Analysis and Visualization of Agricultural Data based on the impact of Climate Change through ETL process.

Presentation by: Zeeshan Ahmed
M.Sc Artificial Intelligence

Module: Methods of Advanced Data Engineering (MADE)
Teaching Team: made-team@group.riehle.org
Fau Erlangen-Nürnberg

Introduction

Background

- Climate change is significantly impacting agricultural productivity and food security worldwide.
- Rising mean surface temperatures affect crop yields differently across regions.
  - **Temperate Regions:**
    - Potential benefits from longer growing seasons (e.g., Turkey, Germany).
  - **Tropical Regions:**
    - Increased risk due to heat stress as crops near optimal temperature limits (e.g., Pakistan, Australia).

Project Objective

- Build an ETL Pipeline:
  - Assess the impact of climate change on agricultural data.
  - Provide valuable insights on temperature variations and crop yield effects in diverse climates.

Key Research Questions

- What impact has climate change based on mean surface temperature variation had on crop yield in countries with diverse climates like Pakistan, Germany, Turkey, and Australia in the last years?
- Have there been any trends or recurring patterns, in the yields of key crops such as carrots, turnips, maize, rice, and grapes?

Data Sources

DataSource-1: Crop Production

- **Data Type:** CSV
- **Licence:** CC BY-NC-SA
- **Source:**
  - This dataset is taken from Food and Agriculture Organization of the United Nations (FAO).

Description:

- **Timeframe**: 1961 to 2019
- **Products**: 173 types (cereals, vegetables, fruits, tree nuts, fiber crops, oil crops, pulses, roots, tubers).
- **Measurement**: Harvested areas: ha, Production quantities: tonnes, Yields: hg/ha.

DataSource-2: Mean Surface Temperature Changes

- **Data Type:** CSV
- **Licence:** CC0 1.0
- **Source:**
  - This data is provided by FAOSTAT and is based on publicly available GISTEMP data from the NASA GISS.

Description:

- **Timeframe**: 1970 to 2021
- **Data Source**: Weather stations, satellites, ocean buoys.
- **Measurement**: Degrees Celsius.
- **+ve Index**: Increase in temperature.
- **-ve Index**: Decrease in temperature.

Methods: ETL Pipeline Overview

ETL Pipeline Construction: Developed using Python3 to extract datasets from two sources:

- Crop production data from datasource-1
- Historical temperature records from datasource-2

Data Transformation and Loading:

- Standardized formats, integrated relevant variables, and filled missing values using linear interpolation and forward filling techniques.
- Merged dataframes using "Country Name" as the key, focusing on matching country names, crops of interest, and year-wise temperature readings.
- Retained the period from 1970 to 2019 to maintain consistency.
- Chose SQLite database format for loading the final data.

Logging and Error Handling:

- Configured logging to track process flow, record errors, and ensure accountability.
- Implemented try-except blocks to manage exceptions, halt processing on critical errors, and log issues effectively.

Environment Cleanliness:

- Implemented procedures to clean up Kaggle dataset files after data extraction.

Technical Setup

Used Technologies:

- Python3 with visual studio, Git, GitHub.
- Libraries / Frameworks: Kaggle, Pandas, unittest, logging, SQLite, Matplotlib

Project Components:

- **project-plan.md:** Document used in conjunction with GitHub's issues feature to plan and manage project tasks.
- **ETL_Pipeline.py:** Python script implementing the Extract, Transform, Load (ETL) process to extract data from sources, transform it to fit operational needs, and load it into a SQLite database.
- **unit_test.py:** Contains unit and system tests to validate the functionality of the ETL pipeline.
- **pipeline.sh & tests.sh:** Scripts serving as entry points for the Continuous Integration (CI) pipeline, managed with GitHub Actions, ensuring automated testing and deployment workflows.
- **data-report.pdf:** Detailed description of the dataset used in the project, including its structure, source, and preprocessing methods.
- **analysis-report.pdf:** Comprehensive report presenting the findings, insights, and conclusions derived from analyzing the integrated dataset, focusing on the impact of climate change on crop yield variability from 1970 to 2019 across selected countries (Pakistan, Germany, Turkey, and Australia), with a specific focus on carrots, turnips, maize, rice, and grapes.

