Temperature vs Crop Yield

*How temperature fluctuations affect the crop yield

Stamatios Karvounis
School of Computing
National College of Ireland
Dublin, Ireland
x18197051@student.ncirl.ie

***Abstract*— This project aims to investigate the impact of yearly fluctuations in temperature and precipitation on agricultural production in California using two large datasets collected over the past two decades. The datasets comprise daily summaries of temperature and precipitation and information on crop yield and animal production. The project employs several programming languages, including Python, and database systems like MongoDB and PostgreSQL, and analyses the data using appropriate programming environments like Ubuntu Terminal and Anaconda Jupyter Notebook. The study provides an understanding of the connection between climate and agriculture in California and demonstrates the use of diverse programming languages and databases for large-scale data analysis. The results of the analyses are presented using Seaborn and Matplotlib Python libraries, providing insights on the correlations between temperature, precipitation, and crop/animal production.**

Keywords—temperature, crop yield, database, MongoDB, PostgreSQL

I. Introduction

The agricultural sector plays a significant role in the economy of the United States, and the impacts of climate change on crop yields and profitability have become a growing concern. [1] As global temperatures continue to rise, it has become crucial to investigate the relationship between temperature rise and crop yields to assist policymakers in making informed decisions for sustainable farming practices.

This project's primary objective is to analyze the impact of temperature fluctuations on crop production volume in California over the last 20 years. California was selected as the main area of analysis, as according to Bauer (2022) [2] California is considered the largest food producer in United States of America(USA), and plays a crucial role in meeting the food demands not only of the United States but also of the world's population.

California plays a significant role in global agriculture, producing a wide range of crops, including high-value crops like tree nuts, stone fruit, and citrus. The state's unique climate, characterized by warm and dry summers and mild, wet winters, is well-suited for agriculture. California's permanent crops account for over 28% of the state's agricultural sector, with annual sales of more than $14 billion. Additionally, California's agricultural industry generates over $100 billion in economic activity annually and exported over $20 billion worth of food, feed, and fiber in 2020. [2]

The project will investigate correlations between average annual temperatures, minimum and maximum temperatures and the quantities of major crops cultivated in the country. It aims to explore if there has been a significant change in crop yield due to temperature rise as well as precipitation. The objective of this proposed study is to examine the historical correlation between climate and crop production and to assess the overall influence of climate on crop yield. This study aims to provide a quantitative analysis of the relationship between climate and crop production over a specific period.

The proposed analysis aims to answer the following research question: *How do temperature fluctuations affect the yield of major crops in California.*

II. Related Work

A relevant study to this project examined how weather affects agricultural production in California by analyzing the relationship between crop yields and monthly temperature and precipitation for 12 major crops. [3] The study aimed to understand how climate impacts crop yield trends and found that this impact varies depending on the crop. The study suggests that California's diverse agriculture may help reduce the effects of climate change on the agricultural sector, but it could still have a significant impact on the economy.

The study that examined the correlation between climate and crop production in California was limited by its outdated timeframe from 1980-2003, which might not fully reflect current irrigation practices and climate changes. Moreover, according to Johnson & Cody (2015) [4] the 2014 drought in California had a notable influence on crop yield, indicating a chance for further analysis using more up-to-date data suggesting that other factors like irrigation systems could be significant for research in combination to the climate changes.

III. Methodology

For the purpose of this research project, two different datasets were used - a semi-structured dataset in JSON format and a structured dataset in CSV format. The datasets were chosen based on the fact that they had a common element that could be used for analysis, allowing to gain valuable insights about the relationship of temperature and crop yield.

The semi-structured dataset was obtained through the National Centers for Environmental Information (NCEI) API, which serves as a comprehensive platform for the dissemination of information related to climate, weather, and oceans across the United States of America. The dataset was stored in the appropriate database, MongoDB, which is designed to handle semi-structured data.

The second source of data is a CSV file was obtained from the Food and Agriculture Organization of the United Nations (FAO), a UN agency that spearheads global initiatives to combat hunger. The FAO is a reliable source of information and data on food and agriculture, which includes detailed statistics on livestock production and crop yield. This file has been stored in a PostgreSQL database. More information will be provided in the next part.

A. Daily Summaries Data

The semi-structured dataset was obtained from the website https://www.noaa.gov/, which serves as a one-stop platform for accessing information on weather, climate, and oceans in the United States of America. To access the data required for this research, the website's application programming interface (API) was used. Creating an account and obtaining an API key were mandatory steps to be able to use the API. These steps facilitated access to the dataset, which contained valuable information on climate and weather conditions.

To obtain weather data for California, an application programming interface (API) provided by the National Centers for Environmental Information (NCEI) was utilized. The API was accessed by setting the API endpoint URL to 'https://www.ncei.noaa.gov/cdo-web/api/v2/datasets' and sending a GET request with the necessary parameters and headers to the API using the Python requests module. As California is considered the largest crop producer in the United States, a California weather station was specified as a parameter to ensure that the obtained data was relevant to the research. The obtained data was then stored in a MongoDB database after setting up a connection to the database and inserting the data. The data acquisition and database storage were performed using Jupyter notebook and terminal.

At this point, I should mention that the implementation of the MongoDB server posed certain challenges in this project. Initially, the server was set up on a Windows operating system, but later it was switched to Linux Ubuntu for better utilization of the terminal. However, this transition resulted in some obstacles and time constraints. I discovered that older versions of MongoDB had already been installed, which caused issues when installing the new versions. For instance, for MongoDB, it was necessary to delete older configuration files to ensure the proper functioning of the new ones. This process consumed a significant amount of time before the errors were identified. It was also noted that using a virtual environment or Docker would have been more efficient, but I had issues using the sudo user privileges.

