Daniyal Mohammad, Jay Sadrani, Payal Surana, Sadhvi Grover
Professor Isaac Vaghefi
CIS 9650
05/12/2024

Executive Summary

The Kaggle dataset that we have selected for our project is from the Global Initiative against Transnational Organized Crime to analyze crime levels and resilience in 193 countries along **3 key pillars: Criminal markets, Criminal actors and Resilience.** The goal is to help policymakers prioritize actions against crime and measure the effectiveness of their efforts. For the purpose of the Study, we decided to link this data with GDP to understand how a country's economic performance relates to different types of crime for the year 2021. We chose this because we wanted an insight into understanding what sort of relationship GDP of countries have on various types of Crimes factors and Resilience factors.

First, we set up Zoom and Google Colab for teamwork as instructed. Using the Kaggle API, we imported our dataset and unzipped the files. Then, we decided to link two datasets with one of its common criteria of Country. Read, understood & analyzed both Excel files and started working on the jupyter notebook using pandas and numpy. After checking the rows and columns, we decided to clean up the data. We started by ensuring the country names in both datasets matched. We compared the country names in the Organized Crime Dataset with those in the GDP per capita dataset. We found and corrected some spelling errors. Later, we replaced the old country names with the new ones in the GDP per capita dataset, making sure both datasets had consistent country names.
In the next step, we renamed the "Country/Area" column in the GDP dataset to simply "Country" to align it with the naming convention in the Organized Crime dataset and updated GDP per capita values to two decimal places for better understanding of data. Rest, everything we felt was in place to take it to the next level of visualization of data. We merged the GDP and Organized Crime dataset. We named the merged dataset as “Organized_data" for ease of use in further study. Later, we re-arranged the Countries alphabetically (A-Z) so that our data is well prepared for our next step of analysis.

To begin our analysis, we used the describe function in our Organized dataset to provide an overview of each column such as the total count, the mean value, standard deviation, the minimum value, and the maximum value. For example, using describe, we learned that the mean for Criminal markets in the 2021 period was around 4.87,with a standard deviation of 1.32. These findings further helped us in our visualization graphs as well.
For visualization, we explored and used both seaborn and plotly. First, from the plotly scatterplot we learned that GDP and Criminal Markets have a weak correlation which means that criminal market ratings vary widely across countries, irrespective of GDP per capita. Second from the bar chart, we learned about the Top 10 Countries with Highest Heroin Trade and Lowest Human Trafficking Crime. Third from the Regplot in seaborn we understand a potential negative

relationship between Criminality and Resilience indicating that regions with higher criminality might exhibit lower resilience. We understood the outlier topic by using a boxplot**.**
However, the scattered distribution of data points implies that resilience scores can be influenced by various factors beyond just criminality. Next, we used Seaborn's FacetGrid to generate a scatter plot grid, examining the association between criminality and GDP per capita across various continents. In Africa, Criminality varies despite the same GDP except for an outlier. Whereas, Europe shows that the higher the GDP, the lower the rate of criminality. In contrast to the Oceania Continent, the rest of the continents i.e Asia and Americas have a higher rate of criminality and low GDP.
We also put together a linear regression table to analyze the relationship between GDP per capita with Criminal Markets, Criminal actors and Resilience. The regression model explains about 35.9% of the variance in the dependent variable, GDP, Per Capita as indicated by the R-squared value. The overall model is statistically significant, as indicated by a low p-value associated with the F-statistic 3.48e-18. Further, since p value of Resilience is 0, it shows that Resilience has significant impact on GDP per capita whereas impact of other factors like Criminal Markets, Criminal actors is less significant.

Overall, there is a lot to consider when looking at this dataset. There could potentially be many other factors that impact Average Organized Crime in Countries. We had a lot of countries in our dataset which sometimes made our visualization graph choices very restricted. Despite all this, we found this dataset to have an interesting perspective on Organized Crime. If we continue analyzing this data, we will carry out similar exercises for the 2023 period and compare the trends.