ESTIMATE THE CROP YIELD USING DATA ANALYTICS

TEAM ID - PNT2022TMID02564

TEAM LEADER– ANURAAGAVI M R (2116191001008)

TEAM MEMBER – ADITYA VENKATESH (2116191001003)

TEAM MEMBER – HARINIVAS M (2116191001020)

TEAM MEMBER – HARIPRIYA K (2116191001021)

ABSTRACT:

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. IBM Cognos Analytics integrates reporting, modelling, analysis, exploration, dashboards, stories, and event management so we can understand our organization\'s data, and make effective decisions. A dashboard helps us to monitor events or activities at a glance by providing key insights and analysis about our data on one or more pages or screens. In this project, we visualize, analyze and gain most of the insights by creating a dashboard. The main challenge in using data in agriculture is identification of effectiveness of data analytics. Efforts are going on to understand how data analytics can agriculture productivity. The present study gives insights on various data analytics methods applied to crop yield prediction and also signifies the important lacunae points in the proposed area of research.

Introduction

I. INTRODUCTION

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. Rainfall conditions also influences the rice cultivation. In this context, the farmers necessarily requires a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is enormous in Indian agriculture. The data when become information is highly useful for many purposes. IBM Cognos Business Intelligence is a web-based integrated business intelligence suite by IBM. It provides a toolset for reporting, analytics, score carding, and monitoring of events and metrics. The software consists of several components designed to meet the different information requirements in a company. IBM Cognos has components such as IBM Cognos Framework Manager, IBM Cognos Cube Designer, IBM Cognos Transformer. Cognos Analysis Studio helps business users get fast answers to business-related queries. Reporting studio allows you to create pixel-perfect reports for your organization. Cognos event studio allows you to assign a specific event that sends a notification to the stakeholder in your organization. Cognos Metric Studio allows you to monitor and analyze business metrics of your organization by building a scorecard environment.

1.1 PROJECT OVERVIEW:

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 [11]. The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and gets best prices for their crops.

1.2 PURPOSE:

Agriculture is the widest economic sector and has an important role regarding the framework of socio-economic fabric of India. Farming depends on various factors like climate and economic factors like temperature, irrigation, cultivation, soil, rain fall, pesticide and fertilizers. Historical information regarding crop yield provides major input for companies engaged in this domain. The estimation of production of crop helps these companies in planning supply chain decision like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield. Farmers experience was the only way for prediction of crop yield in the past days. Technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc.

1. LITERATURE SURVEY:

1.1 EXISTING PROBLEM:

A. M. A. Jayaram and Netra Marad, "Fuzzy interference Systems for Crop Prediction", Journal of Intelligent Systems, 2012, 21(4), pp.363-372[1].

Prediction of crop yield is signi?cant in order to accurately meet market requirements and proper administration of agricultural activities directed towards enhancement in yield. Several parameters such as weather, pests, biophysical and morphological features merit their consideration while determining the yield. However, these parameters are uncertain in their nature, thus making the determined amount of yield to be approximate. It is exactly here that the fuzzy logic comes into play. This paper elaborates an attempt to develop fuzzy inference systems for crop yield prediction. Physio morphological features of Sorghum were considered. A huge database (around 1000 records) of physio morphological features such as days of 50 percent ?owering, dead heart percentage, plant height, panicle length, panicle weight and number of primaries and the corresponding yield were considered for the development of the model. In order to ?nd out the sensitivity of parameters, one-to-one, two-to-one and three-to-one combinations of input and output were considered. The results have clearly shown that panicle length contributes forth yield as the lone parameter with almost one-to-one matching between predicted yield and actual value while panicle length and panicle weight in combination seemed to play a decisive role in contributing for the yield with the prediction accuracy re?ected by very low RMS value.

B. P. Vindya "Agricultural Analysis for Next Generation High Tech Farming in Data Mining" , Anna University, Trichy, Tamilnadu, India, 5 May 2015[2]. Recent developments in Information Technology for agriculture field have become an interesting research area to predict the crop yield [1].

In today's world, the amount of information stored has been enormously increasing day by day which is generally in the unstructured form and cannot be used for any processing to extract useful information using mining technique. This paper presents a brief analysis of data mining methods and agriculture techniques, farm types, soil types, prediction using Multiple Linear Regression (MLR) technique for the selected region. This work mainly focuses on analyzing the agricultural analysis of organic farming and inorganic

farming, time cultivation of the plant, profit and loss of the data and analyzes the real estate business land in a specific area and comparison of irrigated and unirrigated land. It concentrates organic, inorganic and real estate data sets from which the prediction in agriculture will be achieved. The purpose is to estimate difference in efficiency and prediction between organic and inorganic farming. This work aims at finding suitable data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities.

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 Maharashtra, India. The precise quantification of the rice productivity in various climatic conditions can help farmer to understand the optimum condition for the higher rice crop yield [8].

Simulation models based on field experiment are valuable technologies for studying and understanding crop yield gaps, but one of the critical challenge remain with these methods is scaling up of these approach to assess the data collated between different time intervals from the broader geographical regions. Satellite retrieved data have frequently been revealed to present data sets that, by itself or

in grouping with other information and model designs, can precisely determine the yields of crop in agricultural lands. The yield maps developed shall provide an unique opportunity to overcome both spatial and temporal based scaling up challenges and thus improve the ideology of crop yield gaps prediction. A review was conducted to discuss the applications of remote sensing technology to determine the impact and causes of yield gaps. Even though the example discussed by the research group demonstrates the usefulness of remote sensing in the prediction of yield gaps, but also many areas of possible application with respect to the crop yield assessment, prediction and improvement remain unexplored. Study proposed two less complicated, easily assessable methods to determine and quantify the yield gaps between various agricultural fields. First method works closely with the constructive maps representing the average crop yields, it can be used directly to accesses specific crop yield influencing factors for further studies whereas the second method use the remote sensing technology to retrieve the data for providing the useful information regarding the crop yield prediction and estimation [14].

1.2 REFERENCE:

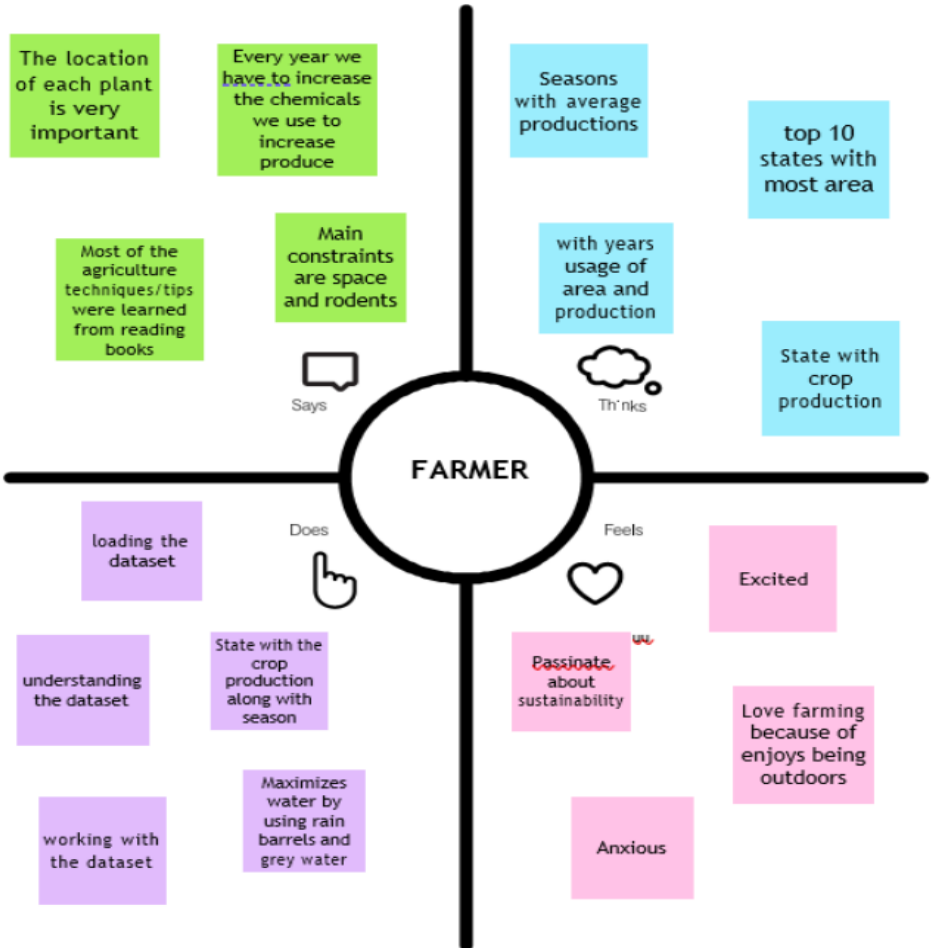
- https://www.degruyter.com/document/doi/10.1515/jisys-2012- 0016/html
- https://ieeexplore.ieee.org/document/8697806
- https://www.researchgate.net/publication/339102917_Big_data_analytics_in_Agriculture

1.3 PROBLEM STATEMENT DEFINITION:

To create a dashboard and perform analysis of crop production in India using IBM Cognos analytic platform. Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. As per this project we will be analyzing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

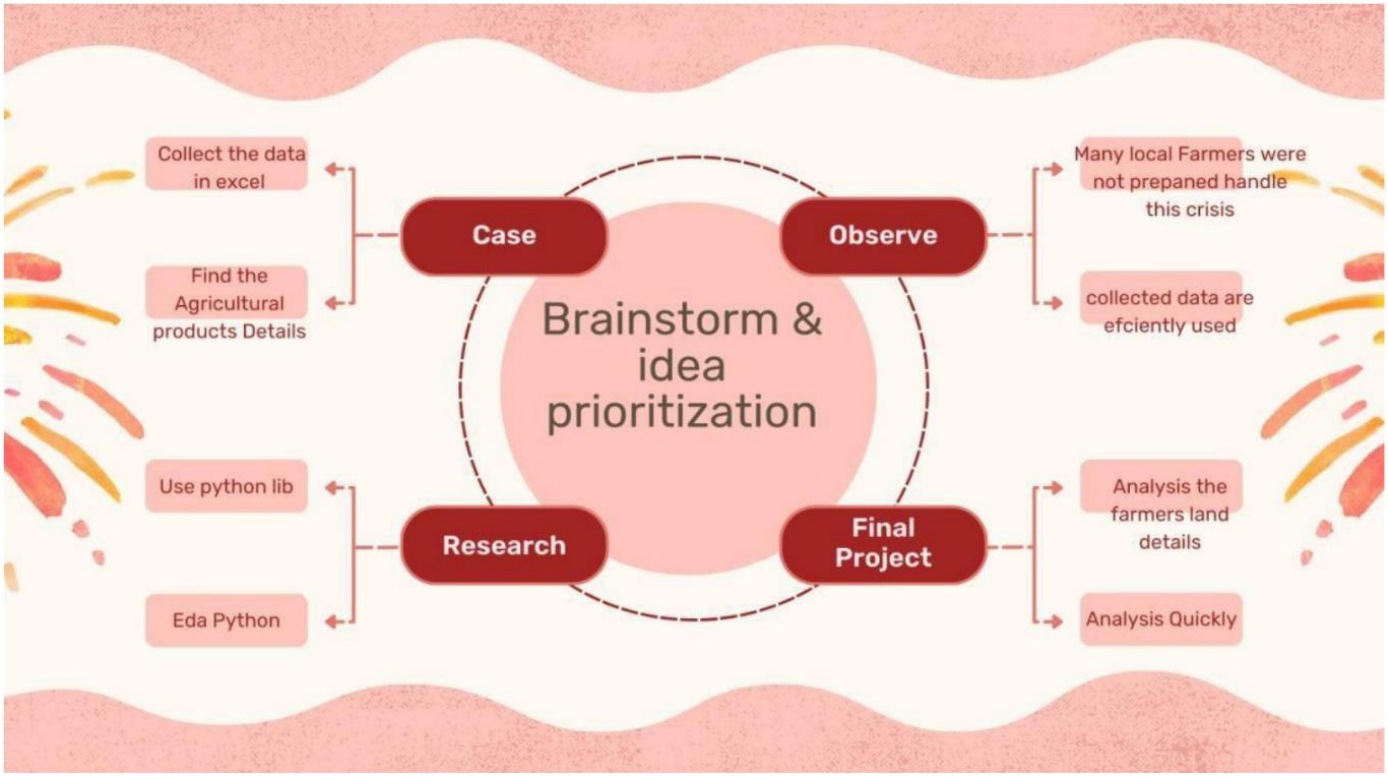
2. IDEATION & PROPOSED SYSTEM:

