“Identifying High Demand Crops for Data Visualization”

A Project Report submitted in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted by

Aditya Ganti(VU21CSEN0100186)
Ponnuri Pavan Kumar (VU21CSEN0100147)
M Praneeth Sai(VU21CSEN0100991)
Kotla Vivek (VU21CSEN0101052)
Siva Ganesh(VU21CSEN0101062)

Under the esteemed guidance of

Dr. T.M.N Vamsi
Associate Professor, Dept. of CSE, GST

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GITAM SCHOOL OF TECHNOLOGY
GITAM (Deemed to be University)
VISAKHAPATNAM

2025

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

GITAM (Deemed to be University)

VISAKHAPATNAM

DECLARATION

I hereby declare that the project report entitled "**Identifying High Demand Crops for Data Visualisation**" is an original work done in the Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University), submitted in partial fulfilment of the requirements for the award of the degree of B.Tech. in Computer Science. The work has not been submitted to any other college or University for the award of any degree or diploma.

Date:

Registration No(s)	Name(s)	Signature
VU21CSEN0100186	Aditya Ganti	[illegible]
VU21CSEN0100147	P. Pavan Kumar	Pavankumar
VU21CSEN0100991	M. Praneeth Sai	M-Praneeth Sai
VU21CSEN0101052	Vivek Kotla	[illegible]
VU21CSEN0101062	Siva Ganesh	G. Siva Ganesh

ACKNOWLEDGEMENT

We want to thank our project guide, Dr. T.M.N Vamsi, Associate Professor Dept. of CSE, for his stimulating guidance and profuse assistance. We are grateful for his constant, ever-ready support, advice, and valuable insights throughout the progress of this project. We consider it a great privilege to have been able to work under their guidance and encouragement.

We want to thank our Project Coordinators, Dr. Godi Apparao, Assistant Professor, Dept. of CSE and Dr. Shaik Khader Ahmed Manoj, Assistant Professor, Dept. of CSE, for helping to complete the project by taking frequent reviews and for their valuable suggestions throughout the progress of this work.

Aditya Ganti(VU21CSEN0100186)

P. Pavan Kumar (VU21CSEN0100147)

P. Praneeth Sai (VU21CSEN0100991)

Vivek Kotla(VU21CSEN0101052)

Siva Ganesh(VU21CSEN0101062)

TABLE OF CONTENTS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

GITAM (Deemed to be University)

q

CERTIFICATE

This is to certify that the project report entitled “XXX” is a bonafide record of work carried out by Student Name (Regd. No.), Student Name(Regd. No.), Student Name(Regd. No.), Student Name(Regd. No.) students submitted in partial fulfillment of requirement for the award of degree of Bachelors of Technology in Computer Science and Engineering / Computer Science and Engineering (AI&ML/DS/CS/IoT).

Date :

Project Guide   Head of the Department

.

1. ABSTRACT

India's agricultural sector faces serious challenges in meeting global production standards for high-demand crops. This project will identify 25 high-demand crops where India's production is lagging behind the international benchmark and provide data-driven insights to inform policy-making and agricultural strategies. We thus conducted a comprehensive analysis using historical and current crop production data that are readily available through sites such as UN Food and Agriculture data. The datasets range from 1961 to 2022, which will enable us to have a comparative study with India and the rest of the world's other large agricultural countries.

We then begin by carrying out an in-depth literature review that shows the existing gaps in agricultural data visualisation research. It was followed by data collection and pre- and post-cleaning to ensure accuracy and reliability. We conducted a gap analysis with advanced data analysis techniques in Python and SQL, benchmarking Indian agricultural output against global standards, taking into account the factors of climate, availability of resources, and potential yield. Key insights are visualised using PowerBI, creating interactive and user-friendly dashboards that allow stakeholders, including farmers, policymakers, and researchers, to make informed decisions based on the data.

This is concluded by developing a robust, comprehensive dashboard showing India's overall performance on the standard benchmarks for agricultural production. Such a tool would then present an avenue to display long-term trends, focus points, and prioritised areas that are considered to increase effective investments in agriculture. Conclusion The Targeted recommendations of this project focus on bridging the gaps in the area of crop production by emphasising higher yield levels so as to contribute to attaining India's aim to ensure it remains agriculturally self-sufficient and highly competitive on a global platform.

Keywords: Agriculture, Data Visualization, Crop Production, Comparative Analysis, PowerBI, Python, SQL, Benchmarking, Decision-Making, High-Demand Crops.

.

2. INTRODUCTION

Agriculture is one of the pillars of the Indian economy, and its contribution to employment and GDP is enormous. Despite being one of the largest agricultural producers in the world, India cannot attain competitive yields for most of the high-demand crops. These gaps in agricultural performance have a significant impact on food security, economic growth, and the livelihood of millions of farmers. Data-informed information is in more considerable demand as an input in agriculture policy and resource allocation optimisation for addressing these concerns. For this purpose, this paper, "Identification of High-Demanding Crops Using Data Visualization," seeks to bridge that gap by using data analytics and visualisation skills to portray areas where agricultural production falls woefully short of worldwide norms.

Some of the tools and systems available in connection with agriculture and crop management that tend to address specific concerns are: Crop Yield Optimization Systems use the method of smart farming with IoT sensors and machine learning to optimise crop yield more efficiently. However, such systems tend to be focused more on domestic improvement in most cases without the corresponding context of global benchmarking. DSSs in agricultural matters help the farmers as well as policy planners in comprehending the weather, soil, and health conditions of the crops. Although practical, the systems are developed for a real-time decision-making capability and have not incorporated the historical comparative analysis of others. Big Data Analytics in Agriculture is gaining acceptance in the sectors, and tools such as Tableau and Qlik view extensive data. They still lack the precision required to benchmark crop production across borders; an essential feature of India's ranking in that regard has to be assessed accurately. Geospatial Analysis Tools, such as GIS, provide deep spatial analysis for crop health monitoring but are not used for comparative study at a global level to identify crops in high demand.

The current systems indicate the possibility of technological advancements in agriculture but still leave a significant gap. Currently, there is no system focused on the identification of crops where Indian production is below the world standard. Most of these tools are developed and oriented towards domestic yield optimisation, soil management, or monitoring in real-time conditions but fail to compare India with major developing global agriculture producers in detail. The absence of such a dedicated system, therefore, prevents stakeholders from basing cropping priorities on what exists on the worldwide level when demand is compared versus how shortfalls are. Policymaking and strategic investment opportunities consequently become less potent because their effectiveness is reduced through suboptimal selection, even if the implementation eventually makes sense.

The title of this project, "Identifying High-Demand Crops for Data Visualization," is an intentional choice reflecting its core objectives. It is justified to emphasise "High-Demand Crops" because agricultural efforts must be prioritised towards crops that hold the most potential for economic and nutritional impact. This project will identify these key crops through data analysis and visualise the gaps between India's production and global standards. This is because it can transform raw data into intuitive insights, thus opening this information to a much greater audience, even those who do not necessarily have some technical background. The core of the project is going to be visualising gaps to provide a clear explanation of India's global performance in agriculture in an actionable way.

With data-driven decision-making on the throne of the world, the agricultural sector has to keep pace with advanced analytics and visualisation techniques to keep its share in the world market. This project addresses the critical gap by undertaking a visual analysis of India's agricultural sector performance, standards of the global world, and crops that significantly impact the economy and food security of the nation. With such high-demand crops and envisioning their position in the international domain, this project would guide stakeholders to understand priorities with strategic interventions in the Indian agricultural sector.

.

3. LITERATURE REVIEW

Table 3.1: Research Papers Referred to

S.No	Title	Authors	Year of Publication	Findings
1.	Agriculture Data Visualisation for Prescriptive Crop Planning	Harshita B P, Amith R, Abhishek S, Rohit Vibhu C	**2017**	This work focuses on why interactive and well-designed visualisations of data make agricultural data complex and, hence, informative. Precision Agriculture: It highlights the role modern technology combined with data analysis can play in improving crop planning and sustainable agricultural practices.
2.	Agriculture Data Visualisation and Analysis using Data Mining Techniques: Application of Unsupervised Machine Learning	Kunal Badapanda, Debani Prasad Mishra, Surender Reddy Salkuti	**2022**	The study demonstrates the use of unsupervised machine learning techniques like K-means clustering and PCA to analyse and optimise agricultural data for improved crop yield planning.
3.	Data Analytics for	Nabila Chergui &	**2022**	Digital Architecture leverages advanced Digital Devices and

	.Crop Management: A Big Data View	Mohand Tahar Kechadi		data analytics to enhance productivity, optimise resource use, and minimise the environmental impact of farming.
4.	Role of Data Visualisation and Big Data Analytics in Smart Agriculture	J Kavitha, Shabnam Kumari, K. Manivannam, Amit Kumar Tyagi	**2024**	The chapter discusses the importance of data visualisation and big data analytics in smart agriculture, highlighting their role in decision-making, resource optimisation, and sustainability.
5.	Intelligent Smart Farming and Crop Visualisation	Rajkumar Rajasekaran, Rajendra Agarwal, Aditya Srivastav, Jolly Masih, Volodymyr Ivanyshyn, Iryna Yasinetska	**2020**	The paper discusses intelligent smart farming and crop visualisation techniques to enhance agricultural productivity and sustainability.

