Project Title :- Crop Production Analysis in India

Dataset :- Crop

By :- Sarthak Priyank Verma (UMIP8357)

Project Link :- Crop Production

Problem Statement :-

The Agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet. This paper presents a novel Business-to-Business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner. This dataset provides a huge amount of information on crop production in India ranging from several years. Based on the Information the ultimate goal would be to predict crop production and find important insights highlighting key indicators and metrics that influence crop production.

Expected Outcomes :-

- **Improved Understanding:**
  Gain deeper insights into the factors influencing crop production in India, including regional variations, seasonal patterns, and agronomic practices.
- **Predictive Modeling:**
  Develop predictive models to forecast crop production based on key indicators and metrics identified through data analysis.
- **Decision Support:**
  Provide decision-makers with actionable insights and evidence-based recommendations for policy formulation, resource allocation, and agricultural planning.
- **Stakeholder Collaboration:**
  Facilitate collaboration among stakeholders in the agricultural sector, including government agencies, farmers, researchers, and industry players, to promote knowledge sharing and innovation.

Targeted Variables :-

- **Crop Production:**
  The primary outcome variable representing the total quantity of crops harvested within a specified area and time frame.
- **Crop Area:**
  The land area under cultivation for various crops, which directly influences production levels and agricultural output.
- **Crop Yield:**
  The productivity of crops per unit area, calculated as the ratio of production to the cultivated area, indicating the efficiency of agricultural practices and crop management.
- **State and District:**
  Geographic units used for disaggregated analysis to identify regional disparities, hotspots of production, and areas for targeted intervention.
- **Season:**
  The time of year when crops are planted, grown, and harvested, influencing production levels and agricultural activities.
- **Crop Type:**
  The specific crops cultivated, providing insights into market demand, crop diversification opportunities, and agricultural sustainability.

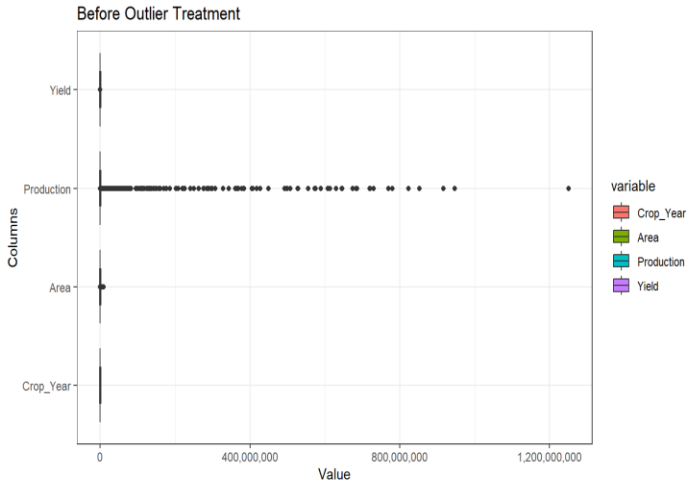
Libraries Used :-

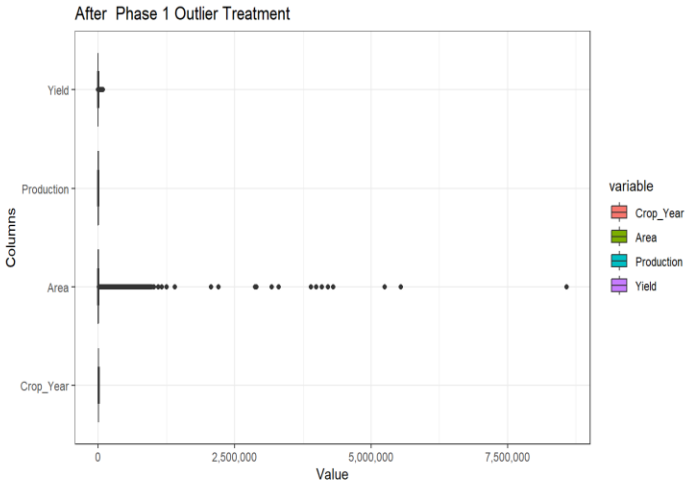
- **dplyr:** for data manipulation tasks like filtering, selecting, mutating, and summarizing data.
- **plotly:** for creating interactive plots and visualizations.
- **tidyr:** for data tidying tasks like reshaping data frames, gathering, and spreading columns.
- **ggplot2:** for creating static plots and visualizations based on the grammar of graphics.
- **reshape2:** simplifies data reshaping and manipulation tasks in R. (melt function)
- **stringr:** used for string manipulation tasks.
- **lubridate:** used for handling date-time objects and operations.
- **gridExtra:** used to make extra grids to combine the plots side by side.

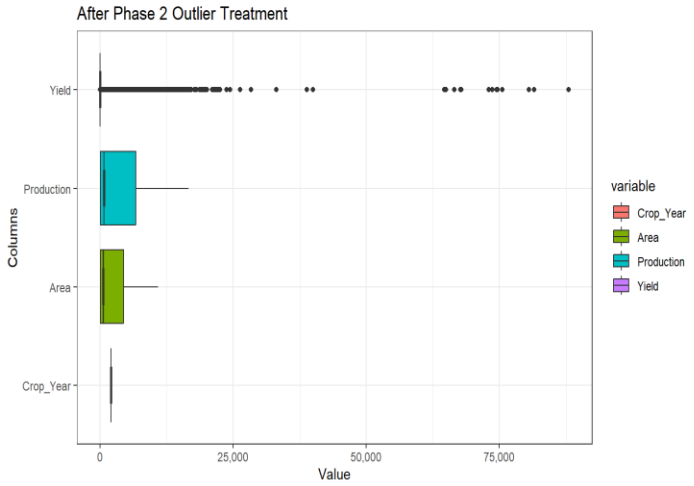
Column Description (Metadata) :-

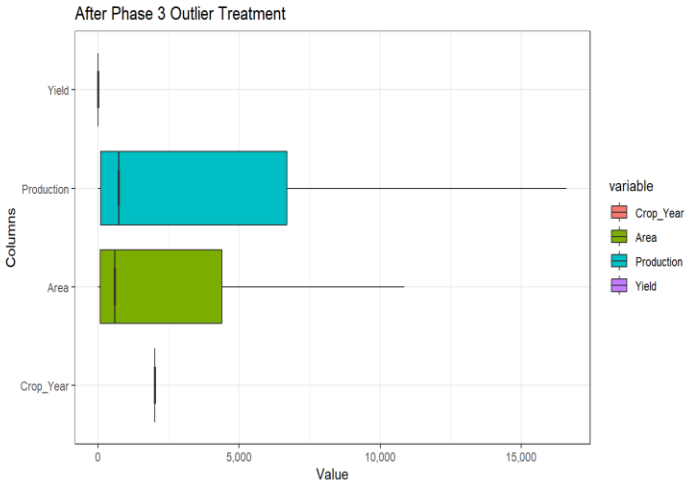
- **Crop_Year:** Year in which the crop data was recorded.
- **Area:** Area of land (in hectares) used for cultivating the crop.
- **Production:** Total production (in tonnes) of the crop.
- **Yield:** Production per unit area (tonnes per hectare).
- **State_Name:** Name of the state where the crops are grown.
- **District_Name:** Name of the district within the state where the crops are grown.
- **Season:** Season during which the crop is grown (e.g., Kharif, Rabi, Summer).
- **Crop:** Name of the crop grown.

