Agriculture Crop Production Analysis

Submitted By

Keshav Yadav

(Reg. No. 12320523)

(Roll No. 02)

(Section: RK23TGA02)

Under the supervision of

Madhav Dubey

COMPUTER SCIENCE AND ENGINEERING

Lovely Professional University

Jalandhar – 144411

Acknowledgment

I would like to thank my mentor and teacher, Madhav Dubey for his constructive guidance, support, and encouragement during the course of completing this project. His valuable suggestions and experience have played a significant role in the success of this work.

I extend my sincere gratitude to the Computer Science and Engineering Department of Lovely Professional University for the required resources and learning environment.

In addition, I thank my friends for their continued encouragement, tolerance, and motivation, which never let me lose hope throughout the process.

Finally, I thank all authors and researchers whose studies have inspired me to acquire knowledge in the domain of Agriculture Crop Production Analysis.

Keshav Yadav

Abstract

Agriculture remains the backbone of India's economy, drastically contributing to the national GDP and employing many of the citizens in the process. However, issues such as climate change, soil degradation, and poor resource utilization affect crop production. This project, Agriculture Crop Production Analysis, aims to analyze historical agricultural data to identify trends, optimize farming practices, and improve productivity.

By doing data science techniques, the project will provide insights into crop yield patterns, resource allocation, and seasonal variations. This analysis will help in better decision-making for farmers, policymakers, and stakeholders in the agricultural sector. The study also explores predictive modeling techniques to forecast crop production, enabling proactive measures to mitigate risks and improve food security.

The findings of this project can contribute to a more sustainable and data-driven approach to agriculture, ensuring higher yields, efficient resource management, and a stable agricultural economy.

Keshav Yadav

Table of Contents

Chapter 1

Introduction

India is a global agricultural powerhouse, more than two-third population i.e 60 percent of population is dependent on agriculture. Agriculture contributes to the economy of the country and employs a large number of people. With the different climate and soil of the country, India provides a number of varieties of crops such as rice, wheat, pulses, and cash crops such as sugarcane and cotton and many more.

Analyzing trends in crop production is important for increasing agricultural practices, productivity, and ensuring food security for the growing population.

This project, Agriculture Crop Production Analysis in India, will make us provide a deep knowledge of agricultural trends, we can contribute to better planning, efficient resource allocation, and improved productivity in Indian agriculture.

1.1 How I Decided This Topic

In discovering different domains where data science could be meaningfully applied, I started by investigating fields that are both impactful and data-intensive. I narrowed down some fields such as healthcare, education, and agriculture. Having scanned through available public datasets, I observed that agricultural data—particularly on crop yields—were rich in details, consistent, and available at the district level. This presented a wonderful chance for thorough analysis.

I then thought about issues where insights derived from data could be useful. Crop farming was one area because it's subject to various quantifiable influences like electricity consumption, fertilizer application, and climatic variables. It was an area where statistical analysis and machine learning methods could be used practically.

After consulting with mentors and going through previous academic projects, I settled on this topic since it provided both practical applicability and analytical richness, making it appropriate for the application of a complete set of data science methods such as regression, time series analysis, and EDA.

1.2 Why This Topic (Motivation)

I was motivated to choose this subject as agriculture is a major industry in the Indian economy and directly influences the lives of millions. Information on the determinants of crop output can aid in promoting agricultural planning, distribution of resources, and yield forecasting. Being a student of data science, I saw an opportunity to apply my analytical skills to real-world problem with social and economic importance. The availability of large district-level data also encouraged me to explore this topic further, knowing that meaningful conclusions could be drawn to guide data-informed agricultural decisions.

1.3 Where I Got the Idea

I conceived this project while researching data science applications in the real world across industries such as health, education, and agriculture. In the process of research, I stumbled upon publicly available agricultural data sets made available by the Government of India and research organizations. These data sets, particularly at the district level, had a high potential for useful analysis. Also, conversations with lecturers and reading past research projects emphasized data-informed decision-making in agriculture. This encouraged me to select crop production analysis as a real-life and significant topic for my project.

1.4 Goals of the Study

- Data Analysis and Exploration - Perform exploratory data analysis (EDA) to discover patterns and trends..
- Optimized resource allocation - Leverage insights from data to optimize resource use, promoting sustainable farming practices.
- Policy recommendations - Offer data-driven recom-

mendations to policy makers for better agricultural planning and decision making.

1.5 Problems to be Solved :

1. **1. Low Agricultural Productivity**
   - Many regions in India experience low crop yields due to inefficient use of fertilizers, water, and electricity.
   - Solution: By analyzing trends and factors affecting productivity, you can identify best practices for improving crop yields.
2. **2. Unequal Resource Allocation**
   - Farmers in some areas overuse fertilizers and electricity, while others face shortages.
   - Solution: Your analysis can help policymakers distribute resources more efficiently based on district-wise needs.
3. **3. Impact of Climate and Seasonal Variations**
   - Farmers struggle with unpredictable weather patterns, affecting crop growth.
   - Solution: Time series analysis can help predict seasonal trends and suggest optimal planting times.
4. **4. Food Security Concerns**
   - Rising population demands better agricultural planning to prevent food shortages.

- Solution: Forecasting crop production helps in strategic planning to ensure food security.

5. **5. Financial Losses Due to Poor Planning**
    - Farmers often suffer losses due to lack of market insights and inefficient crop selection.
    - Solution: Clustering techniques can help categorize high-performing crops and suggest profitable crops based on past data.

