

Tracing the Climate Impact: Examining CO2 Emission Effects on Temperature Patterns in Brazil

Md Jubier Ali

Matriculation No: 23453618

1. Introduction:

This study aims to explore the relationship between CO2 emissions and temperature patterns specifically in Brazil. The project utilizes datasets on CO2 emissions and temperature trends in Brazil to investigate this connection, providing localized evidence of climate dynamics. By focusing on Brazil's unique environmental and industrial context, this study seeks to enhance understanding of how CO2 emissions impact temperature changes, contributing to the global discourse on climate change and its implications.

2. Question:

How do CO2 emissions correlate with temperature trends in Brazil?

3. Data Source:

DataSource1: CO2 and Greenhouse Gas Emissions

- Metadata URL: <https://github.com/owid/co2-data/tree/master>
- Data URL: <https://nyc3.digitaloceanspaces.com/owid-public/data/co2/owid-co2-data.csv>
- Data Type: CSV

This dataset contains data on CO2 emissions (annual, per capita, cumulative and consumptionbased), other greenhouse gases, energy mix, and other relevant metrics.

DataSource2: Earth Surface Temperature Data

- Metadata URL: https://figshare.com/articles/dataset/temperature_csv/3171766/1
- Data URL: <https://figshare.com/ndownloader/files/4938964>
- Data Type: CSV

This dataset contains data from Kaggle, featuring 7 countries and 16 cities.

3.1 Description:

For this analysis, I selected two datasets specific to Brazil: CO2 emissions data from Our World in Data (available on GitHub) and temperature data from Figshare. The CO2 emissions dataset provides detailed historical data on CO2 emissions, including critical indicators such as annual emissions, which are essential for assessing the impact of industrial activities and energy consumption on Brazil's climate. The temperature dataset contains historical temperature records, including monthly and annual averages for various locations, with a focus on Brazil. These datasets enable a focused examination of the relationship between CO2 emissions and temperature trends within Brazil, offering localized insights into climate change dynamics.

3.2 Data Structure:

- **File Format:** Stored as CSV files.
- **Nature of Data:** Semi structured data and not cleaned. There are some null values for all over the world. But for my project which is needed only for Brazil, all the rows were there.

3.3 Data Quality:

- **Accuracy:** Sourced from reputable institutions, reflecting real-world data.
- **Completeness:** Both datasets have some null values. So, they lack completeness.
- **Consistency:** Data is mostly standardized with consistent formats and units.
- **Timeliness:** Historical records are up-to-date and relevant for long-term trend analysis.
- **Relevancy:** Directly aligned with the objective of studying CO2 emissions and temperature changes in Brazil for understanding localized climate change effects.

3.4 License:

Both the CO2 emissions dataset and the temperature dataset are licensed under CC BY 4.0 (Creative Commons Attribution 4.0 International). This license permits the use, sharing, and adaptation of the data as long as proper credit is given to the original creators. To comply with the CC BY 4.0 license, I will ensure that appropriate attribution is provided to the original sources of both datasets in all instances of use, including visualizations, analyses, and publications resulting from this project.

3.5 License Link: Dataset1: <https://github.com/owid/co2-data/tree/master#:~:text=License,data%20before%20use>.

Dataset2:

https://figshare.com/articles/dataset/temperature_csv/3171766/1?file=4938964#:~:text=Climat%20Change%20Processes-,LICENCE,-CC%20BY%204.0

