R.M.D. ENGINEERING COLLEGE

NALIYATHIRAN PROJECT BASED LEARNING

On

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTERPRENUERSHIP

A PROJECT REPORT

Submitted By

Ponarasu A (111519205032)
Tejesh S (111519205051)
Praveen kumar joshi (111519205034)
Sasi Kumar L (111519205040)

INTERNAL MENTOR
Dr. KALPANA
ASSISTANT PROFESSOR,
Department of Information Technology,
R.M.D. ENGINEERING COLLEGE.

INDUSTRY MENTORS
Srikanth
Mahidhar
Mohammed Azhar Uddin

ABSTRACT

In order to address issues with food security and lessen the effects of climate change, it is essential to comprehend global agricultural productivity. One of the top UN Sustainable Development Goals for 2030 is to end hunger, and this goal can assist. In the project, we present a scalable, precise, and lowcost approach to agricultural production prediction utilising openly accessible remote sensing data and machine learning. Several months prior to harvest, our deep learning system can estimate crop yield using just globally accessible covariates with high spatial resolution (county-level). We think our technique may be useful for setting suitable food reserve levels, identifying low-yield locations, and enhancing risk management of derivatives related to crops. In India crop yield is season dependent and majorly influenced by the biological and economic cause of an individual crop. Reporting of progressive agriculture yield in all the seasons is an ample task and an advantageous task for every nation with the respect to assesses the overall crop yield and prediction and estimation at present a common issue world wide is farmers are stressed in producing higher crop yield due to influence of unpredictable climatic changes and significant reduction of water resources world wide. 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, supports to achieve the higher and enhanced crop yield. An explicit rationale model which can effectively applied at various levels of the availability of quality information for identifying data sources to analyze crop yield and measuring yield gaps at definite geographical locations and works based on the rise in titer approach. The model is highly helpful in retrieving the useful data from the available, poor quality, less rigorous data sources or if the data is not available. A case study was discussed on the application of selected model design to quantify the yield gaps of maize crop in the state of Nebraska (USA), and also at the different geographical locations representing the nations Argentina and Kenya at national scale level. Different geographical locations such as Nebraska (USA), Argentina and Kenya were identified to symbolize the distinct scenarios of Agri based data availability and the quality for the selected variables assessed to predict and estimate the crop yield gaps. The definitive aspiration of the planned method is to afford transparent, easily accessible, reproducible and technically sound and strong guidelines for predicting the yield gaps. The proposed guidelines were also relevant for understanding and to simulate the influence of change in climate conditions and usage of cultivable land changes from national to global scales. As indicated, the better understanding of data.

TABLE OF CONTENTS

INTRODUCTION

1.1 Project Overview

India is a predominantly agricultural nation. Agriculture is currently the mostsignificant emerging sector in the actual world and the key industryand economic pillar of our nation. The discipline of agricultural information technology has recently undergone significant changes that have made crop yield prediction an interesting researchtopic. Crop yield prediction is a techniquefor 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 farmersto make right chioce in the selectionof 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 graphsor maps.

1.2 Purpose

Analytics is the interpretation of data patternthat 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.

LITERATURE SURVEY

2.1 Existing problem

- **Crop Yield Prediction Using Machine Learning**: A research group investigated the utilization of various information mining methods which will foreseerice 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 collectedrelated to the principle rice crop yield influencing parameters such as differentatmospheric conditions and various harvestparameters i.e Precipitation rate, minimum, average, maximum and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield for the seasonbetween June to November referredas 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 andthe 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 randomlydistributed into 10 sections where in one data sectionwas used for testing while all

other data sections were utilized for the preparation information. Study reportedthat themethod appliedwas supportive in the preciseestimation 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.

- **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 developedand 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.

- **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 nationwith 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 dueto the influence of unpredictable climatic changes and significant reductionof water resourceworldwide. A studywas carried out to collectthe data on world climaticchanges and the available waterresources which can be used to encourageadvanced and novelapproaches suchas 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 existingconditions, support to achieve the higher and enhanced crop yield.

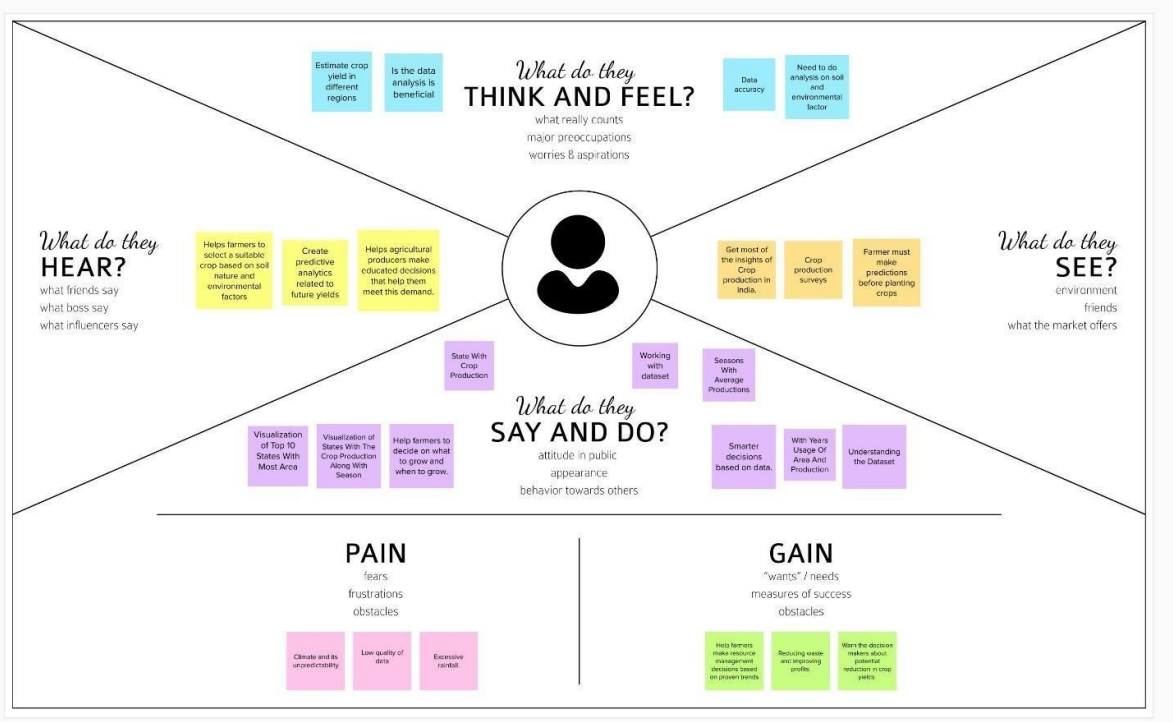
  S. Athmaja, M. Hanumanthappa, and V. Kavitha,a survey of machine learning algorithms has presented effective strategies by for big data analytics. All over the world the agriculturalpeoples gained some advantages throughthe comparative knowledgefrom big data analysis, with machine learning algorithm 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 qualityanalysis to achievehigh crop yield through technology. The main objective of this projectis to predict crop yield that will be extremely useful to farmersto plan for the harvestand sales of harvested grain.

