

**SmartInternz - Project Build-A-Thon 2022**  
**Agriculture Data Analytics in Crop Yield**  
**Estimation using IBM Cognos**

**Dr. M. Karthikeyan**

Cohort 1

Application id:SPS\_APL\_20220093083

## **CONTENTS**

### **1. INTRODUCTION**

1.1. Overview

1.2. Objective

### **2. PROBLEM STATEMENT**

2.1. Present Problem

2.2. Proposed Solution

### **3. PROJECT FLOW**

3.1. Account Setup

3.2. Working With the Dataset

3.2.1. Understanding The Dataset

3.2.2. Loading The Dataset

3.3. Data Visualization Charts

3.3.1. Seasons With Average Productions

3.3.2. Year Wise Usage of Area and Production

3.3.3. Top 10 States with Most Area

3.3.4. State Wise Crop Production

3.3.5. States With the Crop Production Along with Season

3.4. Creating the Dashboard

3.5. Exporting the Analytics

### **4. CONCLUSION**

## **1. INTRODUCTION**

Agriculture is crucial for human survival since it supplies the food which is one of the basic necessities. It is a well-known fact that the majority of population ( $\geq 55$  percent) in India is engaged in agriculture. Due to changes in climatic conditions, there exist obstacles for expanding the crop yield in India. It has become increasingly difficult to reach desired targets in agricultural yield. Factors like climate, geographical conditions, economic and political situations are to be examined which have direct impact on the production, productivity of the crops.

Crop yield prediction is one of the major variables in agriculture techniques. Farmers need knowledge regarding agricultural yield before sowing seeds in their farms to obtain better crop yield. The usage of technology in agriculture has expanded in recent years and data analytics is one such trend that has permeated into the agriculture industry being utilised for management of crop yield and monitoring crop health. The latest advances in the sector of agriculture have made the people to recognise the relevance of Big data. The biggest problem with big data in agriculture is identification of impact and effectiveness of big data analytics.

Efforts are going undertaken to understand how data analytics can be used to improve the productivity in agricultural activities. The study of data connected to agriculture aids in agricultural yield forecast, crop health monitoring and other such related operations. In literature, there exist various research connected to the application of data analytics in the agriculture industry. The present study delivers insights on several data analytics methodologies applied to crop yield prediction.

### **1.1. Overview**

Crop production in India is one of the key sources of income and India is one of the top countries to produce crops. As per this project we will be evaluating certain crucial graphics, making a dashboard and by going through them we will obtain most of the insights on agricultural production in India.

