Title:
Climate Change Patterns in Bangladesh: A Correlation Analysis of Temperature and Rainfall Trends

Introduction:
Climate change is a big deal in worldwide, Bangladeshand it also hitting Bangladesh hard. The weather is getting warmer and the rain patterns are also changed. This project is all about looking at weather data for a long time, to see if there's a connection between temperature and rainfall in . We want to understand why this is happening because it's really important for figuring out how to deal with climate change. By figuring out, hope we can hopefully find ways to make things better for Bangladesh.

Question:
How temperature and rainfall patterns in Bangladesh relate to each other over time, and what can this tell us about dealing with climate change?

Data Sources:

1. Bangladesh Weather Dataset (Dataset 1):

Source: The data is sourced from Kaggle, a platform for hosting datasets and data science projects.

URL:
https://www.kaggle.com/datasets/yakinrubaiat/bangladesh-weather-dataset

Reason for Selection: This dataset was chosen because it provides comprehensive weather data specifically for Bangladesh, covering a wide range of variables over a significant time period.

License: Open source. No restrictions on use.

Obligations: Any use or modification of the dataset requires proper attribution to the dataset source. Additionally, any alterations made to the dataset must be clearly documented and indicated in any subsequent works or analyses.

2. Historical Rainfall Data in Bangladesh (Dataset 2):

Source: This dataset is also sourced from Kaggle.

URL:
https://www.kaggle.com/datasets/redikod/historical-rainfall-data-in-bangladesh

Reason for Selection: It was chosen due to its focus specifically on historical rainfall data in Bangladesh, which is essential for understanding precipitation patterns in the region.

License: Open source. No restrictions on use.

Obligations: Despite being open source, acknowledge the source of the data when using it in any analysis or publication.
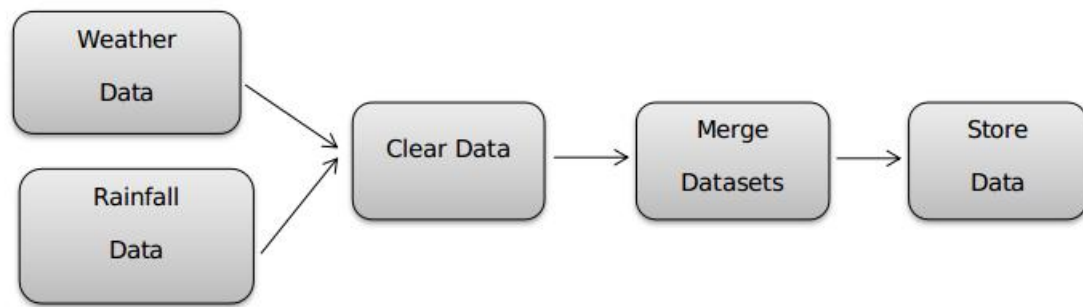
Data Pipeline:

Figure: Data Pipeline

The data pipeline described in the provided code is focused on fetching weather and rainfall data for Bangladesh from Kaggle datasets, merging them based on year and month, and storing the merged data in a SQLite database. Here's an overview of the pipeline and some relevant considerations:

Pipeline Overview:

- The pipeline consists of three main steps:

    1. Fetching weather data
    2. Fetching rainfall data
    3. Merging the datasets based on year and month.

- The pipeline is implemented using Python, utilizing libraries such as pandas for data manipulation, opendatasets for downloading datasets from Kaggle, and SQLAlchemy for database interaction.

Data Transformation and Cleaning:

- The cleaning steps involve dropping the 'Day' column from both weather and rainfall datasets, as it might not be needed or could cause redundancy after merging.

- Merging datasets based on 'Year' and 'Month' columns ensures alignment and simplifies analysis.

- Missing values are handled by dropping rows with missing values after the merge.

Encountered Problems and Solutions:

- One potential problem could be inconsistency or missing data in the Kaggle datasets. This could lead to errors during data processing.

• The 'force=True' parameter in the 'od.download()' function ensures that datasets are re-downloaded, helping to mitigate potential issues caused by outdated or corrupted data.

• Another problem could arise from changes in the structure or format of the input data. However, the pipeline doesn't explicitly handle this, so it may require manual intervention to update the pipeline code accordingly.

Dealing with Errors or Changing Input Data:

• The pipeline doesn't have explicit error handling mechanisms beyond dropping missing values.

• To handle errors more robustly, additional error handling code could be added, such as try-except blocks to catch and log errors during data processing.

• For changing input data, periodic reviews of the pipeline code may be necessary to ensure compatibility with any updates or changes to the input datasets.

Result and Limitations:

Results

Output Data Description

• The output is a merged dataset stored in a SQLite database, combining temperature and rainfall data for Bangladesh.

• The dataset is structured in a tabular format with columns for Year, Month, Temperature, and Rainfall.

• Data quality is maintained by removing redundant 'Day' columns and dropping rows with missing values.

Data Format

• Chose SQLite database format for its efficiency, compatibility, and persistence. It handles medium-sized datasets well and integrates seamlessly with analysis tools.

Limitations

• Differences in data collection methods could lead to inconsistencies.

• Merging by year and month loses finer time granularity (e.g., daily variations).