

Agriculture Data Analytics in Crop Yield Estimation

Using IBM Cognos



Submission for the partial fulfilment of

Project Build-A-Thon - An Exclusive Program for Faculty of India

July-2022

Submitted by

Dr. Sandeep K. Raghuwanshi

(sraghuwanshi@gmail.com)

1. Introduction

India is known to be a land of farmers where the major part of GDP is depends over agriculture. About 70% population of India come from rural area and almost 58% of them depends on agriculture for their livelihood. This sector contributes around 15-20% to the national GDP over a few decades. This field is very diverse and different crops depends over environmental conditions and seasons. A proper analysis of the agriculture crops, with area, season and production can help farmers to maximum profits.

1.1 Overview of Project

This project is an effort to analyse crop production data and produce visualization to understand the insights. The idea is to understand crop production in the different parts of country with seasons. This might help to setup food processing industries at the right place to maximise profits.

1.2 Objectives/Purpose

Followings are the key objectives of this project:

- a. Understanding the broad applications of different visualizations pattern and how to draw them using IBM Cognos Analytics.
- b. Creating meaningful and dynamic dashboard to get maximise insight of dataset.
- c. Understanding the relation between average production of crop and season.
- d. Year wise usage of area and production of crops.
- e. Which are the top-10 states with most area for crop production.
- f. How crops can be categories state wise in the country.
- g. Categorisation of states with crop production along with seasons.

2. Proposed Solution

Data Analytics enables organizations to understand their data to identify insight and generate meaningful information. This helps to make right decisions to solve a problem. This produces business success in many cases. Exploratory data analysis is one of the essential step of data analytics. It is the step to produce multiple chart/graphs for the data to understand their distribution (Univariate analysis) and sometime association ship between two variables (Bivariate analysis).

In the same line the proposed solution is also based on the analytics process where the aim to produce a dashboard of multiple charts to understand the pattern of data to get insights and meaningful information. Following figure shows one of the examples of dashboard.

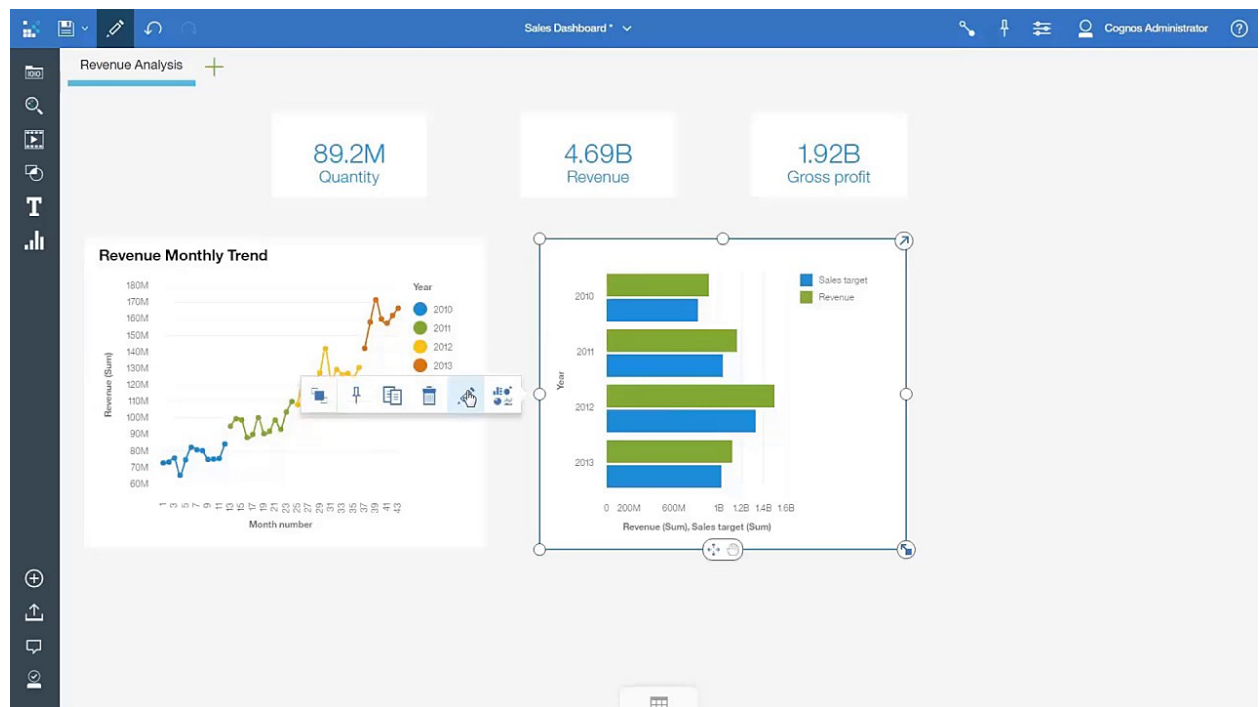


Figure 1: Dashboard to show sales target and revenue month wise

3. Experimental Setup (Tool and Data set)

Following diagram shows the experimental framework:

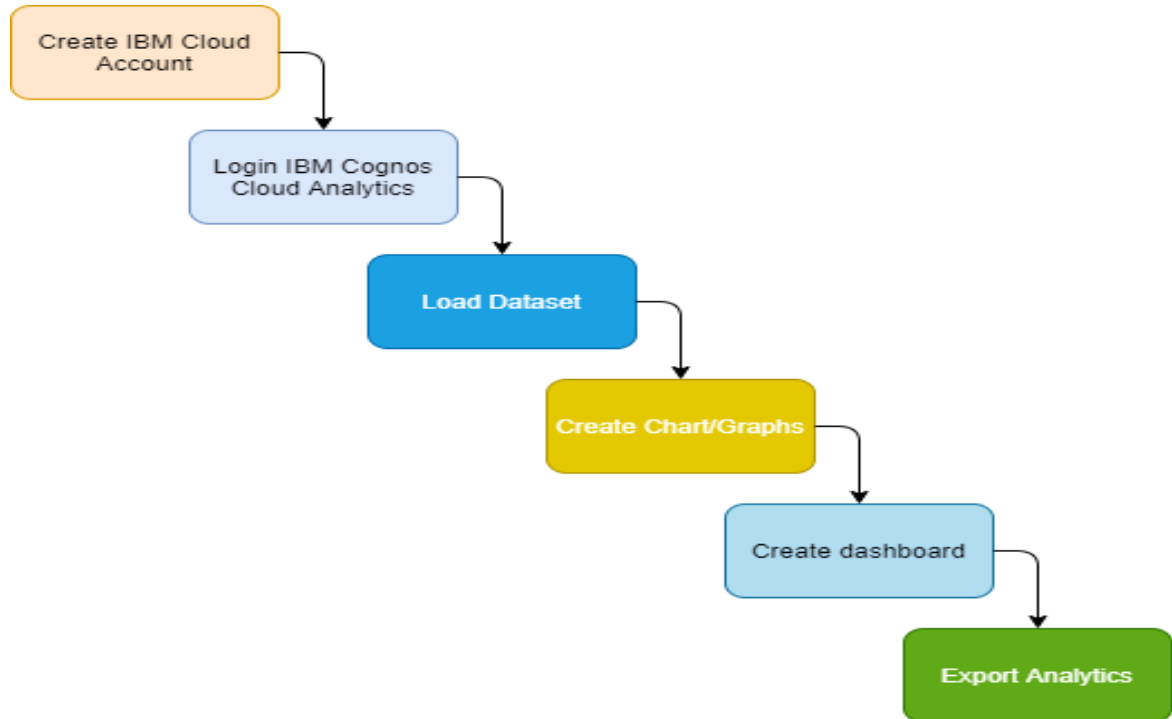


Figure 2: Frame work of experiment

IBM Cognos Analytics- IBM Cognos Analytics is business analytics tool to manage and analysing data intelligently. It is very rich tool in terms of features like data preparation, exploration and sharing the insights. Cognos analytics is a powerful tool which covers descriptive, explorative and predictive techniques to answer business questions. It is well versed with many statistical tests to perform data analysis.[1]



- Web based data modelling.
- Interactive dashboards and enterprise reports.
- Data exploration and prediction.
- Continuous update and support.