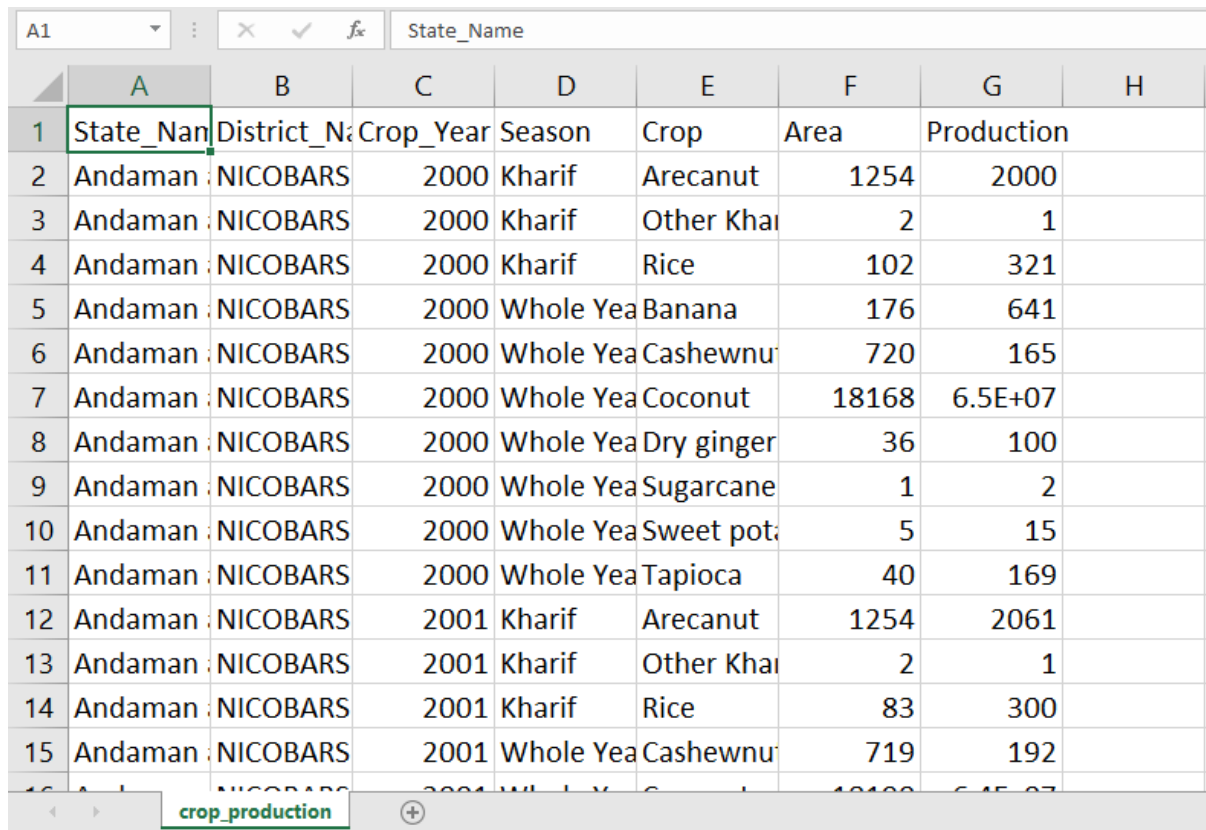
### **1.2. Objective**

The major objective of this project is to visualise data relevant to crop production in India on a dashboard featuring different types of plots produced using IBM Cognos Analytics.

## 2. PROBLEM STATEMENT

### 2.1. Present Problem

Although, we have plenty of data regarding Agricultural crop production in India, this data in its raw form is useless since the raw data itself does not give any insight which can be used for meaningful purposes.



	A	B	C	D	E	F	G	H
1	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production	
2	Andaman	NICOBARS	2000	Kharif	Arecanut	1254	2000	
3	Andaman	NICOBARS	2000	Kharif	Other Kharif	2	1	
4	Andaman	NICOBARS	2000	Kharif	Rice	102	321	
5	Andaman	NICOBARS	2000	Whole Year	Banana	176	641	
6	Andaman	NICOBARS	2000	Whole Year	Cashewnut	720	165	
7	Andaman	NICOBARS	2000	Whole Year	Coconut	18168	6.5E+07	
8	Andaman	NICOBARS	2000	Whole Year	Dry ginger	36	100	
9	Andaman	NICOBARS	2000	Whole Year	Sugarcane	1	2	
10	Andaman	NICOBARS	2000	Whole Year	Sweet potato	5	15	
11	Andaman	NICOBARS	2000	Whole Year	Tapioca	40	169	
12	Andaman	NICOBARS	2001	Kharif	Arecanut	1254	2061	
13	Andaman	NICOBARS	2001	Kharif	Other Kharif	2	1	
14	Andaman	NICOBARS	2001	Kharif	Rice	83	300	
15	Andaman	NICOBARS	2001	Whole Year	Cashewnut	719	192	
16	Andaman	NICOBARS	2001	Whole Year	Coconut	18168	6.5E+07	

As seen above, the raw data for crop production in India contains 246092 data points for 6 features which when represented by using a spreadsheet program doesnot give any valuable insight and is difficult to understand.

