SOULPAGE

IT SOLUTIONS

Agricultural Crop Production India Analysis & Prediction

1997 - 2014

Agriculture Crop Production Analysis

Problem Statement:-

India is one of the Top 3 Agricultural countries where people in India mostly depend on Agriculture. It produces a wide variety of crops every year. India ranks second worldwide in farm output. As per 2018, Agriculture employed 50% of the Indian workforce and contributed 17-18% to the country's GDP. India ranks first globally with the highest net cropped area followed by the US and China. Analyzing crop production here helps us in a better understanding of crops, seasons, statewide and districtwide.

Opportunity:-

The data refers to district wise, crop-wise, season-wise and year-wise distribution on crop covered area and production. The data is being used to study and analyze crop production, production contribution to district/state/country, Agro-climatic zone wise performance, and high yield production order for crops, crop growing pattern, and diversification.

Technologies used:-

- Python, Numpy, Pandas
- Exploratory Data Analysis
- Feature scaling
- Feature Engineering
- Seaborn, Matplotlib

Machine learning models:

- Linear Regression
- Random forest Regression
- XGBoost Regression

Dataset:-

The source of the dataset is here. This dataset contains several records with features like State, District, Year, Season, Crop, Area, and Production with 246091 rows and 7 columns.

Process:-

Use the tools that we have like Numpy, Pandas, Matplotlib, Seaborn and Python to get insights as asked above from the data. Let's answer each and every point one by one.

Clean the data to remove null values, Drop the duplicates and start exploring the dataset useful for our analysis.

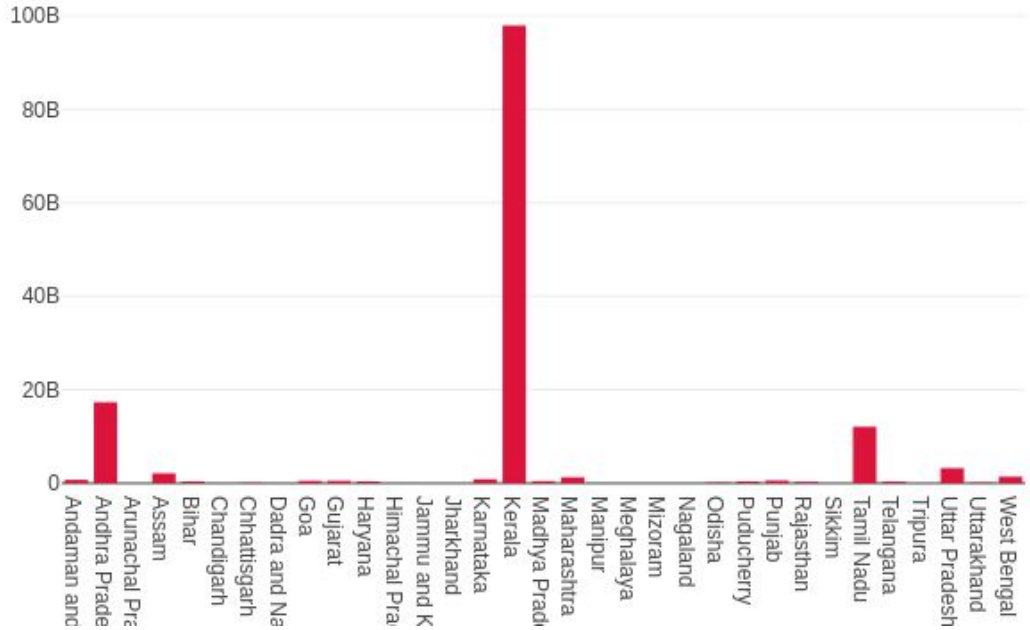
Insights:-

- What are the Top3 states giving the highest production overall
- What is the crop that giving the highest production every state
- Overall, what are the highest crop production Districts
- What is the Year wise total production
- What is the crop with the highest production every year
- Which season gives us more production