Chapter 2

Dataset Description

I have collected data from many sources, which include government websites for getting accurate data. I have collected data from UBAg (Unified Portal of Agricultural Statistics), Open Government Data (ODG) Platform India, Department of Agriculture and Farmers Welfare. These datasets include information about crop production, electricity consumption in agriculture, fertilizer usage and district-level agriculture trend.

2.1 Data Collection

For this project, I have collected data from many government websites:

- UPAg (Unified Portal of Agricultural Statistics), a government-backed initiative providing agricultural information.
- Data.gov.in – A central repository of government datasets, including agricultural statistics.
- Department of Agriculture and Farmers Welfare – Official data on crop yields, land use, and climatic conditions.

2.2 Challenges Faced While Collecting Data

- **Data Inconsistency Across Sources**: Different government agencies used to present the same variables (for example, crop yield or fertilizer use) differently or in different units, and therefore all datasets had to be standardized with great care so that everything remains consistent.
- **Missing or Incomplete Data**: For some districts or seasons, information like electricity consumption was not available.
- **Unstructured or Scattered Data**: The information was in a disjointed form over multiple sources, including PDFs, Excel documents, and separate web pages. It took long and was tiresome to merge and collate this unorganized data into one single and workable format.
- **Lack of Real-Time or Recent Data**: Statistics for a few of the most recent years had not yet been published or were behind schedule, limiting the depth of the time-series analysis. This left it short of important current information for the project.
- **Manual Cleaning Effort**: Due to inconsistent column names, misspelling district names, and differences in data formatting, significant time was spent on data cleaning and pre-processing. Ensuring uniformity and correcting errors were essential steps in maintaining data integrity.

2.3 Reason for Approval of the Project

The crop production analysis project was authorized because it would be relevant in solving some of the major problems facing India's agricultural industry. Through the investigation of the nexus between crop yield, electricity use, and fertilizers, the project offers worthwhile insights into forces shaping agricultural output. The application of real-world datasets and implementation of data science methodologies, including exploratory data analysis, regression, and time-series analysis, makes the project both academically challenging and relevant in practice. Also, the multi-dimensionality of the data, along with meticulous data cleaning and analysis, provides the possibility of significant recommendations that may have the power to influence agricultural practice and policy. This turns the project into a worthwhile and informative contribution to agricultural research.

2.4 Load Forecasting in Agriculture

Load forecasting involves estimating the future demand for agricultural practices:

- Water usage for irrigation.
- Fertilizer and pesticide consumption.
- Electricity demand for farming equipment.
- Crop yield expectations based on weather and soil conditions.

2.5 Datasets Used

a) Crop Production Data (District-wise) :

- Source: UBAg (Unified Portal of Agricultural Statistics).
- Description: Contains information about the production quantity, area, season, and crop type across various districts in India.

b) Electricity Consumption in Agriculture :

- Source: Department of Agriculture and Farmers Welfare.
- Description: Electricity usage details in the agricultural sector, helping to understand energy consumption in farming.

c) Fertilizer Usage in Agriculture Sector :

- Source: Open Government Data (ODG) Platform India (Data.gov.in).
- Description: Types and amounts of fertilizers used in different regions and soil fertility and crop yield.

d) Rewari District Agricultural Data :

- Source: UBAg (Unified Portal of Agricultural Statistics).

- Description: Extracted from the Crop Production Data (District-wise) data.

2.6 Description of Columns in Each Dataset

a) Crop Production Data :

- State – The state where the crops were cultivated.
- District – The district where the data was recorded.
- Crop – The name of the crop (e.g., wheat, rice, sugarcane).
- Season – The farming season (Kharif, Rabi, Zaid).
- Area (Hectares) – The total land area used for cultivation.
- Production (Tonnes) – The total crop production in tonnes.
- Yield (Kg/Hectare) – Crop yield per unit area.

b) Electricity Consumption Data :

- Year – The year of data collection.
- State – The state where electricity was used.
- Electricity Consumption (GWh) – The total electricity consumption in agriculture.

c) Fertilizer Usage Data :

- State – The state where fertilizers were applied.
- Year – The year of data collection.
- NPK (Tonnes) – The amount of nitrogen, phosphorus, and potassium (NPK) fertilizers used.

d) Rewari District Data :

- Year – The year of data collection.
- Crop – The type of crop grown in Rewari.
- Production (Tonnes) – The total production for that crop in the district.
- Area (Hectares) – The total land area used for cultivation.
- Season – The farming season (Kharif, Rabi, Zaid).

Chapter 3

Statistical and Predictive Analysis of Crop Production Data

In this chapter, I am going to do different analysis on the crop production data using statical and predictive analysis. I will begin with a descriptive analysis that analyze the spreadness of the data, central tendency, and skewness of the data. Then I will apply regression analysis to show the relationship between crop production that influences factors such as fertilizer usage and electricity consumption. And then we will do time series analysis in which I am going to show the seasonal patterns of the crop. And at last I will do Hypothesis Testing in which I will do T-Test to see the usage of fertilizers in crop production.

3.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an essential data analysis step whose objective is to outline the primary characteristics of the data through visualizations as appropriate. EDA assists us to view patterns and relationships in the data before official statistical modeling.

- **Data Cleaning** - Dealing with missing values, removing duplicates, and fixing inconsistencies.
- **Visualizations** - Using charts to understand the relationships and distribution of the data.
- **Summary Statistics** - Reviewing such as mean, median, mode, variance, and standard deviation.
- **Correlation Analysis** - Establishing dependencies between different variables.