IDEATION & PROPOSEDSOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming

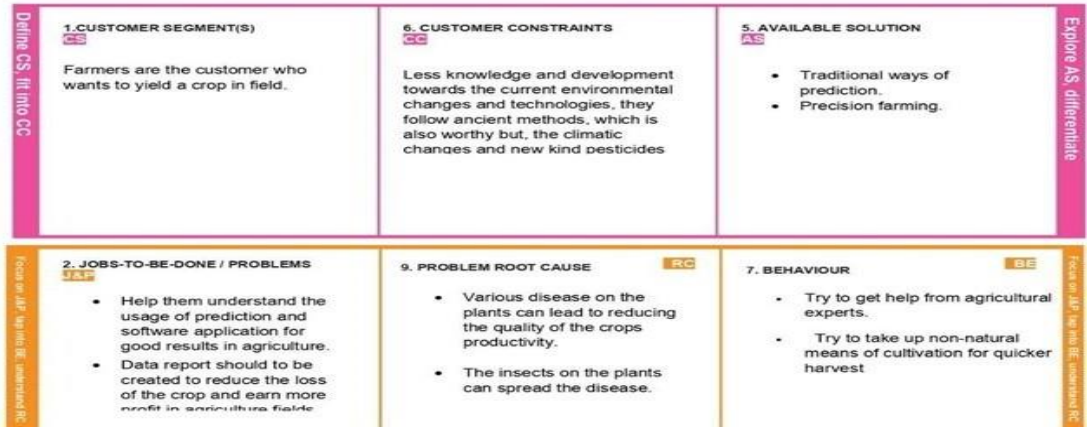
3.3 Proposed Solution

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 cropyield through technology. The main objective of this project is to predictcrop yield that will be extremely usefulto farmers to plan for the harvest and sales of harvested grain.
2.	Idea / Solution description	Provide perfect datareport after deepanalysis of the past data. Helpingthem out to overcome lossin farming and business.
3.	Novelty / Uniqueness	With this solution we can analyse, visualize data and givethe farmers the option to choose whichplant/crop to cultivate in which periodof time/season to earn more profit from the crop yield.
4.	Social Impact / Customer Satisfaction	Perfect data visuals create a large impact in the cropyield. And hence farmers will be able to gainmoreprofit.
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 datavisual reports, farmers will be ableto cultivate crop according to the area, climate, soiland other features that impact the crop yieldand hence enhancing the productivity.

3.3 Problem Solution fit

Project Title: Estimate the Crop Yield Using Data Analytics

Project Design Phase-I - Solution Fit Template



REQUIREMENT ANALYSIS

4.1 Functional requirement

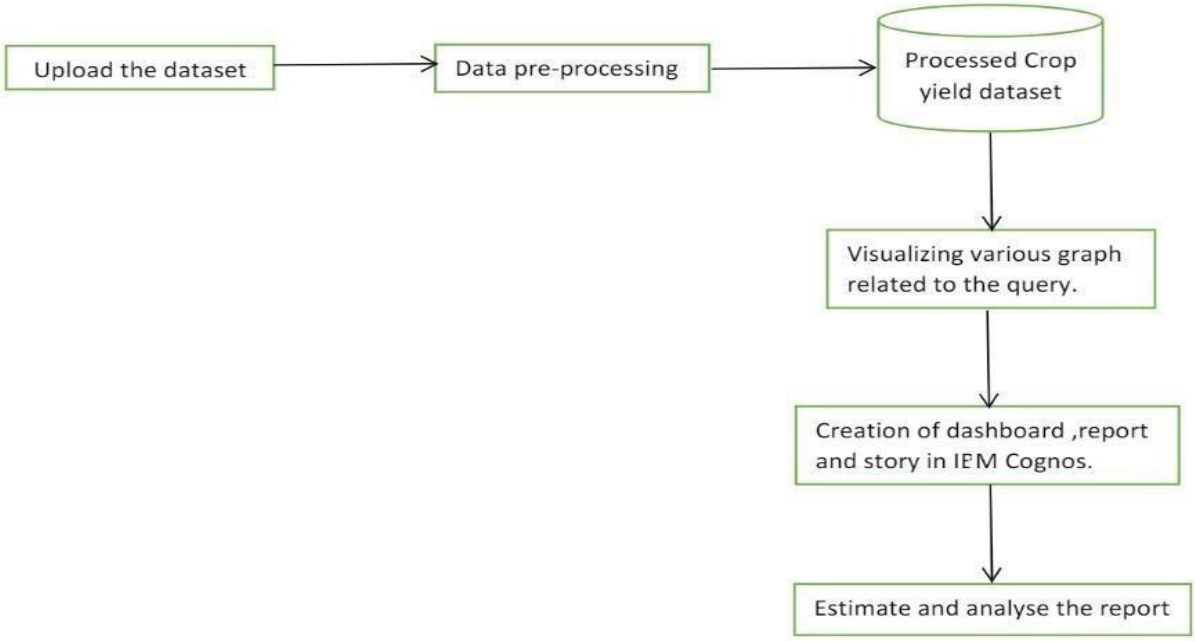
FR No.	Functional Requirement (Epic)	Sub Requirement (Story/ Sub-Task)
FR-1	User Registration	User can register through Forms or Gmailaccount or LinkedIN account.
FR-2	User Confirmation	Confirmation viaEmail or OTP
FR-3	User Profile	User specific information, Farm details, Yield history.
FR-4	Knowledge aboutfactors that influence the yield	Behaviour of cropsand the yieldobtained is highlydependent on factorslike rainfall, temperature, soil type, etc., Hence it is significant to know the impact of these factors on the yieldwith its past history.
FR-5	Estimation module	A prediction of crop yieldis to be done basedon the user's input data (season ,crop ,production ,area).
FR-6	Analysis	An analysis is done on the givendata to gainuseful insights on the crop yield.