In order to facilitate the process of visualizing and working with the database (ETL), the MongoDB Compass tool was employed. This tool provided a user-friendly interface that enabled easy interaction with the database. The database was connected to the data obtained from the API, which allowed for seamless integration of the data into the database for further analysis.

In order to transform and prepare the data from the API for analysis, ETL process was conducted using MongoDB Compass. The process involved the addition of a new pipeline with four stages. The first stage was the 'Match' stage, which was used to remove null values from the data. The second stage was the 'Project' stage, where variables deemed insignificant for the analysis were excluded. These variables included STATION, NAME, ELEVATION, PRCP_ATTRIBUTES, TAVG_ATTRIBUTES, TMAX_ATTRIBUTES, and TMIN_ATTRIBUTES. The third 'Project' stage was used to convert variables from string to their respective values, with PRCP, TAVG, TMIN, and TMAX being converted to Double to enable calculations for a new field. The DATE variable was also converted into the Date type. The final 'Project' stage was used to add a new variable, 'temperatureDifference,' by subtracting TMIN from TMAX.

The resulting dataset for analysis contained the following 6 variables namely DATE, representing the date of the daily summary, PRCP for precipitation, TAVG for the average temperature of the day, TMAX for the maximum temperature of the day, TMIN for the minimum temperature of the day, and temperatureDifference, which represented the difference between the minimum and maximum temperatures for the day.

It is important to note here, that the use of the pipeline for the ETL process allows for the flexibility of choosing different variables for analysis without affecting the original database. For the purposes of this research, the variables selected through the pipeline included DATE, PRCP, TAVG, TMAX, TMIN, and temperatureDifference. These variables were deemed relevant and significant for the analysis at hand. After the ETL process was completed, the resulting dataset was exported to a CSV file, which was then stored on my local hard drive. Later, this CSV file was imported as a new table into the postgreSQL database to enable further analysis.

B. Crop Yield Data

The FAOSTAT dataset offered by the FAO - Food and Agriculture Organization of the United Nations (FAO) is a specialized agency that aims to defeat hunger by providing comprehensive information and data on food and agriculture - is a reliable and high-quality source of information on crop yield and livestock, with a long time series dating back several decades. It is a valuable resource for trend analysis and identifying long-term patterns.

The dataset utilized in this research was exported from the FAOSTAT database, which is available at https://www.fao.org/faostat/en/#data/QCL. The database allows users to select variables according to their specific research needs. Since this analysis focuses on the impact of temperature fluctuations on crop production, the chosen variables were:

- Countries: The United States of America (given that the daily summaries database only includes observations from a weather station in California)
- Elements: Yield
- Items: Crops (Primary) and Livestock (Primary)
- Timeframe: The years 2003 to 2022, which is consistent with the timeframe covered by the daily summaries dataset.

The chosen variables are expected to provide relevant insights into the impact of temperature on crop production in the United States of America. Following the extraction of data from the FAOSTAT dataset, the resulting file was saved in a local hard drive as a csv file. The csv file was subsequently imported into the PostgresQL database for further analysis.

Postgresql was chosen as the database management system to store and analyze the production dataset. The ETL process was carried out using PG Admin 4, a graphical user interface that provided an efficient means of extracting, transforming, and loading the data into the database. This combination proved to be a powerful tool that enabled the entire data management process, from data acquisition to data analysis.

The process of setting up the Postgresql database involved several steps. Firstly, a server was created. Next, a user was created with the appropriate permissions to manage the database. A database was then created to store the data, followed by the creation of a schema to organize the tables. Two tables were created, one for crop production and another for daily summaries. The faostat csv file was used to populate the crop production table, while data from the daily summaries dataset was copied into the daily summaries table. The use of Postgresql and PG Admin 4 allowed for efficient data management and streamlined the process of ETL.

In order to prepare the crop yield dataset for analysis, the following ETL steps were performed. Firstly, distinct values were selected from the dataset to remove duplicate rows. Following this, rows containing NULL values were removed to ensure data integrity. Additionally, columns that were not significant in the analysis domain, including domain_code, domain, area_code, area, year_name, flag_name, and flag_description, were dropped from the dataset. These variables either contained values that were codes or represented the same value in different forms. The data was then grouped by year and item name to facilitate further analysis. Finally, the required table was exported in CSV format for use in further analysis. The resulting dataset contained the following variables: year_code, element_name, item_name, and total_value. Year_code represented the year of the data, which would be used as the common variable to join the FAOSTAT dataset with the daily summaries dataset. Element_name represented the type of data, either yield or carcass weight. Item_name represented the different crops or animals, and total_value represented the corresponding value of yield or carcass weight for each year and item.

C. Merged Data

After successfully extracting and transforming the data from the two sources and loading them into the PostgreSQL database, the next step in my analysis was to join the two datasets together to perform a more comprehensive analysis. I decided to join the two datasets on the year_code variable since this would provide a common point of reference for the two datasets.

To accomplish this task, I needed to utilize the Jupyter Notebook. I used the Anaconda environment to run the Jupyter Notebook and used the Pandas library to import the two CSV files containing the data tables. After importing the CSV files, I merged the two tables using the merge function provided by Pandas, which allowed me to join the two tables based on the year_code variable. This resulted in a new dataset that combined the crop yield and climate data into a single dataset that could be used for further analysis.

The merged dataset was then used for data visualization and to identify insights related to crop yield and the impact of climate conditions on the yield. This process allowed me to gain a deeper understanding of the trends and patterns related to crop yield and the impact of climate conditions on the crop production over time.

