

Methods Of Advanced Data Egnineering

Analysis Report

Muhammad Arbaz

Matrikel Number: 23073711



1 Introduction

Crime rates have an impact on the economy, sociology, and public policy. High unemployment rates in a city or region significantly contribute to increased crime levels, as financial stress and limited opportunities often drive individuals toward illegal activities. This project aims to investigate this relationship by analyzing two datasets: one containing crime statistics and the other providing unemployment rates across US states, to determine whether unemployment influences crime rates or not.

2 Data Sources

The datasets used in this report were generated through an ETL (Extract, Transform, Load) pipeline that processed raw data from online sources, transformed it to address quality issues, and loaded it into a target database for analysis. The first dataset, US Crime Dataset, provides detailed crime reports from various states across the USA, including crime types, victims, perpetrators, and incident details. It allows for the analysis of crime trends over time and by location. The second dataset, Unemployment in America, Per US State, captures unemployment rates by state and year, allowing insights into economic conditions across different regions. The ETL pipeline addressed critical issues such as missing values, inconsistent formats, and data type conversions to ensure the data was accurate and ready for analytical queries.

These two datasets were merged on common attributes like "state," "year," and "month" to create a unified view that enables a comprehensive analysis of the relationship between crime rates and unemployment across the United States. The merged dataset serves as the foundation for exploring correlations between economic conditions for un employment and crime patterns, providing valuable insights for policy-making and public safety strategies.

Sample records from merged dataset are shown in Figure 1.

2.1 Data Structure

The merged dataset provides comprehensive details about crime reports and unemployment rates across different states of the USA. The two data sources have been merged based on the "state," "year," and "month" columns to enable more effective analysis within a single table of records. Each row represents a record corresponding to a specific time and location within a state, combining both crime and employment statistics to facilitate deeper insights. The dataset includes key information such as the victim count, perpetrator count, incident count, total civilian population, and labor force for each state. The "month" column has been standardized into a numeric format to ensure accurate time-based aggregation.

2.2 License

The US Crime Dataset is primarily based on U.S. government data, which falls under the CCO license. The dataset is in the public domain and permits unrestricted use, allowing users to freely access, analyze, and share the data without any licensing restrictions.

The Unemployment in America, Per US State dataset is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) license. This license allows users to copy, distribute, and adapt the dataset for non-commercial purposes, as long as appropriate credit is given to the original creator, any modifications are clearly indicated, and derivative works are shared under the same license. The CC BY-NC-SA license ensures that the dataset remains open and accessible for educational and research purposes, promoting transparency and collaboration among users.

3 Analysis

The merged dataset of crime dataset and unemployment rates across U.S. states provides a rich information to explore various socio-economic

[43]:

	state	year	month	incident	victim_count	perpetrator_count	fips_code	total_civilian_non-institutional_population_in_state	total_civilian_labor_force_in_state	percent_of_sta
18466	West Virginia	1997	3	13	22	4	54	1437254	790117	
6737	Kentucky	1981	10	22	20	0	21	2681000	1655770	
3697	Florida	1994	11	265	2	4	12	11194958	6999058	
4980	Idaho	2001	4	3	0	0	16	965243	675268	
11758	New Jersey	1982	7	81	10	10	34	5714000	3641990	

Fig. 1: Sample records from merged dataset generated from ETL(Extract, Transform, Load) pipeline.

patterns. In this analysis, we aim to investigate the relationship between crime rates and unemployment trends over time, considering factors such as employment levels, population size, and regional differences. The analysis examines both general trends and specific patterns, such as the impact of unemployment on crime rates, the regional disparities in crime occurrences, and seasonal fluctuations in incidents. Respective sections present the methods used for each analysis, the corresponding results, and interpretations of the findings to get meaningful insights.

3.1 Methodology

To analyze the relationship between crime and unemployment rates across various U.S. states, we merged two datasets: the crime data and the unemployment data. The final merged table contains columns such as state, year, month, incident count, victim count, perpetrator count, total civilian population, labor force, employment, and unemployment percentages. We conducted exploratory data analysis to find patterns, correlations, and seasonal trends between these variables. The analysis were focused on understanding how unemployment rates influence crime incidents and identifying any seasonal or state-specific trends.

