How has rainforest degradation in the Brazilian Amazon contributed to carbon emissions (CO2) and climate change?

Introduction

This study explores the connection between rainforest degradation and carbon emissions in the Brazilian Amazon. By analyzing historical data on fire activity and CO2 emissions from 1999 to 2019, the project aims to show how land-use changes and fires contribute to carbon emissions. Understanding this connection is essential for evaluating the Amazon's environmental role, identifying major drivers of climate change, and promoting sustainable forest management and conservation.

Data Sources

1. FAOSTAT Emissions Data

- **Metadata URL:** https://www.fao.org/faostat/en/#data/GT/metadata
- **Data URL:** https://bulks-faostat.fao.org/production/Emissions_Totals_E_Americas.zip
- Description This dataset provides annual CO2 emissions data from land-use changes and forestry activities for all countries in the Americas (1999–2019). For this study, the data has been filtered to include only Brazil and aggregated to represent total yearly CO2 emissions.
- Structure: Tabular format with columns for emission sources, years, and values in metric tons of CO2. The dataset was filtered for relevant activities to ensure consistency.
- Licensing: Publicly available under an open-data license, with proper attribution and documentation.

2. Amazon Fires Dataset

- **Metadata URL:** https://www.kaggle.com/datasets/mbogernetto/brazilian-amazon-rainforest-degradation
- **Data URL:** N/A (Data downloaded using KaggleHub).
- Description: This dataset includes yearly firespot counts in the Brazilian Amazon from 1999 to 2019, highlighting temporal patterns of fire activity critical for analyzing deforestation.
- Structure: Tabular format with year and firespot counts. Data was cleaned to retain only relevant columns and align with the project's timeframe.
- Licensing: Available under an open-data license, with appropriate citations provided.

Data Pipeline

The pipeline, implemented in Python, employs key libraries such as pandas for data manipulation, sqlite3 for efficient storage and querying, urllib.request and zipfile for

downloading and extracting datasets, kagglehub for accessing Kaggle datasets programmatically, and os for file management. The steps are as follows:

1. **Loading Data:** CO2 emissions data from FAOSTAT and firespot data from Kaggle were imported into the project directory.
2. **Preprocessing:**
    - CO2 Data: Filtered to include only Brazil's data from 1999 to 2019. The emissions were aggregated to calculate total yearly CO2 emissions. Temporary files, such as unrelated FAOSTAT files, were removed to keep the pipeline efficient.
    - Firespot Data: The firespot data was aggregated by year, counting the total firespots in the Brazilian Amazon, and was restricted to the 1999–2019 period to match the CO2 emissions data timeframe.
3. **Merging Data:** The datasets were merged on the "Year" column to create a unified dataset containing total yearly CO2 emissions and firespot counts.
4. **Storage:** The final dataset was stored in an SQLite database (data_pipeline.db) for efficient querying and exported as a CSV file for visualization.
5. **Error Handling & Cleanup:** Validation mechanisms ensured data consistency, filtering out rows with missing or zero values. Logs tracked and recorded any processing errors for troubleshooting, Removed temporary files, such as unrelated FAOSTAT CSVs, to keep the workspace clean.