3.1.1 Different Types of Charts and Their Uses

- **Histogram** - Used to show the distribution of a single variable.
- **Box Plot** - To determine outliers and spread of distribution.
- **Bar Chart** - To compare various categories.
- **Pie Chart** - Shows proportions or percentages.
- **Scatter Plot** - Shows relationships between two continuous variables.
- **Line Chart** - Best for observing trends over time.
- **Heat Map** - Used to visualize correlations between multiple variables.

3.2 Descriptive Analytics

Descriptive Analytics will give key characteristics of the data. It includes spr

3.2.1 Spread of Data

In the Spread of Data we will see Range, Variance, Standard Devision, and Interquartile Range (IQR).

- **Range** - $\max(X) - \min(X)$
- **Variance** - $\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N}$
- **Standard Deviation** - $\sigma = \sqrt{\sigma^2}$
- **Interquartile Range (IQR)** - $Q_3 - Q_1$

3.2.2 Central Tendency

Central Tendency measures center of the dataset:

- **Mean** - The average of all data points.
- **Median** - The middle value of the data.
- **Mode** - The most frequently occurring value.

3.2.3 Skewness of Data

Skewness measures the asymmetry of the data distribution:

- **Positive Skew (Right-Skewed):** - The right tail is longer, showing higher values have greater variance.
- **Negative Skew (Left-Skewed)** - The left tail is longer, showing lower values are more spread out.
- **Symmetric Data** - The distribution is approximately equal on both sides of the mean.

3.3 Regression Analysis

Regression analysis is technique to find the relationship between crop production and some other factors such as fertilizer usage, electricity consumption and rainfall. This will help to analyze future crops.

3.3.1 Types of Regression

- **Simple Linear Regression** - Examines the relationship between crop production and a single independent variable (e.g., fertilizer usage).
- **Multiple Linear Regression** - Includes multiple independent variables to improve prediction accuracy.

Chapter 4

Results and Interpretations

4.1 Distribution of cultivation area

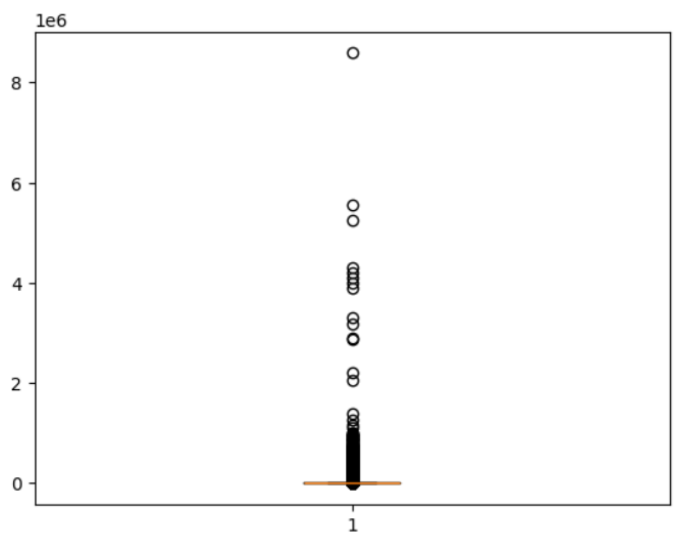

Fig. 1

Insights:

- Most values are very close to zero, showing that a large number of records have small cultivation areas.
- Some entries have very large land areas.
- The data is highly right-skewed, indicating imbalance in land distribution.

- This high number of outliers suggests skewed data — maybe a few districts have much larger land under cultivation.

4.2 Distribution of cultivation area

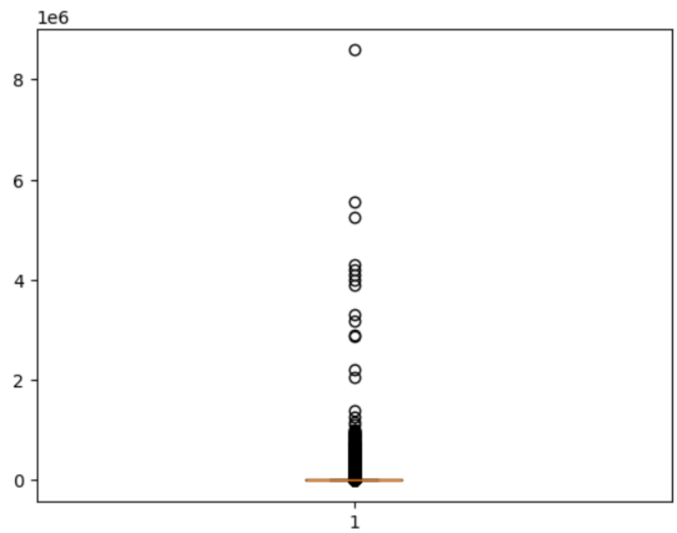

Fig. 2

Insights:

- Many values are low, but some have extremely high production.
- More outliers than Area — showing production varies more dramatically across districts/crops.

