Exploring the Correlation Between Unemployment and Crime Rates in MERCOSUR Countries Over Two Decades (2000-2020)

Raj Sinha, 23196502

Friedrich-Alexander-Universität Erlangen-Nürnberg

Introduction

Unemployment is often linked to crime, as economic instability can drive individuals toward unlawful activities. This study examines the relationship between unemployment and crime rates in MERCOSUR countries (Argentina, Brazil, Paraguay, Uruguay, Colombia ,Guyana). By analyzing data on unemployment and crime, the research aims to uncover patterns and correlations, offering insights to help policymakers address unemployment-driven crime effectively.

1. Main Analytical Questions:

1. What is the strength and direction of the correlation between unemployment and crime rates in MERCOSUR countries?
2. Is there evidence of causation between unemployment and crime, or is the relationship merely correlational?
3. Does the relationship between unemployment and crime differ across geographic areas within the MERCOSUR region?
4. How have the trends in unemployment and crime rates evolved over time, and are there observable patterns or lagged effects between the two variables?

2. Datasets

The Project utilizes two datasets from the World Bank to provide comprehensive and standardized data for exploring the relationship between unemployment and crime rates across MERCOSUR (Argentina, Brazil, Paraguay, Uruguay, Colombia, Guyana) countries. The zipped CSV format ensures easy handling and compatibility with various analytical tools. The accompanying metadata provides detailed descriptions of indicators, methodologies, and data sources, ensuring transparency and clarity for analytical purposes.

2.1. Unemployment Rates Dataset

- Data : https://api.worldbank.org/v2/en/indicator/SL.UEM.TOTL.ZS?downloadformat=csv
- Meta Data: https://databank.worldbank.org/reports.aspx?source=2&type=metadata&series=SL.UEM.TOTL.ZS
- Data Type: Zipped
- CSVLicense: Creative Commons Attribution 4.0 International

This dataset provides annual unemployment rates as a percentage of the total labor force, modeled by the International Labour Organization (ILO). It includes data for countries globally, allowing for cross-country comparisons.

2.2. Crime Rate Dataset

- Data URL: https://api.worldbank.org/v2/en/indicator/VC.IHR.PSRC.P5?downloadformat=csv
- Meta Data: https://databank.worldbank.org/reports.aspx?source=2&type=metadata&series=VC.IHR.PSRC.P5
- Data Type: Zipped
- CSVLicense: Creative Commons Attribution 4.0 International

This dataset records annual crime rates measured by the number of intentional homicides per 100,000 people. It serves as a key indicator for analyzing crime levels across different regions.

2.3 Data License

Both datasets are licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license link. Under this license, it allows to freely use the data by copying, sharing, and modifying. So, I am using the datasets from the World Bank and transforming those datasets for our project. Giving proper credits by acknowledging the World Bank as the source of the data and providing a link to the license. This ensures that others know where the data originated from and can access the original source for further information or verification. By complying with these terms, we contribute to open access and encourage the continued sharing and use of valuable data for research, analysis, and innovation. I will make sure that the World Bank receives proper credit in all the reports and documents.

3. Data Pipeline

This section gives a thorough overview of the pipeline, explaining its main parts, the tools used, the changes made to the data, and how we deal with problems or changes in the data's format.

- **pipeline.py** : This orchestrating module brings together the extract, transform, and load components into a cohesive automated pipeline. It defines the sequence of execution, ensuring that each step is performed in the correct order and that dependencies are met.

The pipeline consisted of 3 parts:

- **extract.py**: This Python code is responsible for extracting data from the original sources. The International Labour Organisation and the UN Office on Drugs and Crime's International Homicide Statistics database served as the primary sources for unemployment and crime data, respectively.

- **transform.py** : The transformation step in the pipeline presented a significant challenge, especially with the crime dataset. Addressing issues such as empty or extensively null values in the crime data required meticulous handling. To ensure a uniform dataset for analysis, the shape of both the crime and unemployment datasets was aligned. Specifically, data from 1960 to 1990 (where unavailable) thus taken from data from 1991 to 2021 into account and were transformed into a consistent format, facilitating seamless exploration and comparison. After 1st round of transformation around 70 countries remained the same in both the crime and unemployment dataset.

