Project Report: Effects of Amazon deforestation on Forest fires and CO_2 emissions in Brazil.

1. Introduction

The Amazon rainforest, often called the planet's lungs, is the largest tropical rainforest on Earth and plays an essential role in regulating global carbon dioxide levels. However, decades of extensive deforestation have severely disrupted its ecosystem, making it increasingly susceptible to forest fires. These fires, frequently linked to human activities such as slash-and-burn agriculture, can also be attributed to the interconnected nature of deforestation and fires. This creates a feedback loop where forest loss leads to drier conditions, fueling fire outbreaks. This study aims to investigate the relationship between deforestation, forest fires, and CO_2 emissions in Brazil, focusing on the Amazon Basin states, using statistical analysis and data visualization to uncover and analyze their complex interconnections.

2. Methods

2.1. Data Sources

This project utilizes three datasets from two sources. Two datasets from the Global Wildfire Information System (GWIS) provide historical data on wildfire burned areas and emission gases (2002–2023) in CSV format, with no missing values, focused on Brazil and its provinces. The third dataset, sourced from Kaggle and contributed by Mariana Boger Netto, contains information on deforestation in Brazil.

1. **Global Monthly Burned Area Dataset (GMBAD) [2002–2023]**
   This dataset contains monthly records of burned areas, including details such as year, country names and codes, and various land cover types measured in hectares. A value of 0 indicates that no wildfire occurred during the recorded period.
2. **Global Monthly Emission Dataset (GMED) [2002–2023]**
   This dataset includes monthly data on burning emissions by pollutant, along with year, country details, and various pollutant gases measured in tons. Similar to GMBAD, a value of 0 signifies that no emissions occurred during the recorded period.
3. **Yearly Deforested Area [2004–2019]**
   This dataset provides annual deforestation data for Brazil (Amazon), sourced from PRODES using satellite imagery. It is structured as a CSV directory and reflects data from reliable Brazilian government sources. The dataset includes key variables such as year, state, and deforestation area. It is of high quality, consistently formatted, and suitable for detailed analysis.

2.2 Data Pipeline

The automated data pipeline, built using Python and libraries like pandas, numpy, requests, zipfile, io, and sqlalchemy, follows an ETL (Extract, Transform, Load) format to handle data extraction, transformation, and database integration.

Extraction: Datasets on Amazon Rainforest degradation (Kaggle) and wildfires (Global Wildfire Information System) are extracted and loaded as pandas dataframes for analysis.

Transformation: The transformation stage cleans and prepares datasets for analysis by removing missing values, dropping irrelevant columns, and renaming columns for clarity (e.g., 'Ano/Estados' to 'Year'). It filters data to include only Brazilian records from 2004–2019 and aggregates monthly wildfire and emissions data into yearly totals. These steps ensure consistent formatting and compatibility for seamless integration and accurate analysis.

Loading: The final stage involves loading the transformed data into a structured SQL database named "brazil_amazon_deforestration_and_wildfire_and_emission_data", ensuring efficient storage for easy retrieval and analysis in subsequent stages of the project.