TREATING THE OUTLIER

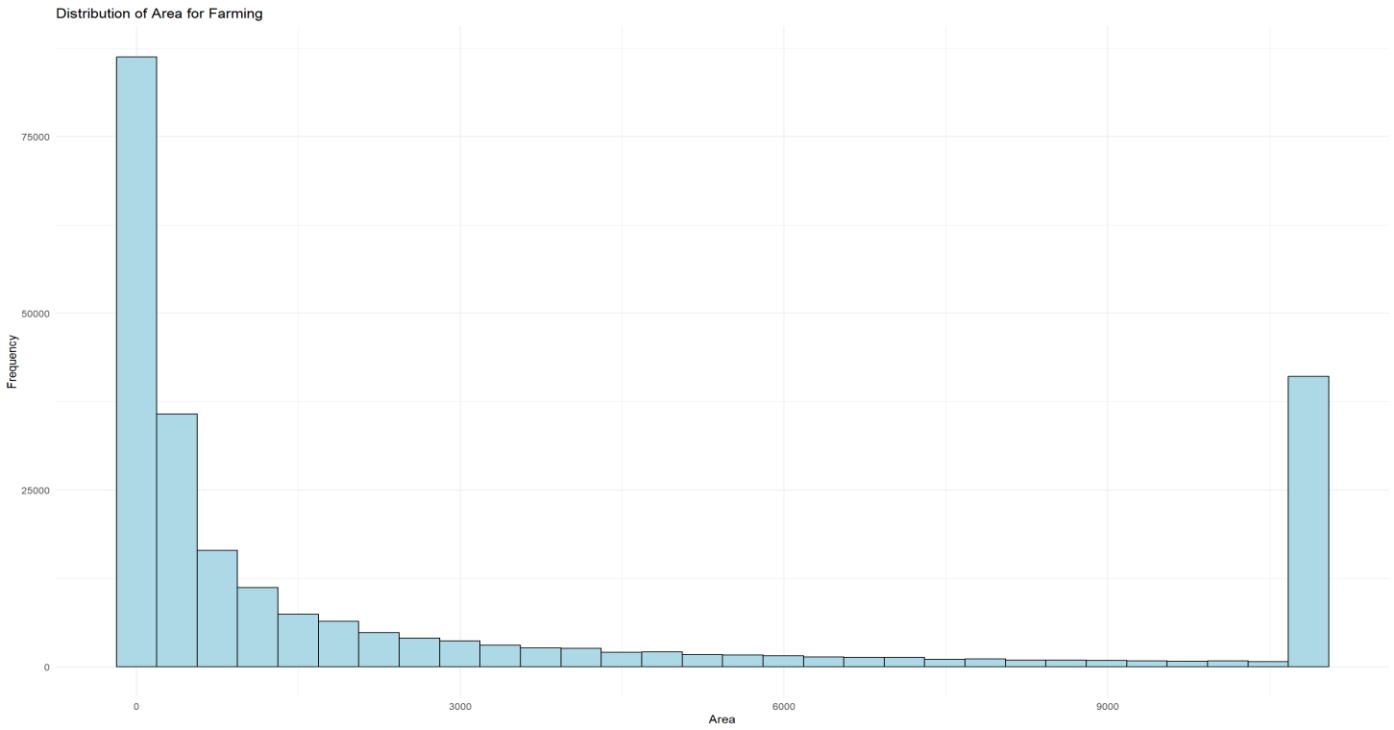




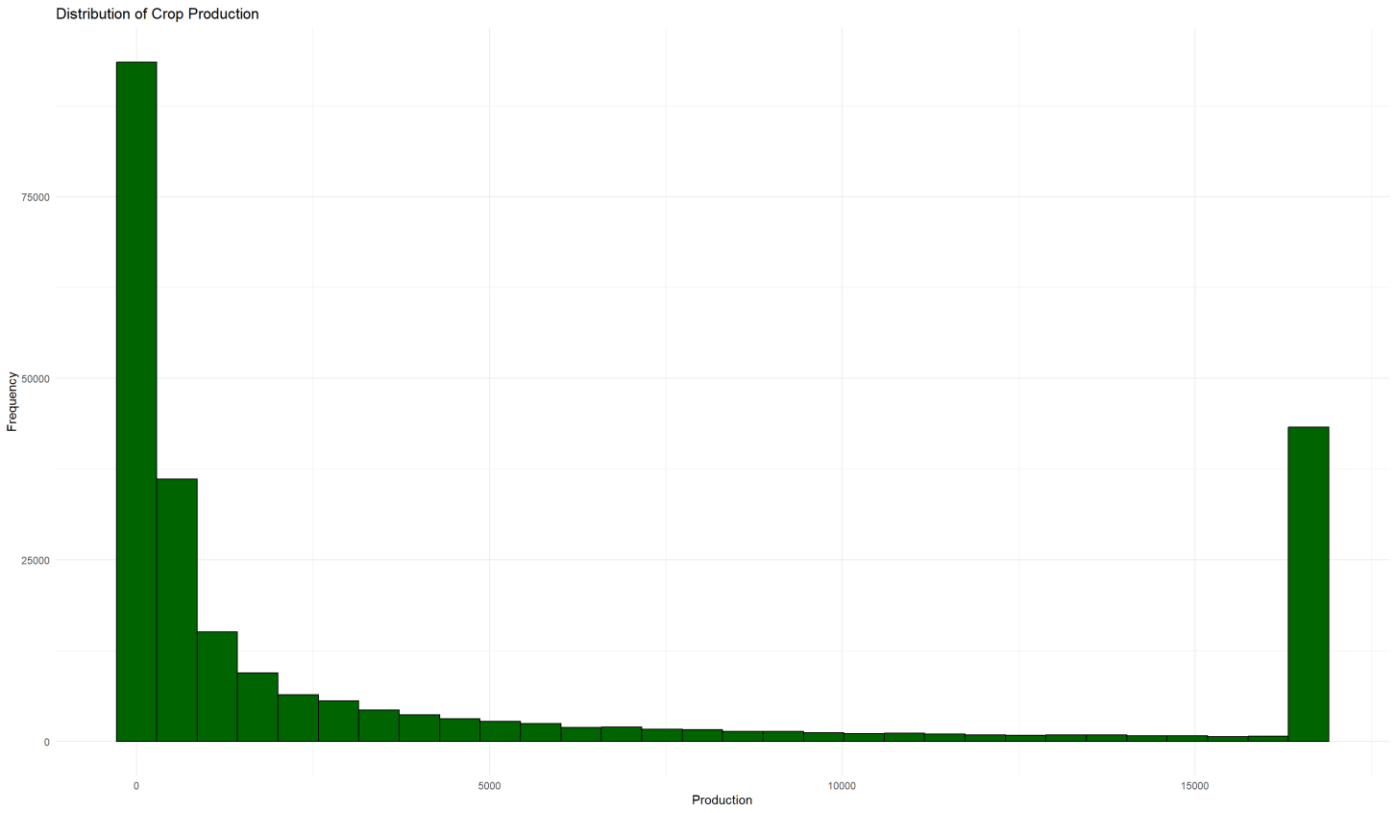




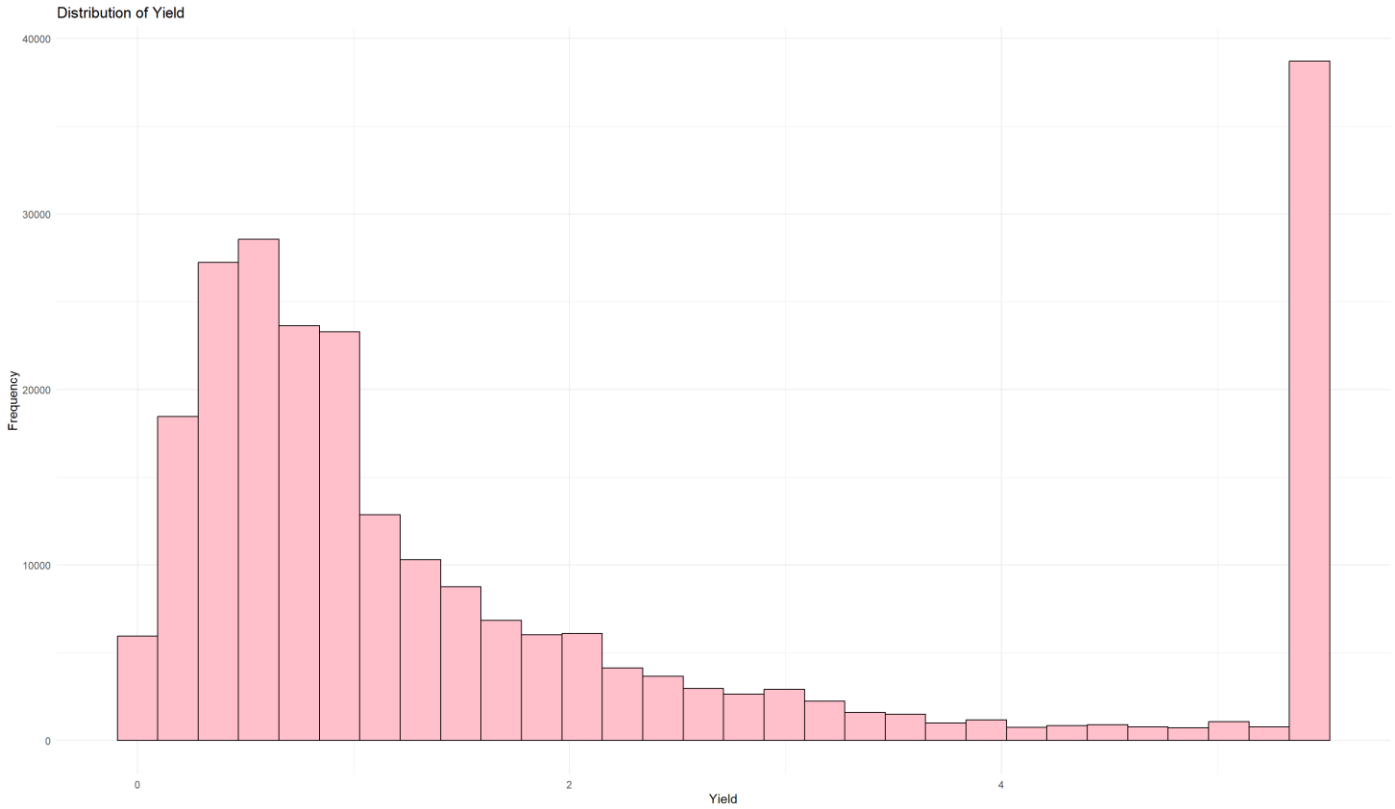
Univariate Analysis



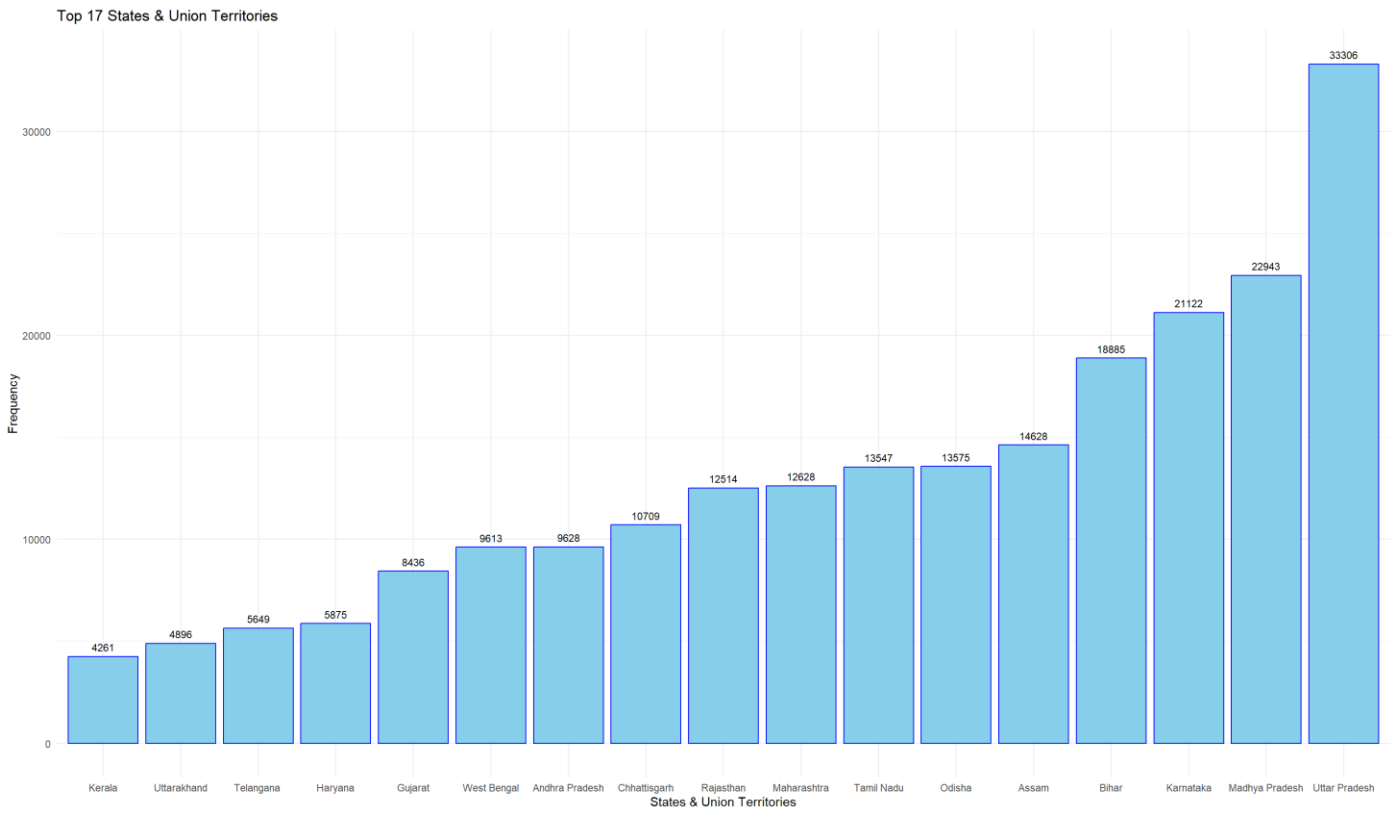
- The distribution of area for farming shows a highly skewed pattern, with most farms having relatively small areas, and a few outliers with very large areas.
- This could indicate the presence of both small-scale and large-scale farming operations in the dataset.



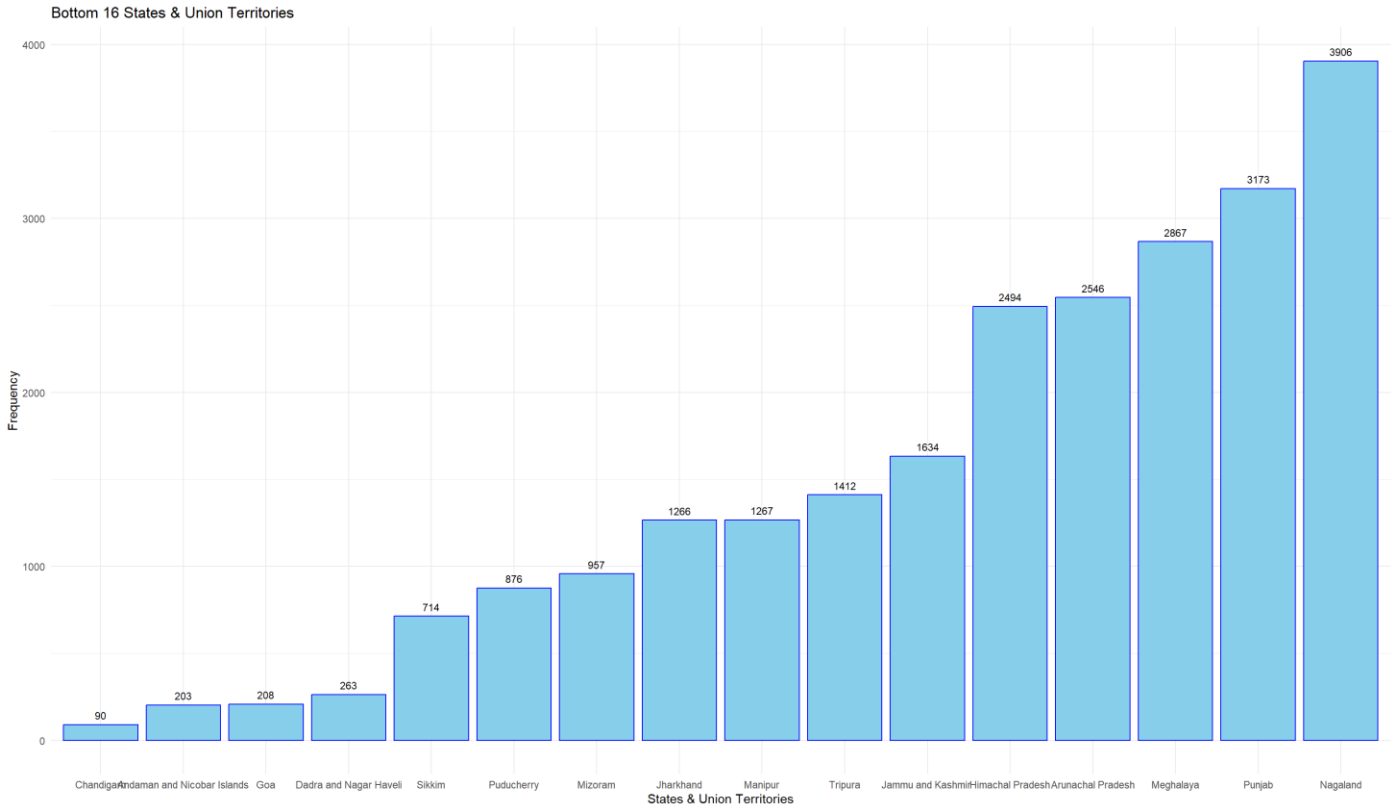
- The distribution of crop production is also highly skewed, with a few areas contributing to a significant portion of the total production.
- This could be due to factors like favorable climate, soil conditions, irrigation facilities, or the use of advanced agricultural practices in those areas.



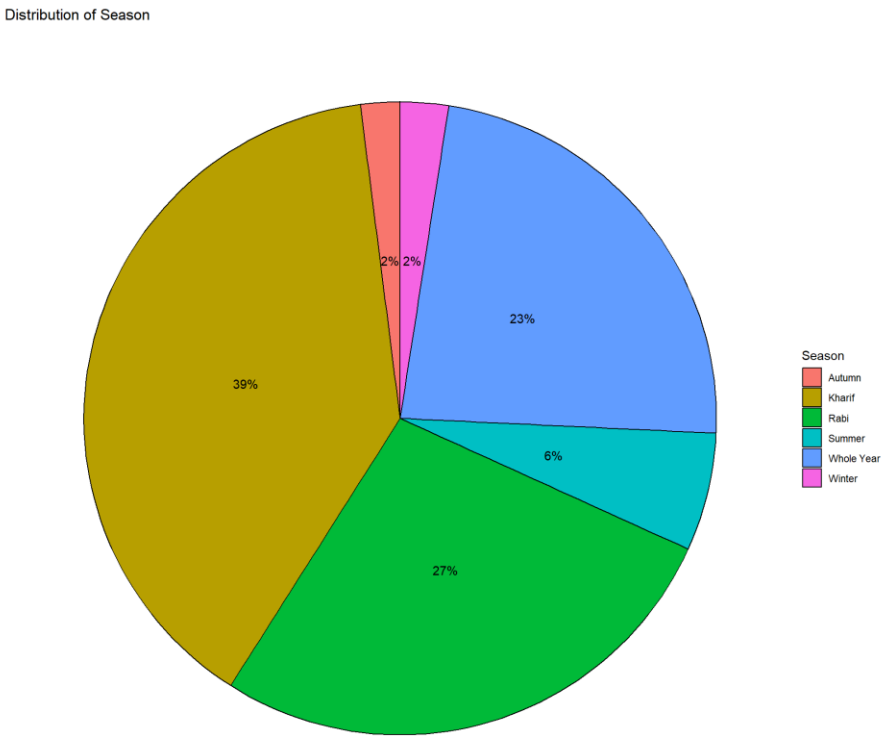
- The distribution of yield shows a more normal distribution, with most areas having moderate yields, and fewer areas with exceptionally high or low yields.
- This could suggest that yield is influenced by a combination of factors, rather than being heavily skewed by a single factor.