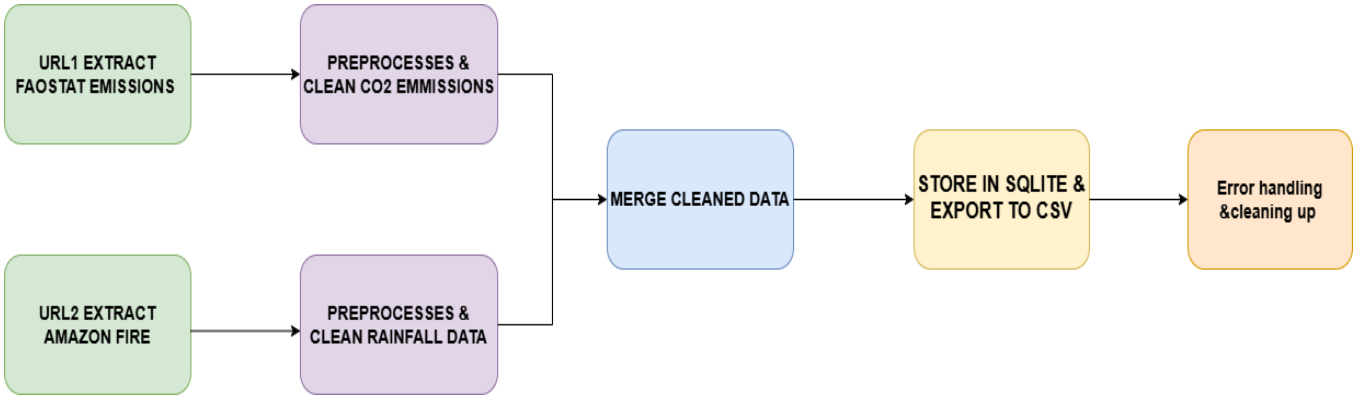


Fig: Automated Data Pipeline

Data Quality and Challenges

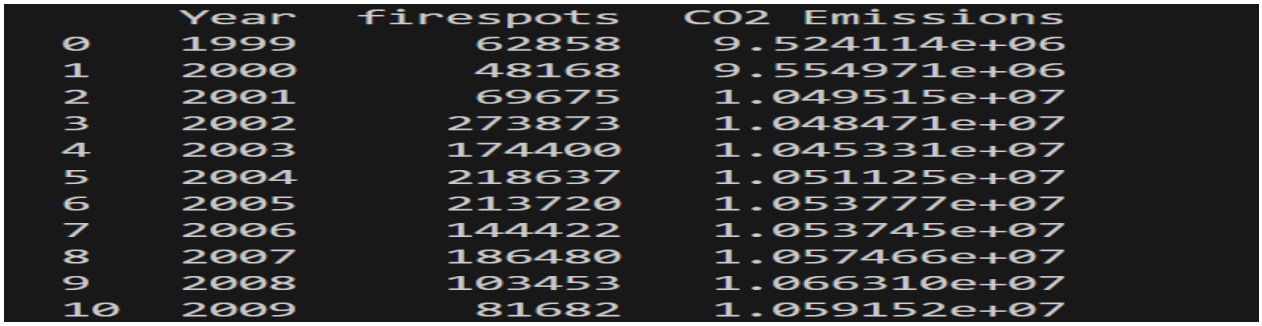
- Quality Metrics: The data reflects real-world events (accuracy), includes necessary information for 1999–2019 (completeness), maintains consistent formats (consistency), and aligns with the analysis period (timeliness).
- Challenges: Data merging required standardizing column names and formats. Filtering addressed inconsistencies and missing values. Dynamic preprocessing ensured adaptability to input changes.

Results:

The pipeline produced an SQLite database (data_pipeline.db) with three tables:

- FAOSTAT_Cleaned: Filtered and aggregated yearly CO2 emissions data.
- Amazon_Fires_Cleaned: Aggregated firespot data.
- Merged_Data: Combined dataset with total yearly CO2 emissions and firespot counts (1999–2019).

The SQLite format was chosen for efficiency, portability, and ease of use, while the CSV export (Merged_Data.csv) enables additional analysis and visualization.



	Year	firespots	CO2 Emissions
0	1999	62858	9.524114e+06
1	2000	48168	9.554971e+06
2	2001	69675	1.049515e+07
3	2002	273873	1.048471e+07
4	2003	174400	1.045331e+07
5	2004	218637	1.051125e+07
6	2005	213720	1.053777e+07
7	2006	144422	1.053745e+07
8	2007	186480	1.057466e+07
9	2008	103453	1.066310e+07
10	2009	81682	1.059152e+07

Limitations:

- **Timeframe:** Limited to 1999–2019, potentially missing recent trends or anomalies.
- **Data Gaps:** Rows with missing or zero values were removed, which could affect the completeness of the total yearly emissions data.
- **Aggregation Trade-Offs:** Aggregating emission reasons into yearly totals simplifies analysis and enhances readability but sacrifices the ability to analyze specific contributors (e.g., forest fires, drained organic soils) independently. This limits deeper exploration into the drivers of emissions.
- **Methodological Bias:** CO2 estimates depend on FAOSTAT's methodology, which may differ from other sources.
- **Correlation Complexity:** Inferring causation between firespots and CO2 emissions requires further statistical modeling. Other factors, such as agricultural practices or non-forest land-use changes, could also contribute to emissions.
- **Regional Focus:** The study is specific to Brazil, limiting broader generalizations about global rainforest degradation.

Conclusion and Next Steps:
This study establishes a foundation for analyzing the relationship between Amazon fires and CO2 emissions. Future work will explore causal relationships using statistical modeling, incorporate recent data, and develop visualizations to support environmental policy decisions.