Methods of Advanced Data Engineering

Kimia Sedighi
23455613

Introduction

The primary objective of this analysis is to investigate the correlation between atmospheric CO2 concentrations and global surface temperature over the past several decades. Understanding these trends is critical in assessing the impact of anthropogenic activities on climate change. This report aims to provide a comprehensive overview of the data, analysis methods, results, and interpretations to support the hypothesis that rising CO2 levels are associated with an increase in global surface temperature.

Used Data

For this analysis, two key datasets were used:

1. **Monthly Atmospheric CO2 Concentrations (1958-2024)**

- Source: IMF (International Monetary Fund) data.
- Structure: The dataset contains monthly records of CO2 concentrations measured in parts per million (ppm) or percent; we used only ppm for our analysis.
- Fields: Date, Value (CO2 concentration), Unit (ppm or percent), etc.
- License: Data usage complies with the terms specified by the IMF data services.

2. **Annual Mean Global Surface Temperature (1961-2023)**

- Source: IMF (International Monetary Fund) data.
- Structure: The dataset contains annual records of global surface temperature anomalies measured in degrees Celsius (°C).
- Fields: Country, Unit, F1961 to F2023 (temperature anomalies for each year).
- License: Data usage complies with the terms specified by the IMF data services.

Analysis

Method

1. Data Retrieval and Preprocessing

   - The datasets were fetched from the ArcGIS REST services in GeoJSON format.
   - The GeoJSON data were converted to CSV format for easier manipulation.
   - The CO2 dataset was filtered for records between 1961 and 2023, and dates were standardized.
   - Both datasets were transformed to remove unnecessary columns and validated for some column values (such as numeric ones). The heading names were also made proper (for example, removing the F prefix in the temperature table).

2. Temperature Data Transformation:

   - The temperature dataset, which had multiple columns for each year, was reshaped using a "melt" operation. This created "Year" and "Temperature" columns.
   - The melted data was then aggregated by year to obtain the average temperature anomaly across countries
   - This produced a dataset with yearly averages, simplifying the analysis of temperature trends over time.

3. Data Storage:

   - Processed data were stored in SQLite databases for efficient querying and analysis.

4. Data Visualization:

   - Line plots were created to visualize trends in CO2 concentrations and global surface temperatures over time.

Results and Interpretation

1. CO2 Concentrations

- The plot (Figure 1) shows a clear upward trend in CO2 concentrations from 1961 to 2023.
- The increase is consistent, with minor seasonal variations likely due to natural processes such as photosynthesis.

2. Global Surface Temperature

- The plot (Figure 2) indicates a significant rise in global surface temperatures from 1961 to 2023.
- There are fluctuations year-to-year, but the overall trend is an increase in temperature anomalies.

