Data Analysis Report

Abdul Basit

23460819

1. Introduction

The relationship between Crime Rate and Unemployment Rate has been a widely debated topic for a long time in Social Sciences. High unemployment rates often lead to a high crime rate due to economic stress and reduced opportunities. This report aims to explore the relationship between the two in the US from 1990 to 2016 and see to what extent do the two factors co-relate.

2. Used Data

2.1 US Unemployment Rate by Country, 1920-2016

This dataset represents the Unemployment Rates in the US broken down by state and month. It is under License type CCO: Public Domain. The dataset is structured as a Time Series dataset where each row represents unemployment rate for a specific state and country for a month and year. I have calculated average rate over the years (Figure 1).

Figure 1

2.2 US Crime Dataset

Each row in this dataset represents a crime instance. It has been broken down by state. It has license type **U.S Government Works** and according to **17 U.S.C. § 105** (read here), works created by U.S. federal government employees as part of their official duties are not eligible for copyright protection. As a result, such works are in the public domain. With columns like Year and Month, we can also aggregate it over the Years and see

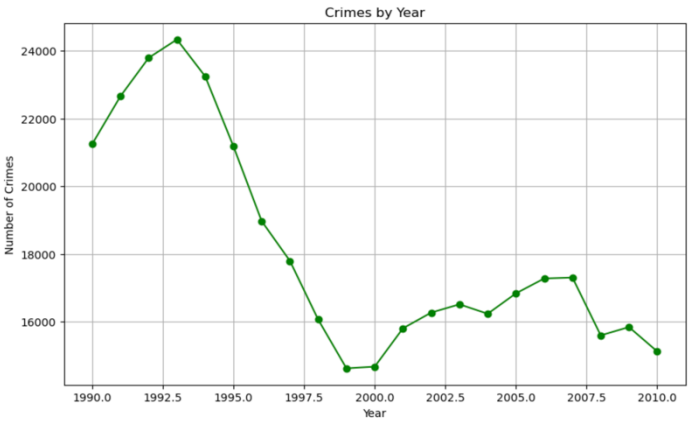

Figure 2

Number of Crime over the Years or over the Months

2.3 Data Pipeline

For the Data Pipeline, I made use of a slightly modified version of the Medallion Architecture. I did not utilize a Bronze Layer, rather I used the Silver and the Golden layers. As the data comes in from the source, I transform it and dump it into the silver layer. This layer contains two tables i.e each dataset has their own tables '**crimes**' and '**unemployments**'. And then using this clean transformed data I prepare the data ready for a specific business usecase, in this case, analysis of corelation between unemployments and crime rates. The merged data is then dumped into the golden layer in a table called '**unemployment_crime_merged**'. Transformation steps include:

- Standardized month names and added month numbers for consistency
- Handled missing data by dropping incomplete rows
- Added verifications for correct data types in Years, Rates, etc.
- Aggregated crime counts by State, Year and Month.
- Calculated Average Unemployment Rate by State, Year and Month.
- Merged Unemployments and Crimes data on State, Year and Month – Which is the final ready to use data for our usecase.

3. Analysis

This section represents the analysis of the data carried out to explore the relationship between the Unemployment and Crime Rates. The analysis is based on the merged dataset **unemployment_crime_merged.**

3.1 Distribution of Data

A histogram was generated to show the distribution of the data across all states and across all time periods. (Figure 3 and Figure 4).

- According to Figure 3, the Crime Rate is Right-Scewed. Indicating that most states have a low Crime Rate. However, there are a few outliers showing relatively high Crime Rate
- Figure 4 shows that the Average Unemploymente Rate is nearly normally distributed. So, except a few

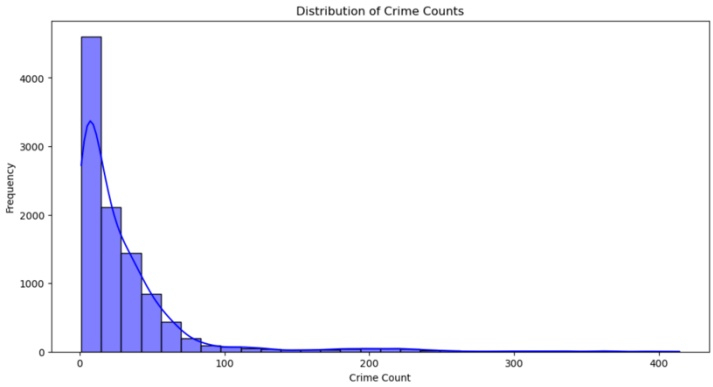

Figure 3

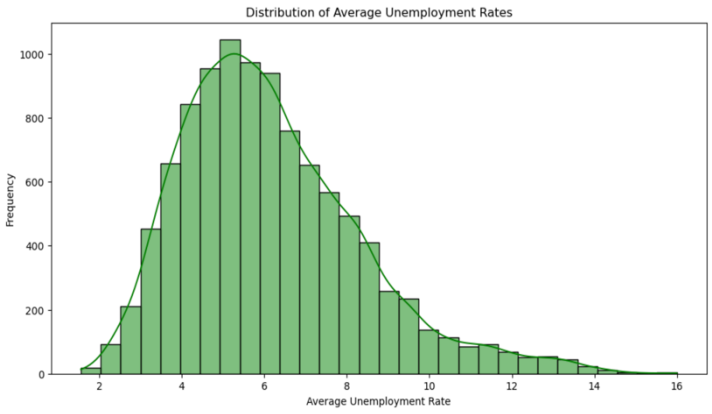

Figure 4

states with unusual high or unusual low unemployment rates, most of the states have an average unemployment rates throughout the years.