Analysis: Impact of Climate Change on Crop Yield in Diverse Climates

Analysis Objectives & Scope

- **Objective:** Analyze the impact of mean surface temperature variations on crop yields from 1970 to 2019.
- **Scope:** Focused on diverse climatic regions: Pakistan, Germany, Turkey, and Australia.
- **Key Crops:** Carrots, turnips, maize, rice, and grapes.
- **Method:** Time series plots used for analysis.

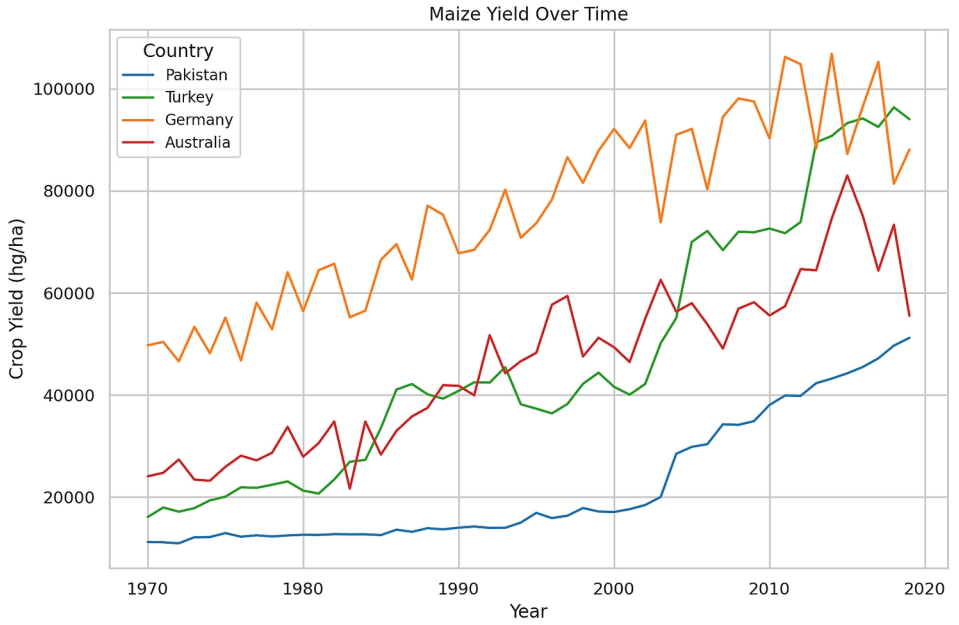
Analysis: Crop Yield Trends

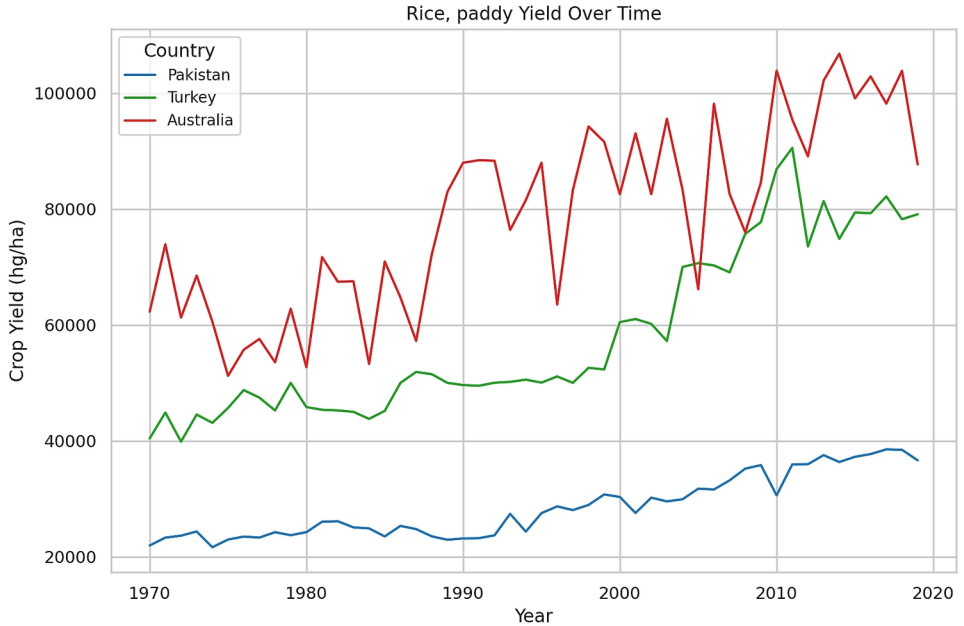
General Overview:

Crop yield trends analyzed for Pakistan, Germany, Turkey, and Australia (1970-2019).