IV. RESULTS AND VISUALIZATIONS

After merging the two datasets, I proceeded with visualizing the data to gain insights on the relationship between temperature, crop yield, and climate conditions. Using Python libraries such as seaborn and matplotlib in the Jupyter notebook from the Anaconda environment, I created the following visualizations.

Firstly, I generated a scatterplot to show the relationship between the total value of the crop yield and the average precipitation for each year during the 20-year period under analysis. This visualization was important to identify any patterns and trends in the data “Fig 1”.

Average Precipitation and Total Crop Yield

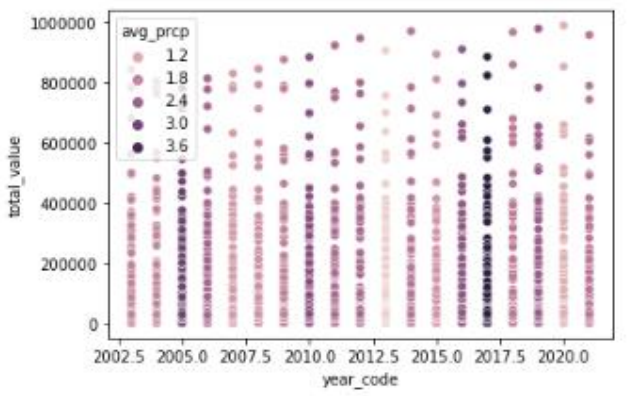


Fig 1. Average Precipitation and Total Crop Yield

Notably, it appears that the average precipitation was lower between 2012 and 2015, with the points in the graph showing a lighter color during this period. This finding aligns with the results of Johnson & Cody (2015) [4] which reported a severe drought in California during this timeframe. This visual analysis confirms the relationship between precipitation and crop yield, highlighting the impact of weather conditions on agriculture.

Next, I created a histogram to display the distribution of the total value of each item over the 20-year timeframe. This allowed me to analyze the frequency of occurrence of different crop yields “Fig 2”.

Crop Yield by Year

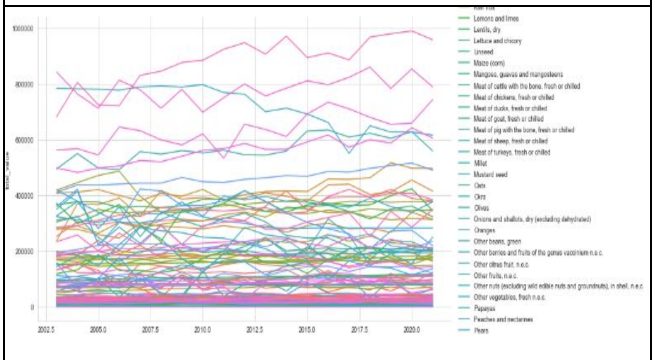


Fig 2. Fluctuation of the crop yield year by year

To further explore the performance of the two different elements (crop yield and carcass weight) over the years, I generated a line plot “Fig 3”.

Crop Yield vs Carcass Weight

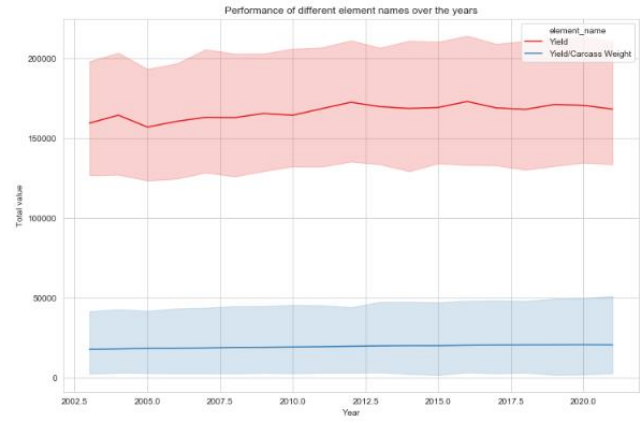


Fig 3. Crop Yield and Carcass weight comparison

Upon analyzing the line plot, it can be observed that the crop yield variable shows a slight increase year by year, while the carcass weight variable appears to be relatively stable throughout the 20-year time frame.

Another line plot was created to display the fluctuation of the temperature and precipitation during the 20-year period. This enabled me to compare the trend with the previous line plot with the elements “Fig 4”.

Temperature and Precipitation

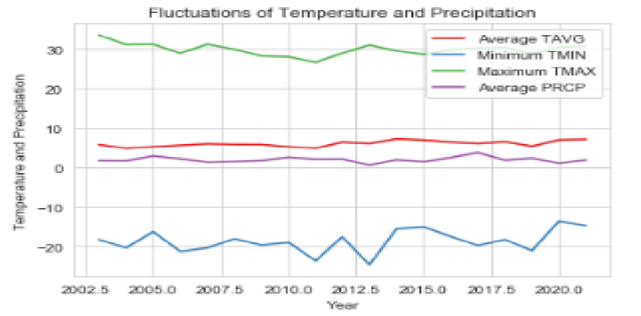


Fig 4. Fluctuation of temperatures(min, max, average) and precipitation

Upon visualizing the fluctuation of temperature and precipitation during the 20-year timeframe, it can be observed that the overall line for the average temperature and precipitation remains relatively stable. However, a closer inspection reveals fluctuations in the minimum and maximum temperatures. The line chart shows some dents in certain years, which may indicate extreme weather events such as heatwaves or cold snaps. These fluctuations in temperature may have a significant impact on crop yields, especially if they occur during critical periods in the crop's growth cycle. Therefore, it is important to consider not only the average temperature and precipitation but also the variability and extremes when studying the relationship between climate and crop yield.