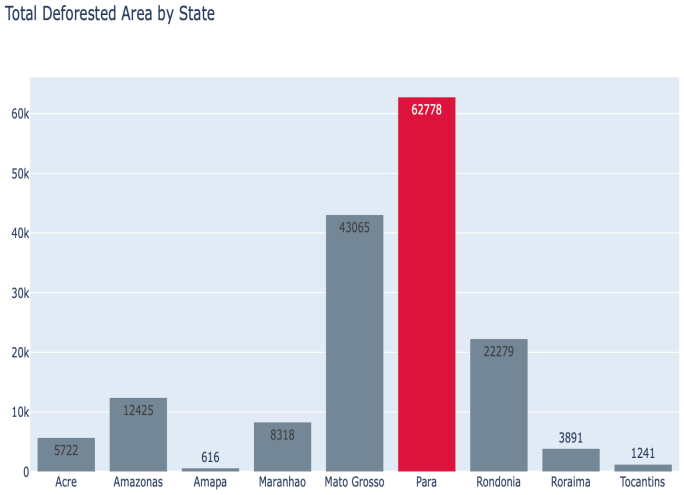
2.3 Data License

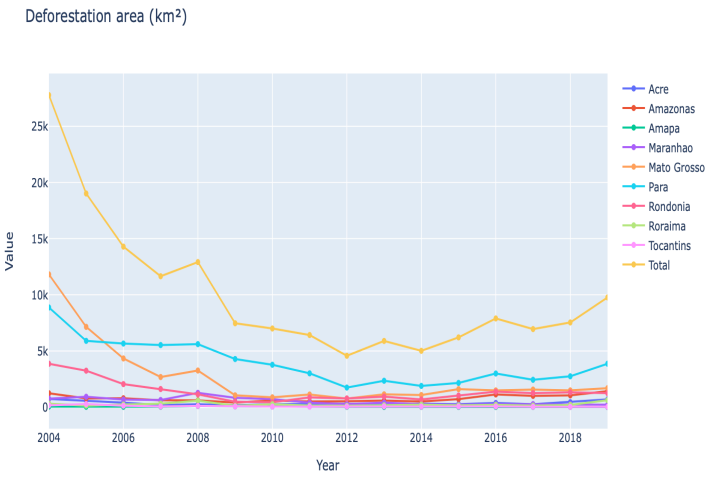
The Amazon deforestation dataset is available under the CC0 Public Domain, allowing unrestricted use for non-commercial and commercial purposes without copyright. For full details, click on License. The Wildfire (Global Wildfire Information System) dataset is available under the Creative Commons Attribution 4.0 International (CC BY 4.0) license (see the second page of the PDF for details). This License allows reuse with proper credit and disclosure of any modifications. The European Commission implements this reuse policy. To comply with the obligations, appropriate credit will be given to GWIS along with a link to the License.

3. Analysis

To address the objective question, I filtered the dataset to focus specifically on the Brazilian states that contain Amazon rainforest coverage. I then performed exploratory data analysis on the datasets, including variables related to wildfire burn areas and emissions. I also visualized the data to gain a deeper understanding and uncover key insights.

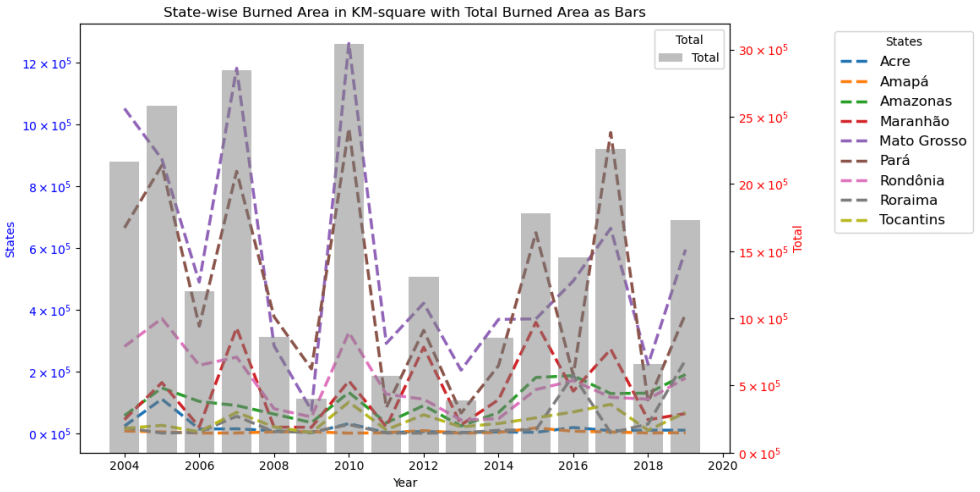
3.1 Deforested Area (2004-2019)