4.2 Non-Functional requirements

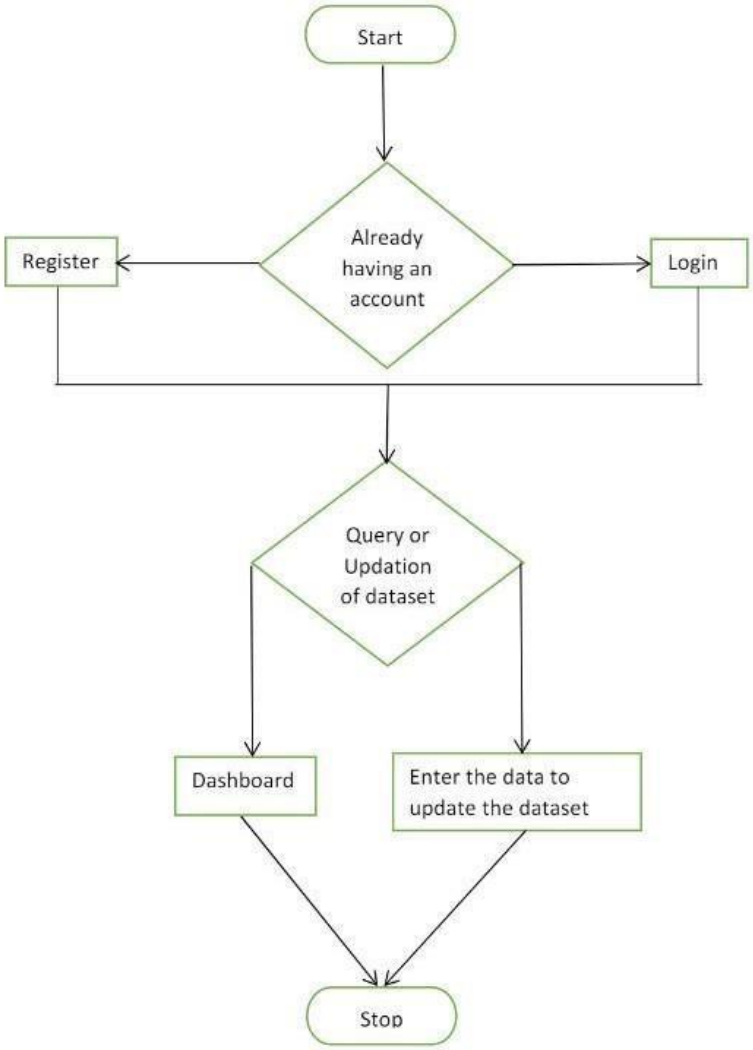
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Provide perfectdata report afterdeep analysis ofthe past data.Helping farmers to overcome loss in farming andbusiness.
NFR-2	**Security**	The user information is protected by the user loginandregistration with a secured password.
NFR-3	**Reliability**	Effectivetool that all farmers can use, makingit reliable by improving the accuracy of the estimation or prediction. This will bridgethe gap between farmers and technology.
NFR-4	**Performance**	Multiple technologies and services that will improvethe usability in agricultural activities.
NFR-5	**Availability**	Both websiteand 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, soiland other features that impact thecrop yield and hence enhancing the productivity.

PROJECT DESIGN

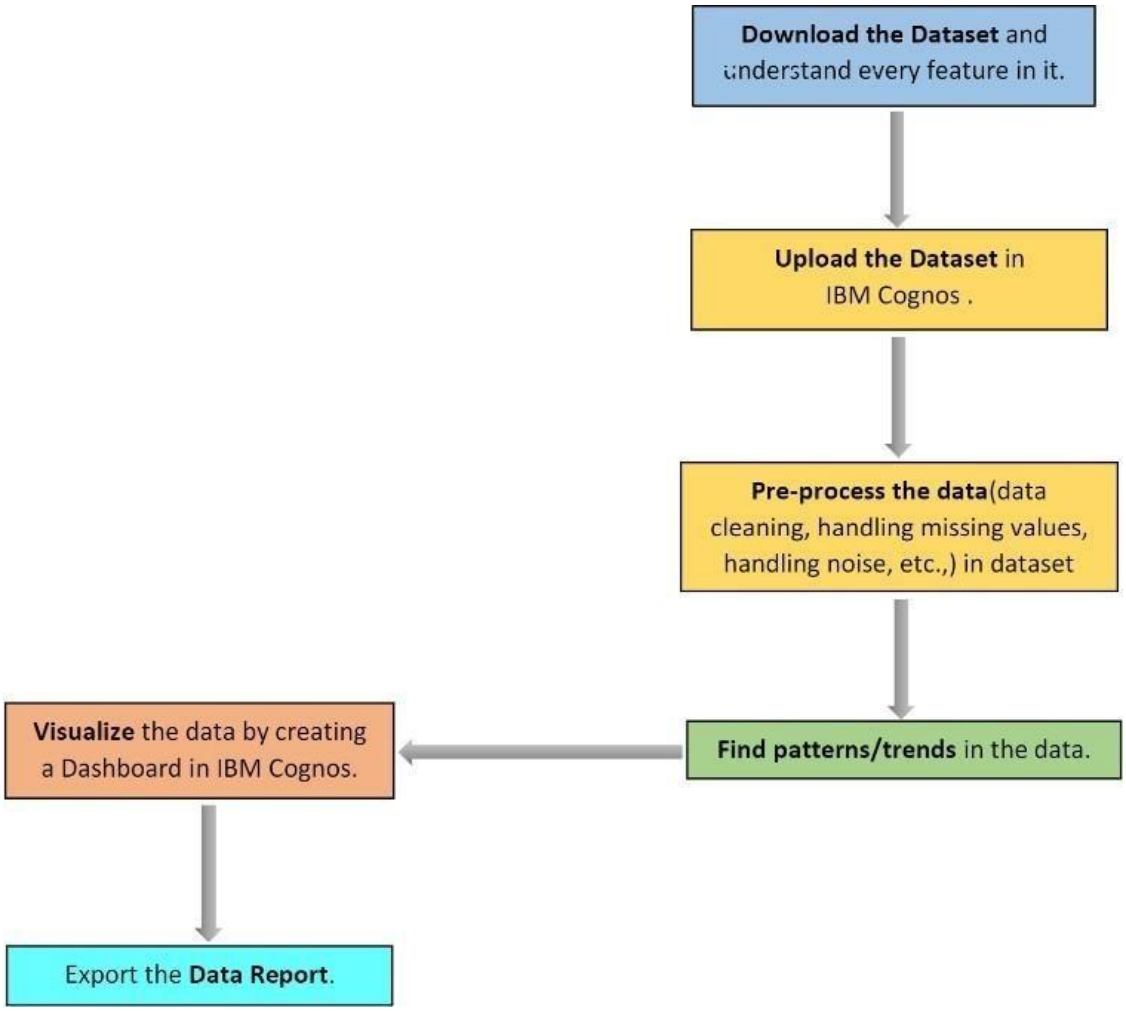
5.1 Data Flow Diagrams



Simplified Flow



5.2 Solution & TechnicalArchitecture



5.3 User Stories

Functional Requirement (Epic)	User StoryNumber	User Story/ Task	Story Points	Priority	Team Members
Registration	USN-1	As a user, I can register for byentering my Agri- id card and request..	2	High	Sasikumar,ponarasu
	USN-2	As a user, I can register for the application through Gmail	2	Medium	Tejesh,Sasikumar
Login	USN-3	As a user, I can Call and requestor Approach fordataset	2	High	Joshi,Sasikumar
Working with the Dataset	USN-4	To work on the given dataset, Understand the Dataset.	4	High	Tejesh,ponarasu
	USN-5	Load the dataset to Cloud platform then Build the required Visualizations	10	High	Sasikumar ,ponarasu

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	Tejesh,Sasikumar
		Showcase theYearly usage of Area in Crop Production.	4	Medium	Tejesh,Sasikumar
		Build a visualization to show case top 10 States in Crop YieldProduction by Area.	4	Medium	Sasikumar,ponarasu
		Build the required Visualization to showcase theCrop Production by State.	4	Medium	Tejesh,ponarasu
		Build Visual analytics to represent the Sates with Seasonal Crop Production using a Textrepresentation.	4	Medium	Tejesh,Sasikumar
Creating Thedashboard	USN-8	Create the Dashboard by using the created visualizations.	20	High	Tejesh,Sasikumar, joshi,ponarasu
Export TheAnalytics	USN-9	Export thecreated Dashboard	20	High	Tejesh,ponarasu

