Agriculture Insights Dashboard

A Power BI–Based Agricultural Data Analysis Project

Aditya Rajput

B.Tech Computer Engineering

2025

1. Abstract

Agriculture still ranks among the most vital elements within India's economy, but it appears that all data associated with production, crop yield, rainfall, and conditions related to climate is dispersed and maintained as unstructured data. As a result, it becomes extremely difficult for anyone interested in the issue to draw meaningful insights.

At present, there are challenges associated with agriculture insights. The project will solve these problems with an interactive insights dashboard in Agriculture created using Power BI. To achieve this, various datasets on agriculture include crop production, crop yield, rainfall per year, and suitability of temperatures. As a result, there will be thorough data cleansing and processing. Use of DAX measures and data modeling will enable the insights dashboard to provide an overarching perspective on agriculture performance.

The insights offered by the dashboard include identifying high-yielding crops, understanding state-wise production variability, identifying seasonal production variability, and understanding the impact of rainfall, temperature, and productivity on agriculture. All these insights help stakeholders analyze and identify ways for improving efficiency in crop management.

From the above, it can be ascertained that this project exemplifies the conversion of intricate agricultural data sets into an insightful and graphical decision-making tool. The utilization of data analytics and visualization capabilities for superior agricultural planning, policy formulation, and data-informed agriculture growth can hardly be underestimated.

2.Introduction

Agriculture sustains the Indian economy substantially and contributes significantly to employment generation, food security, and national income. Not withstanding agriculture's importance for India, agriculture data ends up being unorganized and uninterpretable. Typically, data about crop production, area under cultivation, crop productivity, rainfall variability, and climate variability are maintained on separate spreadsheets and reports. All with all, due to a lack of efficient tools for agriculture data analysis, vital relationships within the data remain unexplored.

Technological advancements made in the area of data analytics and business intelligence tools, for instance, Power BI, have made it possible for farmers to turn unstructured agricultural data into structured and meaningful information. Through these tools, farmers are capable of viewing and interpreting large chunks of data.

The objective of this project will be to develop an Agriculture Insights Dashboard that can combine various datasets like crop production, crop yield, rainfall, and temperature suitability for analysis. It will be working towards addressing the fact that there are no visualization tools available for agriculture.

By employing this dashboard, data interpretations about production at the state level, insights into seasonal fluctuations in crop production, determinations about the effects of climate on agriculture, and determination about high and low production crop types can be achieved. Overall, it can be realized that this project represents an efficient method for enhancing agriculture analysis with visualization.

3.Methodology

The methodology adopted in the project follows a structured data analysis lifecycle, ensuring accurate preparation, meaningful exploration, and effective visualization of agricultural data. This critical process contains the following key stages:

1. Data Collection

The datasets included: several datasets related to Indian agriculture; these are listed below.

- Crop Production
- Crop Yield
- Annual Rainfall
- Temperature Suitability
- Information about Season and Crop Type

These datasets provided comprehensive coverage of agricultural performance, climate factors, and cultivation patterns across various states and years.

2. Data Cleaning and Preprocessing

These inconsistencies, missing values, duplicates, and formatting discrepancies were part of the raw datasets.

Cleaning steps included:

- Removing duplicate and redundant records.
- Handling missing numerical and categorical values
- Standardization of crop names, state names, and season categories
- This is turning text-based numeric fields into appropriate number formats.
- Smoothing unit inconsistencies across datasets

This preprocessing step ensured that the data was dependable and consistent for further analysis.

3. Data Integration and Transformation

These cleaned datasets were then combined into one master table, using the common fields of State, Crop, Year, and Season.

Other changes made included:

- Creating derived metrics such as Yield per Hectare

- Temperature suitability metrics derivation
- Structuring the data model in Power BI

This step allowed direct comparison among the crops, regions, and climate conditions.

4. Exploratory Data Analysis (EDA)

EDA was performed in order to understand the distribution, relationships, and patterns in the dataset.

This included:

- Identification of high and low production states
- Seasonal trend analysis
- Rainfall and temperature characteristics analysis
- Outlier and anomaly detection

EDA insights were used to inform the development of meaningful visualizations.

