Exploratory Data Analysis (EDA) on Indian Rice Cultivation

Introduction

This project analyzes **rice cultivation trends in India** using historical data from 1997 to 2020. The objective is to explore **seasonal variations, state-wise production trends, and the relationship between cultivation area, production, and yield** using various statistical and visualization techniques.

Data Cleaning

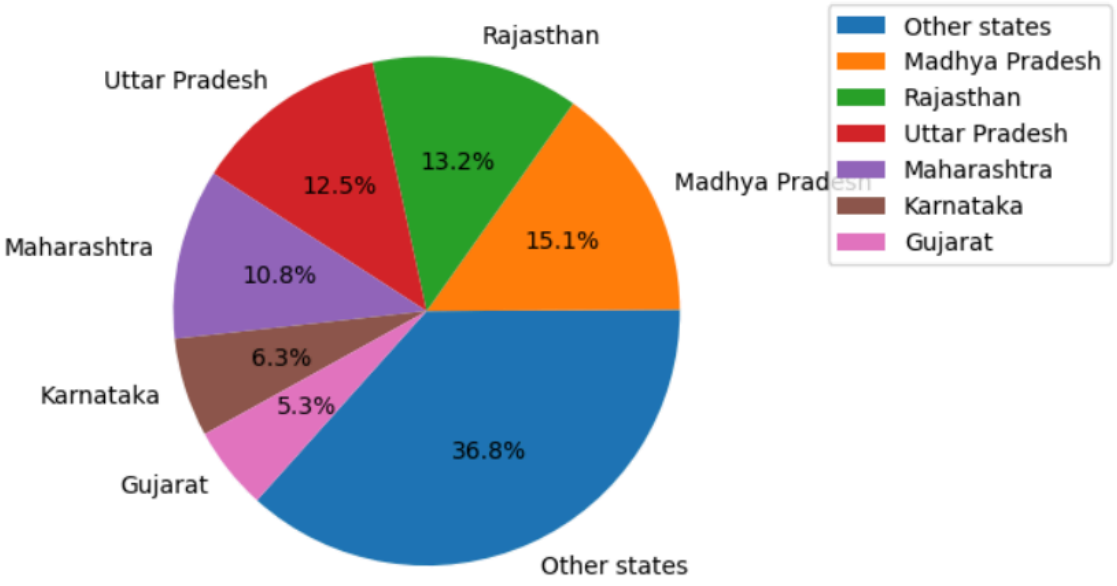
Steps Taken:

1. **Removed Year 2020-21** as it contained data from only one state, which could bias the analysis.
2. **Identified 57 unique crops** grown across India.
3. **Checked for duplicates** and found none.
4. **Removed 32 rows** with missing crop names.
5. **Filled 4,960 missing values in the 'Production' column** with the mean value of the same crop in that district and state.
6. **Final dataset is cleaned and ready for analysis.**

Data Visualization & Key Findings

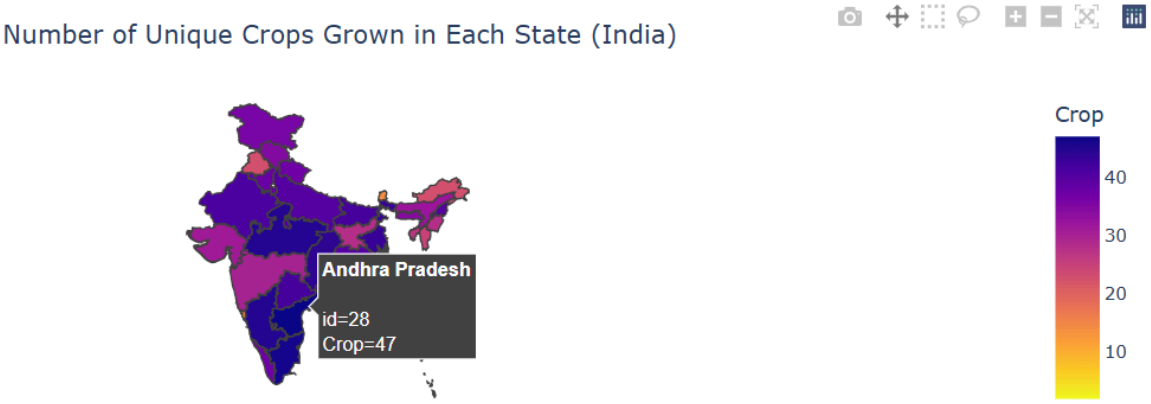
1) State-wise Cultivation Area

- **Top 5 states with the highest cultivation area:**
    - Madhya Pradesh (largest cultivation area)
    - Rajasthan
    - Uttar Pradesh
    - Maharashtra
    - Karnataka
- **Pie Chart Analysis: More than 50%** of total cultivation area is contributed by just **four states** (MP, Rajasthan, UP, Maharashtra).