3.1 Major Takeaways from Literature Review:

1. **Lack of Comparative Analysis in Existing Research**

   One of the key insights from the literature review was that no existing study has compared crop production in India directly to international benchmarks. Most studies were aimed at improving agricultural practice in India by either clever farming techniques for increased crop production or making the best use of available resources, with no analysis of whether India has a comparative advantage in terms of performance globally.

This thre. w to the forefront a major lacuna of the existing literature, which was on the importance of a project focusing on comparative analysis with regards to identifying where the Indian agriculture sector is lagging.

2. Domestic Yield Improvement Focus

A trend exists in the existing body of literature on the importance of techniques for enhancements of domestic crop yield, like clever cropping techniques, precision agriculture, and sustainable farming methods. A good number of them are promising studies. Still, these studies were mainly aimed at addressing the domestic challenges of India's agricultural sector, not considering the competitiveness of India's agricultural sector globally.

It identifies high-demand crops and visualises the data, which fills this gap in linking domestic agricultural practice to global standards that give a more holistic vision of India's agricultural potential.

3. **Emphasis on Data Visualization as a Key Tool**

Most of the reviewed papers highlighted the importance of data visualisation in agriculture to enable better decision-making. Visualisation tools are essential in helping stakeholders rapidly interpret complex agricultural data, thus facilitating better strategic planning.

This further reinforced the decision to use PowerBI and other visualisation tools in the project, affirming that visual representation of data is not only practical but also necessary for effective communication with farmers, policymakers, and researchers.

4. Big Data and Analytics Opportunities

This literature again emphasised the application of big data and analytics in optimising agricultural practice, most especially in managing resources for improved crop yields. Analysing historical data would bring about better resource allocation and crop management.

This result justified the project's methodology, as it makes use of Big Data analytics techniques with the help of tools such as Python and SQL to process massive agricultural datasets, thus arriving at more informative recommendations.

5. Impact of Climate Change on Crop Yields

A recurring theme in the literature was the impact of climate change on crop yields, and several authors noted the lack of consideration of the environment in generalisations of agricultural performance. Clearly, this means including all climate-related variables in any comparative analysis.

The liter.ature review suggested that the climate factor has to be incorporated in the project analysis, as knowing the said factors can better explain why crops are performing less well in India compared to international standards.

6. Higher Interest in Prescriptive Data Analysis

A common feature of all the papers reviewed was a shift from descriptive to prescriptive analytics in agriculture. It has been observed that researchers are no longer only describing crop patterns but suggesting actions for yield improvement based on their data-driven insights.

This aligns with the objectives of the project: identifying not only high-demand crops but also a source of actionable insights to be turned around by stakeholders into policy changes and strategic interventions.

7. Machine Learning and Advanced Analytics

A number of papers discussed how machine learning and sophisticated data analytics can be applied in simulating crop yields and improving production. In fact, these techniques are shown to enhance farm management decisions and resource use efficiency significantly. Although the current project does not adopt machine learning, the paper suggests that possible extensions to this project include the integration of predictive analytics that can augment the results from data visualisation and crop comparison.

4. PROBLEM IDENTIFICATION AND OBJECTIVES

4.1 Problem Identification

India's agricultural sector is strong, but in many crops of great demand, it faces quite a challenge in meeting the world's production standards. It is one of the greatest producers of agriculture, yet many of the most important crops are produced in much lesser quantities than at the international level. This situation not only threatens the potential of the country's economy but also jeopardises the food security and livelihood of millions of farmers. The existing systems and research on localised yield improvement lack a comprehensive analysis of India's performance vis-à-vis the global standards. The gaps that exist in this system must be identified through data-driven analysis, and actionable insights have to be delivered to the stakeholders so they can make informed decisions and decide the areas for improvement. Hence, it tries to bridge the already prevailing gap by identifying which of the demand crops is lagging and thus envisioning those areas as they require better focus.

4.2 Objectives:

To address the issue of low crop yields that are low compared to world benchmarks and to align agriculture in India with global output, the following goals are framed:

- **Objective 1: Identification of High-demand Crops** - The identification, through sound data sets in agriculture, of 25 high-demand crops in which the current output by India is lagging further behind the world benchmarks.

- **Objective 2: Comparative analysis-** Analyzing and comparing Indian agriculture with other advanced nations so as to be in a position to determine the fields of weakness in output and know why this lags.

- **Objective 3: Enhance Data Visualization -** Leverage features such as PowerBI to create an interactive visualisation that will make comparative analysis easier and thus help stakeholders in making well-informed decisions.

- **Obje.ctive 4: Contributing Factors Researching -** Explain underlying factors that may contribute to low agricultural productivity in India. This factor is related to climatic, resource constraint limits, and farming practices, which make sense of the observed gaps in India's agricultural productivity.

- **Objective 5: Informing Decision:** Provide actionable recommendations to policymakers, farmers, and researchers on how to shrink yield gaps to make agriculture more competitive for India.

- **Objective 6: Develop an engaging User Dashboard:** Develop a summarising dashboard to provide for easy interaction between users in digging up trends within the data and to ease the narrowing of search possibilities as well as reporting of options on specific crops as well as by country.

5. PROPOSED SYSTEM

5.1. Overview

This is a system that can determine the most in-demand crops using a data-driven comparison between the production of crops in India and international standards. Such a system involves the assimilation of data from several sources, data preprocessing to get accuracy, and finally, applying visualisationtools to draw inferences that are helpful in making decisions at policy levels as well as strategies in agriculture. The bottom line here is to use data analysis methods and visualisation tools to determine the Indian agricultural scene based on global standards.

5.2. Critical Features of the Developed System

Data Collection:

Crop production data is captured through sources using Kaggle and OurWorldInData since the former builds reliable sources of generating other regions and nations based on focused datasets.

Data collection ensures that a proper collection of data points is carried out based on a comparative analysis system so that data suitable for comparison can be obtained.

Data Preprocessing

The collecte.d data was cleaned and normalised.

Transformed accordingly into an acceptable format in order to carry out a comparison with another.

Preprocessing: Removal of inconsistency, which is usually accompanied by multiple sources of data, ensuring accuracy

This step in itself becomes crucial in having data that are coming from different regions and can easily be comparable.

Comparative Analysis:

The kernel functionality involves the comparison between crop production data in India and its global counterparts.

There is the analysis of holes in crop production where, for sure, India trails or tops other countries.

It analyses the production patterns of 25 high-demand crops and takes insights from data using SQL query and statistical methods.

Detection of High-Demand Crops:

By comparative analytics, the system identifies an individual crop that has a very high demand yet lower-than-expected production in India.

In this procedure, crop clusters are categorised by region and type and globally acknowledged by standards for focusing on definite areas.

Data Visualization :

Using PowerBI as a tool, the system visualised data in an interactive and understanding manner.

These are graphs, charts, and dashboards that would bring out trends, gaps, and performance in the production of crops at regional levels.

The intuitive graphics will aid stakeholders in making decisions.

Testing and Validation:

Through this. step, stakeholders check data accuracy and clarity in visualisation.

This step ensures the output is accurate enough, thus fulfilling the users' expectations.

Final Dashboard Presentation

These insights from the analysis will be presented in a fine-crafted dashboard.

The dashboard is user-friendly and gives a clear view of India's agricultural performance against global standards.

Conclusion and Recommendations:

The system puts together findings and policy recommendations for enhancing crop production.

It does this by bridging the identified gaps and on high-demand crops.

6. PROPOSED SYSTEM ARCHITECTURE

6.1: UML Use Case Diagram.

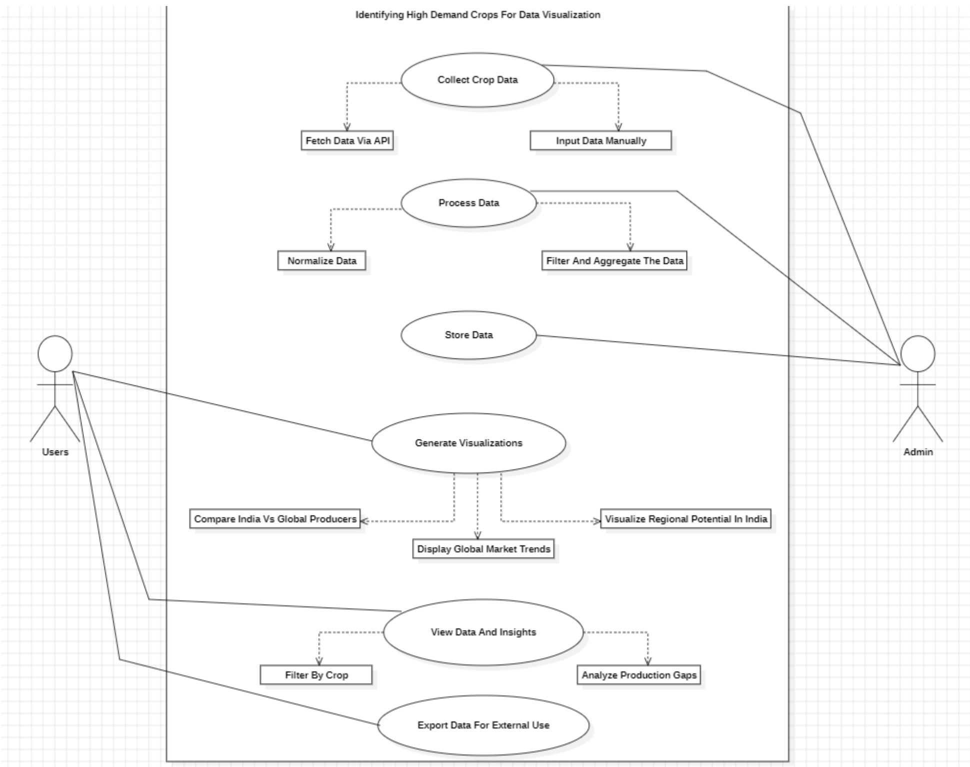


Fig 6.1: Use Case Diagram

Overview.

