Project Report

Date	19 November 2022
Team ID	PNT2022TMID18592
Project Name	Estimate the crop yield using Data Analytics

1. Introduction

1.1 Project Review

India is one of the major players in the agriculture sector worldwide and it is the primary source of livelihood for about 58% of India's population. The discipline of agricultural information technology has recently undergone significant changes that have made crop yield prediction an exciting research topic. Crop yield prediction is a technique for estimating crop yield using many characteristics, including temperature, rainfall, fertilizers, insecticides, and other climatic variables and parameters. Using data analytics to analyze those parameters and provide the patterns or trends that have been followed over the past years in estimating the yield can help farmers to make the right choice in the selection of crop varieties, etc., To make the prediction results clearly we made a visualization in the dashboard and it is helpful in representing the results of the prediction to choose the appropriate crop.

1.2 Purpose

Data analytics are helpful in predicting some results for the future. The analytics on the crop yield is helpful for agriculture and it gives a better outcome for the farmers by giving the high yield of the crops with the help of the visualization, creating the dashboard.

2. Literature Survey

2.1 Existing problem

[1]Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications: Predicting crop yield based on the environmental, soil, water, and crop parameters have been a potential research topic. Deep-learning-based models are broadly used to extract significant crop features for prediction. Though these methods could resolve the yield prediction problem there exist the following inadequacies: Unable to create a direct non-linear or linear mapping between the raw data and crop yield values; and the

performance of those models highly relies on the quality of the extracted features. Deep reinforcement learning provides direction and motivation for the aforementioned shortcomings. Combining the intelligence of reinforcement learning and deep learning, deep reinforcement learning builds a complete crop yield prediction framework that can map the raw data to the crop prediction values. The proposed work constructs a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm to forecast the crop yield. The sequentially stacked layers of the Recurrent Neural network are fed by the data parameters. The Q- learning network constructs a crop yield prediction environment based on the input parameters. A linear layer maps the Recurrent Neural Network output values to the Q-values. The reinforcement learning agent incorporates a combination of parametric features with the threshold that assist in predicting crop yield. Finally, the agent receives an aggregate score for the actions performed by minimizing the error and maximizing the forecast accuracy. The proposed model ef_ciently predicts the crop yield outperforming existing models by preserving the original data distribution with an accuracy of 93.7%.

[2]Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression: The synthetic aperture radar (SAR) interferometric coherence can complement optical data for the estimation of crop growth parameters, but it has not been yet investigated for predicting crop yield. Many studies have used machine learning methods, such as neural networks, random forest, and Gaussian process regression, to estimate crop yield from remotely sensed data. However, their performance depends on the amount of available ground truth data. This study proposed Gaussian kernel regression for rice yield estimation from optical and SAR imagery using a limited amount of ground truth data. The main objective was to investigate the synergetic use of Sentinel-2 vegetation indices and Sentinel-1 interferometric coherence data through Gaussian kernel regression for estimating rice grain yield. The prediction accuracy was assessed using in situ measured yield data collected in 2019 and 2020 over Xinghua county in Jiangsu Province, China. In all cases, Gaussian kernel regression outperformed the probabilistic Gaussian regression and Bayesian linear inference. With the independently used optical and SAR data, better prediction accuracy was achieved with the optical red edge difference vegetation index (RDVI1) (r2 =0.65, RMSE=0.61 t/ha) than with the interferometric coherence (r2 = 0.52 and RMSE = 0.79 t/ha). The highest prediction accuracy can be achieved by combining RDVI1 with interferometric coherence at the heading stage (r2 = 0.81 and RMSE = 0.55 t/ha). The results suggest the value of synergy between satellite interferometric coherence and optical indices for crop yield mapping with Gaussian kernel regression.

[3]Exploiting Hierarchical Features for Crop Yield Prediction Based on 3-D Convolutional Neural Networks and Multikernel Gaussian Process: Accurate and timely prediction of crop yield based on remote sensing data is important for food security. However,

crop growth is a complex process, which makes it quite difficult to achieve better performance. To address this problem, a novel 3-D convolutional neural multi-kernel network is proposed to capture hierarchical features for predicting crop yield. First, a full 3-D convolutional neural network is constructed to maximally explore deep spatial-spectral features from multispectral images. Then, a multi-kernel learning (MKL) approach is proposed for the fusion of intro image deep spatial-spectral features and intersample spatial consistency features. Specifically, we assign a group of nonlinear kernels for each feature in the MKL framework, which provides a robust way to fit features extracted from different domains. Finally, the probability distribution of prediction results is obtained by a kernel-based method. We evaluate the performance of the proposed method on China wheat yield prediction and offer detailed and systematic analyses of the performance of the proposed method. In addition, our method is compared to several competing methods. Experimental results demonstrate that the proposed method has certain advantages and can provide better prediction performance than the competitive methods.

[4] 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 progressive agricultural yield in all seasons is an ample task and an advantageous task for every nation with respect to assessing the overall crop yield prediction and estimation. At present, a common issue worldwide is, farmers are stressed about producing higher crop yields due to the influence of unpredictable climatic changes and significant reduction of water resources worldwide. A study was carried out to collect 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 of the crop yield prediction and estimation. A study imported that the selection and usage of the most desirable crop according to the existing conditions, support achieving higher and enhanced crop yield S. Athmaja, M. Hanumanthappa, and V. Kavitha,a survey of machine learning algorithms has presented effective strategies for big data analytics. All over the world, agricultural people 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 agricultural industry farmers are facing a lot of problems with climate change, and soil quality. So this, makes it the farmers very hard to get higher yields from their crops. The main objective of this project is to predict the crop yield that will be more useful for the farmers to plan the harvest and to get a higher yield.

