Project Report

Date	17 November 2022
Team ID	PNT2022TMID14659
Project Name	Estimate the crop yield using Data Analytics

1. **INTRODUCTION**

1.1 Project Overview

India is a predominantly agricultural nation. Agriculture is currently the most significant emerging sector in the actual world and the key industry and economic pillar of our nation. The discipline of agricultural information technology has recently undergone significant changes that have made crop yield prediction an interesting research topic. Crop yield prediction is a technique for estimating crop yield using many characteristics, including temperature, rainfall, fertilisers, insecticides, and other climatic variables and parameters.Using data analytics to analyse those parameters and provide the patterns or trends that has been followed over the past years in estimating the yield can help farmers to make right chioce in the selection of crop varieties,etc.,To make people gasp and use the knowledge represented we finally put the visualisations made in a dashboard and represent it with the most suitable and appropriate charts or graphs or maps.

1.2 Purpose

Analytics is the interpretation of data pattern that assist decision- making and performance improvement. Agriculture Data analytics in crop yield helps in analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

2. **LITERATURE SURVEY**

2.1 Existing problem

[1] Crop Yield Prediction Using Machine Learning : A research group investigated the utilization of various information mining methods which will foresee rice crop yield for the data collected from the state of Maharashtra, India. A total of 27 regions of Maharashtra were selected for the assessment and the data was collected related to the principle rice crop yield influencing parameters such as different atmospheric conditions and various harvest parameters i.e Precipitation rate, minimum, average, maximum and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield for the season between June to November referred as Kharif, for the years 1998 to 2002 from the open source, Indian Administration records. WEKA a Java based dialect programming for less challenging assistance with information data sets, assigning design outcomes tool was applied for dataset processing and the overall methodology of the study includes,

(1) pre-processing of dataset

(2) Building the prediction model utilizing WEKA and

(3) Analyzing the outcomes.

Cross validation study is carried out to scrutinize how a predictable information mining method will execute on an ambiguous dataset. Study applied 10-fold higher cross validation study design to assess the data subsets for screening and testing. Identified and collected information was randomly distributed into 10 sections where in one data section was used for testing while all other data sections were utilized for the preparation information. Study reported that the method applied was supportive in the precise estimation of rice crop yield for the state of P.

Priya et al., (2018) has proposed a random Forest Algorithm for predicting the crop yield of particular area considering various parameters such as rainfall, seasonal crop (Rabi and Kharif) district-wise, temperature (max.), crop production in terms of Kgs/tonnes. Area for doing research was Tamil Nadu. Dataset record were collected from Indian Government over 15years for rice production. They proved in experimental results that prediction analysis done using Random Forest Algorithm – a supervised machine learning algorithm will help farmer to predict the yield of the crop before cultivating onto the agricultural field. This algorithm run efficiently on large databases with high classification accuracy.

[2] Crop Yield Prediction Using Data Mining Techniques : Raor ane A.A. and Kul karn i R.V., discussed few data mining techniques in their paper. They concluded that efficient technique can be developed and analyzed using the appropriate data, to solve complex agricultural problems using data mining techniques. Also recommend some of the algorithms and statistical methods that give [8]good results in agriculture growth.

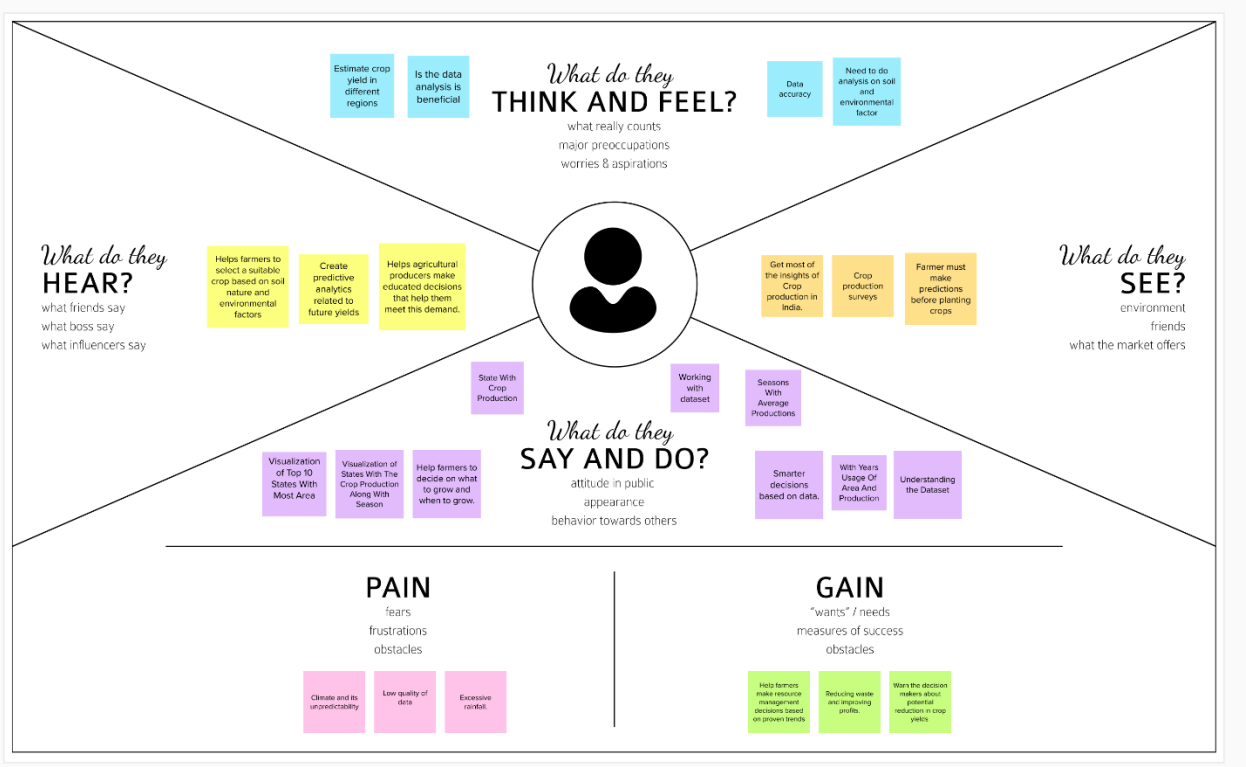
[3] Crop yield prediction using Big Data Analytics: In India crop yield is season dependent and majorly influenced by the biological and economic causes of an individual crop. Reporting of progressive agricultural yield in all the seasons is an ample task and an advantageous task for every nation with respect to assesses the overall crop yield prediction and estimation. At present a common issue worldwide is, farmers are stressed in producing higher crop yield due to the influence of unpredictable climatic changes and significant reduction of water resource worldwide. A study was carried out to collect the data on world climatic changes and the available water resources which can be used to encourage advanced and novel approaches such as big data analytics to retrieve the information of the previous results to the crop yield prediction and estimation. Study imported that the selection and usage of the most desirable crop according to the existing conditions, support to achieve the higher and enhanced crop yield S. Athmaja, M. Hanumanthappa, and V. Kavitha,a survey of machine learning algorithms has presented effective strategies by for big data analytics. All over the world the agricultural peoples gained some advantages through the comparative knowledge from big data analysis, with machine learning algorithm by using huge data the agricultural peoples get some comparative knowledge and changes in regular agriculture