Dataset- (Crop Production in India) The data set is very rich in terms of information ranging from several year, areas and seasons. The data set contains 246091 entries over seven features as State_Name, District_Name, Crop_Year, Season, Crop, Area and Production. The summarized view of the dataset is present on following figure.[2]

```
RangeIndex: 246091 entries, 0 to 246090
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   State_Name      246091 non-null object
1   District_Name   246091 non-null object
2   Crop_Year       246091 non-null int64
3   Season          246091 non-null object
4   Crop            246091 non-null object
5   Area            246091 non-null float64
6   Production      242361 non-null float64
dtypes: float64(2), int64(1), object(4)
memory usage: 13.1+ MB
```

Crop Production in India Dataset

Figure 3

4. Project Flow

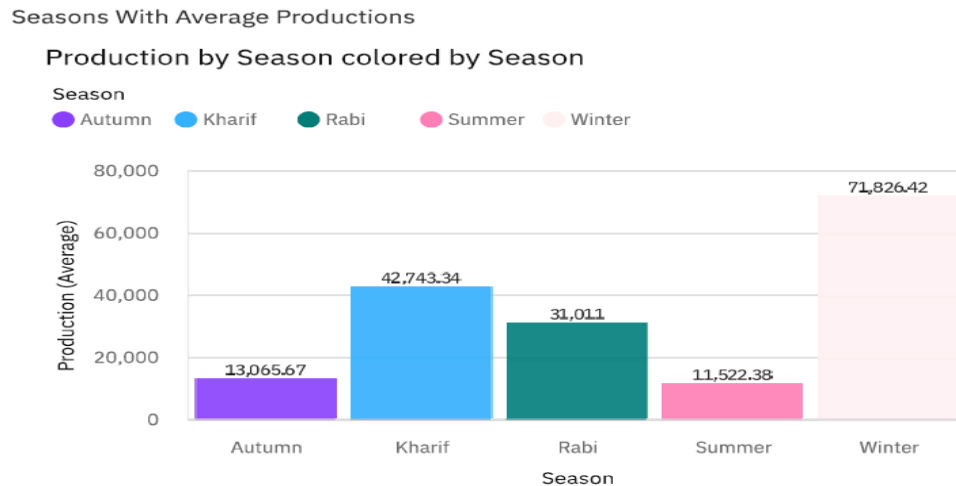
To accomplish the project, we have to create multiple chart/graphs and a dynamic dashboard which summarizes all visualizations at one place. Following steps can summarize the complete project flow:

- Create IBM cloud Account
- Login to IBM Cognos Analytics Cloud
- Data Set (Crop Production)
 - Understand the dataset
 - Load the data on to cloud
- Create Visualization chart/Graphs
 - Seasons with average productions
 - With years usage of Area and Production
 - Top 10 States with most area
 - State with crop production
 - States with the crop production along with season (Text Table)
- Create dashboard
- Export the Analytics

5. Outcome

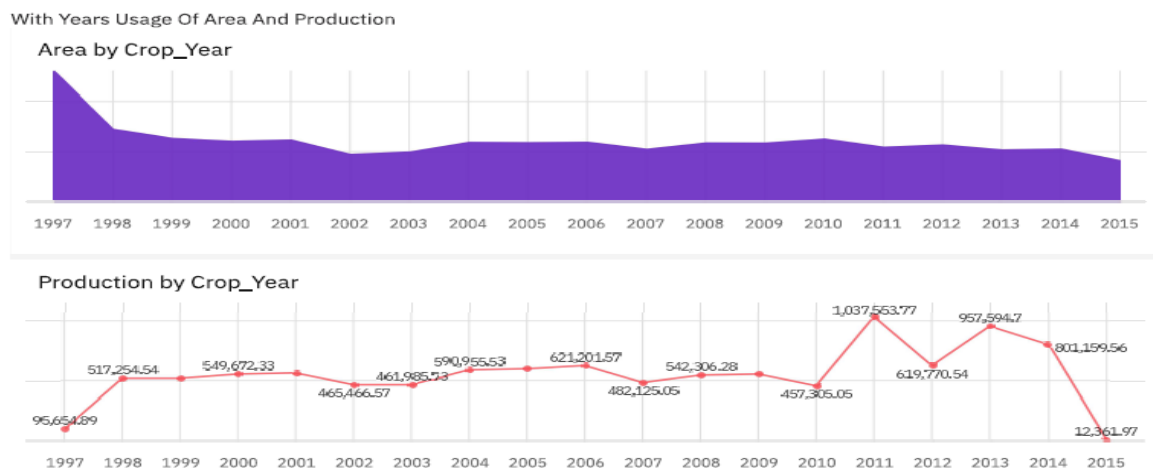
a. Seasons with average productions

The Bar chart present the average production of crop with respect to different season. It is clear from the figure that the maximum average production from winter season. (Excluding whole year column).