5. Power BI Dashboard Development

The following elements were implemented using Power BI:

- KPI Cards: Total Area; Total Production; Total Rainfall
- Charts & Visuals: Bar charts, pie charts, line charts, card visuals, and heat-style visuals
- Filters & Slicers: State, Year, Crop, Crop Type, Season
- DAX Measures: These are used to create totals, averages, and calculated dynamic metrics.

It should be an interactive, intuitive, visually appealing dashboard that enables users to explore the insights dynamically.

6. Abstracting Insight and Interpreting:

Final steps included analyzing the visualizations placed on the dashboard, which could provide the following pattern or insight:

- State-wise production trends
- High-yielding crops
- Seasonal variations in productivity
- Climate impact on agriculture

4. Dataset Overview

This project uses various datasets, including some related to agriculture production, climate, and crop attributes, among others, within India. Datasets have unique information that needs to be incorporated into a complete picture with regard to agriculture performance. Various datasets have been brought together to form an integrated analytical model that can be visualized using Power BI.

1. Crop Production Dataset

It contains data about production at a state level for various years.

Key fields include:

- State Name
- Crop
- Year
- Area under cultivation (in hectares)
- Production (in tons)
- Season (Kharif, Rabi, Whole Year, etc.)

2. Crop Yield Dataset

It contains data on the yield (output per hectare) for various crops and different states.

Purpose: To assess efficiency and generate comparative data with regards to crop variety.

Key fields:

- Crop
- Year
- Yield per hectare

3. Rainfall Dataset

It contains rainfall data on an annual basis for various states.

Key fields

- State
- Year
- Annual Rainfall (mm)

The rainfall information aids in examining the relationship that exists between climate and crop production.

4. Temperature Dataset

This data contains temperature ranges for different types of crops like Kharif, Rabi, and Whole Year.

Key fields:

- Crop Type
- Minimum Temperature
- Maximum Temperature

Temperature suitability helps with climate-based agriculture analysis.

5. Merged Master Dataset

After pre-processing and transforming respective datasets, they were integrated into a common dataset, which enabled cross-dimensional analysis on:

- States
- Crops
- Years
- Seasons
- Climate conditions

Final Dataset Fields included were:

- State
- Crop
- Crop Type
- Season
- Year
- Area
- Production
- Yield
- Rainfall
- Temperature Suitability

The result of combining these datasets is the foundation for the Power BI dashboard.

6. Dataset Size and Structure:

In all the files, the total set includes:

- Thousands of rows spanning several states and types of crops
- 10+ Key Fields That Serve for Visualization and Analysis Goals
- Key Data spanning several years

6. Data Cleaning Steps

Raw agricultural datasets often contain inconsistencies, missing values, formatting errors, and structural differences that must be addressed before meaningful analysis can take place. In this project, extensive data cleaning was performed to ensure high-quality, reliable, and consistent data suitable for integration and visualization. The major cleaning steps included:

1. Removal of Duplicate and Irrelevant Records

All datasets were checked for duplicate entries and unnecessary rows. The repeated entries were gone so that the calculations and the corresponding visualizations were not affected by distortion and were accurate.

2. Handling Missing Values

Several datasets contained missing or incomplete information, especially in rainfall, yield, and temperature fields.
Steps included:

- Replacing missing numeric values with mean/median (where appropriate)
- Filling categorical gaps based on domain knowledge
- Removing rows with insufficient data when no reliable estimation was possible

3. Standardization of Crop, State, and Season Names

Different files used inconsistent naming formats. To ensure uniformity:

- Spelling variations were corrected (e.g., “Uttar Pradesh”, “uttar pradesh”, “Uttar pradesh”)
- Extra spaces, special characters, and capitalization differences were removed
- Crop types and season categories (Kharif, Rabi, Whole Year) were standardized

This step ensured smooth merging across datasets.