The diagram represents the flow of how crop data is collected, processed, stored, visualised, and utilised by Users and Admins to gain insights into high-demand crops in India.

Key Components & Workflow

1. **Collect Crop Data**
    - **Data is obtained via:**
        - **Fetch Data via API:** Automatically gathering crop-related information from external sources.
        - **Input Data Manually:** Admins manually enter data into the system.
2. **Process Data**
    - **The collected data undergoes preprocessing through:**
        - **Normalize Data:** Standardizing the dataset for consistency.
        - **Filter and Aggregate the Data:** Extracting relevant information and summarising it.
3. **Store Data**
    - Processed data is stored in a database for visualisation and further analysis.
4. **Generate Visualizations**
    - **The system creates insights based on the stored data:**
        - Compare India vs. Global Producers: Analyzing how Indian crop production compares to global standards.
        - Visualize Regional Potential in India: Identifying areas in India with high potential for specific crops.
        - Display Global Market Trends: Showing trends in global demand and production.
5. **View Data and Insights**

- **Users can explore and analyse the data through:**
    - **Filter by Crop:** Viewing data based on specific crops.
    - **Analyze Production Gaps**: Identifying shortages and opportunities in Indian agriculture.

6. **Export Data for External Use**
    - Users and admins can download or share insights for decision-making, investments, or further research.

Actors Involved

- **Users**
    - Access visualisation tools.
    - Filter data by crop.
    - Analyze production gaps.
- **Admins**
    - Collect and process data.
    - Manage database storage.
    - Oversee the visualisation and export functionalities.

Purpose of the Diagram:

This flowchart effectively illustrates the data-driven approach to understanding high-demand crops in India. It highlights:

- Data collection from APIs and manual input.
- Data processing through normalisation and filtering.
- Insights generation via comparisons, trend analysis, and visualisation.
- Actionable insights for farmers, policymakers, and investors to optimise agricultural production.

6.2: UML C.lass Diagram.

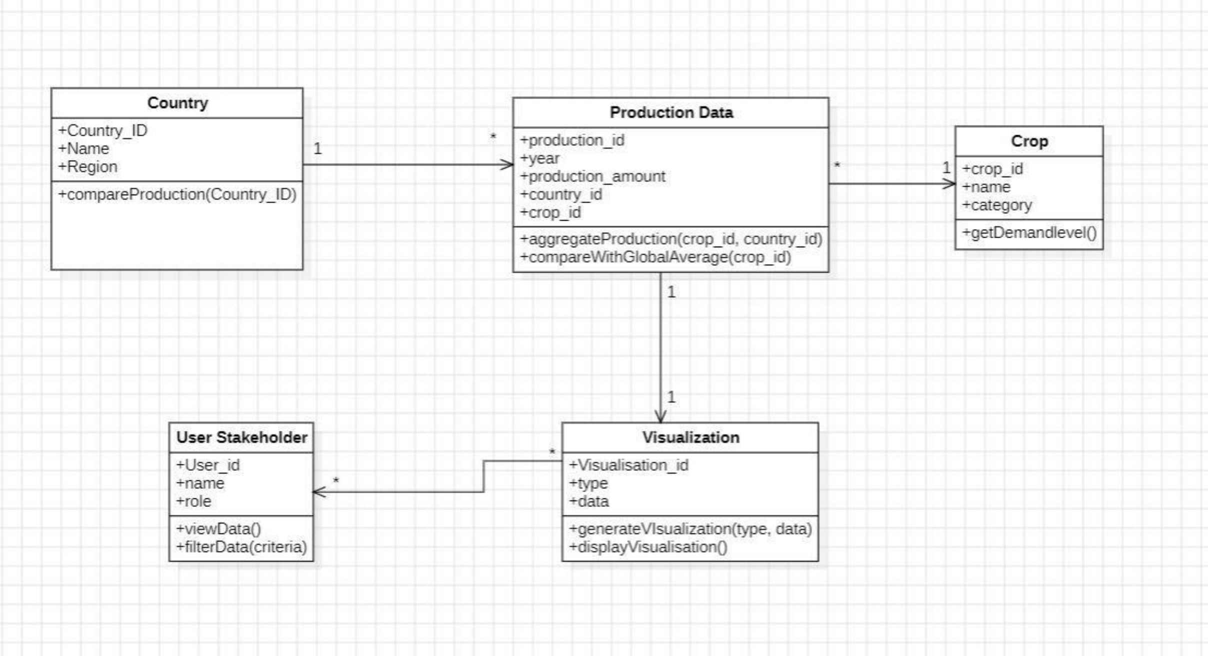


Fig 6.2: Class Diagram

Overview

This class diagram illustrates the relationships between different entities in the system, focusing on how crop production data is structured, analysed, and visualised for insights.

Key Components & Relationships

1. Country (Class)

- **Attributes:**
  - **Country_ID:** Unique identifier for each country.
  - **Name:** Name of the country.
  - **Region:** The geographical region of the country.
- **Method:**
  - **compareProduction(Country_ID):** Compares crop production between countries.

Relationship:

- One-to-Many (1 →*) relationship with Production Data, meaning one country can have multiple production records.

2. Crop (Class)

- **Attributes:**
    - **crop_id:** Unique identifier for each crop.
    - **name:** Name of the crop.
    - **category:** Type of crop (e.g., grains, fruits).
- **Method:**
    - **getDemandLevel():** Determines the demand level of the crop.

Relationship:

- One-to-Many (1 → *) relationship with Production Data, as a crop can be grown in multiple countries.

3. Production Data (Class)

- **Attributes:**
    - **production_id:** Unique identifier for each production record.
    - year: The year of production.
    - **production_amount:** Total amount of production.
    - **country_id:** Foreign key linking to the Country class.
    - **crop_id:** Foreign key linking to the Crop class.
- **Methods:**
    - **aggregateProduction(crop_id, country_id):** Aggregates production data for a specific crop and country.

- **compareWithGlobalAverage(crop_id):** Compares the production amount of a crop with the global average.

Relationship:

- Many-to-One (Many → 1) relationship with Country and Crop.

4. Visualization (Class)

- **Attributes:**
  - **visualisation_id:** Unique identifier for visualisation.
  - **type:** Type of visualisation (e.g., bar chart, pie chart).
  - **data: The dataset used for visualisation.**
- **Methods:**
  - **generateVisualization(type, data):** Generates a visualisation based on the provided type and data.
  - **displayVisualization():** Displays the created visualisation.

Relationship:

- One-to-Many (1 → *) relationship with Production Data, meaning a visualisation can be based on multiple production records.

5. User Stakeholder (Class)

- **Attributes:**
  - **User_id:** Unique identifier for each user.
  - **name:** Name of the user.
  - **role:** Role of the user (e.g., admin, researcher, policymaker).
- **Methods:**
  - **viewData():** Allows users to view production and visualisation data.
  - **filterData(criteria):** Enables users to filter data based on specific criteria.

Relationship:

- One-to-Many (1 → *) relationship with Visualization, allowing multiple users to interact with different visualisations.

Purpose of the Diagram

This class diagram provides a structured representation of how data flows in the system. It helps in:

- Defining the relationships between crop production, country, and visualisation.
- Establishing methods for data aggregation, comparison, and visualisation.
- Outlining the user interaction with the system.

6.3: UML Sequence Diagram.

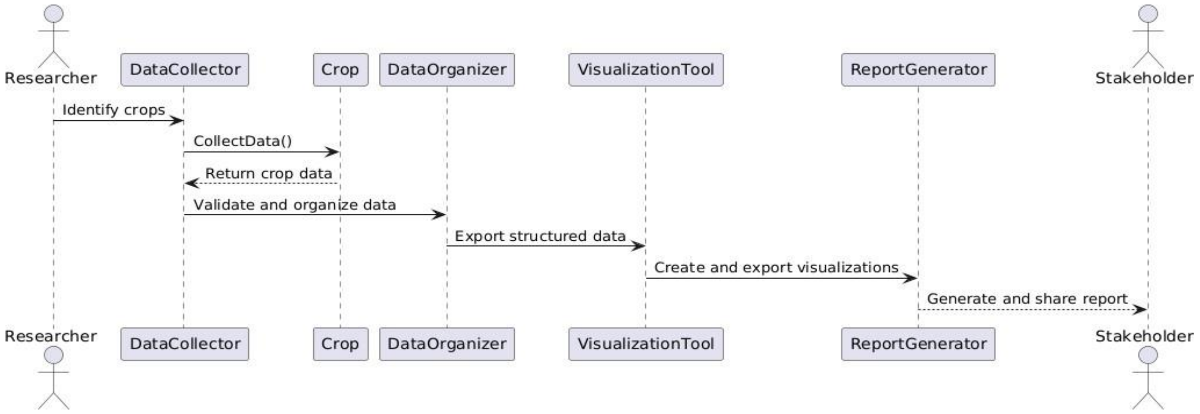


Fig 6.3: Sequence Diagram.

Overview

This sequence diagram illustrates the interactions between various system components and users involved in the process of identifying high-demand crops, organising data, visualising insights, and generating reports.

