data-report

June 5, 2024

1 Annual Surface Temperature and Land Cover Analysis Report

This report details the data processing and analysis performed on the Annual Surface Temperature Change and Land Cover and Land Cover Altering Indicator datasets. The data is sourced from the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT) and National Aeronautics and Space Administration Goddard Institute for Space Studies (NASA GISS).

1.1 Project Goal

The objective of this project is to implement an automated data pipeline that cleans, processes, and analyzes the dataset to provide insights into the effects of land cover on climate regulation and surface temperature changes globally.

1.2 Data Sources

1.2.1 1. Annual Surface Temperature Change

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

1.2.2 2. Land Cover and Land Cover Altering Indicator

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

1.3 License

In preparing this report, I have utilized data provided under the Copernicus Programme, specifically leveraging information from the Copernicus Climate Change Service. The use of such data is governed by a license agreement that grants free, non-exclusive, and perpetual access to Copernicus Products for lawful purposes, which include reproduction, distribution, communication to the public, and modification. In compliance with the licensing terms, appropriate attribution has been given to the Copernicus Programme in all derived data visualizations and analyses presented herein. This attribution ensures transparency in the source of the data, and it is stated explicitly that neither the European Commission nor the ECMWF is responsible for any use of the data as per clauses outlined in the license. This adherence not only reinforces the credibility of our

analysis but also respects the intellectual property rights held by the European Union concerning the Copernicus Products.

1.4 Data Pipeline Overview

The data pipeline consists of 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.

1.4.1 Technologies Used

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

1.5 Setup and Imports

1. Importing required libraries
2. Defining the directory to store database files

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

file_directory = "../data"
```

1.6 Data Pipeline Implementation

[2]:
```
# 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
```

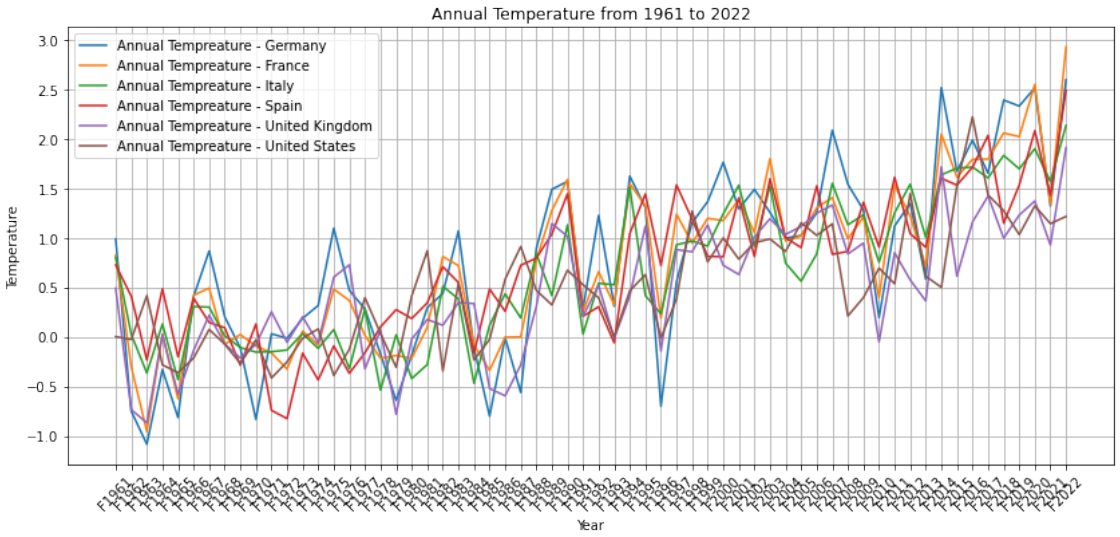
```
def save_to_sql(table, table_name):
    if table is not None:
        table_path = f"{file_directory}/{table_name}.db"
        engine = create_engine(f'sqlite:///{table_path}')
        table.to_sql(table_name, con=engine, if_exists='replace', index=False)
    else:
        print("Error: Table is 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))
```

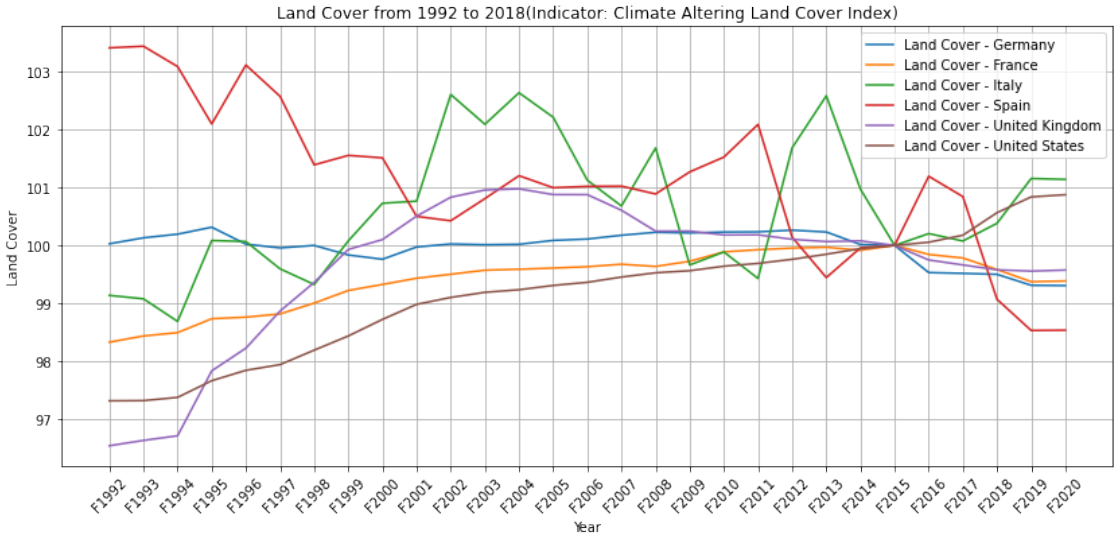
1.7 Results and Discussion

The analysis of surface temperature shows a consistent increase over the last six decades, underscoring the effects of global warming. The land cover data, once processed, will provide insights into how changes in land cover types contribute to carbon sequestration.

[3]:
```
def plot_data(df, title, x_label, y_label, label, trim_index):
    # 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()

plot_data(dataframes[0], "Annual Temperature from 1961 to 2022", "Year",␣
 ↪"Temperature", "Annual Tempreature", 10)
plot_data(dataframes[1], "Land Cover from 1992 to 2018(Indicator: Climate␣
 ↪Altering Land Cover Index)", "Year", "Land Cover", "Land Cover", 11)
```





1.8 Limitations

The datasets are limited to publicly available data and might not capture all geographical nuances.

1.9 Conclusion

This report highlights the crucial role of automated data pipelines in analyzing and interpreting complex datasets. The insights derived from the Annual Surface Temperature and Land Cover data can assist policymakers in understanding and mitigating climate change impacts.