analysis

Incident count

Time-series analysis

GUN VIOLENCE DATA – THE PROCESS

- **Data Preparation**

Load sample data: A sample dataset was created to demonstrate the analysis process. This data included columns for incident date, location, number of people killed and injured in each incident.

Check for missing values: We checked for missing values in the data using the isnull().sum() method. The results showed no missing values in our sample data.

- **Data Analysis**

Yearly trends: We analysed the overall trends of gun violence incidents by year. We grouped the data by the year of the incident date and calculated the sum of the number of people killed and injured for each year.

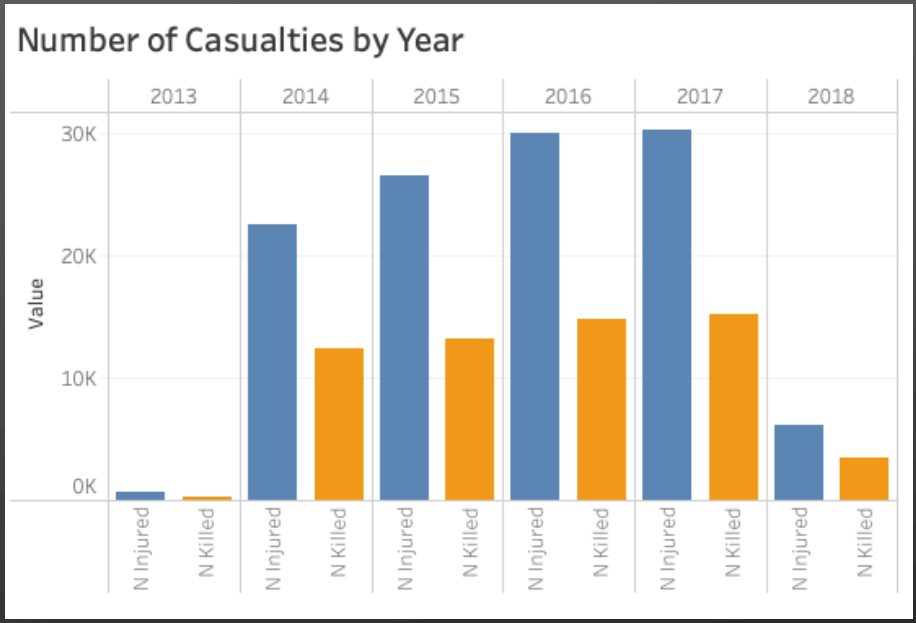
Location analysis: We analysed violence by location by grouping the data by location and calculating the sum of the number of people killed and injured in each location.