Furthermore, I created a combined line plot with the fluctuation of the total crop yield in comparison to the temperature and precipitation for the 20-year timeframe. This visualization was crucial to see how the weather conditions affect crop yield “Fig 5”.

Weather summaries and Crop Yield

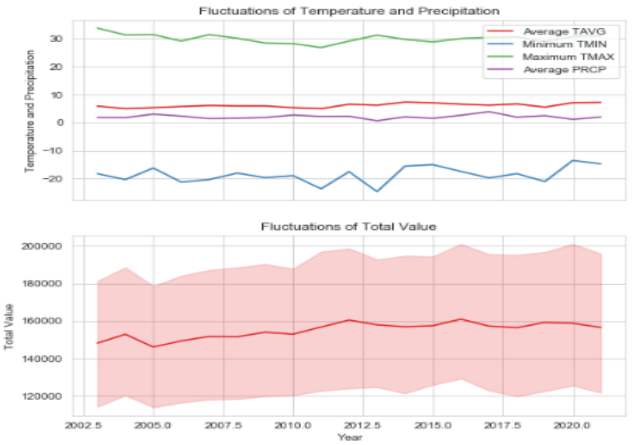


Fig 5. Fluctuation of temperatures(min, max, average) and precipitation vs total crop yield

In this figure, I observed that there is a positive relationship between the total value produced and the minimum and maximum temperatures. Specifically, when the minimum temperature is at its highest (e.g. in 2012), the total value produced appears to be high as well. Similarly, when the maximum temperature is at its highest (e.g. in 2013), the total value produced also appears to be high. This suggests that temperature, particularly extreme temperature, may have a significant impact on crop yield and production. It is worth noting, however, that this relationship might depend on other factors like precipitation or introduction to new irrigation systems due to the drought period that California experienced.

Lastly, I generated a correlation matrix to display the correlation between the variables in the merged dataset. This visualization allowed me to identify any patterns or relationships between the variables and to perform statistical analyses “Fig 6”.

Correlation Matrix

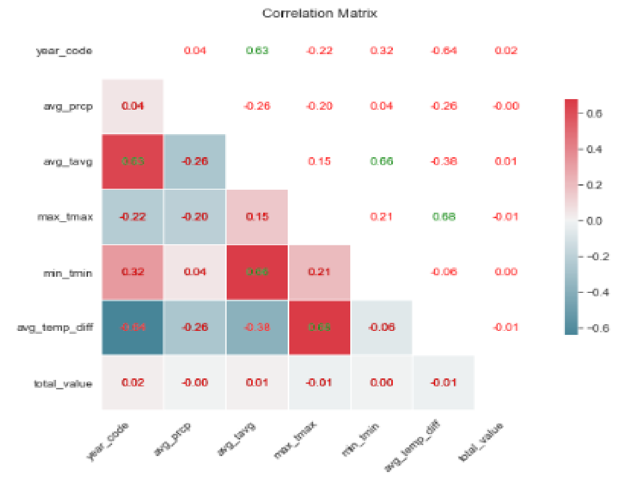


Fig 6. Correlation Matrix for the data frame variables

The correlation matrix shows the pairwise correlations between the variables avg_prcp, avg_tavg, max_tmax, min_tmin, avg_temp_diff, and total_value.

The correlation matrix displays values that range between -1 to 1, where values closer to 1 signify a strong positive correlation, values closer to -1 indicate a strong negative correlation, and values closer to 0 signify little to no correlation between the variables. Upon examining this matrix, I observed a weak negative correlation between avg_prcp and total_value, indicating that as average precipitation rises, there is a slight decrease in the total value of production. In addition, a weak positive correlation was found between avg_tavg and total_value, signifying that as the average temperature rises, there is a slight increase in the total value of production.

V. Conclusion and Future Work

In conclusion, my analysis revealed a weak negative correlation between average precipitation and total value, suggesting a potential threshold level for precipitation beyond which crop production is not significantly benefited. In addition, a weak positive correlation between average temperature and total value was observed, indicating a slight increase in production with rising temperatures, although this relationship is not linear due to fluctuations in minimum and maximum temperatures.

It is important to note that my analysis has limitations, and further research may be needed to fully understand the complexities of climate-crop production relationships. As the database is stored programmatically, it can be easily extended to include additional variables and information for further analysis. This allows for the incorporation of new data sources or the inclusion of additional metrics that may be relevant to the research question.

There are many ways to continue studying the results of this research. Firstly, we could look more closely at how different crops are affected by specific weather conditions, by collecting more detailed data and analyzing the correlations. Secondly, we could study how changes in temperature and precipitation affect the prices of crops, by collecting market price data and analyzing it. In addition, we could use more weather stations in our data collection process, to get more accurate and consistent results across different areas. Lastly, we could analyze the data by season instead of by year to better understand how weather affects crop growth and development.

Overall, there are many ways to continue studying the relationship between weather and agriculture and to expand our knowledge in this area.