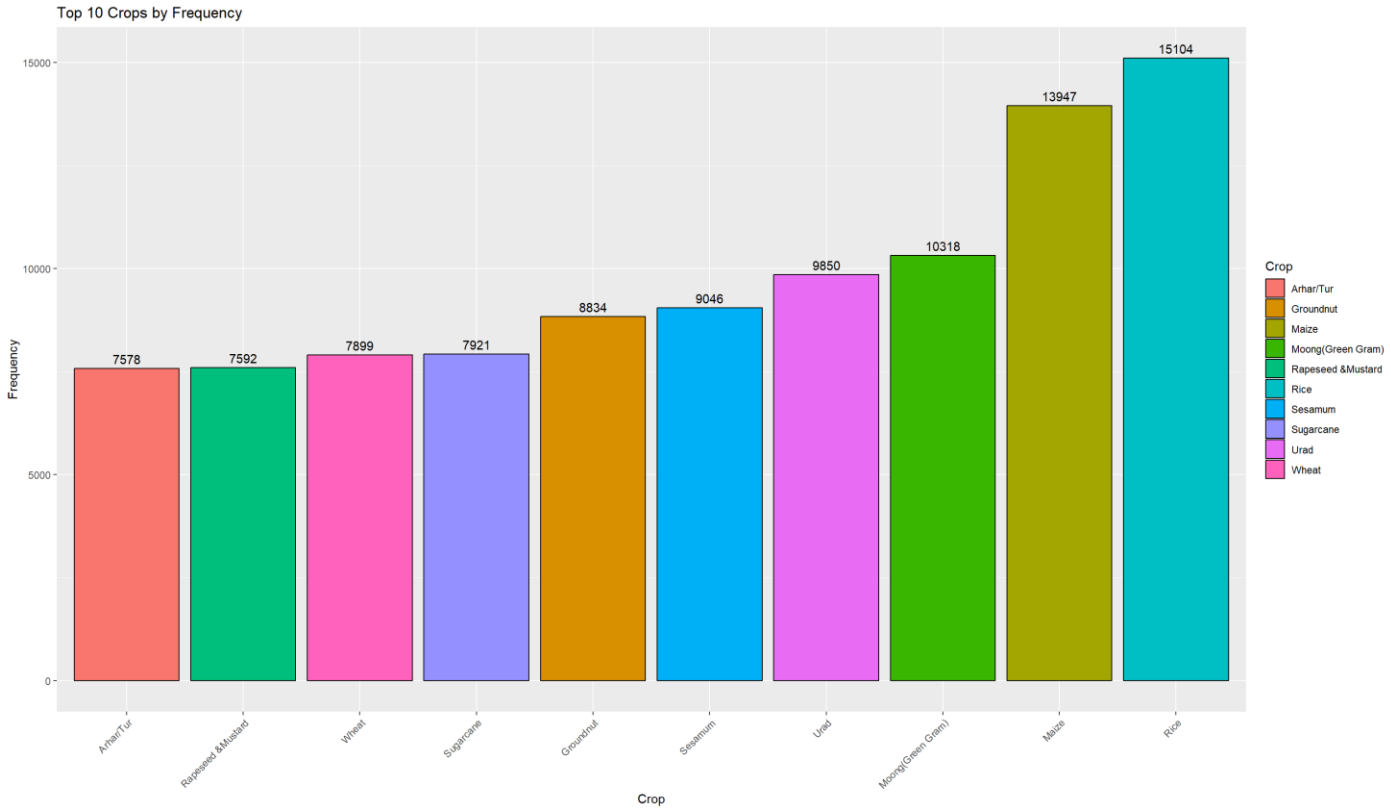
The top 17 states and union territories account for a significant portion of the total crop production, indicating the importance of these regions in the overall agricultural landscape of India.



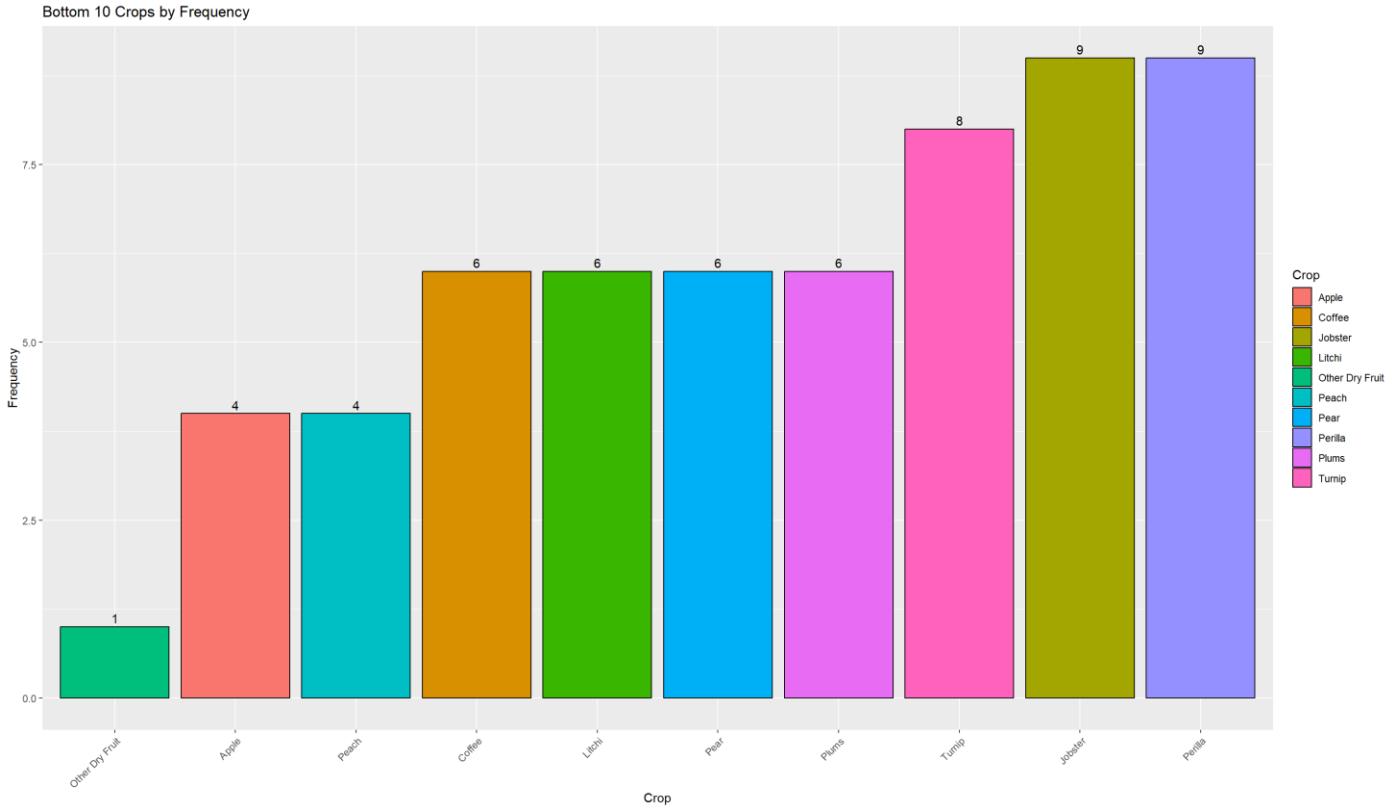
The bottom 16 states and union territories contribute a relatively smaller portion of the total crop production, potentially due to factors like climate, soil quality, irrigation facilities, or other regional challenges.



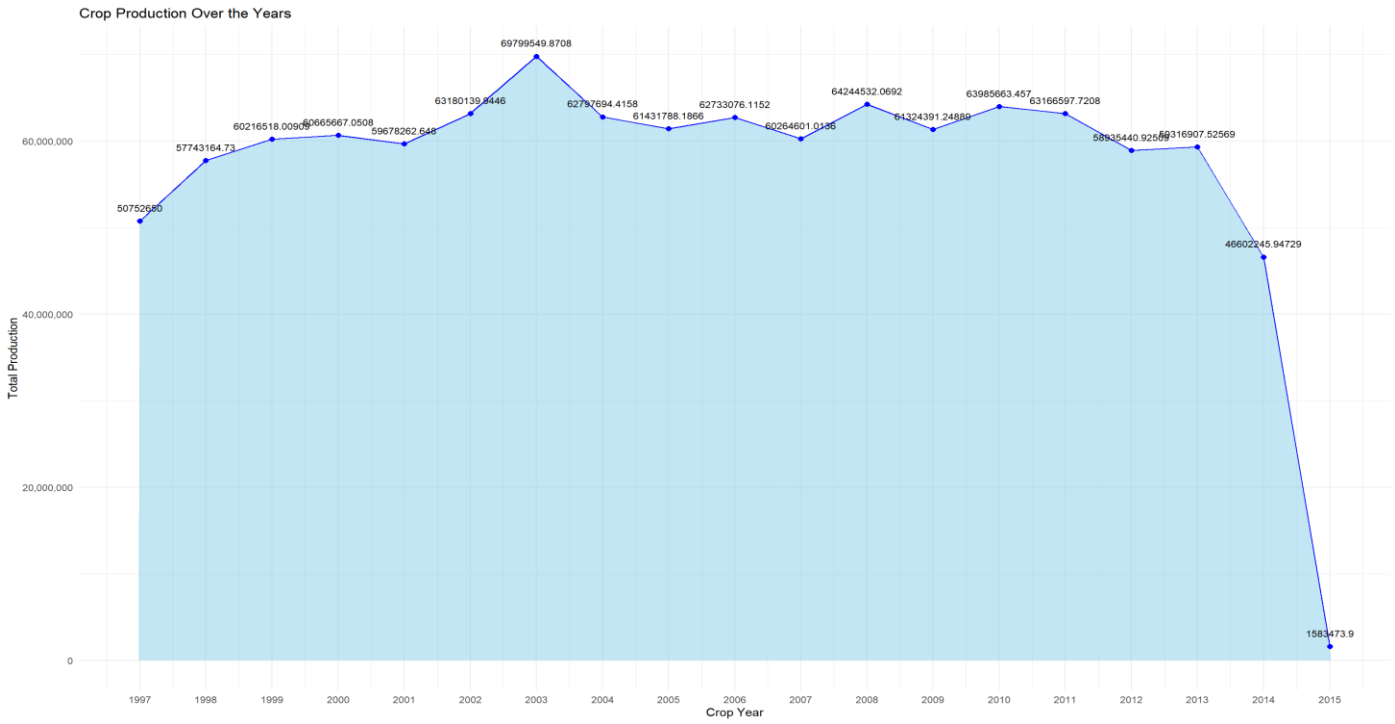
- The majority of crop production occurs during the Autumn season (39%), followed by Winter (23%) and Summer (27%).
- This indicates that the climatic conditions and availability of resources during these seasons are favorable for crop cultivation in India.



