MAHENDRA INSTITUTE OF TECHNOLOGY(AUTONOMOUS)

MAHENDRAPURI, MALLASAMUDRAM(W), THIRUCHENGODE,

NAMAKKAL, TAMIL NADU 637503

ESTIMATE THE CROP YIELD USING DATA ANALYTICS

TEAM ID – PNT2022TMID17541

TEAM LEADER - SINDHU S (319UEC087)

TEAM MEMBER - SARANYA A (319UEC080)

SUBASRI S (319UEC095)

MATHESWARI G (319UEC052)

ABSTRACT

Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Various factors are to be considered which have direct impact on the production, productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. The main challenge in using big data in agriculture is identification of effectiveness of big data analytics. Efforts are going on to understand how big 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.

1.INTRODUCTION

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.

2.LITERATURE SURVEY:

2.1 EXISTING PROBLEM:

a) B M Sagar Agriculture Data Analytics in Crop Yield Estimation May 2, 2018

Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Various factors are to be considered which have direct impact on the production, productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. The main challenge in using big data in agriculture is identification of effectiveness of big data analytics. Efforts are going on to understand how big 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.

B. j Majumdar estimated the crop yield using data analytics may 2008

Agricultural data is being produced constantly and enormously. As a result, agricultural data has come in the era of bigdata. Smart technologies contribute in data collection using electronic devices. In our project we are going to Analyse and mine this agricultural data to get useful results using technologies like data analytics and machine learning and this result will be given to farmers for better crop yield in terms of efficiency and productivity.

2.2 REFERENCE

•https://www.researchgate.net/publication/329467349_Agriculture_Data_Analytics_in_Crop_Yield_Estimation_A_Critical_Review

• **https://www.sciencedirect.com/science/article/pii/S2666285X21000364**

2.3 PROBLEM STATEMENT DEFINTI

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.

3.IDEATION &PROPOED SYSTEM:

3.1 EMPATHY MAP CANVAS:

3.2 IDEATION & BRAINSTORMING:

A) BRAINSTORMING:

3.3 PROPOSED SOLUTION:

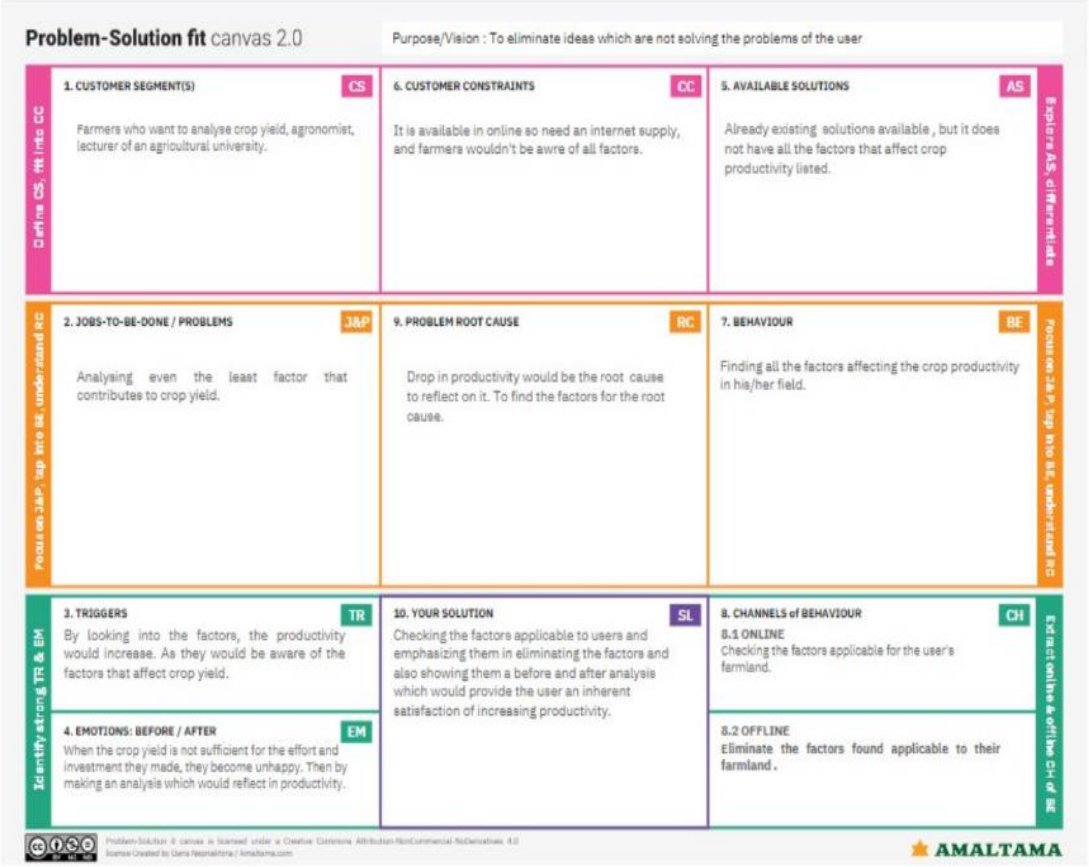
SNO	PARAMETER	DESCRIPTION
1.	Problem statement (problem to be solved)	● Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. ● Big data provides farmers granular data on rainfall patterns, water cycles, fertilizer requirements, and more. This enables them to make smart decisions, such as what crops to plant for better profitability and when to harvest. The right decisions ultimately improve farm yields.
2.	Idea/solution description	● Data analytics can help farmers monitor the health of crops in real-time, create predictive analytics related to future yields and help farmers make resource management decisions based on proven trends. Reducing waste and improving profits