VI. References

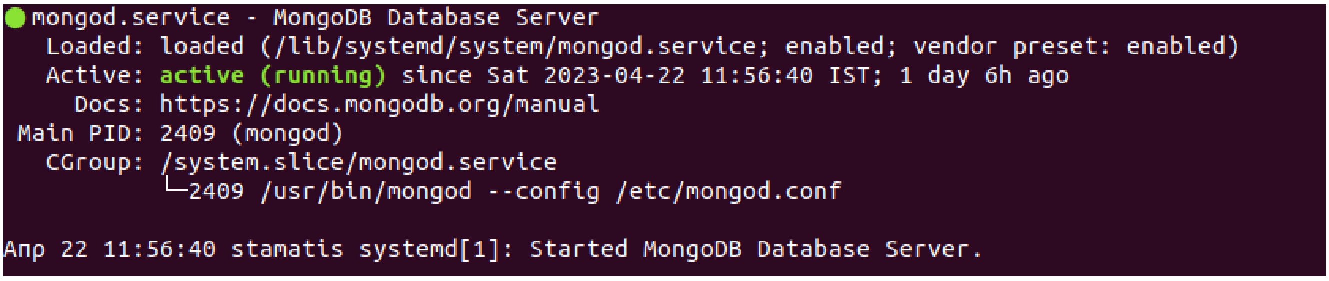
[1] J. Medellín-Azuara and R. E. Josué, "Economic impacts of climate-related changes," Climatic Change, p. S387–S405, 2011.

[2] R. Bauer, "Farm Together," 27 07 2022. [Online]. Available: https://farmtogether.com/. [Accessed 2023].

[3] D. B. Lobell, K. N. Cahill and C. B. Field, "Historical effects of temperature and precipitation on California crop yields," Climatic Change, p. 187–203, 2007.

[4] R. Johnson and B. A. Cody, "California Agricultural Production and Irrigated Water Use," Congressional Research Service, 2015.

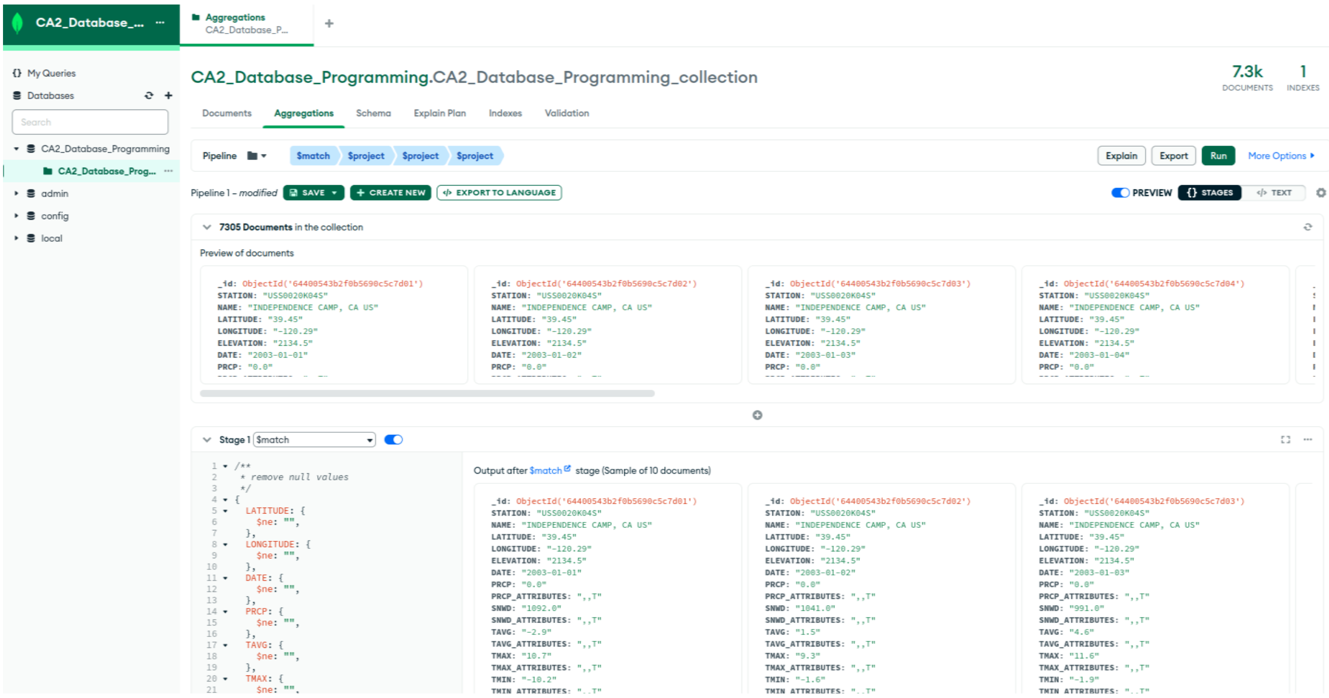
VII. APPENDIX A



```
● mongod.service - MongoDB Database Server
     Loaded: loaded (/lib/systemd/system/mongod.service; enabled; vendor preset: enabled)
     Active: active (running) since Sat 2023-04-22 11:56:40 IST; 1 day 6h ago
       Docs: https://docs.mongodb.org/manual
   Main PID: 2409 (mongod)
     CGroup: /system.slice/mongod.service
             └─2409 /usr/bin/mongod --config /etc/mongod.conf

Апр 22 11:56:40 stamatis systemd[1]: Started MongoDB Database Server.
```

Figure 7. MongoDB Server working after 2 days of struggling with configuration files (happiest day of the project)



CA2_Database_Programming.CA2_Database_Programming_collection

7.3k DOCUMENTS 1 INDEXES

Pipeline 1 – *modified*

7305 Documents in the collection

Preview of documents

```
_id: ObjectId('64400543b2f0b5690c5c7d01')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-01"
PRCP: "0.0"
```

```
_id: ObjectId('64400543b2f0b5690c5c7d02')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-02"
PRCP: "0.0"
```

```
_id: ObjectId('64400543b2f0b5690c5c7d03')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-03"
PRCP: "0.0"
```

```
_id: ObjectId('64400543b2f0b5690c5c7d04')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-04"
PRCP: "0.0"
```