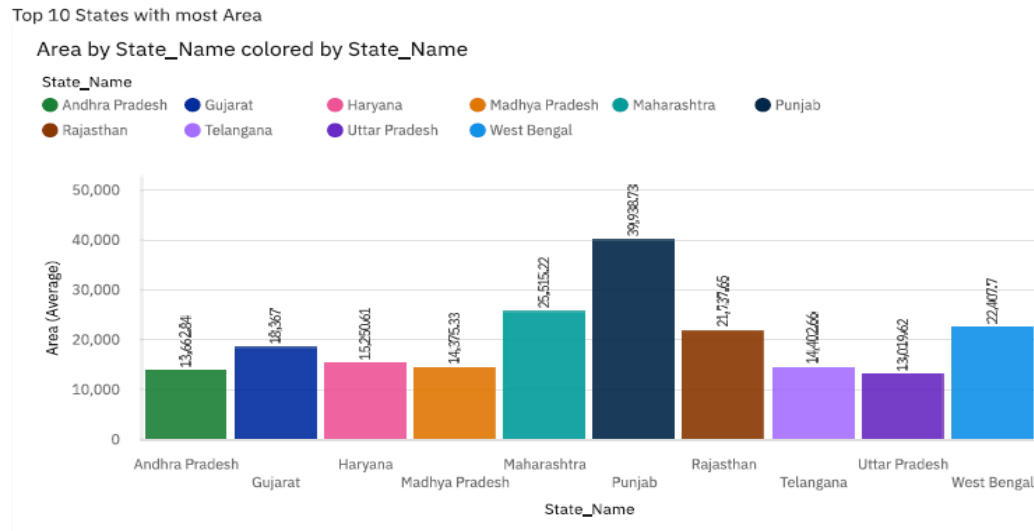
b. With years usage of Area and Production

It is evident from the figure that the highest average production was in the year 2011 and maximum area for the crop was used in 1997.



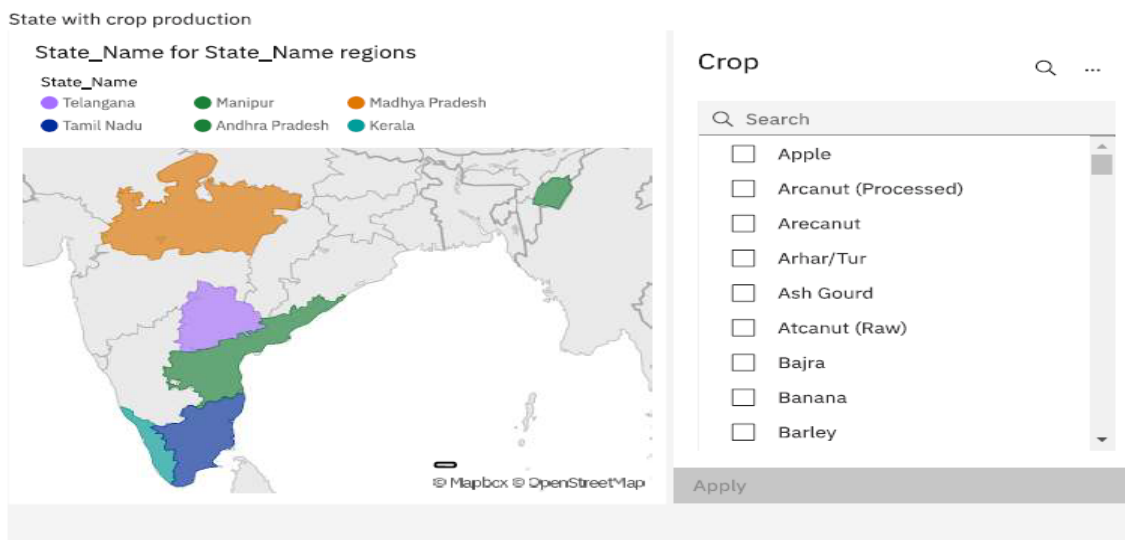
c. Top 10 States with most area

The bar chart shows the top-10 states in the country area wise. Highest crop area is in Punjab and after that Maharashtra is second in tally.