4. Data Pipeline:

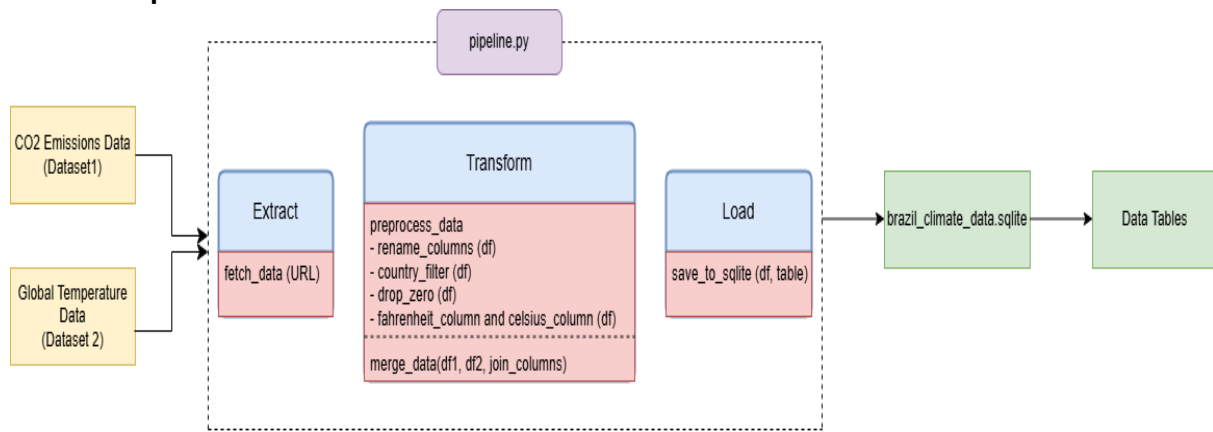


Figure: ETL Pipeline Diagram

The ETL pipeline structure is used in this project. The pipeline diagram for ETL (Extract, Transform, Load) shows how data moves through several processing steps, including extraction, transformation, and loading into a storage system. Let's dissect every element of the ETL pipeline, including the transformation or cleaning stages:

➔ Extract:

- **fetch_data (url):** Fetches CSV data from the specified URL and loads it into a Pandas DataFrame.

➔ Transform:

- **preprocess_data(rename_columns, country_filter, drop_zero, fahrenheit_to_celcius):** Manipulates, rename rows and augments the data..
- **drop_zero_values(df):** Drops rows where the specified column value is 0.0.
- **rename_columns(df):** Renames columns according to a specified mapping.
- **convert_fahrenheit_to_celcius(df):** Converts temperatures from Fahrenheit to Celsius and drops the original Fahrenheit column.
- **merge_data(df1, df2, join_columns):** Joins two DataFrames on specified columns (year, country).

➔ Load:

- **save_to_sqlite(df, table):** Saves the processed DataFrame to an SQLite database.

4.1 Problem encountered and solution:

- **Problem:** There were no data of Celsius.
- **Solution:** Added a method to convert temperatures from Fahrenheit to Celsius and drop the original Fahrenheit column.

5. Result and Limitations:

The output data is a cleaned and transformed DataFrame containing:

- CO2 emissions for Brazil per year.
- Average surface temperature in Celsius for Brazil per year.
- The data is stored in an SQLite database for seamless access and analysis.

5.1 Data Structure of the output:

- **Storage Format:** Stored in an SQLite database named `brazil_climate_data.sqlite`.
- **Data Type:** A cleaned and transformed DataFrame with rows representing Brazil-specific records for CO2 emissions and temperature trends.

5.2 Data Quality of the output:

- **Accuracy:** The output data reflects real-world records for Brazil, ensuring reliability and correctness.
- **Completeness:** All necessary columns and rows for Brazil are included, with no missing or null values after cleaning.
- **Consistency:** The data adheres to standardized formats, ensuring uniformity across all entries.
- **Timeliness:** Though historical, the data is current enough to analyze long-term trends for Brazil.
- **Relevancy:** Focused entirely on Brazil, the data is highly relevant for studying the relationship between CO2 emissions and temperature trends in this specific region.

5.3 Data Format of the output:

I chose SQLite as the output format for my pipeline because it is a lightweight and portable database format, making it easy to store, share, and manage the Brazil-specific transformed data; it supports structured queries, enabling efficient retrieval and analysis of Brazil's climate data; and it is scalable, handling the transformed data efficiently without the overhead of larger database systems.

5.4 Limitations:

There were some missing and null values in the datasets, but all relevant data for Brazil were complete. The global scope required extra preprocessing to isolate Brazil-specific data. Throughout the examined time, Brazil's temporal coverage was stable, but it was inconsistent for global.