- **load.py** : After data transformation, the load module takes charge of storing the processed data in a CSV file. This step ensures that the cleaned and standardized dataset is readily available for subsequent exploration and analysis. The CSV format is chosen for its simplicity, widespread compatibility, and ease of use in various analytical tools.

3.3 **Challenges and Solutions**

- Inconsistent Formats: Resolved using parsing scripts for uniformity.
- Missing Data: Imputed gaps or excluded unreliable records.

3.4 **Meta-Quality Measures**

- Error handling with logs and retries.
- Data validation (completeness, duplicates).
- Adaptability for changing input formats.
- Automated alerts for anomalies.

This pipeline ensures accurate, efficient, and scalable processing for analyzing unemployment and crime data.

4. Analysis

4.1. Chosen Data Format for the Output of the Pipeline

After transforming data, the unemployment and crime data look like this. We then selected target countries in the later stage of our evaluation.

Country Name	1991	1992	1993	1994	1995	1996	1997	1998	1999	...	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
0 Albania	10.307	30.015	25.258	20.840	14.611	13.931	16.876	20.047	20.840	...	13.380	15.870	18.050	17.190	15.420	13.620	12.30	11.470	13.067	12.683
1 Armenia	1.600	1.800	5.300	6.600	6.700	9.300	10.800	9.400	11.200	...	11.298	11.531	11.862	12.177	12.426	12.944	13.21	12.200	12.180	12.729
2 Australia	9.580	10.730	10.870	9.720	8.470	8.510	8.360	7.680	6.870	...	5.220	5.660	6.080	6.050	5.710	5.590	5.30	5.160	6.460	5.120
3 Austria	3.420	3.590	4.250	3.540	4.340	5.280	5.150	5.480	4.700	...	4.870	5.340	5.620	5.720	6.010	5.500	4.85	4.490	5.360	6.180
4 Azerbaijan	0.900	1.800	4.500	6.300	7.200	8.100	9.100	10.000	10.900	...	5.190	4.970	4.910	4.960	5.000	4.960	4.94	4.850	7.160	5.950
5 Benin	1.484	1.510	1.396	1.358	1.237	1.175	1.083	1.018	0.927	...	2.454	2.254	2.111	2.023	1.836	1.643	1.47	1.466	1.671	1.752
6 Bulgaria	11.100	15.300	16.400	12.800	11.100	12.500	13.700	12.200	14.100	...	12.270	12.940	11.420	9.140	7.570	6.160	5.21	4.230	5.120	5.270
7 Bahamas, The	12.170	14.810	12.720	13.830	10.830	11.460	9.760	7.650	7.450	...	14.020	16.180	13.800	12.000	12.700	9.800	10.00	10.051	12.867	11.665
8 Brazil	6.366	6.420	6.030	6.470	7.090	8.030	9.000	10.150	11.130	...	7.250	7.070	6.760	8.560	11.740	12.930	12.46	12.050	13.930	13.340
9 Barbados	22.683	22.940	24.540	22.980	19.660	14.710	14.610	12.400	10.610	...	11.590	11.550	12.170	9.800	8.250	8.620	8.32	8.410	10.105	9.582