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 cardand request..	2	High	Joshi
Sprint-1		USN-2	As a user, I can register for the application through Gmail	2	Medium	Joshi

Sprint-1	Login	USN-3	As a user, I can Call and requestor Approach fordataset	2	High	Joshi,ponarasu
Sprint-1	Working with the Dataset	USN-4	To work on the given dataset, Understand the Dataset.	4	High	Sasikumar,joshi
		USN-5	Load the dataset to Cloud platform then Build the required Visualizations	10	High	Tejesh,ponarasu
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	Joshi,ponarasu
			Showcase theYearly usage of Area in Crop Production.	4	Medium	Tejesh,Joshi
			Build a visualization to show case top 10 States in Crop YieldProduction by Area.	4	Medium	Joshi,Sasikumar
			Build the required Visualization to showcase theCrop Production by State.	4	Medium	Joshi
			Build Visual analytics to represent the Sates with Seasonal Crop Production using a Textrepresentation.	4	Medium	Sasikumar,Joshi
Sprint-3	Creating Thedashboard	USN-8	Create the Dashboard by using the created visualizations.	20	High	Joshi,Tejesh
Sprint-4	Export TheAnalytics	USN-9	Export thecreated Dashboard	20	High	Ponarasu, Tejesh

6.2 Sprint Delivery Schedule

Sprint	Total StoryPoints	Duration	Sprint StartDate	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

WORKING WITH THE DATASET& DATA VISUALISATION

7.1 Understanding the dataset

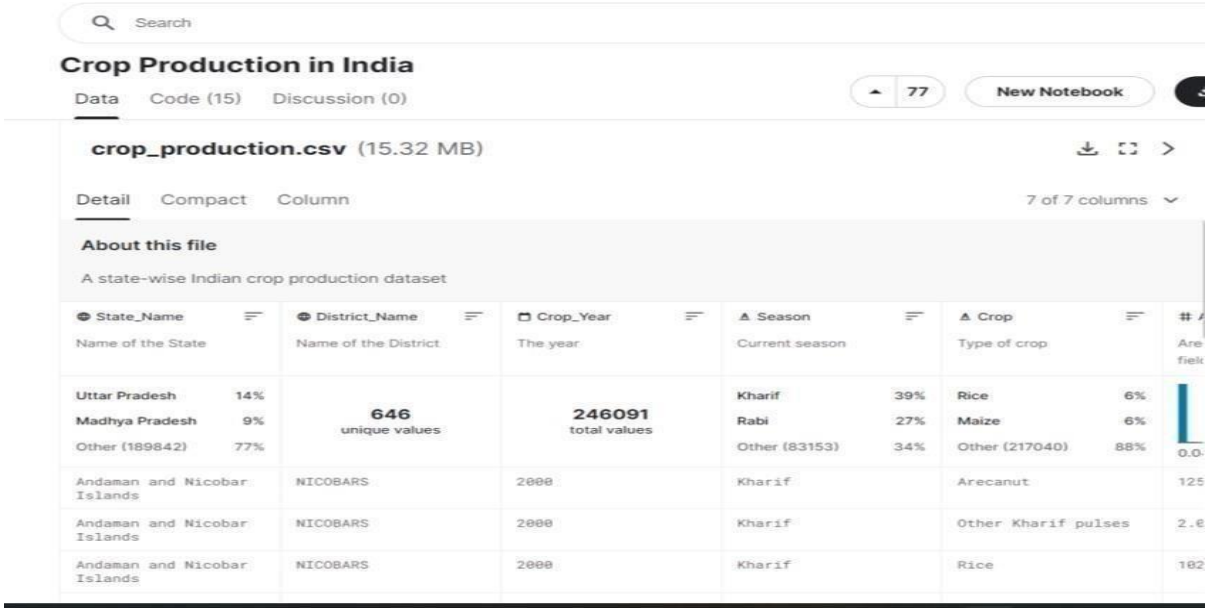
This project is based on a understanding the crop production of India. It has 2,46,092datapoints (rows) and 6 features (columns)describing each crop production relateddetails.

Dataset Link: Dataset

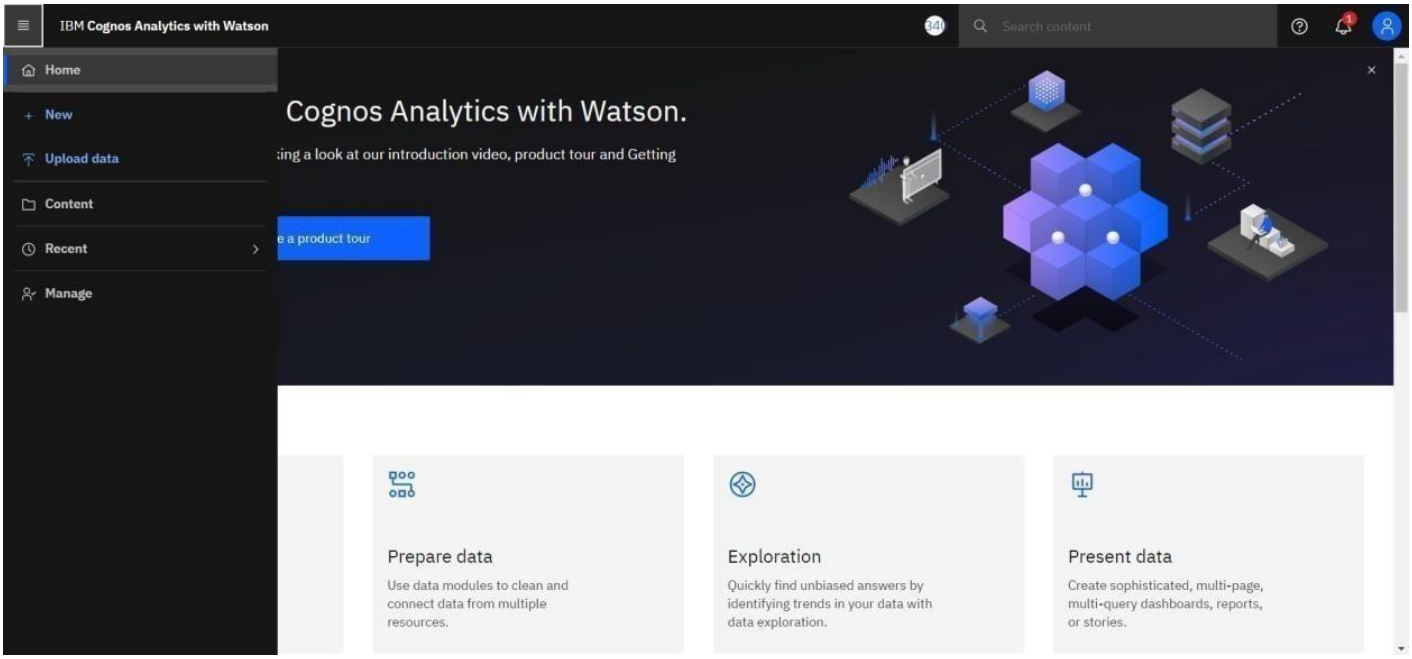
Let's understand the data we're working with and give a brief overview ofwhat each featurerepresents or shouldrepresent 1. StateName - All the IndianState names.