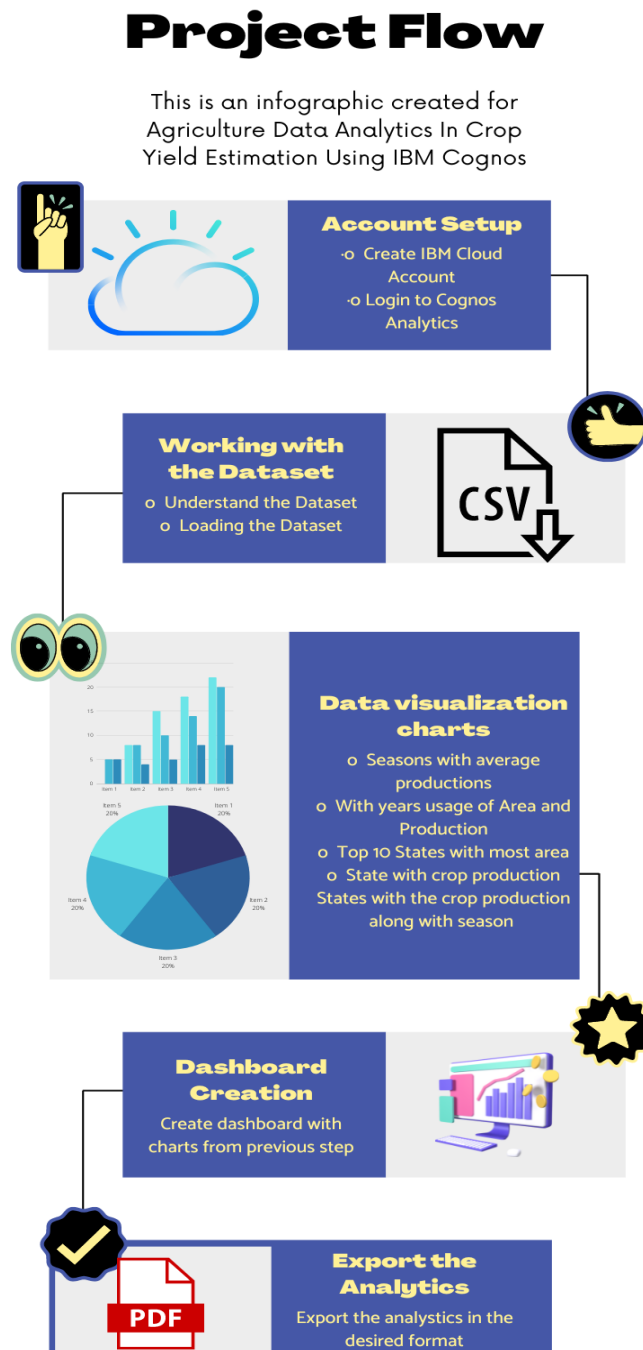
### 2.2. Proposed Solution

A graphical solution is proposed in form of dashboard which represents important information such as seasons with average productions, year wise usage of areas and production, top 10 states with most area, state wise crop production, states with the crop production along with season which can be filtered based on crop in an unambiguous, visually appealing format.

This will make the task of decision making easier for concerned people such as farmers and policy makers.

### 3. PROJECT FLOW

The project flow is explained in the following infographic.



### **3.1. Account Setup**

In order to create dashboard, the IBM account is setup as follows.

IBM Account was created and logged-in through

Link: <https://cloud.ibm.com/registration>

Create Cognos Analytics Account.

Link: [IBM Cognos Account Creation](#)

### **3.2. Working With the Dataset**

In order to analyze the data, we need to download, understand and load the dataset to IBM Cognos Analytics.

#### **3.2.1. Understanding The Dataset**

The dataset is downloaded from Kaggle using the following link.

<https://www.kaggle.com/abhinand05/crop-production-in-india>

The downloaded from Kaggle is extracted to the working folder.

This project is based on understanding the crop production of India. It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

The features given are as follows.

1. State Name - All the Indian State names.
2. District Name - Different District names.
3. Crop Year- contains the crop years.
4. Season – Different seasons for crop production.
5. Area- Total number of areas covered.
6. Production- production of crops.

### 3.2.2. Loading The Dataset

Before you can build a view and analyze your data, you must first connect the data to IBM Cognos. Cognos supports connecting to a wide variety of data, stored in a variety of places.

The data might be stored on your computer in a spreadsheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in your enterprise.

In our case, we will be using csv file downloaded from Kaggle for making our analysis.