Table 1: Sample Output data table for unemployment dataset after Transformation

Country Name	1991	1992	1993	1994	1995	1996	1997	1998	1999	...	2012	2013	2014	2015	2016	2017
0 Albania	7.469486	3.934936	5.780633	3.400123	7.885849	8.314658	40.999238	20.419683	16.582197	...	5.428407	4.295095	4.646160	2.220310	2.742043	2.014340
1 Armenia	5.141486	8.532529	7.751603	4.653626	3.491050	3.455700	3.056778	2.869883	2.807211	...	2.230289	2.171377	2.491410	2.605438	3.035764	2.454484
2 Australia	1.968608	1.723693	1.888664	1.797197	1.977448	1.943790	1.977159	1.795538	2.047775	...	1.069106	1.060065	1.035383	0.990754	0.936181	0.845861
3 Austria	1.528066	1.237116	1.011982	1.108960	0.981466	1.244025	0.828408	0.965413	0.750809	...	0.960896	0.719379	0.491454	0.509117	0.618297	0.795681
4 Azerbaijan	4.735836	7.378975	6.175625	6.379541	5.615179	5.032015	3.549973	3.482113	2.987120	...	2.129549	2.298617	2.422530	2.139204	1.694024	1.807067
5 Benin	7.469486	8.189933	8.509661	10.776662	10.797970	10.087810	9.990271	8.835399	8.600639	...	8.910891	8.430357	8.631802	8.224460	8.945892	1.127497
6 Bulgaria	4.299640	5.329617	5.690042	5.755419	5.827135	5.213432	4.784249	4.626635	4.006223	...	1.882926	1.466795	1.519233	1.723842	1.090261	1.434055
7 Bahamas, The	10.142814	14.540425	12.345840	17.687285	13.351185	16.725013	14.830553	8.835399	8.600639	...	28.791214	30.856995	31.351910	37.178744	28.031967	30.574909
8 Brazil	18.357327	16.656003	17.680355	18.543499	20.793213	23.424882	24.033101	24.526455	24.736250	...	28.197143	28.179904	29.358647	28.356406	29.589154	30.593037
9 Barbados	8.095543	8.454628	8.046116	7.642003	6.481042	10.087810	9.990271	7.586059	8.706531	...	7.963315	8.668485	9.009236	11.147751	7.895237	10.745486