3.	Novelty/Uniqueness	● To increase quality and yields, it is crucial to understand the current nutrient levels of the soil to be able to ascertain which areas require improvement.
4.	Social Impact / Customer Satisfaction	● Farmers are always looking for innovations and information to help them boost production and maximize returns on them products ● The rise of digital technology has offered access to a wide range of opportunities for more informed and accurate actions in the field. ● As a result, big data analytics are proving to be a real game-changer in the industry. ● With the increased availability of data, farmers have more tools than ever at their disposal, offering numerous benefits. Knowledge is not only powerful, but it can make or break an agricultural producer's success.

5.	Business Model (Revenue Model)	● the agricultural entrepreneur to make easier and better decisions based on information. This work aims to sensitize farmers to the importance of adopting data-driven solutions to address the needs of this sector.
6.	6. Scalability of the Solution	● Increases in agricultural productivity lead also to agricultural growth and can help to alleviate poverty in poor and developing countries, where agriculture often employs the greatest portion of the population. As farms become more productive, the wages earned by those who work in agriculture

3.4 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.



4.REQUIREMENT ANALYSIS:

4.1 FUNCTIONAL REQUIREMENT:

FR NO.	FUNCTIONAL REQUIREMENT(EPIC)	SUB REQUIREMENT(STORY/SUB-TASK)
FR-1	Uploading the dataset	Display the Manage Uploads page for the Data Set that will receive the data. Click the Upload file button. Select the file(s) you want to upload then click Upload.
FR-2	Preparation of dataset	Data collection. Relevant data is gathered from operational systems, data warehouses, data lakes and other data sources. □ Data discovery and profiling. □ Data cleansing. □ Data structuring. □ Data transformation and enrichment. □ Data validation and publishing.
FR-3	Exploratory Data Analysis	Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.
FR-4	Building a ML model	□ Contextualize machine learning in your organization. □ Explore the data and choose the type of algorithm. □ Prepare and clean the dataset. □ Split the prepared dataset and perform cross validation. □ Perform machine learning optimization.

		⍰ Deploy the model.
FR-5	Model Evaluation	Model evaluation is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Model evaluation is important to assess the efficacy of a model during initial research phases, and it also plays a role in model monitoring
FR-6	Data Pre-Processing	Model evaluation is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Model evaluation is important to assess the efficacy of a model during initial research phases, and it also plays a role in model monitoring
FR-7	Prediction Output	Predictive analytics is the process of using data analytics to make predictions based on data. This process uses data along with analysis, statistics, and machine learning techniques to create a predictive model for forecasting future events.

4.2 NON-FUNCTIONAL REQUIREMENT:

Following are the non-functional requirements of the proposed solution.