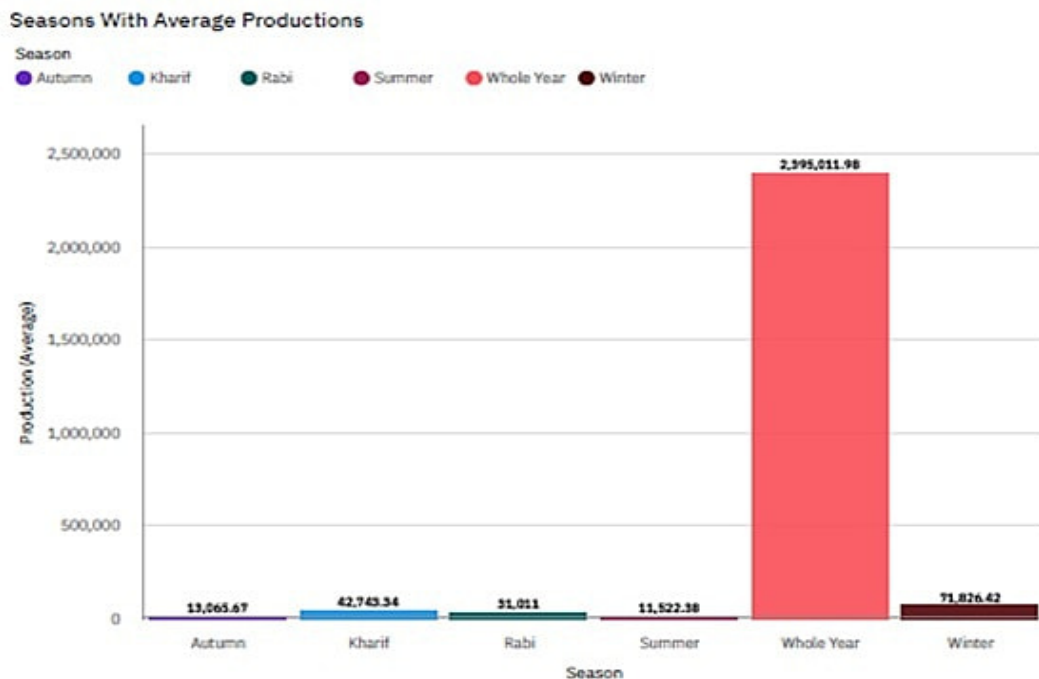
This is done by clicking on clicking on hamburger icon in the top left corner of the IBM Cognos Analytics home page and clicking on upload files. Then, select unzipped csv file from the working location and upload the same.

### 3.3. Data Visualization Charts

Using the Crop production in Indian dataset, we plan to create various graphs and charts to highlight the insights and visualizations. To create the charts, we have to first create a new dashboard. To do this, click on the hamburger icon in the top left corner and then click on New, Dashboard.

#### 3.3.1. Seasons With Average Productions

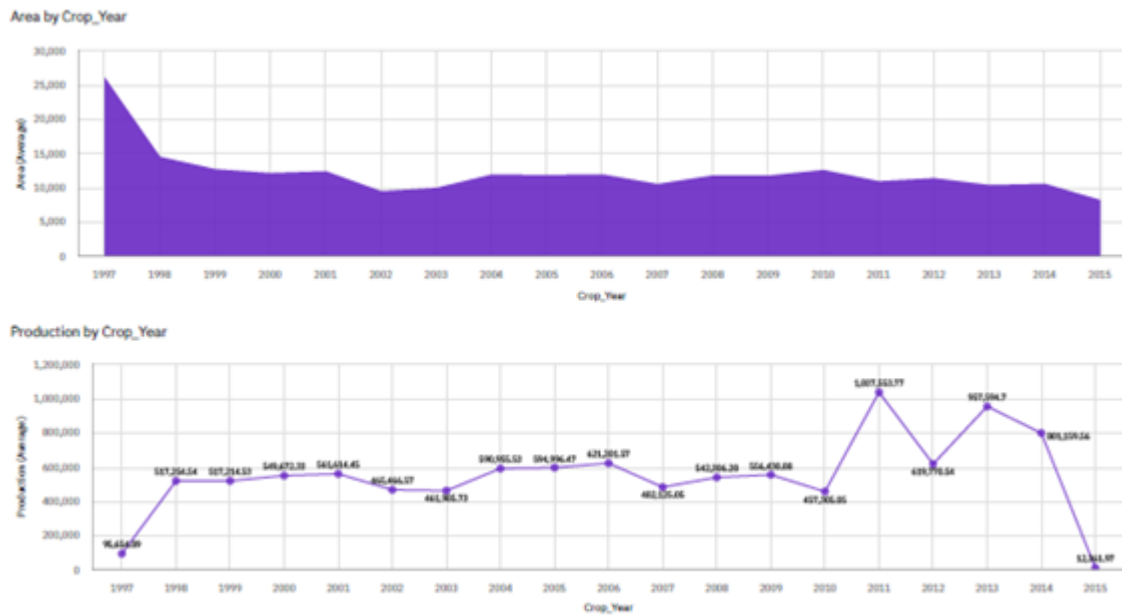
As production of crops depends on different seasons, so let's plot the graphs to visualize the average production based on different seasons. For this purpose, a column chart is used.



