Project Title	AgriData Explorer: Understanding Indian agriculture with EDA
Skills take away From This Project	Python script,data cleaning, EDA, SQL, Power BI
Domain	Agricultural, Mainly focusing on crop production statistics, yields, and farming area in India

Problem Statement:

India's agricultural sector is vital for the economy, but the management of agricultural data remains a challenge due to its complexity, fragmented nature, and lack of easy access. Various stakeholders such as farmers, policymakers, and researchers face difficulties in accessing, analyzing, and making informed decisions based on agricultural data.

The goal of this project is to create a data visualization platform that integrates agricultural data from different states and districts in India. The platform will provide insights into crop production, yields, and areas under cultivation, making it easier for users to identify trends, gaps, and regional disparities. This solution aims to help farmers optimize crop choices, assist policymakers in targeting areas for intervention, and support researchers in analyzing agricultural trends.

Business Use Cases:

Farmers:

- Use Case 1: A farmer can explore historical crop production and yield data to make informed decisions on which crops to cultivate in different seasons based on past performance.
- Use Case 2: The farmer can analyze regional productivity and identify potential areas for improvement, such as soil health or irrigation practices.

Policymakers:

- Use Case 1: Government officials can use the platform to assess regions with low productivity and allocate resources or subsidies to improve crop yields and farmer livelihoods.
- Use Case 2: Identifying crops with fluctuating yields to develop better policy for crop insurance and risk management.

Researchers:

- Use Case 1: Researchers can use the platform to analyze the impact of climate, soil, and irrigation on crop yields over time.
- Use Case 2: The platform can help researchers identify potential areas for agricultural innovation, such as high-yielding varieties or pest management techniques.

Approach:

Data Collection and Cleaning:

- Collect data from government sources, surveys, and agricultural reports.
- Clean the data by addressing missing values, standardizing units, and ensuring consistency across various fields (e.g., converting hectares, tons, and kilograms).

Data Analysis:

- Perform exploratory data analysis (EDA) to understand trends, patterns, and correlations.
- Use statistical methods to identify insights such as high and low production areas, top-performing crops, and yield optimization strategies.

Visualization:

- Create interactive dashboards using tool Plotly for real-time data visualization.
- Develop key charts such as line graphs for time-series data, bar charts for comparing production across regions, and heatmaps to show crop concentration in different states.

User Interface Design:

- Develop a user-friendly interface with filtering options for users to explore the data based on region, crop type, yield, and year.
- Implement interactive elements like dropdowns, sliders, and tooltips to enhance user experience.

Advanced Analytics:

- Integrate machine learning models to forecast crop yields, predict the best areas for cultivation, and offer recommendations for crop diversification.

Power BI Integration:

Steps to Create Data Visualizations in Power BI:

1. Prepare Your Data

- **Source Data:** Ensure you have structured data (CSV, Excel, SQL database, etc.) with clear headers and consistent data formats.
- **Data Fields:** Your dataset should include fields like:
  - **Crop Name** (e.g., rice, wheat, maize)
  - **Region** (e.g., state, district)
  - **Area Cultivated** (in hectares)
  - **Production Volume** (in tons)
  - **Yield** (in kg/hectare)
  - **Year**
- If your data is in separate tables (e.g., crop data, weather data, economic data), use **Power BI's data model** to establish relationships between them.

2. Load Data into Power BI

- Open Power BI Desktop.
- Click on **Get Data** and choose your data source (Excel, CSV, Database, etc.).
- Load your dataset into Power BI.

- Review and clean the data if needed (e.g., checking for missing values or duplicates) using Power BI's **Power Query Editor**.

3. Create a Data Model

- If you have multiple datasets (e.g., crop production data, weather data), define **relationships** between these tables (e.g., linking region names or crop types).
- Go to the **Model View** to establish relationships between tables (drag and drop fields to create relationships).
- Set the **Primary Key** (e.g., region code, crop type) to link tables correctly.