Stage 1 $match

```
/**
 * remove null values
 */
{
  LATITUDE: {
    $ne: "",
  },
  LONGITUDE: {
    $ne: "",
  },
  DATE: {
    $ne: "",
  },
  PRCP: {
    $ne: "",
  },
  TAVG: {
    $ne: "",
  },
  TMAX: {
    $ne: "",
```

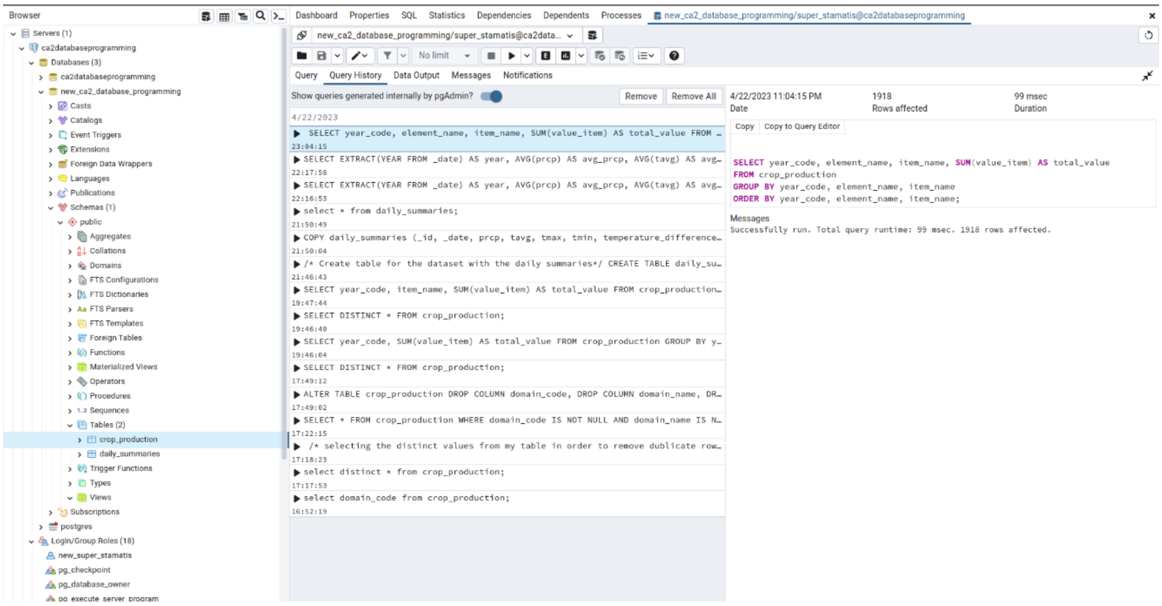
Output after $match stage (Sample of 10 documents)

```
_id: ObjectId('64400543b2f0b5690c5c7d01')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-01"
PRCP: "0.0"
PRCP_ATTRIBUTES: ",,T"
SNWD: "1092.0"
SNWD_ATTRIBUTES: ",,T"
TAVG: "-2.9"
TAVG_ATTRIBUTES: ",,T"
TMAX: "10.7"
TMAX_ATTRIBUTES: ",,T"
TMIN: "-10.2"
TMIN_ATTRIBUTES: ",,T"
```

```
_id: ObjectId('64400543b2f0b5690c5c7d02')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-02"
PRCP: "0.0"
PRCP_ATTRIBUTES: ",,T"
SNWD: "1041.0"
SNWD_ATTRIBUTES: ",,T"
TAVG: "1.5"
TAVG_ATTRIBUTES: ",,T"
TMAX: "9.3"
TMAX_ATTRIBUTES: ",,T"
TMIN: "-1.6"
TMIN_ATTRIBUTES: ",,T"
```

```
_id: ObjectId('64400543b2f0b5690c5c7d03')
STATION: "USS0020K04S"
NAME: "INDEPENDENCE CAMP, CA US"
LATITUDE: "39.45"
LONGITUDE: "-120.29"
ELEVATION: "2134.5"
DATE: "2003-01-03"
PRCP: "0.0"
PRCP_ATTRIBUTES: ",,T"
SNWD: "991.0"
SNWD_ATTRIBUTES: ",,T"
TAVG: "4.6"
TAVG_ATTRIBUTES: ",,T"
TMAX: "11.6"
TMAX_ATTRIBUTES: ",,T"
TMIN: "-1.9"
TMIN_ATTRIBUTES: ",,T"
```

Figure 8. MongoDB Compass, Overview of the Database



Show queries generated internally by pgAdmin?