3 Ideation and Proposed Solution

3.1 Empathy Map Canvas

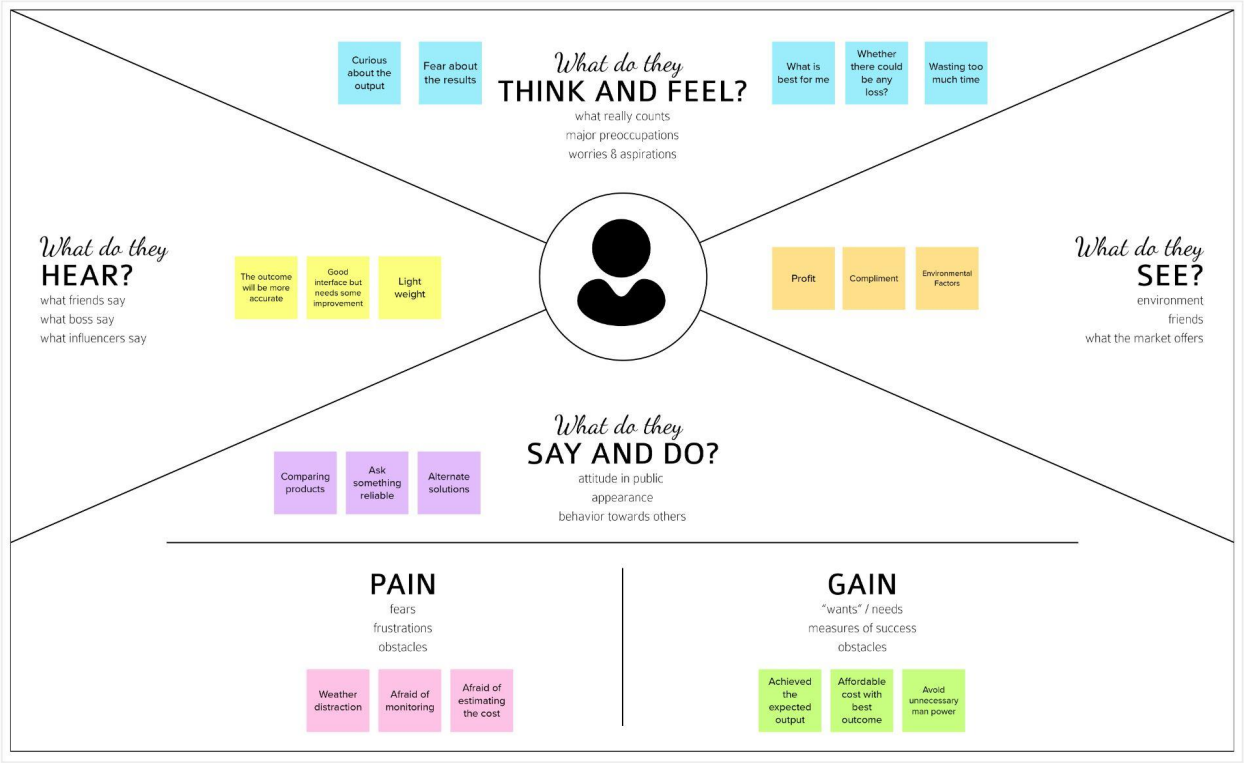
Edit this template
Right-click to unlock

Empathy Map Canvas

Gain insight and understanding on solving customer problems.

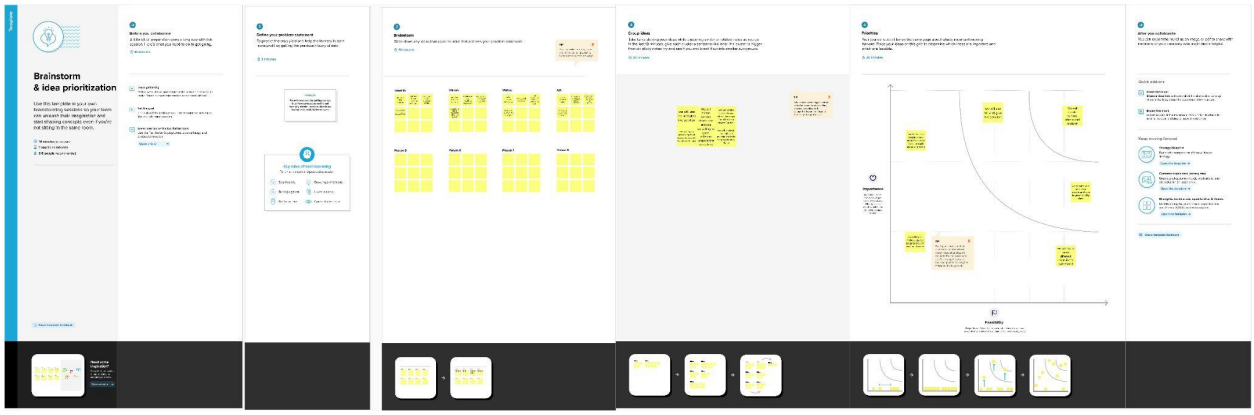
1

Build empathy and keep your focus on the user by putting yourself in their shoes.



Share your feedback

3.2 Ideation and Brainstorming

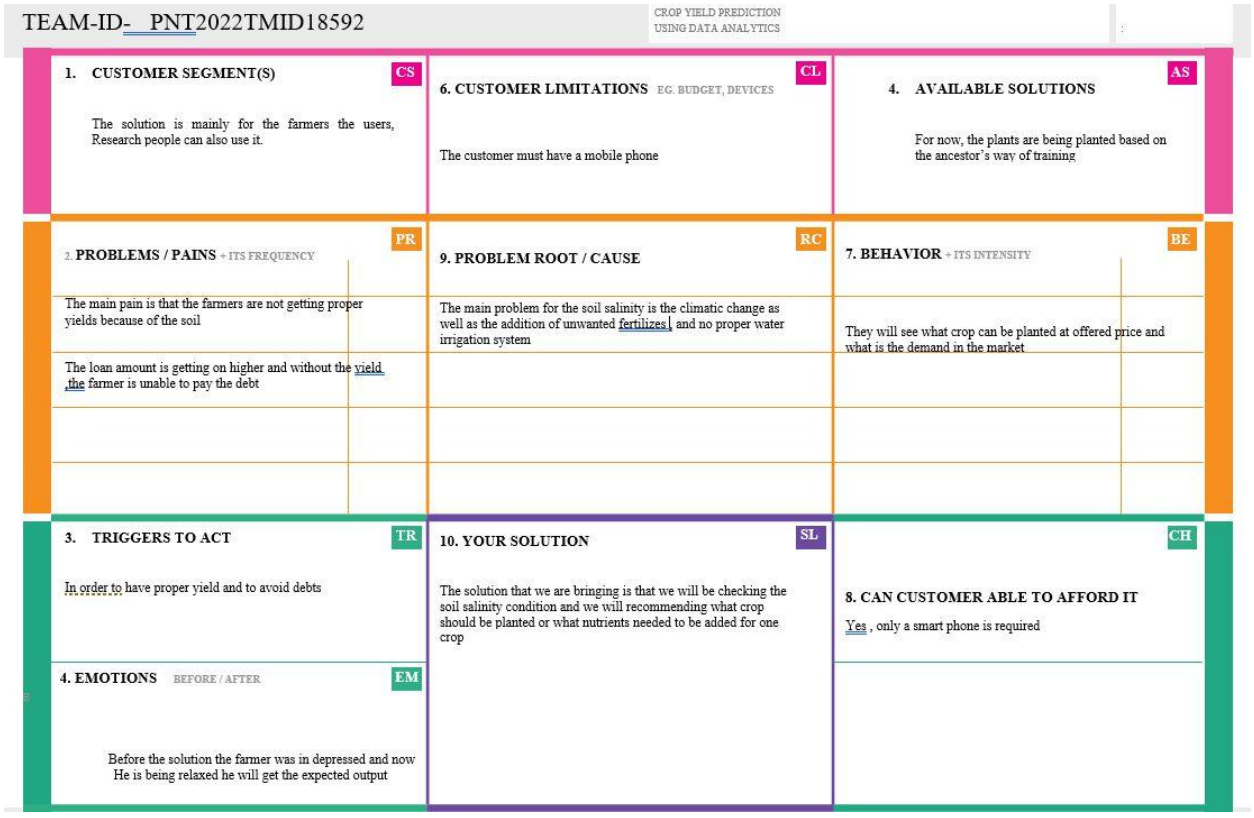


3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	A Farmer is not being able to get a proper yield Because the land has not been used for a long period of time.
2.	Idea / Solution Description	The idea is to find the salinity level of the soil and based on that we will be suggesting what crop will be grown in that land in order to get a good yield also we will be indicating the nutrients required so that the crop required to grow in that land
3.	Novelty / Uniqueness	The uniqueness is that we will be finding the nutrient level of the soil and by the regression process we will say how much the nutrient is to be added.

