Crop production analysis

By Unmilan Baruah

Introduction

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 report uses a dataset which 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. The following section makes a brief mention about this report. Finally, observations are recorded and conclusions are made.

Purpose

- To analyze crop production based on Indian states.
- To study crop production on yearly basis with the help of time series analysis.
- To analyze crop production based on seasons.
- To understand the distribution of crops across India.
- To analyze distribution of crops based on seasons.
- To study distribution of crops on yearly basis with the help of time series analysis.

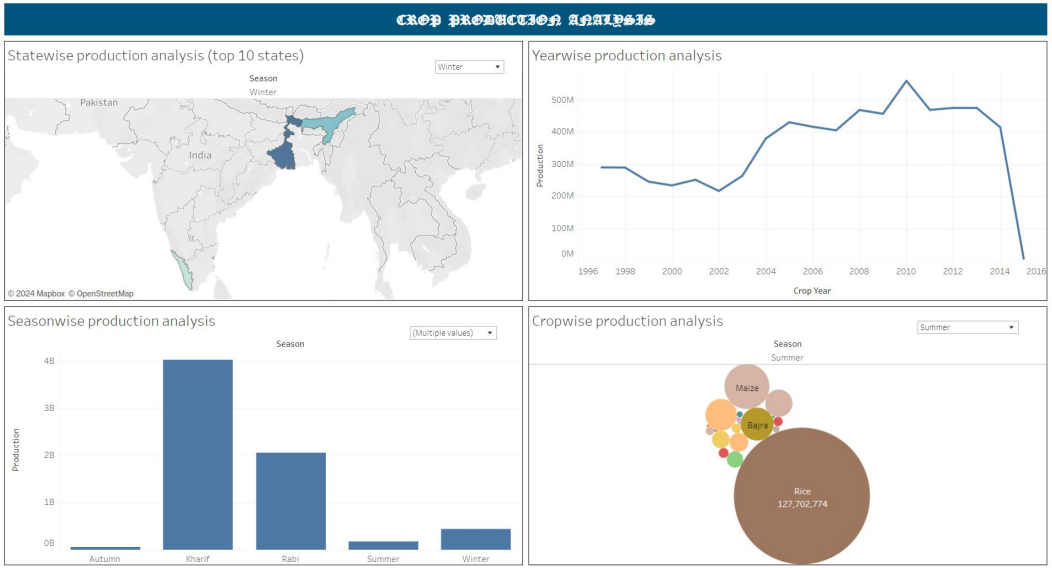
Methodology

1. ETL (Extract, Transform, and Load): Collecting the data, transforming it into something which can be analyzed with accuracy, and loading the processed data for further analysis.
2. Data Analysis
3. Observations
4. Inference

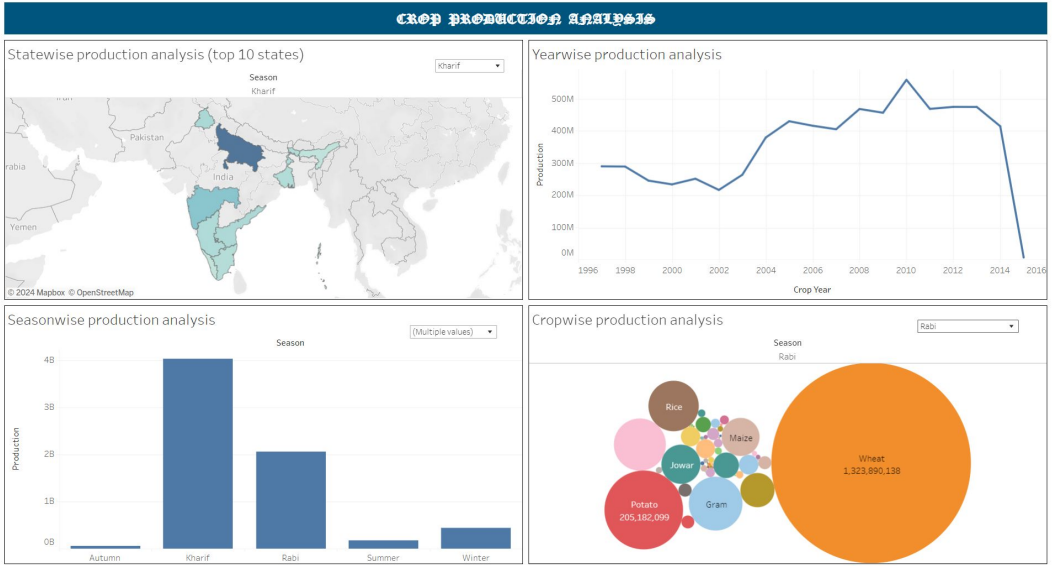
ETL (Extract, Transform, and Load)

- The data source is named as “**Crop production data.csv**” and this CSV file records related to production of crops.
- There are total **246091** entries, and there are missing values in the **Production** column. Therefore, we have to exclude them while doing production analysis using both Python as well as Tableau.
- Moreover, outliers can also be observed in the **Area**, and **Production** column. However, for the accuracy of the analysis and to avoid bias, we won’t remove the outliers. Also, we aren’t making any machine learning model for performing predictive analysis. So, removing the outliers is not an appropriate option.

Data Analysis



Data Analysis



Observations

- Overall production is highest in state of **Uttar Pradesh** during the **Kharif**, and **Rabi** season. While, during the season of **Autumn**, **Summer**, and **Winter**, the production is the highest in **West Bengal**.
- The overall production is highest during the season of **Kharif**, and **Rabi**. Moreover, the production is above 2 Billion during both the seasons.
- Also, overall production of **Wheat**, and **Sugarcane** is the highest during the season of **Rabi**, and **Kharif**, respectively. While the production of **Rice** is the highest during the season of **Autumn**, **Summer**, and **Winter**.
- Overall production of crops is the highest during the year **2010**. While the lowest production is observed during the year **2015**.

Inference

- Clearly, the production of crops is the highest in the state of **Uttar Pradesh**. This is because of the fact that production of crops is significantly higher in **Kharif**, and **Rabi** season.
- Production of crops is visible, for all the five seasons, in the states of **Kolkata**, and **Assam**. This is also evident from the fact that, along with **Kerala**, production is observed in these two states during the harsh **Winter** season.
- In all the five seasons, production of **Sugarcane**, **Rice**, and **Wheat** is significantly higher than all other crops. Overall production of above 1 Billion is observed for each of these three crops. This implies that, people in India are fond of **Sugarcane**, **Rice**, and **Wheat**.

Thank You