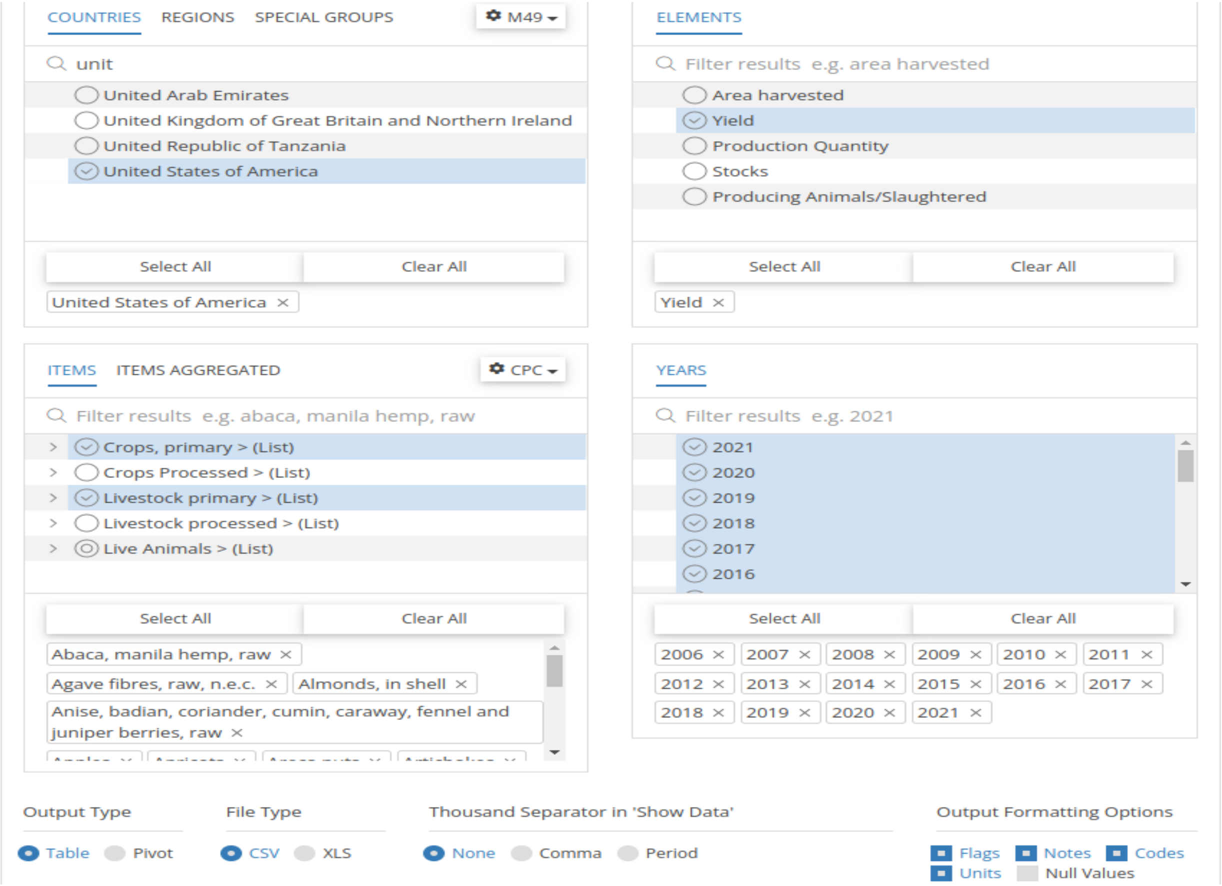
```
4/22/2023
▶ SELECT year_code, element_name, item_name, SUM(value_item) AS total_value FROM …
23:04:15
▶ SELECT EXTRACT(YEAR FROM _date) AS year, AVG(prcp) AS avg_prcp, AVG(tavg) AS avg…
22:17:58
▶ SELECT EXTRACT(YEAR FROM _date) AS year, AVG(prcp) AS avg_prcp, AVG(tavg) AS avg…
22:16:53
▶ select * from daily_summaries;
21:50:49
▶ COPY daily_summaries (_id, _date, prcp, tavg, tmax, tmin, temperature_difference…
21:50:04
▶ /* Create table for the dataset with the daily summaries*/ CREATE TABLE daily_su…
21:46:43
▶ SELECT year_code, item_name, SUM(value_item) AS total_value FROM crop_production…
19:47:44
▶ SELECT DISTINCT * FROM crop_production;
19:46:40
▶ SELECT year_code, SUM(value_item) AS total_value FROM crop_production GROUP BY y…
19:46:04
▶ SELECT DISTINCT * FROM crop_production;
17:49:12
▶ ALTER TABLE crop_production DROP COLUMN domain_code, DROP COLUMN domain_name, DR…
17:49:02
▶ SELECT * FROM crop_production WHERE domain_code IS NOT NULL AND domain_name IS N…
17:22:15
▶ /* selecting the distinct values from my table in order to remove dublicate row…
17:18:23
▶ select distinct * from crop_production;
17:17:53
▶ select domain_code from crop_production;
16:52:19
```

4/22/2023 11:04:15 PM Date | 1918 Rows affected | 99 msec Duration

```
SELECT year_code, element_name, item_name, SUM(value_item) AS total_value
FROM crop_production
GROUP BY year_code, element_name, item_name
ORDER BY year_code, element_name, item_name;
```

Messages
Successfully run. Total query runtime: 99 msec. 1918 rows affected.

Figure 9. PostgreSQL Server, Database, Tables and Query History



COUNTRIES REGIONS SPECIAL GROUPS ⚙ M49

unit

○ United Arab Emirates
○ United Kingdom of Great Britain and Northern Ireland
○ United Republic of Tanzania
✓ United States of America

Select All | Clear All

United States of America ×

ELEMENTS

Filter results e.g. area harvested

○ Area harvested
✓ Yield
○ Production Quantity
○ Stocks
○ Producing Animals/Slaughtered

Select All | Clear All

Yield ×

ITEMS ITEMS AGGREGATED ⚙ CPC

Filter results e.g. abaca, manila hemp, raw

› ✓ Crops, primary > (List)
› ○ Crops Processed > (List)
› ✓ Livestock primary > (List)
› ○ Livestock processed > (List)
› ○ Live Animals > (List)

Select All | Clear All

Abaca, manila hemp, raw ×
Agave fibres, raw, n.e.c. × | Almonds, in shell ×
Anise, badian, coriander, cumin, caraway, fennel and juniper berries, raw ×