4.	Social Impact / Customer Satisfaction	The Customer that is farmer will feel stress-free and will avoid the tension of crop quality and help to avoid getting an unwanted loan.
5.	Business Model (Revenue Model)	Since it's just an app for getting information about the usage of land there is no business motive. But it can be added as a subscriptive pack for research purposes in the future
6.	Scalability of the Solution	As of now, it is a very urgent need for the farmers how to use the land this will help to take unwanted loans from the bank for the loss and suffering.

3.4 Problem Solution Fit



4. Requirement Analysis

4.1 Functional Requirement

FR.No	Functional Requirement(Epic)	Sub Requirement(Story /Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Profile	Log in Access the profile
FR-4	Give the required data	Take the data given by the user as the input for the analysis
FR-5	Analysis	Analyse the yield of the crop from the data given by the user
FR-6	FR-6 Estimation or Predict the data	an Estimate the crop yield from the analysis, using the software from the data given by the user

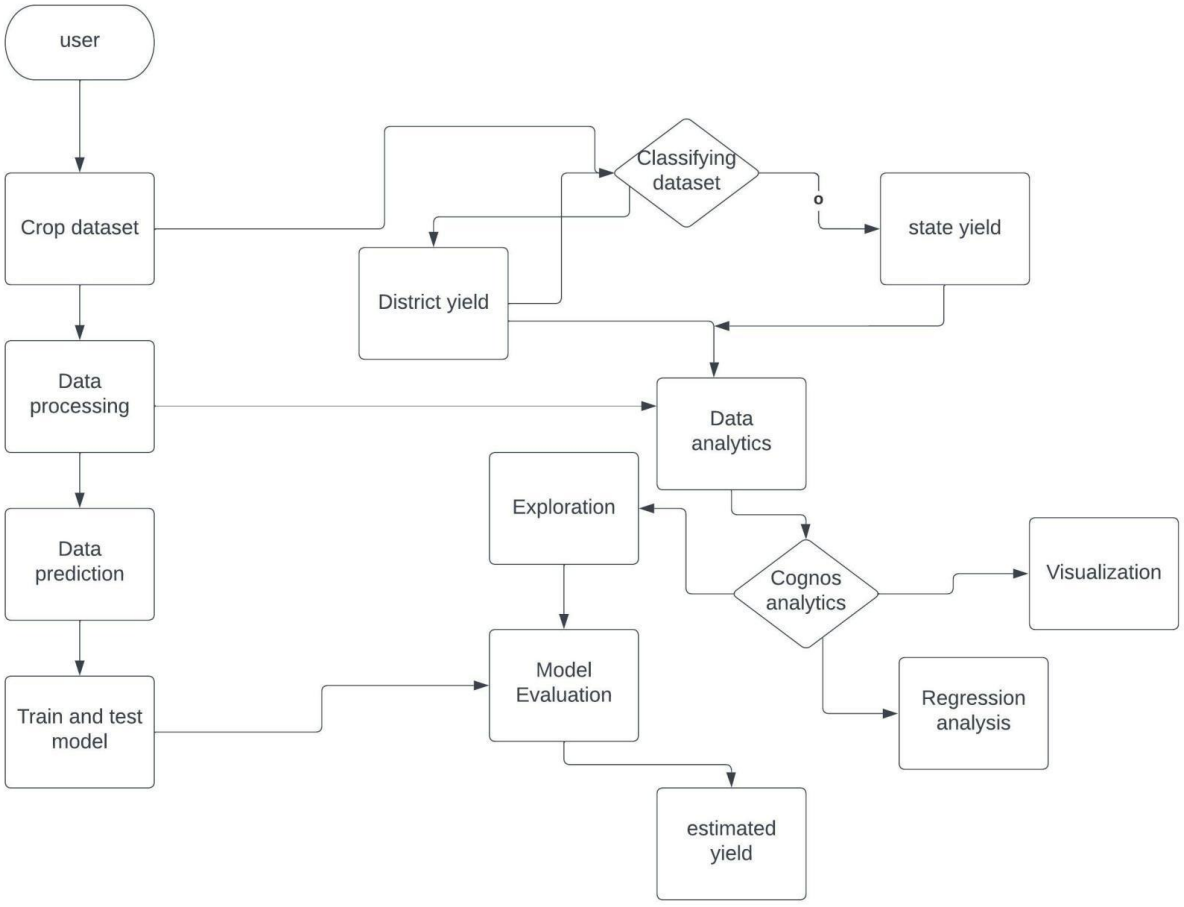
4.2 Non-Functional Requirements

NFR.No	Non-Functional Requirement	Description
NFR-1	Usability	Crop recommendations are created and saved, then these recommended crops are sown by farmers for increased crop yield.
NFR-2	Security	The software keeps user's information more secure.
NFR-3	Reliability	Creating interactive dashboards which is easy to understand and useful for the users.

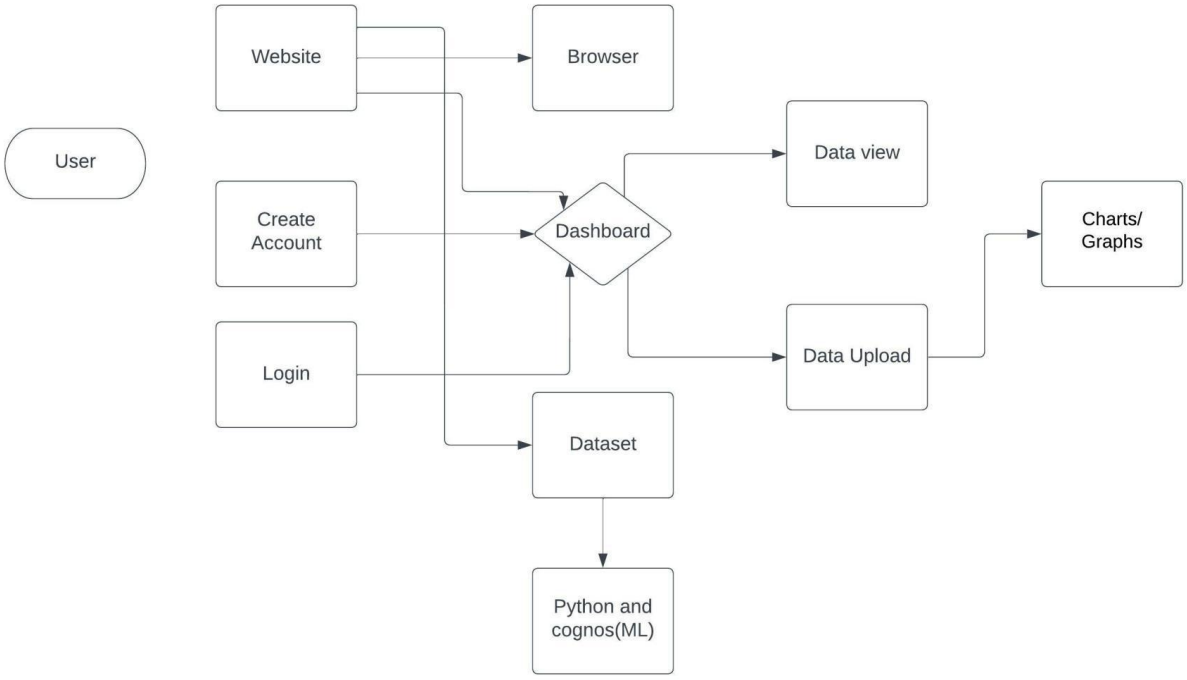
NFR-4	Performance	It is user-friendly software and has high performance.
NFR-5	Availability	The software application is easily available for every user and access is easy for them.
NFR-6	Scalability	The proposed system allows the implementation of a flexible methodology that can be used to estimate the yield of crops in different types of lands.