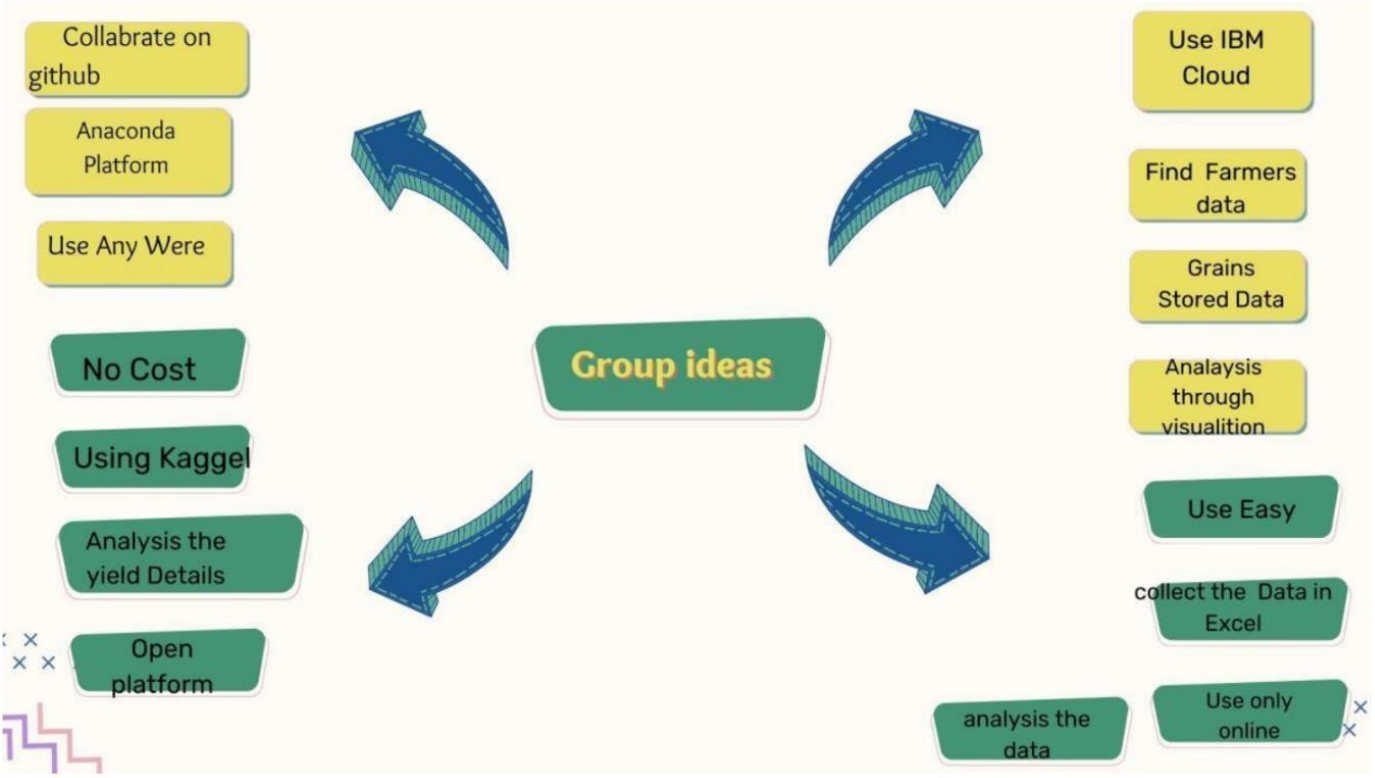
2.1 EMPATHY MAP CANVAS:



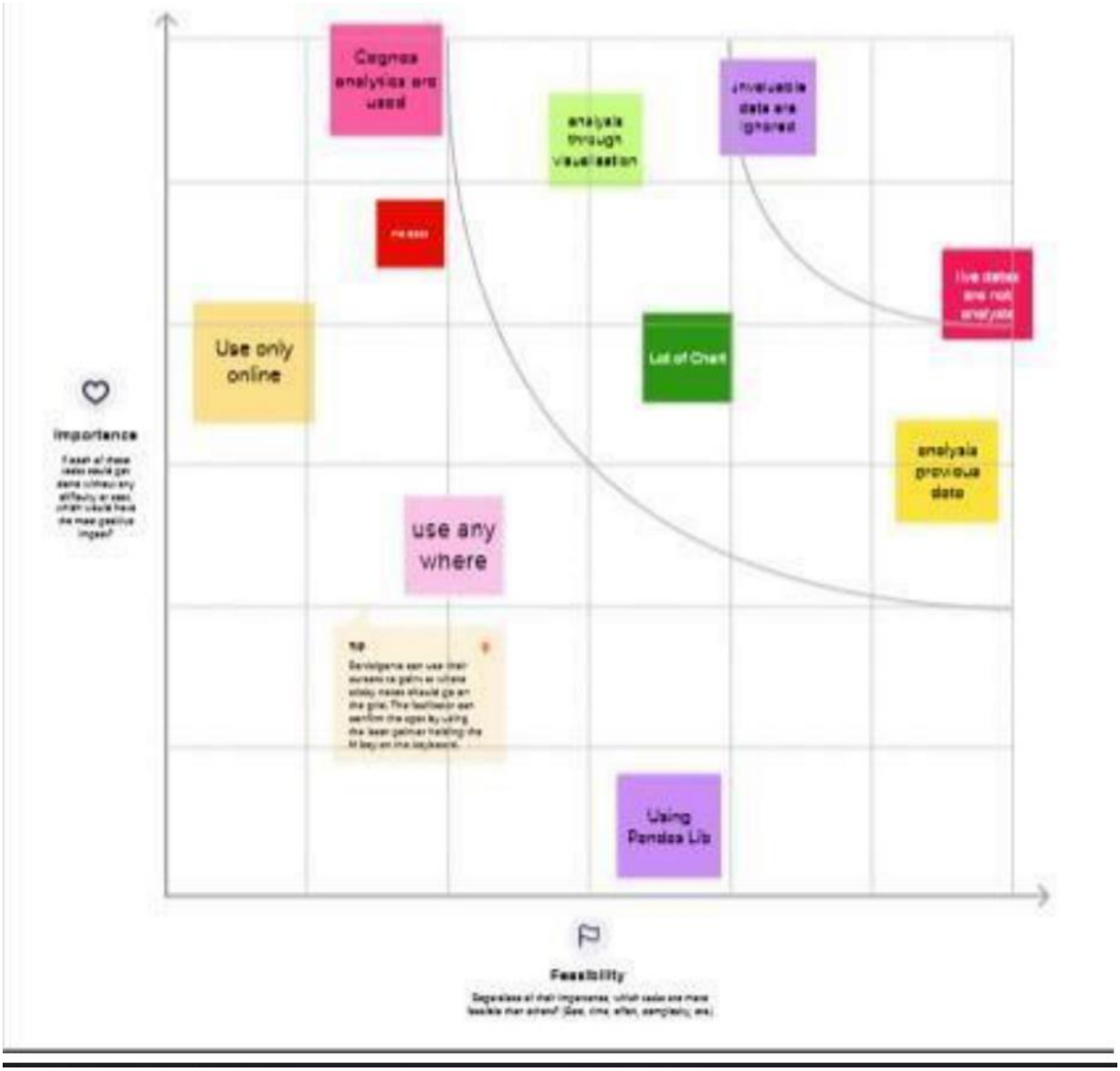
2.2 IDEATION & BRAINSTORMING:

A) BRAINSTORMING:





B) IDEA PRIORITIZATION:

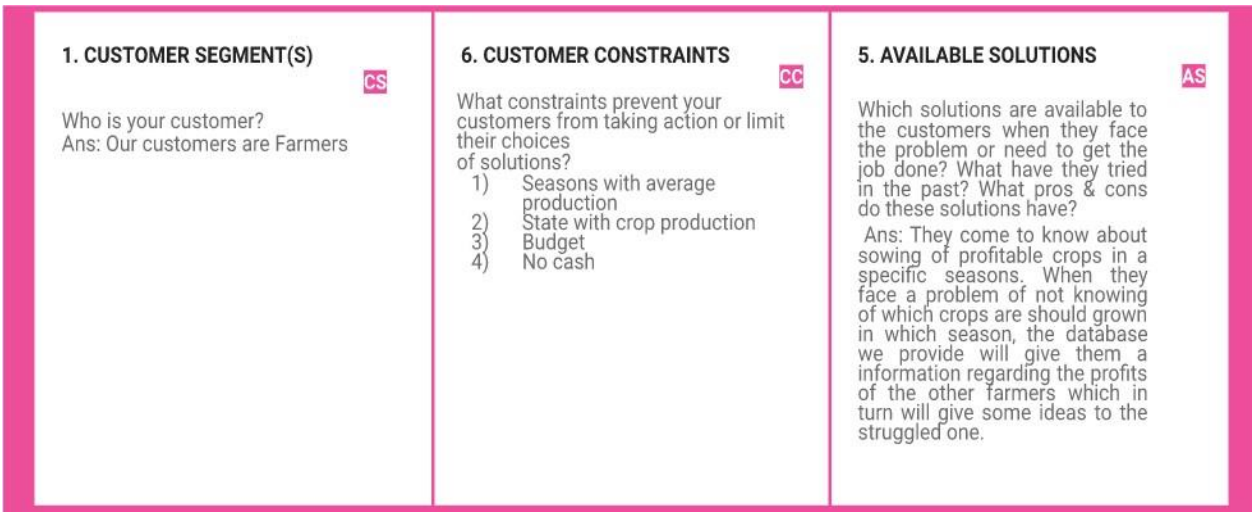


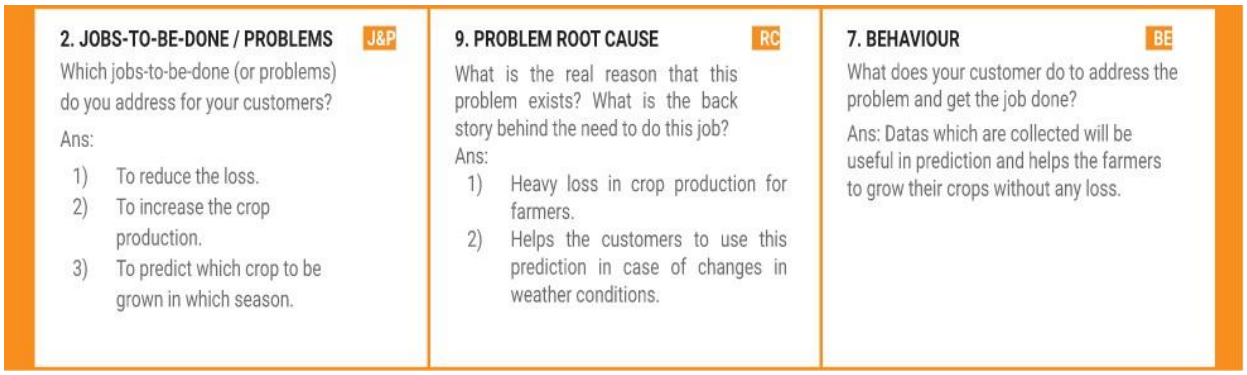
2.3 PROPOSED SOLUTION:

S.No	Parameter	Description
1.	Problem Statement (Problem to be solved)	Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. As per this project we will be analyzing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.
2.	Idea / Solution description	The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and get best prices for the crops.
3.	Novelty / Uniqueness	If the production of a crop observes a declining trend then they can plan to implement the **schemes** at an **early stage**. This in return will **save** the state from **shortage of a product**
4.	Social Impact / Customer Satisfaction	It is used to monitor progress towards a global set by governments, non-governmental organizations, and other stakeholders.
5.	Business Model (Revenue Model)	According to the revenue side it will yield more revenue to the farmers as well as to the governments.
6.	Scalability of the Solution	The data which are present in the datasets will be up to date. So it will help the customer to act accordingly

2.4 PROBLEM SOLUTION FIT:

The Problem-Solution Fit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioral patterns and recognize what would work and why.