Data Analysis:-

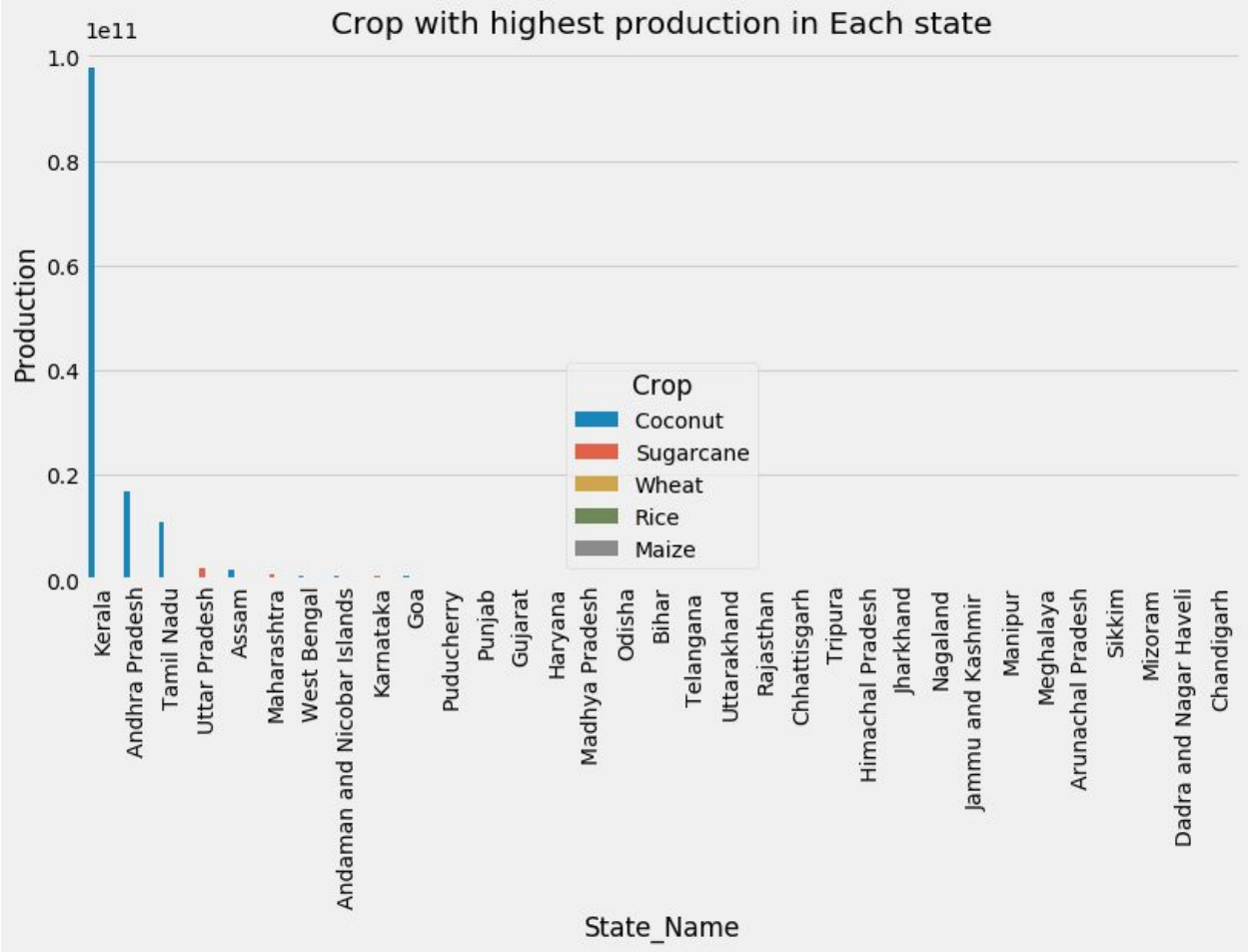
There are a total of 29 states and 4 union territories available in the dataset with their crop production for every year, season, and districts.