Key Components & Interactions

1. Research.er (Actor)

- The Researcher initiates the process by identifying crops for analysis.
- Sends a request to DataCollector to gather crop-related data.

2. DataCollector (Class)

- Receives the request from the researcher to collect crop data.
- Calls the Crop class to retrieve crop data.
- Returns the collected data to the DataCollector.
- Passes the data to DataOrganizer for validation and structuring.

3. Crop (Class)

- Stores and manages crop-related data.
- Provides crop data to DataCollector when requested.

4. DataOrganizer (Class)

- Validates and organises the collected crop data.
- Exports structured data to the VisualizationTool**.**

5. VisualizationTool (Class)

- Receives structured data from DataOrganizer.
- Creates and exports visualisations based on the structured data.
- Sends the visualisation output to the ReportGenerator.

6. ReportGenerator (Class)

- Receives visualisations from the VisualizationTool.
- Generates and shares reports with Stakeholders.

7. Stakeholder (Actor)

- Receives the final report from the system.

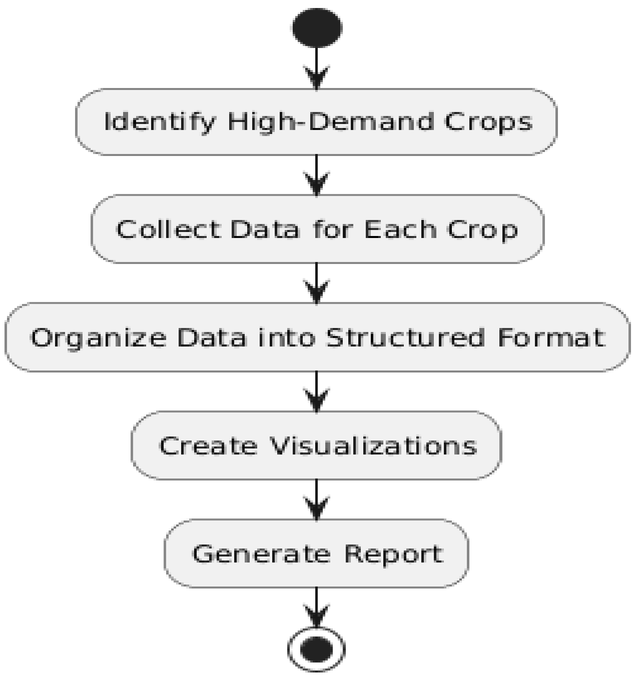
- Uses.the report for decision-making and analysis.

Purpose of the Diagram

This sequence diagram helps in the following:

- Understanding the flow of data from data collection to report generation.
- Defining the interactions between system components.
- Mapping user involvement in the system.
- Ensuring efficient data organisation and visualisation for stakeholders.

6.4: UML Activity Diagram.



6.4: Activity Diagram.

Overview

This activity diagram illustrates the step-by-step workflow for identifying high-demand crops, collecting data, organising information, visualising insights, and generating reports.

Key Steps & Workflow

1. Start (Initial Node)

- The process begins with the identification of high-demand crops.

2. Identify High-Demand Crops (Activity)

- The system or researcher identifies crops that have high global demand but insufficient production in India.

3. Collect Data for Each Crop (Activity)

- **Data is gathered for each identified crop, including:**
  - Production levels
  - Market demand
  - Global vs. local supply comparison

4. Organize Data into Structured Format (Activity)

- The collected data is validated and structured into a format suitable for analysis.
- This ensures consistency and facilitates data processing.

5. Create Visualizations (Activity)

- Based on structured data, visual representations such as bar charts, graphs, and maps are generated.
- These visualisations help identify trends and insights.

6. Generate Report (Activity)

- The system compiles the visualised insights into a report.
- The report provides decision-making support for stakeholders like policymakers, researchers, and farmers.

7. End (Final Node)

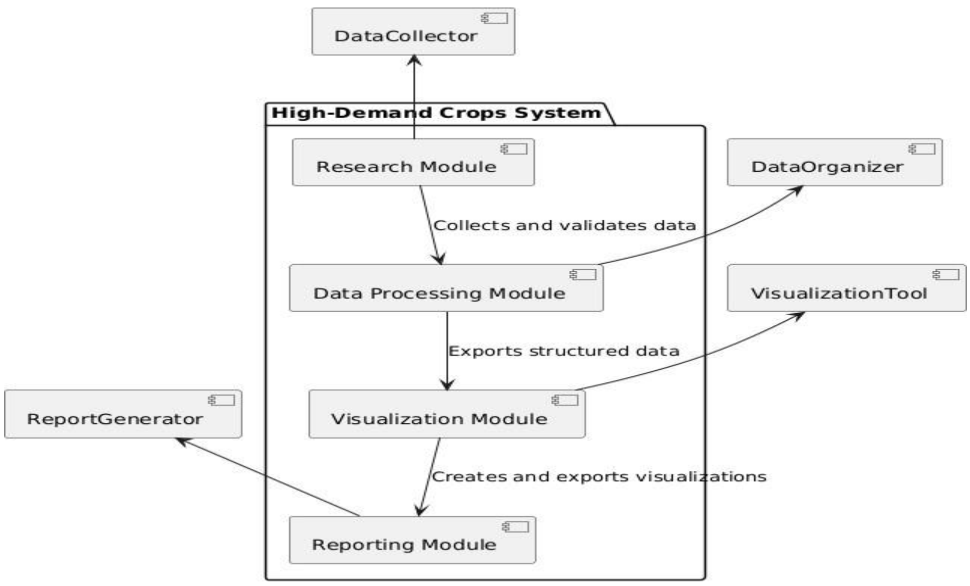
- The process concludes once the report is generated.

Purpose of t.he Diagram

This activity diagram helps in the following:

- Understanding the sequential workflow of data collection, processing, and reporting.
- Ensuring smooth transition between various steps.
- Identifying key stages where data is analysed and transformed into meaningful insights.
- Improving efficiency in handling and structuring agricultural data.

6.5 : UML Component Diagram.



6.5 : Component Diagram.

Overview

The diagram represents the architecture of the system by showing the modules responsible for data collection, processing, visualisation, and reporting.

Key Components & Interactions

1. High-Demand Crops System (Main System)

- This is the central system that contains several key modules responsible for different functionalities.

2. Research Module

- **Role:** Collects and validates raw data on high-demand crops.
- **Interaction:** Works with the DataCollector to retrieve relevant data.

3. Data Processing Module

- **Role:** Organizes and structures the collected data.
- **Interaction:**
    - Receives validated data from the Research Module.
    - Sends structured data to DataOrganizer for further structuring.

4. Visualization Module

- **Role:** Generates meaningful visualisations from structured data.
- **Interaction:**
    - Receives structured data from the Data Processing Module.
    - Uses VisualizationTool to create graphical representations.

5. Reporting Module

- **Role**: Compiles and formats the final report.
- **Interaction:**
    - Collect visualisations from the Visualization Module.
    - Works with ReportGenerator to generate a final report.

External Components

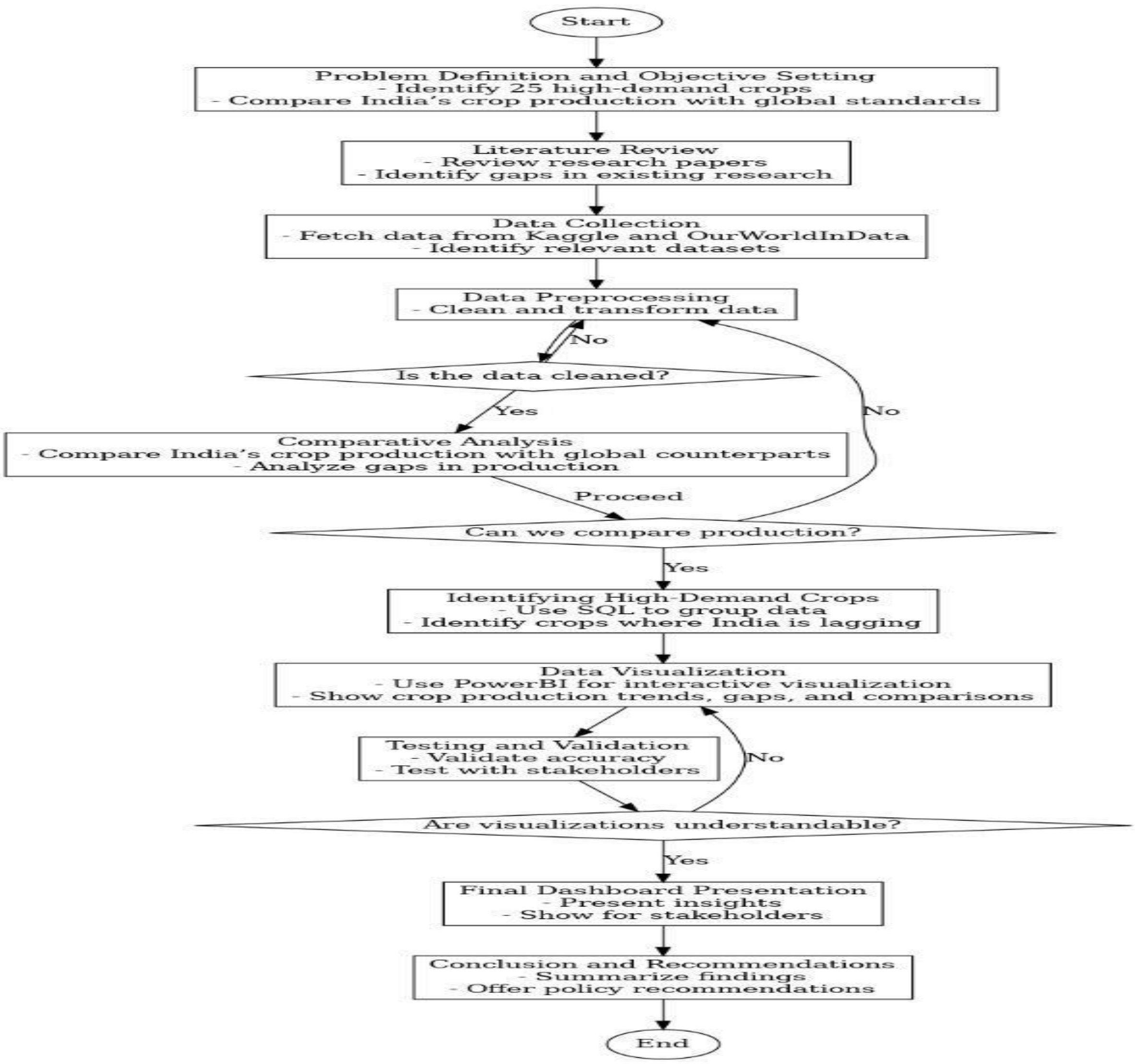
- **DataCollector:** Helps in fetching raw data for the system.
- **DataOrganizer:** Supports structuring and organising data.