Table 2: Sample output data table for crime dataset after transformation

4.2 Long Term Trends

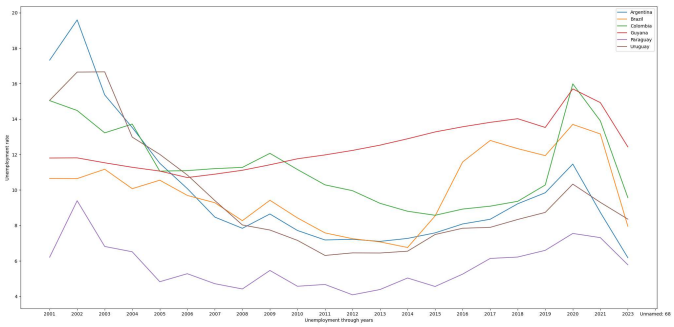

Fig. 1. a) Line plot of unemployment of different countries over years

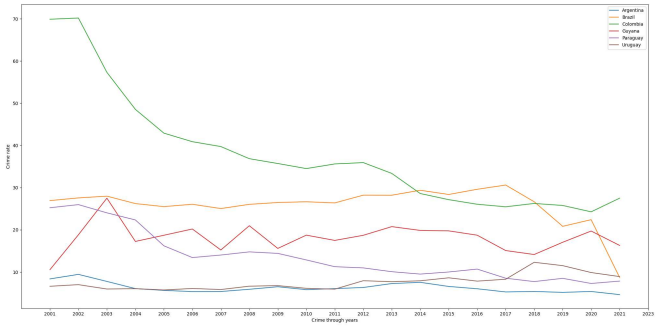

Fig. 1. b) Line plot of crime of different countries over years

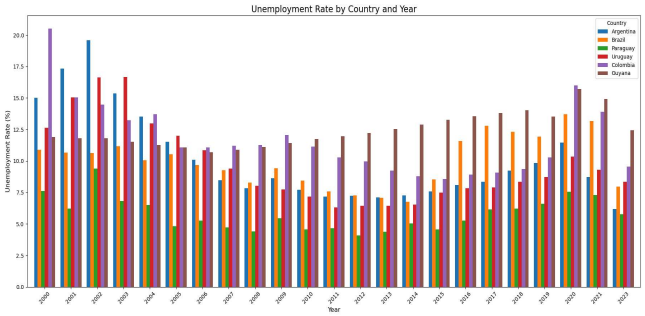

Fig. 2. a) Bar plot of unemployment rate of different countries

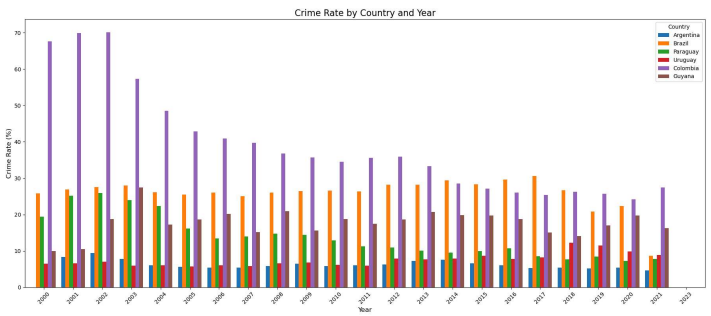

Fig.2. b) Bar plot of the crime rate of different countries

- **Paraguay** consistently demonstrates the lowest and most stable unemployment rates across all years, highlighting its economic steadiness.
- **Argentina and Colombia** show significant improvements, with steep declines in unemployment from high levels in the early 2000s.

- **Guyana and Brazil** exhibit more volatility, with notable fluctuations in unemployment rates over the years.
- Uruguay maintains moderate rates, showing economic resilience, especially in the 2010s.
- The global impact of economic events, such as the 2008 financial crisis and the 2020 pandemic, is reflected in rising unemployment rates during these years across most countries.
- **Colombia** stands out with the highest crime rates initially but has shown a remarkable improvement over the years.
- **Paraguay** and **Argentina** have the lowest and most stable crime rates.
- **Brazil** and **Uruguay** show moderate crime levels with gradual improvements.
- **Guyana** demonstrates notable variability in crime trends without a consistent pattern.