The analysis involved calculating key metrics like the total number of incidents and victims for each state, monthly crime trends, and the correlation between unemployment rates and crime incidents and regional and seasonal trends.

3.2 Results

We found several findings during our analysis:

Crime Incidents Count by State: The bar plot showed significant changes in average crime incidents across states. States with larger populations naturally reported higher crime incidents. Florida reported the highest number of crime incidents. The plot suggests the need for further investigation into regional factors driving crime. The bar plot of the top 10 states with higher crime rates is shown in the Figure 2.

Yearly crime and unemployment trends: The analysis shows the yearly trends of crime

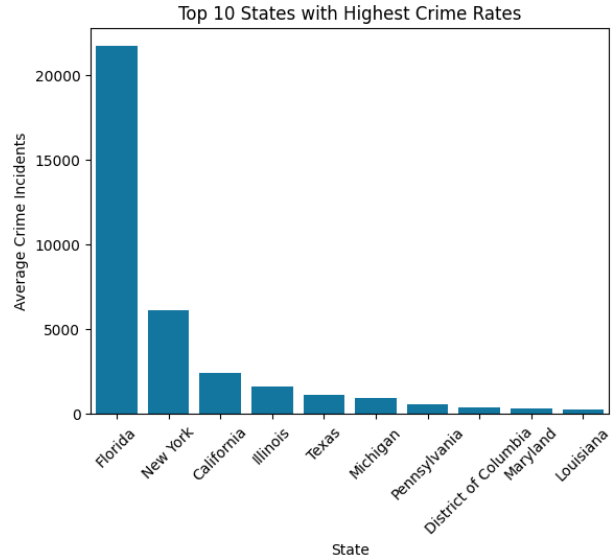


Fig. 2: The bar plot shows the top 10 states with the highest crime rates.

incidents and unemployment rates. The line plot demonstrates that crime incidents tend to increase during periods of higher unemployment after the year 2000, with some fluctuations in specific years. However, unemployment rates do not appear to have improved over the years. The yearly trend is shown in Figure 3.

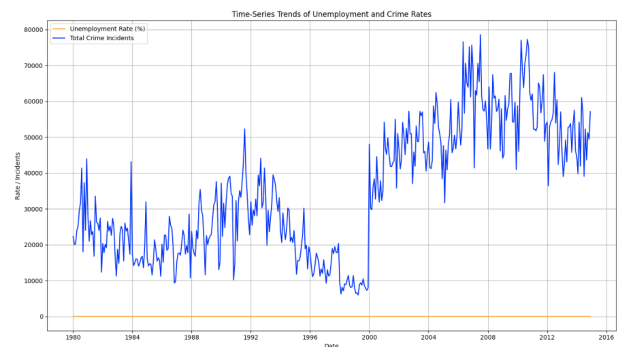


Fig. 3: The line plot shows the yearly seasonal trend of crime incidents and unemployment rates.

Crime Incidents vs. Unemployment Rate: The plot shown in Figure 4 illustrates the relationship between unemployment rates (%) and the number of crime incidents across the top five U.S. states with higher crime rates: California, Florida, Illinois, New York, and Texas. Florida

shows significantly higher and more variable crime incidents compared to the other states, as indicated by the orange line and its wide shaded area. The remaining states display lower and more stable crime rates, with minimal correlation to unemployment rates. The shaded regions around the lines represent variability or uncertainty in the data.

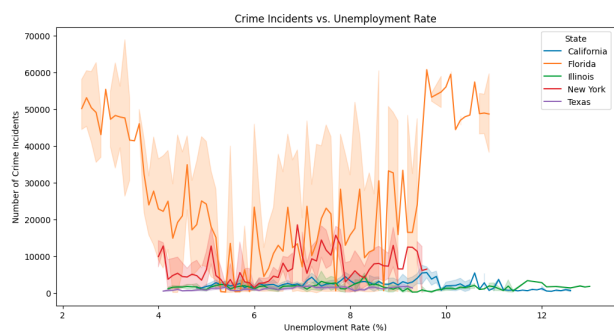


Fig. 4: The relationship between unemployment rates and crime incidents across the top five U.S. states with higher crime rates.