Crop production in Every state from 1997 to 2015

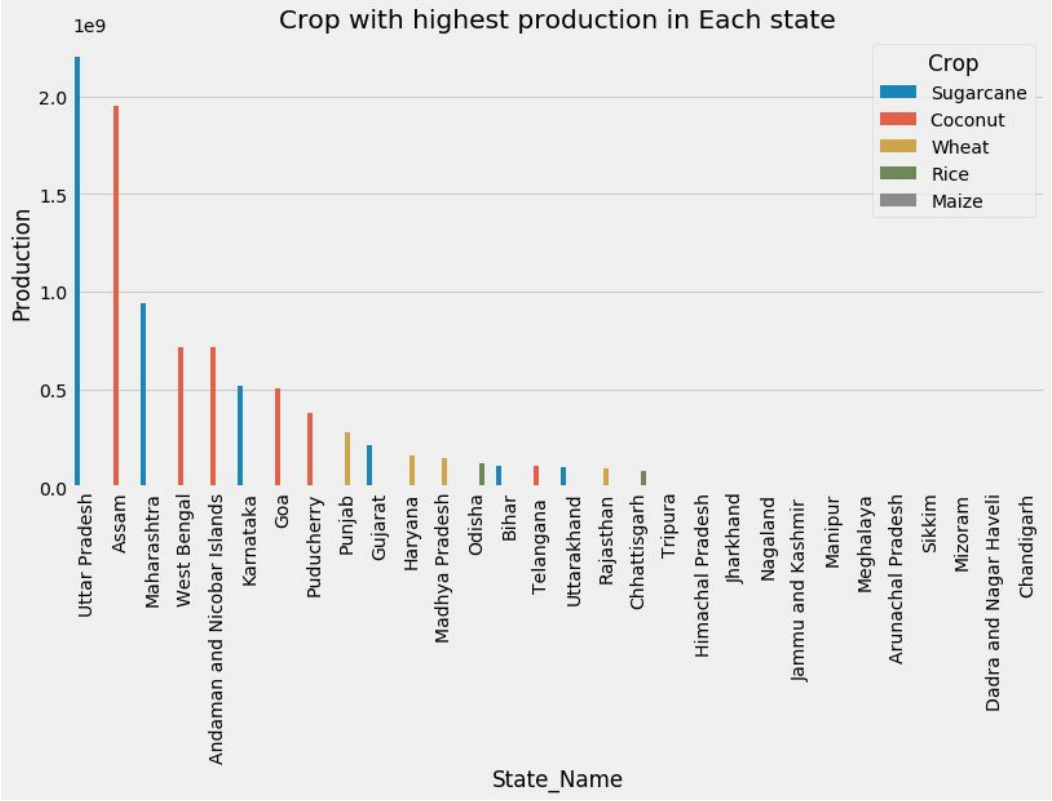


It's interesting to see Kerala ranks first in crop production over the years as it is also the highest literacy rate state in India. Andhra Pradesh ranks second in the highest crop production overall from 1997 to 2015 followed by Tamil Nadu.