4.3 State wise Production

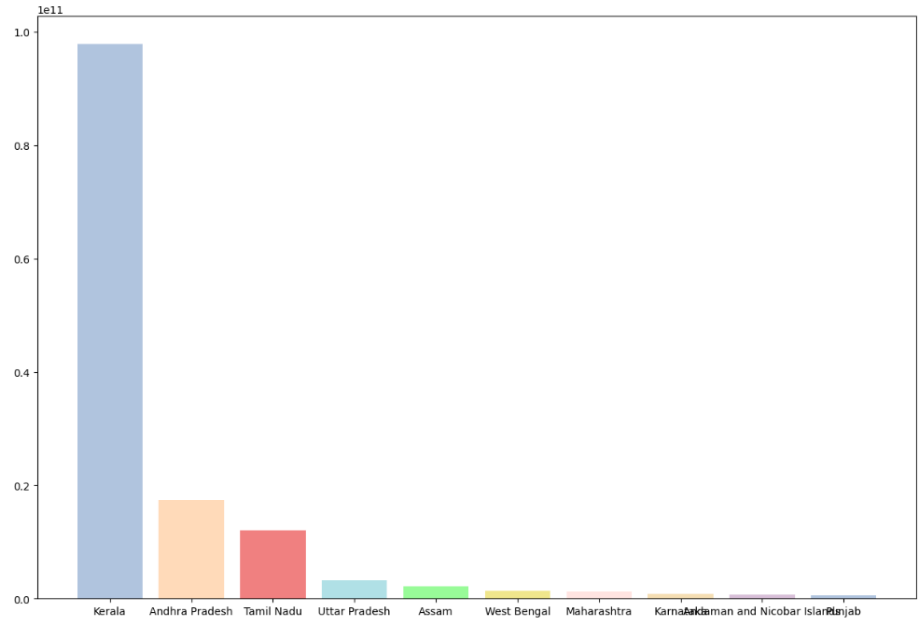

Fig. 3

Insights:

- Each bar is representing a state.
- The length of the bar shows the total crop production.
- The highest production state wise is Kerala.

4.4 Year wise Production

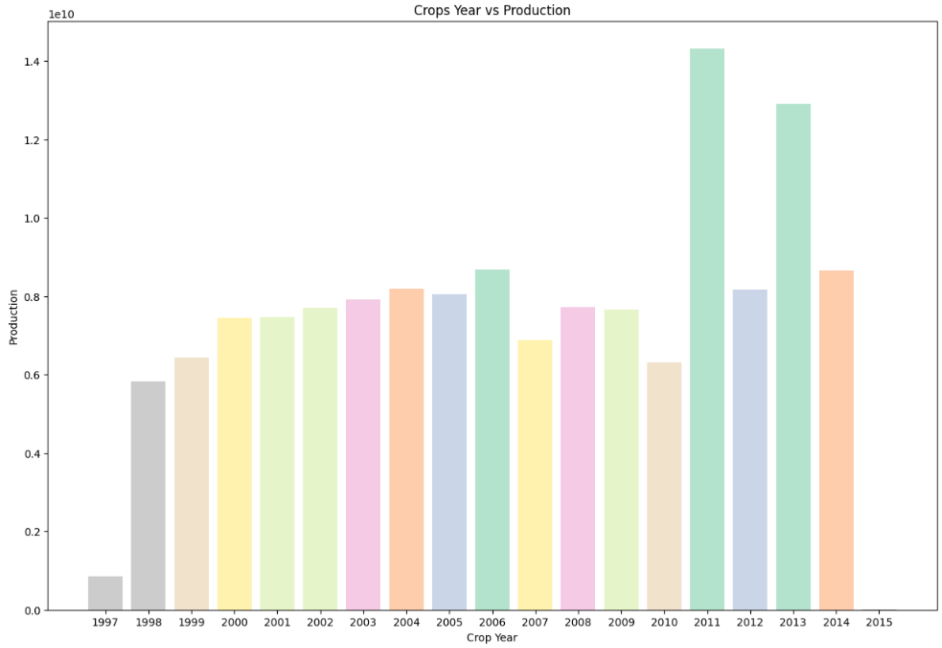

Fig. 4

Insights:

- Height of each bar represents the total crop production in that year
- 2011 saw the highest crop production.
- 2015 saws the lowest crop production may be due to missing or inconsistent data.
- 1997 One of the lowest full data years and low crop production year.

4.5 Season wise Production

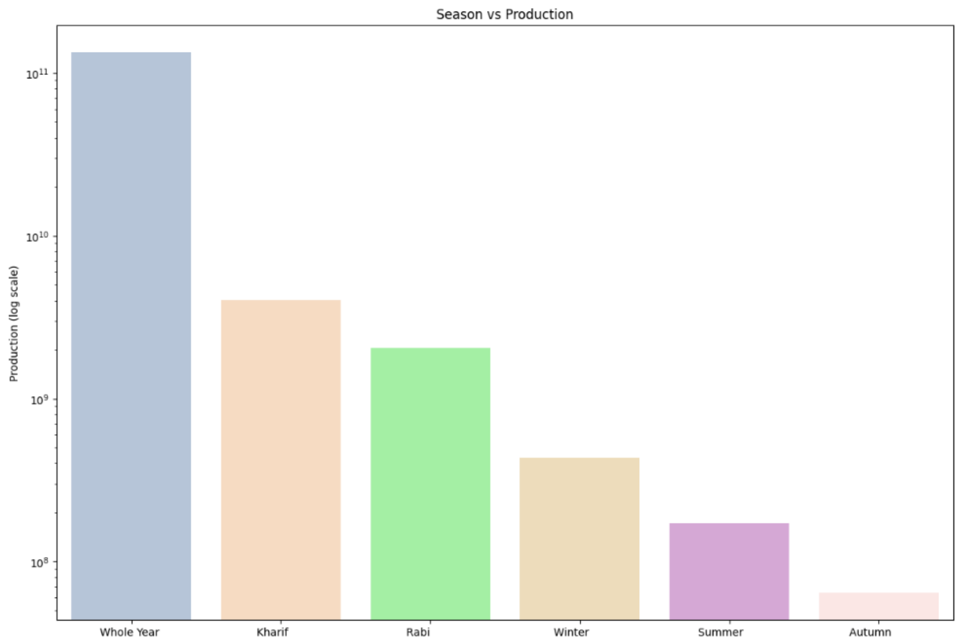

Fig. 5

Insights:

- Top crop categories which shows high production values are Whole Year(Annual growing plants), Kharif and Rabi crops.
- These crop are generally dependent on monsoons.