### 3.3.2. Year Wise Usage of Area and Production

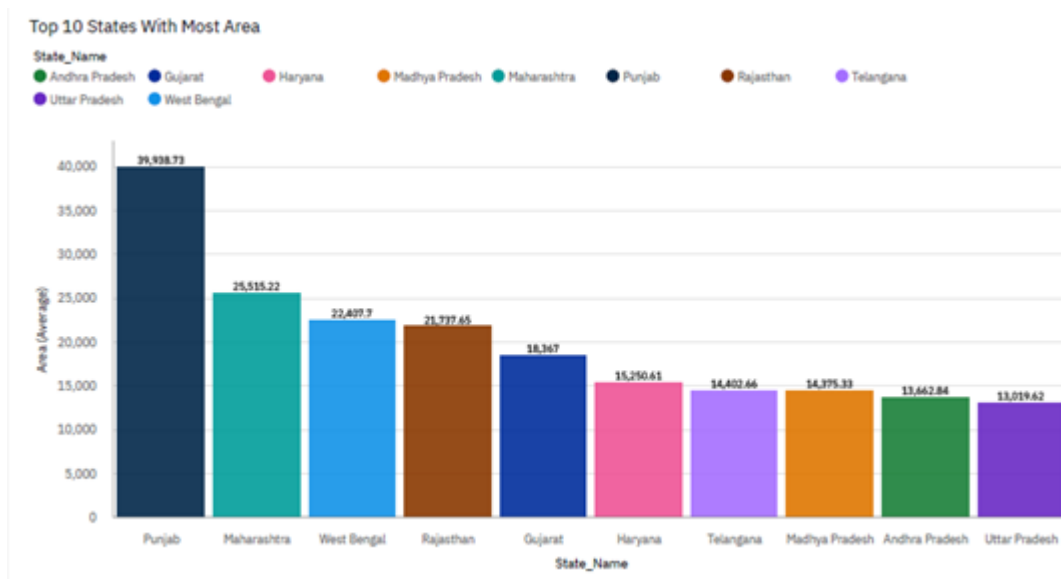
In our dataset we also have a year's columns by which we will plot a line and area graphs to see the change in these both data with respect to increase in years.

For this purpose, we use area chart for year wise usage of area and line chart for year wise production.



### 3.3.3. Top 10 States with Most Area

As we have an area data in our dataset, we will be plotting some graphs to visualize the top 10 Indian states with the most area. Again, we use column chart for this purpose and sort the same in descending order using the top 10 feature.

