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.
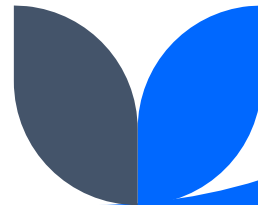
Data Analysis
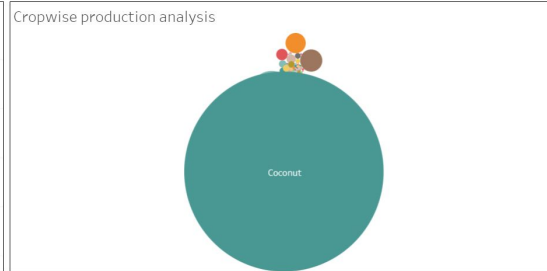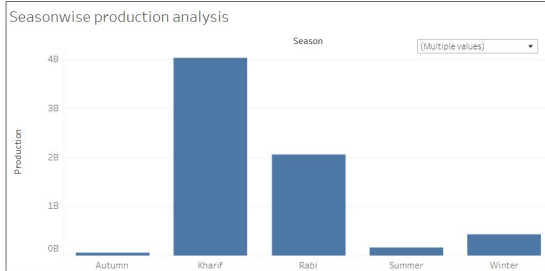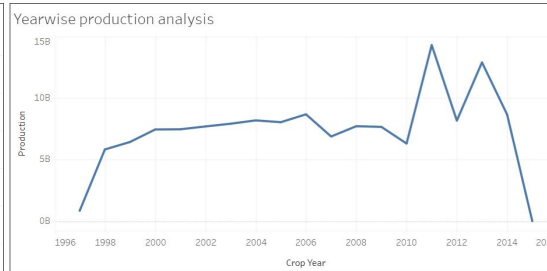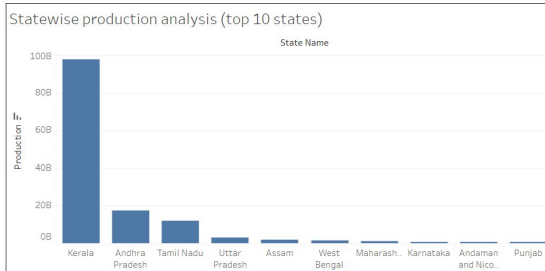
Data Analysis
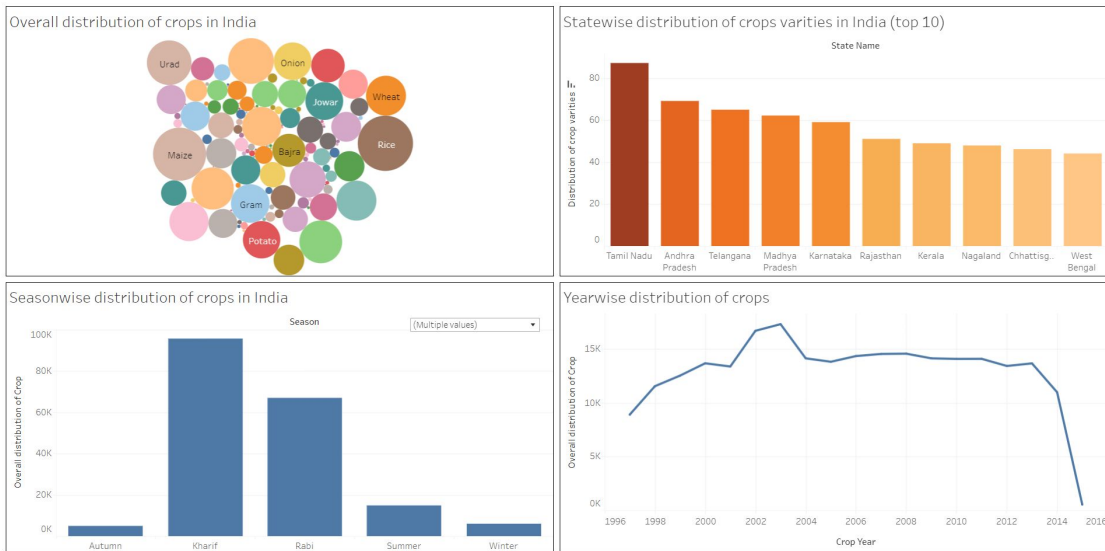
Observations

- Production is highest in **Kerala** state, followed by **Andhra Pradesh**. Top three states with highest production are in South India.
- Excluding the **Whole Year**, the production is highest in **Kharif** season and followed by **Rabi** season.
- The production is highest for the year **2011**, followed by the year **2013**. The production is found out to be the least in **2015**.
- Production of **Coconut** is the highest overall, followed by **Rice** and then by **Wheat**.
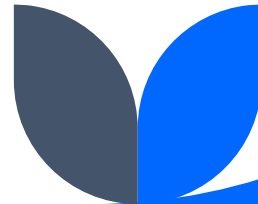
Observations

- The analysis clearly indicates that crop distribution is highest for **Rice**, and followed by **Maize**.
- When it comes to producing varieties of crops, then **Tamil Nadu** is the highest followed by **Andhra Pradesh**.
- Based on seasons and excluding the **Whole Year**, the distribution of crops is observed to be the highest in **Kharif** season, followed by **Rabi**.
- Distribution of crop is highest in the year **2003**, followed by the year **2002**.

Inference

- Even though the overall production is highest in **Kerala**, the production of different crop varieties is higher in **Andhra Pradesh** and **Tamil Nadu** compared to crop production in **Kerala**. Moreover, it is clear that production in **Kerala** is the highest only because of **Coconuts**.
- Crop with highest production is **Coconut**, but the most produced or the preferred crop in India is **Rice** and then it is **Maize**. High production of **Coconut** is mainly understandable from the fact that overall production is highest in **Kerala**.
- Both overall production and distribution of crops is highest in **Kharif**, followed by **Rabi** season. This implies that production of crops in India is mainly visible during the **Kharif** season, followed by **Rabi**.

Thank You