4.6 Category wise Production

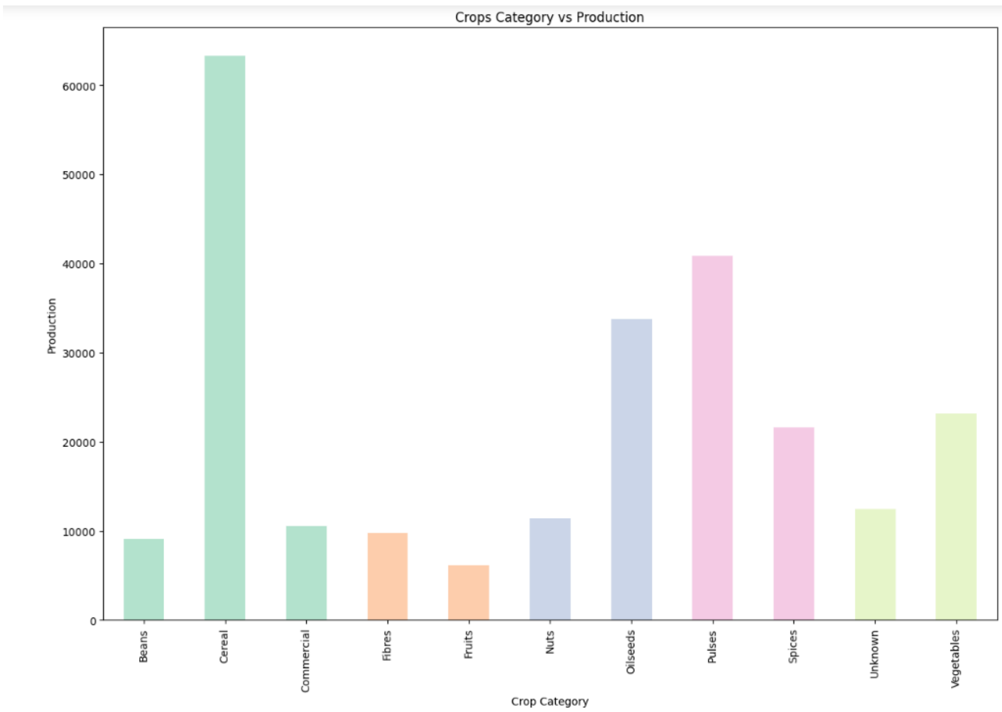

Fig. 6

Insights:

- This graph is showing the the production of the crop category.
- In this graph the highest production crop category is Cereal followed by Pulses.

4.7 Different proportion of Crop Categories for India

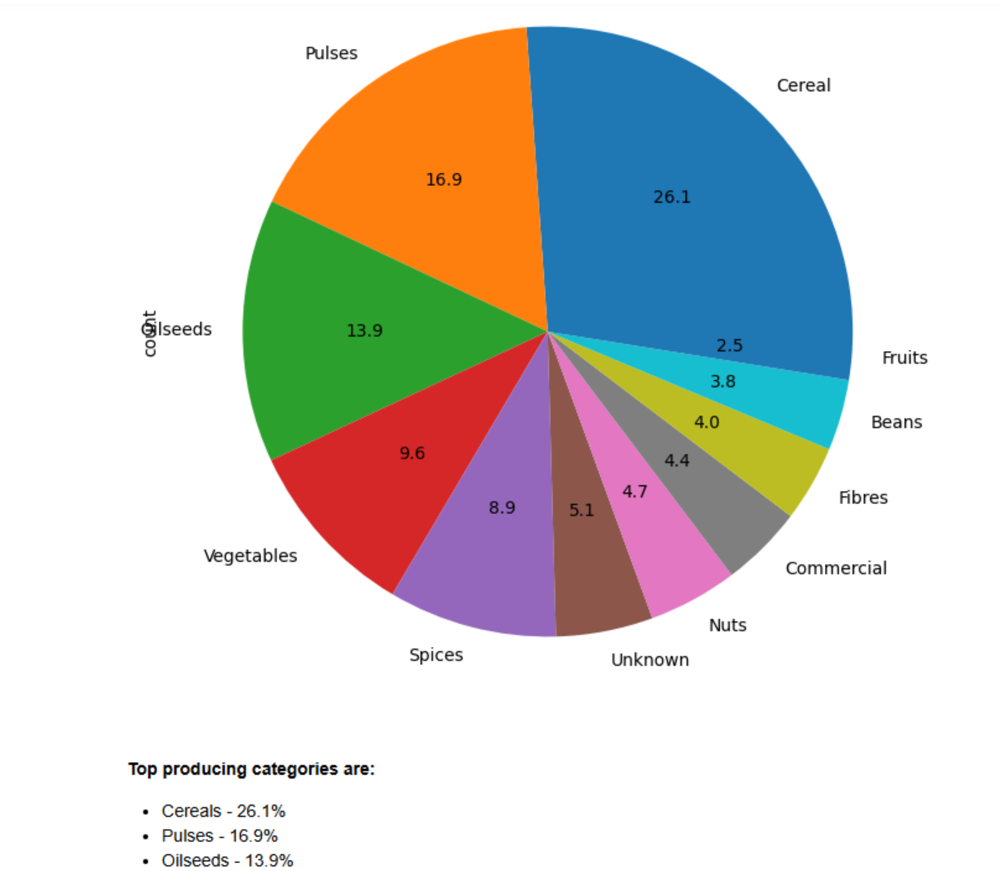

Fig. 7

Insights:

- The top three (Cereals, Pulses, Oilseeds) make up 56.9 percent of total production — over half.
- Top producing categories are:
- Cereals - 26.1
- Pulses - 16.9

- Oilseeds - 13.9

4.8 State wise crop production with different categories of crops

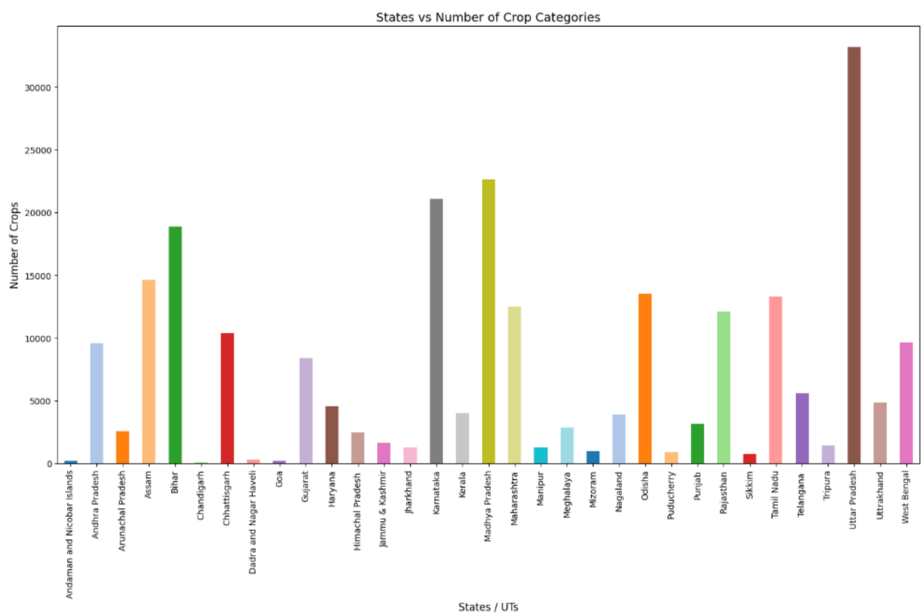

Fig. 8

Insights:

- Uttar Pradesh leads by a large margin — indicating high crop diversity.
- These states are likely to have:
- Higher agricultural activity,
- Greater land under cultivation

4.9 Boxplot for Outliers for Electricity Consumption

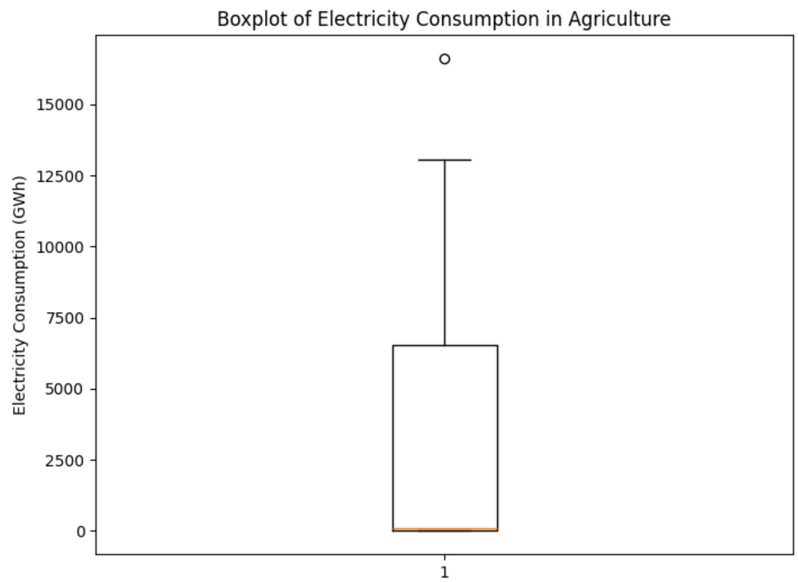

Fig. 9

Insights:

- The boxplot shows that most states have low to moderate electricity consumption for agriculture.
- A few states are outliers, consuming much higher electricity (above 15,000 GWh).
- The median electricity consumption is closer to the bottom, indicating that many states consume relatively less electricity compared to the highest ones.