5. Project Design

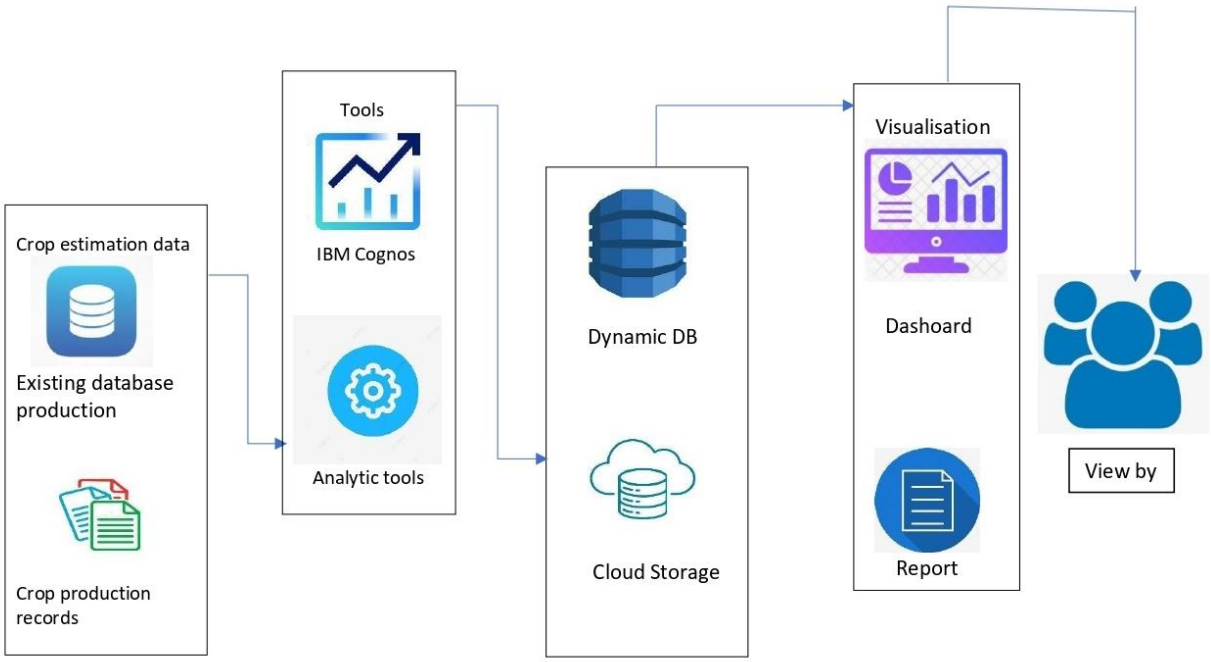
5.1 Data Flow Diagram



Simplified View



5.2 Solution & Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user and Laptop users)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	Can use the methods provided in the dashboard		Medium	Sprint-1
	Invest	USN-7	With help of desired results obtained from application, making profit or loss	Gain or Loss	High	Sprint-2
Administrator	Updating data		Collecting the data and storing it	Checking and updating dataset	High	Sprint-1
Customer (Web User)	Accessing the resources	USN -8	Using my own credentials for accessing the data	These resources cannot be accessed by others but only me	High	Sprint -1
	Satellite Visioning	USN -9	Having a view with geographic data		Medium	Sprint-2
Customer tools	Tools	USN -10	Analysis is performed by tools like cognos analytics	Ease of accessing the tools	High	Sprint 2

6. Project Planning and Scheduling

6.1 Sprint Planning and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	Register by entering my id card and request.	2	High	Vikram A K Vishnu Pragadeesvaran S Ajit Rajkumar M S Vasanth Kumar R
		USN-3	Register for the application through mail	2	Medium	Vikram A K Vishnu Pragadees varan S
	Login	USN-4	Call and request or Approach for the dataset	4	High	Ajit Rajkumar MS Vasanth Kumar R
	Working with the Dataset	USN-5	To work on the given dataset and Understanding the given Dataset.	2	High	Vikram A K Vishnu Pragadeesvaran S Ajit Rajkumar M S Vasanth Kumar R
		USN-6	Loading the dataset to Cloud platform then Build the required Visualizations.	10	High	Vikram A K Vishnu Pragadeesvaran S
Sprint-2	Data Visualization Chart	USN-7	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	Ajit Rajkumar M S Vasanth Kumar R
			*Showcase the Yearly usage of Area in Crop Production.	4	Medium	Vikram A K Vishnu Pragadeesvaran S
			Build a visualization to show case top 10 States in Crop Yield Production by Area.	4	Medium	Ajit Rajkumar MS Vasanth Kumar R
			Build the required Visualization to showcase the Crop Production by State.	4	Medium	Ajit Rajkumar MS Vishnu pragadeesvaran
			Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation.	4	Medium	Vikram A K Vasanth Kumar R
Sprint-3	Creating The dashboard	USN-8	Create the Dashboard by using the created visualizations.	20	High	Ajit Rajkum ar M S Vikram A K
Sprint-4	Export The Analytics	USN-9	Export the created Dashboard	20	High	Vishnu Pragadeesvaran S Vasanth Kumar R

6.2 Project 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	7 Days	24 Oct 2022	30Oct 2022	20	29 Oct 2022
Sprint-2	20	7 Days	31 Oct 2022	06 Nov 2022	20	05 Nov 2022
Sprint-3	20	7 Days	07 Nov 2022	13 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 an understanding of crop production in India. Download the dataset from the below link. It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production-related detail. 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.

