

Exploring CO₂, GDP, and Electricity Access Across Eight South American Countries

Introduction

This little report explores three indicators : CO₂ emissions per capita, GDP (constant 2015 US\$), and access to electricity across eight South American countries from 2016 to 2020.

The objective is to understand how these indicators vary across countries and time, investigate potential correlations, and identify patterns of growth, difference, and consistency.

All visualizations were created using Python (pandas, seaborn, matplotlib), and the dataset was cleaned and structured manually as part of this data science mini project.

Dataset Overview

The dataset was collected from the [World Bank SDG database](#), which provides global development indicators aligned with the Sustainable Development Goals (SDGs).

For this project, data from **eight South American countries** was selected: **Argentina, Brazil, Chile, Paraguay, Uruguay, Bolivia, Peru, and Ecuador.**

The time frame was limited to **2016 – 2020**, and the analysis focused on the following three indicators:

- **CO₂ emissions (metric tons per capita)**
- **GDP (constant 2015 US\$)**
- **Access to electricity (% of population)**

Each country has yearly data for each indicator, resulting in a small but manageable panel dataset suitable for exploratory analysis.

Note: The data of CO₂ emissions for all countries in the year 2020 are missing. These missing values were preserved and handled appropriately during the visualization process.

Data Cleaning Summary

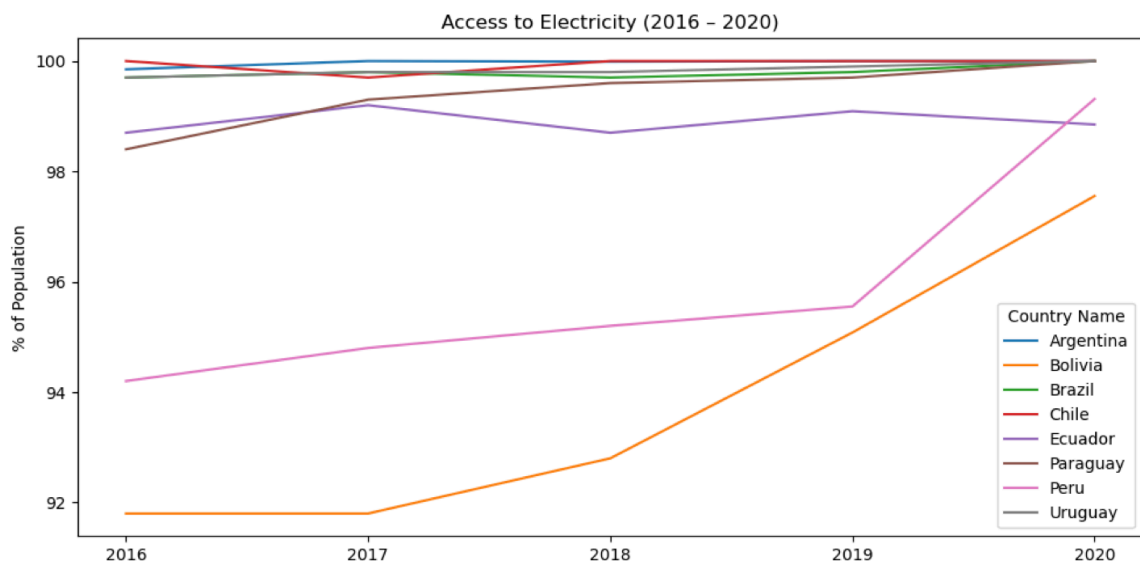
The raw dataset initially included 29 rows, containing a mix of actual indicator values and metadata. These were filtered to retain only the **24 valid rows** (8 countries × 3 indicators).

The following cleaning steps were performed:

- Removed empty rows and metadata entries
- Dropped redundant columns (e.g., Country Code , Series Code)
- Extracted year values and converted them to integers
- Standardized missing values (e.g., converted non-numeric strings to NaN)
- Verified data types and consistency across columns
- Restructured the dataset using `melt()` and `pivot_table()` to achieve a tidy format.
- Computed **GDP growth rates (%)** from 2017 to 2020 using `pct_change()`

By the end of this stage, the dataset was fully prepared for visualization, with clearly defined structures and preserved missing values.