- **Visu.alizationTool:** Aids in creating interactive data visualisations.
- **ReportGenerator:** Assists in compiling and exporting the final report.

Purpose of the Diagram

- Illustrates system architecture with different modules.
- Shows dependencies between modules and external tools.
- Ensures scalability by clearly separating concerns.
- Enhances maintainability by structuring the workflow efficiently.

6.6. Proposed Process Flow



.

Fig 6.6: Proposed Process Flow.

6.6.1. Description

The Methodology Flowchart is the graphical illustration of all steps, from problem identification to the final recommendations, to detail how data is gathered, processed, and applied for informed decisions regarding crop production in India vis-a-vis global standards.

6.6.2. Key Steps in the Flowchart

1. **Problem Definition and Objective Setting:**

List 25 high-demand crops and set objectives to compare India's crop production with global standards.

2. Literature Review:

Reviewing research papers to identify existing gaps so that the study is focused and relevant.

3. Data Collection:

Datasets are fetched from reliable sources like Kaggle and OurWorldInData, focusing on identifying relevant data for analysis.

4. Data Preprocessing:

Data cleaning and transformation for accuracy and validation checks to ensure that the data is ready for analysis.

5. Comparative Analysis:

Compare the crop production by India with the global standard to identify which areas India was lagging.

6. Identifying High-Demand Crops:

Going one step deeper by finding crops that have high demand. Use SQL to aggregate and categorise data to see the gaps.

7. Data Visualization:

Application of Power BI for interactive and visual trends, gaps, and comparison in detail for crop production

8. Testing and Validation:

- It tests the accuracy of the visualisations and the feedback collected from the stakeholders to see whether it is precise and reliable.

9. Final Dashboard Presentation:

Critical insights gathered from the analysis are compiled into a final dashboard for stakeholders.

10. Conclusion and Recommendations:

This process will conclude by summarising findings and giving recommendations for crop production improvement based on the analysis.

Decision Points:

Data Cleaning Validation:

It tests whether data is cleaned correctly; if not, then it moves back to the preprocessing step.

Comparison for Production:

It checks whether it can compare validly with the global data. If it cannot, it may require furth. er preprocessing.

Validation for Visualization:

It is a check that ensures visualisations are understandable and clear.

7. TOOLS AND TECHNOLOGIES USED

7.1 Overview

For this project, several tools and technologies will be used strategically to compare high-demand crops in India with global standards. Each of these tools has been selected for its capability in data analysis, visualisation, and project management. A detailed overview of each of these is given below.

7.1.1 Process Tools and Technologies to be Used

1. Data Collection and Storage:

Kaggle: The website will serve as the primary source for collecting high-quality crop data sets. With this vast agricultural database in our hands, the whole analysis becomes more feasible through Kaggle.

OurWorldInData: This data source is essential because it provides us with global datasets. Hence, a comparative study will be conducted between crop production in India and international benchmarks.

SQL (Structured Query Language): SQL will be used to query and organise large datasets. It will help extract, group, and aggregate data to make it easier to spot patterns and insights.

1. Data Preprocessing and Analysis:

Python: Python will be used to clean and preprocess data. Its libraries, such as Pandas and NumPy, will be essential for transforming raw data into a structured format that is appropriate for analysis.

Excel: This will be applied in the preliminary data exploration, work with smaller d.atasets, and simple calculations or checks before applying Python for the detailed analysis.

2. **Data Visualization:**

Microsoft Power BI: This is the primary tool that will be used for data visualisation since it enables the user to create interactive dashboards that give insights into the trends of crop production in India as compared to the global data.

8. IMPLEMENTATION (Pseudocodes)

8.1 Technology Stack:

Implementation of the project involved a mix of tools and technologies to facilitate efficient data processing, analysis, and visualization. The technologies applied include:

Excel: Excel was utilized for preliminary data gathering and structuring. It facilitated efficient data cleaning, formatting, and rapid visualization at the preliminary stages of analysis.

Python: Python was utilized extensively for data preprocessing to perform tasks like dealing with missing values, data transformation, and dataset preparation for SQL-based analysis. Pandas and NumPy libraries were used for these purposes.

SQL: SQL was utilized to query the dataset and derive insights by comparing India's crop production with other nations. It allowed for effective filtering, aggregation, and comparison of big data points to determine patterns and trends.

PowerBI: PowerBI was utilized to develop interactive dashboards that graphically represented the major insights obtained from the dataset. The dashboard effectively pointed out crops where India's production was behind, as well as other trends.

HTML/CSS: HTML and CSS were utilized to create a web interface that incorporates the PowerBI dashboard. The interface created a convenient platform for navigation through the insights, with the addition of report downloads for greater convenience.

.

The combination of these technologies provided a smooth flow of work from data preparation to visualization, making the project more efficient and accurate.

8.2 Data Processing

The data set used in this project had comprehensive data on crop production in 10 countries between 2016 and 2019. It had features like crop names, total production value, area harvested, yield per hectare, regional distribution, soil requirements, and climatic conditions. To make the data set clean, organized, and ready for analysis, the following data processing activities were carried out:

Data Cleaning: Missing values, duplicates, and inconsistencies were checked in the dataset. Missing values were managed using proper methods like forward fill in the case of sequential data and mean imputation for numerical missingness. Duplicate entries were detected and eliminated to avoid redundancy in data.

Data Transformation: Some categorical fields such as 'Regions', 'Soil Requirement', and 'Climatic Conditions' were standardized to ensure uniformity. For instance, spelling differences and inconsistent terms (e.g., "Tropical" vs "Trop.") were standardized to allow for consistency while analyzing.

Data Filtering: In order to concentrate on the main goal — finding crops where India's output was behind — the dataset was filtered to remove all data except India's crop data. This subset was then contrasted with other countries' data to bring out areas where India performed poorly.

Data Aggregation: As the dataset had more than one entry for crop production by different regions and years, data was aggregated with the help of SQL queries. This helped to ensure that values of crop production were grouped for proper year-wise and region-wise comparisons.

Data Preparation for PowerBI: The cleaned and formatted dataset was exported in a PowerBI-compatible format. Proper column naming conventions and data types were maintained to allow easy integration with PowerBI visualizations.

8.2.1. Pseudocode for Data Aggregation:

```
START
    LOAD dataset from Excel
    CHECK for missing values
    IF missing values found:
        HANDLE using forward fill or mean imputation
    ENDIF
    GROUP data BY 'Crop' and 'Year'
    SUM the 'Total Production' column for each group
    OUTPUT the aggregated data as a new DataFrame
STOP
```

Fig 8.1: Pseucdocode for Data Aggregation

8.2.2 Pseudocode for Excel Integration into MySQL

```
START
    IMPORT required libraries
    LOAD dataset from Excel
    CONNECT to SQL database
    CREATE table structure in SQL
    INSERT data row by row
    CLOSE database connection
STOP
```

Fig 8.2: Pseudocode for Excel Integration into MySQL

8.3 PowerBI Insights

The PowerBI dashboard was created to provide end-to-end insights from the crop analysis dataset, with a purpose of determining high-demand crops in ten countries between 2016 and 2019. The steps followed in creating the dashboard were:

Data Preparation:

Importing the dataset into PowerBI was done after a thorough preprocessing in SQL to remove relevant data points. Major filtering parameters were:

- Crop types with major production gaps in India as opposed to other nations.
- Climate conditions favoring the cultivation of such crops in India.
- Local data for focused analysis of cultivation patterns.

The features of data modeling in PowerBI were used to establish connections among different data tables to facilitate smooth interactions in the dashboard. Calculated columns, measures, and DAX were used to add data granularity and create actionable insights.

Dashboard Design:

The dashboard was designed to provide easy navigation and simplicity in data visualization. The major features are:

- **Interactive Filters:** Users dynamically filter data based on country, type of crop, and climate condition to hone in on trends.
- **Comparative Analysis Visuals:** Bar charts, line graphs, and geography heat maps call out production gaps and opportunities for growth.
- **Summary Cards:** Present KPI such as total crop output, highest-performing crops, and opportunity areas.