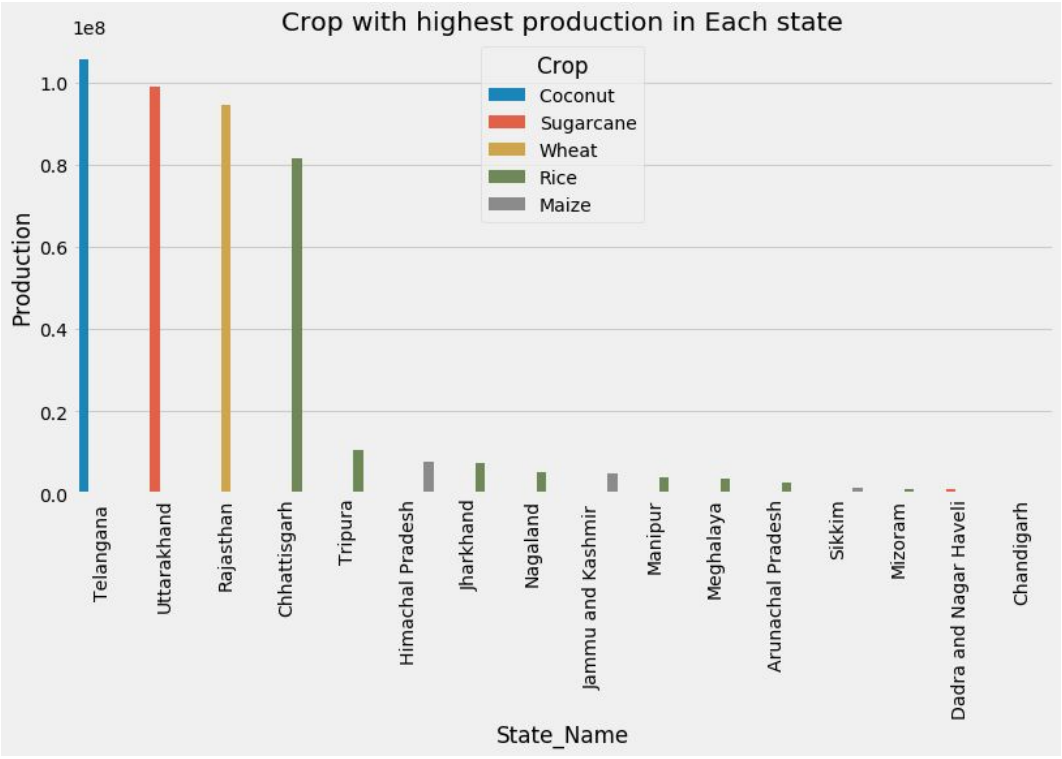
Now it would be interesting to see what is the crop that is giving huge production all through the years. We know that Kerala is the highest crop production state and Coconut is the crop that is produced in it.



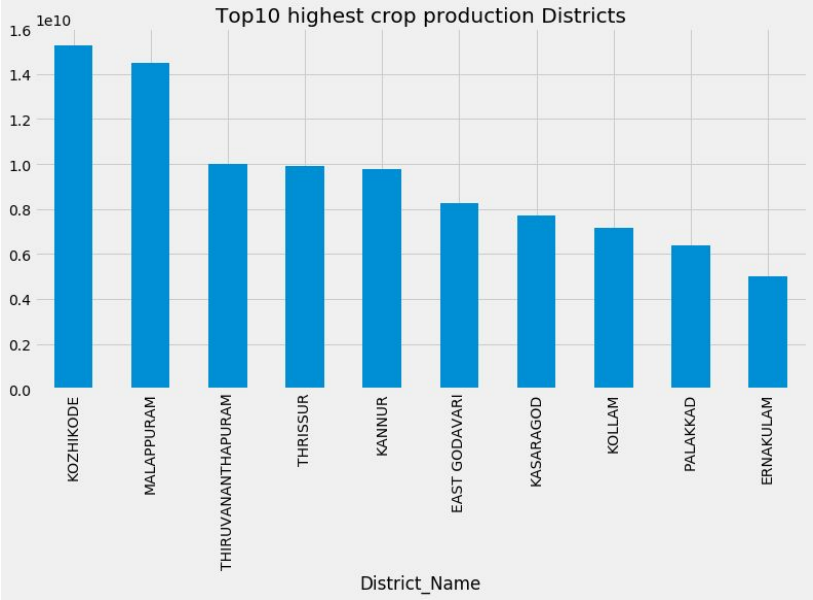
Coconut in Kerala, Andhra Pradesh, Tamil Nadu are the crops that giving the highest production in these 3 states. Below we can also see all the other states with the crops mostly produced in them.