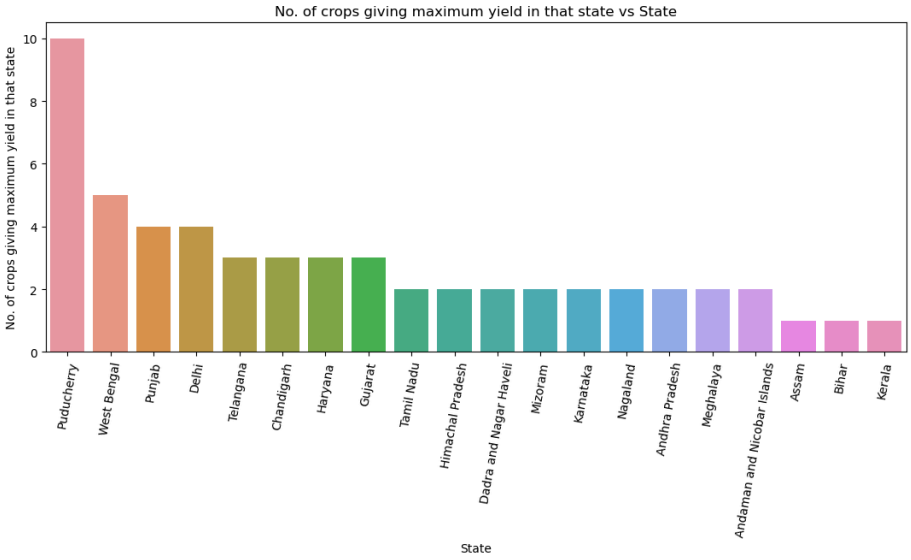
2) Crop Distribution by State

- **Andhra Pradesh grows the highest number of different crops (47),** followed by Tamil Nadu, MP, Karnataka, and West Bengal.
- **Choropleth Map Analysis:** The **southern part of India has more diverse agriculture** than northern states.

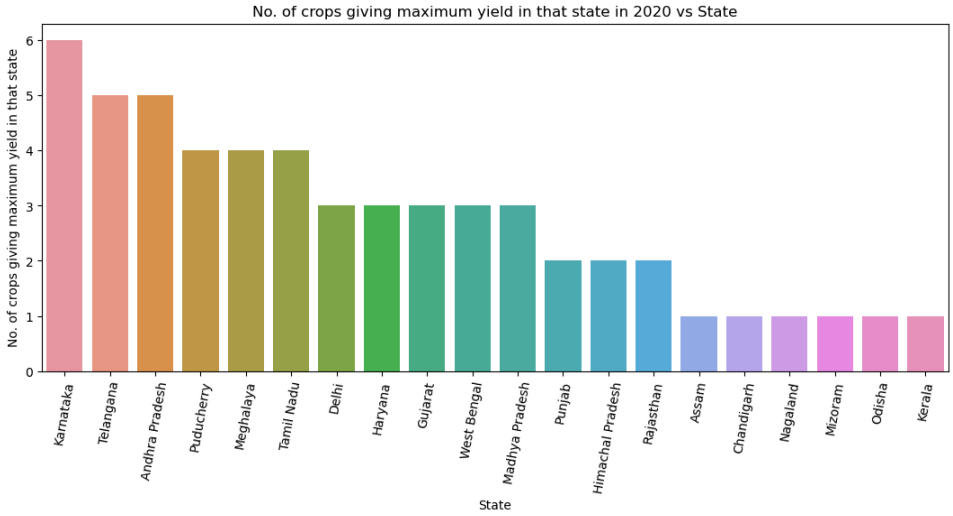


3) High-Yield Crop Analysis (1997-2020 vs. 2019-20)

- Puducherry had 10 crops with the highest yield (1997-2020).