1. District Name -Different District names.
2. Crop Year- containsthe crop years.
3. Season – Different seasons for crop production.
4. Area- Total numberof areas covered.
5. Production- production of crops.

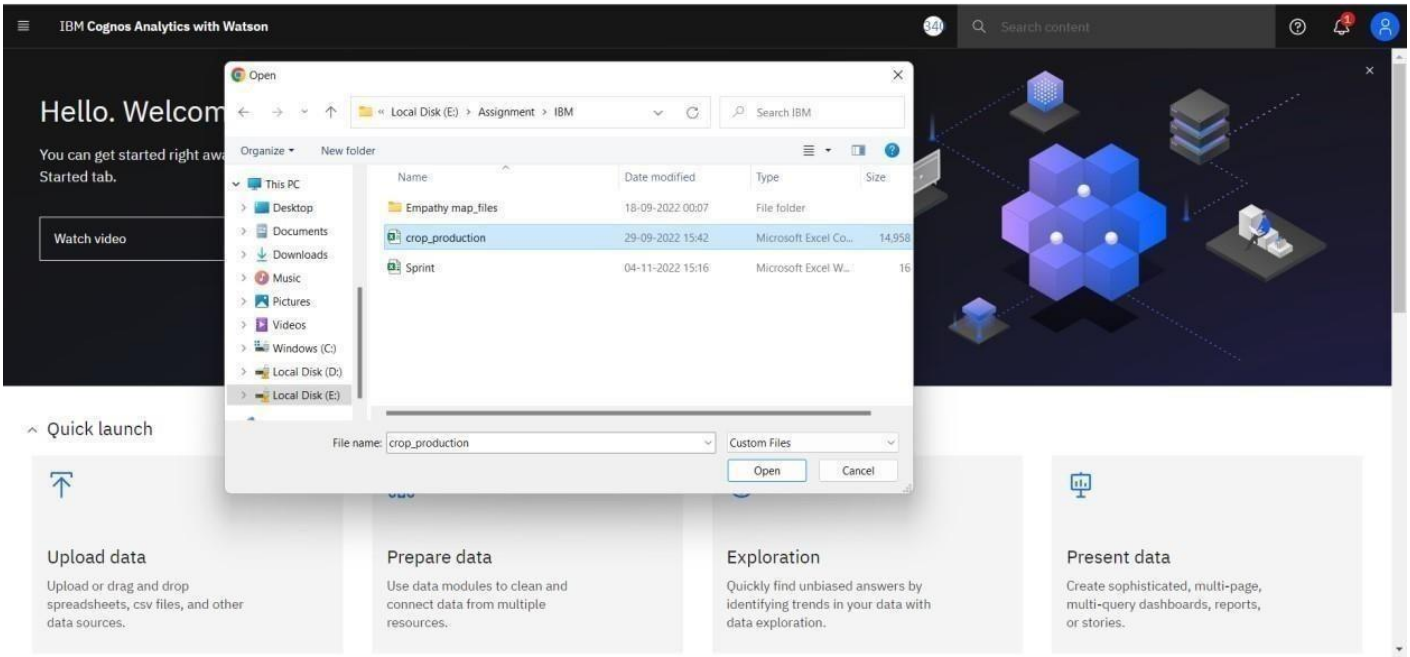
The data format is as shownin the below image:



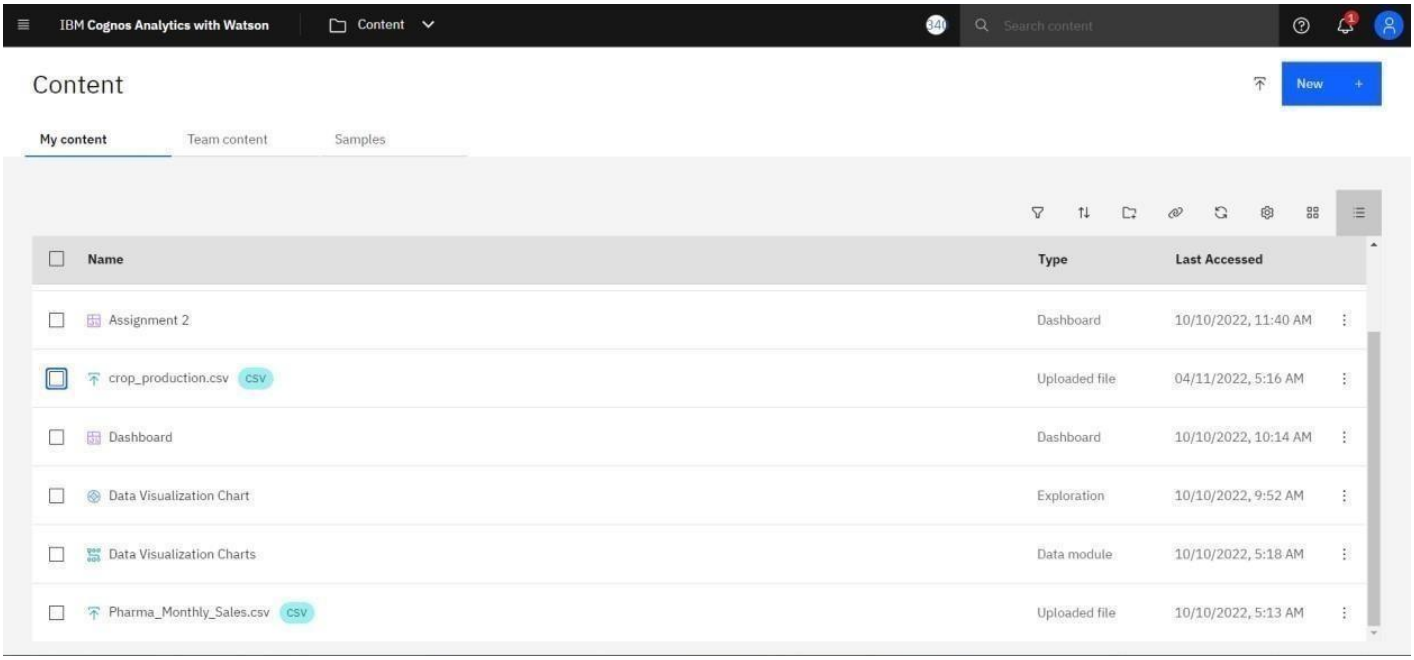
7.2 Loading the dataset



- Click the open menu in the top left corner.
- Select the **Upload Data** inthe menu and select the Datasetthat you want upload.