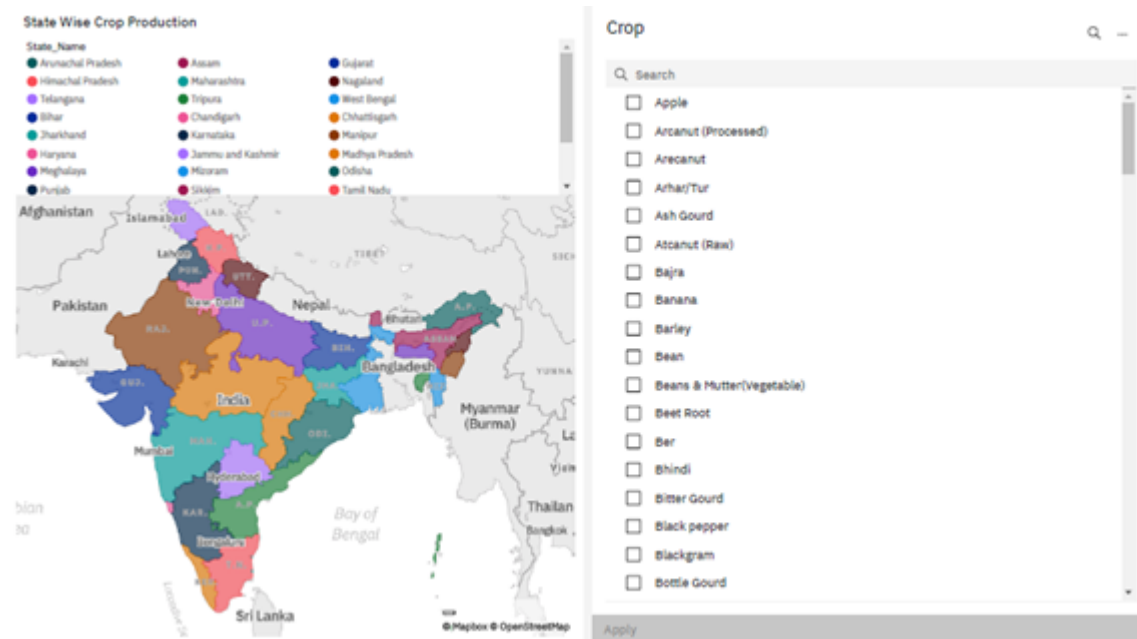




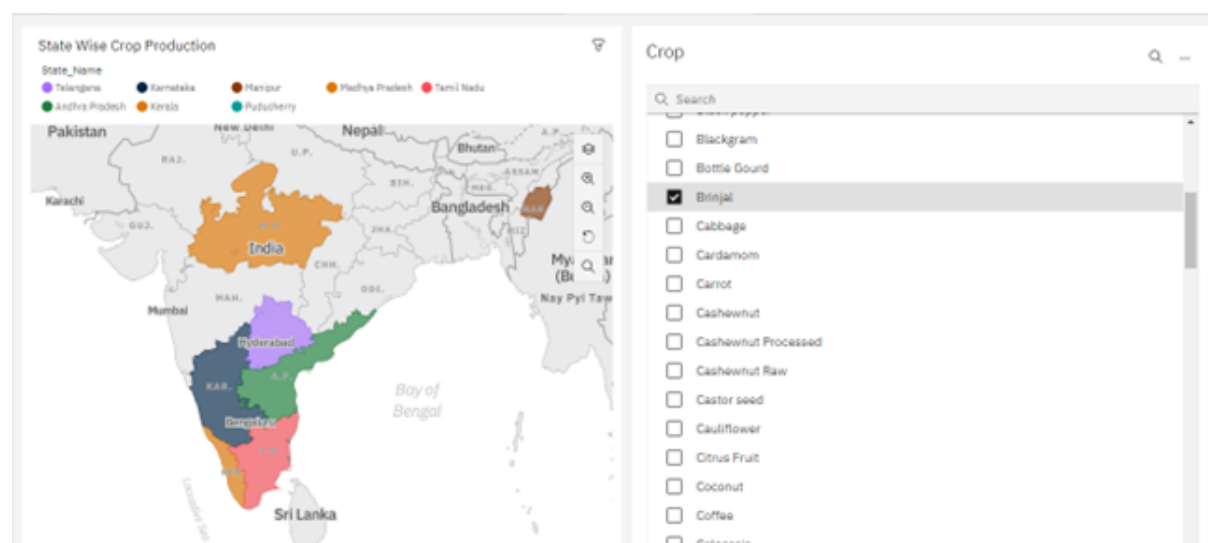
### 3.3.4. State Wise Crop Production

There are so many different crops produced in Indian and most of us don't know which crop is belongs to which state so we will be plotting and highlight the states in map according to different crops.

A map type chart is used for this purpose and a filtered is added to canvas based on the type of crop. This can be used to display which states produces a particular crop.



For example, if we select brinjal and click apply, the map will show the states producing different states that produce brinjal in different colours and the list of states is given in text format



### 3.3.5. States With the Crop Production Along with Season

Taking forward the previous plot we will be fetching the state name and showing it in a text table whenever different crops are chosen.

