

analysis-report

July 3, 2024

1 Annual Surface Temperature and Land Cover Analysis Report

1.1 Introduction

Climate change is the most critical issue in contemporary times at a global level. Surface temperature change and land cover change are the two main facets of climate change. This report attempts to present a detailed analysis regarding the role of land cover in regulating the global climate and changes in surface temperature. We are going to set up an automated data pipeline that cleans, processes, and analyzes these datasets for meaningful insights using the data goaltending from the Food and Agriculture Organization Corporate Statistical Database and NASA Goddard Institute for Space Studies.

1.2 Used Data

The analysis utilizes two primary datasets:

1.2.1 Annual Surface Temperature Change:

- **Source:** FAOSTAT, based on NASA GISS data.
- **Period:** 1961-2021
- **Purpose:** To analyze global temperature changes using a 1951-1980 baseline.

1.2.2 Land Cover and Land Cover Altering Indicator:

- **Source:** FAOSTAT
- **Period:** 1992-2018
- **Purpose:** To understand the impact of different types of land cover on climate regulation and carbon sequestration.

1.3 Data License

The datasets are used under the Copernicus Programme, which grants everybody free, non-exclusive, and perpetual access for lawful purposes, including reproduction and distribution. Proper attribution is given to the Copernicus Programme, ensuring compliance with their licensing terms.

1.4 Analysis

1.4.1 Data Pipeline Overview

The data pipeline implemented for this project includes the following steps:

1. **Data Acquisition:** Fetching CSV files from provided URLs.
2. **Data Cleaning:** Removing missing values to ensure data quality.
3. **Data Storage:** Storing cleaned data in a SQLite database for further analysis.

1.4.2 Technologies Used

- **Python** for scripting
- **Pandas** for data manipulation
- **SQLAlchemy** for database operations
- **Matplotlib** for visualization

1.4.3 Data Pipeline Implementation

The pipeline functions to fetch, clean, and store data from the specified sources. Here's an overview of the pipeline implementation:

```
[1]: import io
import requests
import pandas as pd
from sqlalchemy import create_engine
import matplotlib.pyplot as plt

file_directory = "../data"

# Function to run the entire data pipeline
def run_pipeline(data_url, table_name):
    table = fetch_process_data(data_url)
    save_to_sql(table, table_name)
    return table

# Function to fetch and process data
def fetch_process_data(url):
    try:
        response = requests.get(url).content
        dataframe = pd.read_csv(io.StringIO(response.decode('utf-8')))
        dataframe = dataframe.dropna() # Data cleaning step
        return dataframe
    except Exception as e:
        print(f"Error: {e}")
        return None

# Function to save data to SQL database
def save_to_sql(table, table_name):
    if table is not None:
        table_path = f"{file_directory}/{table_name}.db"
        engine = create_engine(f'sqlite:/// {table_path}')
        table.to_sql(table_name, con=engine, if_exists='replace', index=False)
    else:
```

```

        print("Error: Table is empty")

# Data sources with their respective URLs
data_sources = {
    "surface_temperature": "https://opendata.arcgis.com/datasets/
↳4063314923d74187be9596f10d034914_0.csv",
    "land_cover": "https://opendata.arcgis.com/datasets/
↳b1e6c0ea281f47b285addae0cbb28f4b_0.csv"
}

dataframes = []
# Running the data pipeline for each dataset
for table_name, url in data_sources.items():
    dataframes.append(run_pipeline(url, table_name))

def plot_data(df, title, x_label, y_label, label, trim_index, output):
    # Filter for specific countries
    countries = ['Germany', 'France', 'Italy', 'Spain', 'United Kingdom',
↳'United States']
    filtered_df = df[df['Country'].isin(countries)]

    # Select the columns from the given trim_index onward for the years
    year_columns = df.columns[trim_index:]

    plt.figure(figsize=(14, 6))
    for country in countries:
        country_data = filtered_df[filtered_df['Country'] == country]
        if trim_index == 11:
            country_data = country_data[country_data['Indicator'] == 'Climate_
↳Altering Land Cover Index']
            yearly_data = country_data[year_columns].mean(axis=0)
            plt.plot(yearly_data, label=f'{label} - {country}')

    plt.title(title)
    plt.xlabel(x_label)
    plt.ylabel(y_label)
    plt.xticks(rotation=45)
    plt.legend()
    plt.grid(True)
    # plt.show()
    plt.savefig(f'{output}_plot.png')
    plt.close()

plot_data(dataframes[0], "Annual Temperature from 1961 to 2022", "Year",
↳"Temperature", "Annual Temperature", 10, "temperature")

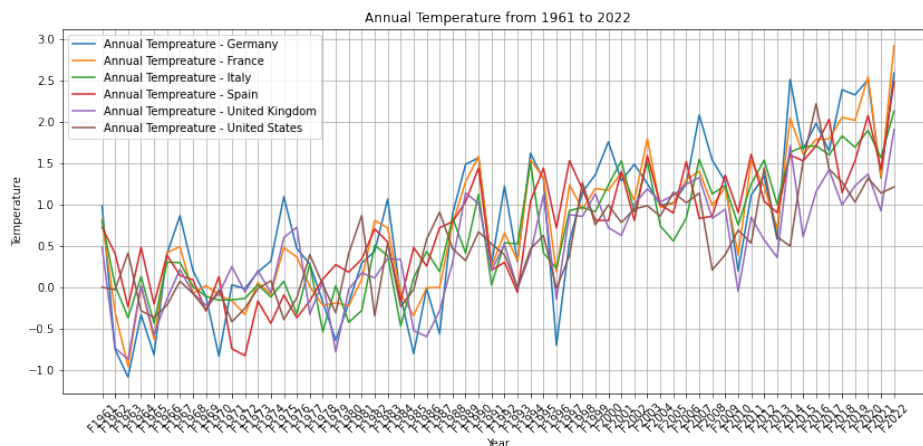
```

```
plot_data(dataframes[1], "Land Cover from 1992 to 2018(Indicator: Climate_
↳Altering Land Cover Index)", "Year", "Land Cover", "Land Cover", 11,↳
↳"land_cover")
```