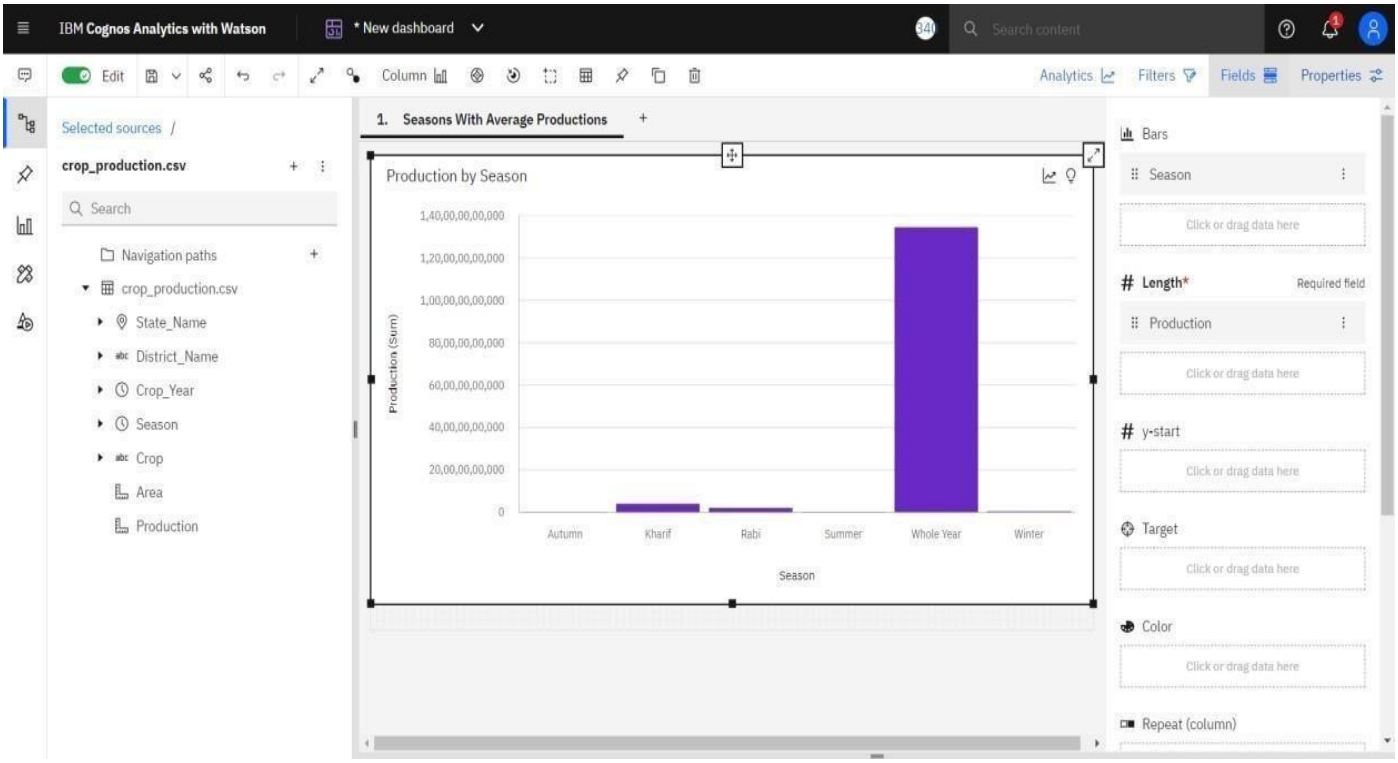


- Once the Dataset is Uploaded it will be displayed in content.