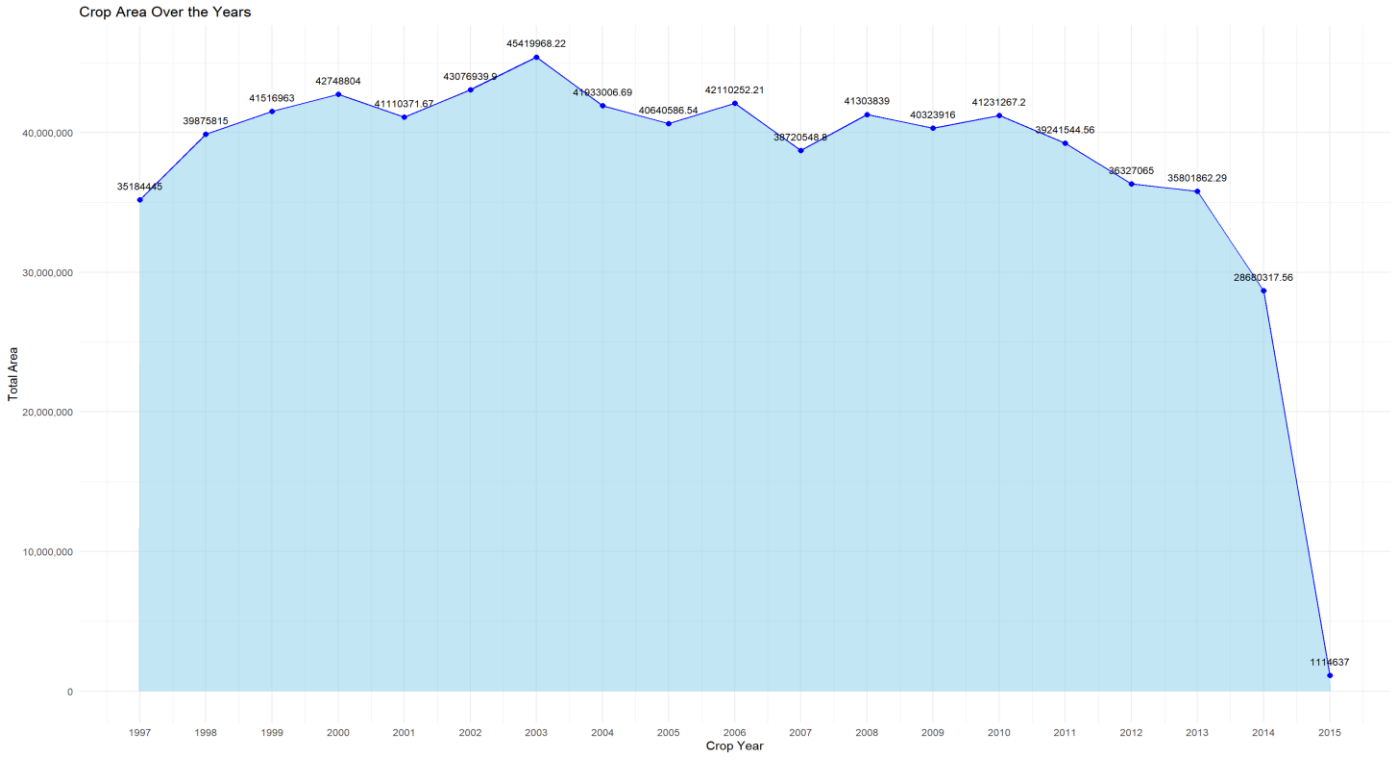
Rice, Maize, and Wheat are among the most frequently cultivated crops, suggesting high demand and production of these staple grains.



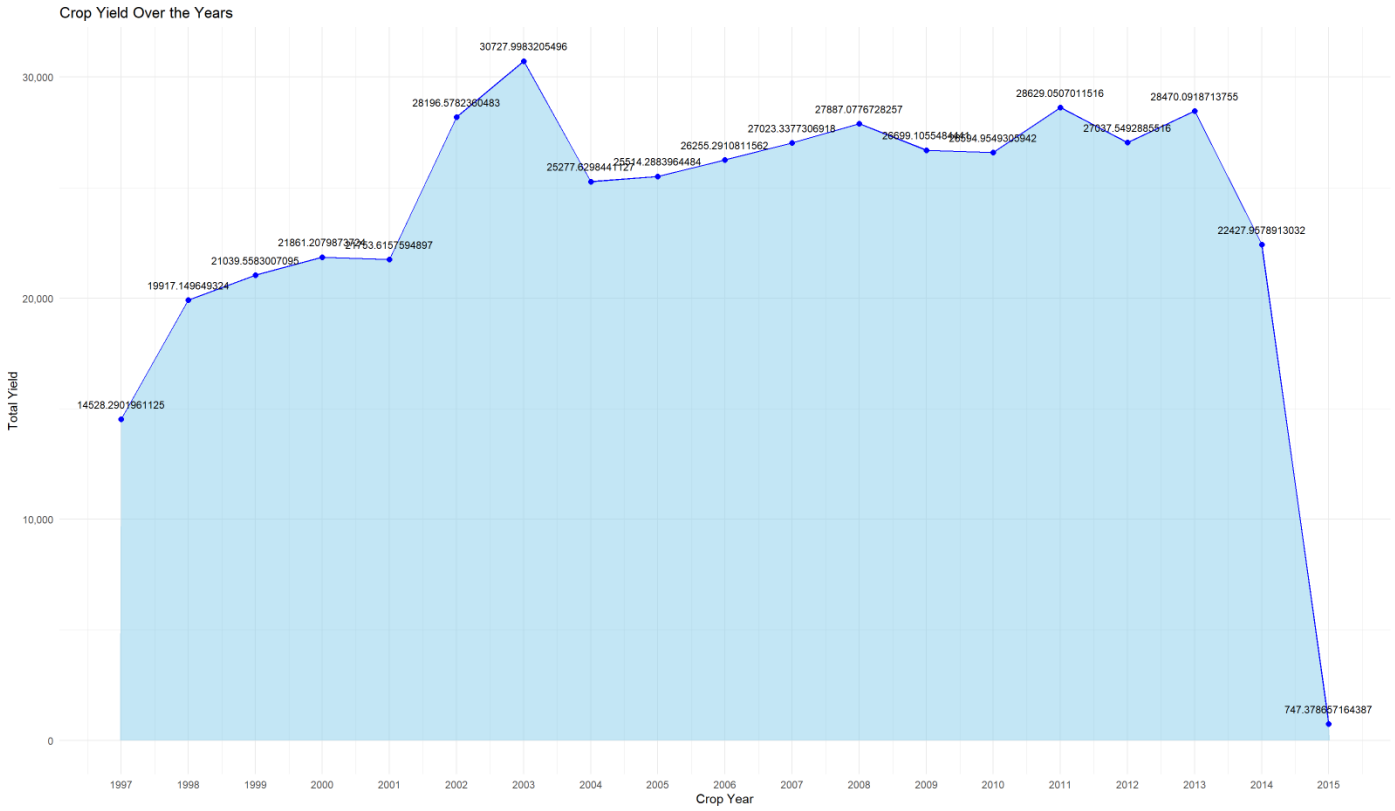
Crops like Apple, Coffee, Litchi, and Peach have relatively low frequencies, indicating potential opportunities for diversification or specialization in these niche crops.



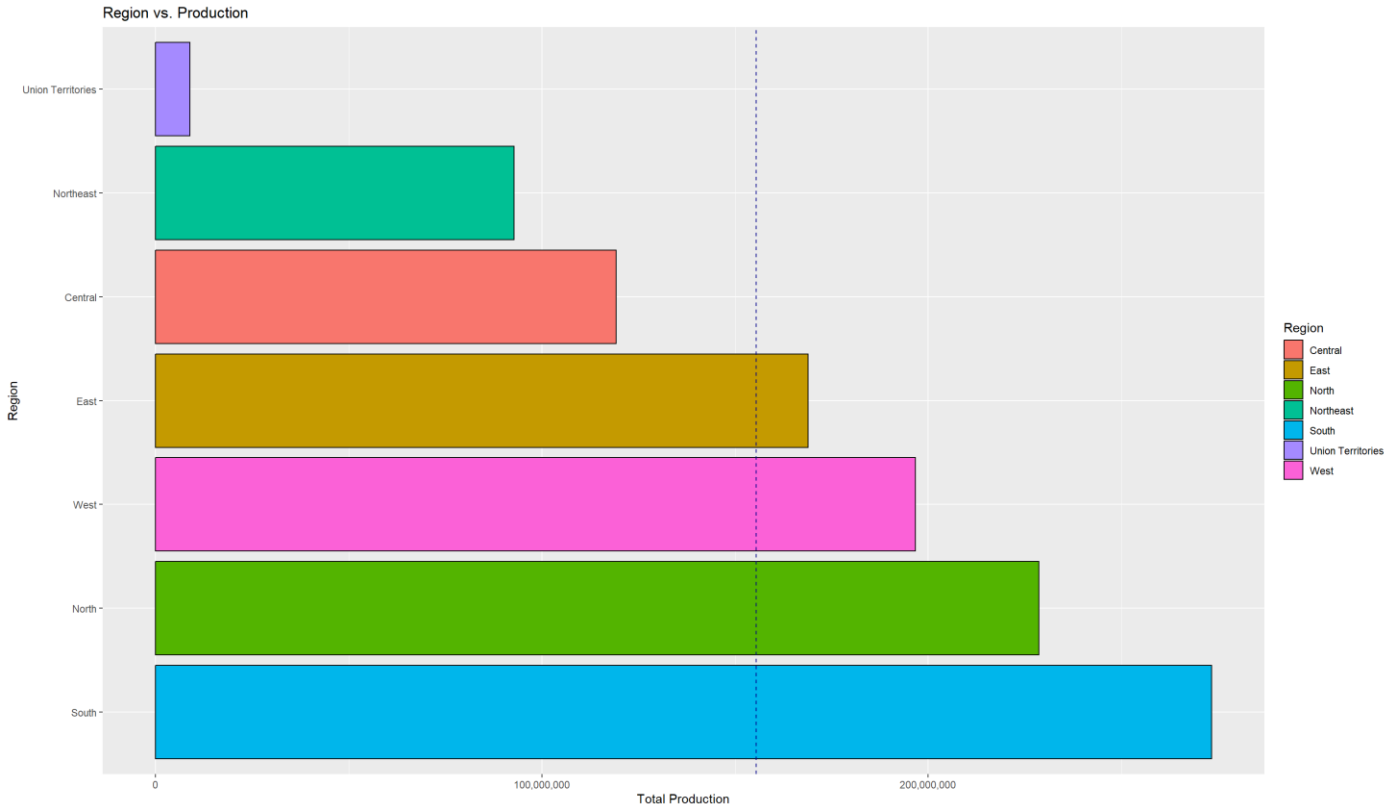
- Crop production in India has experienced significant fluctuations over the years, with peaks in certain years and dips in others.
- The highest production peak was observed around 2003, followed by a decline in subsequent years, indicating the potential impact of factors like weather conditions, policy changes, or resource availability.
- The most recent years (2013-2015) show an upward trend, suggesting improved agricultural practices or favorable conditions.

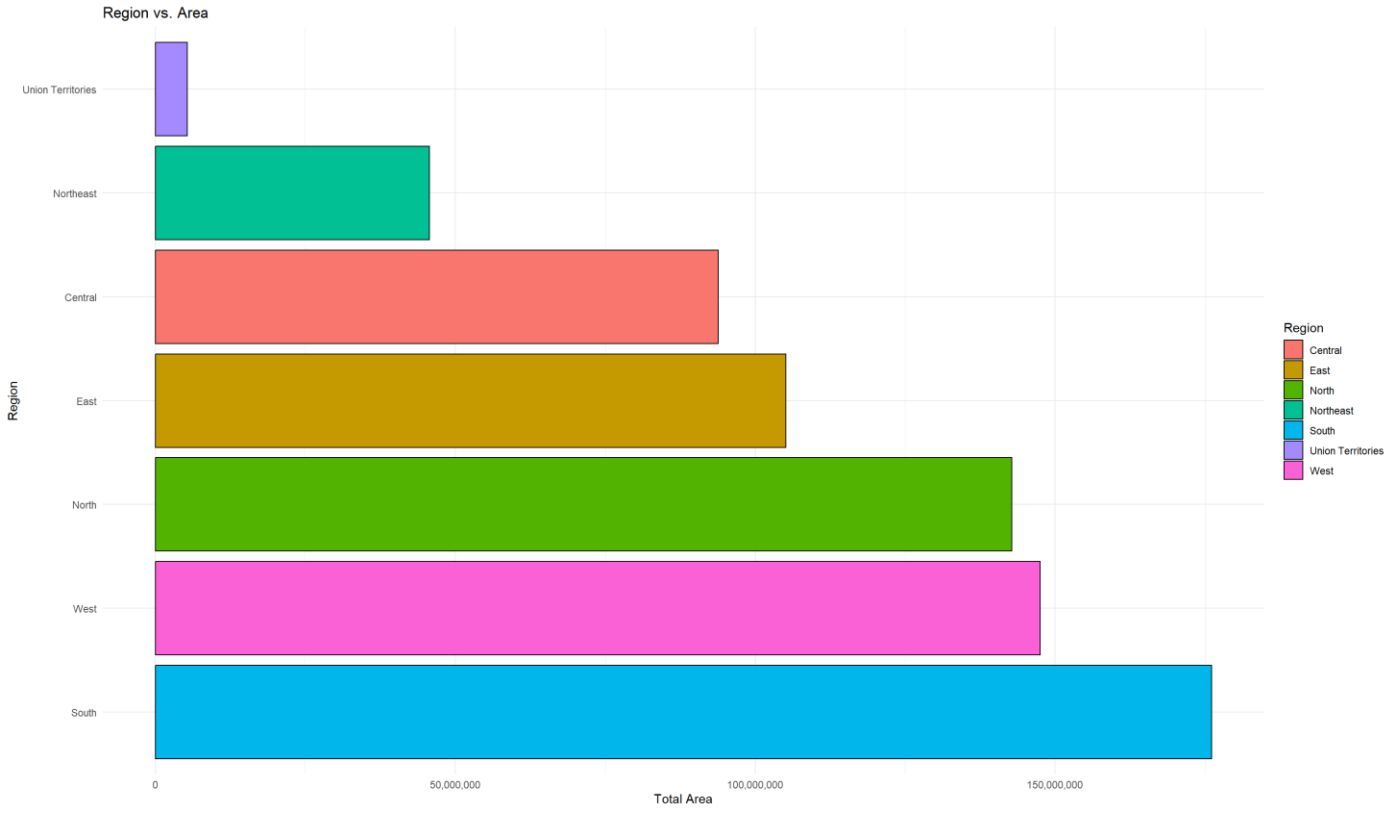


- The total area under cultivation has also fluctuated over the years, with a peak around 2003-2004 and a gradual decline afterward.
- The declining trend in crop area could be attributed to urbanization, industrialization, or other factors leading to land-use changes.
- However, the recent years (2013-2015) show an increase in crop area, indicating potential efforts towards expanding agricultural land or implementing more efficient land-use practices.

