Analysis of the effect of forest and land areas on global temperature change and global warming

Main Question:

How do changes in forest and land cover affect global temperature changes and global Earth warming?

Data Sources:

To answer the main question of this project, two data sources from International Monetary Fund are used.

Data source 1: Annual Surface Temperature Change

- **Metadata URL:** https://climatedata.imf.org/datasets/406331492 3d74187be9596f10d034914/explore
- **Data URL:** https://opendata.arcgis.com/datasets/4063314923d74187be9596f10d034914_0.csv
- **Data Type:** CSV
- **Description:**
  This dataset presents annual estimates of mean surface temperature change measured with respect to a baseline climatology, corresponding to the period 1951-1980. This indicator presents in Degree Celsius, for the years 1961 – 2021 by country and for World.
  This data is provided by the Food and Agriculture Organization Corporate Statistical Database and is based on publicly available GISTEMP data from the National Aeronautics and Space Administration Goddard Institute for Space Studies.

Datasource2: Forest and Carbon

- **Metadata URL:** https://climatedata.imf.org/datasets/66dad9817da847b385d3b2323ce1be57/explore
- **Data URL:** https://opendata.arcgis.com/datasets/66dad9817da847b385d3b2323ce1be57_0.csv
- **Data Type:** CSV
- **Description:**
  Forests contain the largest stock of land-based carbon. Changes in forests, both in their extent and condition, can lead to low carbon storage and result in faster global warming.
  This dataset presents the data on land area, forest area, and carbon stocks in forests, for the years 1992 – 2020, Sourced from the Food and Agriculture Organization Corporate Statistical Database.

  The dataset includes the following indicators:

  - **Forest area:** The extent of forested land in various countries and regions.
  - **Land area:** Total land area for each country and region.
  - **Carbon stocks in forests:** Estimates of carbon stored in living biomass within forests.
  - **Share of forest area:** The indicator can be considered as identical to global SDG indicator 15.1.1 "Forest area as a proportion of total land area".
  - **Index of forest extent:** The index shows the magnitude of the forest area of a given year with reference to the base year 1992 that is depicted as 100.
  - **Index of carbon stocks in forests:** The index shows the magnitude of the carbon stocks in living biomass in forests of a given year with reference to the base year 1992 that is depicted as 100.

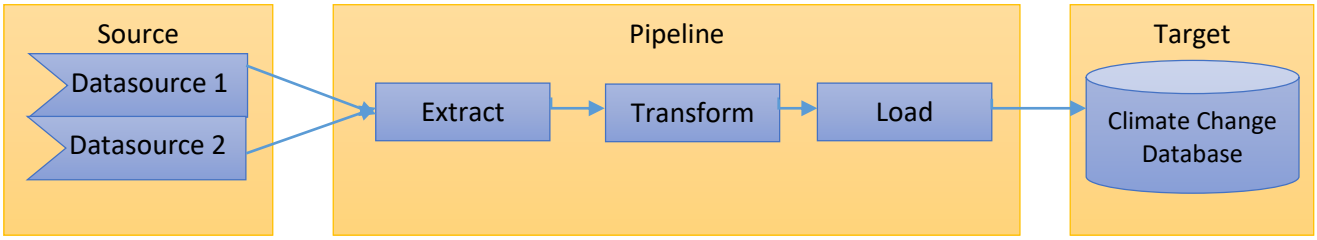
Data License:

Both datasets provided are sourced from the International Monetary Fund(IMF), licensed under an open data license CC BY-NC-SA 3.0 IGO, allowing sharing and adaptation with attribution for non-commercial purposes. More information about the license is available on this page.

Data Pipeline:

The data pipeline used in this project is an ETL pipeline, which includes three main stages: Extracting, transforming, and loading the data into a SQLite database, as shown in the following diagram.

Several functions are implemented to run this pipeline. In the first stage, extraction, data from the two data sources is extracted and stored into respective dataframes. In the next stage, transformation, data cleaning methods are applied to the dataframes. Methods such as deleting unnecessary columns, correcting data errors, data imputation and removing duplicate data. And in the last stage, loading, the cleaned dataframes are loaded into their corresponding tables into the Climate Change SQLite database.



Result:

The output of this data pipeline is a database which contain tables Annual_Surface_Temperature, and Forest_and_Carbon, which are the cleaned version of the main data sources. And additional seven tables, which each of them presented data of specific indicators needed to analyses and answer the main question of the project: Forest area, Index of forest extent, Land area, Share of forest area, Carbon stocks in forests, Index of carbon stocks in forests and Temperature.