PowerBI Embed Integration:

To facilitate smooth integration of the PowerBI dashboard into the web interface, the PowerBI Embed Token approach was adopted. This approach provides secure and authorized access to the dashboard. The integration process entailed the following steps:

- **PowerBI Service Configuration:** The dashboard was published to the PowerBI Service workspace, where permissions for access were controlled.
- **Embed Token Generation:** An embed token was created to authorize and authenticate user login into the dashboard securely through PowerBI's REST API.
- **Frontend Integration:** An embed token was integrated into the web app via JavaScript and HTML. The PowerBI iframe URL was embedded in a specific part of the frontend UI to provide seamless user experience.

8.4 Front End Deployment

The frontend interface was designed to provide a seamless and user-friendly experience for visualizing crop analysis insights. The implementation involved the following key elements:

User Interface Design:

The web interface was developed in HTML and CSS for a professional and clean appearance. Major design elements are:

- **Navigation Bar:** A bold header "Crop Analysis Dashboard" for effective branding.
- **Content Sections:** Efficient sections for description, PowerBI dashboard integration, and contact details.
- **Consistent Theme:** Green and white was the color scheme used to achieve a clean and natural theme.

PowerBI Dashboard Integration:

In order to integrate the PowerBI dashboard securely:

- The iframe tag was utilized to embed the PowerBI report directly onto the web page.
- The src attribute inside the iframe points to the PowerBI dashboard URL, and the latest insights are dynamically displayed.
- The button for Download Report allows users to download a detailed report directly through the provided PowerBI Embed URL.

Deployment Strategy:

Frontend was deployed by:

- Hosting the HTML and CSS files on a secure web server.
- Verifying the PowerBI Embed URL was properly set to enable authorized access.
- Applying responsive design concepts to ensure optimal viewing on different screen sizes and devices.

8.4.1 Pse.ucdocode for PowerBI Frontend Integration

```
START
    CREATE HTML structure with:
        - HEADER titled "Crop Analysis Dashboard"
        - SECTION for project overview
        - PLACEHOLDER for PowerBI Dashboard using an iframe
        - BUTTON to download the report

    INSIDE the iframe tag:
        SET 'src' attribute to the PowerBI Embed URL
        ENABLE fullscreen option
        STYLE iframe for optimal display size

    WHEN 'Download Report' button is clicked:
        INITIATE file download using PowerBI Export URL
        IF download successful:
            DISPLAY "Download Successful" message
        ELSE:
            DISPLAY "Error in downloading the report"

    USE CSS media queries to:
        - ADJUST layout for mobile and desktop devices
        - ENSURE the iframe scales appropriately on different screen sizes

STOP
```

Fig 8.3: Pseudocode for PowerBI Frontend Integration

9. RESULTS & DISCUSSION

The project implementation provided useful insights on trends in crop production, India's ranking in world agricultural production, and areas of growth opportunity. The findings were obtained through rigorous data analytics with SQL, visualized via PowerBI, and displayed through a friendly HTML/CSS interface. This section provides the major findings and addresses their implications.

9.1 Identification of Crops Where India is Lagging

By SQL-based analysis, the data was filtered to compare India's crop production with other nations. The analysis showed a number of crops where India's production was much lower even with favorable climatic and soil conditions.

Key Findings:

- Crops like Barley, Rye, and some Oilseeds had significant production deficits in India relative to top producers.
- In a few instances, Indian regions with the best climatic conditions for certain crops were not utilized efficiently, hinting at scope for better agricultural practice.
- India's production in some of the crop categories was significantly less even with a greater cultivated area, showing potential inefficiency in the utilization of resources.

Discussion:

These results imply that India could be helped with specific enhancements to crop choice, farming methods, and resource usage. By detecting these gaps, policymakers and agronomists are able to address improving production in key areas.

9.2 Visual Insights via PowerBI Dashboard

The PowerBI dashboard gave a neat, interactive space to view the analysis. The most notable features of the dashboard were:

- **Comparative Bar Charts:** India's crop production compared to leading producers.
- **Heatmaps:** Showcasing regional variations in crop growth and growth opportunities.
- **Summary Metrics:** Giving concise suggestions of India's standing by crops produced, overall production, and growth trends.

Discussion:

The visual depiction made complicated data easy to understand, which was available to stakeholders like policymakers, researchers, and farmers. The interactive functionalities enabled user.s to dynamically filter insights by crop type, year, and region, facilitating data-driven decision-making.

9.3 Frontend Interface for Improved Accessibility

An exclusive HTML/CSS interface was created to securely embed the PowerBI dashboard. Some of the notable features were:

- **PowerBI Embed Integration:** Providing hassle-free access to insights through secure embedding methods.
- **Download Report Option:** Allowing users to download the most important findings as PDF or CSV files for offline use.
- **Responsive Design:** Providing the best user experience on all devices by adjusting the layout for various screen sizes.

Discussion:

The frontend interface improved user interaction by integrating interactive visualizations with a minimal and intuitive design. This allowed the insights to be accessed and understood by both technical and non-technical users.

10. CONCLUSION AND FUTURE SCOPE

10.1 Conclusion

This project effectively fulfilled the task of determining crops where India is behind in production using thorough data analysis, visualization, and an intuitive web interface. Utilizing Excel, Python, SQL, and PowerBI, the project yielded actionable insights that point out India's agricultural performance compared to other countries.

Through data analysis, the following major findings were brought to light:

- Crops like Barley, Rye, and some Oilseeds showed notable production deficiencies even though India has favorable climatic and regional conditions.
- India's lower yield in some crops may have implied inefficiencies in farming practice, resource use, or technology adoption.
- Regional heatmap analysis revealed locations in India that are underutilized for the

production of high-potential crops, implying possibilities for scaling up agricultural activity.

The PowerBI dashboard played an integral part in displaying these insights in easy-to-understand manner. By providing interactive filtering capabilities and visual comparisons, the dashboard enabled users to delve into trends and spot opportunities for growth in an efficient manner.

Front-end interface development added an extra layer of accessibility by incorporating the PowerBI dashboard within a simple and easy-to-understand layout. A Download Report option was also added to ensure that insights were easily downloadable for offline use.

As a whole, this project effectively combines data visualization with real-world insights to offer useful advice to policymakers, agricultural bodies, and researchers in making informed decisions that can better India's crop production results.

10.2 Future Scope

While the existing implementation successfully serves the project goals, there are some upgrades that can enhance performance, scalability, and user interaction in subsequent versions.

Integration of Real Time Data:

Future releases can integrate real-time data pipelines that automatically update PowerBI dashboard with current crop production records. This would provide dynamic insights and aid timely decision-making.

AI/ML Model Integration:

Integrating machine learning models can enhance forecasting by anticipating crop demand, yield patterns, and possible risks from climate trends and soil types. These forecasting predictions can assist policymakers in strategizing for maximizing crop yields.

Improved Filtering and Analysis:

Integrating enhanced filtering capabilities in PowerBI — including multi-level drilldowns, region-based trends, or soil-type-based analysis — can give farmers and agricultural scientists more insightful information.

Better Fron.t-End Features:

Subsequent versions may add more features to the user interface like interactive tooltips, visual markers, and tailored user options to make the user experience even better.

Mobile Optimization:

Although the present interface is responsive, specific mobile optimization can enhance usability for farmers and on-site farm personnel to enable them to view insights on the move.

Cloud Deployment for Scalability:

Having the entire system deployed on a cloud platform like AWS, Azure, or Google Cloud would enhance performance, scalability, and data protection.

Stakeholder Collaboration:

Through collaboration of the system with agricultural associations, government agencies, and research organizations, this project can develop as an extensive decision-support system to direct large-scale agricultural planning.

11. REFERENCES

[1] Rajasekaran, R., et al. (2020). Intelligent smart farming and crop visualisation. IJMP

[2] Badapanda K., et al (2022). Agriculture data visualisation and analysis using data mining techniques: application of unsupervised machine learning. TELKOMNIKA (Telecommunication Computing Electronics and Control) & Technology, 10, 2250-1770.

[3] Chergui, N., & Kechadi, M, T. (2022). Data analytics for crop management: a big data view. Journal of Big Data, SpringerOpen

[4] J, Kavitha, et al. (2024). Role of Data Visualization and Big Data Analytics in Smart Agriculture. IGI Global

[5] B, P, Harshita, et al. (2017). Agricultural Data Visualization for Prescriptive Crop Planning. International Journal of Computer Trends and Technology (IJCTT) – Volume 49 Number 3 July 2017

[6] Crop sta.tistics FAO - All countries – Kaggle Dataset containing crop data from the year 1961 to 2019 across all countries.

[7] Agricultural Production - Our World in Data – A website containing data and visualisations for all countries from the year 1961 to 2022.

[8] Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India. (2019). Agricultural statistics at a glance 2019.

[9] National Portal of India. (2016). *Agricultural statistics at a glance 2016*.

[10] Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India. (n.d.). *Area, production & yield - Reports*.

[11] Open Government Data (OGD) Platform India. (n.d.). *Crop production statistics dataset*.

[12] Indian Agricultural Statistics Research Institute. (n.d.). *Production and productivity in Indian agriculture during past 70 years*.

[13] ShivLab. (n.d.). *Integrating Power BI into your web application: Step-by-step guide*.

[14] Imenso Software. (n.d.). *Seamless integration of Power BI with web applications: A guide*.

[15] Microsoft. (n.d.). *Embed a report in a secure portal or website*.