2.2 Problem Statement Definition

In the agriculture sector the farmers are facing difficulties in analysing the demand in market and soil quality analysis to achieve high crop yield through technology. The main objective of this project is to predict crop yield that will be extremely useful to farmers to plan for the harvest and sales of harvested grain

3. **IDEATION & PROPOSED SOLUTION**

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	In the agriculture sector the farmers are facing difficulties in analysing the demand in market and soil quality analysis to achieve high crop yield through technology. The main objective of this project is to predict crop yield that will be extremely useful to farmers to plan for the harvest and sales of harvested grain.
2.	Idea / Solution description	Provide perfect data report after deep analysis of the past data. Helping them out to overcome loss in farming and business.
3.	Novelty / Uniqueness	With this solution we can analyse, visualize data and give the farmers the option to choose which plant/crop to cultivate in which period of time/season to earn more profit from the crop yield.
4.	Social Impact / Customer Satisfaction	Perfect data visuals create a large impact in the crop yield. And hence farmers will be able to gain more profit.
5.	Business Model (Revenue Model)	We can increase/enhance crop production and other raw materials. Also, Increase in productivity will result in increase of Revenue for the farmers.
6.	Scalability of the Solution	With the data visual reports, farmers will be able to cultivate crop according to the area, climate, soil and other features that impact the crop yield and hence enhancing the productivity.

3.4 Problem Solution fit

Project Title: Estimate the Crop Yield Using Data Analytics

Project Design Phase-I - Solution Fit Template
Team ID: PNT2022TMID14659

Define CS, fit into CC

1.CUSTOMER SEGMENT(S) CS

Farmers are the customer who wants to yield a crop in field.

6. CUSTOMER CONSTRAINTS CC

Less knowledge and development towards the current environmental changes and technologies, they follow ancient methods, which is also worthy but, the climatic changes and new kind pesticides

5. AVAILABLE SOLUTION AS

- Traditional ways of prediction.
- Precision farming.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS J&P

- Help them understand the usage of prediction and software application for good results in agriculture.
- Data report should to be created to reduce the loss of the crop and earn more profit in agriculture fields

9. PROBLEM ROOT CAUSE RC

- Various disease on the plants can lead to reducing the quality of the crops productivity.
- The insects on the plants can spread the disease.

7. BEHAVIOUR BE

- Try to get help from agricultural experts.
- Try to take up non-natural means of cultivation for quicker harvest

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS TR

- Seeing their crops are being infected by disease and facing huge loss in quality.

4. EMOTIONS: BEFORE / AFTER EM

Before: Most of the famers in India have Stress, Loosing Self Confidence.

After : Gain of Self Confidence.

10. YOUR SOLUTION SL

- The solution for the problem, creating data report using past datasets.
- Creating IBM Cognos dashboard could make them better understand easily.

8.CHANNELS of BEHAVIOUR CH

- Trying to use pesticides and fertilizers that increase gain but cause harm.
- Irrigation channel changes.

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

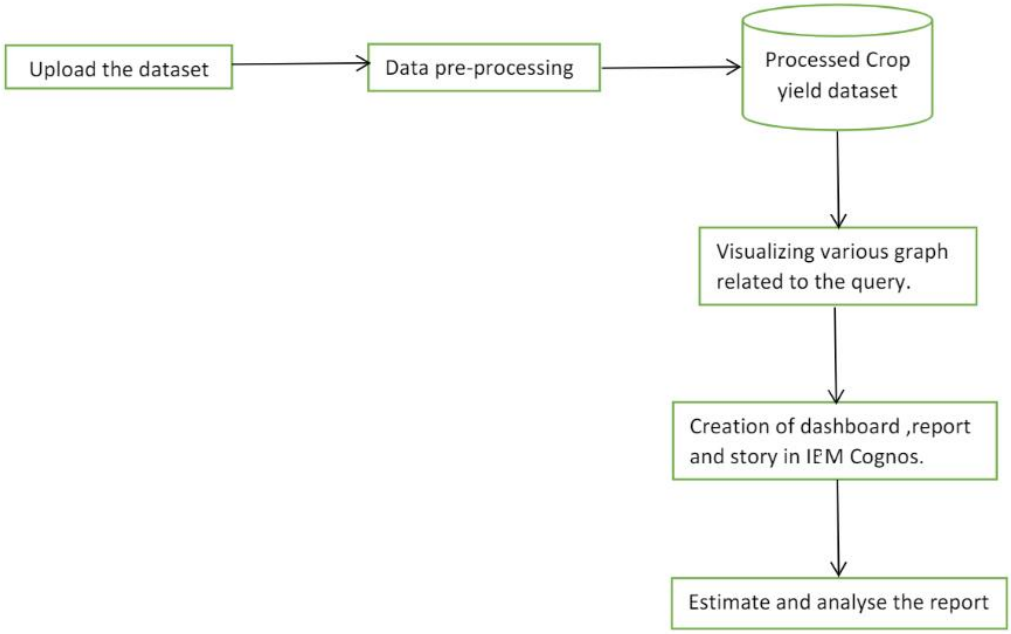
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	User can register through Forms or Gmail account or LinkedIN account.
FR-2	User Confirmation	Confirmation via Email or OTP
FR-3	User Profile	User specific information, Farm details, Yield history.
FR-4	Knowledge about factors that influence the yield	Behaviour of crops and the yield obtained is highly dependent on factors like rainfall, temperature, soil type, etc., Hence it is significant to know the impact of these factors on the yield with its past history.
FR-5	Estimation module	A prediction of crop yield is to be done based on the user's input data (season ,crop ,production ,area).
FR-6	Analysis	An analysis is done on the given data to gain useful insights on the crop yield.