4. Converting Text-Based Numeric Fields

Some datasets stored numeric values (e.g., rainfall, production, area) as text due to formatting issues.
All numeric fields were cleaned by:

- Removing commas and symbols
- Converting data types to integers or decimals
- Ensuring consistent units (e.g., mm for rainfall, tons for production)

5. Fixing Discrepancies Between Files

Since the datasets came from different sources, formatting and structural inconsistencies existed across files.
These were corrected by:

- Aligning column names
- Matching field formats
- Ensuring uniform year representation
- Validating row alignment for merging

6. Outlier Detection and Correction

Extreme values—such as unusually high production or rainfall—were reviewed. Outliers caused by entry errors were corrected or removed to maintain dataset accuracy.

7. Creation of Calculated Columns

New fields were created to support analysis, including:

- **Yield per Hectare = Production / Area**
- Temperature suitability metrics based on crop type

These derived fields added analytical depth to the dashboard.

9. Final Validation

The merged dataset was checked for:

- Logical consistency
- Numerical correctness
- Missing field dependencies
- Proper alignment across all attributes

This step ensured that the final dataset was accurate and ready for visualization.

7.Exploratory Data Analysis

Exploratory Data Analysis (EDA) procedures were conducted on the agriculture datasets before proceeding with creating the dashboard. EDA enabled an understanding of the structure and distribution of data that would be vital for informing the visualization process. The main components of EDA will be highlighted below:

1. Comprehending Data Distribution

The preliminary analysis involved exploring ways and means to analyze the distribution of important variables like production, area, yield, rainfall, and temperature.

By employing summary statistics (mean, median, minimum, and maximum), it was determined that:

- Large production variability across the states
- Varieties of Crops Produced
- Variation in rainfall distribution among regions

It marked the beginning of understanding the datasets.

2. State-Wise Production Analysis

Bar charts and aggregation methods were employed on a state-by-state basis to analyze variability in crop production.

Findings included:

- Some States like Uttar Pradesh and Tamil Nadu have maintained high production levels.
- “Other” provinces have shown considerably lower production levels, suggesting room for improvement

It aided in creating visualization designs which depicted the top and bottom-performing states.

3. Crop Yield Trends

The yield factors were compared based on the efficiencies achieved by various crops.

EDA discovered:

- Yield per hectare for sugarcane stood at ___

- Some core food plants, like rice and corn, have moderate levels of production.
- Some plants have shown high variability and are thus very susceptible to climatic and seasonal changes.

4. Seasonal Trend Analysis

The data was categorized based on seasonal categories (Kharif, Rabi, Whole Year, and so on) to analyze seasonal production.

Key observations:

- Kharif Crops dominated Total Crops Production
- Rabi and Whole Year Crops were more stable with less fluctuation
- This enabled inclusion of seasonal distribution graphics on the dashboard.

5. Rainfall and Climate Relationship Analysis

- EDA assessed changes in rainfall level per state and compared them with any possible correlations with production.
- Insights included:
- The regions with more rainfall normally tended to have greater production.
- Some of these states had lower production levels, yet they experienced high rainfall. It created a need for more analysis
- Moreover, temperature suitability was analyzed against crop types to check for compatibility.

6. Annual Crop Trend Analysis

Historical crop numbers and production information were assessed for trends.

Findings included:

- Eminent peaks of cultivation ≈ 2002-2003
- There were some years with troughs, perhaps because of regional climate changes and government regulations.
- Trend analysis justified inclusion of 'Crops in a Year' visualization.