4.3 Heatmap of unemployment and crime

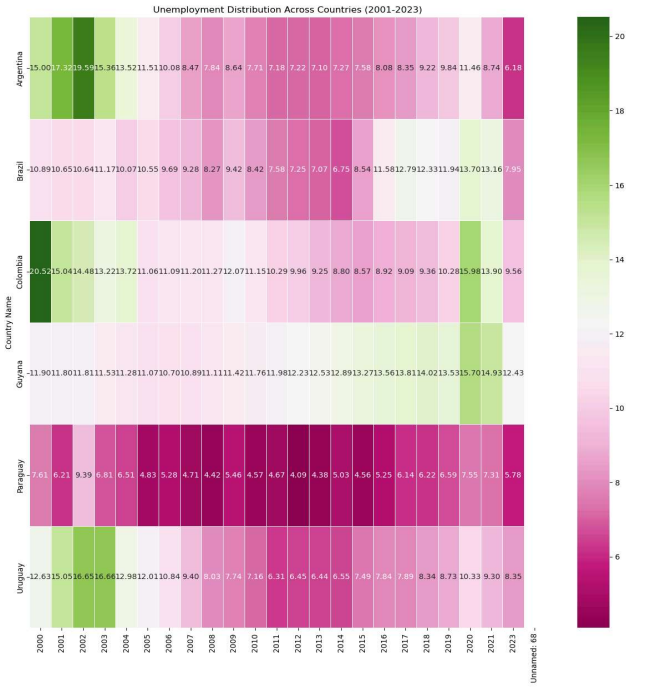

Fig. 3. a)Heatmap for Unemployment data for different countries

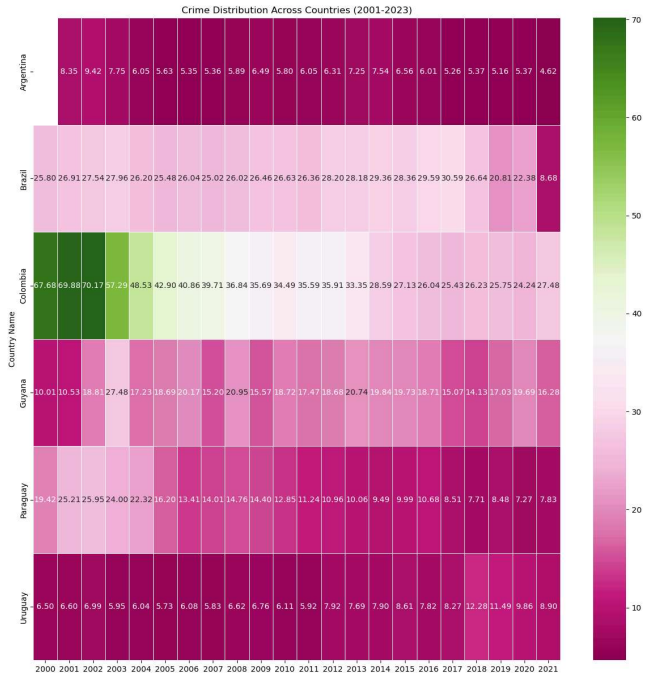

Fig.3. b) Heatmap for crime data for different countries

Unemployment

Significant Improvements: Argentina shows the most substantial improvement, with unemployment rates dropping dramatically over two decades.
Stable Trends: Paraguay maintains consistently low unemployment rates, highlighting economic stability.
Moderate Fluctuations: Brazil, Guyana, and Uruguay exhibit moderate fluctuations, with recent improvements post-2020.
Persistent Challenges: Colombia's rates have improved but remain higher than other countries, indicating ongoing economic challenges.

Crime

Colombia: Despite starting with the highest crime rates, it has shown remarkable improvement over the years.
Argentina and Uruguay: Stand out for their low and stable crime rates, reflecting a relatively safe environment.
Brazil: Shows moderate crime rates with some fluctuations, indicating room for improvement.
Paraguay and Guyana: Exhibit steady and relatively low crime rates, with minor fluctuations over time.

4.4. Variablity in unemployment and crime rate of different countries using Violin Plot

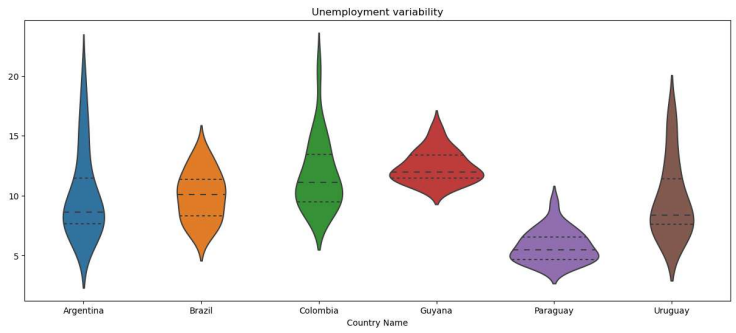

Fig. 4. a) Unemployment Variability

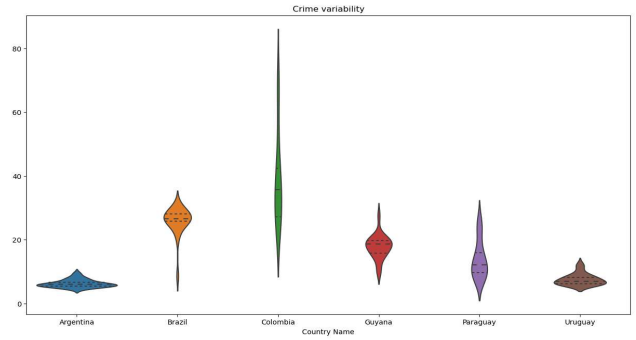

Fig. 4. b) Crime rate variability

Unemployment:

High Variability: Argentina stands out with the most variability, indicating significant fluctuations in unemployment over the observed period.
Stability: Paraguay and Guyana show the most consistent unemployment rates, reflecting stability in their labor markets.
Moderate Variability: Brazil, Colombia, and Uruguay exhibit moderate variability, with unemployment rates showing some fluctuations but remaining relatively balanced.

Crime:

High Variability: Colombia's wide range suggests notable disparities or fluctuations in crime rates.
Stability: Argentina, Uruguay, and Guyana have the most stable crime patterns, with tightly clustered data.
Moderate Trends: Brazil and Paraguay show moderate fluctuations but remain relatively consistent.