Maize and Rice:

- **Maize Yield:**
  - **Germany:** Highest average yield at 75,972 hg/ha.
  - **General Trend:** Fluctuations with an overall upward trend
- **Rice Yield:**
  - **Australia:** Highest average yield at 78,909 hg/ha.
  - **General Trend:** Similar fluctuations and upward trend as maize.





Analysis: Crop Yield Trends

General Overview:

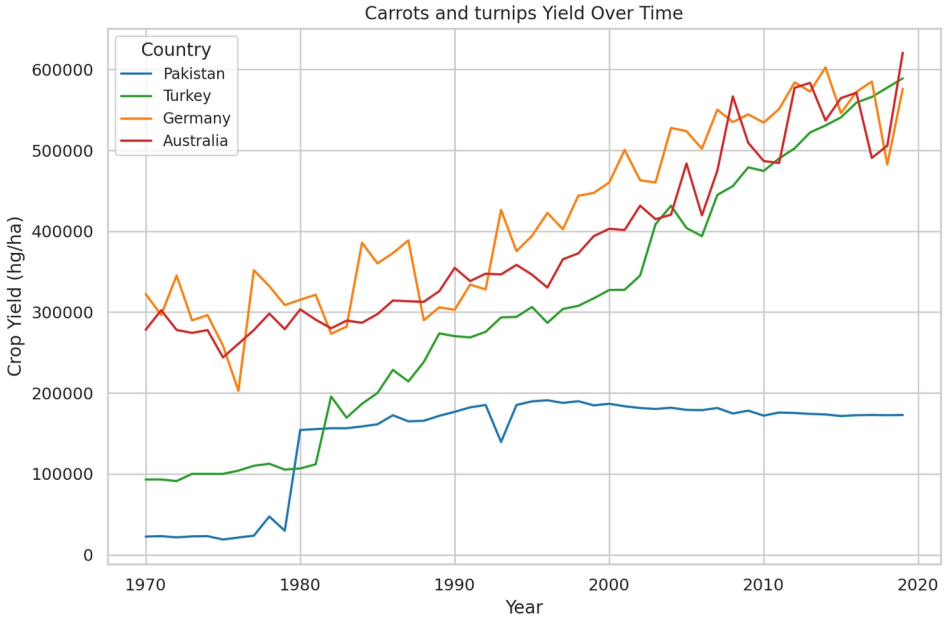
Crop yield trends analyzed for Pakistan, Germany, Turkey, and Australia (1970-2019).

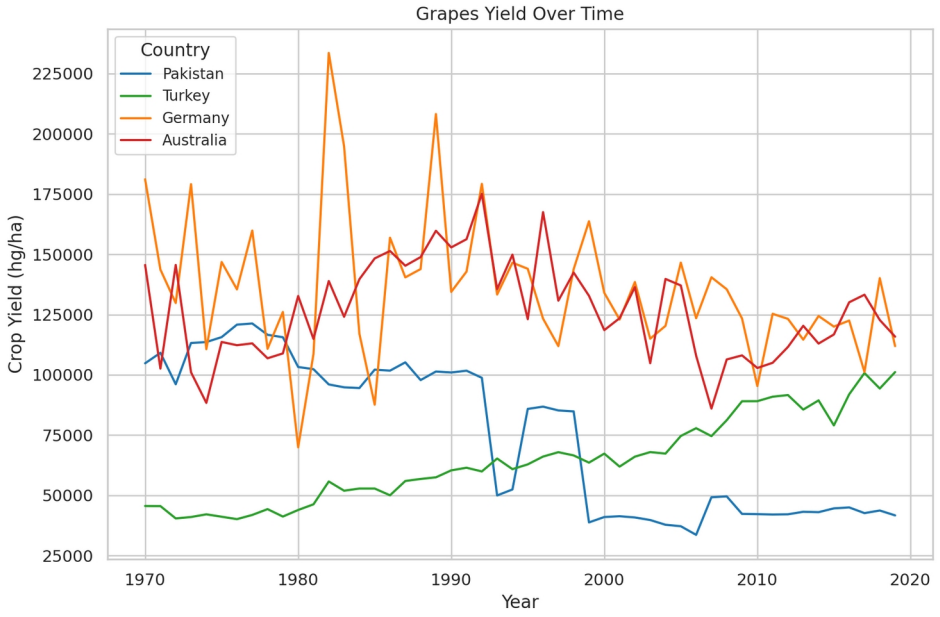
Carrots and Turnips:

- **Upward Trend:** Reached up to 600k hg/ha in Turkey, Germany, and Australia by 2019.
- **Decline in Pakistan:** Consistent decline, yields remained below 200k hg/ha from 1975 to 2019.

Grapes:

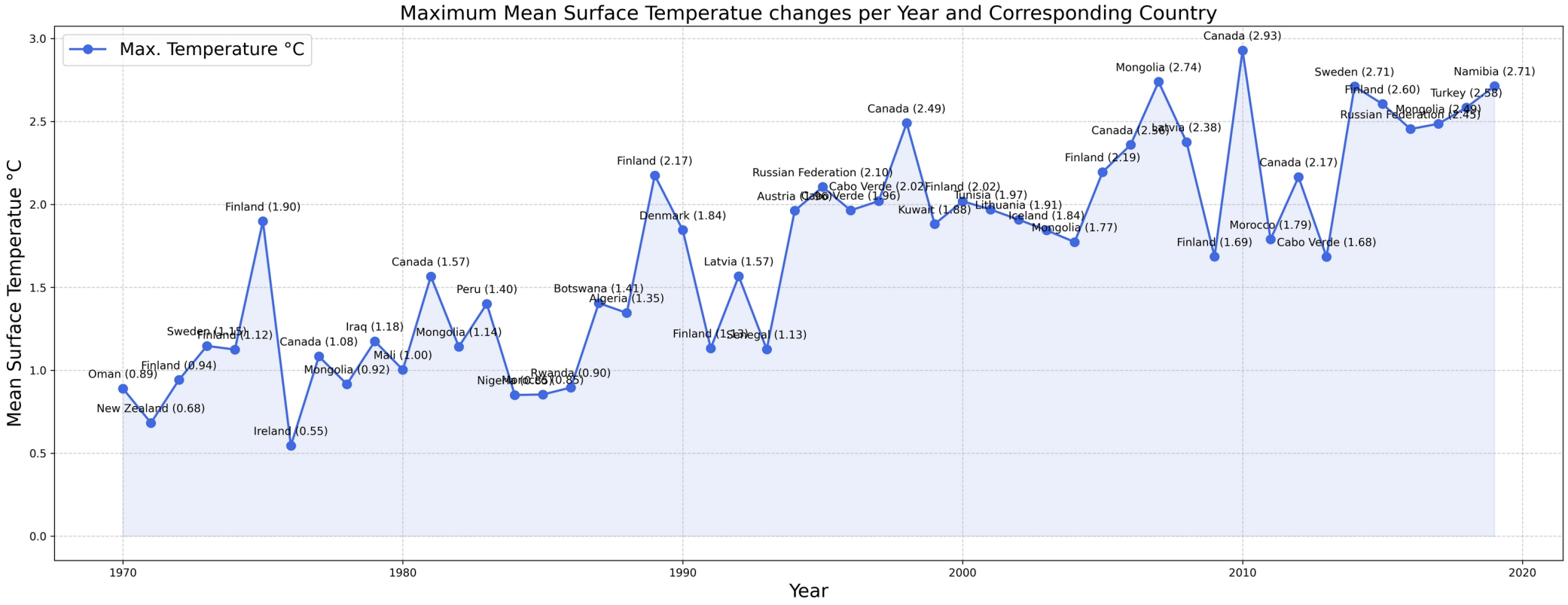
- **Variability:** Yields showed significant variability.
- **Influence:** Strongly affected by temperature changes and climate fluctuations.