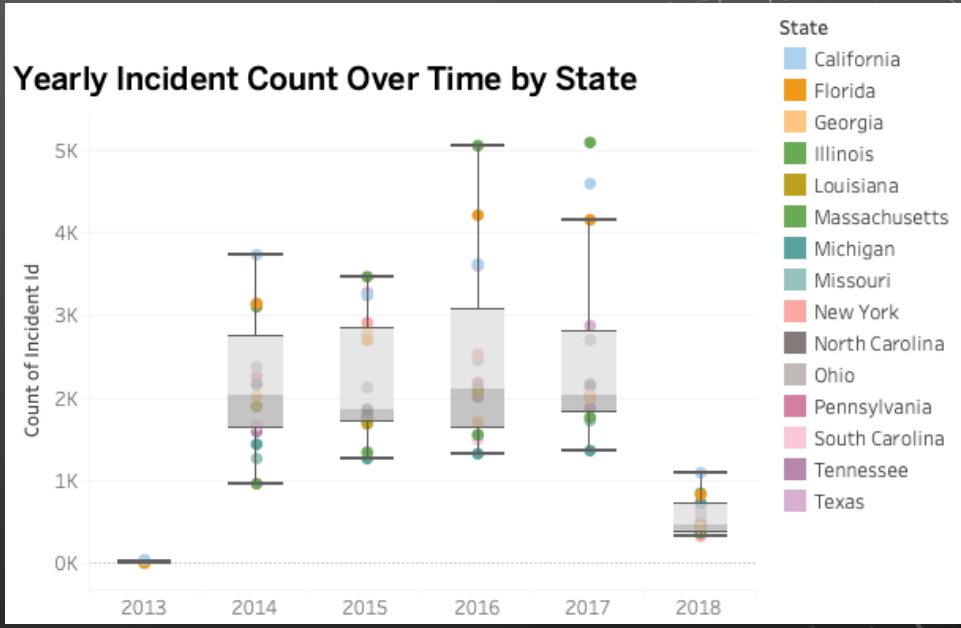
- **Results**

Yearly violence: The analysis showed yearly variations in the total number of people killed and injured in gun violence incidents.

Violence by location: The analysis revealed variations in the number of people killed and injured across different locations

GUN VIOLENCE DATA – THE ANALYSIS

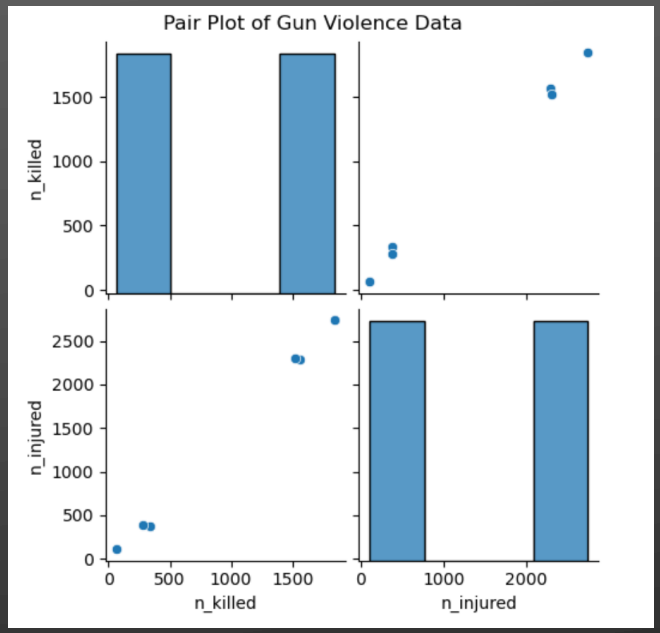




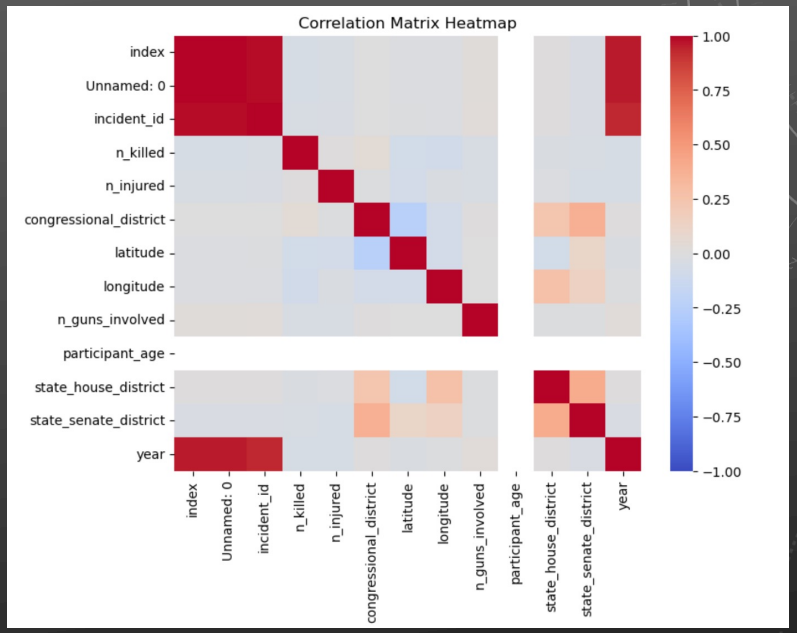
There is a sharp rise in the incident count from 2013 to 2014. This significant increase suggests a potential change in reporting practices, legislation or actual increase in gun violence incidents.