4.2 Non-Functional requirements

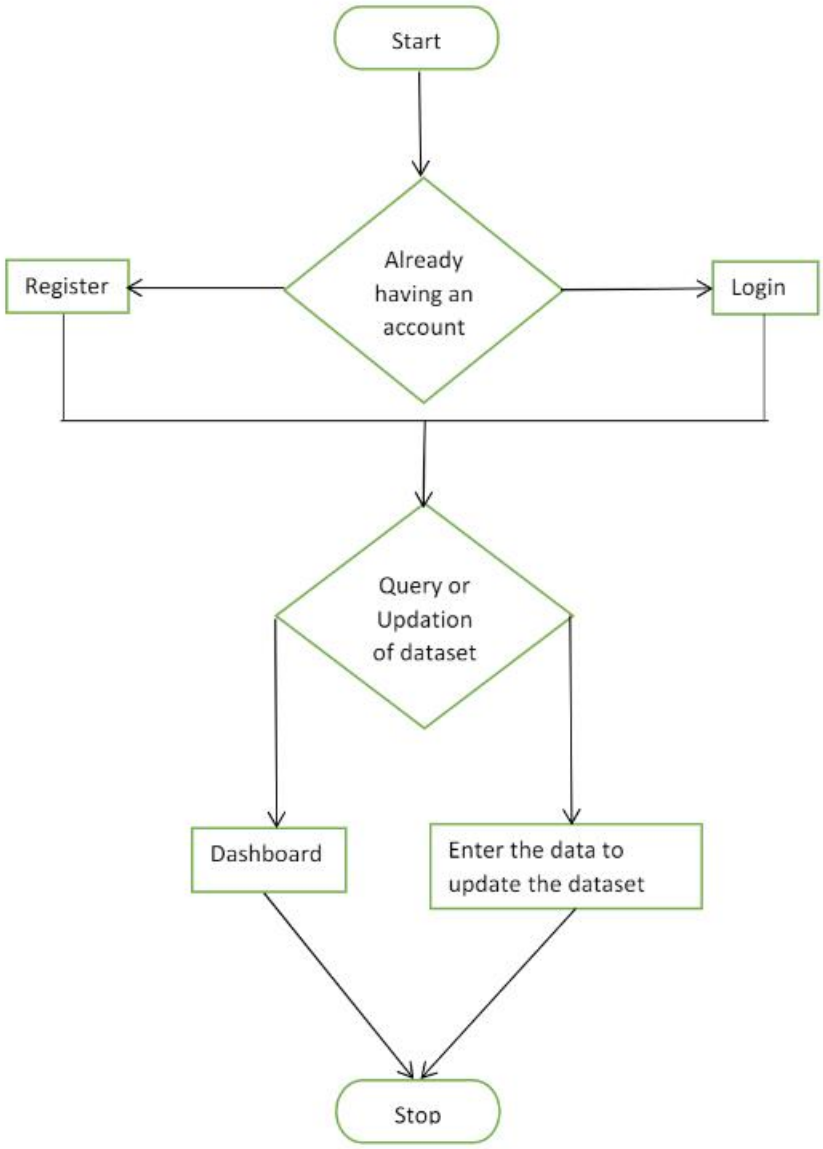
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Provide perfect data report after deep analysis of the past data. Helping farmers to overcome loss in farming and business.
NFR-2	**Security**	The user information is protected by the user login and registration with a secured password.
NFR-3	**Reliability**	Effective tool that all farmers can use, making it reliable by improving the accuracy of the estimation or prediction. This will bridge the gap between farmers and technology.
NFR-4	**Performance**	Multiple technologies and services that will improve the usability in agricultural activities.
NFR-5	**Availability**	Both website and mobile application interface and developed in local language and the content is available in localized language.
NFR-6	**Scalability**	With the data visual reports, farmers will be able to cultivate crop according to the area, climate, soil and other features that impact the crop yield and hence enhancing the productivity.

5. PROJECT DESIGN

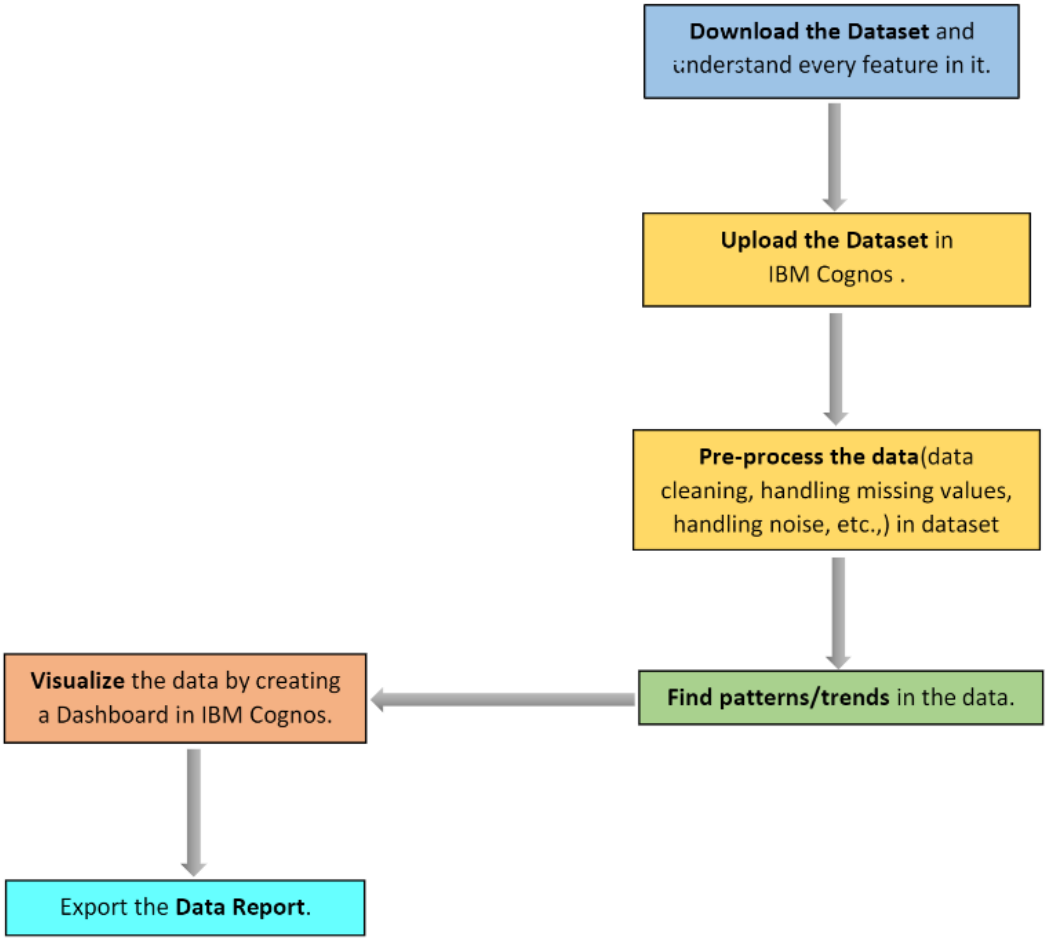
5.1 Data Flow Diagrams



Simplified Flow



5.2 Solution & Technical Architecture



5.3 User Stories

Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Registration	USN-1	As a user, I can register for by entering my Agri - id card and request..	2	High	Praveena R
	USN-2	As a user, I can register for the application through Gmail	2	Medium	Vadhani RK
Login	USN-3	As a user, I can Call and request or Approach for dataset	2	High	Sneka V Ravina C
Working with the Dataset	USN-4	To work on the given dataset, Understand the Dataset.	4	High	Praveena R Vadhani RK
	USN-5	Load the dataset to Cloud platform then Build the required Visualizations	10	High	Sneka V Ravina C