Exploratory Visualizations



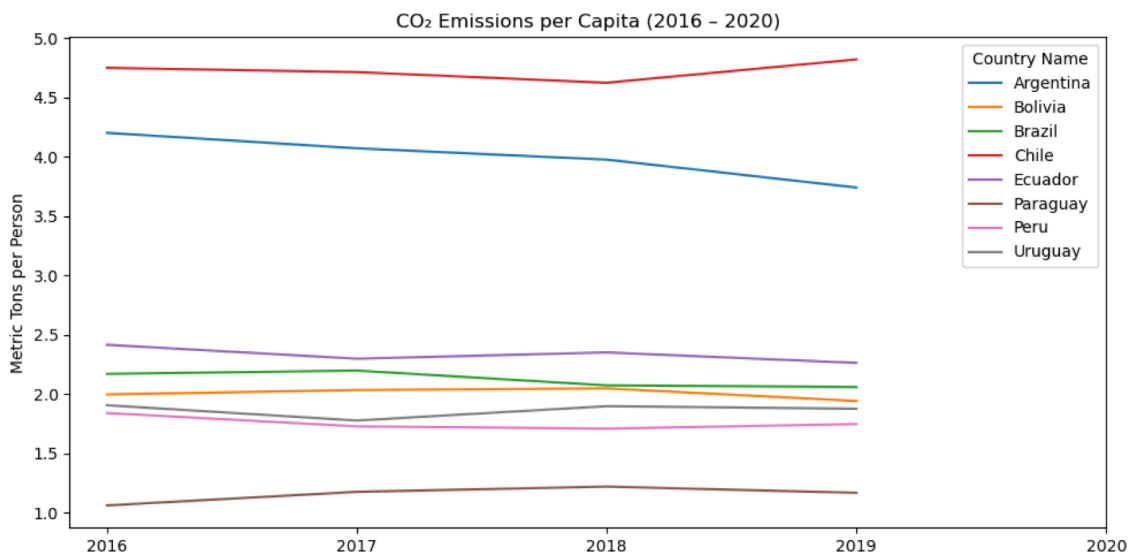
Observations:

- All countries had access rates above 92% starting in 2016.
- Bolivia and Peru showed steady improvements over the five years.
- Ecuador maintained a high level between 98% – 99%, with slight yearly fluctuations.
- Paraguay and other countries reached nearly full coverage (100%) by 2020.

Summary:

South America has made strong progress in universal electricity access.

By 2020, most countries achieved or nearly achieved full national coverage.

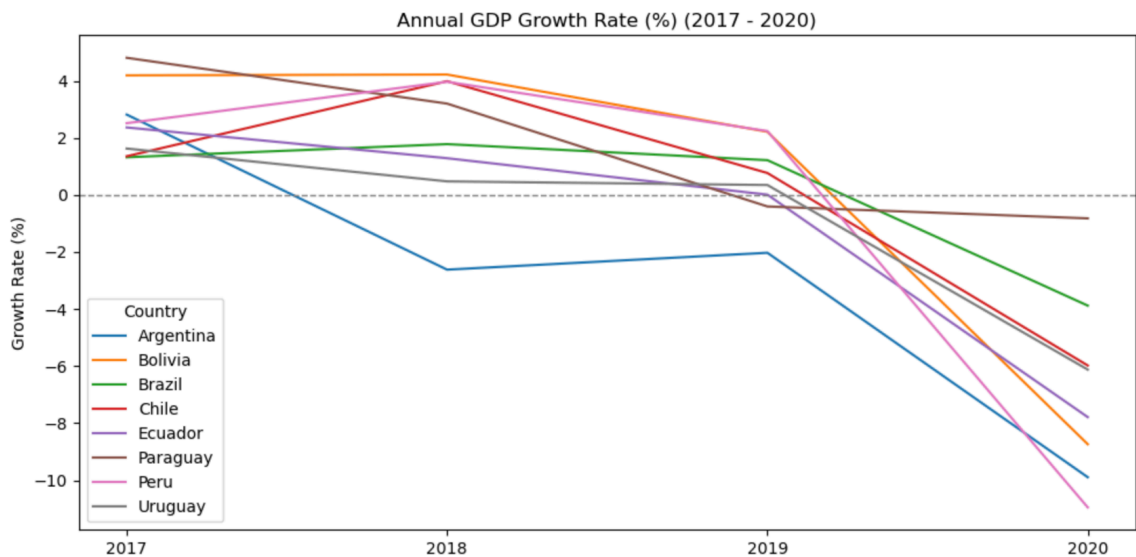


Observations:

- Chile had the highest per capita emissions, with a rising trend.
- Argentina ranked second but showed consistent decline.
- Paraguay remained the lowest and most stable.