The most noticeable trend is the steep decline in incident counts from 2017 to 2018. This sharp drop could be attributed to various factors, including improved law enforcement, successful intervention programs or changes in data collection methods.

GUN VIOLENCE DATA – THE ANALYSIS



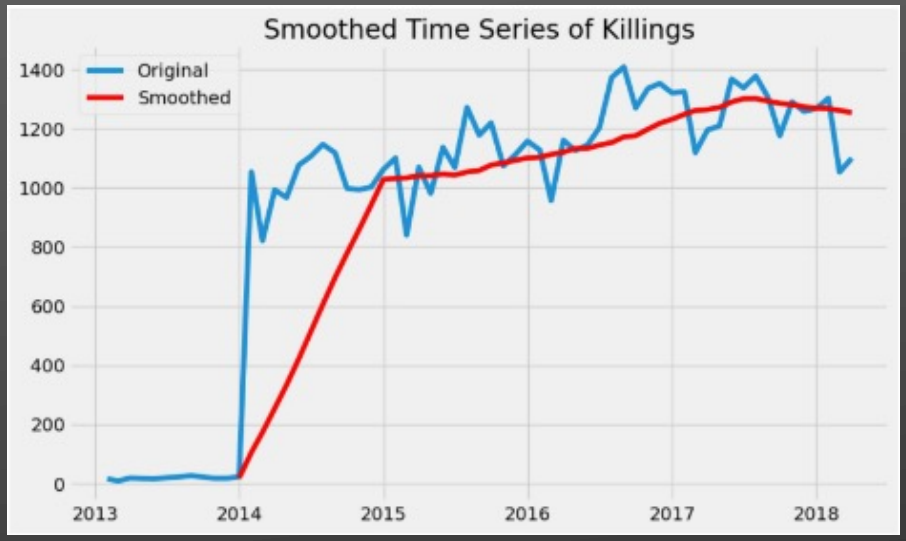
The pair plot provides a visual overview of the distribution of each variable and their relationships. The diagonal shows the distribution of each variable, while the off-diagonal plots show scatterplots of variable pairs. Variables that show strong correlations in scatterplots are worth further exploration, like n_killed and n_injured.



Strong correlations can indicate potential areas for further analysis or potential causative relationships. For example, if 'n_killed' and 'n_injured' are strongly correlated, this may warrant further investigation into factors that simultaneously affect both.

Weak correlations might suggest that other variables or non-linear relationships should be explored. By visually examining the heatmap, we can quickly identify which pairs of variables have significant relationships. This can guide subsequent analyses, such as regression modelling or hypothesis testing, to better understand the underlying patterns in the data.