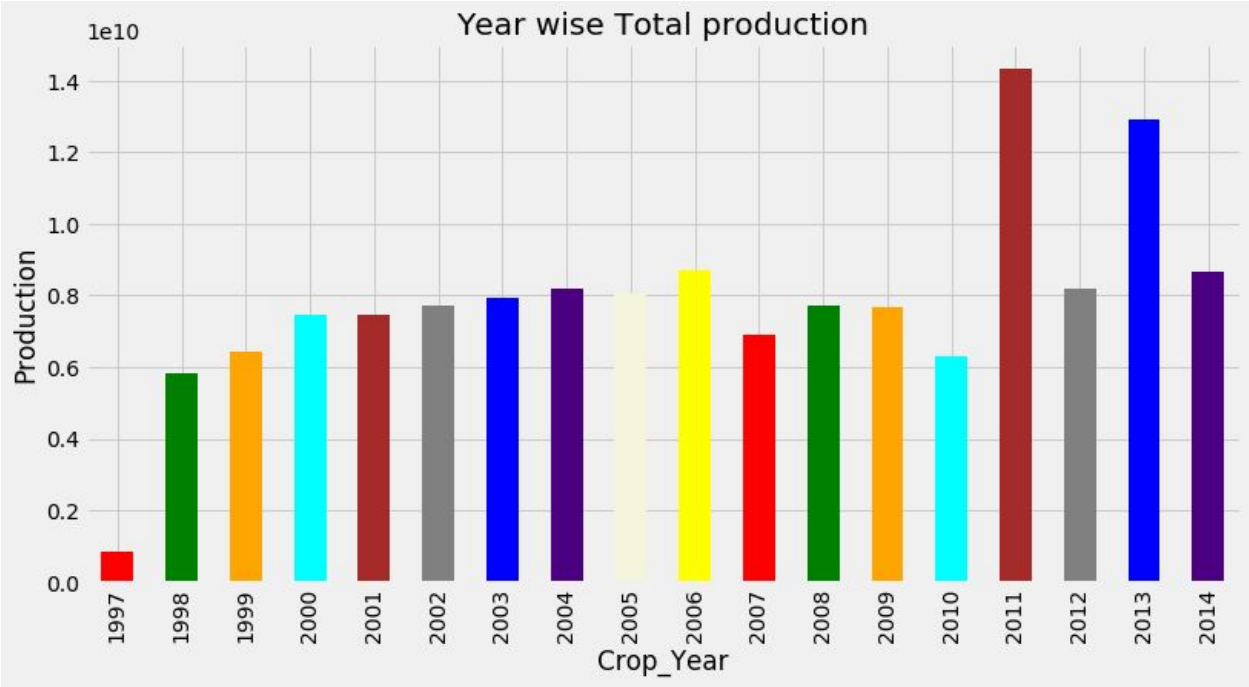
States from Telangana to Chandigarh with their crop production are given below.