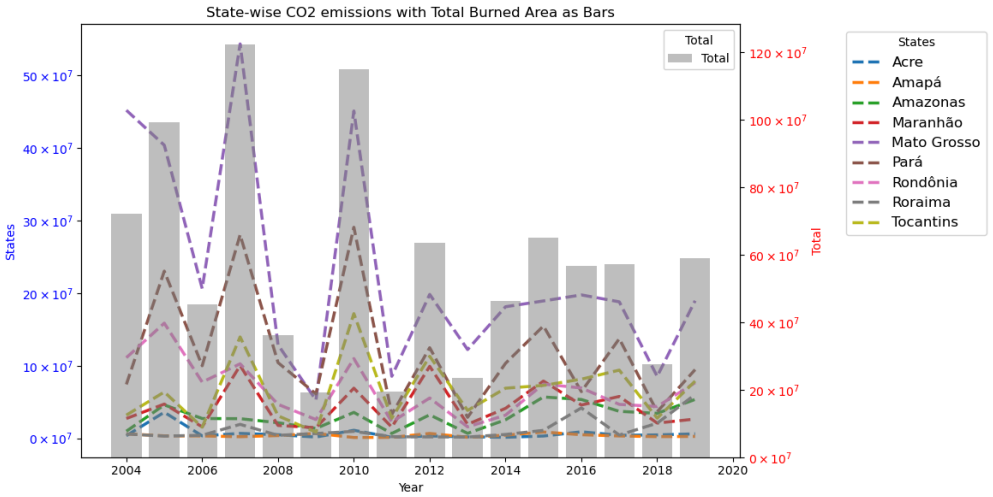




These graphs show statewise and total deforestation that occurred from 2004 till 2019 in Brazil states that come under Amazon Basin. The most deforestation of the rainforest occurred in Para.

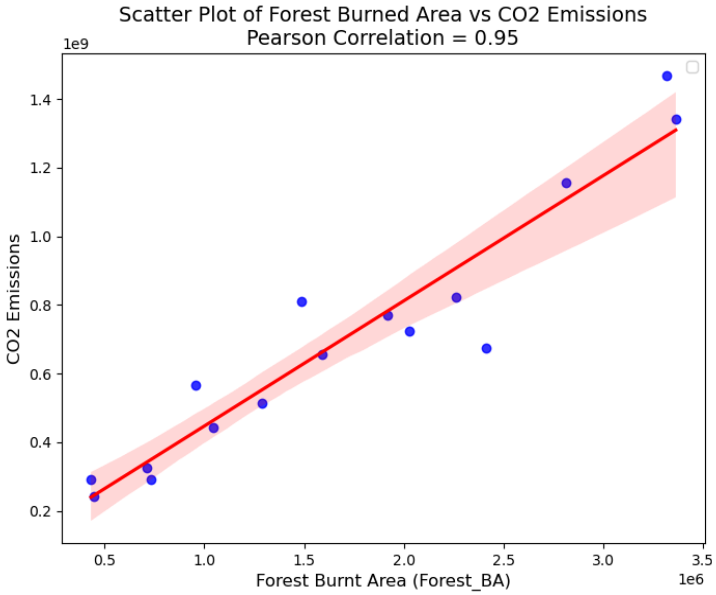
3.2 Forest Burned Area and CO_2 emissions