GUN VIOLENCE DATA – TIME SERIES ANALYSIS

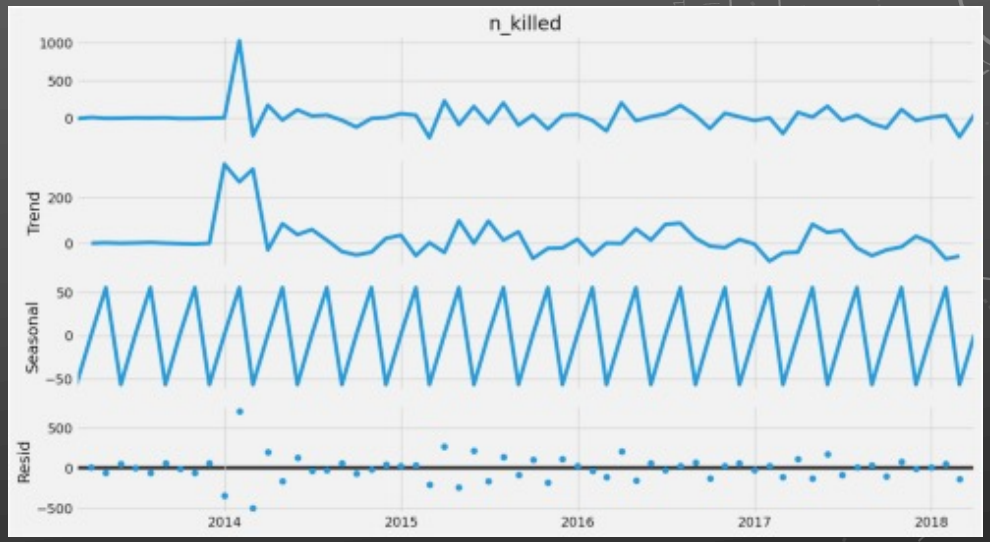


Observation

- **Original Data (Blue Line)**: The blue line represents the original time series data, which shows the raw values of killings over time.
- **Smoothed Data (Red Line)**: The red line represents the smoothed time series data, obtained by applying a rolling mean with a window size of 12.

Conclusion:

The smoothed time series helps in identifying the overall trend and cyclical patterns in the data by reducing the noise and short-term fluctuations. This can be particularly useful for gaining insights into long-term movements and making more accurate forecasts or decisions based on the data.



Observation

- The decomposition of the differenced time series helps isolate the various components contributing to the observed variability. By examining the trend, seasonal and residual components separately, we can gain a clearer understanding of the underlying patterns in the data. This decomposition is particularly useful for further time series analysis, such as forecasting or anomaly detection, as it allows us to model each component individually. The trend and seasonal components reveal the systematic structure in the data, while the residual component provides insights into the remaining unexplained variability.

GUN VIOLENCE DATA – CONCLUSION

What went well

Data Cleaning and Preparation:

- The process of cleaning and preparing the gun violence data was effective. We successfully handled missing values, dropped irrelevant columns, and converted necessary data types. This allowed for a smooth analysis process with minimal errors and inconsistencies.

Time Series Analysis:

- The time series decomposition and differencing was successful in making the series stationary, which is a key requirement for accurate time series forecasting. The seasonal decomposition provided clear insights into the underlying trend, seasonal and residual components of the data.

What Didn't Go Well

Data Challenges:

- One of the primary challenges was dealing with the large size and complexity of the dataset. The gun violence data had many variables, some of which were irrelevant or highly correlated, leading to potential noise in the analysis.

Time Series Forecasting:

- Although differencing and decomposition made the series more stationary, forecasting models were not fully developed due to the complexity of the time series data and limitations in model selection. More advanced techniques like ARIMA or SARIMA could have been explored further.

Challenges Faced

Complexity of the Data:

- The gun violence dataset was complex, with multiple dimensions that required careful consideration. Balancing the trade-off between removing noise and preserving valuable information was a constant challenge.

What Could Be Improved

Data Enrichment:

- Integrating additional data sources, such as socioeconomic or demographic data, could provide a more holistic view of the factors contributing to gun violence. Overcoming API limitations or exploring alternative data sources would be crucial in this regard.

Advanced Modelling:

- Future analysis could incorporate more advanced time series models, such as ARIMA or machine learning-based forecasting models, to improve predictive accuracy.
- In clustering, experimenting with different algorithms like hierarchical clustering or DBSCAN might yield more insightful results, especially with complex data.

Final Thoughts

- Overall, the analysis provided valuable insights into the patterns and trends of gun violence in the dataset. While the project faced some challenges, particularly with data complexity and external data integration, the findings were still significant. With improvements in data modelling, feature engineering, and external data integration, future analyses could yield even deeper insights into the factors driving gun violence and help inform policy decisions aimed at reducing such incidents.

FURTHER INFORMATION ABOUT MY WORK

- Full projects on Github:
- https://github.com/Ezeugonna

Please click on the link to open my Github projects


Thank you