Regional Crime Trends and Unemployment Rate: The regional trends between crime rate and unemployment rate is shown in Figure 5. This plot shows trends in crime incidents (red lines) and unemployment rates (blue lines) from 1980 to 2015, divided by U.S. regions: Midwest, Northeast, South and West. In most regions, crime incidents increases more than unemployment rates, with the South experiencing a significant spike in crime incidents around the late 1990s to early 2000s. The Midwest, Northeast, and West show lower and relatively stable crime incident trends. Unemployment rates remain consistently lower across all regions, indicating no clear correlation with crime incidents in this timeframe. After 2000, crime incidents in most regions show a decline or remain steady, while unemployment rates remain consistently lower and stable across all regions, indicating no strong correlation between the two trends.

Heatmap Correlation: The heatmap shows the relationships between various factors, including crime-related variables (e.g., incident count, victim count, perpetrator count) and unemployment rates (percentage of the labor force unemployed in a state). It indicates relatively low correlations between unemployment rates and crime-related variables. The heatmap of the correlation plot is shown in Figure 6.

State-wise Analysis of Crime Victim Counts: The bar chart illustrates the total crime victim count for each state in the United States. The states are displayed on the x-axis, and the corresponding total victim counts are shown on the y-axis. California, Texas, and New York have the highest victim counts, while states like South Dakota,

North Dakota, and Wyoming have the lowest. The bar plot is shown in Figure 7.

3.3 Interpretation

The analysis of crime incidents and unemployment trends across U.S. states and regions provided valuable insights. States with larger populations, such as Florida, California, and Texas, exhibited significantly higher crime incident counts, with Florida reporting the highest. It shows the influence of population size on crime rates while also highlighting the need to consider socioeconomic conditions and urban density. Yearly trends revealed that crime incidents tended to increase, particularly after 2000, suggesting a indirect link between economic instability and crime. However, unemployment rates remained relatively unchanged over the years. This indicates persistent economic challenges that might indirectly contribute to criminal activity.

The relationship between unemployment rates and crime incidents in the top five states showed considerable variation. Florida showed notably higher and more variable crime rates compared to other states, indicating that state-specific dynamics such as social factors and economic policies significantly influence crime levels. Regional trends further emphasized this complexity, with the South experiencing a considerable spike in crime during the late 1990s and early 2000s, while the Midwest, Northeast, and West exhibited more stable patterns. Across all regions, unemployment rates remained consistently lower and showed minimal direct correlation with crime trends.

The findings show that unemployment might influence crime patterns. However, other variables, such as regional characteristics, cultural factors, and policy interventions, also play a role. These results highlight the importance of localized strategies that address both economic and regional differences to effectively reduce crime rates and improve community safety.

Also, the heatmap shows the less connections between crime-related factors, such as incident counts, victim counts, and perpetrator counts, and unemployment rates in different states as the correlation analysis shows weak correlations between unemployment rates and crime-related variables, indicating that unemployment alone does not strongly influence crime patterns in this dataset.

4 Conclusion

The analysis of unemployment and crime data across the different states of United States reveals a weak correlation between these variables, indicating that unemployment alone may not significantly