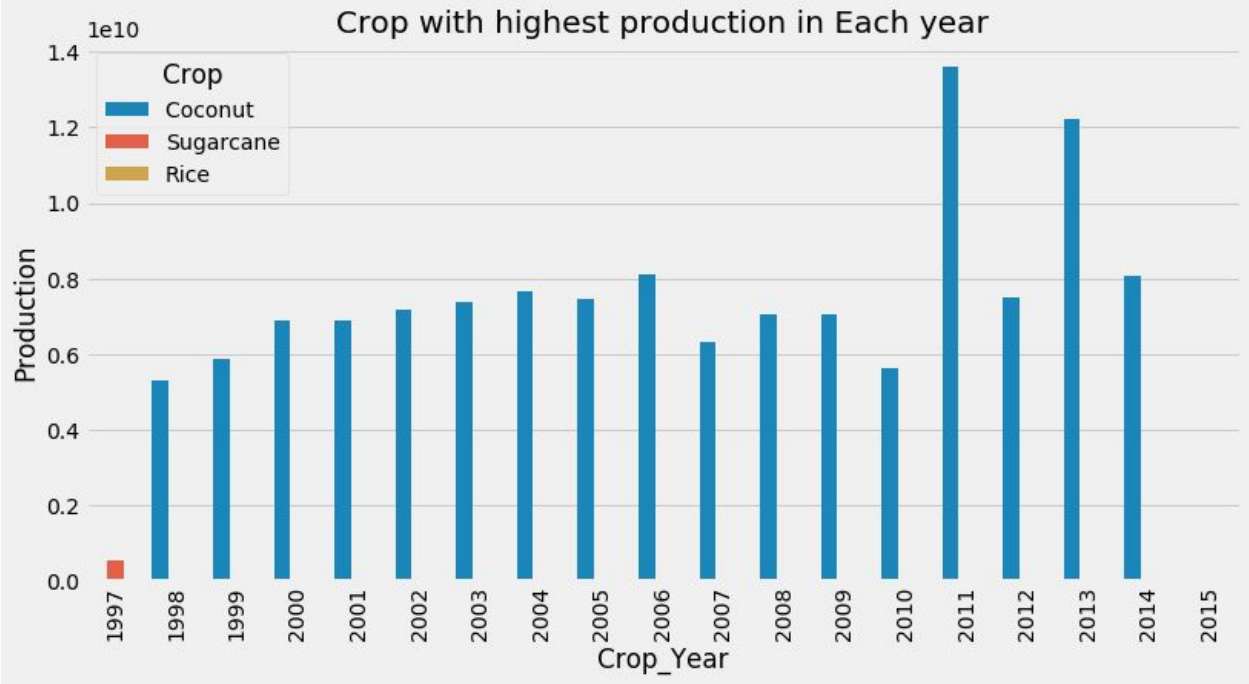
After looking at states with the highest crop production, we might be thinking about districts. We have the Top 4 highest crop production districts from Kerala are Kozhikode, Malappuram, Thiruvananthapuram, and Thrissur. we can also see from the graph the Top 10 highest Crop production Districts in India from 1997 to 2015.



Now How much crop production is happening every year. We have more crop production that occurred in the year 2011 and in the year 2013.

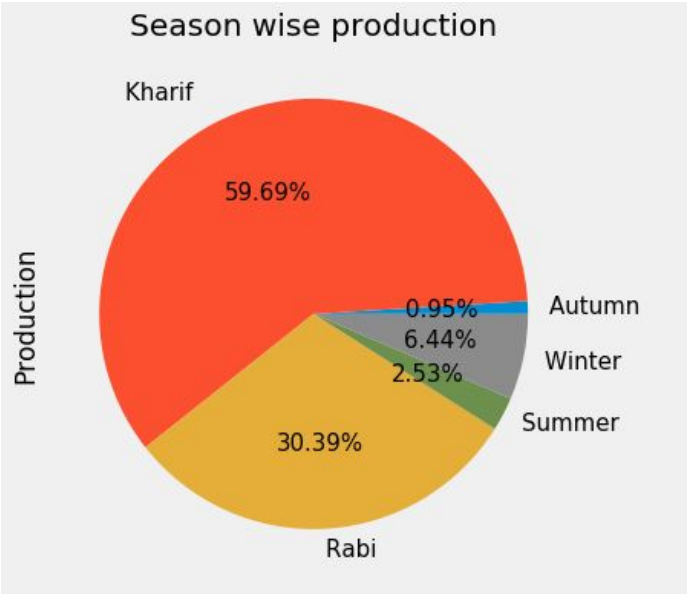


Further, it would be interesting to see which crop is giving the highest production every year.