Analysis: Maximum Mean Surface Temperature (1970-2019)

- **Overview**: Maximum mean surface temperatures recorded across various countries.
- **Trends**: Clear upward trend in mean surface temperatures.
- **Notable Observations**: Significant peaks, especially in Canada, reaching up to 2.93 °C.

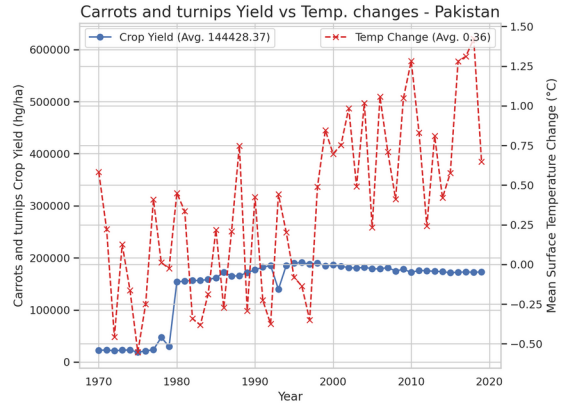


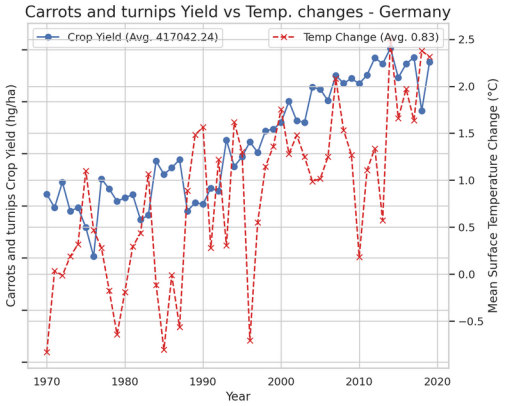
Analysis: Correlation Between Crop Yield and Temperature Change (1970-2019)

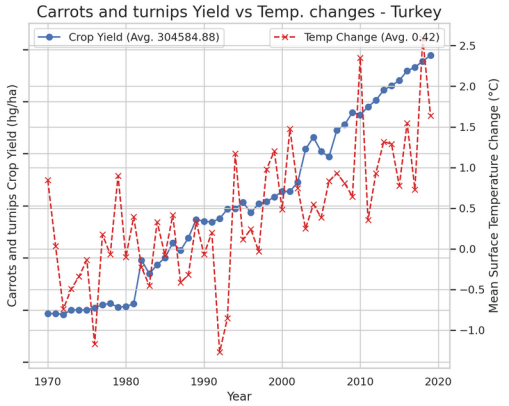
Analysis: Correlation between crop yields and mean surface temperature changes approximately by 1-2 °C for Pakistan, Germany, Turkey, and Australia.

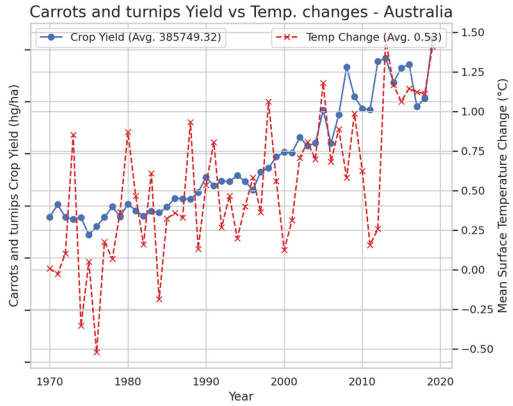
Carrots and Turnips Yield vs. Temperature Changes

- **Germany:** Achieved yields around 580,000 hg/ha by 2019 due to moderate temperature increases extending growing seasons.
- **Turkey:** Recorded yields of about 590,000 hg/ha by 2019, showing positive correlation with temperature changes.
- **Australia:** Experienced yields exceeding 500,000 hg/ha with improved growing conditions from temperature increases.
- **Pakistan:** Yielded below 200,000 hg/ha despite temperature rises, attributed to heat stress and water scarcity in its tropical climate.