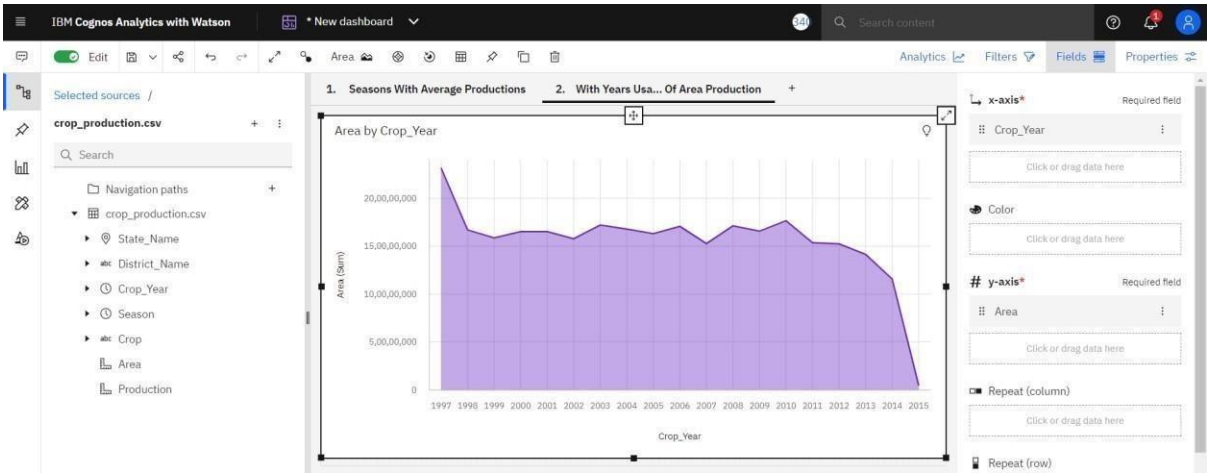


7.3 Visualization charts

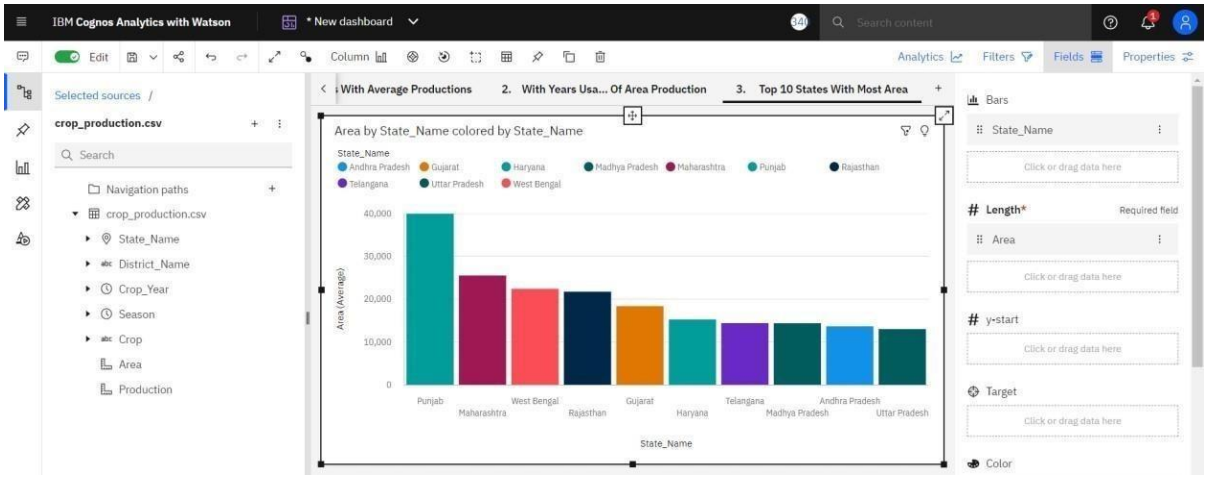
- Seasons with average productions



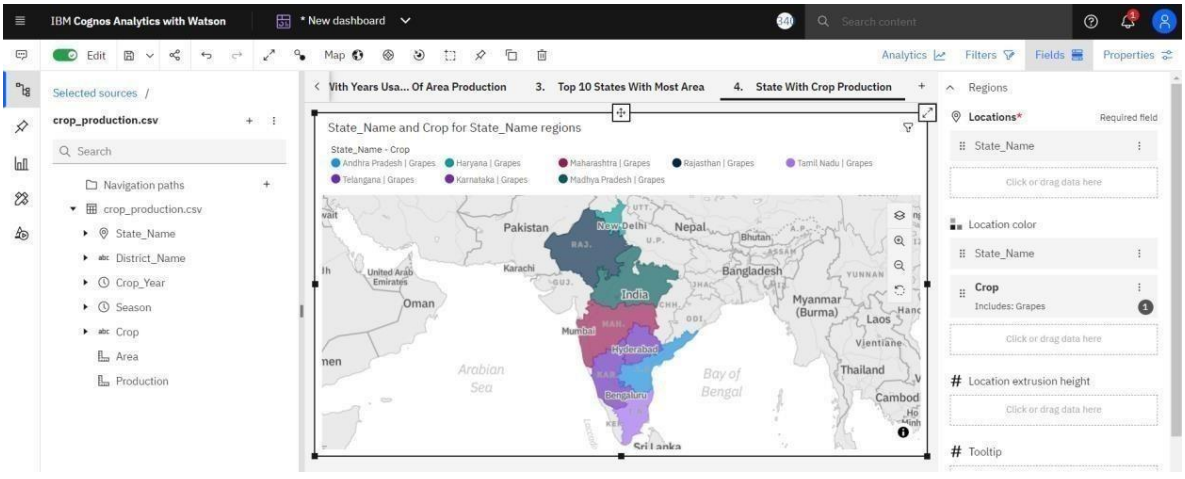
- With years usage of Area and Production