3.2 Unemployment Rate vs Crime Count in a Scatter

In this section, we see a scatter plot to visualize the relationship between Unemployment Rate and Crime Rate as a corelation. According to Figure 5, the scatter plot shows a weak corelation between the two. In technical terms, the **corelation coefficient** was found to be **0.24.** States with a high unemployment rates tend to have a higher crime rate but the positive corelation is just not that strong.

Figure 5

3.3 Time Series Analysis

In this section, we see the relationship between Unemployment and Crime Rate over a time series analysis. Until now, both our analysis results 3.1 and 3.2 negate any strong relationship between Unemployment Rates and Crimes Rates. And as shown in Figure 6, the Time Series analysis also shows that there is no clear parallel trend between the two.

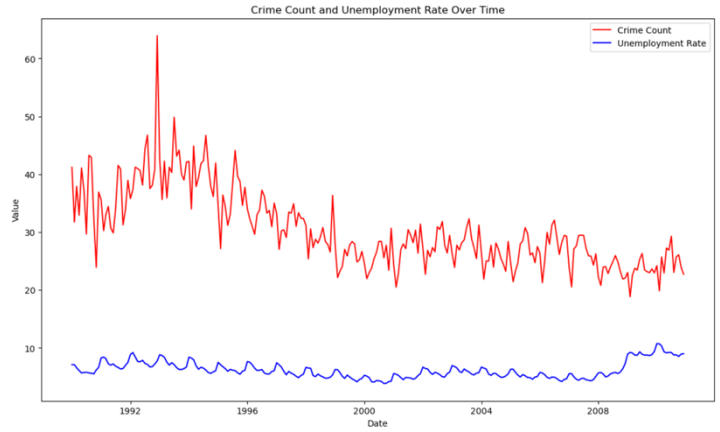

Figure 6

3.4 Grouped By State

Until now, there is no evidence that the Unemployment Rate and Crimes are positively corelated to each other. In the final step of this Analysis, I have grouped the data over State to see if I am missing something in the analysis.

But as evidence in Figure 7, even grouped by State, there is no positive corelation between the Unemployment and Crime rates data. Except for some outliers as labeled in the Data (**California, Texas and New York**) there are no states that

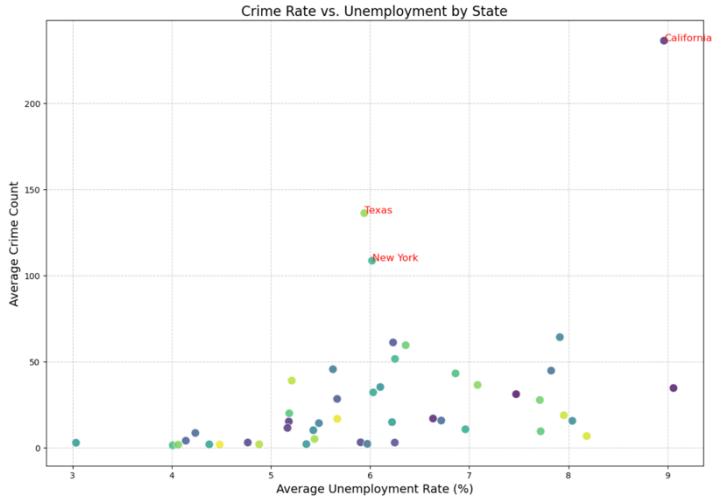

Figure 7

support the positive or negative corelation.

4. Conclusion and Findings

Based on a comprehensive analysis, I can conclude that there is not a straight forward or a strong enough relationship between Crime Rate and Unemployment. In some states, there is actually a corelation, for example, in Figure 7, the state of California shows a positive corelation with Unemployment and Crime Rate. But the greater part of the data suggests a very weak corealtion.

This shows that there are a lot of other factors at play that contribute to a state's crime rate than just the unemployment rate in the state.

4.1 Limitations

- Economical challenges such as unemployment may take time to effect the crime rates, which the analysis may not fully capture.
- The analysis was done on a dataset merged on State level, which may hide the variations in local areas.
- This analysis did not account for other social factors such as poverty rates, law enforcement, population trends, etc. These factors can highly contribute to the crime rates.

4.2 Further Research

Further work can be done as an extension of this to find out what are the factors that can contribute the the Crime Rates in addition to the Unemployment levels in a society. These can be factors such as population trends, law enforcement in the area, access to education, literacy rate, income inequality, etc.

5. References

https://www.kaggle.com/datasets/jayrav13/unemployment-by-county-us
[Unemployment Dataset]

https://www.kaggle.com/datasets/mrayushagrawal/us-crime-dataset
[Crime Rate Dataset]

https://www.copyright.gov/title17/92chap1.html - 105 [Copyright Law]

https://learn.microsoft.com/en-us/azure/databricks/lakehouse/medallion
[MS Azure – Medallion Architecture]