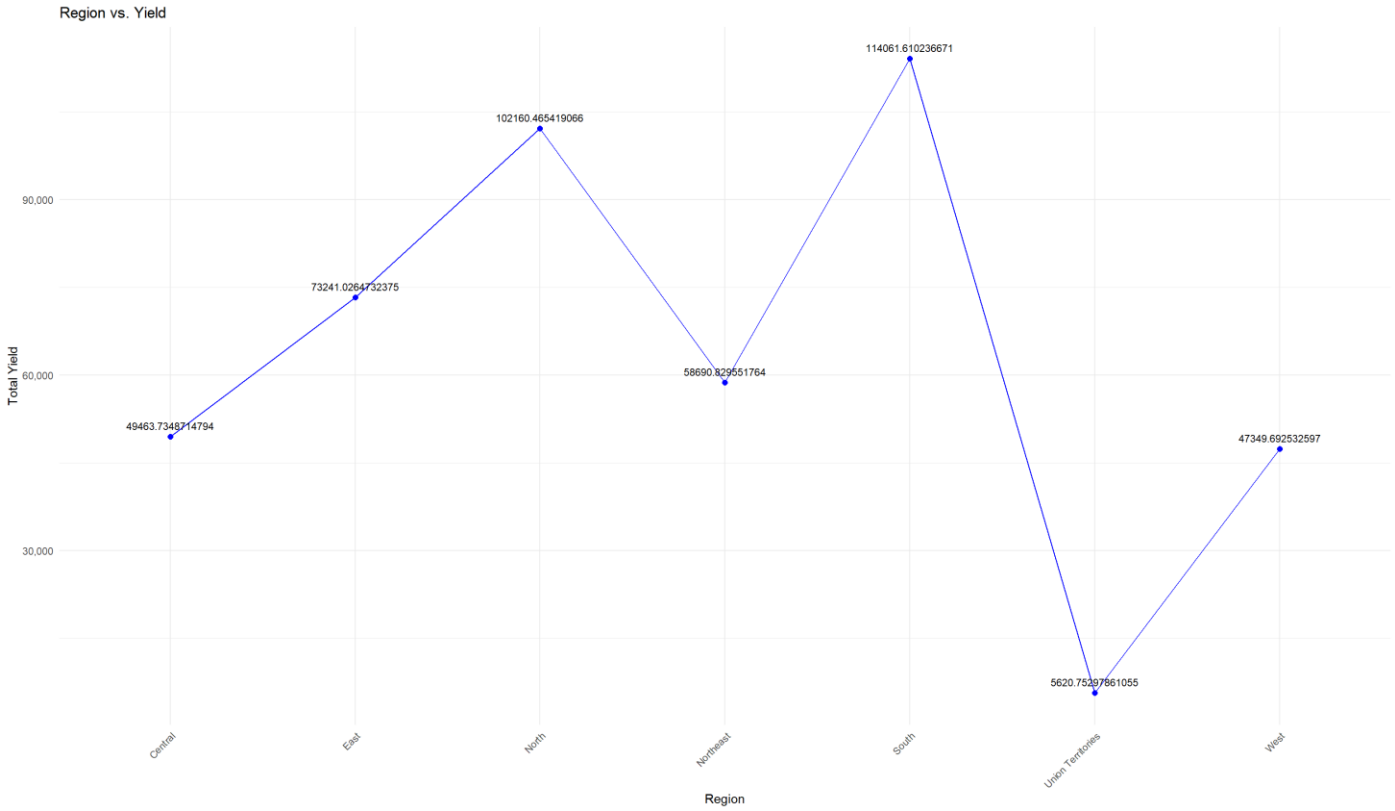


- We can observe that the crop yield over the years has shown a significant increase, with some fluctuations. The yield reached its peak in the year 2003 and then declined until 2007.
- After that, it started increasing again, indicating an overall positive trend in crop production over the years.

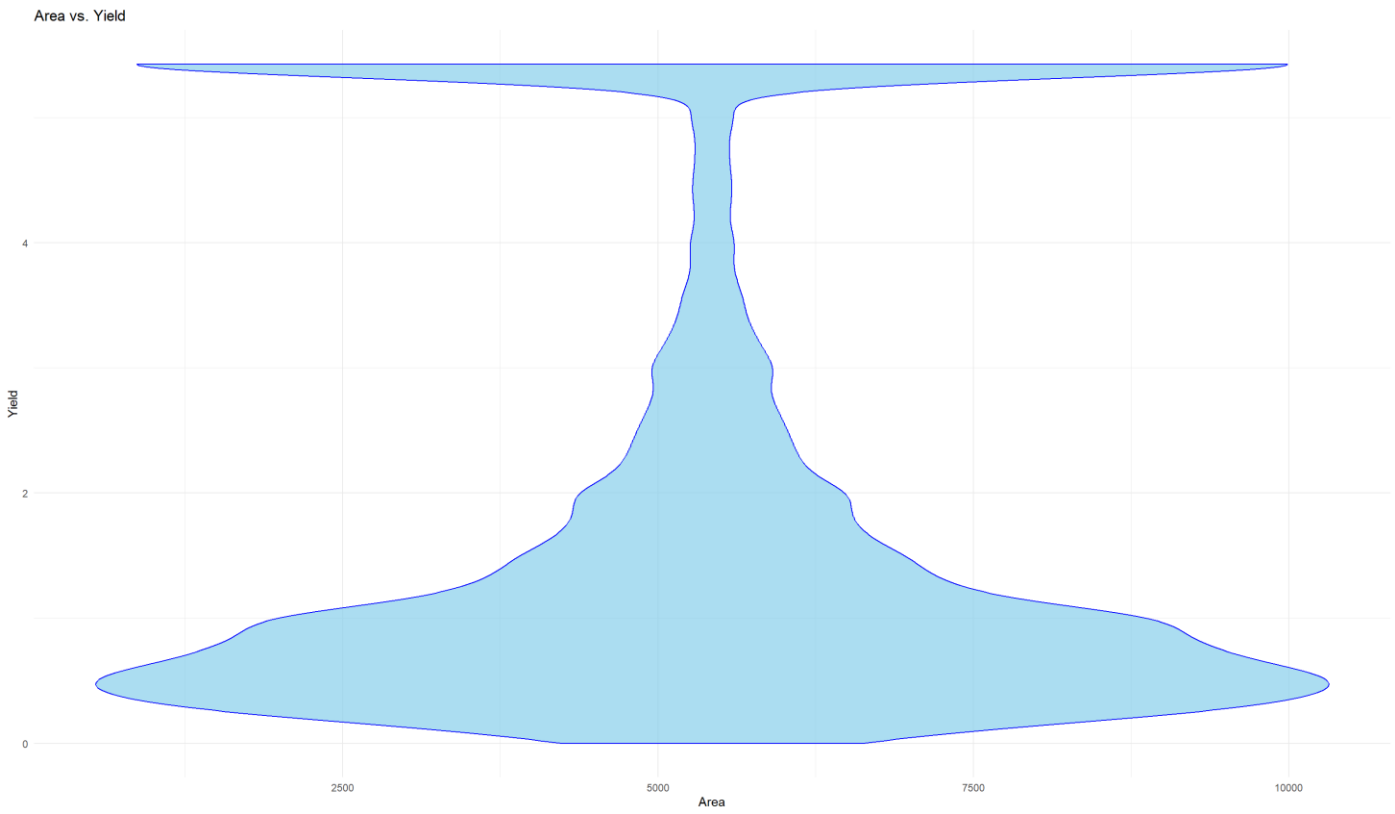




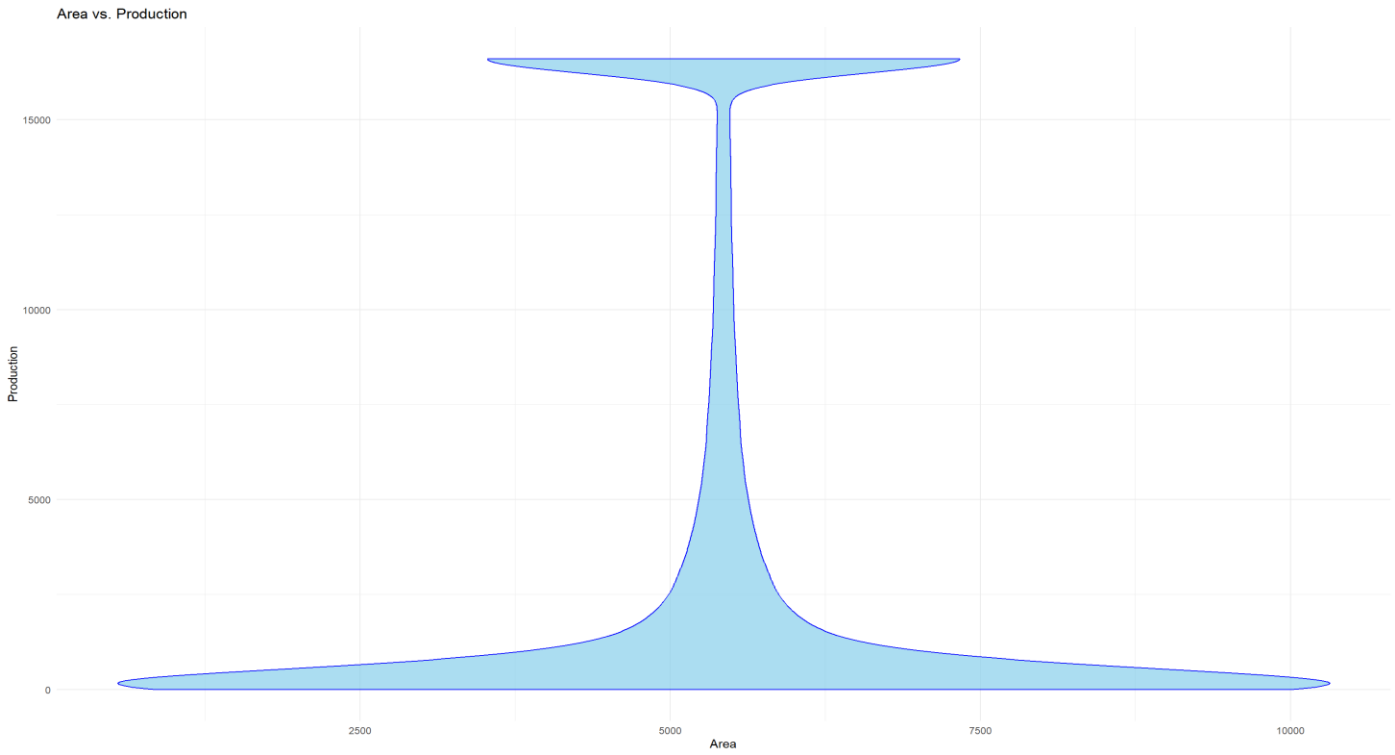
- Both provide insights into the regional distribution of crop production and cultivated area.
- The South region appears to have the highest production, followed by the North and Northeast regions.
- However, the South region also has the largest cultivated area, which could be a contributing factor to its high production.



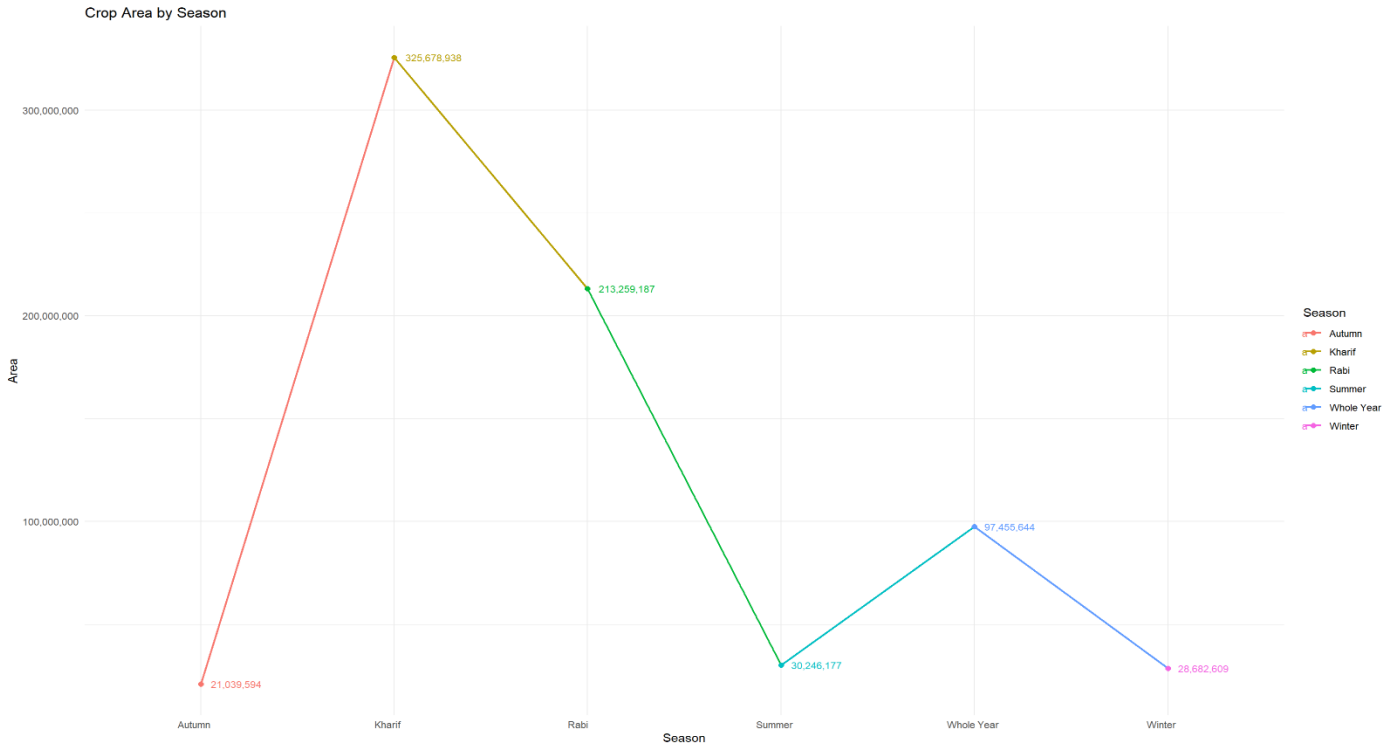
- The Union Territories region exhibits the highest yield per unit area, followed by the Northeast and Central regions.
- This could indicate factors such as favorable climatic conditions, advanced agricultural practices, or the cultivation of high-yielding crop varieties in these regions.