States With The Crop Production Along With Season

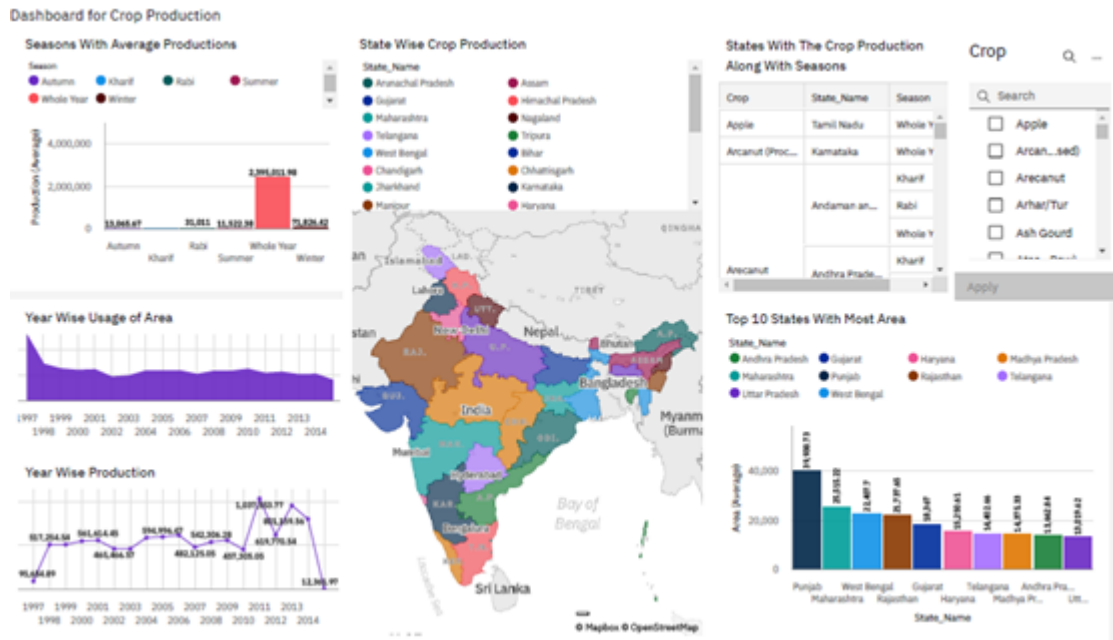
Crop	State_Name	Season
Apple	Tamil Nadu	Whole Year
Arcanut (Processed)	Karnataka	Whole Year
Arecanut	Andaman and Nicobar ...	Kharif
		Rabi
		Whole Year
	Andhra Pradesh	Kharif
		Whole Year
	Assam	Rabi

For example, when brinjal is selected, the table shows a list of states that produce brinjal and the seasons in which it is produced in each state.

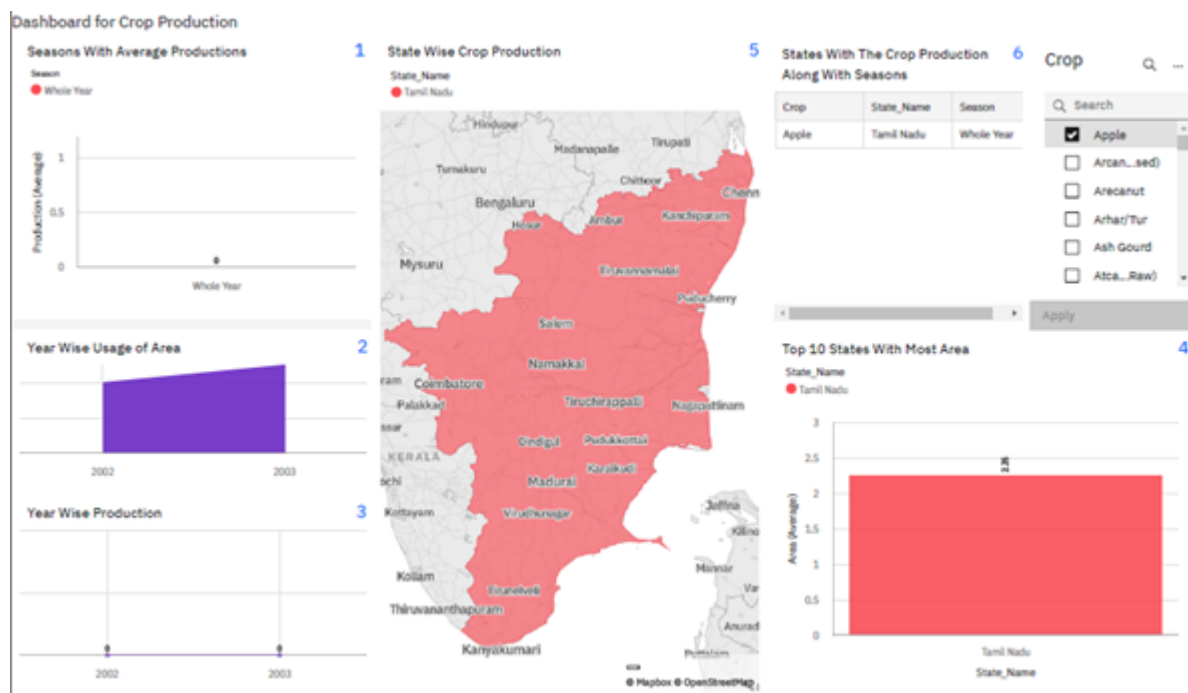
States With The Crop Production Along With Season		
Crop	State_Name	Season
Brinjal	Andhra Pradesh	Kharif
		Rabi
	Karnataka	Whole Year
	Kerala	Whole Year
	Madhya Pradesh	Whole Year
	Manipur	Whole Year