[16] Microsoft. (n.d.). *Embed content in your Power BI embedded analytics application*.

[17] Process Street. (n.d.). *How to embed Power BI report in web application*.

[18] SQLShack. (n.d.). *How to embed a Power BI report server report into an ASP.Net web application*.

[19] Maya Insights. (n.d.). *How to embed Power BI report in web application*.

[20] Microsoft. (n.d.). *Power BI embedded analytics client APIs*.

[21] Microsoft. (n.d.). *Power BI REST APIs for embedded analytics*.

[22] Microsoft. (n.d.). *Embedding Power BI content in web applications*.

[23] Microso. ft. (n.d.). *Power BI JavaScript API reference*.

[24] Microsoft. (n.d.). *Power BI developer samples*.

[25] Microsoft. (n.d.). *Power BI embedded playground*.

[26] Microsoft. (n.d.). *Power BI embedded analytics capacity planning*.

[27] Microsoft. (n.d.). *Power BI security whitepaper*.

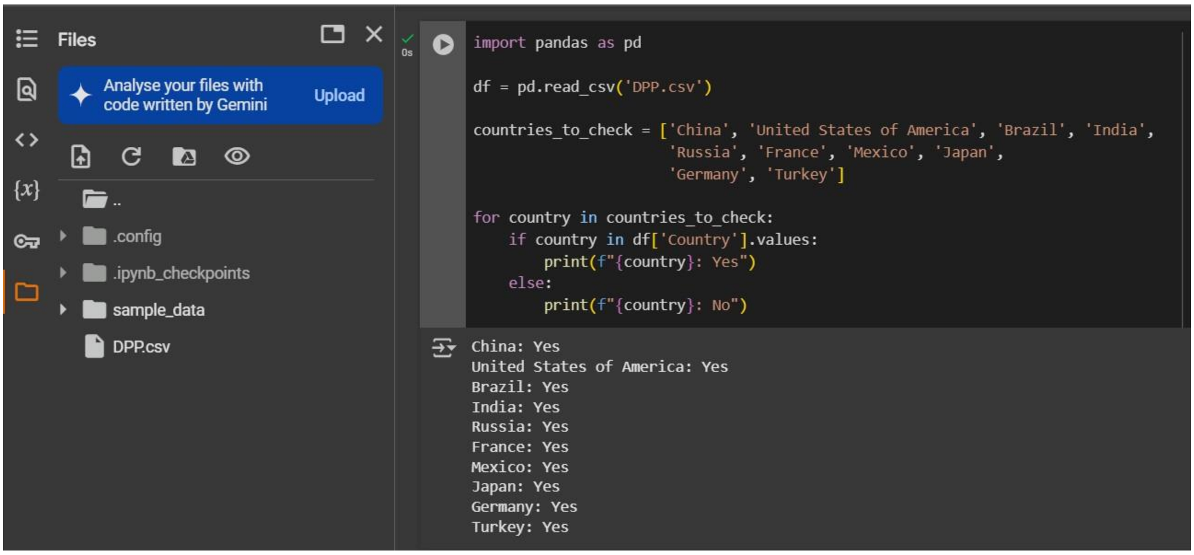
[28] Ramasamy, R. (n.d.). *Crop statistics FAO - All countries*. Kaggle.

[29] Food and Agriculture Organization of the United Nations. (n.d.). *FAOSTAT: Crops and livestock products*.

12. Annexure 1:

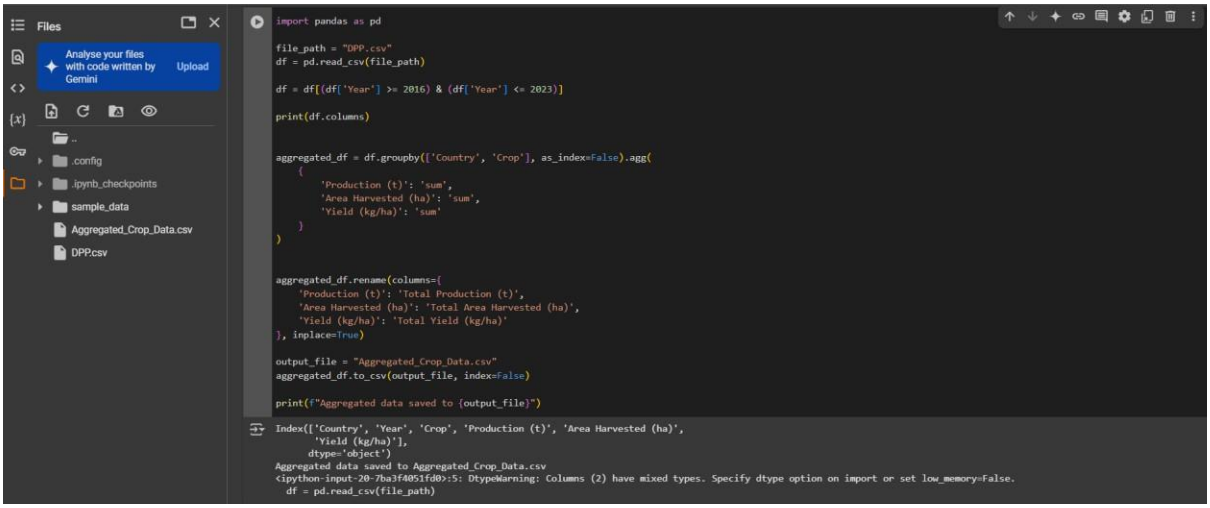
12.1 : Python Program To Check Whether The Dataset Contains All The Required Countries Are Not. Especially India.

Dataset: DPP.csv



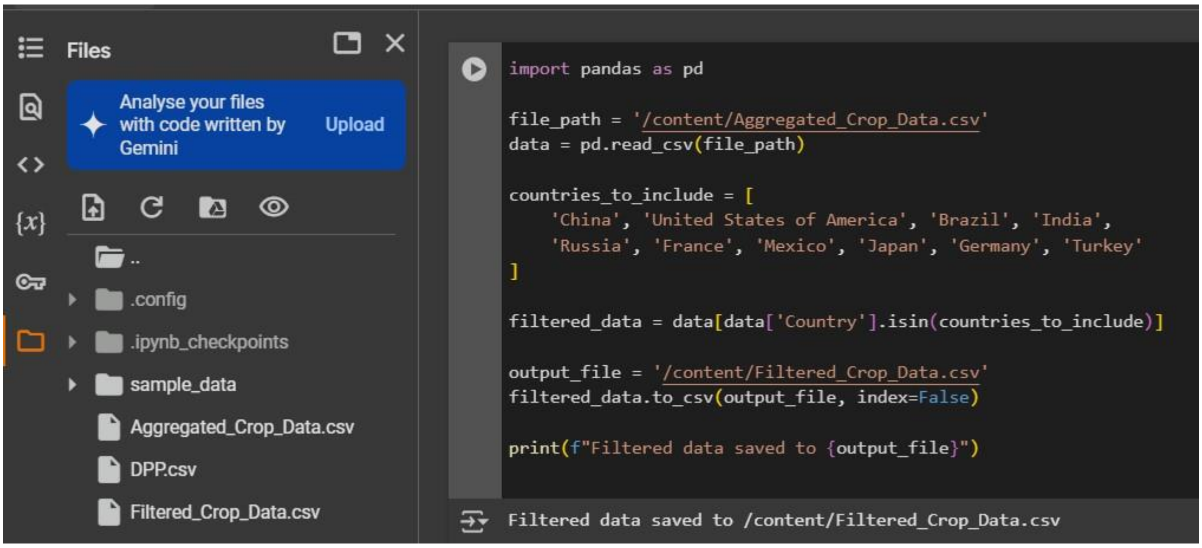
12.2: Python. Program To Aggregate The Dataset From The Year 2016 To 2023, Check The Total Production, Total Area Harvested, and Total Yield from The Aggregated Dataset.

Dataset: Aggregated_Crop_Data.csv



12.3: Python Program For Filtering The Required Countries.