Data Visualization Chart	USN-6	Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations. Build a Visualization to showcase Average Crop Production by Seasons.	4	Medium	Praveena R Vadhani RK
		Showcase the Yearly usage of Area in Crop Production.	4	Medium	Sneka V Ravina C
		Build a visualization to show case top 10 States in Crop Yield Production by Area.	4	Medium	Praveena R Vadhani RK
		Build the required Visualization to showcase the Crop Production by State.	4	Medium	Sneka V Ravina C
		Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation.	4	Medium	Praveena R Vadhani RK
Creating The dashboard	USN-8	Create the Dashboard by using the created visualizations.	20	High	Ravina C Sneka V Vadhani RK Praveena R
Export The Analytics	USN-9	Export the created Dashboard	20	High	Ravina C Sneka V Vadhani RK Praveena R

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for by entering my Agri - id card and request..	2	High	Praveena R
Sprint-1		USN-2	As a user, I can register for the application through Gmail	2	Medium	Vadhani RK
Sprint-1	Login	USN-3	As a user, I can Call and request or Approach for dataset	2	High	Sneka V Ravina C
Sprint-1	Working with the Dataset	USN-4	To work on the given dataset, Understand the Dataset.	4	High	Praveena R Vadhani RK
		USN-5	Load the dataset to Cloud platform then Build the required Visualizations	10	High	Sneka V Ravina C
Sprint-2	Data Visualization Chart	USN-6	Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations. Build a Visualization to showcase Average Crop Production by Seasons.	4	Medium	Praveena R Vadhani RK
			Showcase the Yearly usage of Area in Crop Production.	4	Medium	Sneka V Ravina C
			Build a visualization to show case top 10 States in Crop Yield Production by Area.	4	Medium	Praveena R Vadhani RK

			Build the required Visualization to showcase the Crop Production by State.	4	Medium	Sneka V Ravina C
			Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation.	4	Medium	Praveena R Vadhani RK
Sprint-3	Creating The dashboard	USN-8	Create the Dashboard by using the created visualizations.	20	High	Ravina C Sneka V Vadhani RK Praveena R
Sprint-4	Export The Analytics	USN-9	Export the created Dashboard	20	High	Ravina C Sneka V Vadhani RK Praveena R

6.2 Sprint Delivery Schedule

Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date (Planned)**	**Story Points Completed (as on Planned End Date)**	**Sprint Release Date (Actual)**
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

7. WORKING WITH THE DATASET & DATA VISUALISATION

7.1 Understanding the dataset

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

Dataset Link: Dataset

Let's understand the data we're working with and give a brief overview of what each feature represents or should represent 1. State Name - All the Indian State names.

2. District Name -Different District names.

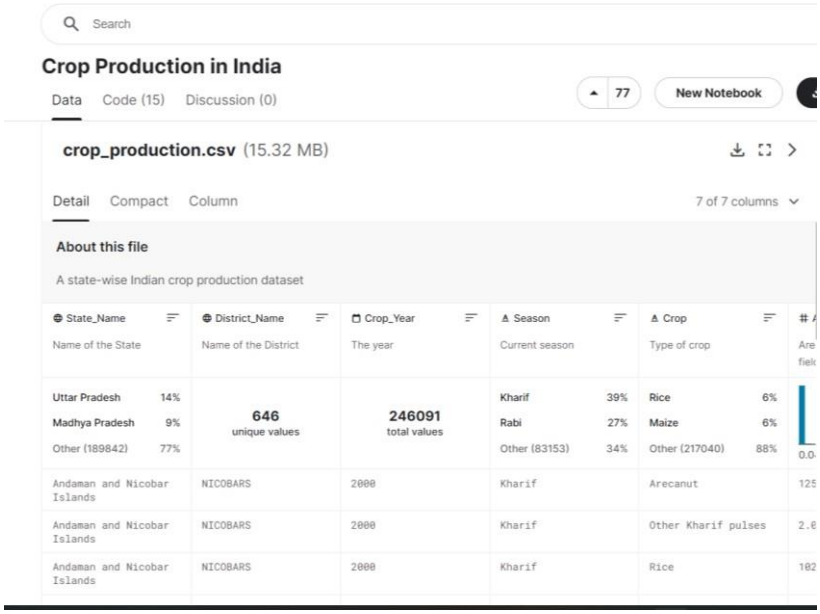
3. Crop Year- contains the crop years.

4. Season – Different seasons for crop production.

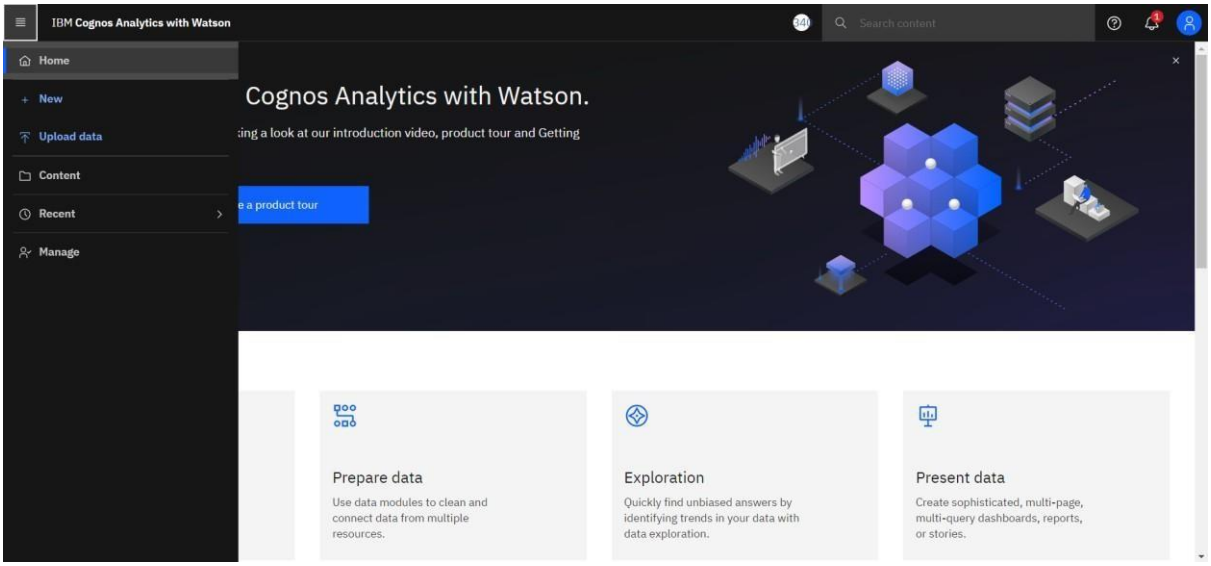
5. Area- Total number of areas covered.