FR NO.	NON-FUNCTIONAL REQUIREMENT	DESCRIPTION
NFR-1	Usability	It helps the farmers to monitor the health of the crops in real time, create predictive analysis related to future yield.
NFR-2	Security	Data security functions to prevent data breaches, reduce risk of data exposure and ensure the ongoing safe and secure use of private data by minimizing exposure risk.
NFR-3	Reliability	The reliability of the data determines whether or not businesses can make good decisions with it. If the data is unreliable, It cannot be trusted, which makes it useless to the organizations

NFR-4	Performance	Regularly evaluating the performance of our organization can help us understand how much progress we're making toward our goals. A performance analysis is a tool you can use to check important metrics of crop yield for very month or year and make plans for adjustment and improvement.
NFR-5	Availability	Data availability in crop yield prediction is a term used by computer storage, manufacturers and storage service providers to describe how data should be available at the required level of performance in situations of predicting data used for crop yield ranching from normal through disastrous.

NFR-6	Scalability	The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field being used for management of crop yield and monitoring crop health. The recent trends in the domain of agriculture have made the people to understand the significance of big data

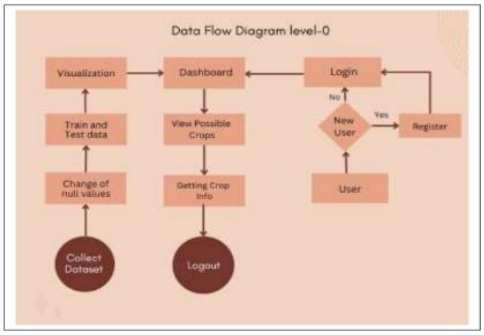
5. PROJECT DESIGN:

5.1 DATA FLOW DIAGRAM:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Example: (Simplified)

Example: DFD Level 0 (Industry Standard)



5.2 SOLUTION AND TECHNOLOGY ARCHITECTURE:

Technical Architecture for Handwritten Digit Recognition System:

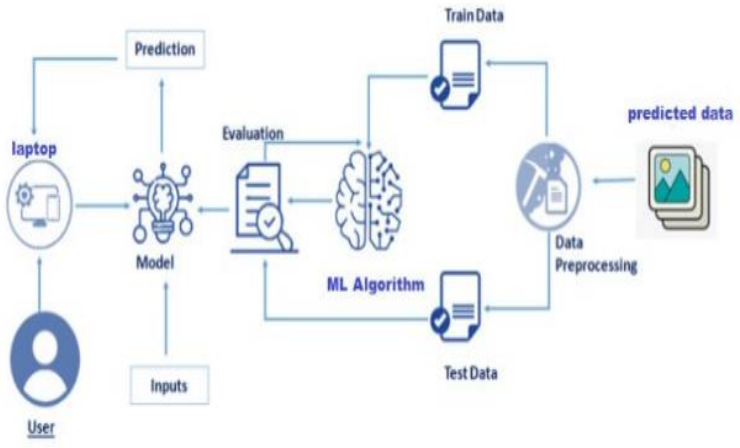


TABLE-1: COMPONENTS & TECHNOLOGIES:

User Interface	**Component**	**Description**	**Technology**
1.	User Interface	User interacts with the application using a web app	python
2.	Database	Data Type, Configurations etc.	MySQL, NoSQL, Excel etc.
3.	Cloud Database	Database Service on Cloud	IBM Cognos analytics.
4.	File Storage	Storage of user files in cloud account	IBM Block Storage or Other Storage Service or Local File system
5.	Machine Learning Model	Machine learning model is used to identify the estimation of crop yield using datasets	prediction Model.
6.	Infrastructure (Server / Cloud)	Application Deployment on Local System / AI Local Server Configuration AI Server Configuration	Local, Cloud Foundry, Kubernetes, etc.

