

Analyzing the temperature change and crop yield dynamics in Ireland from 2009 to 2023

The project aims to explore and understand the relationship between temperature change and crop yield in Ireland. Understanding this relationship is crucial for following reasons:

- Selecting the appropriate crops can help to improve the overall yield.
- Increases awareness of temperature impact on crops.

1 Question

How is temperature change impacting the crop yield over the years and if there is any pattern that can be realized by visualization?

2 Data Sources

2.1 Temperature change countrywise

- **Metadata URL:** <https://www.fao.org/faostat/en/#data/ET/metadata>
- **Data URL:** https://bulks-faostat.fao.org/production/Environment_Temperature_change_E.Europe.zip
- **Data Type:** CSV (Structured and Tabular Dataset)
- **Description:** The CSV file contains temperature change data for Ireland from 1961 to 2023.
- **License:** Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO (CC BY-NC-SA 3.0 IGO)

2.2 Crop Yield (2008 - 2023) - Ireland

- **Metadata URL:** <https://data.gov.ie/dataset/aqa04-crop-yield-and-production/resource/ca2113ee-d9f3-4654-acc6-777213ae9330>
- **Data URL:** https://ws.cso.ie/public/api.restful/PxStat.Data.Cube_API.ReadDataset/AQA04/CSV/1.0/en
- **Data Type:** CSV (Structured and Tabular Dataset)
- **Description:** The CSV file contains data of crop yield in Ireland.
- **License:** Creative Commons Attribution 4.0

2.3 Compliance with Data Source Licenses

- **Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO (CC BY-NC-SA 3.0 IGO):**
 - **Attribution:** Proper credit has been given to the original authors. The sources are cited with links to the original data and the license.
 - **NonCommercial:** The data is used strictly for educational and research purposes, with no commercial use.
 - **ShareAlike:** Any derivatives of this data will be shared under the same license.
 - **License URL:** <https://creativecommons.org/licenses/by-nc-sa/3.0/igo/>
- **Creative Commons Attribution 4.0 (CC BY 4.0):**
 - **Attribution:** Proper credit has been given to the original authors. The sources are cited with links to the original data and the license.
 - **License URL:** <https://creativecommons.org/licenses/by/4.0/>

3 Data Pipeline

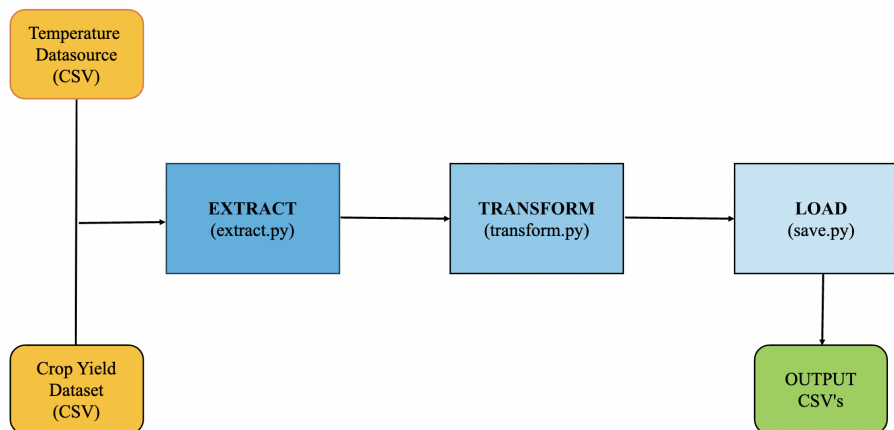


Figure 1: Data Pipeline

3.1 Running the Pipeline

The `pipeline.py` script manages the data processing workflow, serving as an ETL pipeline for both *temperature* and *crop yield* data.

3.1.1 Data Extraction

The data extraction module, `extract.py`, contains functions to retrieve temperature data and crop yield data from a given URL.

- **`get_temperature_data()`:** Downloads a ZIP file from the given URL, extracts the specified CSV file, and returns it as a Pandas DataFrame.
- **`get_crop_yield_data()`:** Downloads and returns the crop yield data from the given URL as a Pandas DataFrame.

3.1.2 Data Transformation

The data transformation module, `transform.py`, contains methods for processing temperature and crop yield data.

- **process_temp_data():** This process involves taking the temperature DataFrame and refining it to collect data. The data is classified into temperature change and standard deviation categories, which are then subdivided into annual, seasonal, and meteorological variations. Once the irrelevant columns are deleted to improve clarity, the data is organized into a dictionary. The columns considered for data analysis are *months* and *Years from 1961-2023*.
- **process_crop_data():** This function filters the crop yield DataFrame to focus on crop production and yield per hectare data. It removes unnecessary columns to highlight important data and structures the manipulated information into a dictionary with unique keys representing different factors of crop production. The columns considered for data analysis are *Statistic Label*, *Year*, *Type of Crop*, *VALUE*.

3.1.3 Saving Dataframes

The `save.py` module manages the saving of DataFrames to CSV files in the data directory.

- **save_temperature_data_to_csv():** This method stores temperature change and standard deviation DataFrames in CSV files within a `data` directory. Inputting a dictionary that includes DataFrames for various temperature change and standard deviation categories is required.
- **save_crop_data_to_csv():** This approach stores crop production and crop yield per hectare DataFrames in CSV files within the `data` folder. It requires a dictionary that includes DataFrames for these specific categories as an input too.

4 Results and Limitations

4.1 Preferred Data Format for Pipeline Output

The reason why the pipeline outputs are stored in CSV format is due to its simplicity, compatibility, and human readability, which makes it perfect for handling structured datasets. The column *Year*, serves as a pivotal factor for correlation analysis in both temperature and crop yield datasets.

4.2 Evaluation of the accuracy of data and Possible Challenges

- **Temperature Data:** The accuracy of temperature data is reliant on the uniformity of temperature monitoring across different regions of Ireland. Discrepancies may arise due to variations in monitoring practices or equipment calibration among different cities. I had to classify the dataset into *temperature change* and *standard deviation* categories, further subdividing it into *annual*, *seasonal* and *meteorological* variations for accurate analysis.
- **Crop Yield Data:** Similarly for this dataset, it had to be divided into *crop production* and *yield per hectare data* csv's for further analysis