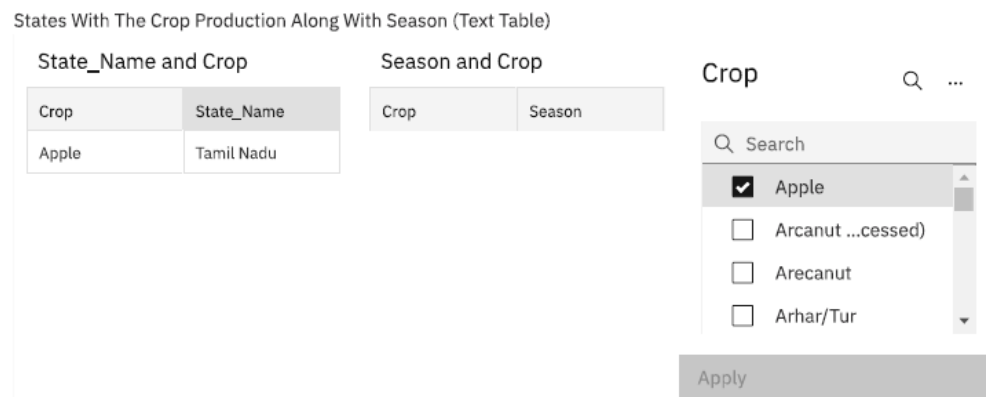
d. State with crop production

This is dynamic visualization where one can select crop and graph displayed corresponding states.



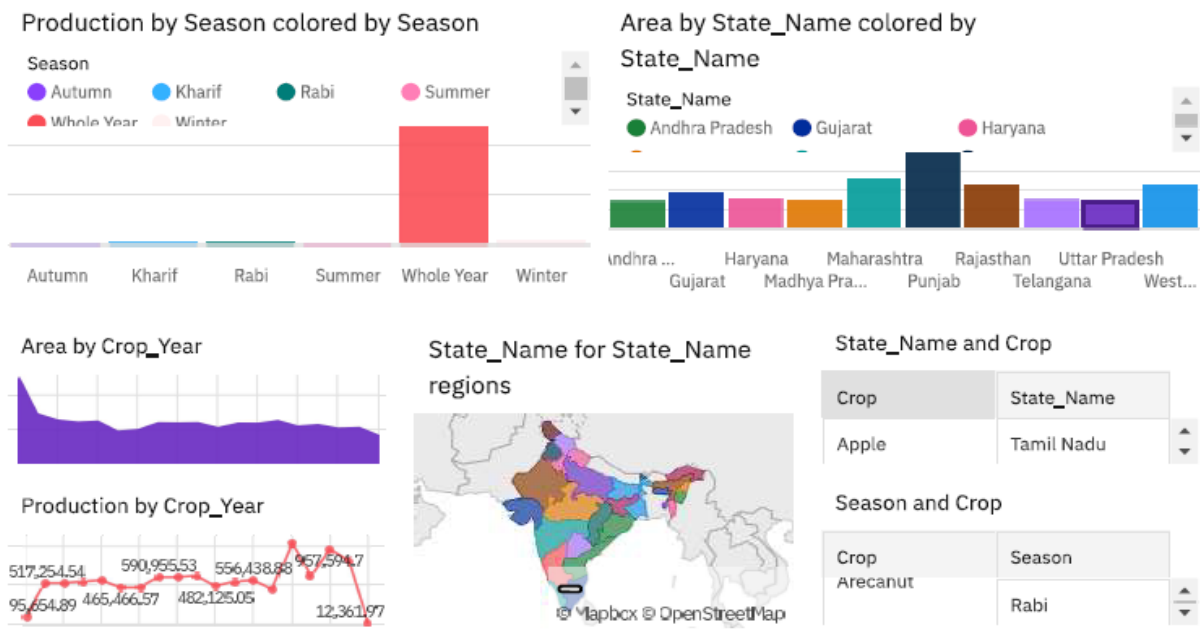
e. States with the crop production along with season (Text Table)

Sometime the information can be filtered in a table form and more presentable. Following graph does the same and filter the crop information with respect to State_Name and Season.



Final Dashboard

Dashboard



6. Conclusion and Future Scope

The final dashboard is beautiful and present different information in summarized way. It is dynamic visualization which changes all the chart/graphs which any filter applied to any of the graph. The Analytics helps to get the insight of various crop production with states and season. It helps to analyse the season pattern and can be used understand the possibility that one particular crop can be produce in different part of country. State wise analytics helps to understand the business prospective of a particular crop with industry setup.

More visualizations can be plotted to get more insights of data and to understand the pattern of crop production in different parts of country, season and area wise. (i.e., yield analysis.), Are there specific district within state contribute most of the production, How the land area is changing over the years etc.

7. References

[1] www.cloud.ibm.com

[2] <https://www.kaggle.com/datasets/abhinand05/crop-production-in-india>
