Detailed Data Report: Impact of Temperature Change on Various Factors of Climate Change in Germany

Question

What is the impact of temperature change on various factors of climate change in Germany?

Data Sources

Data Source 1: Climate Change Indicators for Germany

- **Source:** World Bank Climate Change Indicators for Germany
- **Metadata URL:** World Bank Climate Change Indicators for Germany
- **Data URL:** Climate Change Indicators Data

Data Source 2: Kaggle - Climate Change: Earth Surface Temperature Data

- **Source:** Kaggle, Berkeley Earth
- **Metadata URL:** Kaggle Climate Change: Earth Surface Temperature Data
- **Data URL:** Earth Surface Temperature Data

Data Structure and Quality

- **Climate Change Indicators for Germany:**
  - Format: CSV
  - Structure: Tabular with columns for Year, CO2 emissions, energy consumption, renewable energy, and other indicators.
  - Quality: Generally high, but some missing values and potential outliers need to be addressed.
- **Earth Surface Temperature Data:**
  - Format: CSV
  - Structure: Tabular with columns for date, average temperature, temperature uncertainty, and country.
  - Quality: High coverage and granularity, but requires filtering for Germany-specific data.

Licenses and Usage

- **Climate Change Indicators for Germany:** Provided under an open-data license. The usage complies with the World Bank's data policies, ensuring proper attribution and no commercial use.
- **Earth Surface Temperature Data:** Available under a public domain license on Kaggle. Attribution is given to Berkeley Earth as required, and data is used for educational and non-commercial purposes.

Data Pipeline

Overview

The data pipeline involves collecting, cleaning, transforming, and analyzing the datasets to address the main research question. The pipeline is implemented using Python, leveraging libraries such as pandas for data manipulation, matplotlib, and seaborn for visualization.

Steps and Transformations

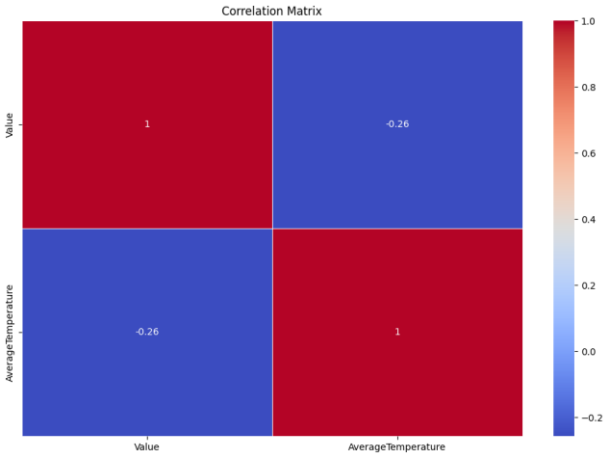
1. **Data Collection:**
    - Download datasets from provided URLs.
    - Load datasets into pandas DataFrames.
2. **Data Cleaning:**
    - Convert date columns to datetime format.
    - Handle missing values using forward fill method to maintain continuity.
    - Filter temperature data to include only records specific to Germany.
3. **Data Transformation:**
    - Merge datasets based on common attributes like date and country.
    - Create new features if necessary (e.g., year from date).
4. **Exploratory Data Analysis:**
    - Generate visualizations to identify trends and correlations.
    - Calculate summary statistics to understand data distribution.
5. **Error Handling and Adaptability:**
    - Implement try-except blocks to handle potential errors during data loading and transformation.
    - Design the pipeline to automatically update with new data, accommodating changes in input data structure or values.

Problems and Solutions

- **Missing Values:** Some indicators had missing values. Applied forward filling to impute missing values.
- **Data Filtering:** Ensured that only Germany-specific temperature data was included by filtering the dataset based on the country column.
- **Data Merging:** Managed merging of datasets by using common attributes and resolving any inconsistencies in data types and formats.

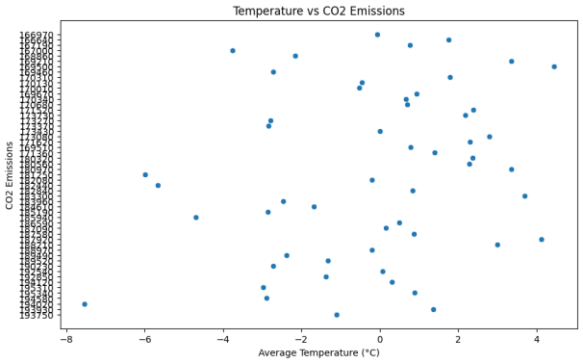
Result and Limitations

Output Data



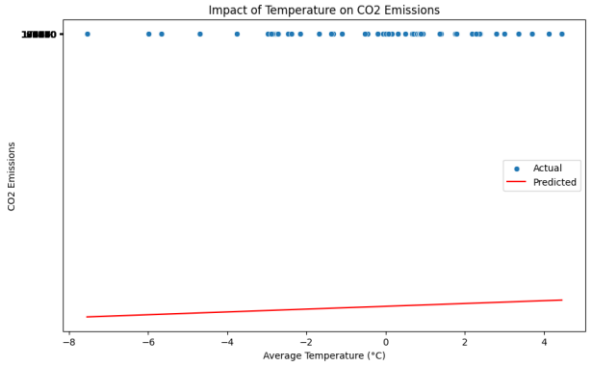
- **Data Structure and Quality:**
  - Format: Merged CSV containing both climate indicators and temperature data.
  - Structure: Tabular with columns for year, average temperature, CO2 emissions, energy consumption, renewable energy, and other indicators.
  - Quality: High-quality data with addressed missing values and consistent formats.

Data Format



- **Chosen Format:** CSV for ease of use and compatibility with various data analysis tools and platforms.
- **Reason:** CSV is widely supported, making it easy to share and integrate with other systems or software for further analysis or visualization.

Critical Reflection



- **Data Limitations:**
  - Potential biases in historical data collection methods.
  - Missing values and imputed data may introduce some level of uncertainty.
- **Anticipated Issues:**
  - Changes in data collection methods or definitions over time could affect trend analysis.
  - External factors influencing climate indicators that are not captured in the datasets (e.g., policy changes, economic shifts).

Conclusion

The data pipeline effectively processes and merges datasets to analyze the impact of temperature changes on various climate indicators in Germany. The resulting data is of high quality, ready for detailed analysis and interpretation. However, potential data limitations and external influences should be considered when drawing conclusions from the analysis.