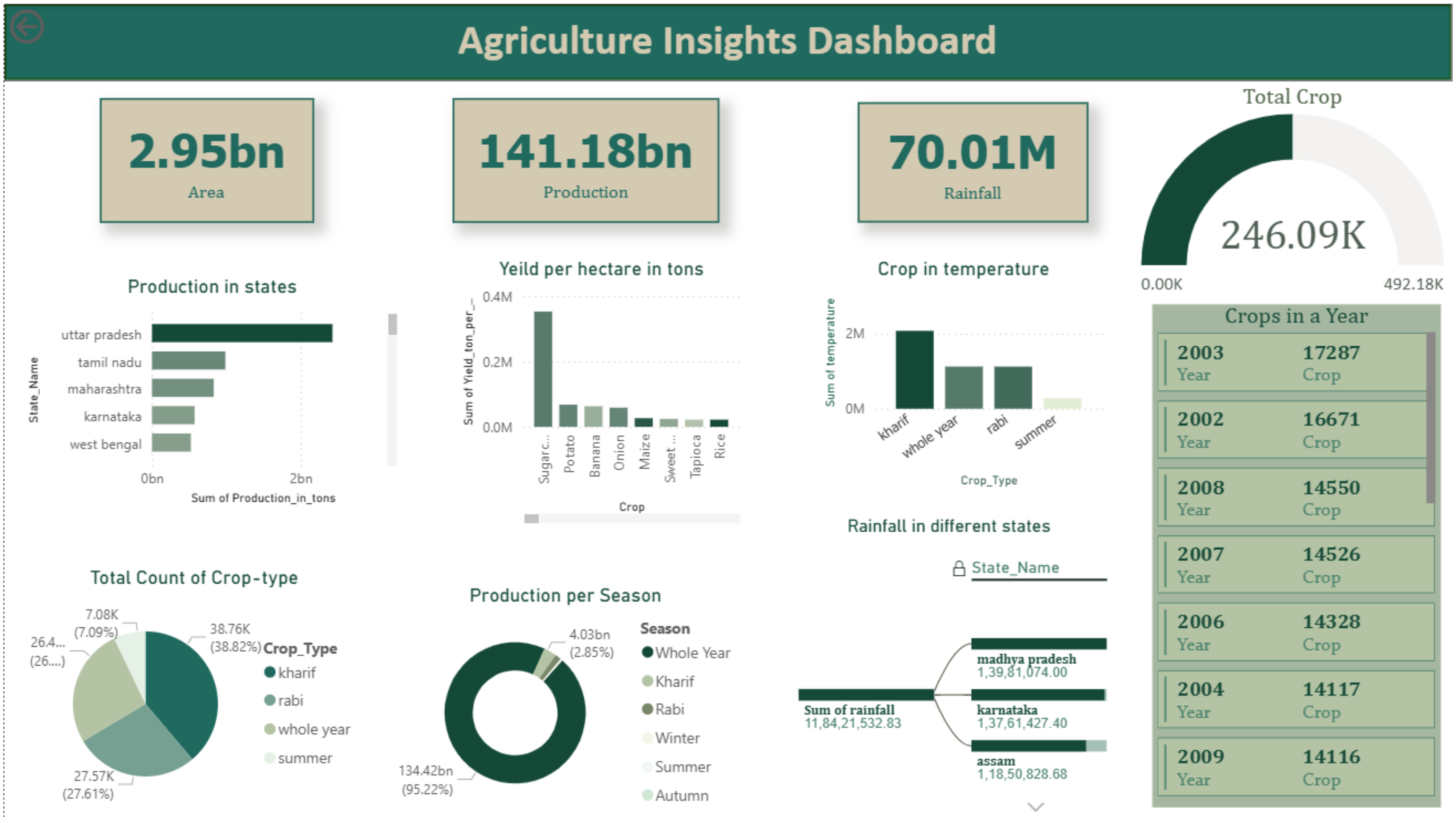
7. Insights to Guide Visualization

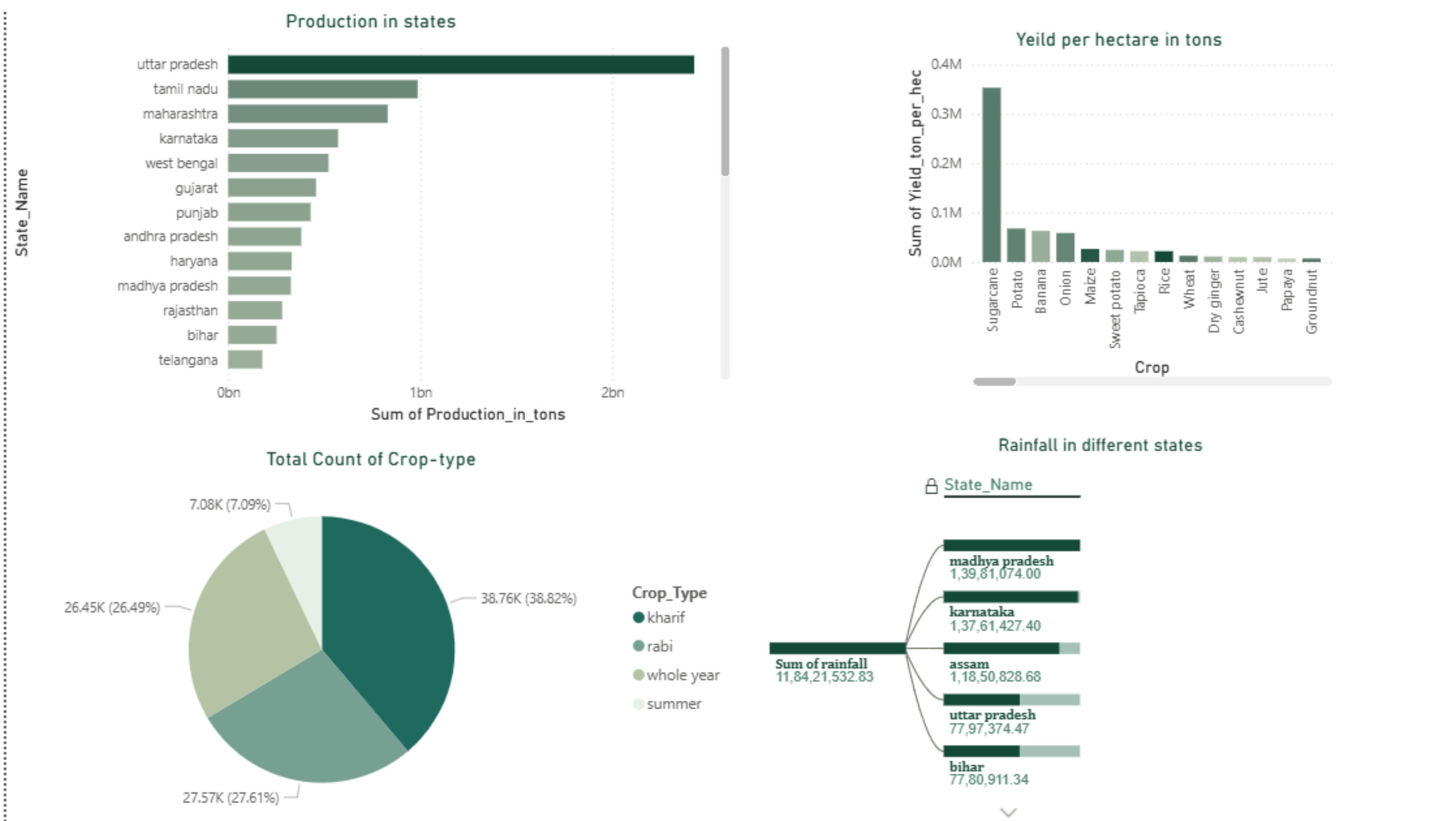
The result of EDA directly impacted the design of the dashboard because it aided in selecting:

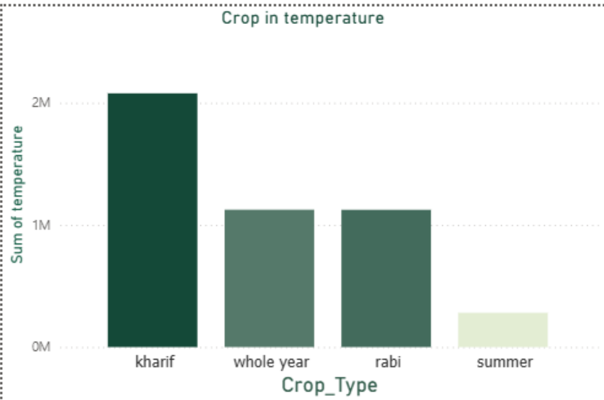
- Features to display on Which KPIs (area, production, rainfall) should be shown
- Which charts were best suited for representing state-wise and crop-wise disparities