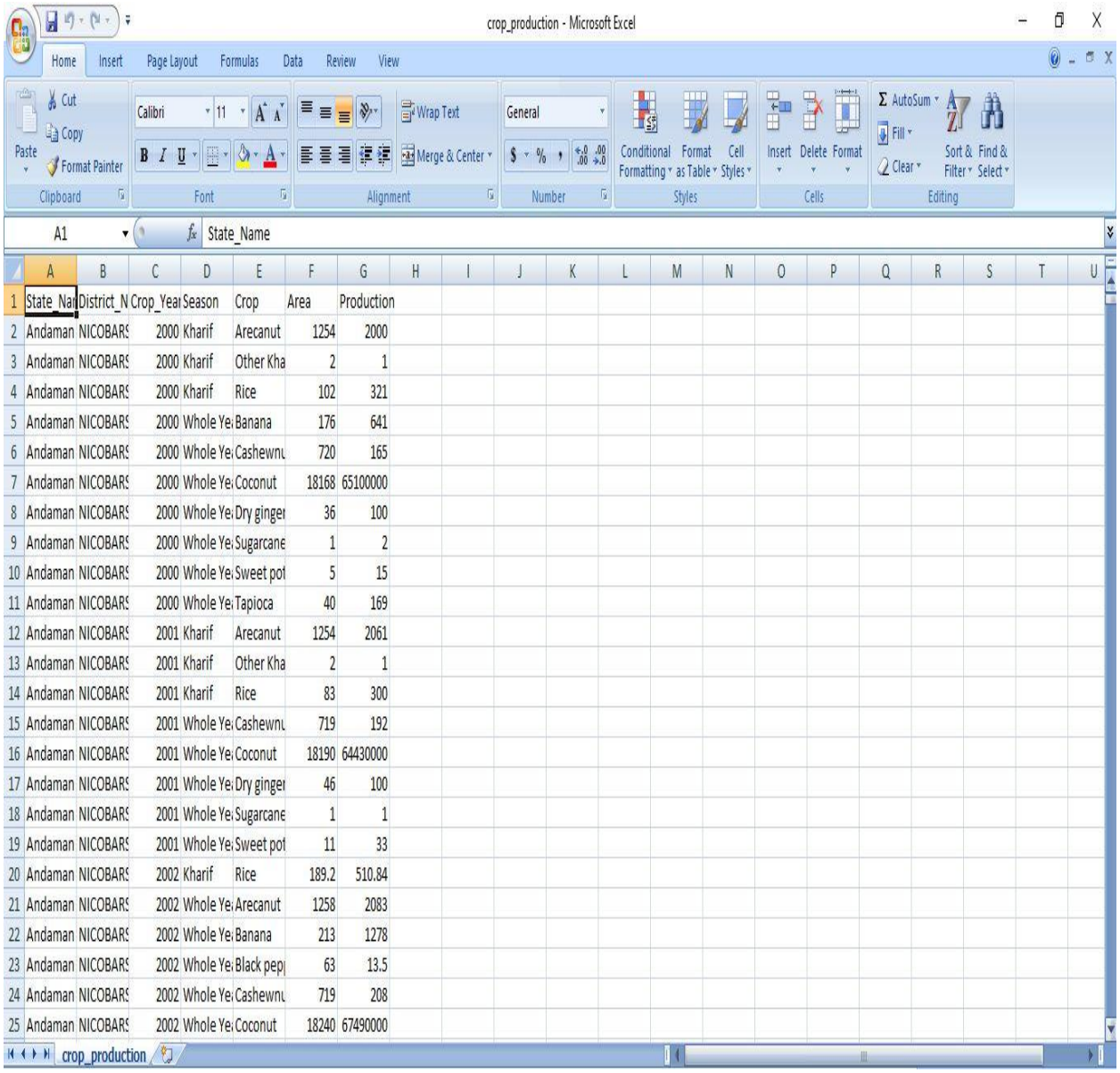
3. Crop Year- contains the crop years.

4. Season – Different seasons for crop production.

5. Area- Total number of areas covered.

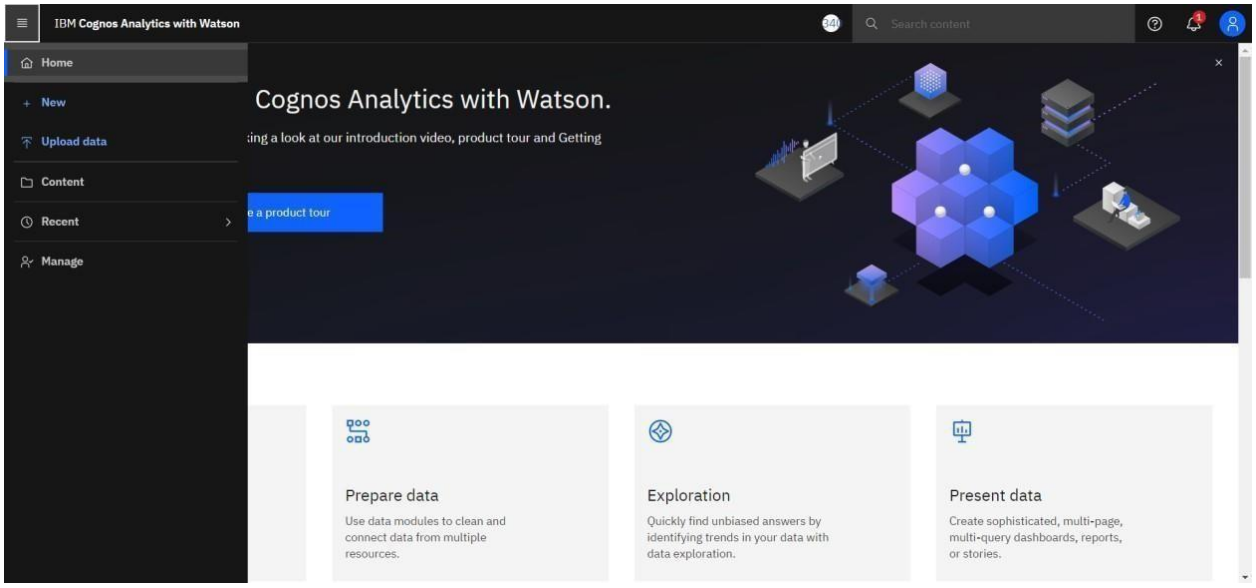
6. Production- production of crops.