4.5 Correlation between unemployment and crime

1.Unemployment and Crime Relationship for Different Countries over the years

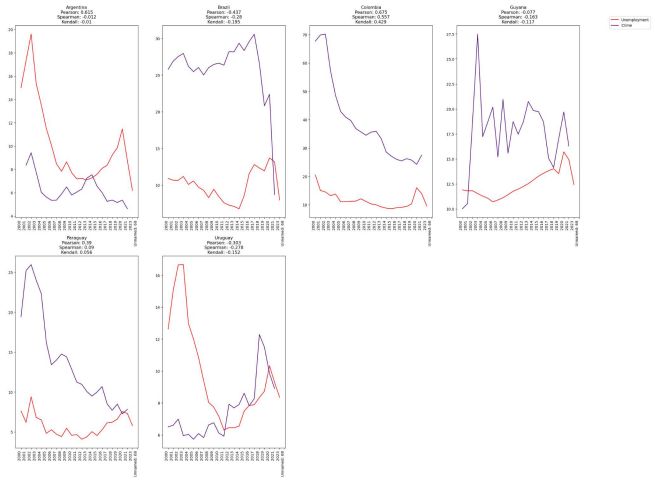

Fig. 5. a) Unemployment & Crime correlation aggregated over the years

2.Unemployment and Crime Relationship for countries in specific year

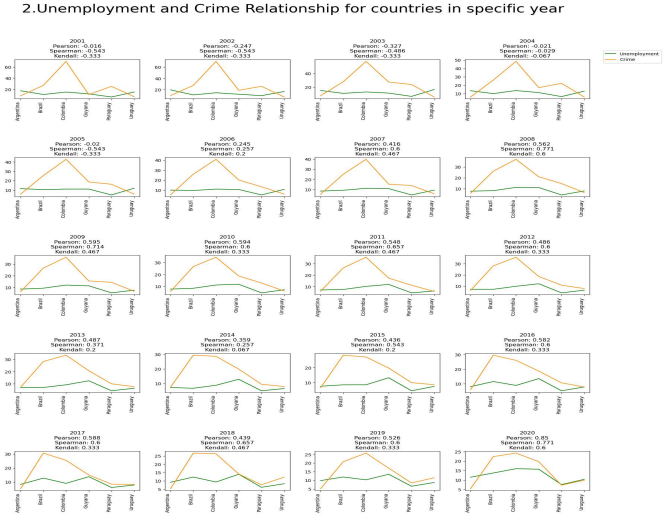

Fig. 5. b) Unemployment & Crime correlation country-wise

- **Positive Correlation**: Argentina and Colombia show that crime decreases with unemployment.
- **Negative Correlation**: Brazil and Uruguay indicate slight crime stabilization or reduction with lower unemployment.
- **Weak Trends**: Guyana and Paraguay show little relationship between unemployment and crime.
- **Strength of Correlation**: The relationship between crime and unemployment strengthens over time, particularly post-2010.
- **Country-Specific Trends**: Colombia and Argentina show the most consistent alignment between the two variables, while Brazil and Paraguay display more variability.
- **Economic Impact**: Years with economic disruptions (e.g., 2020) exhibit stronger correlations, suggesting that unemployment strongly influences crime during such periods.

5. Conclusions:

- Countries like **Colombia** exhibit a strong positive relationship between unemployment and crime, suggesting economic factors significantly influence crime levels.
- **Brazil** and **Uruguay** show a moderate negative relationship, where decreasing unemployment corresponds with declining crime rates.
- **Guyana** and **Paraguay** reflect weak correlations, with stable crime rates less influenced by unemployment fluctuations.
- **Argentina** demonstrates a mixed relationship, indicating additional socio-economic or political factors may play a role in influencing crime trends.

Thus, we can conclude that it is not necessary that if unemployment is high then crime rate will be high and vice versa. Both rates depend on various things like socio-economic factors, political factors, education,, etc.

6. Limitations

Addressing unemployment and crime data limitations can deepen insights into their relationship and inform policy.

Unemployment Data:

- Seasonal jobs (e.g., agriculture) and unpaid caregiving roles often go unrecorded.
- Individuals not seeking work due to barriers (e.g., discrimination or mobility) are excluded.

Crime Data:

- Variations in homicide data sources and societal definitions affect accuracy.
- The analysis might overlook regional variations within countries

Causation vs. Correlation:

Correlation doesn't imply causation; other factors may influence both unemployment and crime.

7. Future works

- **Multifactor Analysis:** Study how policies, education, and the economy jointly impact unemployment and crime.
- **Policy Impact:** Evaluate policies to guide crime prevention strategies.
- **AI Methods:** Apply machine learning to uncover complex patterns.