3. TRIGGERS TR

What triggers customers to act?

Ans: The data which are present in the datasets in an appropriate manner. Due to the diagrammatic representation of the models, the farmers who are not well educated can also be able to work on this and get benefits.

4. EMOTIONS: BEFORE / AFTER EM

How do customers feel when they face a problem or a job and afterwards?

Before: Depressed-> Due to heavy loss in crop production and revenue.

After: Satisfied-> Due to loss reduction in crop production and increment in revenue.

10. YOUR SOLUTION SL

Our projects will enlighten the lives of many farmers by providing suitable information regarding the crop yield in a particular season, which in turn gives many profits to them even from the government side by giving agricultural loans etc,..

8. CHANNELS of BEHAVIOR CH

8.1 ONLINE

What kind of actions do customers take online?

In online mode they are given access to use all the options like crop yield, production, pesticides, agricultural loan, database, etc,..

8.2 OFFLINE

What kind of actions do customers take offline?

In Offline mode they won't be able to access some of the resources, but they can see the databases of the crop production.

3. REQUIREMENT ANALYSIS:

3.1 FUNCTIONAL REQUIREMENT:

S No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN Registration through Whatsapp
2	User Confirmation	Confirmation via Email Confirmation via OTP Confirmation via physical Letter.
3	User Profile	User Details Form Details
4	Required Data	The previous year crop yield data set Farm yield methodology User data of the farmer Details of the Seasons and the Regions
5	Analysis	Clean, Analyze the data by means of set of past data of the multiple users which is farmers. Visualizing the datasets using IBM Cognos
6	Estimation	Creating the perfect data module through attractive stories, dashboard and reports to increase the understandability of data.

3.2 NON – FUNCTIONAL REQUIREMENT:

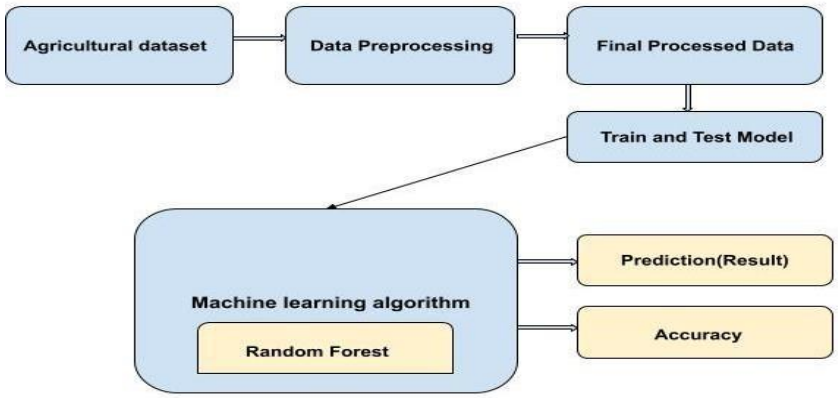
S No.	Non-Functional Requirement	Description
1	**Usability**	From the given datasets , analysis is done and a report is created. Accordingly, sowing of crops is recommended.
2	**Security**	Usage of IBM COGNOS, will provide secure user information(Data Visualization)
3	**Reliability**	Using the interactive data visual dashboards , we can easily understand the data reports.
4	**Performance**	Interaction makes better performance between all users and impresses by the data visuals advice.
5	**Availability**	The dashboard is easily available and accessible in smart phones and PC's.
6	**Scalability**	Prediction of crops for the forthcoming year can be done .It gives you a variety of crops to choose from our region. Also to know the better profitability of crops.

4. PROJECT DESIGN:

4.1 DATA FLOW DIAGRAM:

Project flow describes a preset sequence of activities required to plan, produce, deliver and maintain project product, along with information, materials, and resources required by the project. Project flow is a convenient way to define and plan projects.

Project flow for estimating the crop yield using data analytics is shown below.



4.2 SOLUTION AND TECHNOLOGY ARCHITECTURE:

The Deliverables hall include the architectural diagram as below and the information as per the table 1 & table 2.

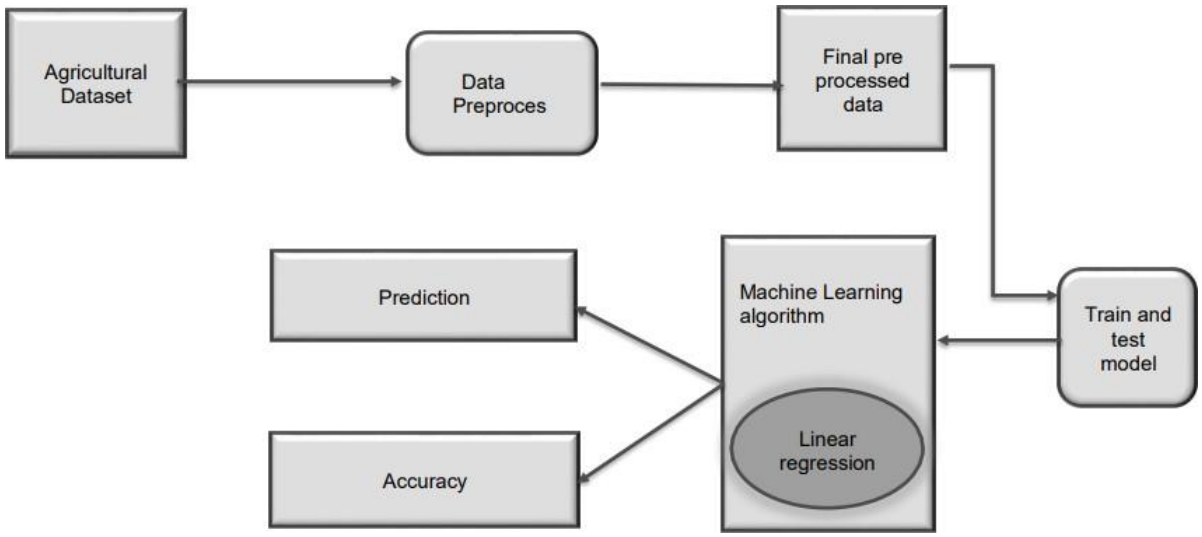


Table-1: Components & Technologies:

S. No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App ,Chat bot etc.	HTML, CSS, JavaScript.
2.	Applicationlogic1	Login as a user in the application	Java/Python
3.	Applicationlogic2	Login as admin in the application	IBM Watson STT service
4.	Applicationlogic3	Login as merchants in the application	IBM Watson Assistant
5.	Database	Data related to crop production in previous and also crop data.	MySQL , NoSQL , etc.
6.	Cloud Database	Database Service on Cloud	IBMDB2,IBM Cloudant etc.
7.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local File system
8.	ExternalAPI-1	Weather API are application programming interface that allow you to connect to large databases.	IBM Weather API ,etc.

9.	ExternalAPI-2	Soil testing is a quick and accurate method to determine the relative acidity of the soil and the level of several essential nutrient needed.	Soil API, etc.
10.	Machine Learning Model	It is mostly used for finding out the relationship between variables and forecasting	Linear Regression
11.	Infrastructure(Server/Cloud)	Application Deployment on Local System/Cloud Local Server Configuration CloudServerConfiguration:l1	Local , Cloud Foundry ,Kubernetes ,etc.

Table-2:ApplicationCharacteristics:

S. No	**Characteristics**	**Description**	**Technology**
1.	Open-Source Frameworks	Bootstrap is a free ,open source front-end development frame work	Bootstrap ,React etc.,
2.	Security Implementations	Improves user experience and provides greater security.	Authentication etc.
3.	Scalable Architecture	A3-tier architecture where in application gets data from various sources ,manipulates it, stores the min IBM Cloud and Cognos.	IBM Cloud, IBM Cognos.
4.	Availability	The application is being developed is made available to all users	Cognos Analytics

5.	Performance	Multiple technologies and services that will improve the usability in agriculture activities.	Robots, IOT agriculture sensors.

4.3 USER STORIES:

User Story Number	User Story / Task
USN-1	Understanding the data set .
USN-2	Loading the data set.
USN-3	Convert the data into required format
USN-4	Explore the data's which is uploaded in the IBM cognos
USN-5	Creating the data visualization chart
USN-6	Creating a dashboard
USN-7	Estimation of accuracy using random forest algorithm
USN-8	Export the analytics

5. PROJECT PLANNING AND SCHEDULING:

5.1 SPRINT PLANNING AND ESTIMATION:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Working with the data set	USN-1	Understanding the data set .	10	Medium	Anuraagavi
Sprint-1	Working with the data set	USN-2	Loading the data set.	10	High	Aditya Venkatesh
Sprint-2	Prepare the data	USN-3	Convert the data into required format	10	Medium	Harinivas
Sprint-2	Data exploration	USN-4	Explore the data's which is uploaded in the IBM cognos	10	Medium	Haripriya
Sprint-3	Data visualization	USN-5	Creating the data visualization chart	10	High	Anuraagavi
Sprint-3	Dashboard	USN-6	Creating a dashboard	10	High	Aditya Venkatesh
Sprint-3	Visualization	USN-7	Estimation of accuracy using random forest algorithm	10	High	Harinivas
Sprint-4	Export	USN-8	Export the analytics	10	High	Haripriya

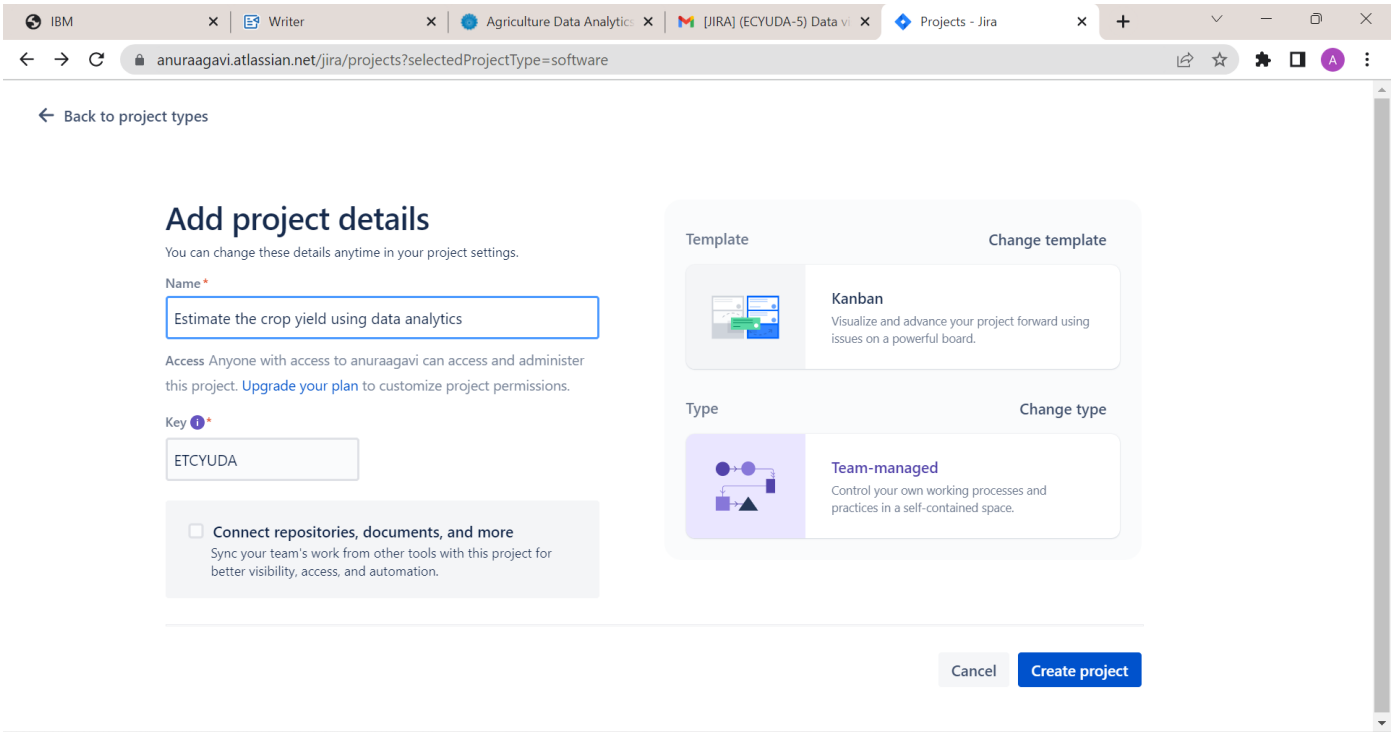
5.2 SPRINT DELIVERY SCHEDULE:

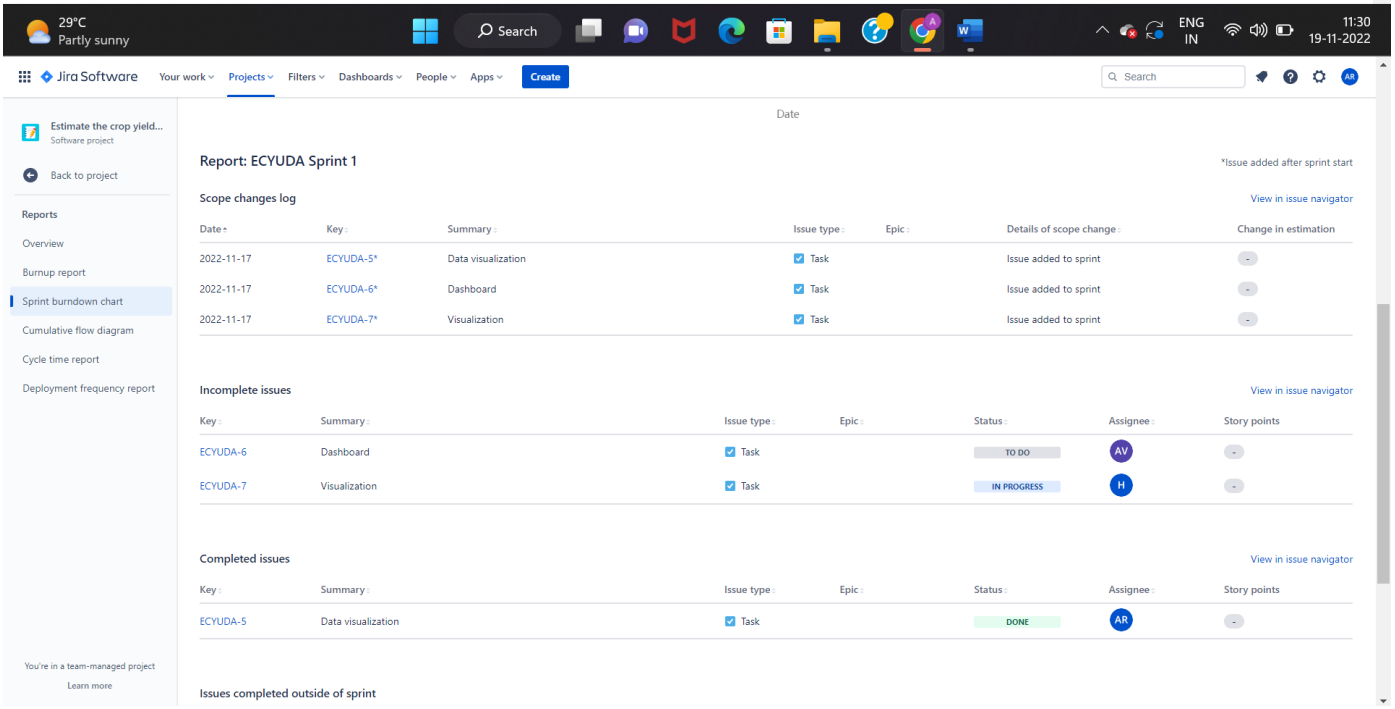
Sprint	Total story points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	4 Days	01 Nov 2022	04 Nov 2022	20	08 - 09 Nov 2022
Sprint-2	20	5 Days	05 Nov 2022	10 Nov 2022	20	10 - 13 Nov 2022
Sprint-3	20	4 Days	11 Nov 2022	14 Nov 2022	20	14 - 16 Nov 2022

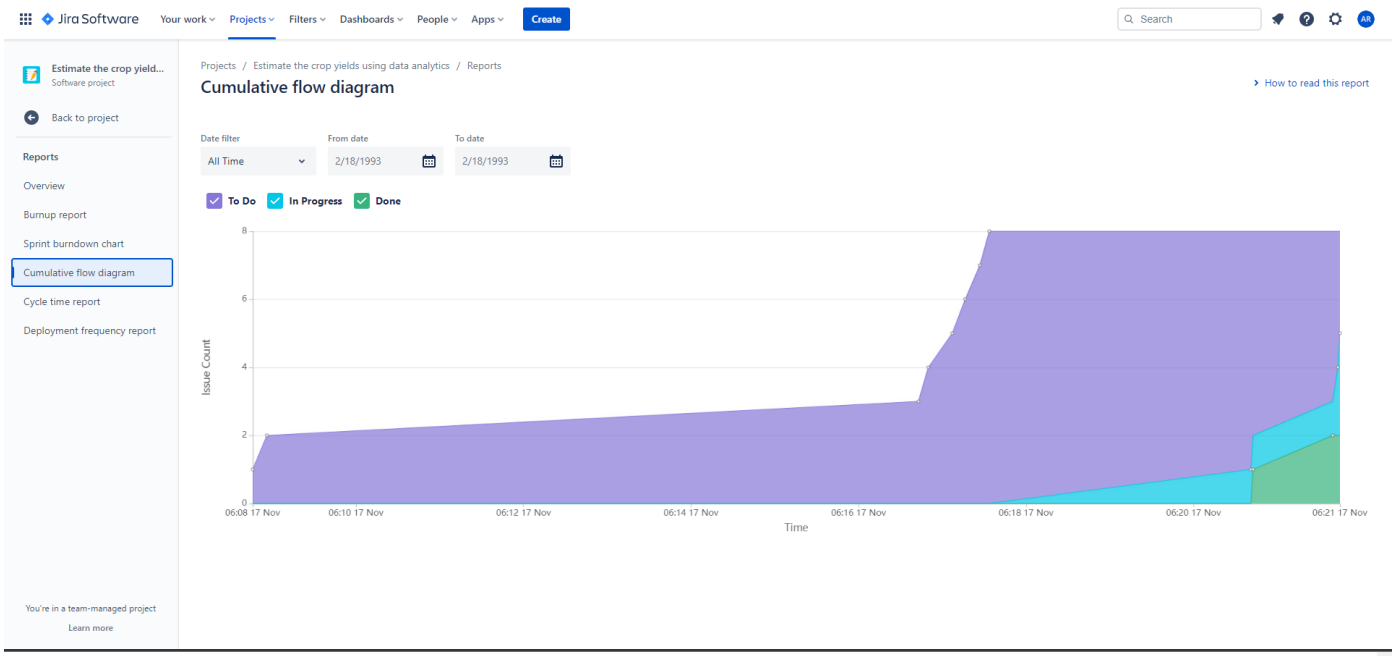
Sprint-4	20	4 Days	15 Nov 2022	19 Nov 2022	20	17 - 19 Nov 2022

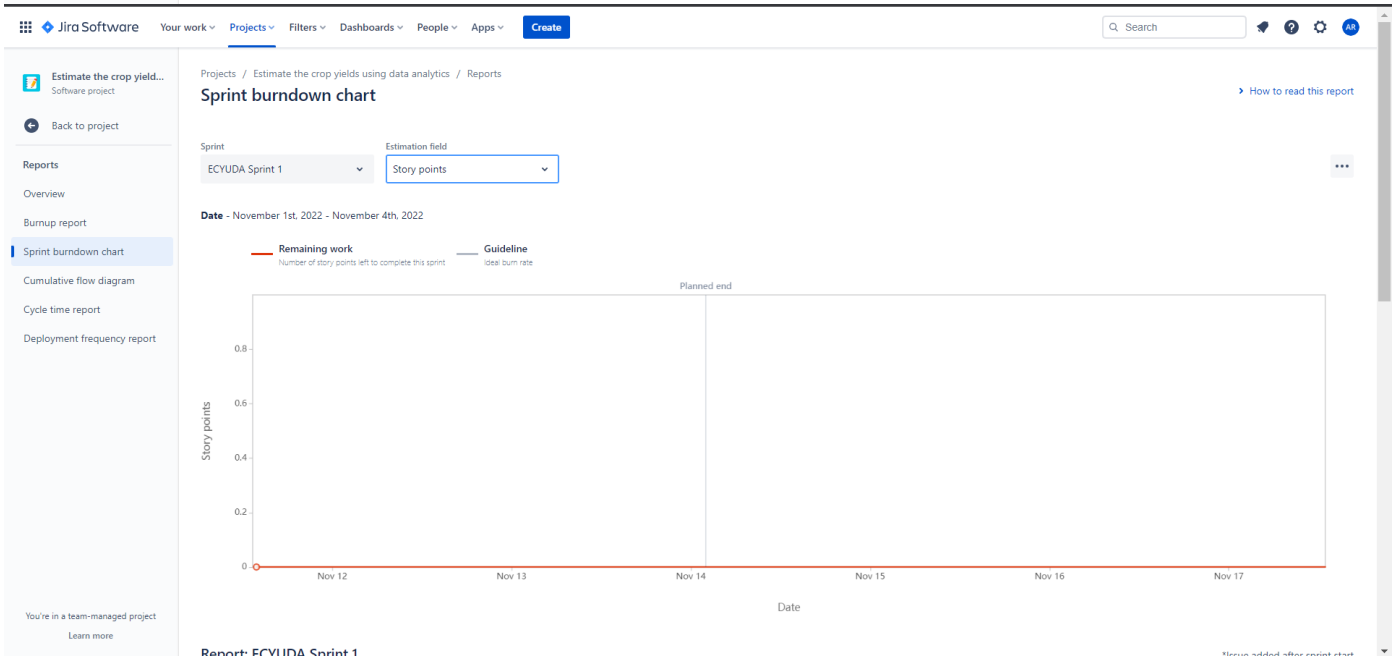
5.3 REPORT FROM JIRA:

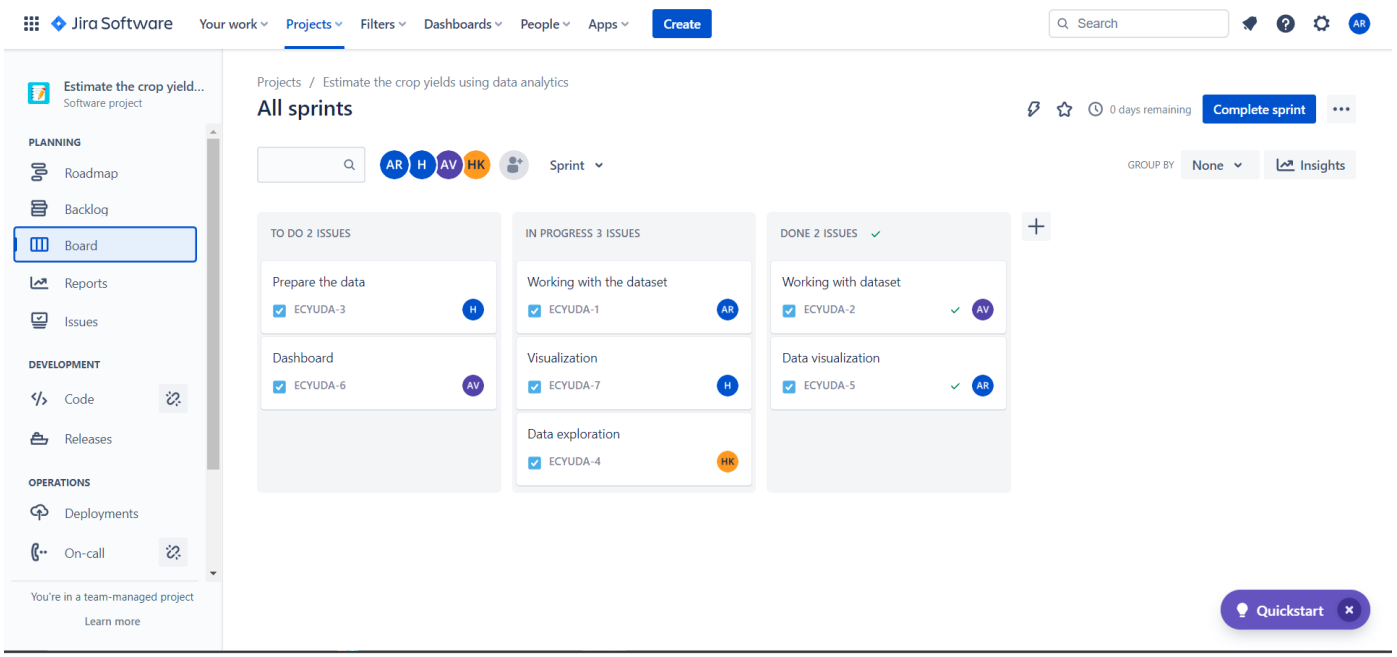
A) SPRINT :

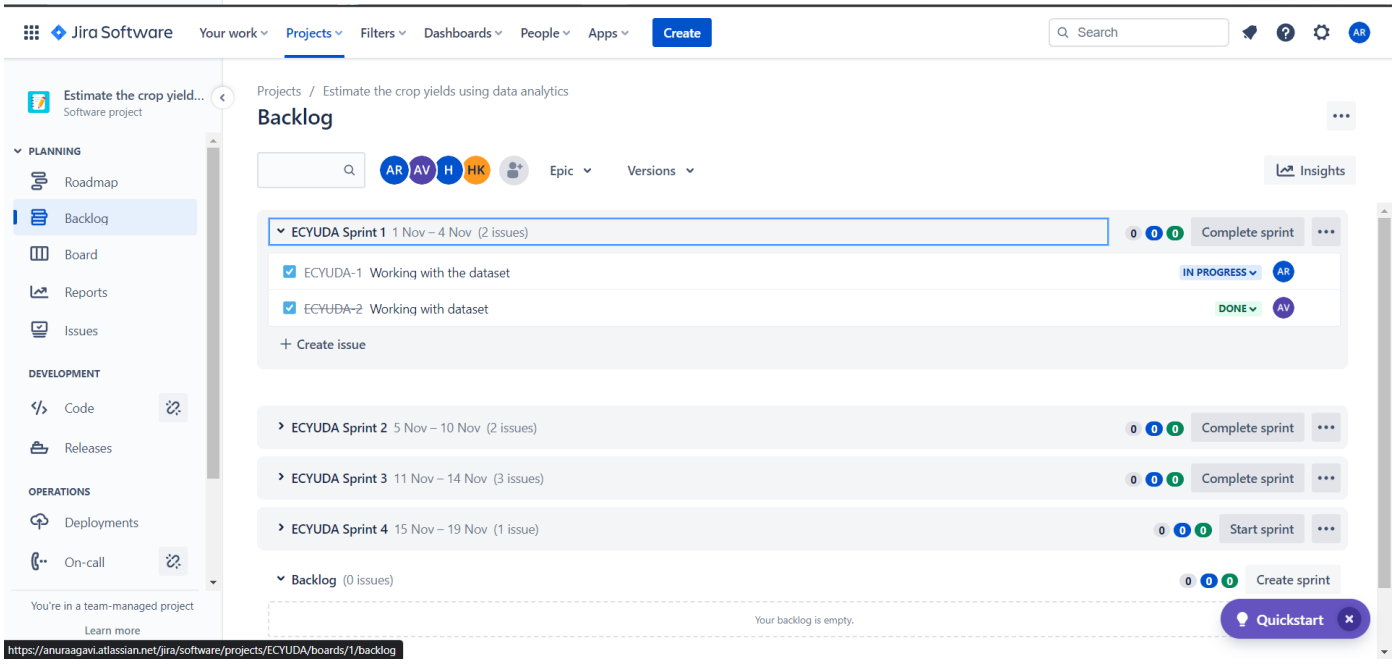


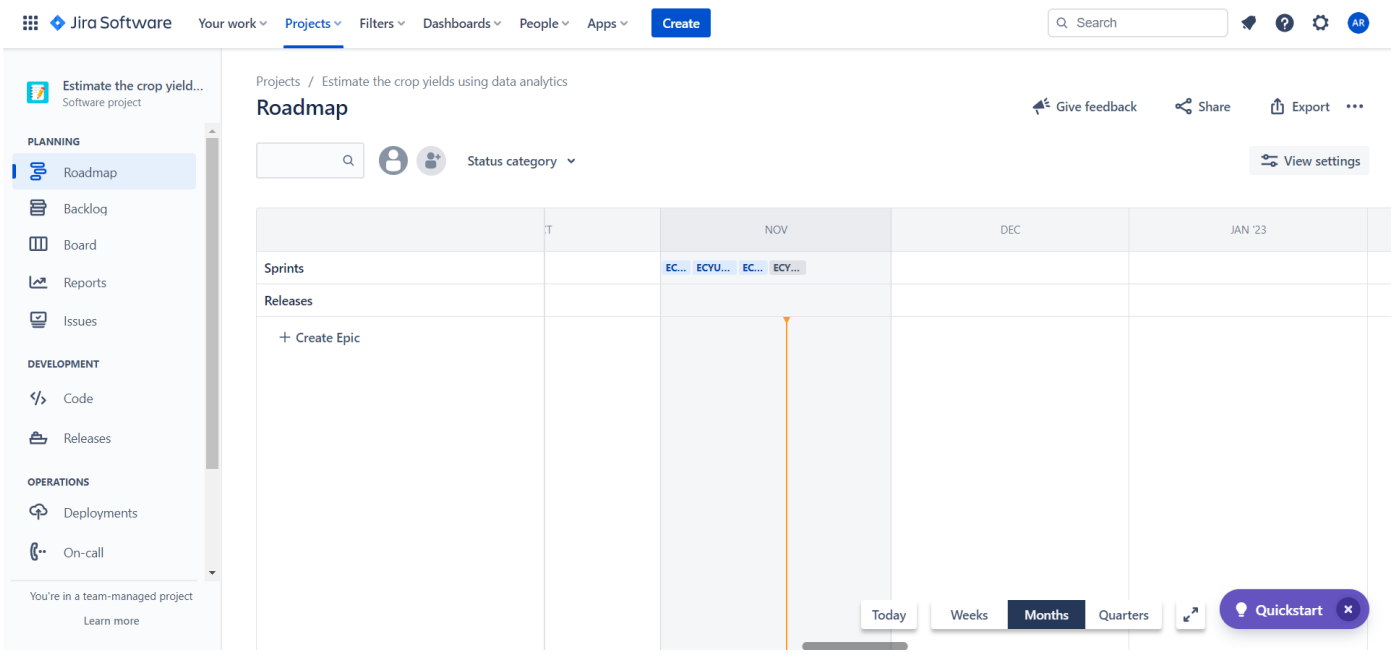



Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate the crop yield...
Software project
Back to project
Reports
Overview
Burnup report
Sprint burndown chart
Cumulative flow diagram
Cycle time report
Deployment frequency report
You're in a team-managed project
Learn more
Projects / Estimate the crop yields using data analytics / Reports
Cumulative flow diagram
How to read this report
Date filter
All Time
From date
2/18/1993
To date
2/18/1993
To Do
In Progress
Done
Issue Count
0
2
4
6
8
06:08 17 Nov
06:10 17 Nov
06:12 17 Nov
06:14 17 Nov
06:16 17 Nov
06:18 17 Nov
06:20 17 Nov
06:21 17 Nov
Time

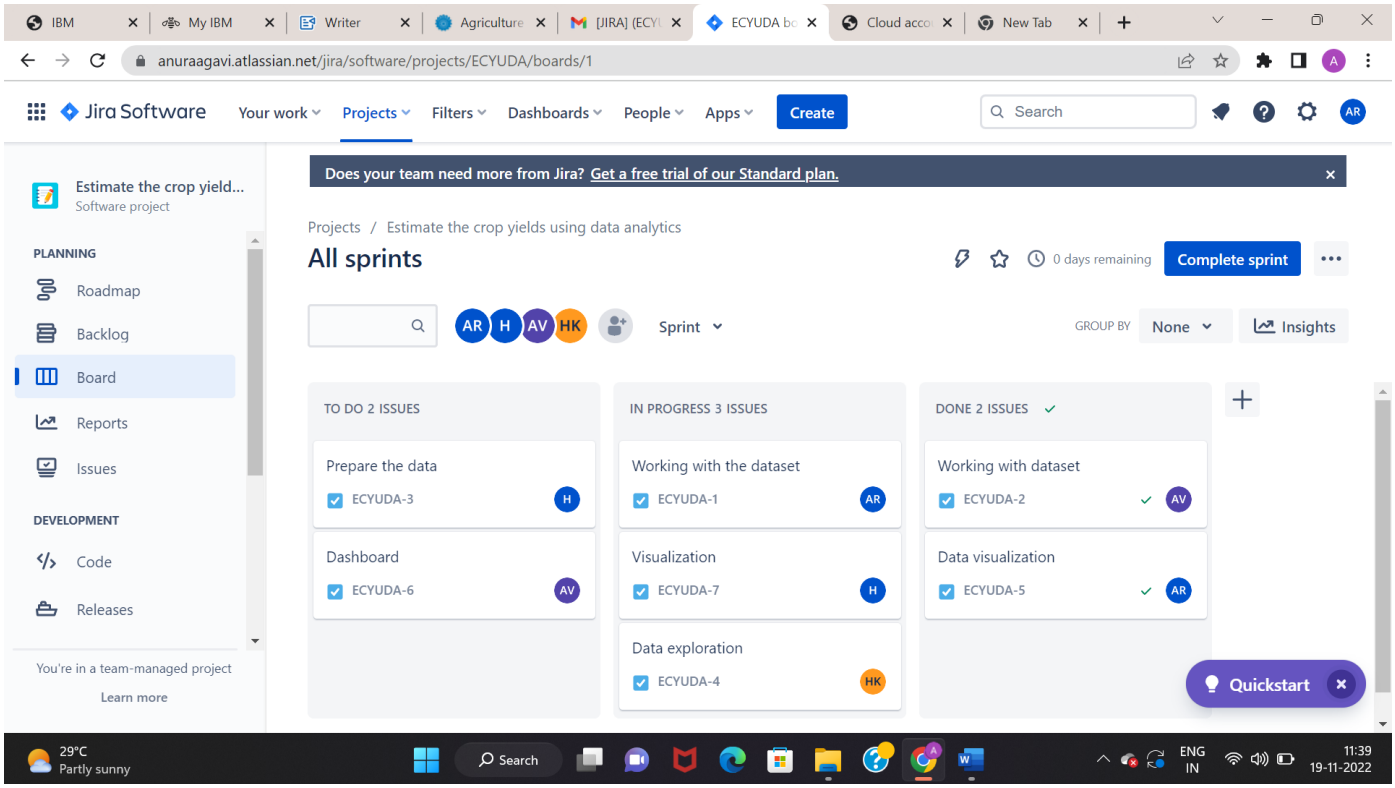

Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate the crop yield...
Software project
Back to project
Reports
Overview
Burnup report
Sprint burndown chart
Cumulative flow diagram
Cycle time report
Deployment frequency report
You're in a team-managed project
Learn more
Projects / Estimate the crop yields using data analytics / Reports
Sprint burndown chart
How to read this report
Sprint
ECYUDA Sprint 1
Estimation field
Story points
Date - November 1st, 2022 - November 4th, 2022
Remaining work
Number of story points left to complete this sprint
Guideline
Ideal burn rate
Planned end
Story points
0
0.2
0.4
0.6
0.8
Nov 12
Nov 13
Nov 14
Nov 15
Nov 16
Nov 17
Date
Report: ECYUDA Sprint 1