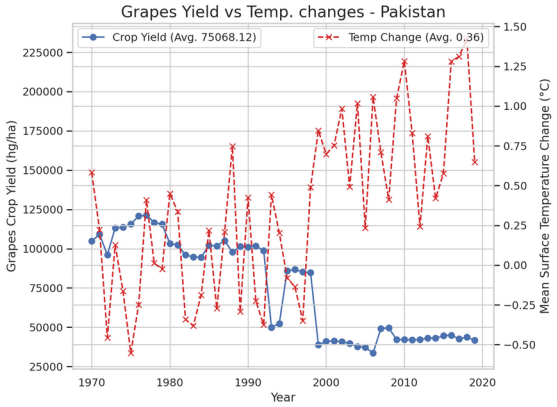


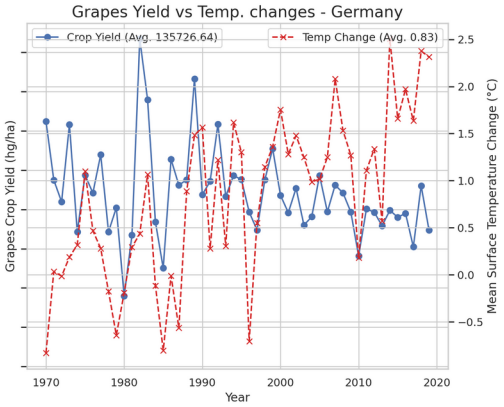
Analysis: Correlation Between Crop Yield and Temperature Change (1970-2019)

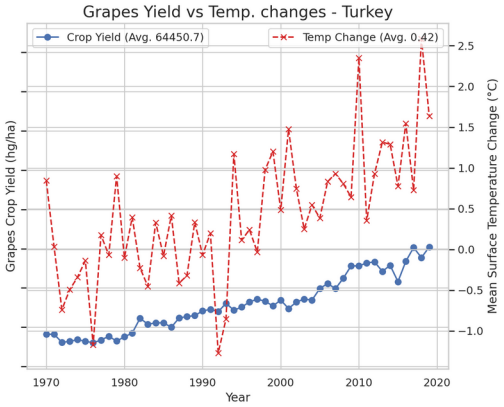
Analysis: Correlation between crop yields and mean surface temperature changes approximately by 1-2 °C for Pakistan, Germany, Turkey, and Australia.

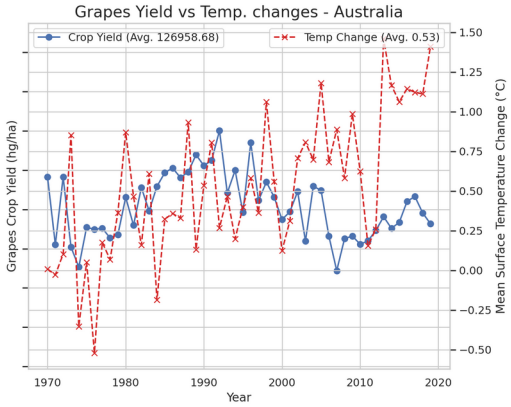
Grapes Yield vs. Temperature Changes

- **Germany:** Averaged 135,726 hg/ha, benefiting from temperature fluctuations extending the growing season.
- **Turkey:** Averaged 64,450 hg/ha, showing stable yields with positive correlation to moderate temperature increases.
- **Pakistan:** Averaged 75,068 hg/ha, with mixed results and negative correlations during extreme temperature changes.
- **Australia:** Averaged 126,958 hg/ha, initially resilient to temperature changes but later experienced declines due to high variations.