Summary:

Argentina and Chile exhibit the most distinct patterns in CO₂ emissions. Paraguay remains consistently low, while other countries are relatively stable.



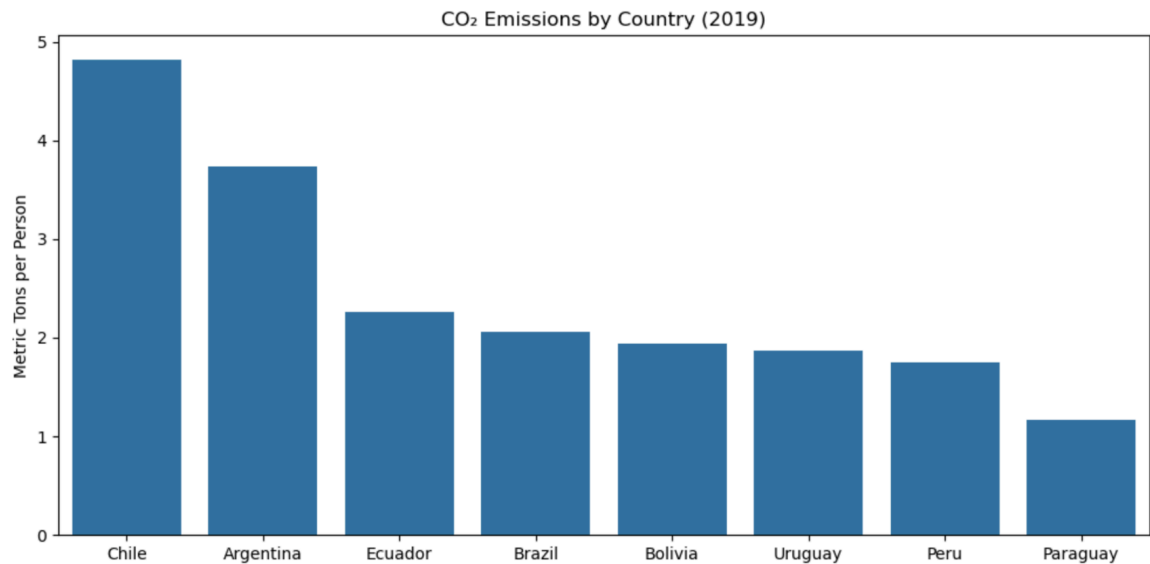
Observations:

- Argentina showed negative growth from 2017 onward, worsened by the 2020 pandemic.
- All other countries declined in 2020, reflecting pandemic impact.
- Paraguay experienced the least contraction, hovering close to 0%.

Summary:

Most South American countries experienced pandemic-driven economic contraction in 2020.

Argentina faced prolonged recession, while Paraguay showed relative stability.