YEARS

Filter results e.g. 2021

✓ 2021
✓ 2020
✓ 2019
✓ 2018
✓ 2017
✓ 2016

Select All | Clear All

2006 × | 2007 × | 2008 × | 2009 × | 2010 × | 2011 ×
2012 × | 2013 × | 2014 × | 2015 × | 2016 × | 2017 ×
2018 × | 2019 × | 2020 × | 2021 ×

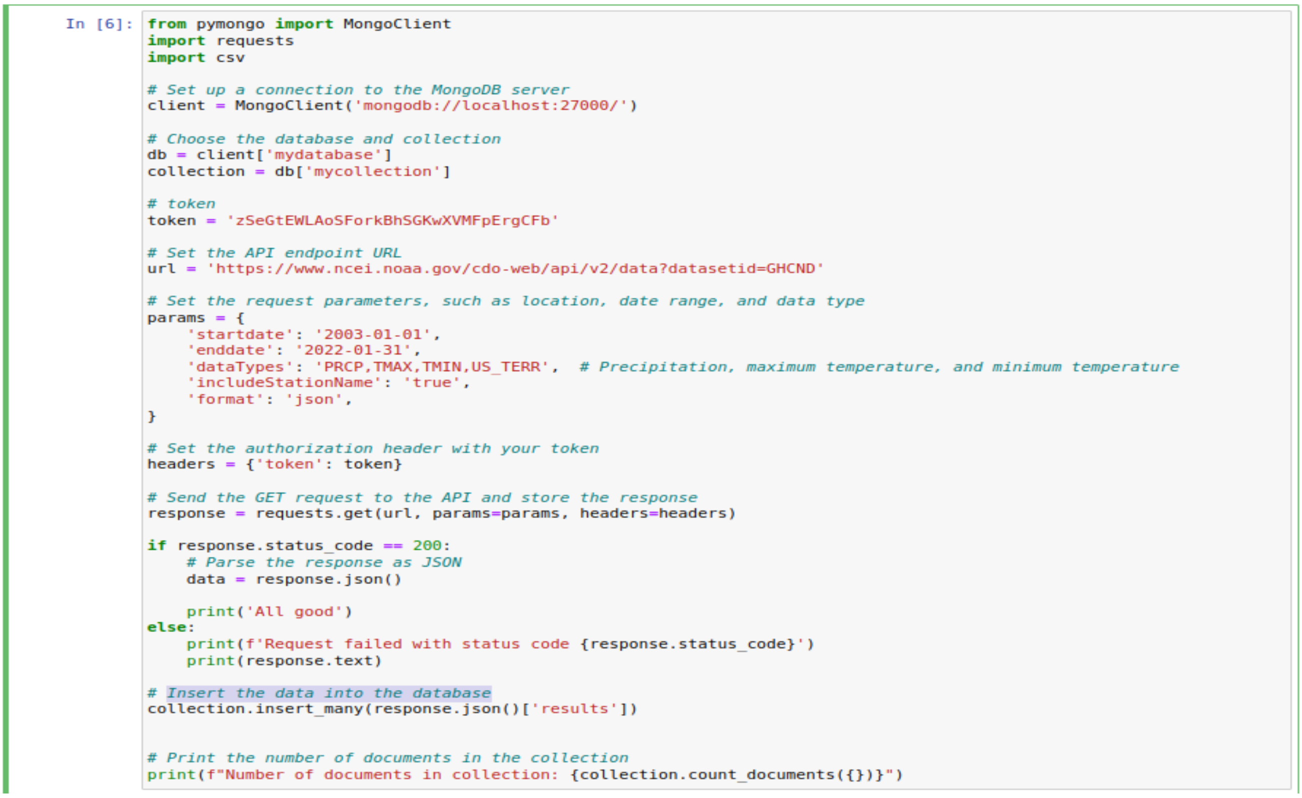
Output Type: ● Table ○ Pivot

File Type: ● CSV ○ XLS

Thousand Separator in 'Show Data': ● None ○ Comma ○ Period

Output Formatting Options: ☑ Flags ☑ Notes ☑ Codes ☑ Units □ Null Values

Figure 10. Variable selection for the Crop Yield dataset from FAOSTAT



```
In [6]: from pymongo import MongoClient
        import requests
        import csv

        # Set up a connection to the MongoDB server
        client = MongoClient('mongodb://localhost:27000/')

        # Choose the database and collection
        db = client['mydatabase']
        collection = db['mycollection']

        # token
        token = 'zSeGtEWLAoSForkBhSGKwXVMFpErgCFb'

        # Set the API endpoint URL
        url = 'https://www.ncei.noaa.gov/cdo-web/api/v2/data?datasetid=GHCND'

        # Set the request parameters, such as location, date range, and data type
        params = {
            'startdate': '2003-01-01',
            'enddate': '2022-01-31',
            'dataTypes': 'PRCP,TMAX,TMIN,US_TERR',  # Precipitation, maximum temperature, and minimum temperature
            'includeStationName': 'true',
            'format': 'json',
        }

        # Set the authorization header with your token
        headers = {'token': token}

        # Send the GET request to the API and store the response
        response = requests.get(url, params=params, headers=headers)

        if response.status_code == 200:
            # Parse the response as JSON
            data = response.json()

            print('All good')
        else:
            print(f'Request failed with status code {response.status_code}')
            print(response.text)

        # Insert the data into the database
        collection.insert_many(response.json()['results'])


        # Print the number of documents in the collection
        print(f"Number of documents in collection: {collection.count_documents({})}")
```

Figure 11. Code for getting the data from NCEI API and storing in MongoDB server.