The data format is shown in the image below

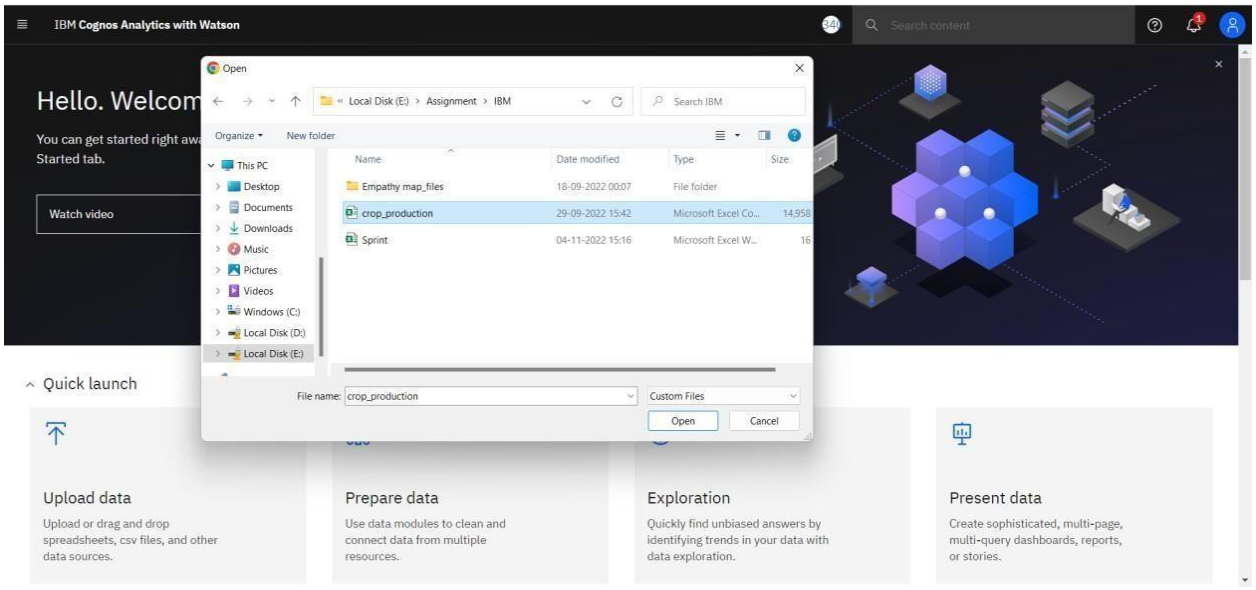


	A	B	C	D	E	F	G
1	State_Nar	District_N	Crop_Year	Season	Crop	Area	Production
2	Andaman	NICOBARS	2000	Kharif	Arecanut	1254	2000
3	Andaman	NICOBARS	2000	Kharif	Other Kha	2	1
4	Andaman	NICOBARS	2000	Kharif	Rice	102	321
5	Andaman	NICOBARS	2000	Whole Ye	Banana	176	641
6	Andaman	NICOBARS	2000	Whole Ye	Cashewnu	720	165
7	Andaman	NICOBARS	2000	Whole Ye	Coconut	18168	65100000
8	Andaman	NICOBARS	2000	Whole Ye	Dry ginger	36	100
9	Andaman	NICOBARS	2000	Whole Ye	Sugarcane	1	2
10	Andaman	NICOBARS	2000	Whole Ye	Sweet pot	5	15
11	Andaman	NICOBARS	2000	Whole Ye	Tapioca	40	169
12	Andaman	NICOBARS	2001	Kharif	Arecanut	1254	2061
13	Andaman	NICOBARS	2001	Kharif	Other Kha	2	1
14	Andaman	NICOBARS	2001	Kharif	Rice	83	300
15	Andaman	NICOBARS	2001	Whole Ye	Cashewnu	719	192
16	Andaman	NICOBARS	2001	Whole Ye	Coconut	18190	64430000
17	Andaman	NICOBARS	2001	Whole Ye	Dry ginger	46	100
18	Andaman	NICOBARS	2001	Whole Ye	Sugarcane	1	1
19	Andaman	NICOBARS	2001	Whole Ye	Sweet pot	11	33
20	Andaman	NICOBARS	2002	Kharif	Rice	189.2	510.84
21	Andaman	NICOBARS	2002	Whole Ye	Arecanut	1258	2083
22	Andaman	NICOBARS	2002	Whole Ye	Banana	213	1278
23	Andaman	NICOBARS	2002	Whole Ye	Black pep	63	13.5
24	Andaman	NICOBARS	2002	Whole Ye	Cashewnu	719	208
25	Andaman	NICOBARS	2002	Whole Ye	Coconut	18240	67490000

7.2 Loading the Dataset



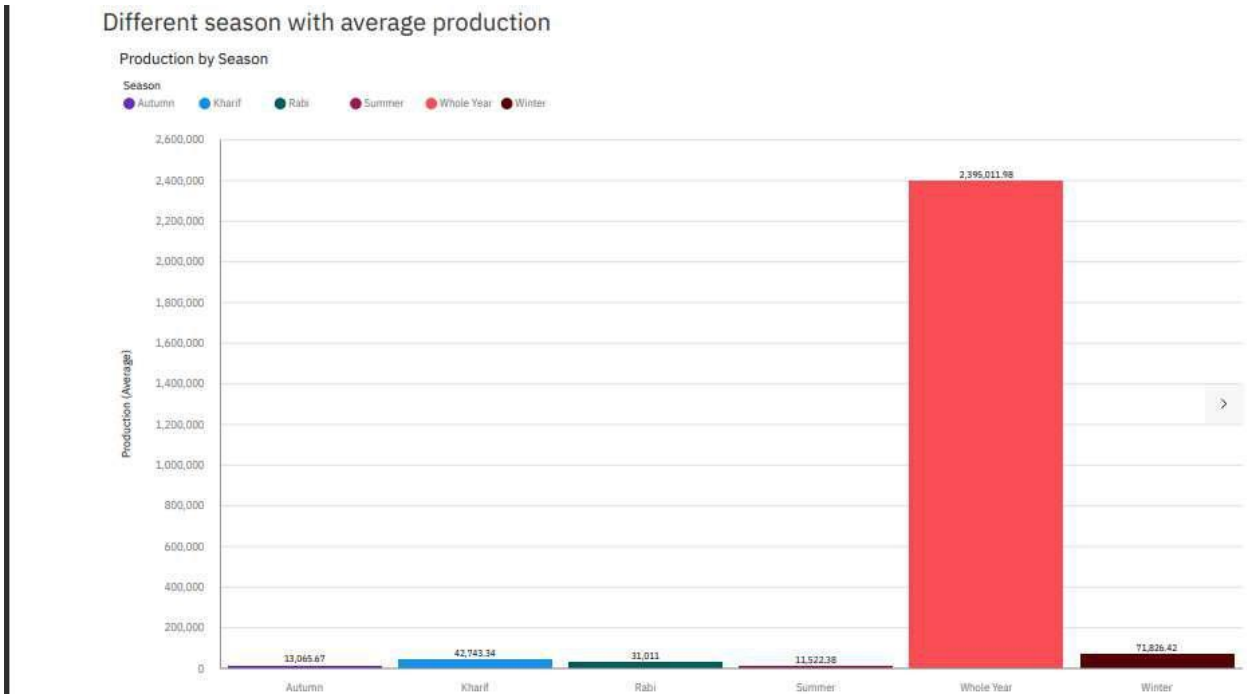
Click the open menu in the top left corner.