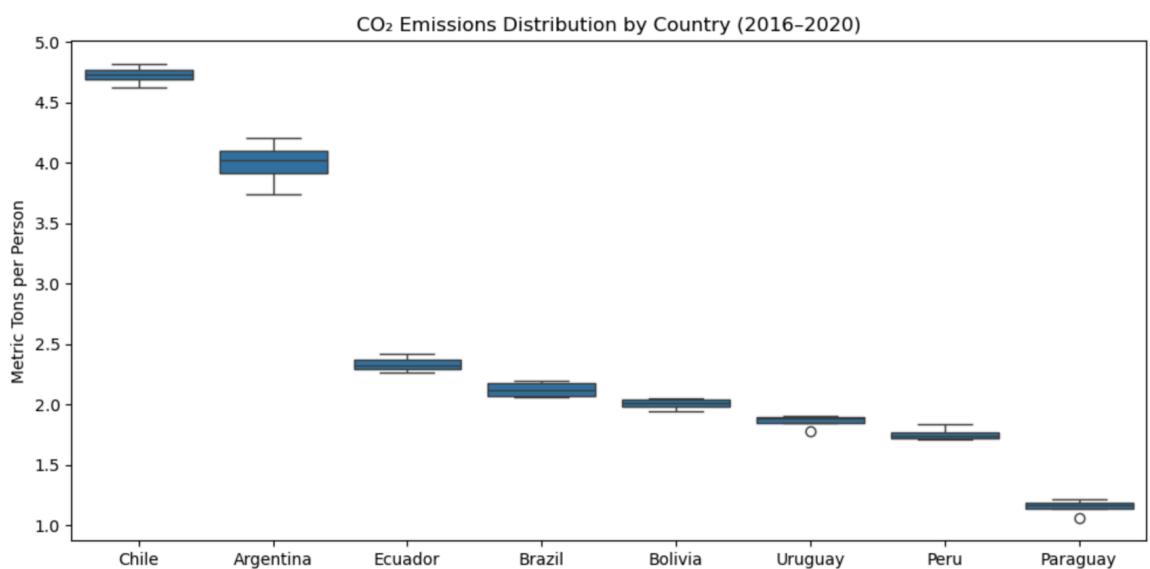


Observations:

- Chile (around 5 tons) and Argentina (around 4 tons) had the highest emissions.
- Other countries clustered around ~2 tons.
- Paraguay emitted just over 1 ton, creating a drop from others.

Summary:

Emission levels follow economic scale, but Paraguay stands out as an outlier.

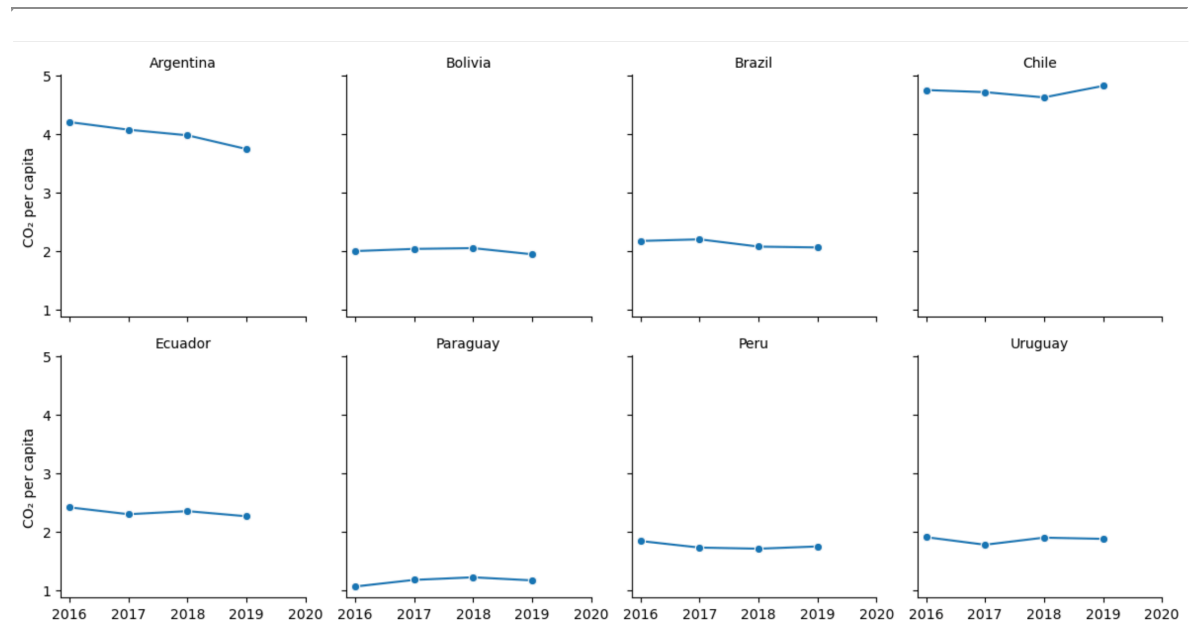


Observations:

- Country ranking was consistent across five years, with no overlaps.
- Most boxplots are compressed, indicating low variation.
- Argentina had the widest spread (around 0.5 ton difference).

Summary:

CO₂ emissions were generally stable, with Argentina showing most variation.