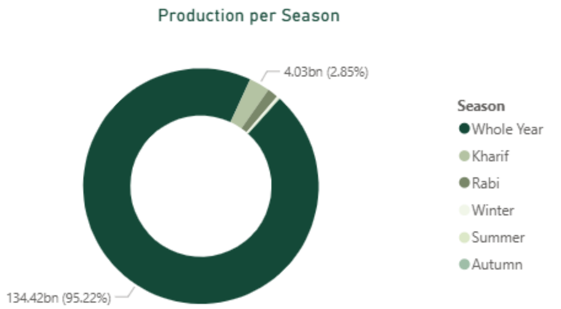
- Representing Season and Climate Relationships

8.Visualizations









Crops in a Year

Year	Crop
2003	17287
2002	16671
2008	14550
2007	14526
2006	14328
2004	14117
2009	14116
2011	14071
2010	14065
2005	13799
2000	13658
2013	13650

9. Insights

The Agriculture Insights Dashboard allows visualization and consideration of production, yield, rainfall, and temperature variables affecting agriculture in India. The insights obtained from these interactive dashboards and analysis include:

1. Uttar Pradesh Dominates Crop Production

It can be seen that among all states, Uttar Pradesh itself accounts for the maximum production, followed by Tamil Nadu, Maharashtra, and then Karnataka. It shows that there are very good agriculture capabilities and supportive climate/region factors.

2. Sugarcane Exhibits Unusually High Yield

Yield analysis shows that Sugarcane boasts the maximum yield per hectare and outshines major crops like rice, corn, and bajra. The result highlights the high productivity and contribution made by sugarcane in agriculture.

3. Kharif Season Dominates Total Production

According to seasonal analysis, Kharif contributes more towards total production due to rainfall and a large area. Rabi and Whole Year will be more stable but will have lower production.

4. Rainfall affects production significantly at the state level.

Those with greater annual rainfall, like Madhya Pradesh and Karnataka, are likely to display greater production levels because rainfall plays a very vital role in crop production. Arid regions have low rainfall and, therefore, lower production.

5. Temperature Suitability Varies Depending on Crop Type

- Kharif season crops prefer warmer temperatures
- Rabi crops are grown at low temperatures
- Full-year plants need Moderate and Constant Temperature

It shows an understanding about the role of climate-appreciative agriculture planning.

6. Number of Harvests Varies Annually

Looking at historical data, there have been some fluctuations in the number of crops, with maximum farming activities taking place around some specific years (for instance, 2002-2003).

7. Whole-Year Crops Offer Production Stability

Although Whole-Year Crops may not be among the biggest producers, they are usually responsible for stable production regardless of the season.

8. Production Distribution Shows High Regional Imbalance

The gap that emerged between high and low food-producing states suggests regional inequality in food production. It leads to an indication that there might be room for improvement with resource allocation and climate-resistant food production.

10. Conclusion

The Agriculture Insights Dashboard project clearly illustrates that data that might be fragmented and unstructured within agriculture can be given useful and meaningful interpretation via data analytics and visualization tools. The project brings together various sets of data, including crop production, harvest per hectare, rainfall distribution, and temperature suitability.

A dashboard for Power BI, created within this research work, will enable dynamic analysis and exploration of agricultural phenomena across different states and seasonal changes. The dashboard will be useful for insights related to understanding successful states, seasonality changes in production, understanding the effects of climate on production rates, and understanding production efficiency for different types of agriculture. All these insights emphasize a critical relationship between agriculture production and factors like rainfall and climatic changes.

By applying data cleansing and processing, exploration, and visualization, it can be understood that structured data processing plays an integral role in making data more relevant and accurate. It can act as an informing tool with potential benefits for farmers, policymakers, researchers, and various stakeholders.

On a concluding note, it can be seen that this project highlights not only the importance of data analytics tools in agriculture but also showcases why tools such as Power BI are useful in unscrambling and then developing agriculture.