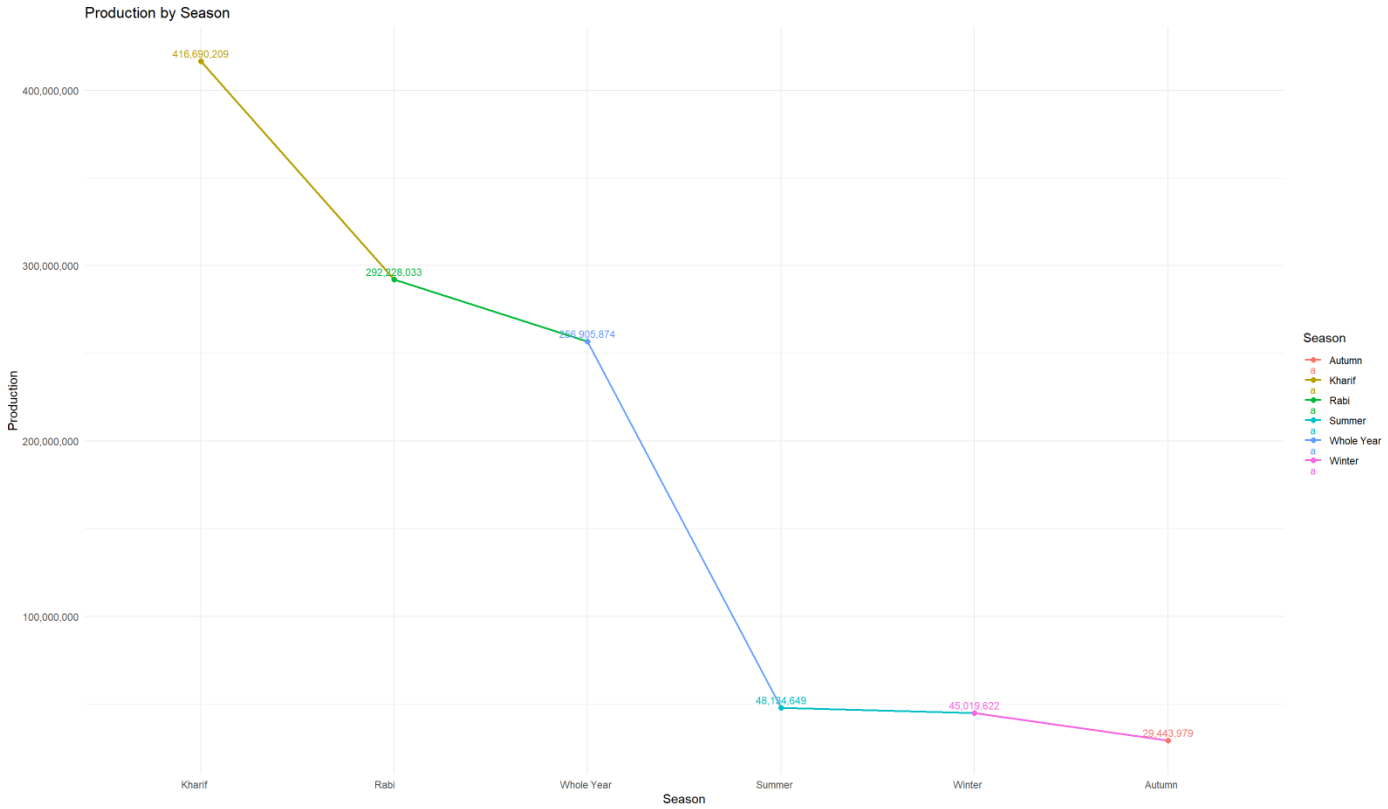
- The curve shows a distinct pattern, indicating that as the cultivated area increases, the yield initially rises, reaches a peak, and then starts declining.
- This could be attributed to factors such as diminishing returns, resource constraints, or environmental factors affecting larger cultivation areas.



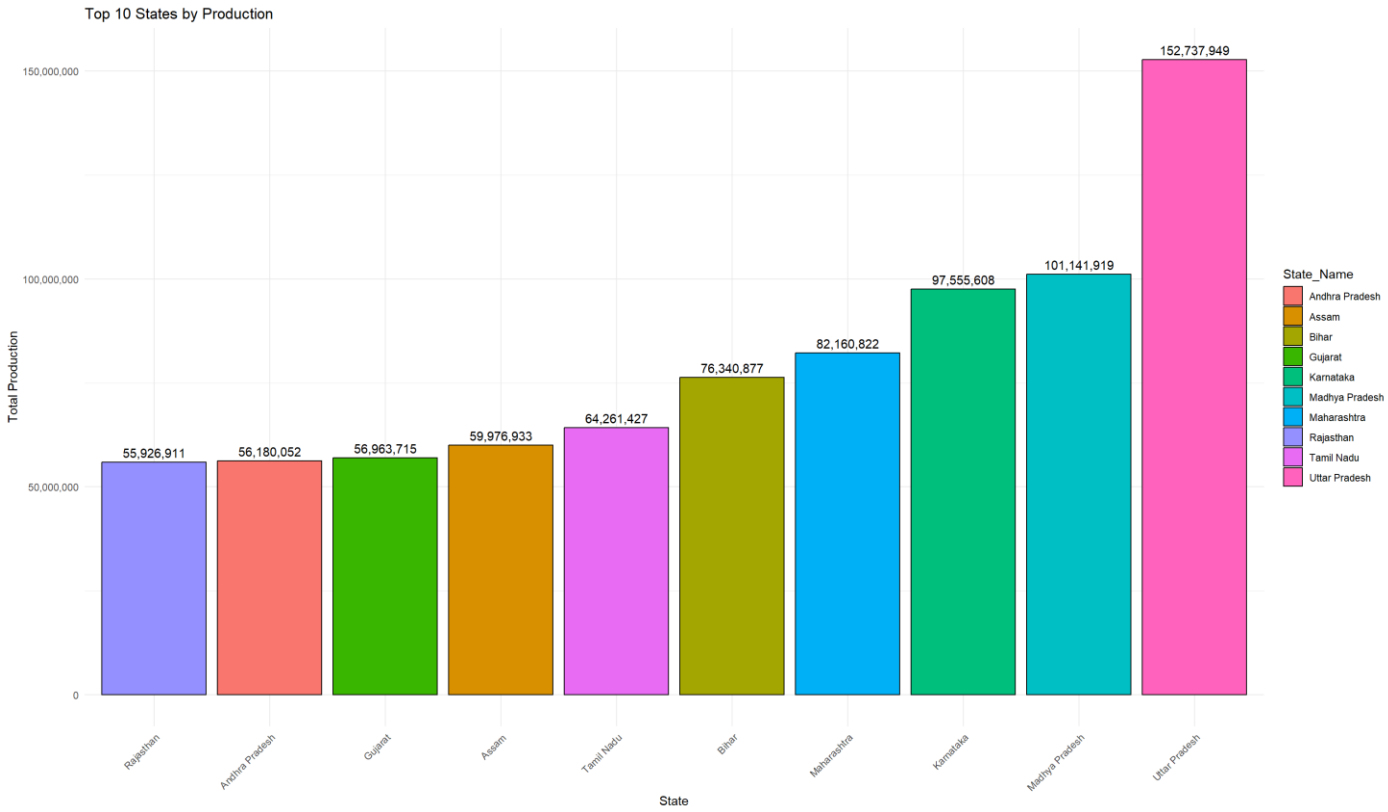
- The curve shows that crop production is maximized within a specific range of cultivated area, beyond which production decreases.
- This could indicate factors like diminishing returns, resource constraints, or potential over-cultivation at larger areas.
- Understanding the optimal area for maximizing production can help in efficient land utilization and resource allocation.



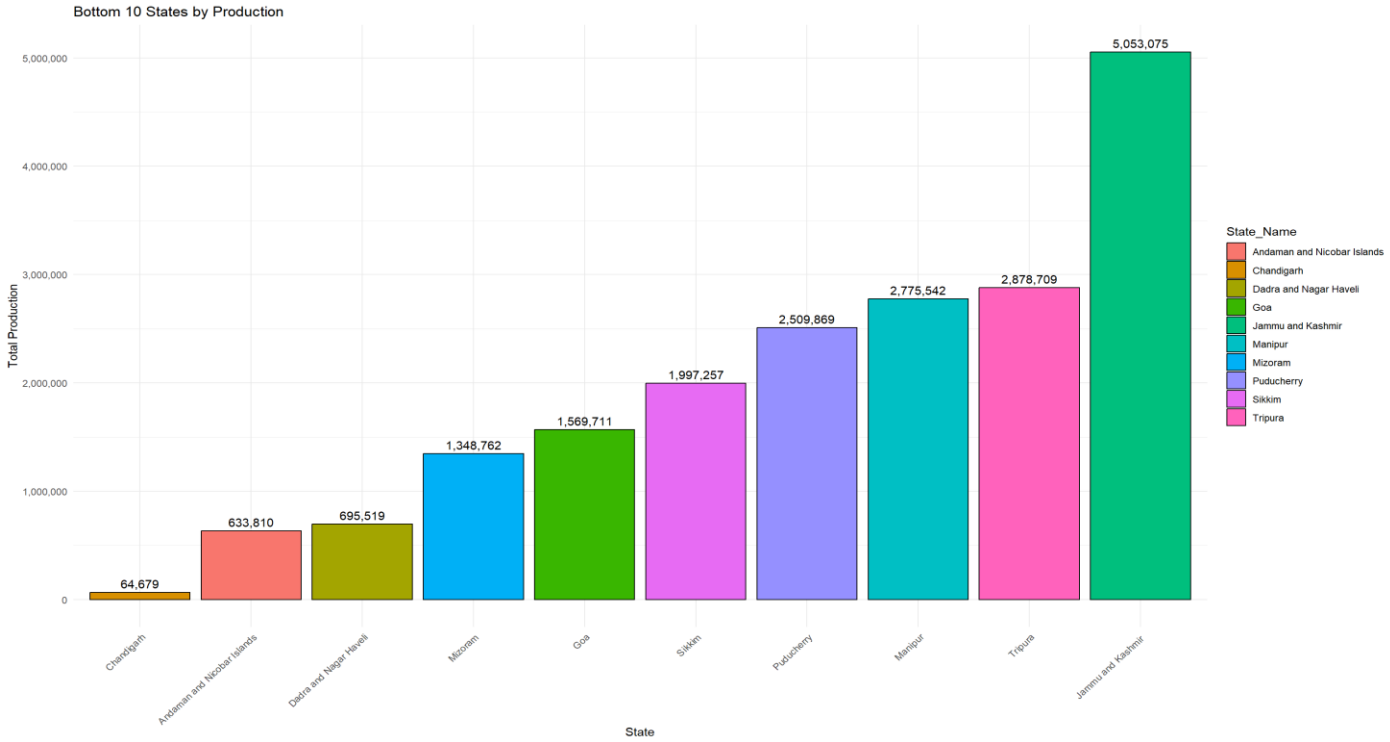
- The highest crop area is during the Kharif (monsoon) season, followed by Rabi (winter) and Autumn seasons.
- This aligns with the typical agricultural patterns in India, where major crops are cultivated during the monsoon and winter seasons.
- Identifying the peak cultivation seasons can aid in planning logistics, labor requirements, and resource management.