### 3.4. Creating the Dashboard

Once the charts are created on different tabs in Cognos analytics, they can be pulled them into a dashboard.

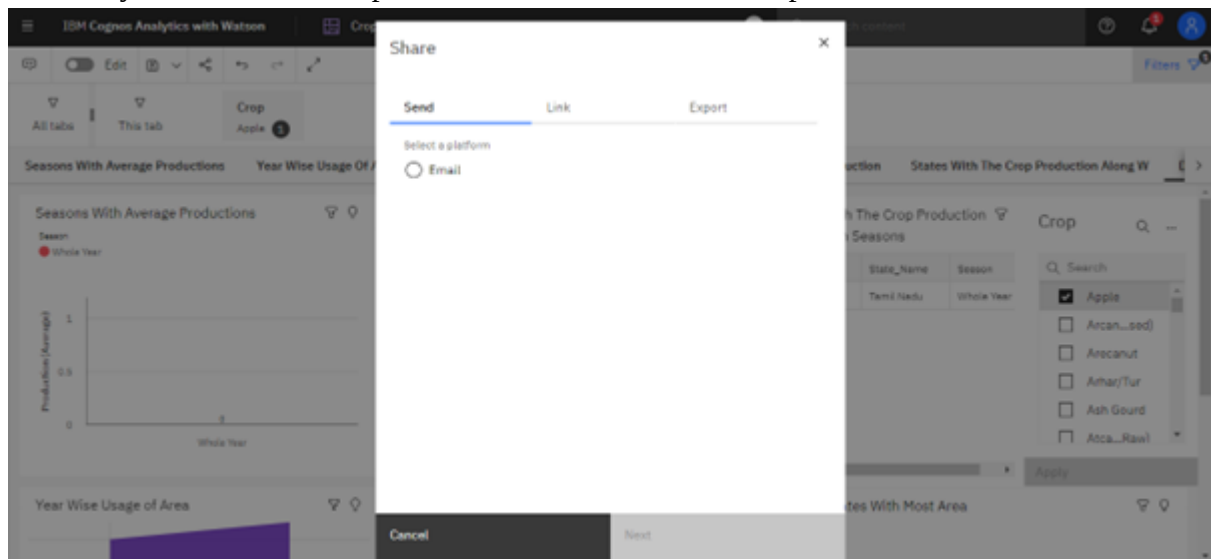


This dashboard can show the data such as seasons with average productions, year wise usage of areas and production, top 10 states with most area, state wise crop production, states with the crop production along with season for different crops when they are selected based on the crop filter on the top right corner of the dashboard. For example, the following figure shows the data for apple.



### 3.5. Exporting the Analytics

The analytics can now be exported to desired format such as pdf or a shareable link or email.



### 4. CONCLUSION

In this project, IBM Cognos Analytics is used to convert raw data into meaningful plots and then the plots are pulled into a dashboard.

The dashboard created is capable of displaying seasons with average productions, year wise usage of areas and production, top 10 states with most area, state wise crop production, states with the crop production along with season for different crops when they are selected based on the crop filer on the top right corner of the dashboard.

By using the dashboard, useful insights can be gained for different crops.

### 5. REFERENCES

- [1] <https://www.kaggle.com/abhinand05/crop-production-in-india>
- [2] <https://youtu.be/2gWcJy7wF9E>
- [3] <https://youtu.be/Mqw9YfLtQIk>
- [4] <https://youtu.be/q36JRI4E6oY>
- [5] <https://youtu.be/3H-3DmRKQP8>
- [6] <https://youtu.be/blgUU-yZo6A>