6. Production- production of crops.

The data format is as shown in the below image:

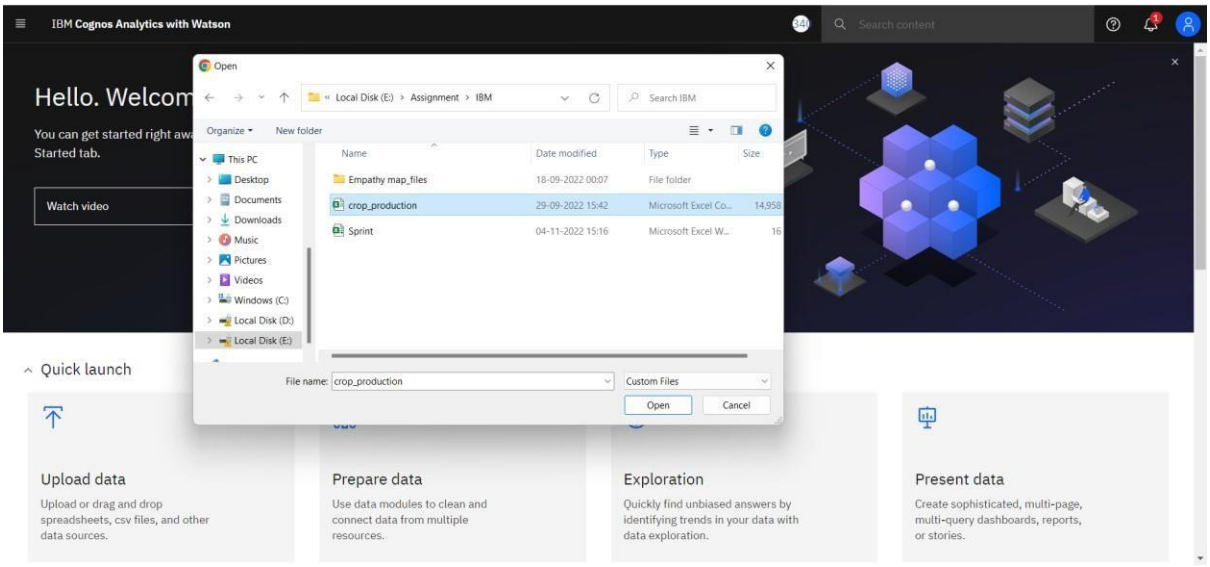


7.2 Loading the dataset

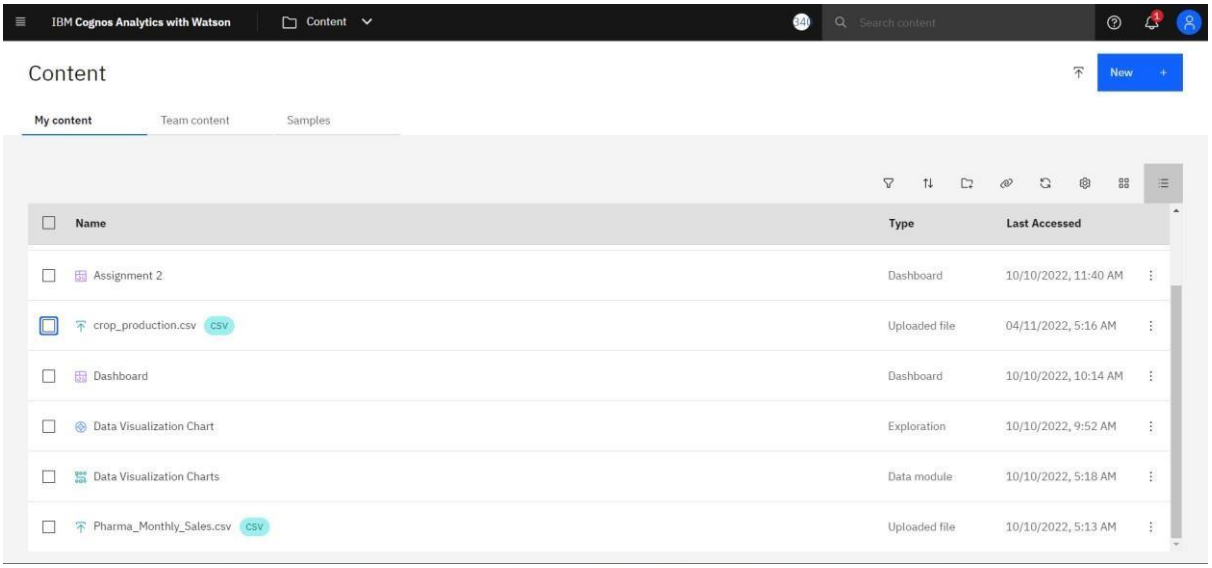


- Click the open menu in the top left corner.

- Select the **Upload Data** in the menu and select the Dataset that you want upload.