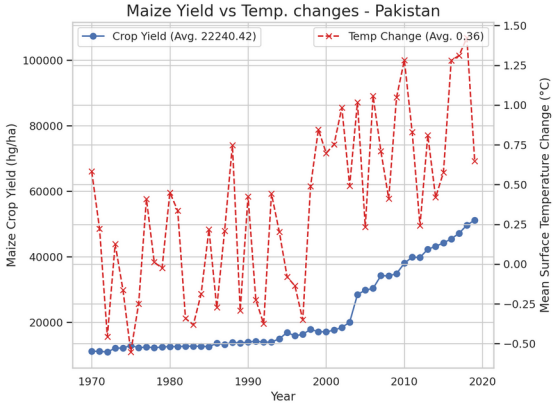


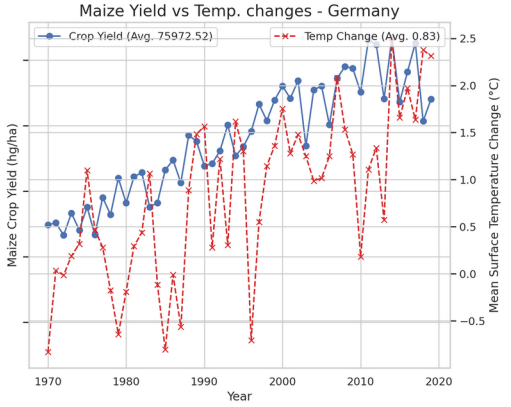
Analysis: Correlation Between Crop Yield and Temperature Change (1970-2019)

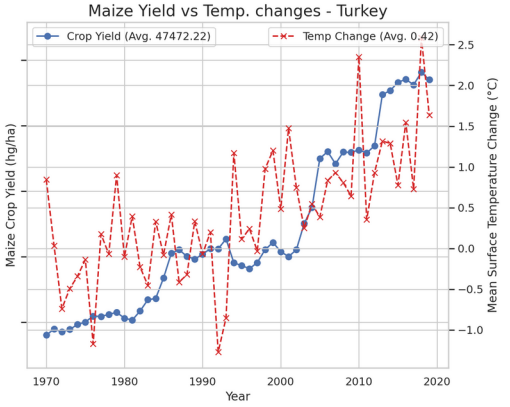
Analysis: Correlation between crop yields and mean surface temperature changes approximately by 1-2 °C for Pakistan, Germany, Turkey, and Australia.

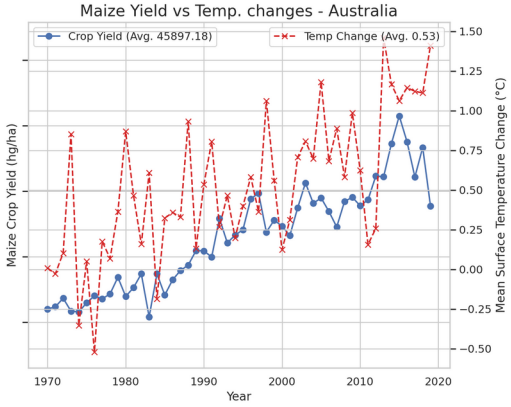
Maize Yield vs. Temperature Changes

- **Germany**: Maintained leadership with average yields of 75,972 hg/ha, leveraging favorable climatic conditions.
- **Turkey**: Achieved robust maize yields averaging approximately 47,472 hg/ha, benefiting from increasing temperatures.
- **Pakistan**: Experienced a steady increase in maize yield, averaging about 22,240 hg/ha, with fluctuations due to regional climate variations.
- **Australia**: Recorded average maize yields of 45,897 hg/ha, showing resilience to climatic changes with occasional fluctuations.









Analysis: Correlation Between Crop Yield and Temperature Change (1970-2019)

Analysis: Correlation between crop yields and mean surface temperature changes approximately by 1-2 °C for Pakistan, Germany, Turkey, and Australia.