4. Design the Dashboard

- **Add Visualizations:** Power BI provides several visualization options to display agricultural data effectively. Below are the types of visualizations that you can use:
  - **Geographical Heatmaps:**
    - Use **Map** visualizations (e.g., **Filled Map** or **Shape Map**) to display regional data.
    - This visualization can show the **crop yield**, **production**, or **area cultivated** by state or district.
    - Color-code the map to show performance differences (e.g., red for low yield, green for high yield).
  - **Trend Lines/Bar Charts:**
    - Use **Line Charts** to display crop production or yield over time (e.g., trends of rice production over the last decade).
    - **Bar/Column Charts** can be used for comparing **crop production** across different regions for a specific year.
  - **Scatter Plots:**
    - Use a **Scatter Plot** to visualize relationships between two variables (e.g., correlation between crop area and production).
    - This can also be used to analyze how inputs (e.g., fertilizer usage) affect output (crop yield).
  - **KPI Indicators:**
    - Use **KPI (Key Performance Indicator)** visualizations to show if the production goals are met. For instance, you can show the target rice yield against the actual yield.

■ Add **cards** or **gauges** to show metrics like total area cultivated, total production, or average yield.

5. Interactive Filters and Slicers

● Add **Slicers** to let users filter data based on different attributes, such as:
  ○ **Crop Type** (e.g., rice, wheat, maize)
  ○ **Region** (states, districts)
  ○ **Year**
● Slicers help in drilling down into specific data points. For example, selecting a specific state will update the map and trend charts with data for that state.
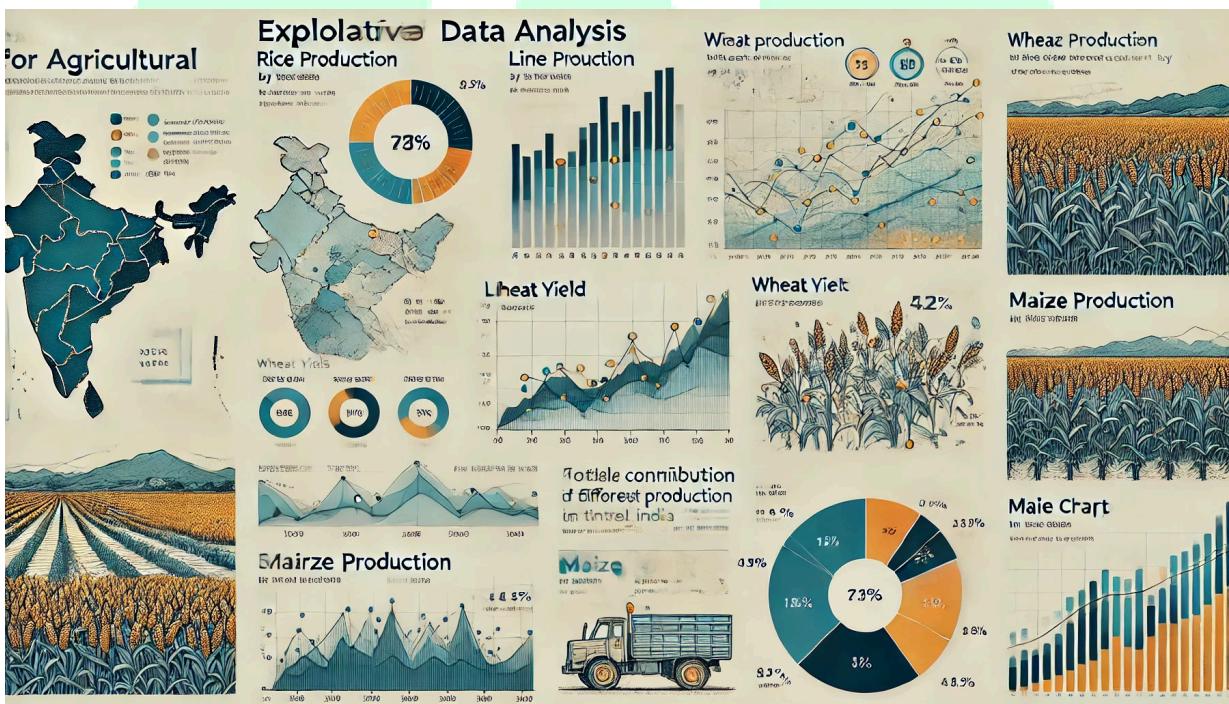
6. Enhance the Visualizations:

● **Formatting:** Customize the look and feel of your visuals by adjusting colors, labels, titles, and tooltips.
● **Interactivity:** Add interactions between visuals. For instance, clicking on a region on the map should update the bar chart with crop production data for that region.
● **Dynamic Titles:** Use dynamic titles that change based on slicer selections, so users know what data they are looking at.

Exploratory Data Analysis (EDA):

● Top 7 RICE PRODUCTION State Data(Bar_plot)

● Top 5 Wheat Producing States Data(Bar_chart)and its percentage(%)(Pie_chart)

● Oil seed production by top 5 states

● Top 7 SUNFLOWER PRODUCTION State

● India's SUGARCANE PRODUCTION From Last 50 Years(Line_plot)

● Rice Production Vs Wheat Production (Last 50y)

● Rice Production By West Bengal Districts

● Top 10 Wheat Production Years From UP

- Millet Production (Last 50y)

- Sorghum Production (Kharif and Rabi) by Region

- Top 7 States for Groundnut Production

- Soybean Production by Top 5 States and Yield Efficiency

- Oilseed Production in Major States

- Impact of Area Cultivated on Production (Rice, Wheat, Maize)
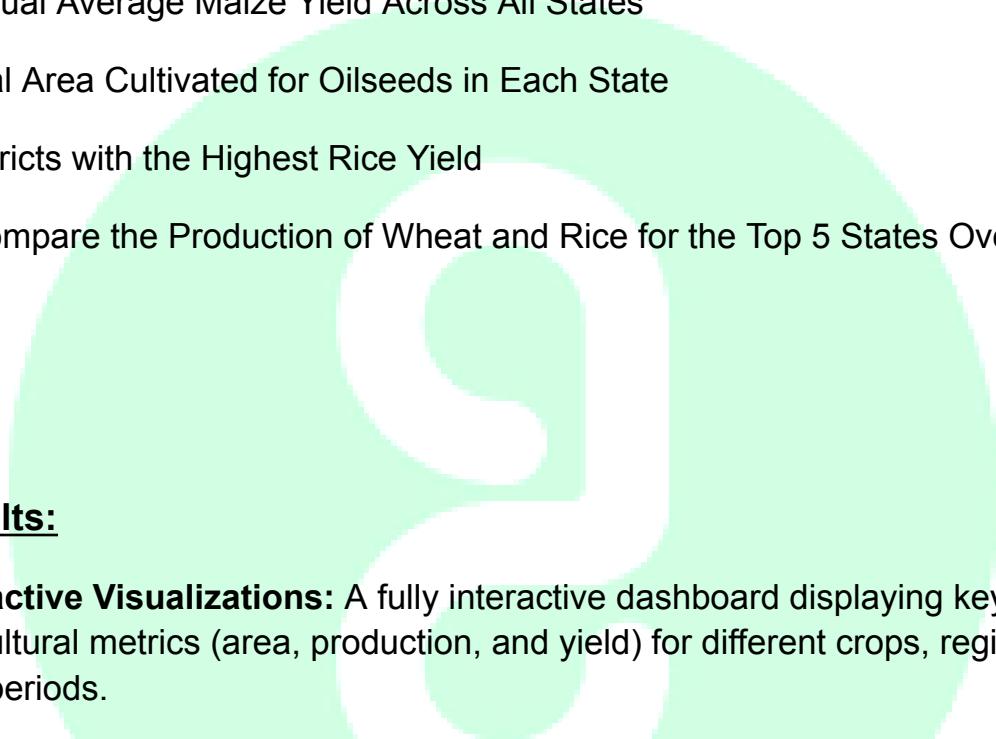
- Rice vs. Wheat Yield Across States

SAMPLE PICTURE OF EDA:

Questions to be answered:(SQL)

Answer the below questions through your analysis and visualizations:

1.Year-wise Trend of Rice Production Across States (Top 3)

2.Top 5 Districts by Wheat Yield Increase Over the Last 5 Years

3.States with the Highest Growth in Oilseed Production (5-Year Growth Rate)

4.District-wise Correlation Between Area and Production for Major Crops (Rice, Wheat, and Maize)

5.Yearly Production Growth of Cotton in Top 5 Cotton Producing States

6.Districts with the Highest Groundnut Production in 2020

7.Annual Average Maize Yield Across All States

8.Total Area Cultivated for Oilseeds in Each State

9.Districts with the Highest Rice Yield

10.Compare the Production of Wheat and Rice for the Top 5 States Over 10 Years

Results:

Interactive Visualizations: A fully interactive dashboard displaying key agricultural metrics (area, production, and yield) for different crops, regions, and time periods.

Trend Identification: Identified patterns in crop production, areas with the highest and lowest yields, and the effects of seasonality on agricultural outcomes.

Recommendations: Based on data insights, the platform provides recommendations for farmers on crop selection and best practices for improving productivity in specific regions.

Project Evaluation metrics:

- **Accuracy of Data Visualizations:** Ensuring the visual representations are clear, accurate, and easy to interpret.
- **User Engagement:** Measuring how frequently users interact with different filters, charts, and reports to assess the usefulness of the tool.
- **Performance:** Evaluating the speed and responsiveness of the dashboard (e.g., loading time, filter application speed).
- **Data Completeness:** Ensuring that all relevant fields are present and accurate in the dataset (e.g., crop areas, yields, production numbers).
- **User Satisfaction:** Conducting surveys or interviews to understand user experience and gather feedback for improvement.

Technical Tags:

Python,SQL, Power BI, Data Cleaning, Data Analysis, Data Visualizations

Data Set:

Data Set Link**: 🗓 ICRISAT-District Level Data**

Data Set Explanation:

1. Dist Code:

- Description: A unique identifier for each district.
- Data Type: Integer or String (unique to each district)

2. Year:

- Description: The year in which the data was recorded for a given crop.
- Data Type: Integer (e.g., 2020, 2021, etc.)

3. State Code:

- Description: A unique identifier for each state in India.
- Data Type: Integer or String (unique to each state)

4. State Name:

- Description: The name of the state in India (e.g., Uttar Pradesh, West Bengal).
- Data Type: String

5. Dist Name:

- Description: The name of the district within the state.
- Data Type: String

Crop and Agricultural Data Fields

6. RICE AREA (1000 ha):

- Description: The area under rice cultivation in a given district or state, measured in thousand hectares (1000 ha).
- Data Type: Numeric (e.g., 2000 for 2000 hectares)

7. RICE PRODUCTION (1000 tons):

- Description: The total rice production in a district or state, measured in thousand tons.
- Data Type: Numeric (e.g., 500 for 500 tons)

8. RICE YIELD (Kg per ha):

- Description: The rice yield per hectare, expressed in kilograms per hectare.
- Data Type: Numeric (e.g., 3000 kg/ha)

9. WHEAT AREA (1000 ha):

- Description: The area under wheat cultivation in a given district or state, measured in thousand hectares.
- Data Type: Numeric

10. WHEAT PRODUCTION (1000 tons):

- Description: The total wheat production in a district or state, measured in thousand tons.

- Data Type: Numeric

11. WHEAT YIELD (Kg per ha):

- Description: The wheat yield per hectare, expressed in kilograms per hectare.
- Data Type: Numeric

Other Crop Data Columns

The dataset also contains data for other crops such as Sorghum, Pearl Millet, Maize, Finger Millet, Barley, Chickpea, Pigeonpea, Groundnut, Sesamum, Rapeseed and Mustard, Safflower, Castor, Linseed, Sunflower, Soybean, and Cotton. Each of these crops has the following data points:

1. Crop Area (1000 ha): The area cultivated with that particular crop (in hectares).
2. Crop Production (1000 tons): The total production of the crop (in tons).
3. Crop Yield (Kg per ha): The yield per hectare for that crop (in kilograms per hectare).