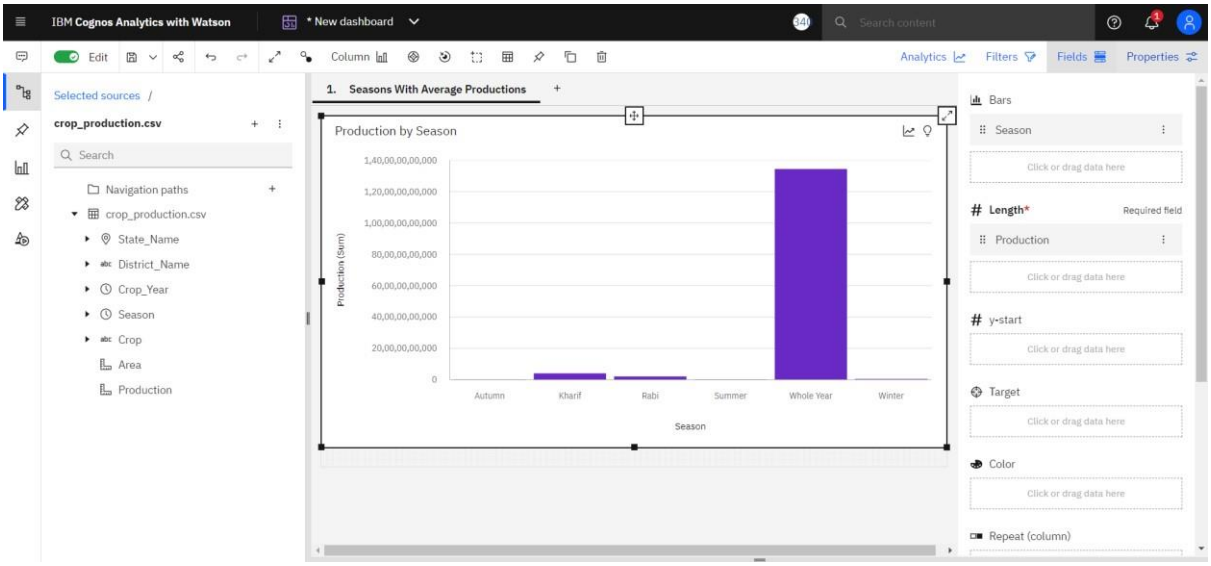


- Once the Dataset is Uploaded it will be displayed in content.