4.10 Total Electricity Consumption by State

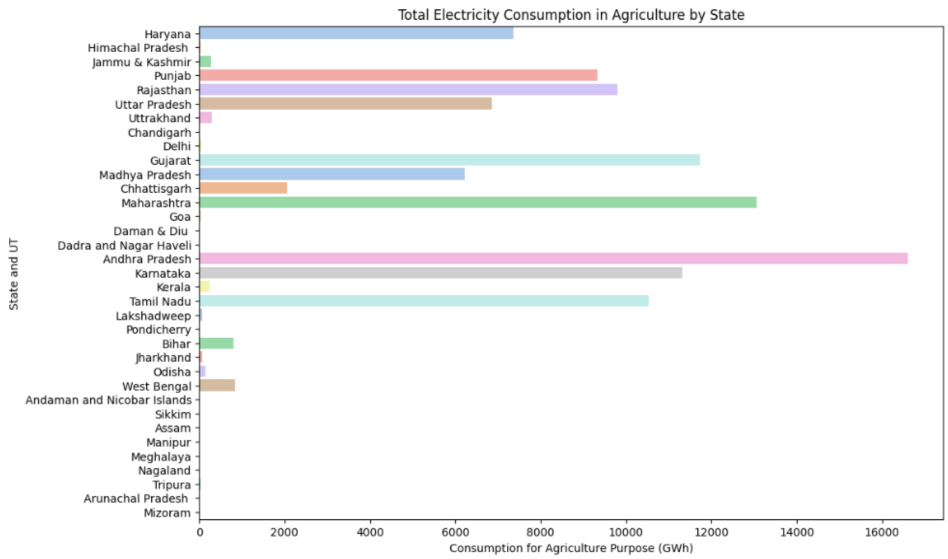

Fig. 10

Insights:

- Andhra Pradesh has the highest electricity consumption for agriculture among all states.
- Many northeastern and smaller states (like Sikkim, Mizoram, Arunachal Pradesh) have very low electricity consumption in agriculture.

4.11 Distribution of Electricity Consumption

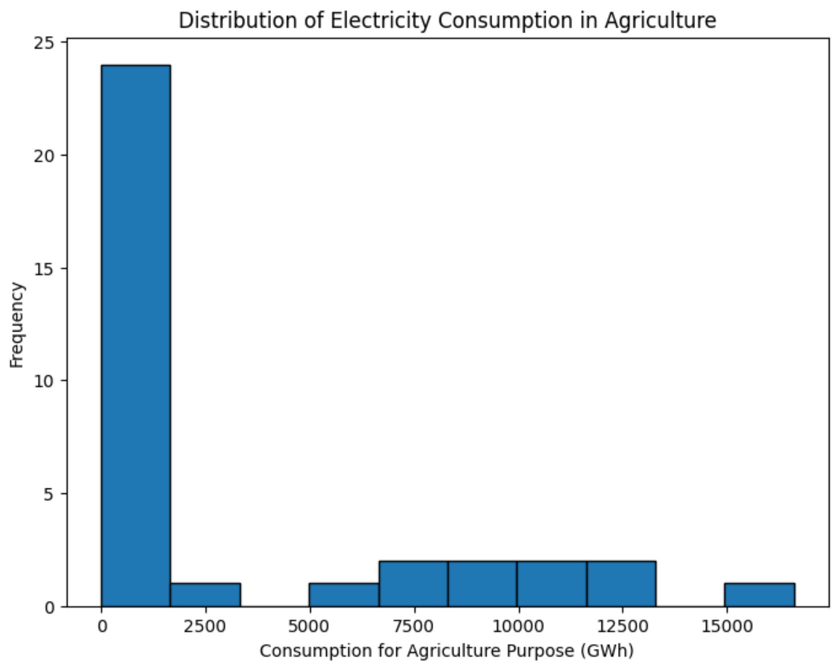

Fig. 11

Insights:

- The distribution is highly right-skewed — most states have low electricity consumption for agriculture.
- Only a few states consume very high amounts of electricity (above 10,000 GWh)
- Majority of the states have consumption below 2500 GWh.

4.12 Production vs Electricity Consumption

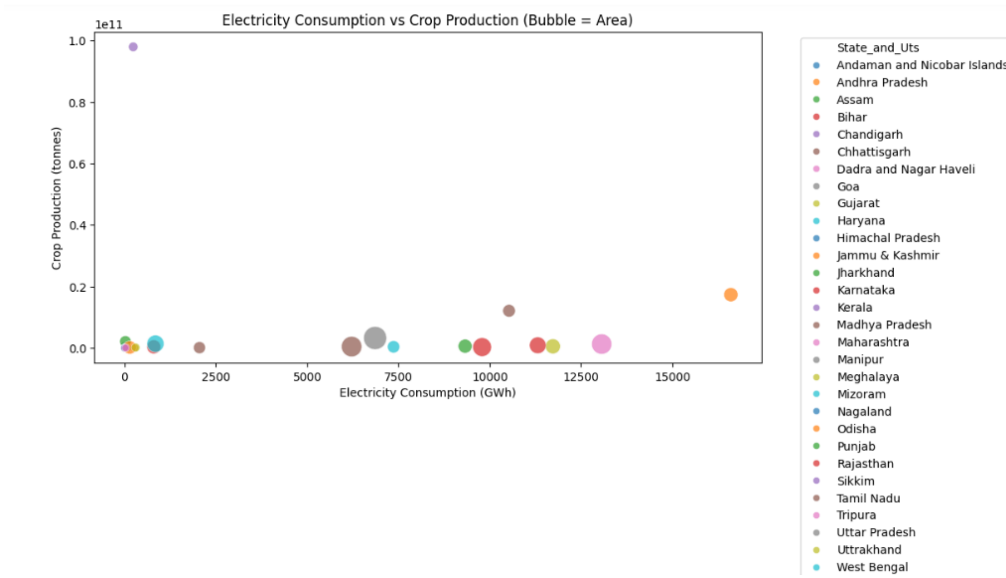

Fig. 12

Insights:

- Uttar Pradesh or West Bengal shows an extremely high crop production, even though its electricity usage is moderate.
- Some states (like Andhra Pradesh, Maharashtra, Tamil Nadu) show high electricity consumption but moderate crop production.
- Larger bubbles near the top-left indicate large agricultural area with less electricity input.