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate the crop yield...
Software project
PLANNING
Roadmap
Backlog
Board
Reports
Issues
DEVELOPMENT
Code
Releases
OPERATIONS
Deployments
On-call
You're in a team-managed project
Learn more
Projects / Estimate the crop yields using data analytics
All sprints
0 days remaining
Complete sprint
Sprint
GROUP BY
None
Insights
TO DO 2 ISSUES
Prepare the data
ECYUDA-3
Dashboard
ECYUDA-6
IN PROGRESS 3 ISSUES
Working with the dataset
ECYUDA-1
Visualization
ECYUDA-7
Data exploration
ECYUDA-4
DONE 2 ISSUES
Working with dataset
ECYUDA-2
Data visualization
ECYUDA-5
Quickstart


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate the crop yield...
Software project
PLANNING
Roadmap
Backlog
Board
Reports
Issues
DEVELOPMENT
Code
Releases
OPERATIONS
Deployments
On-call
You're in a team-managed project
Learn more
Projects / Estimate the crop yields using data analytics
Backlog
Epic
Versions
Insights
ECYUDA Sprint 1 1 Nov – 4 Nov (2 issues)
Complete sprint
ECYUDA-1 Working with the dataset
IN PROGRESS
ECYUDA-2 Working with dataset
DONE
Create issue
ECYUDA Sprint 2 5 Nov – 10 Nov (2 issues)
ECYUDA Sprint 3 11 Nov – 14 Nov (3 issues)
ECYUDA Sprint 4 15 Nov – 19 Nov (1 issue)
Start sprint
Backlog (0 issues)
Create sprint
Your backlog is empty.
Quickstart
https://anuraagavi.atlassian.net/jira/software/projects/ECYUDA/boards/1/backlog


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate the crop yield...
Software project
PLANNING
Roadmap
Backlog
Board
Reports
Issues
DEVELOPMENT
Code
Releases
OPERATIONS
Deployments
On-call
You're in a team-managed project
Learn more
Projects / Estimate the crop yields using data analytics
Roadmap
Give feedback
Share
Export
Status category
View settings
NOV
DEC
JAN '23
Sprints
Releases
Create Epic
Today
Weeks
Months
Quarters
Quickstart

SPRINT :

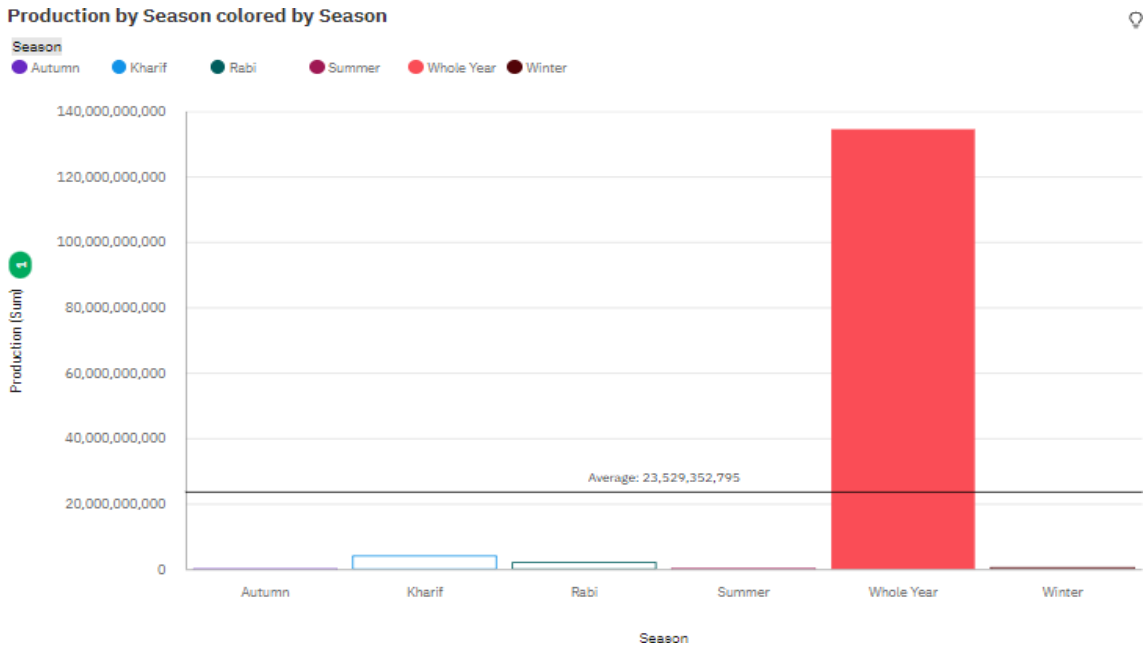

anuraagavi.atlassian.net/jira/software/projects/ECYUDA/boards/1
Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate the crop yield...
Software project
PLANNING
Roadmap
Backlog
Board
Reports
Issues
DEVELOPMENT
Code
Releases
You're in a team-managed project
Learn more
Does your team need more from Jira? Get a free trial of our Standard plan.
Projects / Estimate the crop yields using data analytics
All sprints
0 days remaining
Complete sprint
Sprint
GROUP BY
None
Insights
TO DO 2 ISSUES
Prepare the data
ECYUDA-3
Dashboard
ECYUDA-6
IN PROGRESS 3 ISSUES
Working with the dataset
ECYUDA-1
Visualization
ECYUDA-7
Data exploration
ECYUDA-4
DONE 2 ISSUES
Working with dataset
ECYUDA-2
Data visualization
ECYUDA-5
Quickstart
29°C
Partly sunny
Search
ENG
IN
11:39
19-11-2022

6. CODING AND SOLUTIONING:

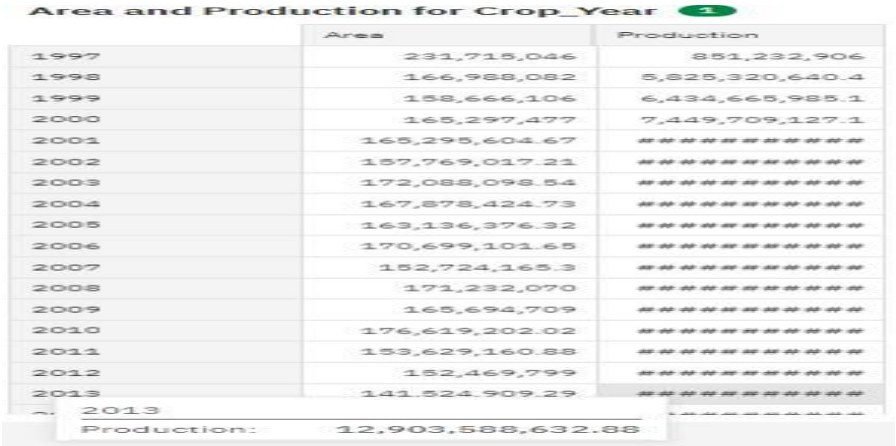
6.1 FEATURE 1:

A) DATA EXPLORATION:

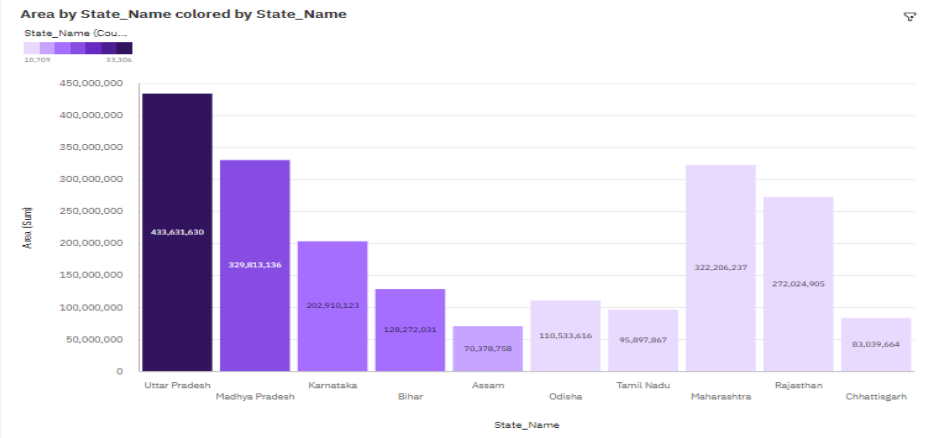
1) Seasons with average production:



2) With years usage of area and production:



3) Top 10 states with the most area.



4) States with crop production with seasons:

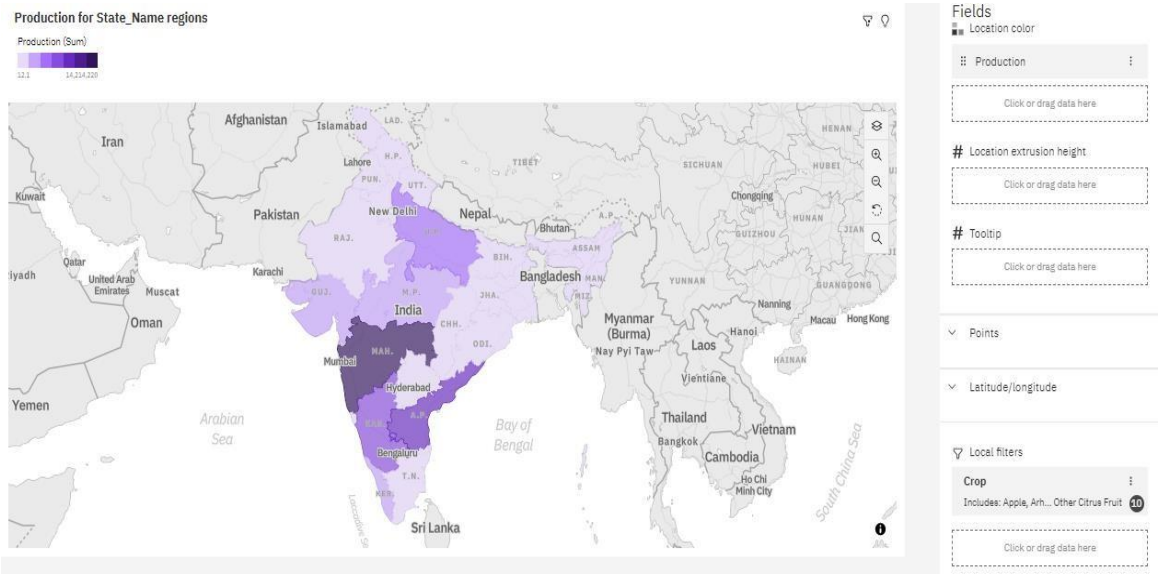
Production for Season, State_Name and Crop

Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Andaman and Ni...	Arecanut	(no value)	14,500	5,800	(no value)	27,735.81	(no value)	**48,035.81**
	Arhar/Tur	(no value)	(no value)	104	(no value)	(no value)	(no value)	**104**
	Banana	(no value)	(no value)	(no value)	(no value)	97,424.65	(no value)	**97,424.65**
	Black pepper	(no value)	(no value)	120	(no value)	604.5	(no value)	**724.5**
	Cashewnut	(no value)	(no value)	310	(no value)	1,374.79	(no value)	**1,684.79**
	Coconut	(no value)	(no value)	(no value)	(no value)	717,790,000	(no value)	**717,790,000**
	Dry chillies	(no value)	(no value)	575	(no value)	3,443.3	(no value)	**4,018.3**
	Dry ginger	(no value)	(no value)	1,850	(no value)	10,825.6	(no value)	**12,675.6**
	Groundnut	(no value)	(no value)	14.4	(no value)	(no value)	(no value)	**14.4**
	Maize	(no value)	(no value)	367.62	(no value)	(no value)	(no value)	**367.62**
	Moong(Green Gr...	(no value)	(no value)	575.5	(no value)	(no value)	(no value)	**575.5**
	Other Kharif puls...	(no value)	649	(no value)	(no value)	(no value)	(no value)	**649**
	Rice	23,916	199,090.06	(no value)	(no value)	(no value)	(no value)	**223,006.06**
	Sugarcane	1,332.95	(no value)	(no value)	(no value)	29,305.72	(no value)	**30,638.67**
	Sunflower	(no value)	(no value)	2.4	(no value)	(no value)	(no value)	**2.4**
	Sweet potato	(no value)	(no value)	923	(no value)	2,142.35	(no value)	**3,065.35**