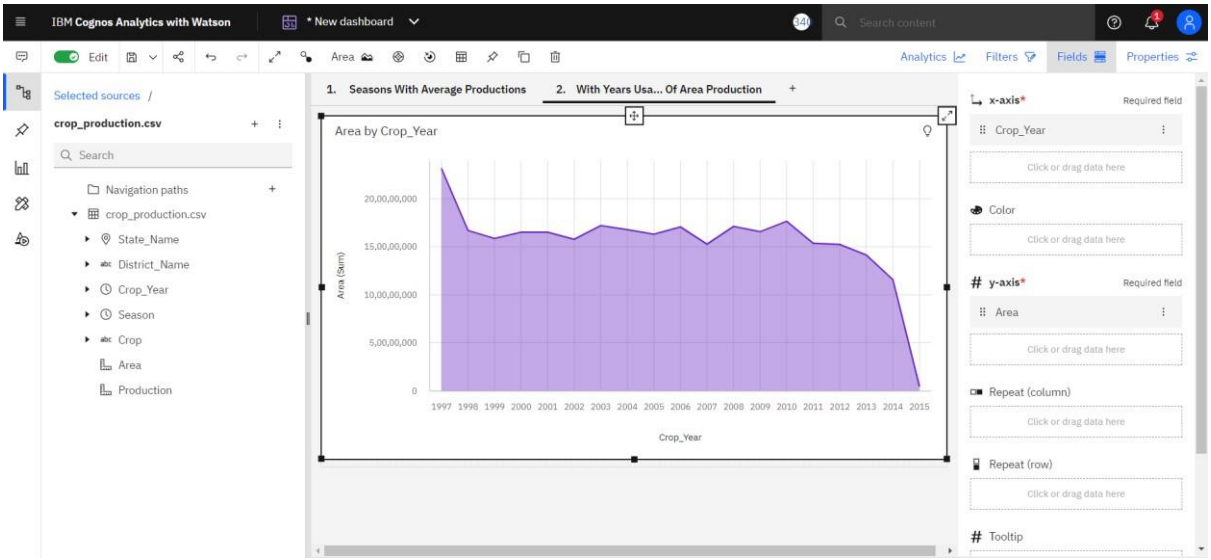


7.3 Visualization charts

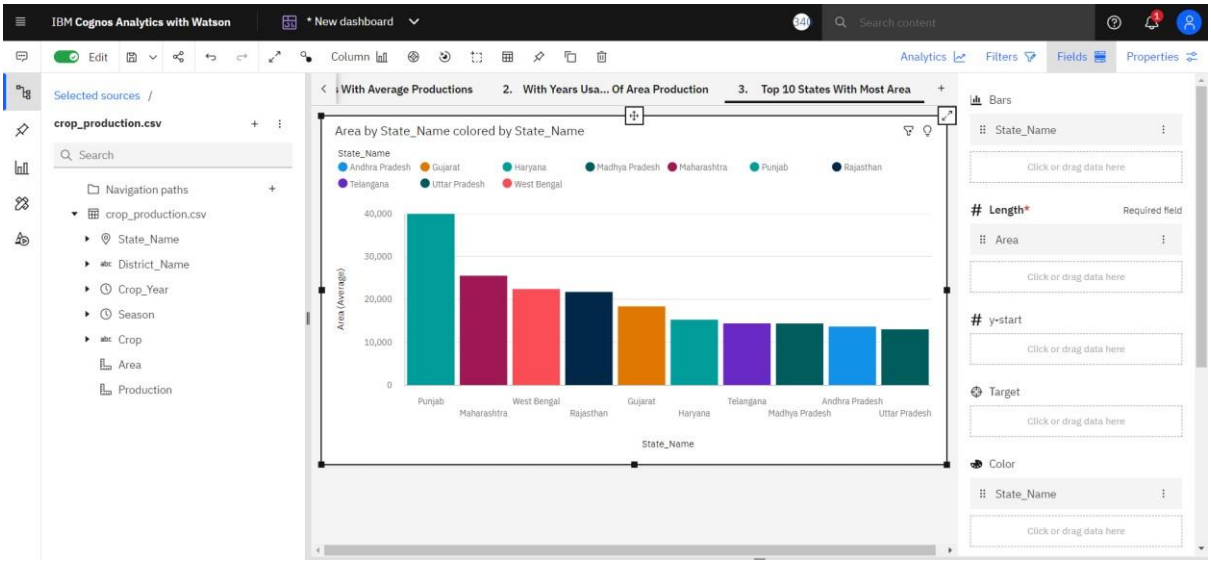
- Seasons with average productions



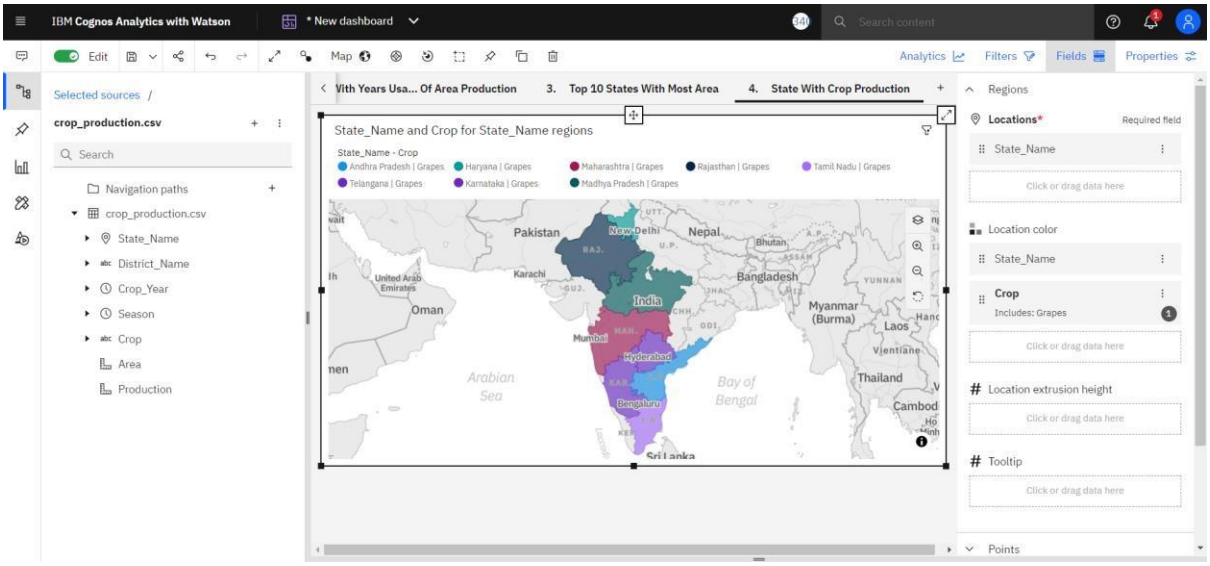
- With years usage of Area and Production



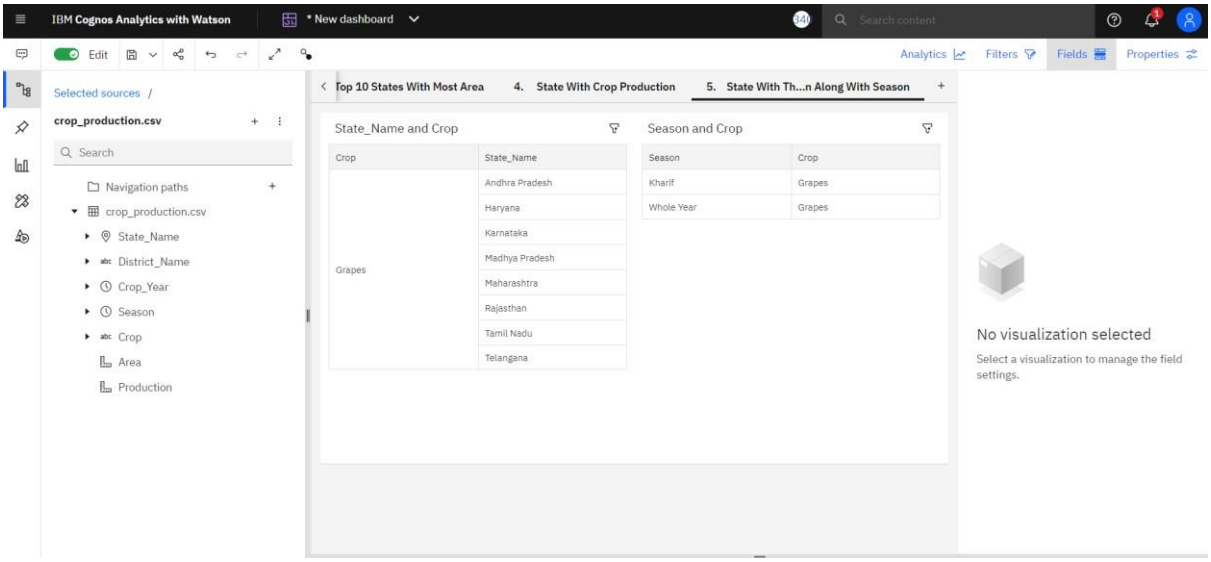
- Top 10 States with most area



- State with crop production

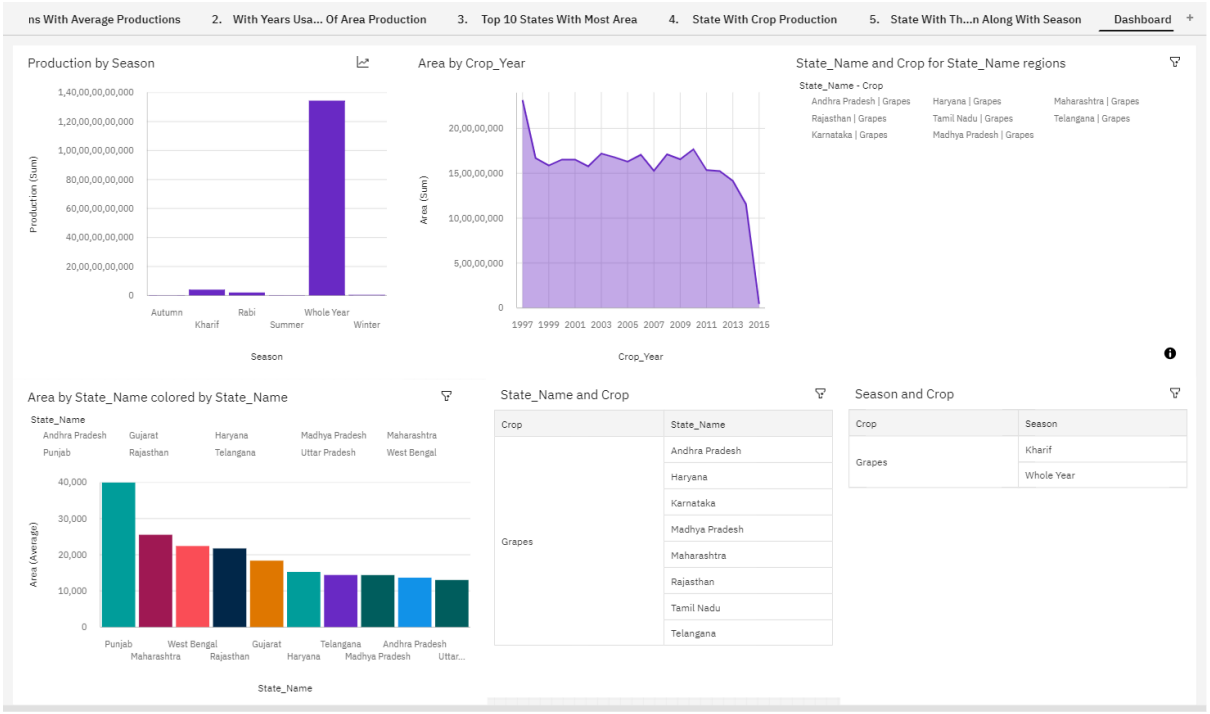


- States with the crop production along with season (Text Table)