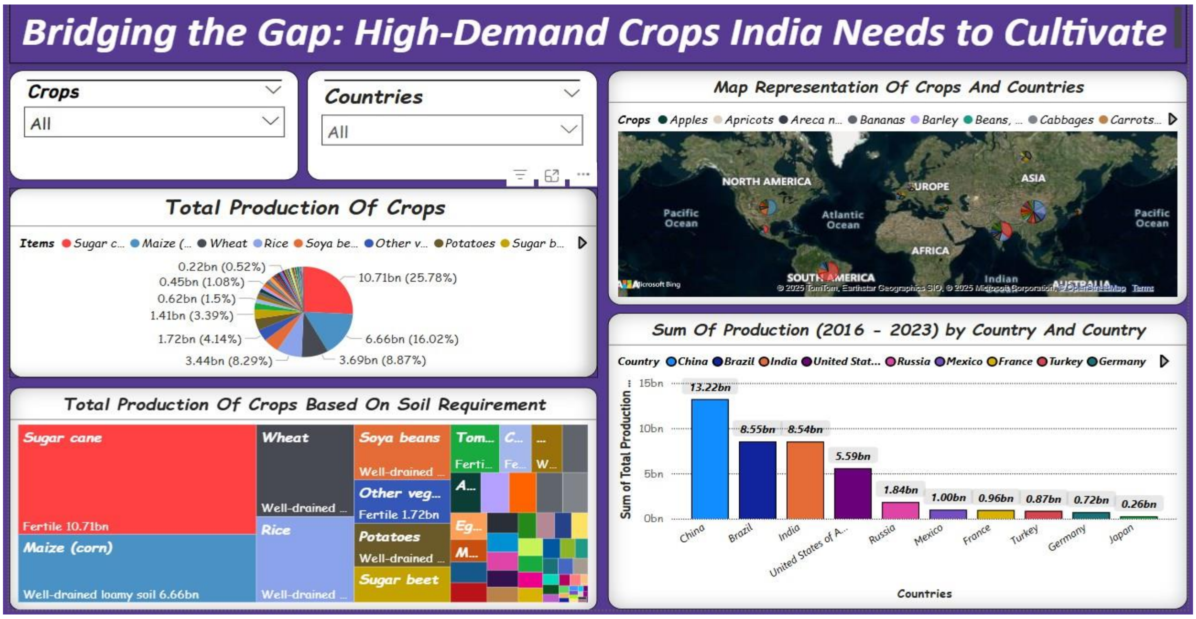
Dataset: Aggregated_Crop_Data.csv



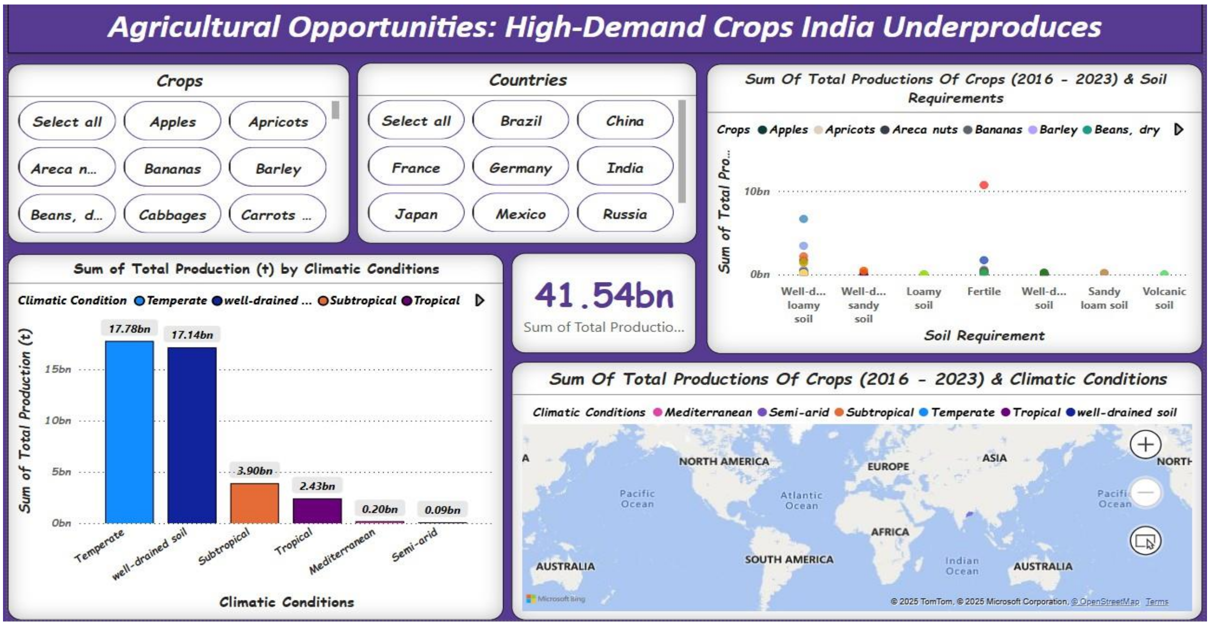
.

13. **Annexure 2:** Final Power BI Dashboard.

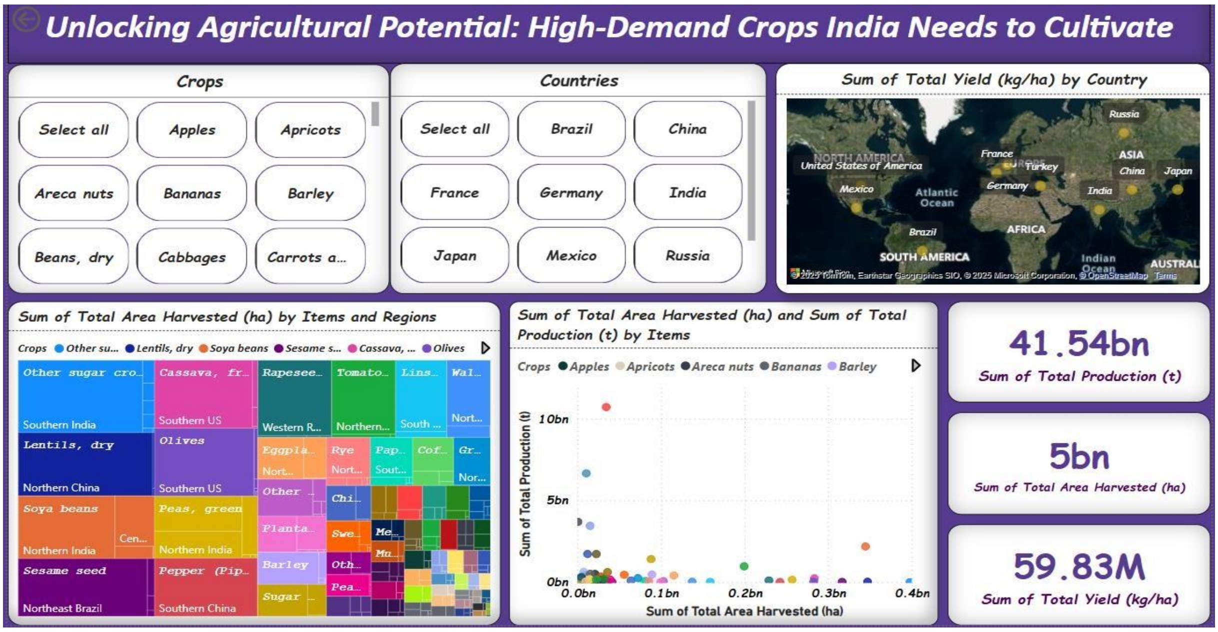
13.1 : A visual dashboard highlighting global crop production trends, soil requirements, and India's position in high-demand crop cultivation.



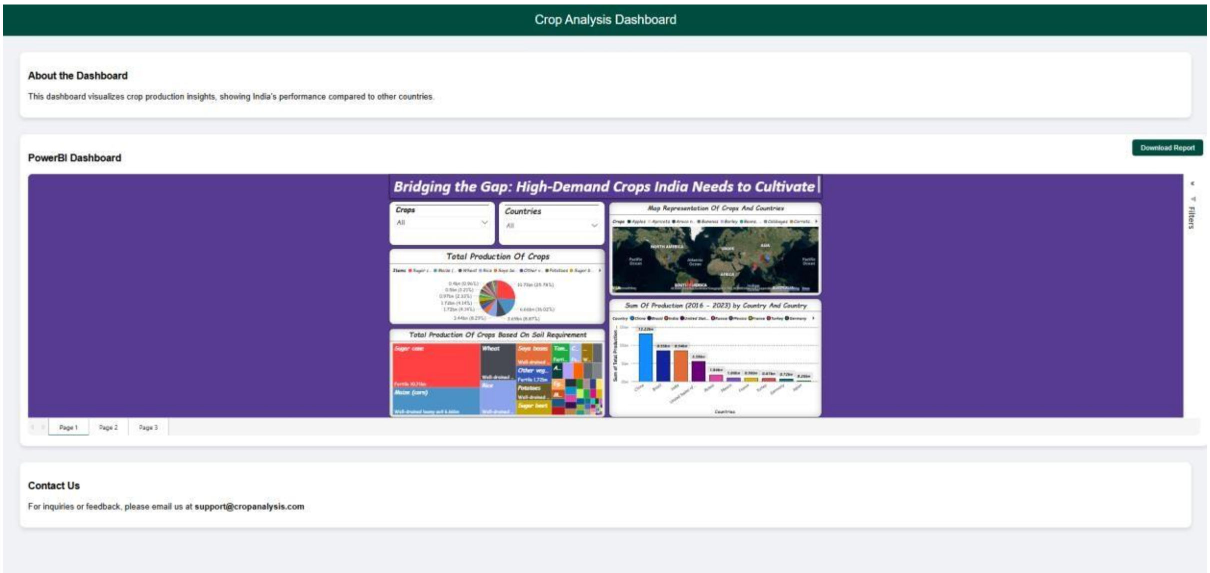
13.2 : A concise dashboard showcasing high-demand crops that India underproduces, emphasizing climatic conditions, soil requirements, and global production insights.



13.3 : A focused dashboard highlighting high-demand crops India needs to cultivate, featuring insights on production, harvested area, yield, and regional trends.



13.4 : Using Few Web Development Frameworks Making The Dashboard Public To Get Some Insights.



Crop Analysis Dashboard
About the Dashboard
This dashboard visualizes crop production insights, showing India's performance compared to other countries.
PowerBI Dashboard
Download Report
Bridging the Gap: High-Demand Crops India Needs to Cultivate
Crops
All
Countries
All
Map Representation Of Crops And Countries
Total Production Of Crops
Total Production Of Crops Based On Soil Requirement
Sum Of Production (2016 - 2023) by Country And Country
Filters
Page 1
Page 2
Page 3
Contact Us
For inquiries or feedback, please email us at support@cropanalysis.com