Methods of Advanced Data Engineering

Analysis-Report

in the context of the Module “MADE”

at Friedrich-Alexander-Universität Erlangen-Nürnberg
at the Department Informatik
Professorship for Open-Source Software

Principal Supervisor:	Dirk Riehle
:	
Author:	Mubashir Ul Hassan Rothernburger Str.145 90439 Nuremberg +49 17634386460 mubashir.hassan@fau.de 23079239
Submission:	16th January 2025

1 Introduction

1.1 Question

How do climate conditions (Rise in Temperature) influence agricultural productivity in South and North America?

1.2 Purpose and Motivation

Climate change is causing rising temperatures that are increasingly impacting the agricultural sector. This analysis investigates the effect of the increase in temperature on agricultural crop production in North and South America over 52 years (1961–2013) as a function of the stable atmospheric boundary layer. It looks at countries such as Argentina, Brazil, Canada, and the US, examining how variation in average temperature impacts agricultural statistics like production, harvested area, and yield.

1.3 Overview of Dataset

Data Sources:

(1) **Climate Data (Climate Change: Earth Surface Temperature Data):**The average monthly temperature data is downloaded from Kaggle, which contains global land temperature records.

Link to Data: Download here

(2) **Agricultural Data (FAOSTAT):**The source for agriculture data is FAOSTAT and the data covers agriculture, stocks and animal production. These data are by year and cover a diverse set of agricultural goods in the America.

Link to Data: Download here

Data Structure and Meaning

Climate Data Columns	Description
`dt`	Date (monthly)
`AverageTemperature`	The temperature recorded for the month
`AverageTemperatureUncertainty`	Uncertainty of the temperature measurement
`Country`	The country where the data was recorded
`Year`	The year of the observation (derived from `dt`)

Table 1 Climate Data Columns

Agricultural Data Columns	Description
`Year`	Year of agricultural data
`Element`	Type of metric (e.g., Production, Area Harvested, Yield)
`Value`	The value of the agricultural metric for that year
`Country`	The country of the agricultural data point

Table 2 Agricultural Data Columns

1.4 Data Processing

(1) **Date Conversion and Aggregation:** This first step is to assume that the monthly temperature data is aggregated into annual average temperature.

(2) **Country Mapping:** Country names are standardized using a generic dictionary. This ensures consistency in both datasets when merge occur.

(3) **Merging Climate and Agricultural Data:** Climate data (including temperature) is merged manually with agriculture data (including production and yield) based on year and country to facilitate correlation analysis between temperature and agricultural metrics.

(4) **Handled Missing Values:**Data cleaning steps also handle missing values to prepare the merged dataset for analysis.

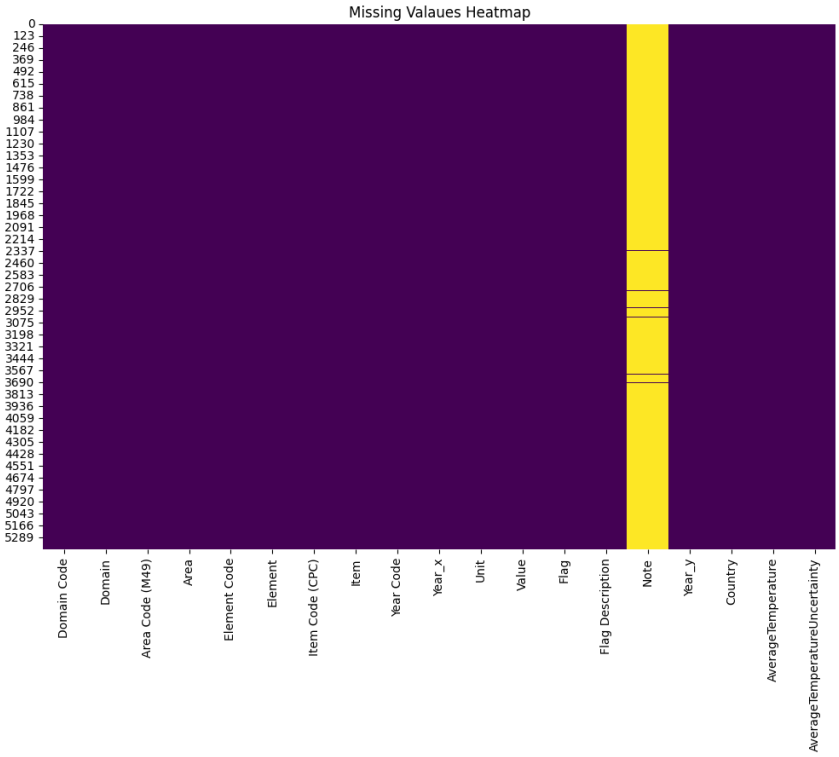

Figure 1 Missing values HeatMap

2 Analysis

2.1 Methodology

Goal was to investigate temperature impact on how agricultural variables (e.g. production, area harvested, yield). The key steps included:

2.1.1 Data Cleaning and Aggregation:

Monthly temperature data was then aggregated by **year**. Then calculated average temperature and average temperature uncertainty for each country and year. This was merged with the annual agricultural production data.

2.1.2 Exploratory Data Analysis (EDA)

The first analysis I did was the correlation between average temperature and agricultural metrics across all countries.