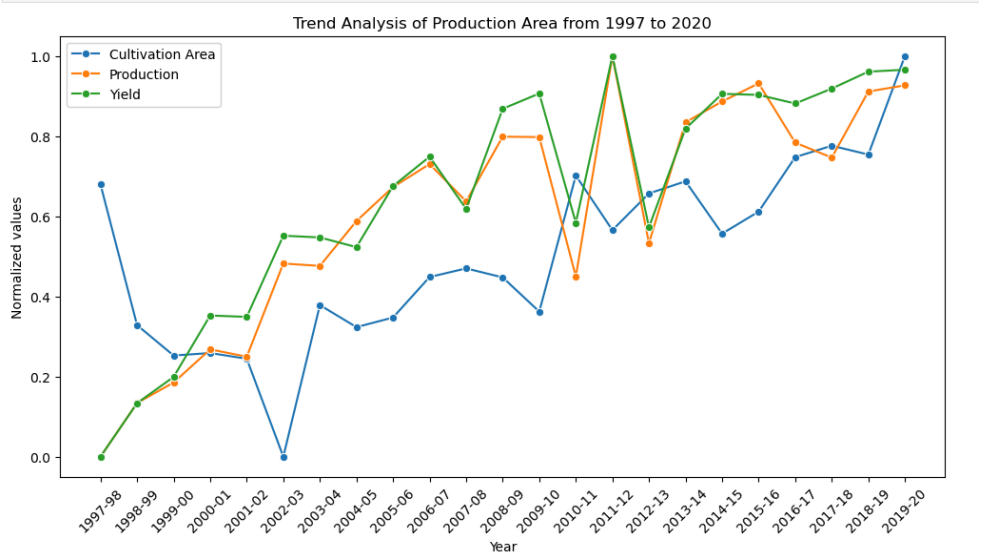
- **In 2019-20:**
    - **Karnataka (6 crops), Telangana (5 crops), and Andhra Pradesh (5 crops) emerged as top states.**



    - States like Puducherry, Punjab, and West Bengal showed a decline in high-yield crop production.

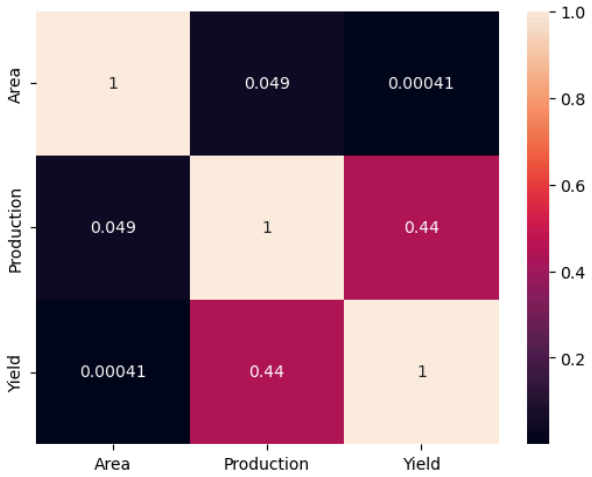
4) Trend Analysis of Area, Production, and Yield (1997-2020)

- **Key Observations from Line Plot Analysis:**
    - **2002-03:** Sudden **drop in cultivation area**, but **high production led to better yields.**
    - **2010-11 & 2012-13:** Drastic **drop in production**, leading to lower yields.
    - **2016-2020:** Crop yields remained **relatively stable**.



5) Correlation Analysis (Heatmap)

- **Correlation between Area & Yield = 0.00041** → No strong relationship.
- **Correlation between Production & Yield = 0.44** → **Moderate positive correlation**, meaning higher production tends to increase yield but is not a direct factor.

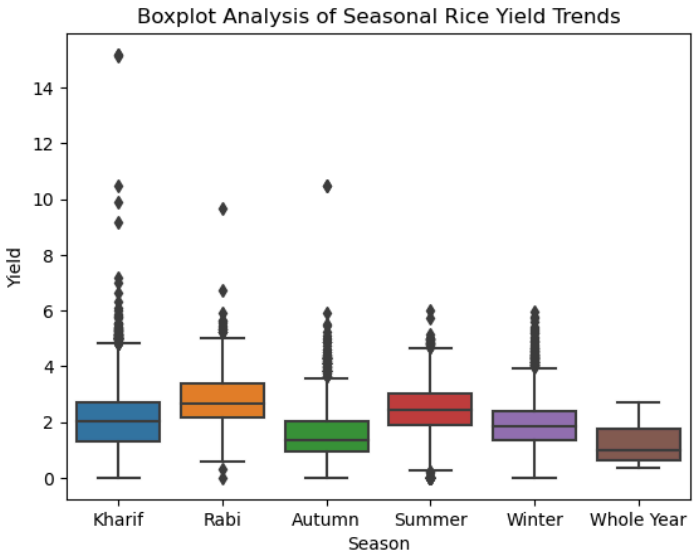


- **Insight:** Increasing the cultivated area does not necessarily improve crop yield; other factors like irrigation, soil quality, and farming techniques play a bigger role.