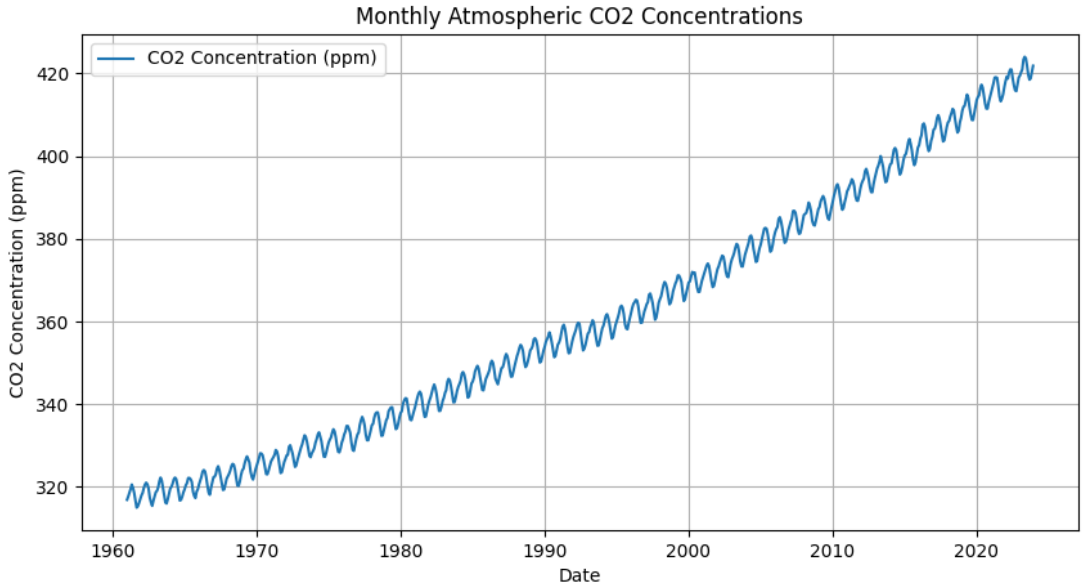

Figure 1

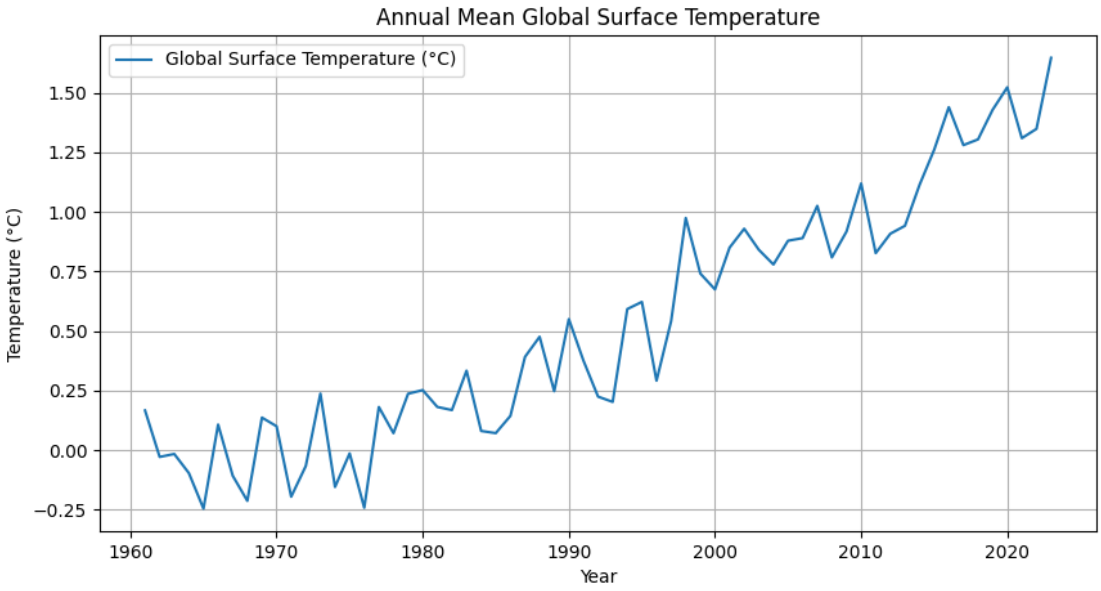

Figure 2

Conclusions

The analysis supports the hypothesis that rising atmospheric CO2 levels are associated with an increase in global surface temperatures. The data clearly shows that CO2 concentrations have steadily increased over the past several decades, and this trend corresponds with a noticeable rise in global surface temperatures.

While the findings are robust, it is essential to acknowledge some limitations:

- The analysis does not account for other factors influencing global temperatures, such as volcanic activity, solar radiation variations, or other greenhouse gases.

- The datasets, though comprehensive, are limited to the recorded periods and may not capture earlier variations.

Future research could include more variables and extend the temporal range to provide a deeper understanding of climate dynamics. Nonetheless, the current analysis underscores the critical role of CO2 in global warming and emphasizes the need for mitigating actions to address climate change.