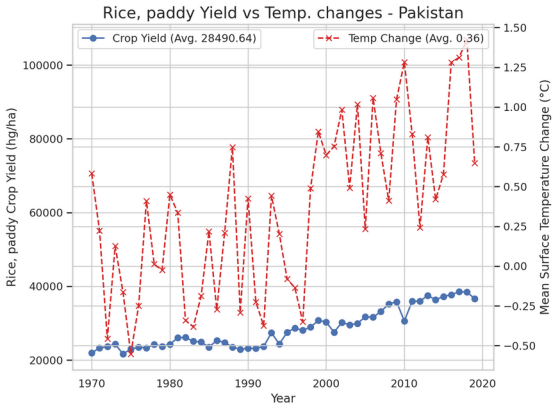
Rice Yield vs Temperature Changes

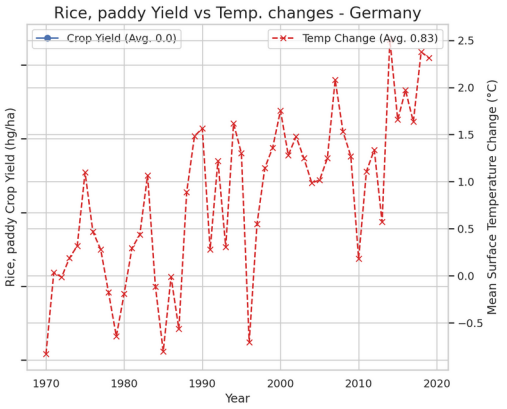
Germany: Did not produce rice, focusing on other crops within its agricultural practices.

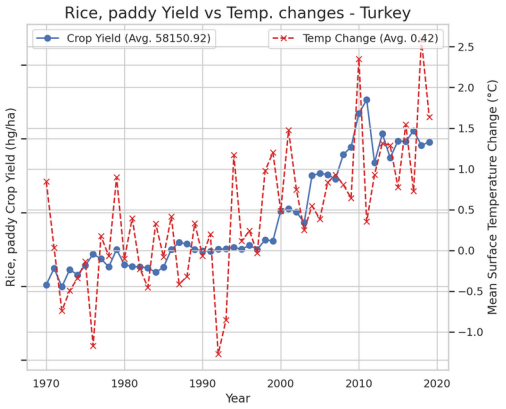
Turkey: Demonstrated consistent rice yields, averaging around 58,150 hg/ha, supported by favorable temperature changes.

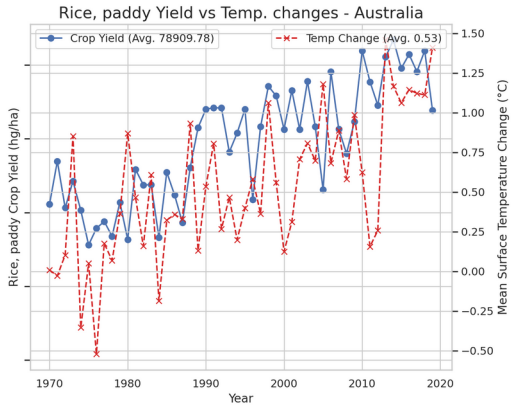
Australia: Experienced fluctuations in rice yields but achieved the highest average yield of approximately 78,909 hg/ha.

Pakistan: Showed a steady increase in rice yield, averaging about 28,490 hg/ha, benefiting from moderate temperature increases.









Discussion and Conclusion

- The analysis from 1970 to 2019 reveals significant impacts of rising mean surface temperatures on crop yields in Pakistan, Germany, Turkey, and Australia.
- **Maize and Wheat**: These crops showed adaptability to gradual temperature increases.
- **Carrots, Turnips, and Grapes**: More affected by extreme temperature variations.
- Countries with strong agricultural practices like Germany and Turkey sustained or increased yields despite climatic changes.

The ETL framework emphasizes the need for climate-resilient agricultural practices to ensure food security amid global climate shifts.

Limitations

- **Potential Issues in Data**: Data consistency challenges due to possible variations in source data. Need for adjustments if input data structures or schemas change over time.
- **Data Scope:** The analysis did not consider other climatic factors such as precipitation and extreme weather.
- **Regional Variability:** Local agricultural practices and policies, which also influence yields, were not accounted for.
- **Ongoing Changes:** Findings are based on historical data; future trends may differ due to continuing climate change.

Exploring Future Research Directions

Climate-Resilient Agricultural Practices

- What are the optimal strategies for integrated water management to enhance crop resilience in water-stressed regions?

Impact of Extreme Weather Events

- What are the long-term effects of extreme weather events, such as droughts and floods, on crop yields in diverse climatic regions?

Technological and Genetic Solutions

- How can advancements in biotechnology and genetics contribute to developing climate-resilient crop varieties?

Thank you for your attention!

For more information, visit my github repository:
https://github.com/ZeeshanAhmed13/made-template-23432274