- **Univariate Analysis**: Univariate allows for the interpretation of the distribution and statistics of one variable. Conducted univariate analysis on the primary variables under consideration, Average Temperature and Agricultural Production (or Yield and Area Harvested).
- **Bivariate Analysis**: Bivariate analysis is the analysis of the relationship between twovariables. Demonstrated correlation between temperatureand agricultural indices such as production, area harvested and yield.
- **Multivariate Analysis**: Multivariate analysis is a method used to understand the relationships between multiple variables at once. Example: Correlation

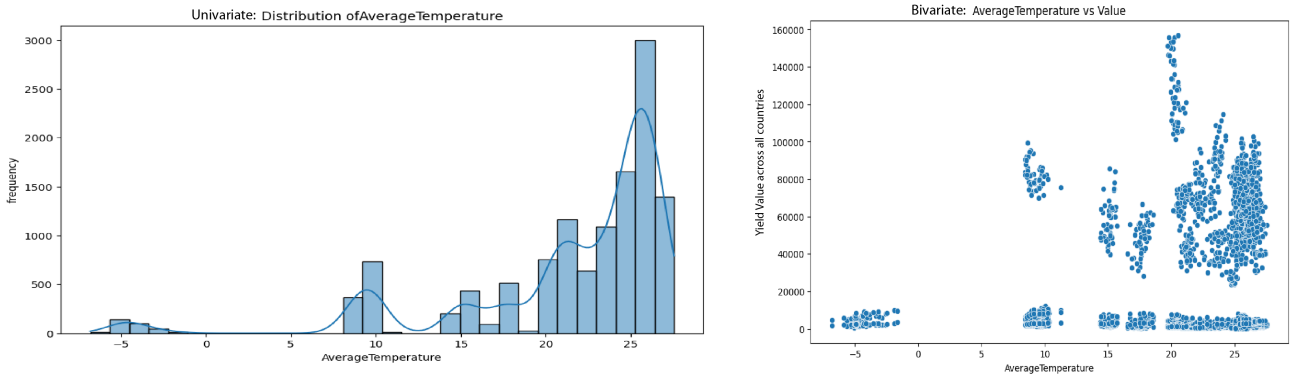

Figure 2 Uni and Bi-Variate Analysis

3 Correlation Analysis:

Calculated the correlation between temperature and the agricultural metrics (Production, Area Harvested, Yield) for each country and dataset as a whole.

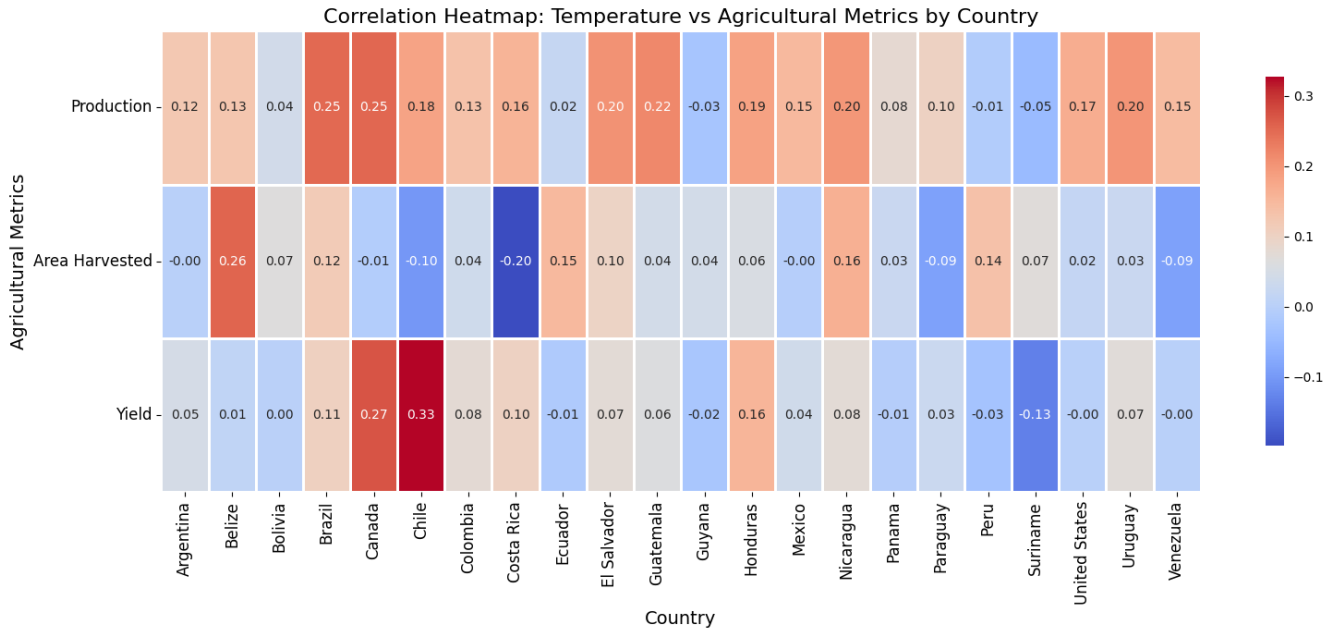

Figure 3 Correlation Heatmap

3.1 Results and Interpretation

- **Production:** For several countries, there exists a positive correlation of temperature with production, where :Argentina (0.1229) and :Brazil (0.2525) showing weak to moderate positive relationships.
- **Area Harvested:** The correlations are, overall, weak suggesting there is not much effect of temperature on area harvested, as in Argentina (-0.0007) or Belize (0.2554).
- **Yield:** Climate Change & Yield, the results are mixed,with Chile (0.3269) has a positive result while Suriname (-0.13326) has a negative impact.

3.2 Conclusion

Temperature's influence on agricultural production is somewhat mixed, with weak correlation on Area Harvested and Heterogeneous effects on Yield and Production, indicating that not everything is directly related to temperature.other factors for example, the regional climate conditions, the sort of crops grown, and farming practices are more likely responsible for agricultural production changes.