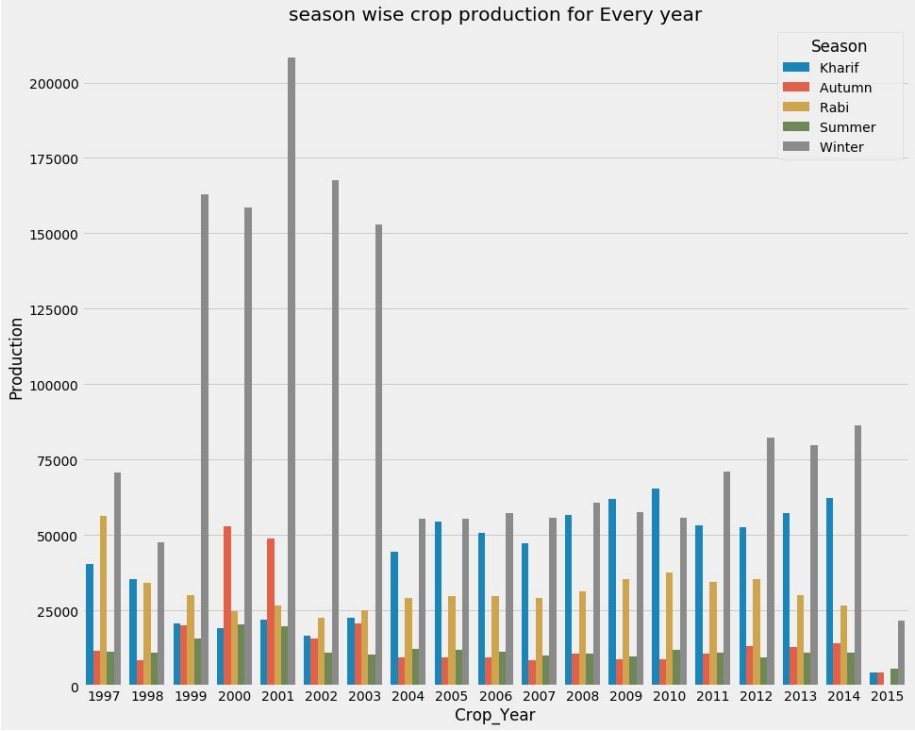
As we have only a few samples of the year 2015 with only 34 districts out of 500 approximately for all the other years. But in 1997 we can see that sugar cane is mostly produced in that year.

If we could look at the Season wise crop production, we will easily identify that Kharif the season with 59.69% production from 1997 to 2015.

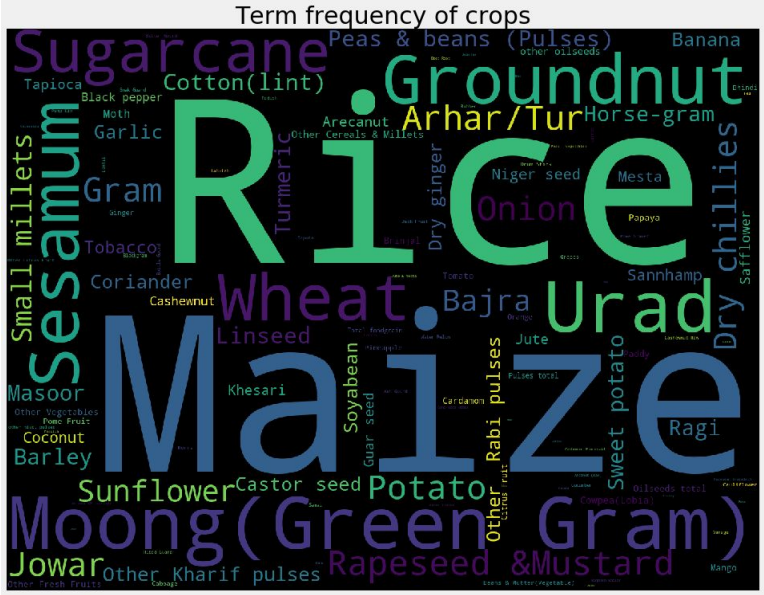


Overall we have 6 seasons named Kharif, Rabi, Summer, Winter, Autumn and Whole Year. including Whole year as the crop has 95.22 % crop production overall the years.