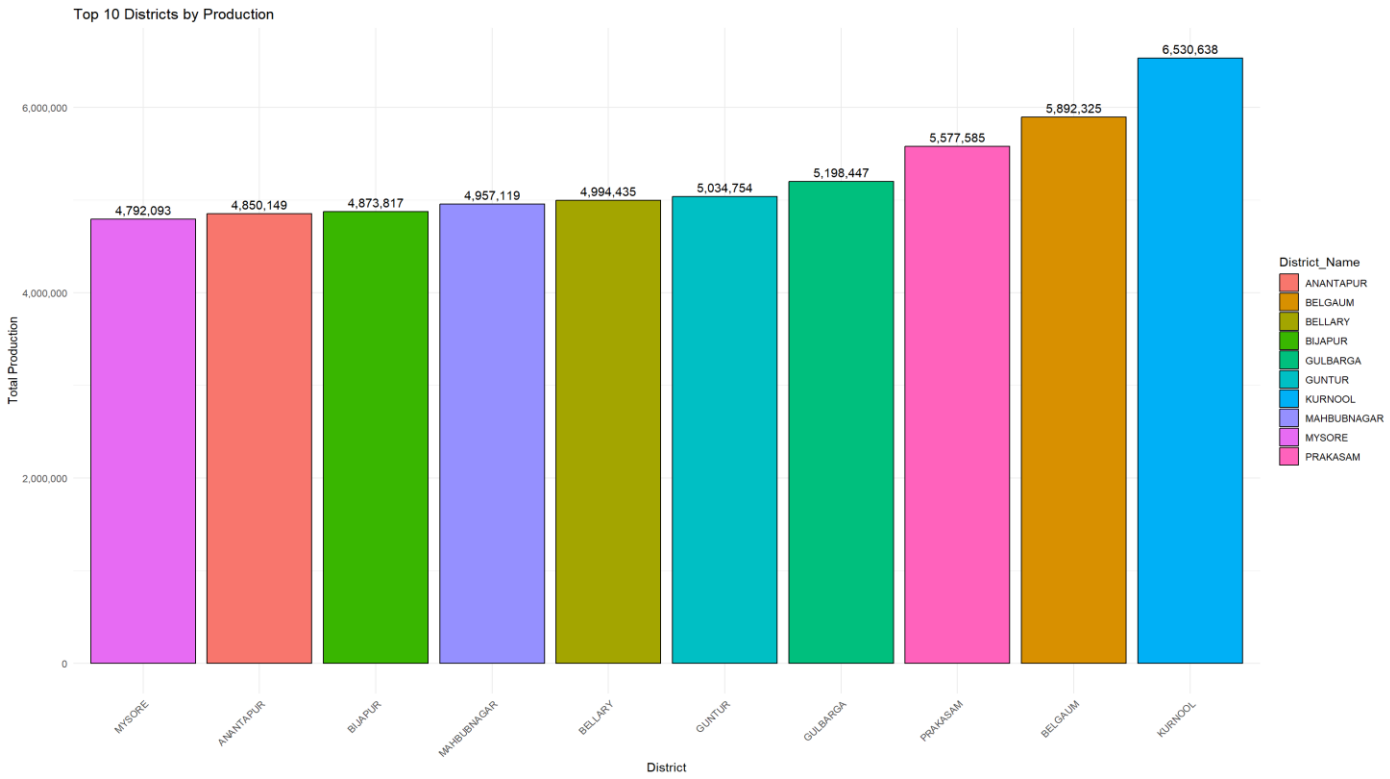
- Production follows a similar trend as crop area, with the highest production during the Kharif season and the lowest during Autumn.
- This highlights the importance of monsoon rainfall and favorable climatic conditions for crop yields.
- Understanding seasonal variations can help in forecasting demand, storage requirements, and potential supply gaps.



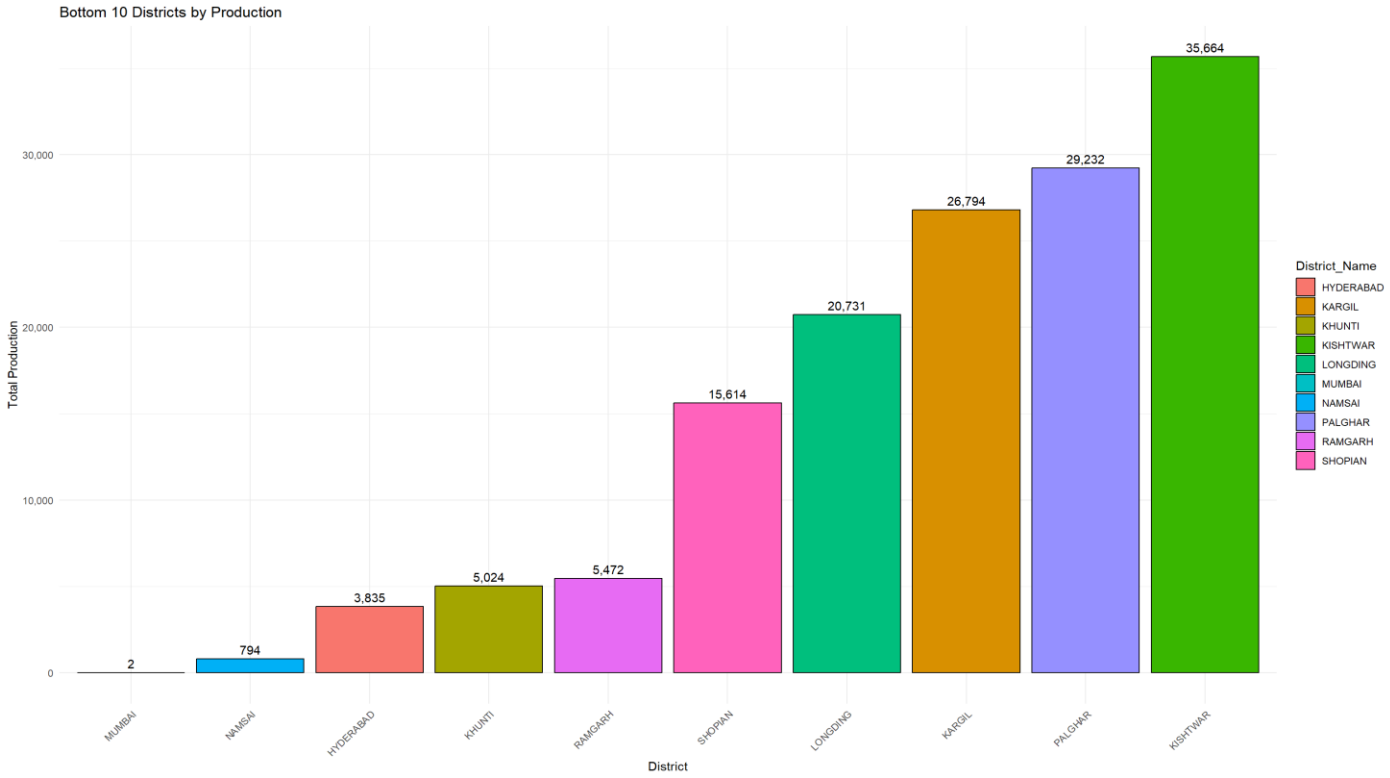
- Uttar Pradesh, Andhra Pradesh, and Assam are the top three states in terms of crop production.
- These states may have favorable climatic conditions, better agricultural practices, or more extensive cultivation areas.
- Identifying high-producing states can guide investment decisions, infrastructure development, and targeted agricultural policies.



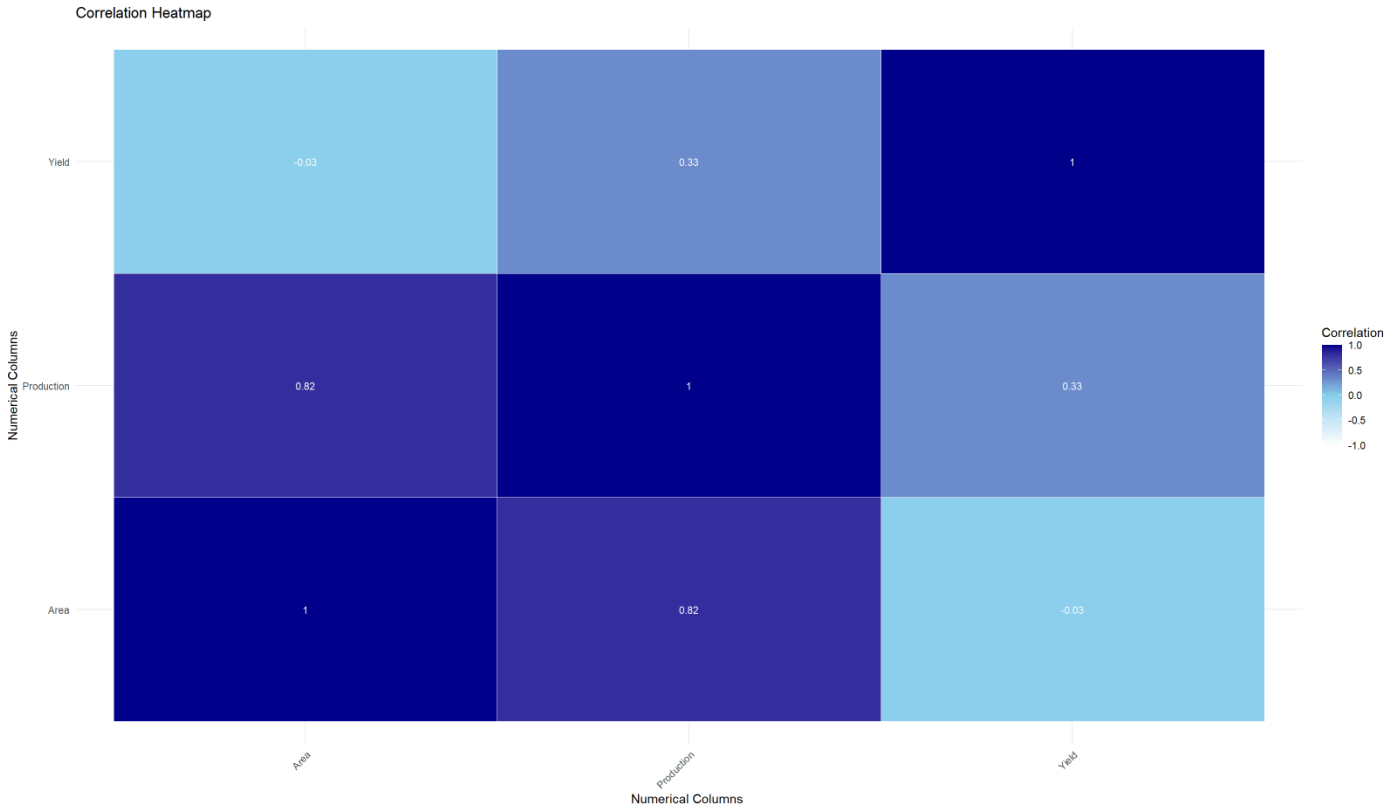
- Tripura, Andaman and Nicobar Islands, and Chandigarh have the lowest crop production among the states/UTs.
- These regions may face challenges like limited arable land, unfavorable climatic conditions, or resource constraints.
- Providing targeted support, introducing advanced agricultural techniques, or focusing on alternative industries could be explored for these regions.



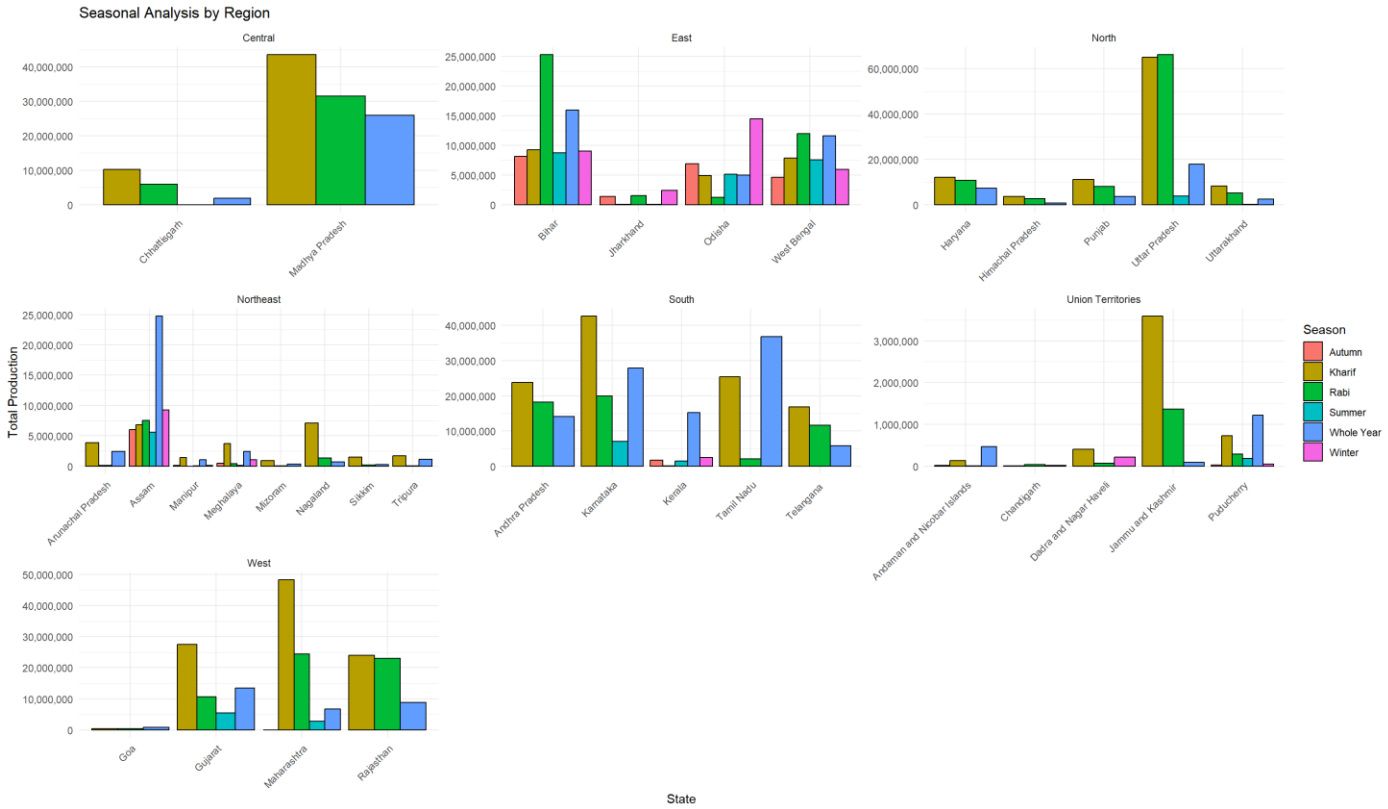
1. Kornool district in Andhra Pradesh has the highest production among the districts, followed by Belgaum in Karnataka and Bellary in Karnataka.
2. The top 3 districts account for a significant portion of the total production, highlighting their importance in the agricultural landscape.
3. Several districts from Uttar Pradesh, like Ghaziabad and Gulbarga, are also among the top producers.
4. The production levels vary significantly among the top 10 districts, suggesting differences in factors such as land area, climatic conditions, and agricultural practices at the district level.