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

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

7.3 Visualization Charts



States with crop production along with seasons

State_Name and Crop

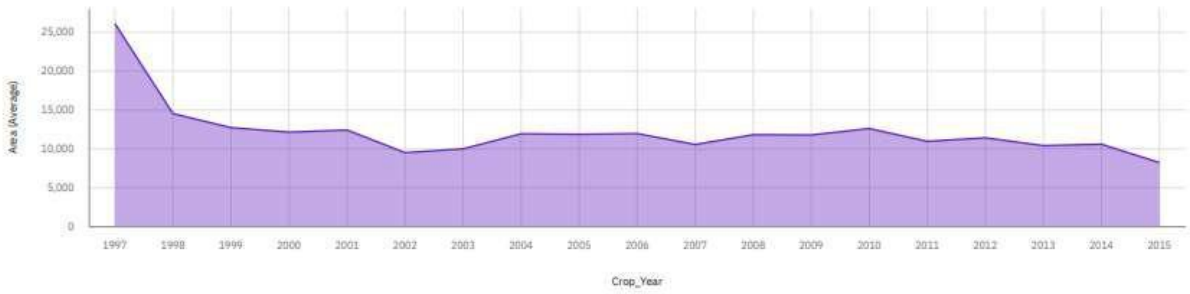
Crop	State_Name
Grapes	Andhra Pradesh
	Haryana
	Karnataka
	Madhya Pradesh
	Maharashtra
	Rajasthan
	Tamil Nadu
	Telangana