But in which season we are getting more crop production. Winter is the season in every year giving more crop production other than all.

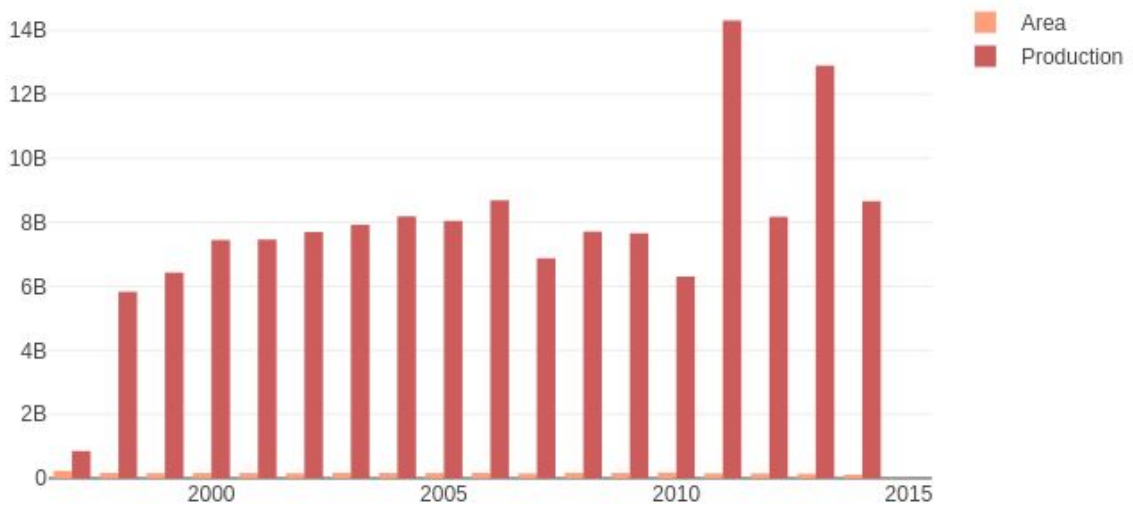


And from the word cloud, we can see that Most grown crop is Rice, with Maize in Second places.



Year-wise for how much production is happening with respect to Area. The below graph talks about Area vs production in all the years.

Area vs Crop production 1997 - 2015



Model:-

Overall we got an idea of the Indian Agricultural crop production system. Out of all the features we have, we choose Season, Crop, Area to predict the Crop Production.

As the problem is the regression model, let's use all the available Regression models. Before that make sure the whole data is split to Train and Test sets of 70:30 ratio. We have applied Linear Regression, Ridge regression, Lasso Regression, and Random Forest Regression.

Now as our values in Features are categorical, use Categorical encoding to convert categories to integers. Once converted, use One-Hot encoding to make the model to get a better understanding of data.

Out of all, we got Random Forest with more accuracy of 61.57%. As we have values with a Huge range from Thousands to ones, apply Standard Scaling/Feature scaling to bring all the values into the same scale. Now Random Forest Regressor accuracy increased from 61.57% to 63.13%.

We also did Hyperparameter tuning with the estimator to 100, and we got 1% accuracy increase. We also have another Regression model named XGBRegressor which is the Ensemble method of Boosting.
After applying XGBRegressor we got the accuracy of 66.85% accuracy. We also performed Hyperparameter tuning with learning rate to 0.01, estimators to 1200, depth to 4 which increased the model accuracy to 67.70%.

Conclusion:-

We want our model to predict the Crop Production depending on the Season, Crop, Area. we applied all the available models and Finally, XGBRegressor is the model with the Highest accuracy of 67.70% than the other models in predicting crop production.