**Exploratory Data Analysis**

1. **Project explanation**

Agriculture sector is one of the most significant sectors of Indian Economy. It is a crucial contributor accounting about 15% of the GDP. In agriculture sector where farmers and agribusinesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them.

An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Exploratory Data Analysis are necessary for accomplishing practical and effective solutions for this problem. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions.

1. **Problem Statement Description**

This dataset provides a huge amount of information on crop production in India ranging from several years across different states in India. Based on the Information the ultimate goal would be to predict crop production using powerful machine learning techniques.

1. **Aim**

The main goal of Exploratory Data Analysis is to identify errors in data sets. Give a better understanding about the relationship between various attributes in the dataset. Thus, we will be able to schedule all the other processes accordingly.

1. **Business Context**

Historical crop yield information is also important for supply chain operation of companies engaged in industries. These industries use agricultural products as raw material, livestock, food, animal feed, chemical, poultry, fertilizer, pesticides, seed and paper.

 An accurate estimate of crop production helps these companies in planning supply chain decision like production scheduling.

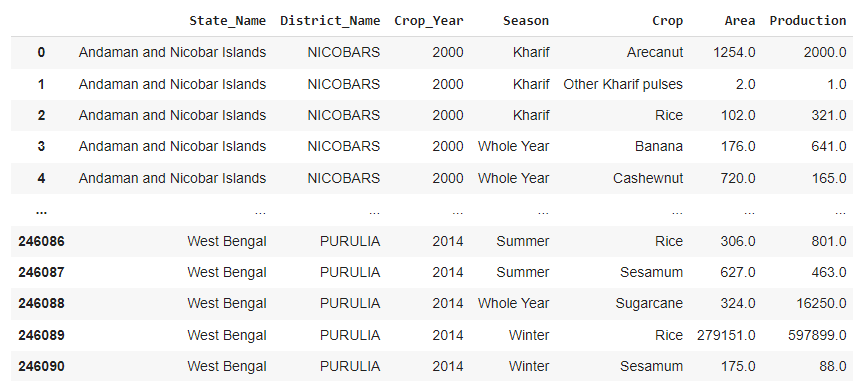
Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates.

There are 2 factors which are helpful for the farmers and the government in decision making namely:

1. It helps farmers in providing the historical crop yield record with a forecast reducing the risk management.
2. It helps the government in making crop insurance policies and policies for supply chain operation.

Exploratory Data Analysis plays a vital role in the analysis of data. Exploratory Data Analysis is the computing process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database system.

1. **Data Explanation.**

****

This is the basic view of the dataset.

**State-wise Crop Production in India for 2000 to 2014**

It also gave information about different seasonal crops at district level and area of cultivation along with total crop production. India being agriculture rich country, this data will have lots of minor and major facts which will help in charting a next successful agriculture revolution after 1965.

Doing an exploratory data analysis of this dataset would give insights into Indian agriculture status: state-wise, district-wise, crop-wise, area-wise and levels of productions. A complete analysis will paint a beautiful story of this important aspect of India.

**Dimensions of the dataset is 246091 rows and 7 columns.**

| **Colum’s** | **Description**: |
| --- | --- |
| State\_Name | Name of the respective states in India |
| District\_Name | Name of the respective Districts in India |
| Crop\_Year | Data is recorded in between the year of 2000 - 2014 |
| Season | This include all the season that is present in various part of the India. |
| Crops | The variety of Crops that are grown in different part of India during different seasons. |
| Area | Area at which a particular crop is grown during a particular time of the year and the unit is in hectare |
| Production | Production is total harvest measured in tons per hectare |

Seasons:

| **Seasons** | **Duration** |
| --- | --- |
| Kharif | Kharif crops are sown early-May – usually at the beginning of the first monsoon rains |
| Rabi | Rabi crops are sown around mid-November – preferably after the monsoon rains. |
| Whole Year | Throughout the Year. |
| Summer | April to June |
| Winter | December to January |
| Autumn | September and October |

1. **Learning Outcomes**

* Should be able to discover patterns in data
* Should be able to spot anomalies
* Learn how to visualize the Data
* Able to manipulate data according to the needs

1. **Skills** **Required**

* Problem Solving
* Good Analytic Skills
* Data visualization
* Coding Skills such as Python and SQL.