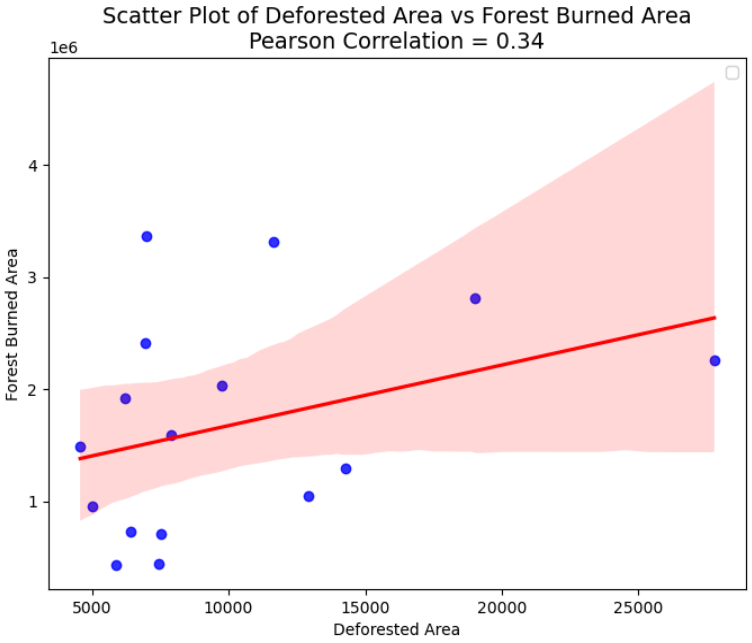


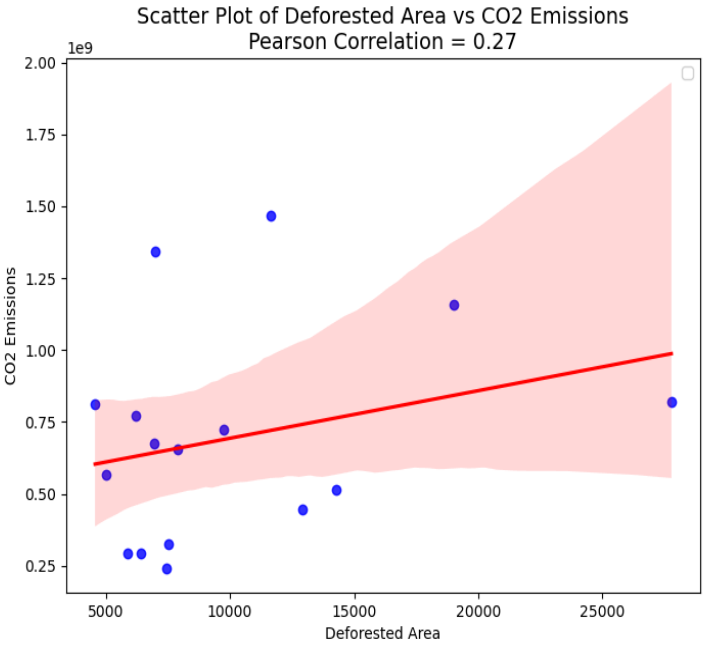
The line graph shows the state-wise distribution, and the bar chart shows the total forest area burned and CO_2 emissions. The left-hand(blue) y-axis in both graphs shows state measurements, and the right-hand(red) shows total measurements. The two highest states for forest burned area and CO_2 emissions are Mato Grosso and Para, respectively.

3.3 Pearson Correlation



By finding the Pearson Correlation between Forest Burned Area and CO_2 emissions to be "0.95," which indicates a strong correlation, I came to the conclusion that an increase in Forest burned areas also increases CO_2 emissions. In my initial hypothesis, I predicted this to be the case. However, I also hypothesised that increased deforestation gives rise to more forest fires and, hence, more CO_2 emissions, which was not entirely true.





However, I also predicted that increased deforestation would lead to more forest fires and, hence, more CO_2 emissions, which was not entirely true. As seen in these graphs, the Pearson Correlation is positive, but it's too weak to deduce that the rise in forest burned area and CO_2 is majorly contributed by the deforestation that is happening in the Amazon Basin.

4. Conclusion

The analysis of deforestation, wildfire burned area, and CO_2 emissions in the Amazon (2004–2019) provided valuable insights but *did not fully answer the initial question*. While Para had the highest deforestation, Mato Grosso showed greater forest burned area and CO_2 emissions, contradicting the hypothesis that deforestation directly drives these outcomes. As shown in the further analysis, the weak correlation between deforestation, burned area, and emissions suggests other factors, such as climate, human activities, and policies, play significant roles.

4.1 Limitation

- The analysis does not fully explore all the factors contributing to the spread of wildfires, such as human activities, land use changes, and specific climatic conditions. It also lacks evidence of how much CO_2 emissions occurred due to wildfires and how much was due to human activities.
- The deforestation data is aggregated yearly, which may overlook detailed temporal variations. Additionally, the effects of deforestation on a forest can manifest over several years, further complicating the analysis. Due to the limited availability of data and the annual aggregation of deforestation records, I also had to aggregate CO_2 emissions data and forest burned area into yearly values. As a result, the dataset was too limited to generate a meaningful correlation heatmap, reducing the ability to capture nuanced relationships between the variables.

4.2 Future Work

Future work should use higher-resolution temporal data, such as monthly records, to capture short-term variations and extreme events. Incorporating additional variables like climate patterns and land-use changes could provide a deeper understanding of the factors driving forest fires and CO_2 emissions.