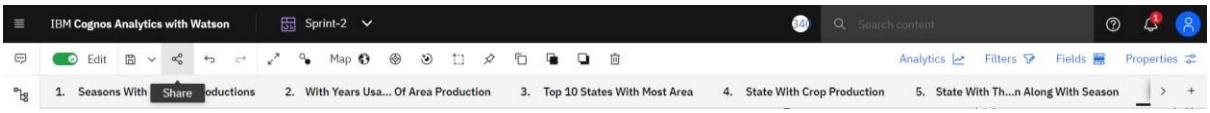
8. CREATING THE DASHBOARD AND EXPORT THE ANALYTICS

8.1 Creating the Dashboard

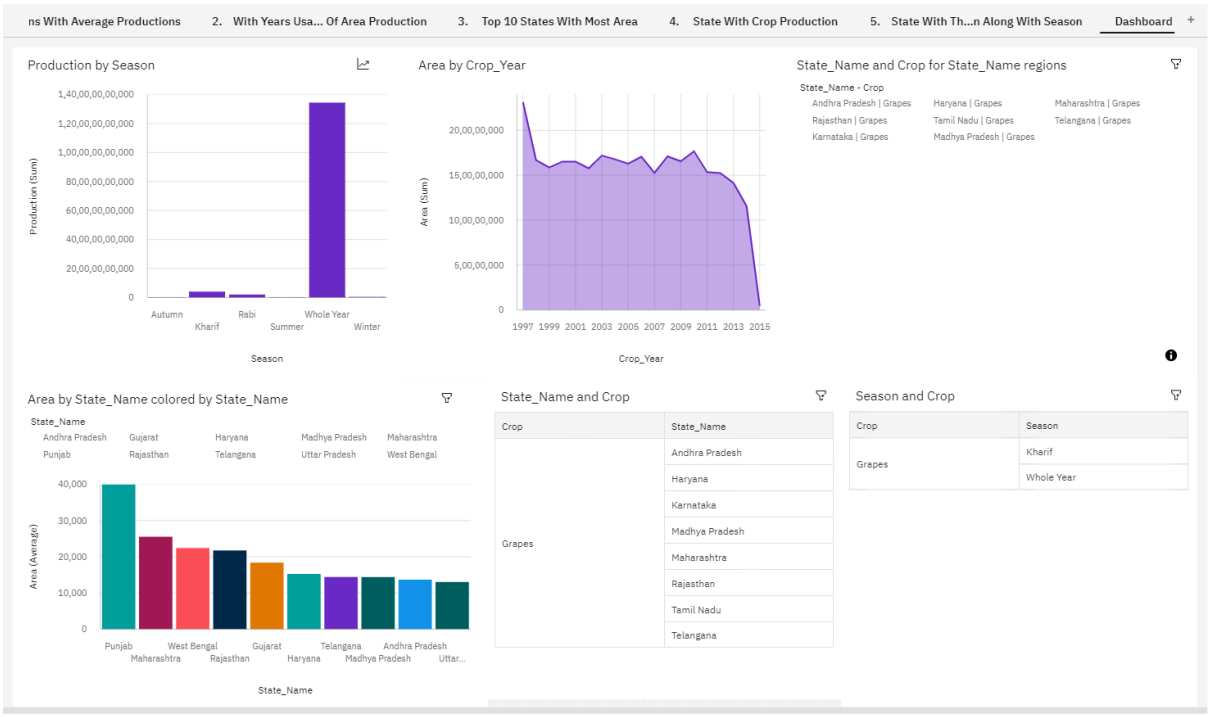


8.2 Export the Analytics

- Click on the **share icon**



- You can share using Email or Link or Export as pdf.
- Click the Export tab in the Share dialog box.
- You can change the page size and Orientation setting then click **Export**.



9. ADVANTAGES & DISADVANTAGES

Advantages:

One can easily analyse and understand trends in croppping pattern, seasonal behaviour of land in various areas with the created dashboard.

With no prior skills and knowledge about the tools that we use for analysis,anyone (literate or illiterate)can easily infer the knowlegde that we represent in various charts or graphs or maps. So that it would be helpful to farmers to make appropriate decisions in the future.

Disadvantages:

Not all factors influencing the crop yield are being considered for the analyis as we have only taken visible factors into account for the analysis.

10. CONCLUSION

The productivity of agriculture has slightly increased as a result of technology's introduction.
New ideas like digital agriculture, smart farming, precision agriculture, etc.
have been made possible by the innovations. From the analysis dashboard, it has been noted that analyses of agricultural productivity and the detection of hidden patterns utilising data sets related to
seasons and crop yields have been conducted.Using IBM Cognos, we have observed and conducted analysis
on various crops grown, area, and productions in various states and districts, including
1) Seasons with average productions. We learn from these analytics which seasons have higher average production and which have lower production.
2)Production split up per crop year. We learn from this study which years have high and low production.
3) District-based production. With the help of these analytics, we may identify the states and districts that farm the chosen crops.
4) Production by area.This will allow us to estimate the yield and determine how much land needs to be planted. After creating the dashboard,
study was done to determine which state, which year, and how much crop area will be produced.

11. FUTURE SCOPE

Farming is the means of survival as humans require food that is obtained only through farming directly or indirectly.With the growing human population, it is critcal to analyse the production in farming every year. So, that we can know the right time,right place and right crop to be cultivated considering all the factors that influence the crop production.

It can be concluded that the research in the field of agriculture with reference to using IT trends like data analytics is in its infancy. As the food is the basic need of humans, the requirement of getting the
maximum yields using optimal resource will become the necessity in near future as a result of growing population. The survey outcomes indicate the need for improved techniques in crop yield analytics.
There exists a lot of research scope in this research area.

12. APPENDIX

Source Code:
https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FSprint-2&action=view&mode=dashboard&subView=model00000184429ff804_00000000

GitHub Link:
https://github.com/IBM-EPBL/IBM-Project-10421-1659179435

Project Demo Link:
https://github.com/IBM-EPBL/IBM-Project-10421-1659179435/blob/main/Demo%20video.mp4