1.5 Results and Discussion

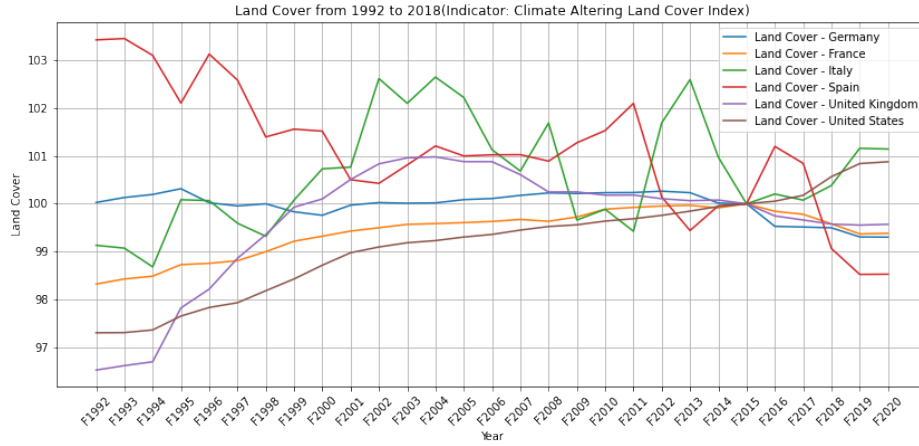
1.5.1 Surface Temperature Analysis

Surface temperature data for 1961-2021 indicate very well the continuous rate of increase of global temperatures, hence confirming global warming. This next figure clearly depicts it with annual temperature trends for some countries: This plot will clearly indicate the rise in temperatures for all regions considered here, hence evidencing a huge warming pattern over six decades. Insights on this kind of information are important to understand the full RAMifications of climate change globally.



1.5.2 Land Cover Analysis

The dynamics of different land cover contributing toward carbon sequestration and climate regulation are portrayed by the 1992–2018 land cover data. Land cover dynamics is shown in the figure below for selected indicators: The plot highlights the dynamics of land cover types with effects on the indicators of climate change. Especially, the dynamics in forest cover and agricultural land use are key drivers associated with carbon sequestration. These findings help identify critical areas for policy intervention and land management practices to mitigate climate impacts.



1.5.3 Limitations

The datasets are limited to publicly available data and might not capture all geographical nuances, which could affect the granularity of the analysis.

1.6 Conclusions

It evidences the critical role of automated data pipelines in processing and analyzing complex datasets. The insights derived from the Annual Surface Temperature and Land Cover data are very important for the policymakers to understand and mitigate the effects of climate change. This work has, despite these inadvertent findings, intrinsic limitations to the analysis based on data availability and coverage, thus indicating that the next lines of research require even more comprehensive sources of data.