

PROJECT REPORT

Analyzing the Impact of Climate Change on Crop Production and CO2 Emissions in the USA (1990-2015).

INTRODUCTION :

Climate change is a significant global issue due to its far-reaching impacts on our environment, weather patterns, ecosystems, and human society. These impacts include rising temperatures, extreme weather events, and increasing sea levels, as well as escalating CO2 emissions. This project aims to analyze historical data on CO2 emissions and crop production in the USA from 1990 to 2015 to gain insights into how climate change has affected these areas over time.

DATASOURCES :

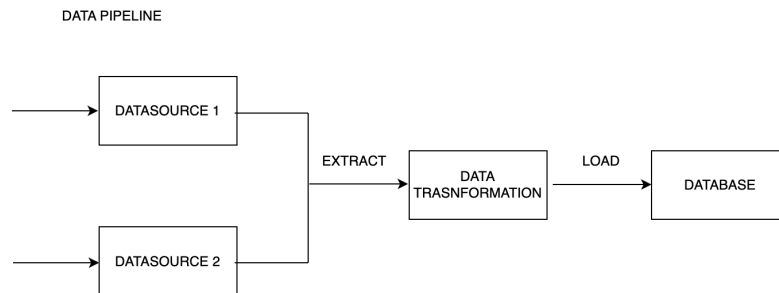
Crop Production & Climate Change

- **Description:** This dataset, sourced from Kaggle, provides information on crop yields, harvested areas, and production quantities for key crops such as wheat, maize, rice, and soybeans. Crop yields are measured in tonnes per hectare, with data available from 2010-2016.
- **Domain-Specific Values:** The data includes metrics like harvested production per unit area and total production in tonnes.
- **Data License Compliance:** The dataset is used in accordance with Kaggle's licensing terms.

CO2 Emissions

- **Description:** This dataset, also from Kaggle, offers a comprehensive overview of CO2 emissions by country from 1960 to 2023. It includes annual CO2 emission data for all countries worldwide.
- **Domain-Specific Values:** The data covers annual CO2 emissions in metric tonnes.
- **Data License Compliance:** The dataset is used in accordance with Kaggle's licensing terms.

DATA PIPELINE :



In this project, a pipeline is created using the Python programming language. The pipeline follows the processes listed below.

1. EXTRACTION :

This step includes extracting the data from the source. We have a CSV dataset from Kaggle. The python code will download the datasets into a temporary directory using api authentication, and it will then load them into a pandas dataframe.

2. TRANSFORMATION :

The transformation of data involves several critical processes to ensure it is clean, filtered, and consistent. In this project, data cleaning encompassed renaming dataset columns for clarity, and converting time and year fields to a consistent datetime format as a best practice. Data for the USA region from 1990 to 2015 was filtered. Using group by, I calculated the year's averages and combined the datasets into a single table based on the date.

3. LOAD :

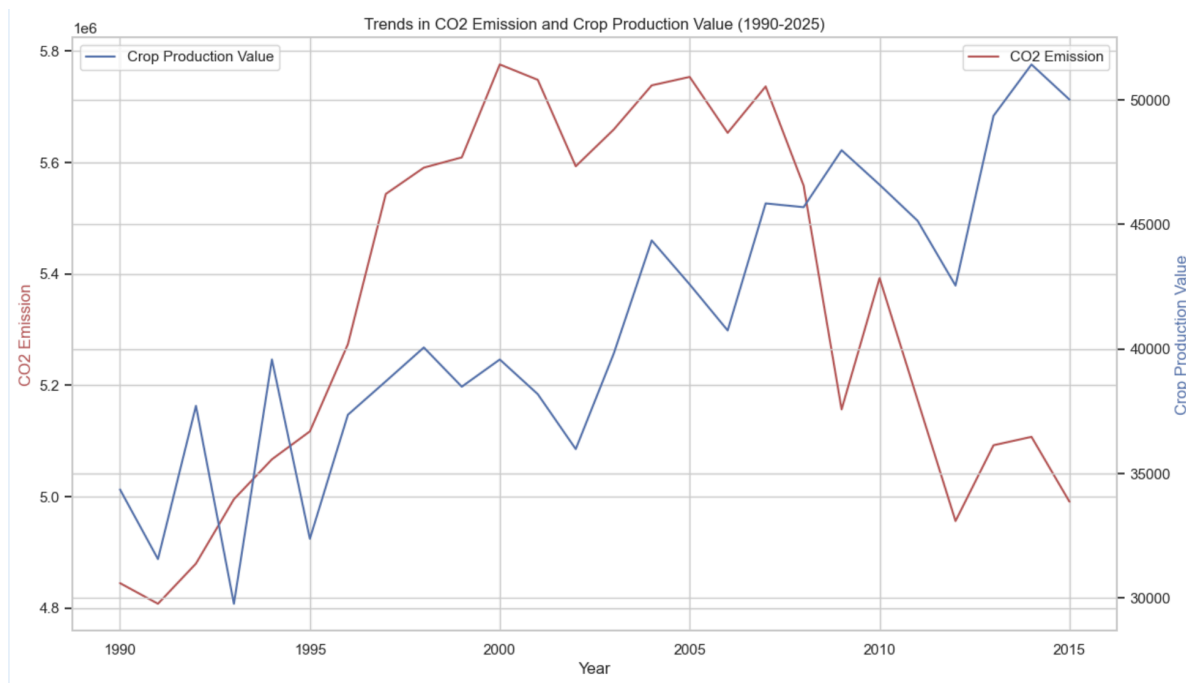
This is the final step of the pipeline, which involves storing the data in a suitable format within the database.

PROBLEMS ENCOUNTERED :

- 1) **Initial Approach:** Initially, datasets were saved to GitHub and accessed directly in the code, which proved cumbersome.
- 2) **API Authentication:** Later, API authentication was used for a more streamlined process.
- 3) **Best Practices:** Good programming practices were implemented after a relevant lecture.
- 4) **Date Inconsistency:** Challenges with date format consistency were addressed during data transformation.

ANALYSIS :

The analysis focused on examining the trends in crop production and CO2 emissions over the 25-year period. Methods included statistical analysis and visualization to interpret the data patterns.



METHOD:

Statistical analysis and data visualization using Python libraries (Pandas, Matplotlib).

RESULTS:

- Showed variability over the period, influenced by factors such as industrial activities and changes in energy consumption.CO2 emissions showed a VARIABILITY.
- Demonstrated a consistent increase despite variability, driven by advancements in agricultural technology and practices, as well as natural climate variability.

CONCLUSION :

The analysis indicates a correlation between rising CO2 emissions and changes in crop production over the examined period. However, it is crucial to note the limitations of the datasets, particularly the crop production data's limited timeframe.