Global Air Temperatures and Precipitation Patterns Analysis

The goal of this report is to examine trends in global air temperature for the period of 1922-2021 and explore changes in the global rate of precipitation for the years 1951-2021. Furthermore, it also elaborates on how both climate variables are related to one another.

Main Questions

- How have global air temperatures changed from 1922 to 2021?
- What are the observed changes in global precipitation patterns from 1951 to 2021, and how do these changes correlate with the trends in global air temperatures?

Data Sources

There are two sets of data being used to understand long-term climatic tendencies. Such long periods help in determining whether there have occurred drastic changes and trends in climate information. Also, both sources are public, means that these can be used, copied, distributed and adapted for any purposes.

1.Global Air Temperature Mean Dataset:

- **MetaData URL:** https://opendata.dwd.de/climate_environment/CDC/observations_global/CLIMAT/monthly/qc/air_temperature_mean/
- **Data URL:** https://opendata.dwd.de/climate_environment/CDC/observations_global/CLIMAT/monthly/qc/air_temperature_mean/historical/01001_192201_202112.txt
- **Data Type:** txt
- **License Type:** CC BY 4.0
- **License Consent:** This license gives a free license to distribute and use and adapt the licensed work.
- **Citation:** The Deutscher Wetterdienst. It is a public institution.
More information can be found here https://www.dwd.de/.
More information about the **License** can be found here
https://opendata.dwd.de/climate_environment/CDC/Terms_of_use.pdf

Specifically, this dataset identifies monthly global air temperature anomalies from the period of 1922 through 2021. Anomalies are concrete deviations from long-term average temperature thus giving a true picture of temperatures rising or falling with time.

2. Precipitation Total Dataset:

- **MetaData URL:** https://opendata.dwd.de/climate_environment/CDC/observations_global/CLIMAT/monthly/qc/precipitation_total/historical/
- **Data URL:** https://opendata.dwd.de/climate_environment/CDC/observations_global/CLIMAT/monthly/qc/precipitation_total/historical/01001_195101_202112.txt

- **Data Type:** txt
- **License Type:** CC BY 4.0
- **License Consent:** This license gives a free license to distribute and use and adapt the licensed work.
- **Citation:** The Deutscher Wetterdienst. It is a public institution.
  More information can be found here https://www.dwd.de/.
  More information about the **License** can be found here
  https://opendata.dwd.de/climate_environment/CDC/Terms_of_use.pdf

The source data includes monthly precipitation for the years between 1951 and 2020. One obvious characteristic is that it contains information on precipitation occurrence for 70 years, which makes it possible to assess the tendencies in climatic changes.

Data Pipeline

The data pipeline is designed to extract, transform, and load (ETL) climate data from online sources into a local SQLite database. In detail, the pipeline retrieves historical precipitation data and temperature data from the DWD Open Data and then cleans it up for being interested metric and then saves it locally in an SQLite database for further data analysis. **Python, Pandas, SQLAlchemy, SQLite** are used as technology in my analysis.

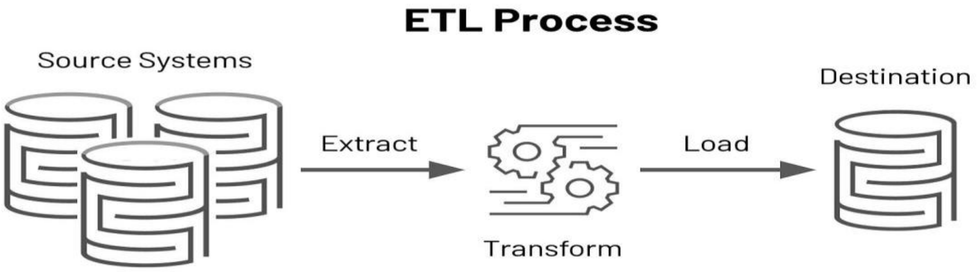

Figure 1: ETL

Transformation and Cleaning Steps

Data Extraction:

Data to be downloaded was obtained from DWD open data using the URLs that were present in the code.The data should now be stored in Pandas DataFrames using pd. read_csv() at the beginning of each data reading process**.**

Data Transformation:

The current implementation of the pipeline focuses on data retrieval and storage because my datasets are quite sufficient for my analysing. That's why there is no need for any cleaning steps. However, some potential steps I know regarding data transformation , such as, Data Type Conversion,Handling Missing Values,Data Normalization,Filtering.

Data Loading:

Exported the cleansed and transformed DataFrame into an SQLite database through the applying the to_sql method of SQLAlchemy.

Problems Encountered and Their Solutions

Issues:

1.Initialization Error: I encountered a problem during initialization because of a typo, in the class constructor (init should be _init_).

2.URL Fetching: Fetching data from URLs, without error handling could cause problems if the URLs become unreachable or if the data format changes.

Solutions:

1. Initialization Issue: Rectified the constructor method to "__init__."

2.Fetching URLs: Implemented error management strategies to smoothly address problems such as connection failures or alterations, in data structure.

Error Dealing

Use try except blocks to enclose data retrieval and database operations for catching and managing exceptions.

Result and Limitations

Final Output:

A Sqlite database “Final_Data” is created to save output data in a table.

Reason:

SQLite is a serverless database engine that's simple to set up making it ideal, for small to medium sized applications. It keeps all the data in one file, which makes managing and moving the database easier. Because there is no need for a server process it reduces overhead.

Quality of Output

As the datasets are clean and precise, the output of these datasets has also quality.

Potential Limitation for Final Report

Assumptions and Simplifications: Simplifications made in models may not capture all the complexities of climate systems