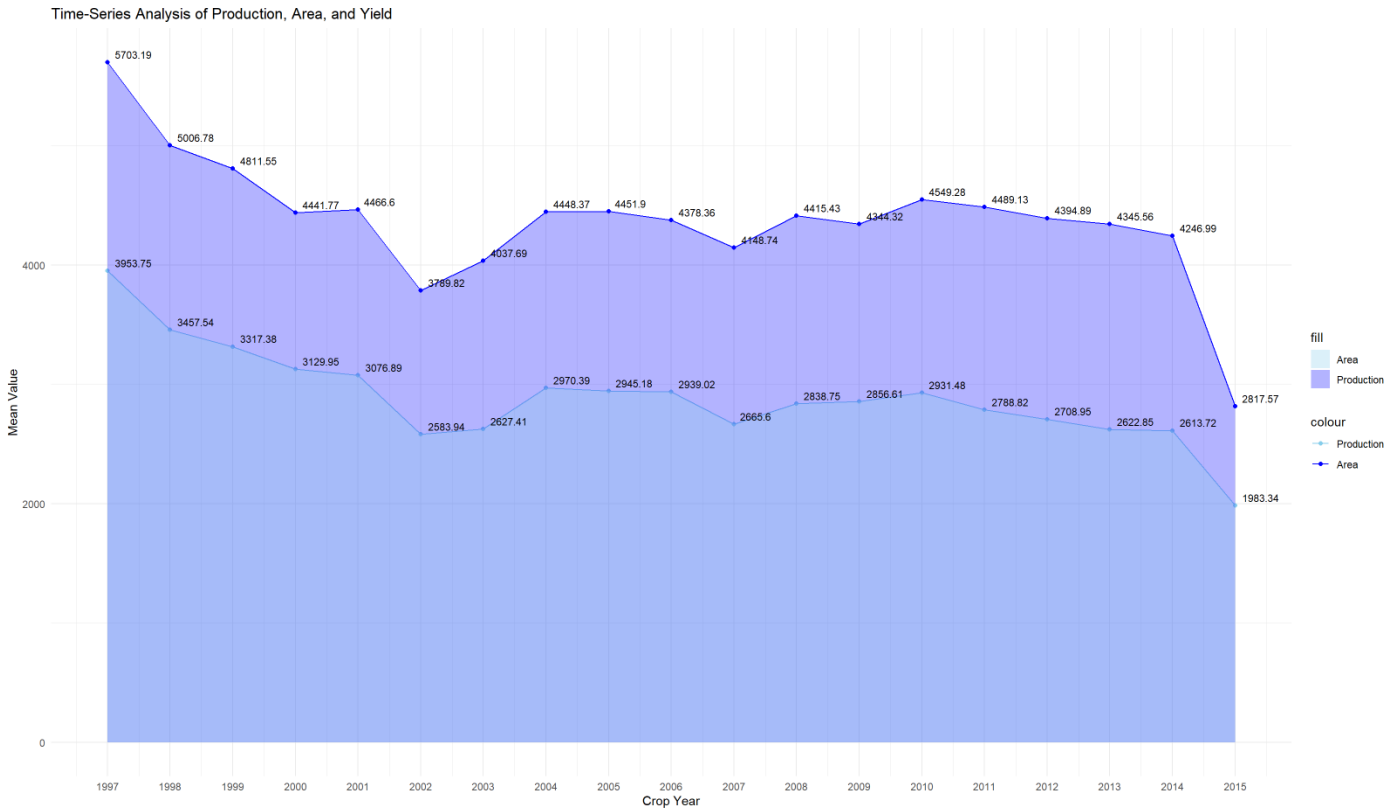
1. Prachint is the district with the lowest production, followed by Kargil and Mumbri.
2. Several districts from the state of Jammu and Kashmir, like Kargil and Kishtwar, are among the bottom 10 producers.
3. Districts like Hyderabad and Namsai also have relatively low production levels.
4. The low production levels in these districts could be due to factors like limited arable land, challenging terrain, or a focus on other economic activities within the district.



- There is a strong positive correlation (1.0) between Area and Production, indicating that as the cultivated area increases, crop production also increases proportionally.
- Yield has a moderate positive correlation (0.33) with Production, suggesting that increasing yield can contribute to higher production, but it is not the sole factor.
- There is no correlation (-0.03) between Area and Yield, implying that increase

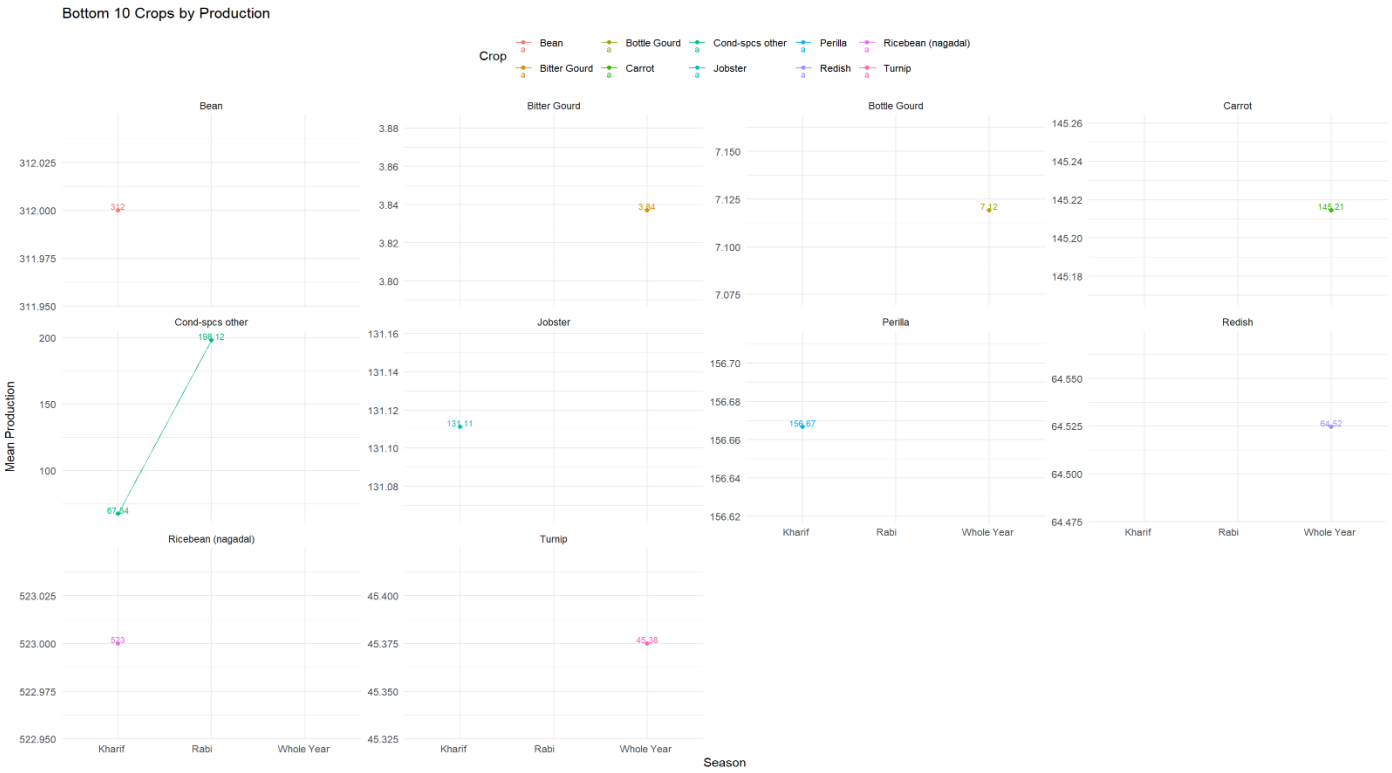


- The highest production across most regions occurs during the Kharif season (monsoon season), indicating the significance of monsoon rains for crop cultivation.
- The Union Territories have relatively low production levels compared to other regions.
- The North and Central regions have the highest production levels, likely due to favorable climatic conditions and large cultivated areas.



- There has been a steady decline in crop production and area since 1997, with some fluctuations in between.
- The yield has remained relatively stable over the years, suggesting that improving yield alone may not be sufficient to increase overall production.
- Identifying and addressing the factors contributing to the decrease in cultivated area could be crucial for increasing crop production.

- Rice, Wheat, and Sugarcane are among the top crops in terms of production, indicating their significance in Indian agriculture.
- For most crops, production is highest during the Kharif season, emphasizing the importance of monsoon rains.
- Identifying and promoting high-yielding crop varieties, especially for major crops like Rice and Wheat, could potentially increase overall production.



- The production levels of the bottom 10 crops are relatively low compared to the top crops, indicating potential for improvement.
- Certain crops like Carrot, Bitter Gourd, and Bottle Gourd exhibit minimal variations in production across seasons, suggesting potential for year-round cultivation.
- Identifying factors influencing the low production of these crops and implementing targeted interventions could help increase their yields and overall production.

Overall Summary :

- The dataset provides comprehensive information on crop production in India, encompassing various years, states, districts, crops, and seasons.
- Univariate analysis reveals skewed distributions in area, production, and yield, highlighting disparities in agricultural practices across different regions.
- Top states and union territories contribute significantly to crop production, while seasonal patterns indicate the dominance of the Kharif season.
- Bivariate and multivariate analyses uncover correlations between variables and regional disparities in production, area, and yield.
- Key insights include the impact of climatic conditions, land availability, and agricultural practices on crop production trends.

Important Insights :

1. **Regional Disparities**:

- There are significant regional disparities in crop production, with certain states and districts emerging as top producers, while others lag behind.
- Understanding the factors contributing to these disparities, such as climate, soil quality, and agricultural practices, is crucial for targeted interventions and policy formulation.

2. **Seasonal Patterns**:

- Seasonal variations play a pivotal role in crop production, with the Kharif season (monsoon season) dominating production levels.
- Identifying seasonal trends and their implications can help in resource allocation, production planning, and market forecasting.

3. **Correlation Analysis**:

- Strong positive correlations between crop area and production signify the importance of land availability in driving agricultural output.
- Moderate correlations between yield and production underscore the significance of improving crop productivity to boost overall production levels.

4. **Top Crops and Bottom Crops**:

- Analysis of top and bottom crops reveals insights into production trends, market demand, and cultivation practices.
- Identifying high-performing crops and addressing challenges faced by low-performing crops can guide investment decisions and policy interventions.

5. **Time-Series Analysis**:
- Time-series analysis highlights trends and fluctuations in crop production, area under cultivation, and yield over the years.
- Recognizing long-term patterns can inform strategic planning, risk management, and adaptation strategies in the face of climate change and other environmental factors.

6. **State and District-Level Analysis**:
- Disaggregated analysis at the state and district levels uncovers hotspots of production, areas of improvement, and opportunities for growth.
- Targeting resources and interventions based on localized insights can enhance efficiency and effectiveness in agricultural development efforts.

7. **Implications for Policy and Practice**:

- The insights generated from the analysis provide valuable inputs for policymakers, researchers, and practitioners involved in agricultural development.
- Evidence-based decision-making, informed by data-driven insights, can lead to more effective interventions, sustainable practices, and inclusive growth in the agricultural sector.

Key Metrics:

- **Crop Production:**
  Total production of crops across different regions and seasons.
- **Crop Area:**
  Total land area under cultivation for various crops.
- **Crop Yield:**
  Production per unit area, indicating crop productivity.
- **Regional Production:**
  Production levels in top and bottom states, districts, and regions.
- **Seasonal Production:**
  Production variations across different seasons.

Factors Affecting Key Metrics:

- **Climatic Conditions:** Monsoon rainfall, temperature, and humidity influence crop growth and yield.
- **Soil Quality:** Soil fertility, composition, and moisture content impact crop productivity.
- **Agricultural Practices:** Farming techniques, irrigation methods, and use of fertilizers affect crop production.
- **Land Availability:** Arable land availability and land-use practices determine the extent of cultivation.
- **Crop Selection:** Choice of crops based on suitability to local conditions and market demand.

Strategies to Be Used:

- **Enhance Agricultural Infrastructure:**
  Invest in irrigation systems, storage facilities, and transportation networks to support farming activities.
- **Promote Sustainable Practices:**
  Encourage adoption of organic farming, water conservation techniques, and eco-friendly pest management methods.
- **Crop Diversification:**
  Introduce high-value crops, specialty crops, and cash crops to diversify agricultural production and minimize risks.
- **Technology Adoption:**
  Harness advancements in agricultural technology, such as precision farming, remote sensing, and IoT-enabled solutions, to optimize resource utilization and improve productivity.
- **Capacity Building:**
  Provide training, education, and financial support to farmers for skill development, access to markets, and adoption of best practices.