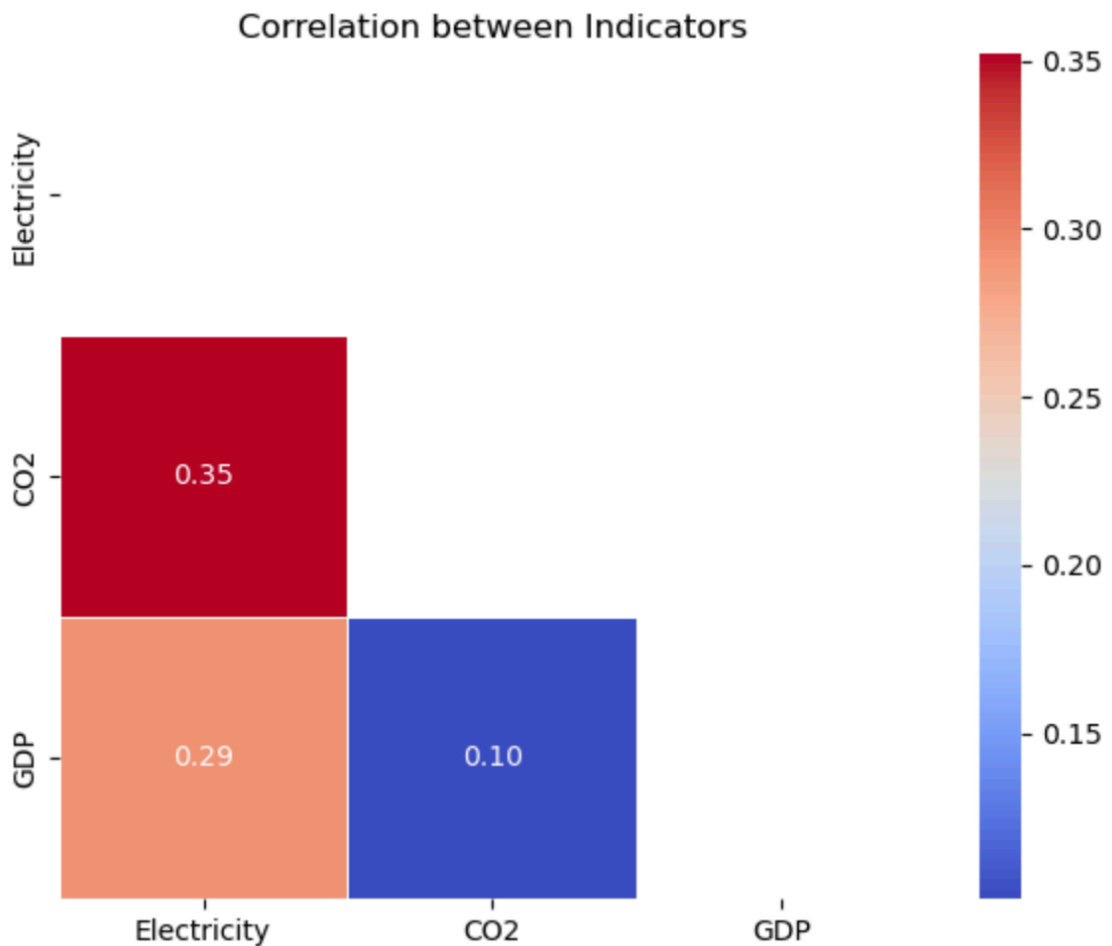


Observations:

- Argentina's emissions decreased over time.
- Chile had flat or slightly rising emissions.
- Other countries were mostly stable with minor fluctuations.

Summary:

Facet view confirms earlier patterns, reinforcing national-level stability.



Observations:

- All correlations were relatively weak.
- GDP vs CO₂ had the weakest correlation.

Summary:

No strong linear relationships were found among the three indicators.

Limitations

This analysis was limited in both size and depth. Only three indicators were considered, and the dataset excluded all CO₂ emission data for the year 2020. Additionally, the study covered only eight countries, making the findings specific and not generalizable.

Future Work

- Incorporating additional sustainability indicators (e.g., renewable energy, poverty rates, education)
- Expanding the time range to include more historical data
- Applying predictive modeling to forecast trends