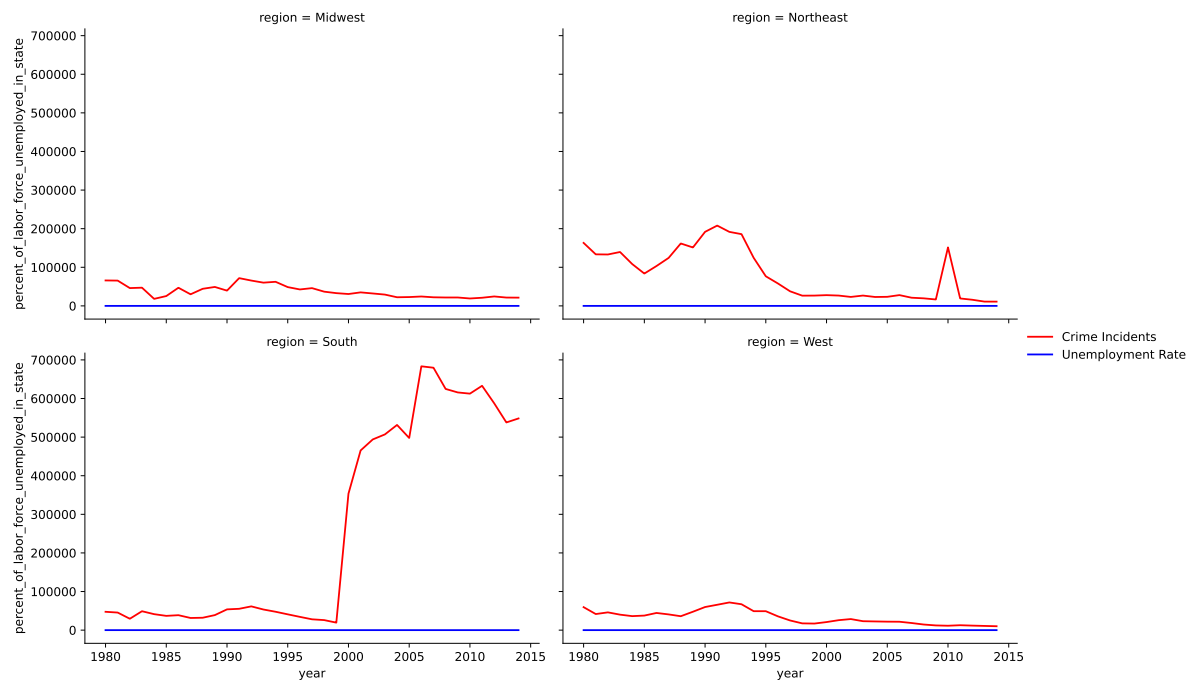


Fig. 5: Regional trends in crime rates and unemployment rates from 1980 to 2015 across U.S. regions, highlighting minimal correlation with unemployment.

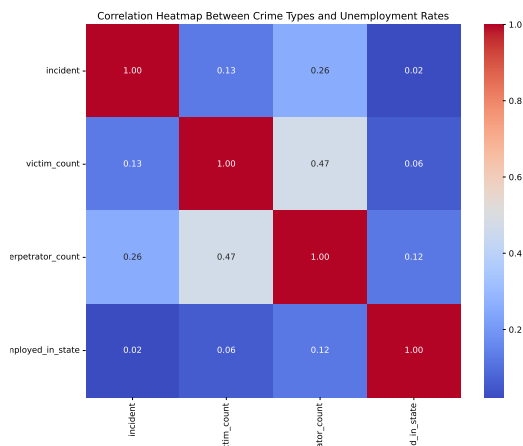


Fig. 6: Heatmap of correlations between crime-related variables (incident count, victim count, perpetrator count) and unemployment rates, showing relatively low correlations.

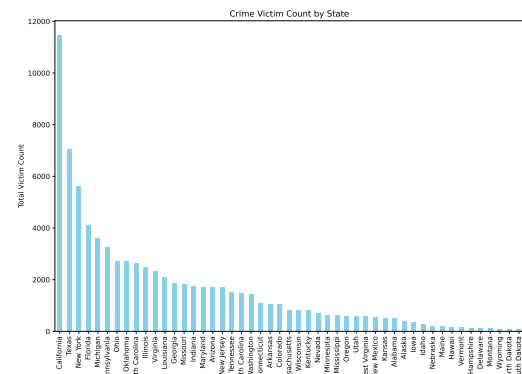


Fig. 7: Total crime victim count per U.S. state, with states on the x-axis and corresponding victim counts on the y-axis.

References

drive crime trends. States with larger populations, such as Florida, California, and Texas, report higher crime numbers, influenced by population density. Regional differences are observed, with the South experiencing notable crime spikes, while the Midwest and Northeast show more stable or declining trends, reflecting the influence of factors other than unemployment rates.

These findings reveals that crime dynamics are shaped by a combination of factors, including economic policies, educational opportunities, and social infrastructure.