The Role of Infrastructure in Economic Growth: Analyzing Electricity Production and GDP Growth

Md Sohan Sorker
23074752

Economic growth has a significant impact on infrastructure, especially in developing nations like Latin America. This study looks at the correlation between GDP growth (yearly percentage) and power generation (kWh per capita) in a few countries in Latin America between 2009 and 2014. The study looks for trends, patterns, and connections between economic growth and energy infrastructure through analysis of two World Bank databases. Governments and investors will benefit from the insights produced, which will assist them maximize infrastructure development for the purpose to encourage sustainable growth.

I. QUESTION

How does electricity production (kWh per capita) correlate with economic growth (GDP growth) in Latin American countries from 2009 to 2014?

This inquiry aims to determine whether increased electricity production—a major indication of energy infrastructure—contributes to economic expansion. By examining the relationship between these two factors, the research aims to offer practical insights into how infrastructure supports economic stability and development.

II. DATA SOURCES

Datasource1: World bank (Electricity production (kWh per capita)) [1]
* Data Type: Zip->CSV
*Description: This dataset contains the electricity production per capita for Latin American countries, which is a direct indicator of energy infrastructure.

Datasource2: World bank (GDP Growth (annual %)) [2]
* Data Type: Zip->CSV
*Description: This dataset measures the annual GDP growth (in constant prices) for the same countries, which reflects the economic performance over time.

A. Data Structure

The Electricity Production dataset is structured with temporal variables (year), categorical variables (country name and country code), and continuous variables (electricity production (kWh per capita)).

The GDP Growth dataset is structured with temporal variables (year), categorical variables (country name and country code), and continuous variables (GDP growth (annual %)).

B. Data Quality

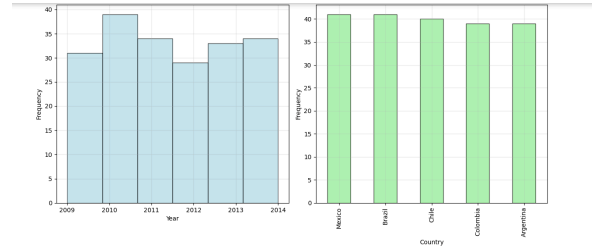The Electricity Production dataset is given below:

Fig. 1. Categorical features from the Electricity Production

The visualizations show the distribution of data in the Electricity Production dataset across Years (2009–2014) and Countries (Mexico, Argentina, Colombia, Brazil, and Chile).

Temporal Distribution (Year): The histogram shows a balanced distribution across all years, providing good coverage for analyzing electricity production trends over time.

Geographic Distribution (Country): The bar chart shows an even representation across the selected countries, ensuring fair comparisons and avoiding bias toward any region.
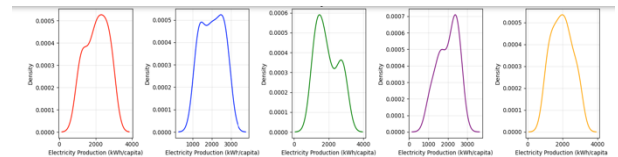

Fig. 2. *Continuous* features from the Electricity Production

The KDE plots illustrate the distribution of Electricity Production (kWh per capita) for Brazil, Mexico, Argentina, Colombia, and Chile:

1. Brazil: The two peaks suggest shifts in electricity production over time.

2. Mexico: A smooth, bell-shaped curve reflects steady and reliable production levels.

3. Argentina: A broader distribution shows greater variation in production.

4. Colombia: A sharp and centered curve highlights consistent production.

5. Chile: A moderately spread-out curve indicates some variability in production.

These patterns provide a clear view of electricity production trends across the countries.

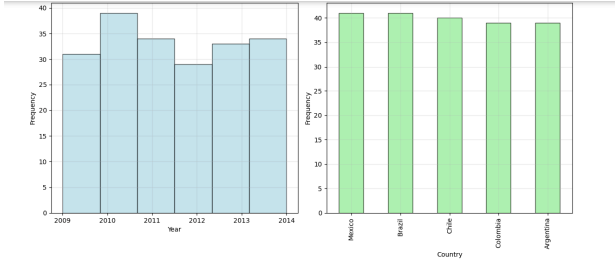The **GDP Growth dataset** is given below:

Fig. 3. Categorical features from the GDP Growth

The visualizations show the distribution of GDP Growth (annual %) data across Years (2009–2014) and Countries (Brazil, Mexico, Argentina, Colombia, and Chile).

- Year: The data is evenly spread across all years, ensuring a good coverage for analyzing economic trends.

- Country: The bar chart shows a balanced representation of the countries, allowing fair comparisons without bias towards any one country.
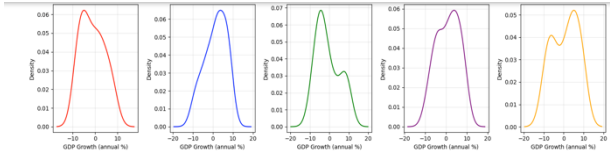

Fig. 4. Continuous features from the GDP Growth

The KDE plots show the distribution of GDP Growth (annual %) for Brazil, Mexico, Argentina, Colombia, and Chile:

1. Brazil: The distribution is roughly bell-shaped, suggesting steady economic growth with values centered around the mean.