Production for Season, State_Name and Crop 1

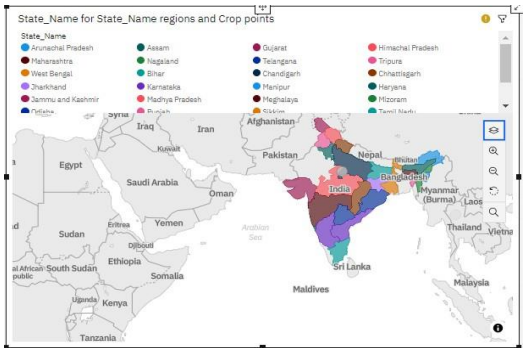
Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Haryana	Peas & beans (P...	(no value)	(no value)	19,624	(no value)	(no value)	(no value)	**19,624**
	Potato	(no value)	(no value)	(no value)	(no value)	3,621,500	(no value)	**3,621,500**
	Rapeseed &Must...	(no value)	(no value)	10,803,800	(no value)	(no value)	(no value)	**10,803,800**
	Rice	(no value)	49,318,300	(no value)	(no value)	(no value)	(no value)	**49,318,300**
	Sannhamp	(no value)	29	(no value)	(no value)	1,800	(no value)	**1,829**
	Sesamum	(no value)	18,379	(no value)	(no value)	(no value)	(no value)	**18,379**
	Soyabean	(no value)	(no value)	(no value)	(no value)	(no value)	(no value)	**(no value)**
	Sugarcane	(no value)	(no value)	(no value)	(no value)	112,680,900	(no value)	**112,680,900**
	Sunflower	(no value)	18,900	146,500	(no value)	(no value)	(no value)	**165,400**
	Sweet potato	(no value)	(no value)	(no value)	(no value)	16,900	(no value)	**16,900**
	Tobacco	(no value)	(no value)	(no value)	(no value)	(no value)	(no value)	**(no value)**
	Turmeric	(no value)	(no value)	(no value)	(no value)	965	(no value)	**965**
	Urad	(no value)	11,318	(no value)	(no value)	(no value)	(no value)	**11,318**
	Wheat	(no value)	(no value)	158,647,000	(no value)	(no value)	(no value)	**158,647,000**
	other oilseeds	(no value)	(no value)	(no value)	(no value)	(no value)	(no value)	**(no value)**
	Summary	**(no value)**	**88,593,481**	**173,272,098**	**(no value)**	**119,408,311**	**(no value)**	**381,273,890**
	Arhar/Tur	(no value)	591	(no value)	(no value)	(no value)	(no value)	**591**

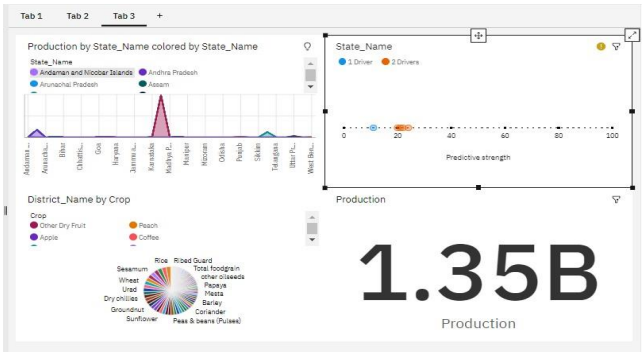
5) States with crop production:

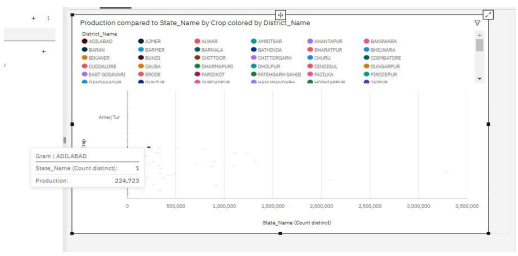


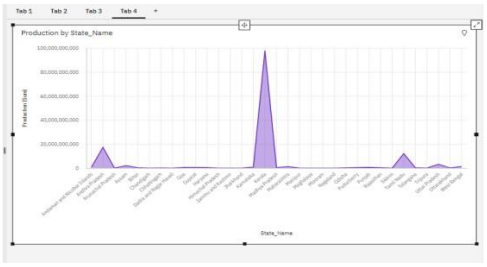
6.2 FEATURE 2:

B) CREATING THE DASHBOARD:









6.3 CODING:

- https://colab.research.google.com/drive/1iXAe3LhSw18iDA8CwMgONYc7Byqfr627?usp=sharing

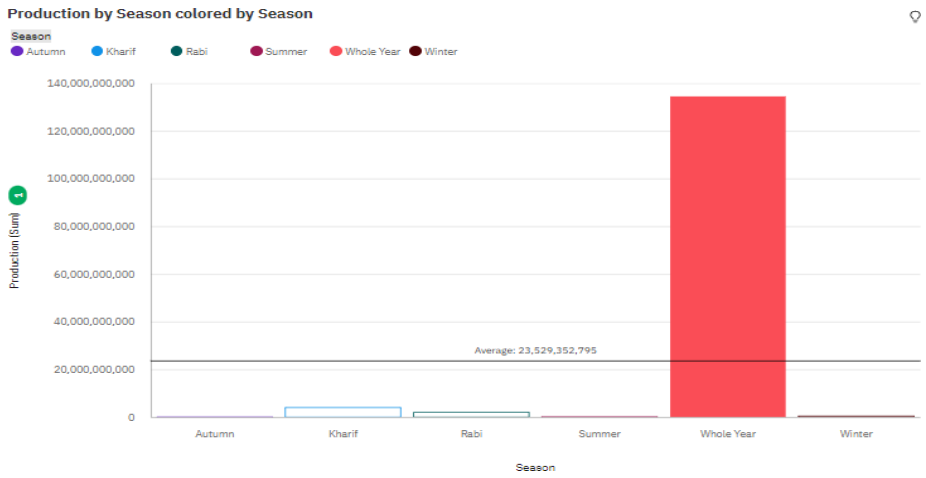
7. TESTING:

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visulizations / Graphs - 6
2.	Data Responsiveness	With minimal delay our model is responding to the large sets of inputs from the user.
3.	Amount Data to Rendered (DB2 Metrics)	We rendered 70% of data from the dataset.
4.	Utilization of Data Filters	We have chosen the particular state (Maharashtra) and year (>=2004) for easy analysis.
5.	Effective User Story	No of Scene Added - 9
6.	Descriptive Reports	No of Visulizations / Graphs - 6

8. RESULTS:

A) DATA EXPLORATION:

1) Seasons with average production:



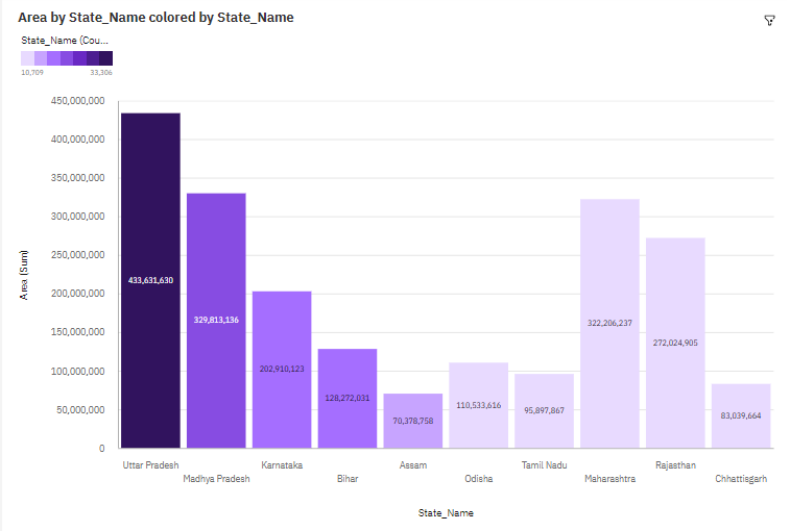
2) With years usage of area and production:

Area and Production for Crop_Year 1

	Area	Production
1997	231,715,046	851,232,906
1998	166,988,082	5,825,320,640.4
1999	158,666,106	6,434,665,985.1
2000	165,297,477	7,449,709,127.1
2001	165,295,604.67	##########
2002	157,769,017.21	##########
2003	172,088,098.54	##########
2004	167,878,424.73	##########
2005	163,136,376.32	##########
2006	170,699,101.65	##########
2007	152,724,165.3	##########
2008	171,232,070	##########
2009	165,694,709	##########
2010	176,619,202.02	##########
2011	153,629,160.88	##########
2012	152,469,799	##########
2013	141,524,909.29	##########

2013
Production: 12,903,588,632.88

3) Top 10 states with the most area.

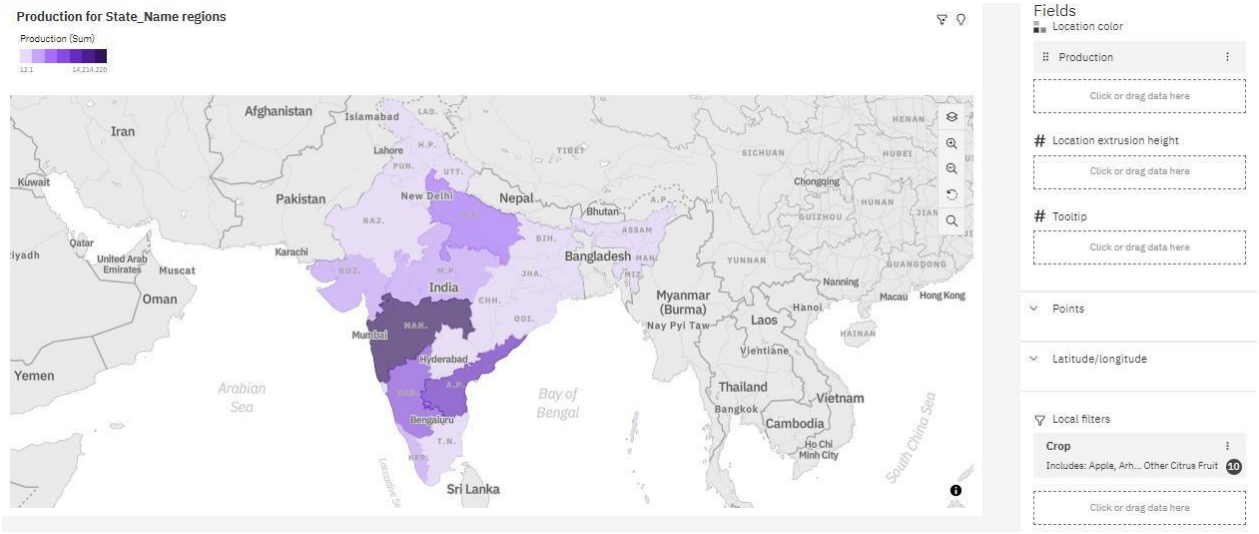


4) States with crop production with seasons:

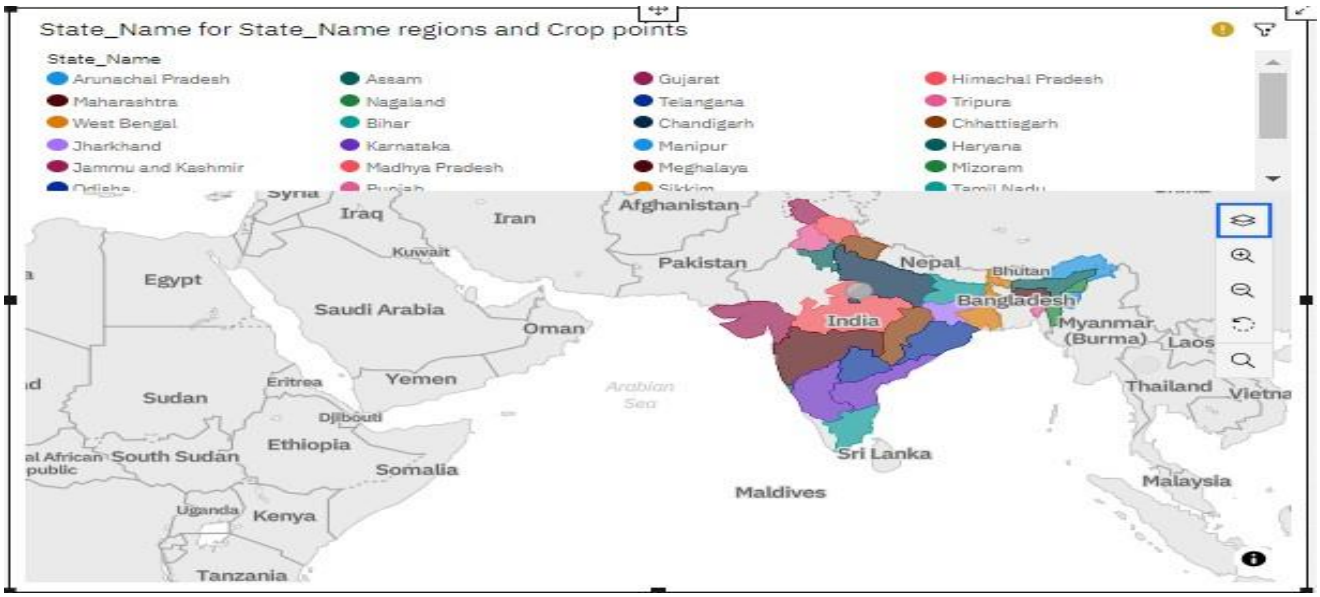
Production for Season, State_Name and Crop

Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Andaman and Ni...	Arecanut	(no value)	14,500	5,800	(no value)	27,735.81	(no value)	**48,035.81**
	Arhar/Tur	(no value)	(no value)	104	(no value)	(no value)	(no value)	**104**
	Banana	(no value)	(no value)	(no value)	(no value)	97,424.65	(no value)	**97,424.65**
	Black pepper	(no value)	(no value)	120	(no value)	604.5	(no value)	**724.5**
	Cashewnut	(no value)	(no value)	310	(no value)	1,374.79	(no value)	**1,684.79**
	Coconut	(no value)	(no value)	(no value)	(no value)	717,790,000	(no value)	**717,790,000**
	Dry chillies	(no value)	(no value)	575	(no value)	3,443.3	(no value)	**4,018.3**
	Dry ginger	(no value)	(no value)	1,850	(no value)	10,825.6	(no value)	**12,675.6**
	Groundnut	(no value)	(no value)	14.4	(no value)	(no value)	(no value)	**14.4**
	Maize	(no value)	(no value)	367.62	(no value)	(no value)	(no value)	**367.62**
	Moong(Green Gr...	(no value)	(no value)	575.5	(no value)	(no value)	(no value)	**575.5**
	Other Kharif puls...	(no value)	649	(no value)	(no value)	(no value)	(no value)	**649**
	Rice	23,916	199,090.06	(no value)	(no value)	(no value)	(no value)	**223,006.06**
	Sugarcane	1,332.95	(no value)	(no value)	(no value)	29,305.72	(no value)	**30,638.67**
	Sunflower	(no value)	(no value)	2.4	(no value)	(no value)	(no value)	**2.4**
	Sweet potato	(no value)	(no value)	923	(no value)	2,142.35	(no value)	**3,065.35**

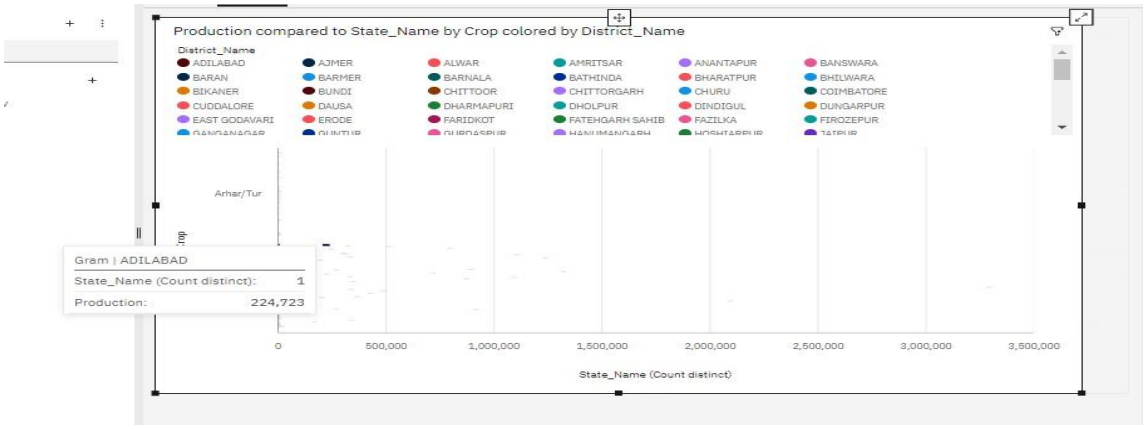
5) States with crop production:

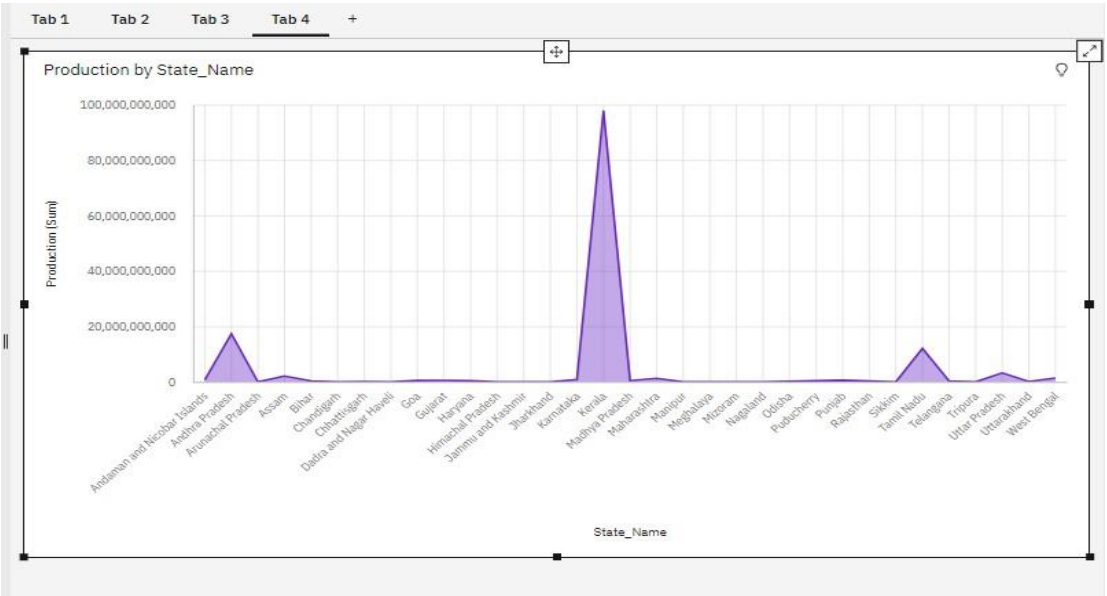


B)DASHBOARD:




Tab 1
Tab 2
Tab 3
Production by State_Name colored by State_Name
State_Name
Andaman and Nicobar Islands
Andhra Pradesh
Arunachal Pradesh
Assam
State_Name
1 Driver
2 Drivers
0
20
40
60
80
100
Predictive strength
District_Name by Crop
Crop
Other Dry Fruit
Peach
Apple
Coffee
Rice
Ribed Guard
Total foodgrain
other oilseeds
Papaya
Mesta
Barley
Coriander
Peas & beans (Pulses)
Sunflower
Groundnut
Dry chillies
Urad
Wheat
Sesamum
Production
1.35B
Production


Production compared to State_Name by Crop colored by District_Name
District_Name
ADILABAD
AJMER
ALWAR
AMRITSAR
ANANTAPUR
BANSWARA
BARAN
BARMER
BARNALA
BATHINDA
BHARATPUR
BHILWARA
BIKANER
BUNDI
CHITTOOR
CHITTORGARH
CHURU
COIMBATORE
CUDDALORE
DAUSA
DHARMAPURI
DHOLPUR
DINDIGUL
DUNGARPUR
EAST GODAVARI
ERODE
FARIDKOT
FATEHGARH SAHIB
FAZILKA
FIROZEPUR
Arhar/Tur
Gram | ADILABAD
State_Name (Count distinct): 1
Production: 224,723
0
500,000
1,000,000
1,500,000
2,000,000
2,500,000
3,000,000
3,500,000
State_Name (Count distinct)


Tab 1
Tab 2
Tab 3
Tab 4
Production by State_Name
100,000,000,000
80,000,000,000
60,000,000,000
40,000,000,000
20,000,000,000
0
Production (Sum)
Andaman and Nicobar Islands
Andhra Pradesh
Arunachal Pradesh
Assam
Bihar
Chandigarh
Chhattisgarh
Dadra and Nagar Haveli
Goa
Gujarat
Haryana
Himachal Pradesh
Jammu and Kashmir
Jharkhand
Karnataka
Kerala
Madhya Pradesh
Maharashtra
Manipur
Meghalaya
Mizoram
Nagaland
Odisha
Puducherry
Punjab
Rajasthan
Sikkim
Tamil Nadu
Telangana
Tripura
Uttar Pradesh
Uttarakhand
West Bengal
State_Name

9. ADVANTAGES:

- Predicting productivity of crop in various climatic conditions can **help farmer and other partners in essential basic leadership as far as agronomy and product decision**.
- This model can be used to select the most excellent crops for the region and also its yield thereby improving the values and gain of farming also.
- This will help the policy makers of the state to determine the budget.
- If the production of a crop observes a declining trend then, they can plan to implement the schemes at an early stage. This in return will save the state from shortage of the product.
- Monitors the growth of healthy crops.
- Helps the government to frame the government policies.
- Yield data helps the farmer to determine how much they should plant next year.
- Helps the farmer in Seed Selection, Pest Management, Irrigation Scheduling, etc,…

CHALLENGES:

Challenges are the major basis which imminent the negative impacts on current project. Some of the challenges faced during crop yield prediction are:

- Choosing appropriate dataset, after choosing dataset tuning of the parameters which makes project more efficient to get the desired results.
- Model must be trained by taking consideration of less computational efficiency and power.
- Increase of error rate due to dynamically changing the environment.

10. CONCLUSION:

Our project will make policy maker of the state to determine the budget. If the production of a crop observes a declining trend then, they can plan to implement the schemes at an early stage. This in return will save the state from shortage of the product. Monitors the growth of healthy crops. Helps the government to frame the government policies. 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. The analysis of agricultural productivity and the uncovering of hidden patterns utilizing data sets related to seasons and crop yields have been noted in the literature. Using IBM Cognos, we have observed and conducted analysis regarding various crops grown, areas, and productions in various states and districts. "**The scope of the project is to determine the crop yield of an area by considering dataset with some features which are important or related to crop production such as temperature, moisture, rainfall, and production of the crop in previous years. To predict a continuous value, regression models are used."**

11. FUTURE SCOPE:

Our future scope is to add many more geographical features and predict using those features.

12. APPENDIX:

- **GITHUB LINK - https://github.com/IBM-EPBL/IBM-Project-20534-1659749523**