Other Agricultural Data

12. OILSEEDS AREA (1000 ha):

- Description: The area under oilseed crops (e.g., Soybean, Groundnut, etc.), measured in thousand hectares.
- Data Type: Numeric

13. OILSEEDS PRODUCTION (1000 tons):

- Description: The total oilseed production in the district or state, measured in thousand tons.
- Data Type: Numeric

14. OILSEEDS YIELD (Kg per ha):

- Description: The yield of oilseeds per hectare, expressed in kilograms per hectare.

- Data Type: Numeric

15. SUGARCANE AREA (1000 ha):

- Description: The area cultivated with sugarcane in a district or state.
- Data Type: Numeric

16. SUGARCANE PRODUCTION (1000 tons):

- Description: The total sugarcane production in a district or state, measured in thousand tons.
- Data Type: Numeric

17. SUGARCANE YIELD (Kg per ha):

- Description: The sugarcane yield per hectare, expressed in kilograms per hectare.
- Data Type: Numeric

18. COTTON AREA (1000 ha):

- Description: The area under cotton cultivation, measured in thousand hectares.
- Data Type: Numeric

19. COTTON PRODUCTION (1000 tons):

- Description: The total cotton production in a district or state, measured in thousand tons.
- Data Type: Numeric

20. COTTON YIELD (Kg per ha):

- Description: The cotton yield per hectare, expressed in kilograms per hectare.
- Data Type: Numeric

Other Crops (Fruits, Vegetables, Pulses, etc.)

In addition to the main crops like rice, wheat, and oilseeds, the dataset contains data on:

1. Fruits Area (1000 ha): Area under fruit cultivation.
2. Vegetables Area (1000 ha): Area under vegetable cultivation.
3. Fruits and Vegetables Area (1000 ha): Combined area under fruits and vegetables cultivation.
4. Potatoes Area (1000 ha): Area under potato cultivation.
5. Onion Area (1000 ha): Area under onion cultivation.
6. Fodder Area (1000 ha): Area under fodder crops.

Project Deliverables:

Source Code:

- Python scripts for data extraction and cleaning.
- SQL queries for creating and populating tables in the database.

SQL Database:

- Structured and normalized database containing the cleaned data.

Power BI Reports:

- A set of interactive reports and dashboards showcasing key insights.

Documentation:

- Detailed documentation explaining the process, challenges faced, and solutions implemented.

Project Guidelines:

- Use best practices for SQL database design and normalization.
- Follow Power BI best practices for creating interactive, user-friendly dashboards.

References:

Project Live Evaluation Metrics	Project Live Evaluation
EDA Guide	Exploratory Data Analysis (EDA) Guide
Capstone Explanation Guideline	Capstone Explanation Guideline
GitHub Reference	How to Use GitHub.pptx
HOW TO ESTABLISH SQL CONNECTION	PYTHONSQLCODE_TAMIL.ipynb
Power BI tutorial	PPB - The Ultimate Power BI
Connect MySQL to Power BI	Connect MySQL to PowerBI
Steps to Create a Free Power BI Work Account	Steps to Create a Free Power BI Work Account

POWERBI_MYSQL_CONN_ISSUe	[POWERBI_MYSQL_CONN_ISSUE](#)
Project Orientation (Tamil)	📄 Project Orientation Session : Ag…
Project Orientation (English)	🎬 Agri Data Explore_Eng

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.
Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (4:00PM to 5:00PM)

Booking link :https://forms.gle/XC553oSbMJ2Gcfug9

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.
Note: This form will Open only on Saturday (after 2 PM) and Sunday on Every Week

Timing:
For DS and AIML
Monday-Saturday (05:30PM to 07:00PM)

Booking link : https://forms.gle/1m2Gsro41fLtZurRA

Evaluation Metrics : [Project Live Evaluation](#)

Approval Workflow

Created By:	Verified By:	Approved By:
Gomathi A		