4.13 Regression Line Code

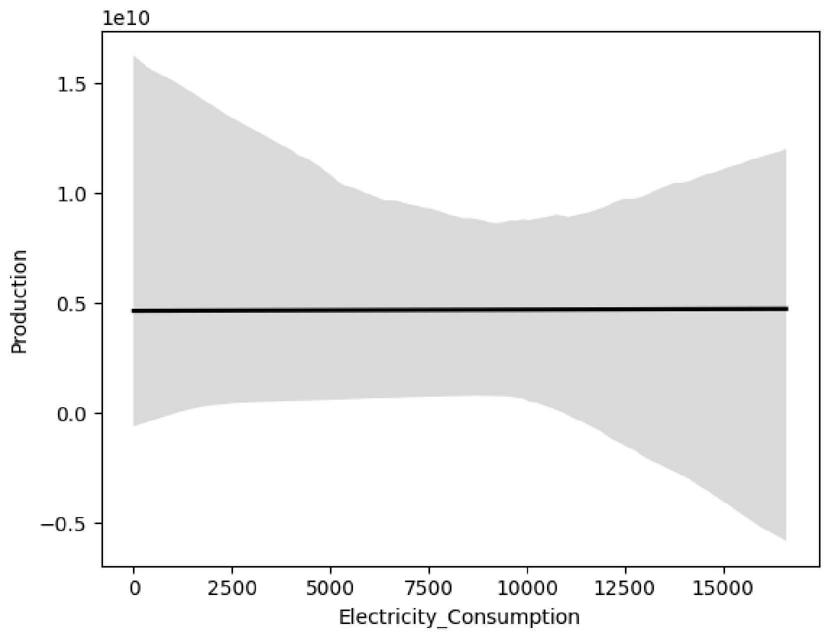

Fig. 13

Insights:

- The black line is almost horizontal, this confirms no significant linear relationship between electricity consumption and crop production.
- Wide Confidence Band, The grey shaded area is very wide, especially at the extremes. This means there is high uncertainty in predicting crop production based on electricity usage.
- Electricity consumption alone is not a strong predictor of crop production.

Chapter 5

Conclusion

This project conducted an extensive descriptive analytics analysis of crop production and electricity usage in Indian agriculture with the purpose of extracting meaningful insights from district-wise and state-wise data. The main aim was to know how agricultural productivity differs across regions and how electricity consumption is related to crop production.

Important Insights from the Crop Production Analysis:

1. **State and District-Level Variation:** The analysis showed a substantial degree of variation in crop production between various states and districts of India. Uttar Pradesh, Punjab, and Maharashtra consistently recorded higher aggregate production values, which pointed towards their dominant position in national agriculture.

2. **Seasonal and Yearly Trends:** Through examination of production season-wise and year-wise, it was seen that the Kharif season has the largest contribution to annual output. The yearly pattern revealed the steady rise in production for some periods and subsequent fluctuations owing to climate, policy, or

economic reasons.

3. **Distribution and Outliers:** Box plots and histograms indicated that cultivation area and production distributions are positively skewed. Few districts are responsible for the majority of crops, with numerous districts reporting low production. Several outliers were evident, suggesting regions with high or low productivity levels.

4. **Crop Dominance:** There are some crops like Rice, Wheat, and Sugarcane, which are being cultivated in almost every district and generate most of the production. Meanwhile, various crops have areas with niche productions. This implies there is insufficient diversification of cultivation of crops across many places.

5. **Area vs Production Correlation:** A high correlation between area and production was seen, as expected. But there were some districts that had more production with comparatively lesser cultivation area, which shows more efficiency or yield due to superior farming practices, soil quality, or irrigation.

Important Observations from Electricity Consumption Analysis:

1. **State-Wise Electricity Usage:** Agricultural electricity consumption is extremely uneven between states. Punjab, Andhra Pradesh, and Tamil Nadu had much

greater electricity consumption, most probably because of their extensive irrigation systems and reliance on electric pumps.

2. **Electricity Consumption and Agricultural Production:** Although there was no quantitative one-to-one relationship, graphical comparisons between electricity consumption and production levels indicated that greater electricity consumption tends to be linked with greater crop production, particularly in irrigation-based areas.

3. **Difficulties in Even Supply of Uniforms:** Some states that were productive had relatively moderate electricity use, reflecting perhaps efficient practices or potential constraints to electricity availability. This highlights the need for the even supply of electricity to advance agriculture in every region.

Concurrent Insights through Crop and Electricity Data:

- It was possible to see clearly that infrastructure plays a central function in optimizing farm output.
- Areas with improved electricity access had improved year-to-year consistency in production.
- Policy implications are that investment in sustainable energy sources, advanced irrigation methods, and crop diversification for balanced distribution needs to be undertaken to guarantee farmer welfare and food security.

Conclusion:

This study gives a clear understanding of how crop production and electricity use are connected in Indian farming. It not only looks at the current situation but also opens up opportunities for future research in areas like predicting crop yields and energy use using machine learning, and analyzing data based on location. The findings highlight how important it is to use data to make better decisions that can improve farming, save energy, and help create better policies. Moving forward, using these methods will be key to ensuring sustainable growth, fair resource distribution, and stronger agriculture in India.