Season and Crop

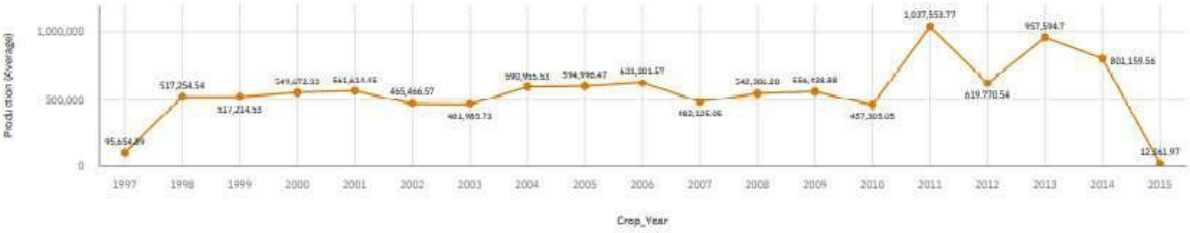
Crop	Season
Grapes	Kharif
	Whole Year

With years usage Area and Production

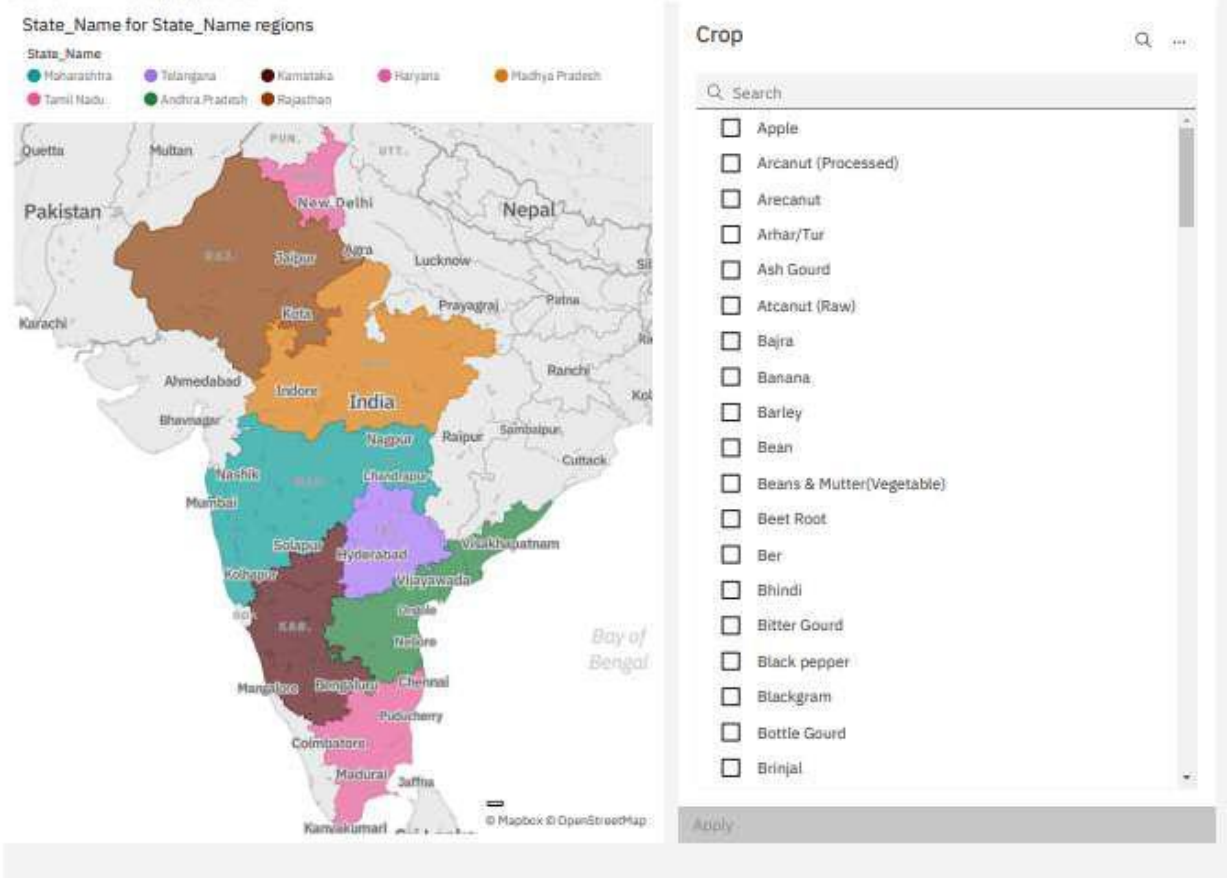
Area by Crop_Year



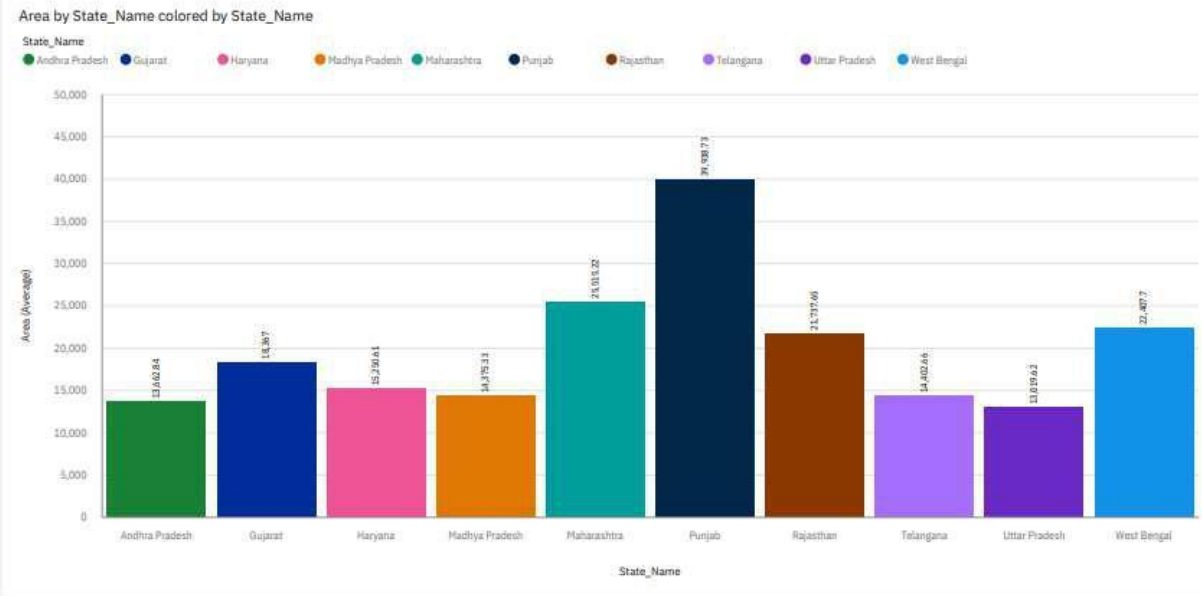
Production by Crop_Year



State with crop production

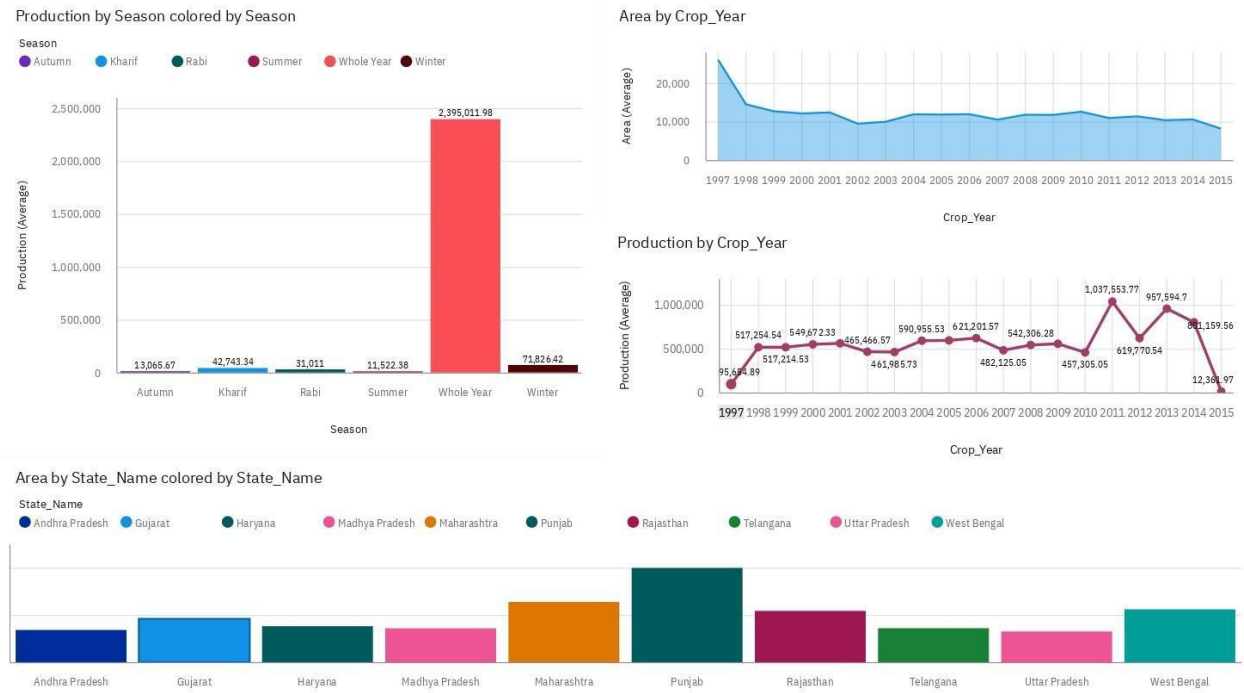


Top 10 States with most area



8. Creating Dashboard and Export the Analytics

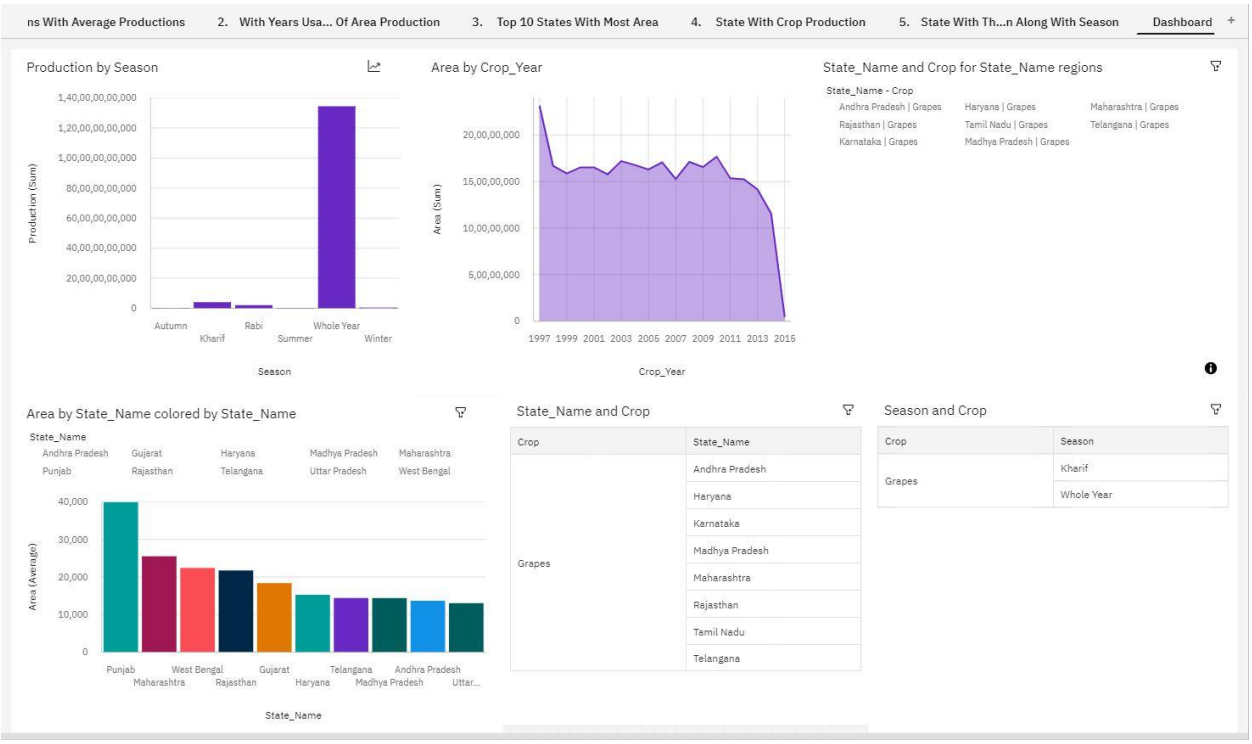
8.1 Creating the Dashboard



8.2 Exporting 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 and disadvantages

Advantages

One can easily analyze and understand trends in cropping patterns, and seasonal behavior 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 knowledge 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 analysis 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 innovations. From the analysis dashboard, it has been noted that analyses of agricultural productivity and the detection of hidden patterns utilizing data sets related to seasons and crop yields have been conducted. Using IBM Cognos, we have observed and conducted analysis on various crops grown areas, 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, the study was done to determine which state, which year, and how much crop area will be produced.

11. FUTURE SCOPE

Agriculture is the major means of survival for humans as well as for other living things. It is very difficult to analyze crop production and the population of humans is rapidly increasing. So we should know the correct timing, crop, and place for the cultivation. As food is the basic need of humans, the requirement of getting the maximum yields using optimal resources will become a necessity in near future as a result of a growing population.

12. APPENDIX

Source Code

Github Link

Projecr Demo Link