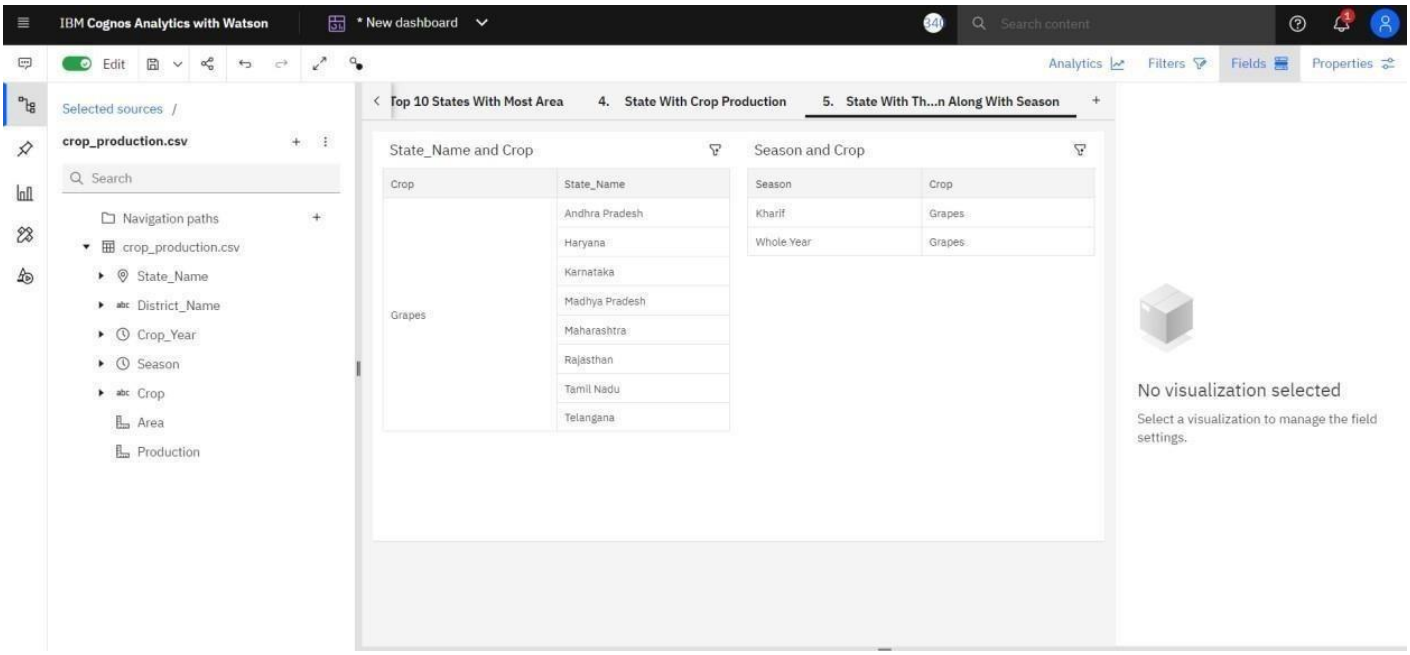
- Top 10 States with most area



- State with crop production

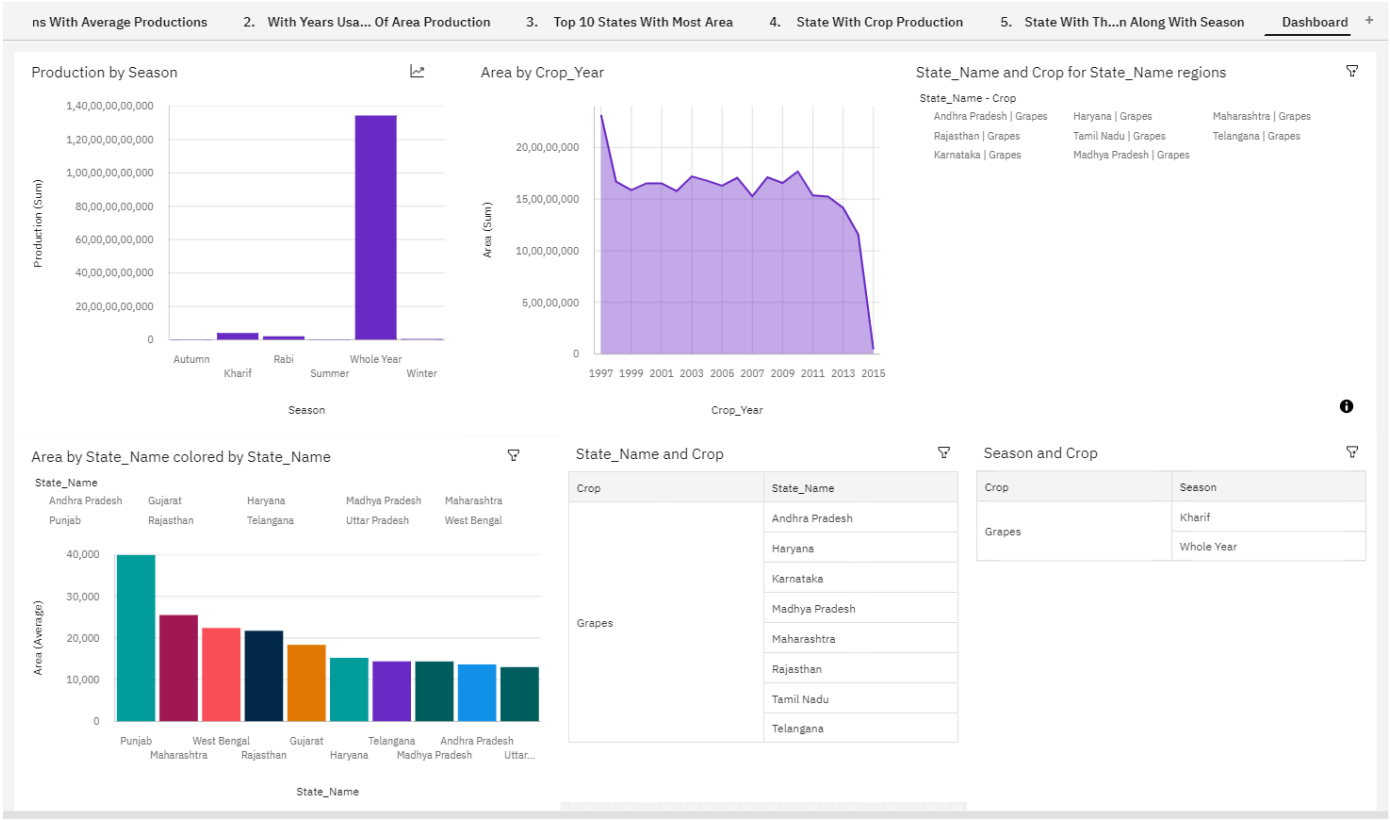


- States with the crop production along with season (Text Table)



CREATING THE DASHBOARD AND EXPORT THE ANALYTICS

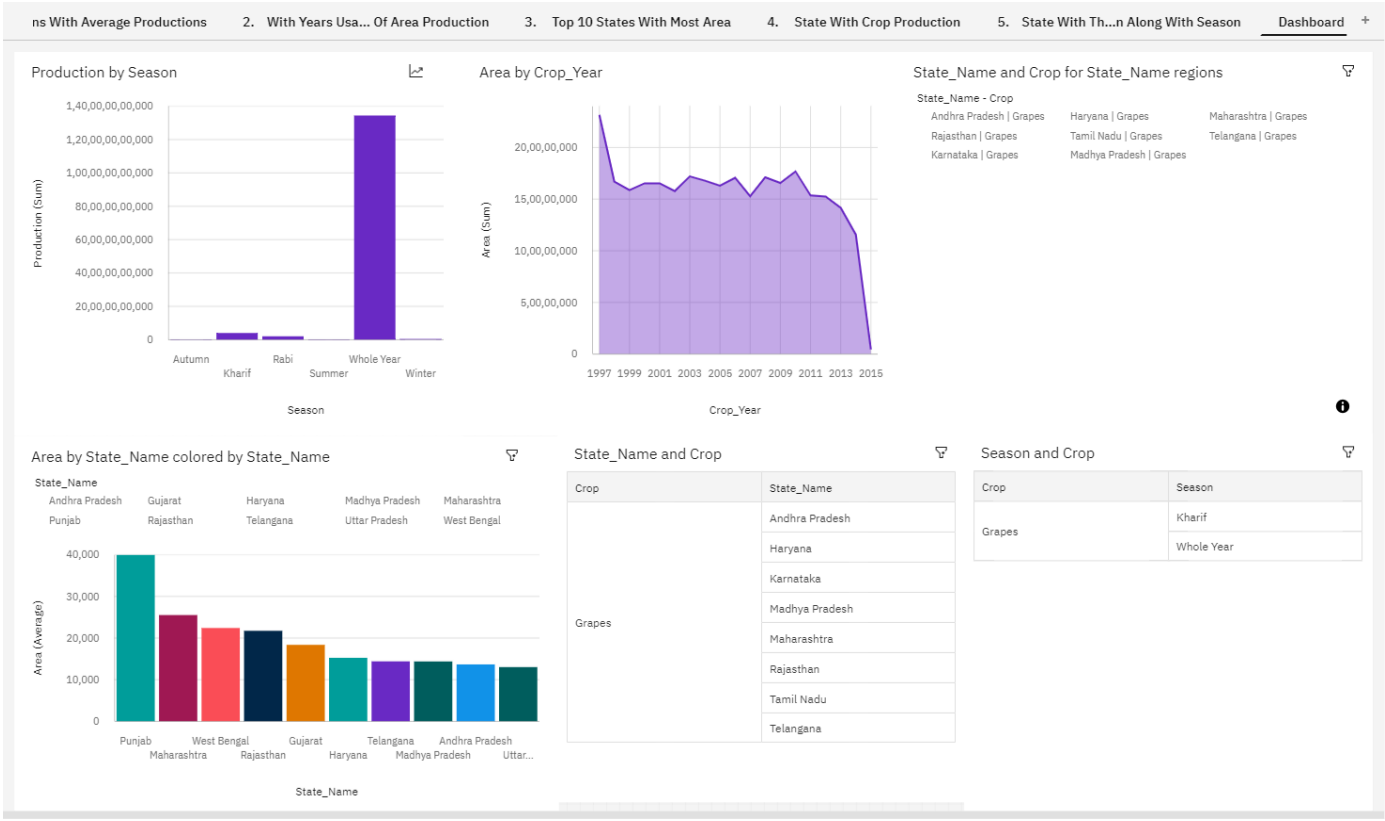
8.1 Creating the Dashboard



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 settingthen click **Export**.



ADVANTAGES & DISADVANTAGES

Advantages:

One can easily analyseand understand trendsin 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 chartsor graphs or maps. So that it wouldbe 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 takenvisiblefactors into account for theanalysis

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 detectionof hidden patternsutilising data sets related to seasons and crop yields have been conducted.Using IBM Cognos, we have observed and conducted analysis
on variouscrops grown, area, and productions in various states and districts, including

1. Seasons with average productions. We learn from these analytics which seasons have higher averageproduction and whichhave lower production.
2. Production split up per crop year. We learn from this studywhich 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 doneto determine which state, which year, and how much crop area will be produced.

FUTURE SCOPE

Farming is the means of survival as humans requirefood that is obtained only through farmingdirectly 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 resourcewill 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.

Source Code:

GitHub Link:

https://github.com/IBM-EPBL/IBM-Project-34008-1660230447

Project Demo Link:

https://drive.google.com/file/d/1HOqF_5a-m0PZabe8ppF4nV6ZT5ZxMj4A/view?usp=sharing