# Analyzing the Impact of Climate Change on Agricultural Productivity in Latin America

(Name: Zannatul Ferdousi,  Mat. Num: 23456320)

1. **Question:**

***How does climate change impact agricultural productivity in Latin America from 2015 to 2021?***

2. **Objective**

This project aims to analyze key metrics, including greenhouse gas emissions, land use, and crop production, to identify patterns and trends related to agricultural sustainability.

3. **Data Source**

**Source 1: FAOSTAT Agricultural Data**

**Data Source Name:** Crops and livestock products

**Description:** Provides crop and livestock production data globally, including metrics such as yield, production volume, and harvested area.

**Reason for Selection:** Comprehensive dataset on agricultural productivity, covering Latin American countries at a granular level.

**Structure and Quality:** The data is presented as a wide-format CSV, with rows representing countries and columns for years (Y1961 to Y2022). It includes metadata on crops, regions, and measurement units, ensuring clarity.

**License:** The data is published by FAO under an open data license. Obligations include attribution and non-commercial use.

**Compliance:** Obligations fulfilled by citing the source: FAOSTAT Open Data License.

**Source 2: World Bank Climate Data**

**Data Source Name:** Average precipitation in depth

**Description:** Contains historical climate data, including greenhouse gas emissions and land use statistics for countries globally.

**Reason for Selection:** Provides a strong foundation to study climate change and its direct implications on agricultural productivity.

**Structure and Quality:** The dataset is structured as a wide-format CSV with rows for countries and columns for years (1960 to 2022).

**License:** Distributed under the World Bank Open Data Terms of Use, permitting reuse with attribution.

**Compliance:** Obligations fulfilled by citing the source: World Bank Open Data Terms.

## 4. Data Pipeline

**Overview:**

The data pipeline automates data retrieval, transformation, and cleaning. It integrates datasets from FAOSTAT and the World Bank, aligning them to focus on Latin American countries and the period 2015–2021. The pipeline is implemented in Python, utilizing libraries like pandas for data processing and SQLite3 for structured storage.

**Steps and Technologies:**

- **Data Download:**
  1. Automated download of ZIP and CSV files from source URLs.
  2. ZIP files are extracted, and the primary datasets are identified and loaded.
- **Filtering and Cleaning:**
  1. Focused on Latin American countries.
  2. Filtered years to retain only 2015–2021.
- **Transformation:**
  1. Reshape data from wide format (years as columns) to long format (years as rows) using pandas.melt().
  2. Standardizes year formats (Y2015 → 2015) for consistency.
- **Error Handling:**
  1. Validates data integrity by skipping rows with missing values in critical columns.
  2. Logs and reports errors during file extraction or processing steps.
- **Final Storage and Output:**
  1. Cleaned datasets are saved in long-format CSVs and an SQLite database.
  2. Output includes metrics like crop production volume, greenhouse gas emissions, and land use data.


Fig 01: Pipeline Workflow

**Workflow:** Data Download → Data Extraction → Filtering (Countries & Years) → Wide-to-Long Transformation → CSV & SQLite Output

## 3. Problems and Solutions

**Issue:** Discrepancy in year formats (e.g., Y2015 in FAOSTAT vs. 2015 in World Bank). Missing data for some countries in specific years.

**Followed Solution:** Implemented standardization logic to harmonize year formats across datasets. Excluded incomplete rows while documenting data gaps for transparency.

## 4. Results and Limitations

**i. Structure:** Long-format tables with columns:

**FAOSTAT:** Area, Item, Element, Unit, Year, Value

**World Bank:** Country Name, Country Code, Indicator Name, Indicator Code, Year, Value

**ii. Format:**

**CSV:** Portable and easy to visualize in tools like Excel or Python.

**SQLite:** Efficient for querying and integrating into data analysis pipelines.

**iii. Content:**

**FAOSTAT:** Agricultural production metrics such as crop yield and volume.

**World Bank:** Climate indicators like precipitation and land use.

**iv. Data Quality:**

**Completeness:** Only rows with valid data for the specified countries and years are included.

**Accuracy:** Transformation steps preserve original values while improving structure for analysis.

**v. Limitations:**

Annual data may not capture seasonal variations or short-term trends. Some countries lack complete data for all years, potentially affecting regional analysis. Updates or changes in source data structures could necessitate adjustments to the pipeline.

## 5. Conclusion:

The project successfully automated the integration and transformation of agricultural and climate datasets for Latin American countries, providing a clean, unified structure for analysis. The long-format outputs enable seamless exploration of trends in agricultural productivity and environmental factors between 2015 and 2021. Despite minor limitations, such as missing data for some countries, the resulting datasets are robust and ready for further insights into regional sustainability and resource management.