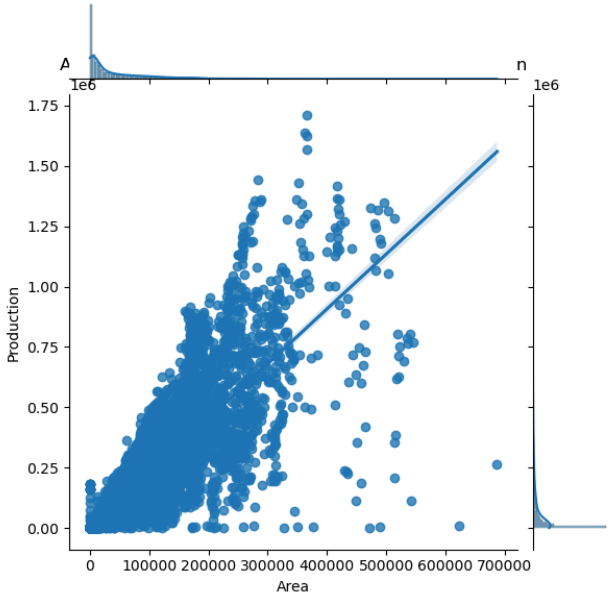
6) Crop-Specific Analysis: Rice

- **Seasonal Trends:**
  - Rice is mostly grown in the Kharif season, followed by Summer, Autumn, Winter, Rabi, and Whole Year.

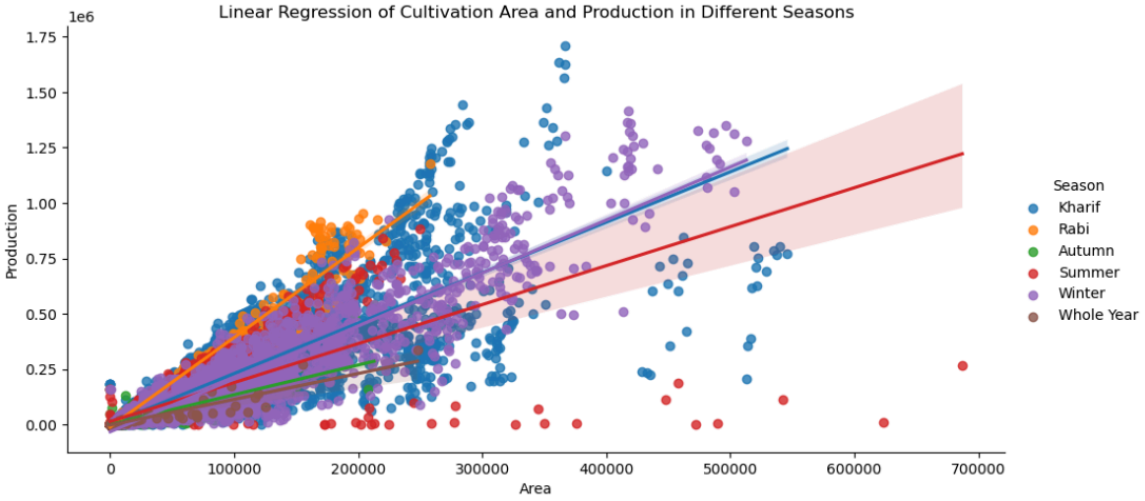
    - Rabi season had the highest average rice yield, likely due to controlled irrigation and better soil conditions.
- **State-wise Rice Yield (2019-20):**
    - **Puducherry** had the **highest** rice yield.
- **Box Plot Analysis:**
    - Yield variation is highest in the Kharif season, whereas Rabi shows the most stable yield.



- **Joint Plot (Area vs. Production):**
    - Higher area does not always mean higher production, as efficiency varies by region.



- **LM Plot (Regression Analysis by Season):**
    - **Rabi season** has the **highest regression trendline**, indicating better efficiency.



- **Strip Plot (Yearly Yield Trends):**
    - Rice yield reached its peak in some areas between 2004-07**.**

Conclusion & Key Takeaways

1. Madhya Pradesh, Rajasthan, and UP dominate the area of cultivation, but high-yield crops are found in Karnataka, Andhra Pradesh, and Telangana.
2. Southern states have diversified their crop production more than northern states.
3. Puducherry had the highest yield in 2019-20, despite being a small state.
4. Yield fluctuations were observed in 2010-11 and 2012-13 due to production drops.
5. Area expansion does not directly increase yield → Efficiency depends on irrigation, soil quality, and agricultural techniques.
6. Rabi season consistently shows higher rice yield than other seasons.