2. Mexico: The curve is smooth, indicating moderate and consistent GDP growth fluctuations over the years.

3. Argentina: The distribution is wider, reflecting significant fluctuations and volatility in economic growth.

4. Colombia: The distribution is sharp, indicating more consistent and stable growth.

5. Chile: The curve is more spread out, indicating some variation in GDP growth rates over the observed period.

These patterns offers the economic stability and growth trends across these countries, showing each country's economic performance fluctuates over time.

C. License

Both datasets used in this project, Electricity Production (kWh per capita) and GDP Growth (annual %), are licensed under the World Bank Open Data License [3]. This license allows the data to be used, shared, and modified for any purpose as long as the World Bank is properly credited as the source.

III. DATA PIPELINE

The data pipeline follows a systematic ETL (Download, Transform, Load) architecture to automate data processing. Below breakdown each step:
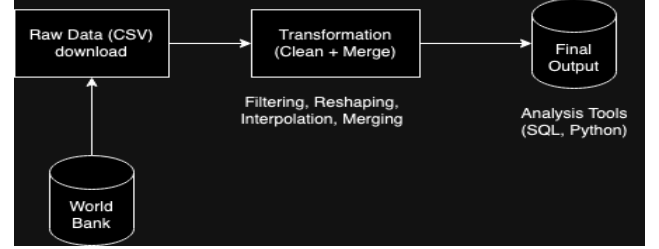

Fig. 5.   The ETL data pipeline architecture

A. Data Download

The first step involves downloading the raw datasets in CSV format using the World Bank API. The datasets are stored locally in a predefined folder.

B. Data Cleaning and Transformation

Removed any unnecessary metadata and header rows from the raw CSV files. Filtered the data to include only Latin American countries from 2009 to 2014. Ensured that data was reshaped from wide format (with years as columns) into a long format (with rows for each year). Filled missing data for some countries.

Both datasets were normalized to ensure consistency in measurement units and scales across countries. The datasets for Electricity Production and GDP Growth were combined using country and year as common columns.

C. Data loading

The cleaned and transformed data was stored in a SQLite database for efficient querying and scalability.Additionally, the processed data was exported to CSV format for use in external analysis tools.

IV. RESULT AND LIMITATIONS

A. Pipeline Output

The output of the pipeline is a cleaned and combined dataset. Its merges data from the electricity production and GDP growth datasets for Latin American countries.

The data is stored in both SQLite database and CSV formats for versatility in analysis and sharing.The output of the data set bellowed:

Country Name	Country Code	Indicator Name	Indicator Code	Year	Value
Argentina	ARG	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	2705.26146589616
Bolivia	BOL	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	539.422090547634
Brazil	BRA	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	2217.79002848663
Chile	CHL	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	3272.59731503103
Colombia	COL	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	1077.12537044961
Costa Rica	CRI	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	1853.64164691205
Cuba	CUB	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	1326.72623373809
Dominican Republic	DOM	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	1309.50520692317
Ecuador	ECU	Electric power consumption (kWh per capita)	EG.USE.ELEC.KH.PC	2009	1101.10958121143

Fig. 6. The Electricity Production

Country Name	Country Code	Indicator Name	Indicator Code	Year	Value
Argentina	ARG	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	-5.91852507634947
Bolivia	BOL	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	3.35699957430565
Brazil	BRA	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	-0.125812002161169
Chile	CHL	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	-1.1180372326977
Colombia	COL	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	1.13964864548062
Costa Rica	CRI	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	-0.873455942787785
Cuba	CUB	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	1.4513054307407
Dominican Republic	DOM	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	2009	0.946155167790621

Fig. 7. The GDP Growth

B. Benefits of using SQLite

SQLite ensures efficient storage, quick querying, and seamless integration with Python, making it ideal for managing the Electricity Production and GDP Growth datasets. Its portability and scalability enhance data accessibility and support future expansions.

C. Limitations

Data Gaps: Countries with significant missing data were excluded, which may limit the generalizability of findings.

External Factors: The analysis does not account for other factors affecting GDP growth, such as trade, political stability, or external economic shocks.

Timeliness: The datasets reflect historical data and may not capture recent trends or developments in energy infrastructure.

V. CONCLUSION

To sum up, the data pipeline effectively used Python and an ETL framework to handle the extraction, transformation, and loading of the GDP Growth and Electricity Production datasets. The pipeline has drawbacks, including difficulties dynamically adjusting to new data inputs, despite its strength in processing structured datasets.

REFERENCES

[1] World Bank, "Electricity use (kWh per capita)," [Online]. Available: https://api.worldbank.org/v2/en/indicator/EG.USE.ELEC.KH.PC?downloadformat=csv. [Accessed: 28-Nov-2024].
[2] World Bank, "GDP growth (annual %)," [Online]. Available: https://api.worldbank.org/v2/en/indicator/NY.GDP.MKTP.KD.ZG?downloadformat=csv. [Accessed: 28-Nov-2024].
[3] World Bank, "Terms of Use for Datasets," World Bank Open Data Licensing, [Online]. Available: https://data.worldbank.org/about/terms-of-use. [Accessed: 28-Nov-2024].