Table-2: Application Characteristics:

S .No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Machine learning frameworks is used to train a predictive model	Py Torch, Open-cv, Google Collab.
2.	Scalable Architecture	The website traffic limit must be scalable enough to support 2 lakhs users at a time	3-tier
3.	Availability	Minimizing errors and maximizing accuracy	distributed servers
4.	Performance	Program should be less time consuming	number of requests per second

5.3 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

6.PROJECT PLANNING AND SHEDULING:

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	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Sindhu Maheswari
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Subasri ,Matheswari
Sprint-2		USN-3	As a user, I can register for the application through Google	2	Low	Saranya, Subasri
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	low	Saranya, Sindhu

Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Saranya, Mathesw ari
Sprint-3	Dashboard	USN-6	As a user, I can freely use my dashboard and explore the features	2	High	Sindhu, Subasri
Sprint-2		USN-7	Assauer, Assauer Akanistha credentials to access the resources of my application	2	High	Mathesw ari, Sindhu
Sprint-3		USN-8	Performance of Data manipulation s on the application	1	High	Sindhu, Subasri
Sprint-3	Visualizations	USN-9	I can create dashboards with particular datasets	2	High	Mathesw ari Sindhu
Sprint-4		USN-10	Predictive analysis can be done	1	High	Mathesw ari Subasri

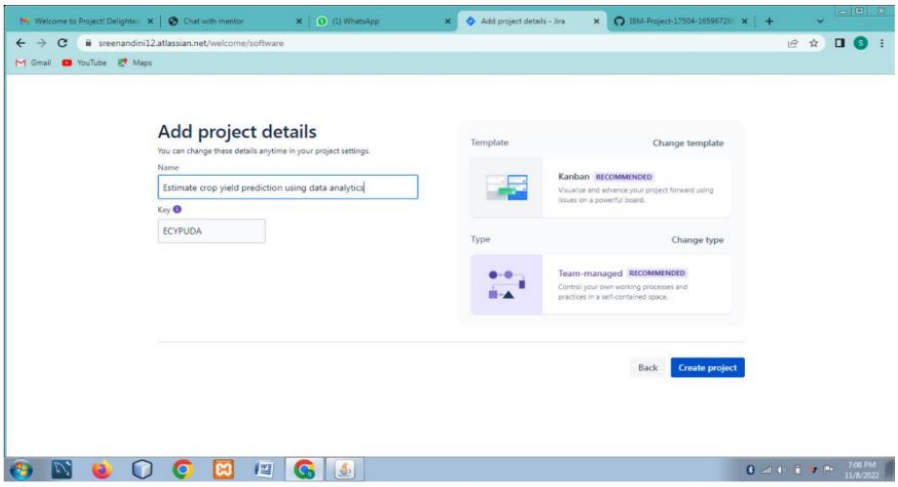
Sprint-3		USN-11	I can create stories with particular datasets	2	High	Subasri, Saranya
Sprint-4		USN-12	I can deliver and export reports according to the dashboards and stories created	2	High	Matheswari, Saranya

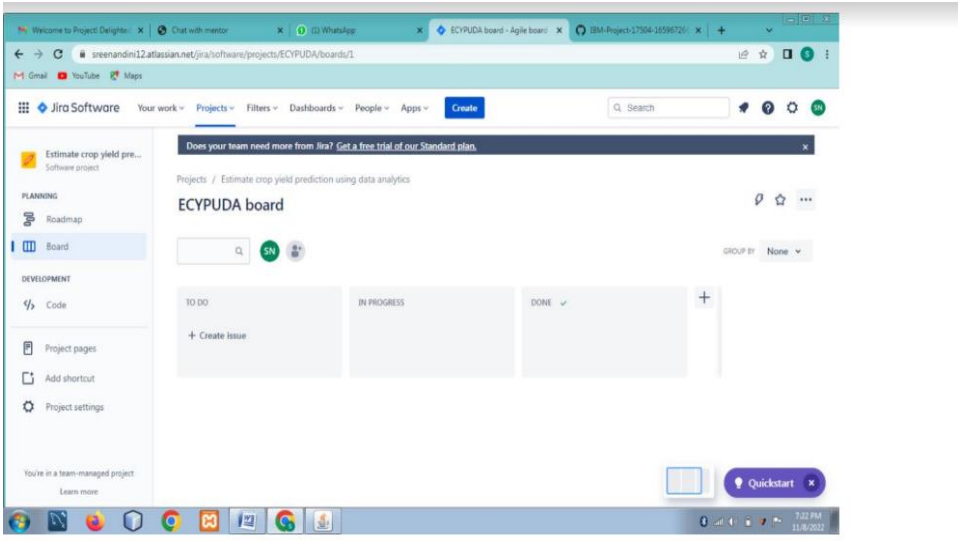
6.2 SPRINT DELIVERY SCHEDULE:

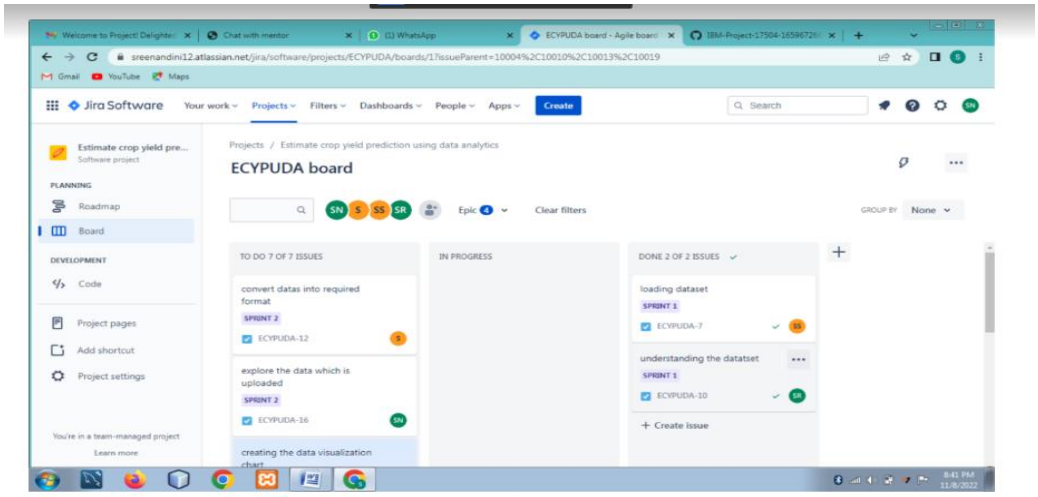
SPRINT	TOTAL STORY POINTS	DURATION	SPRINT START DAATE	SPRINT END DATA(PLANNED)	STORY POINTS COMPLETED (AS PLANNED END DATE)	SPRINT RELEASE DATE(ACTUAL)
SPRINT-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Nov 2022
SPRINT-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
SPRINT-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
SPRINT-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

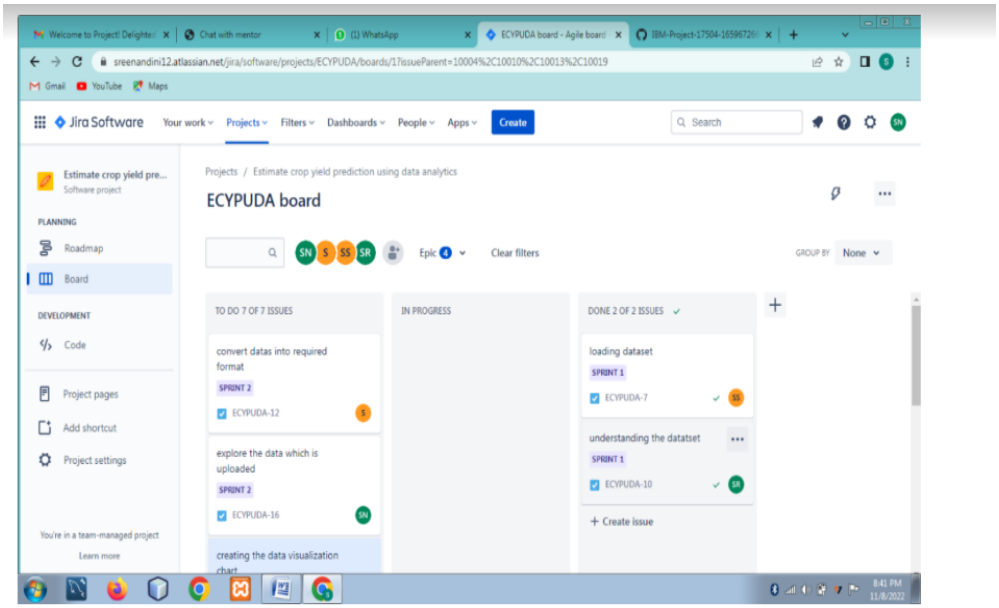
6.3 REPORT FROM JIRA:

A) SPRINT 1:

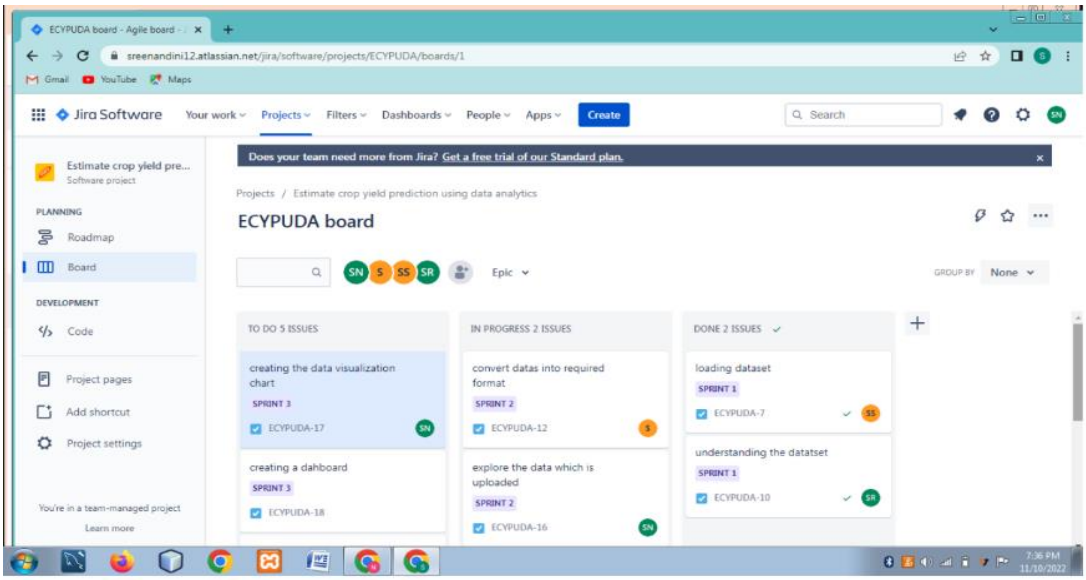


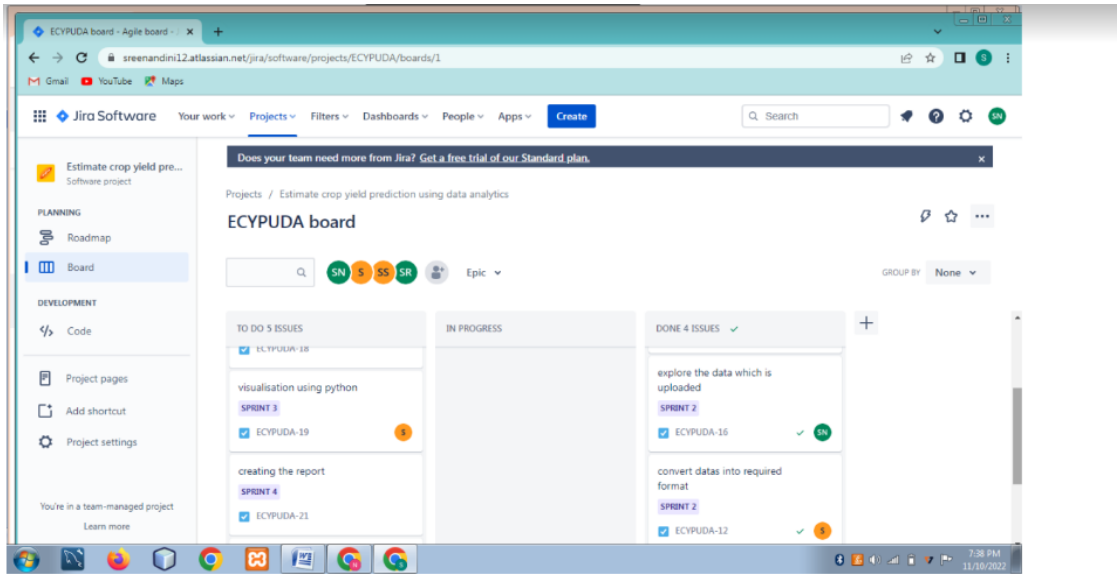


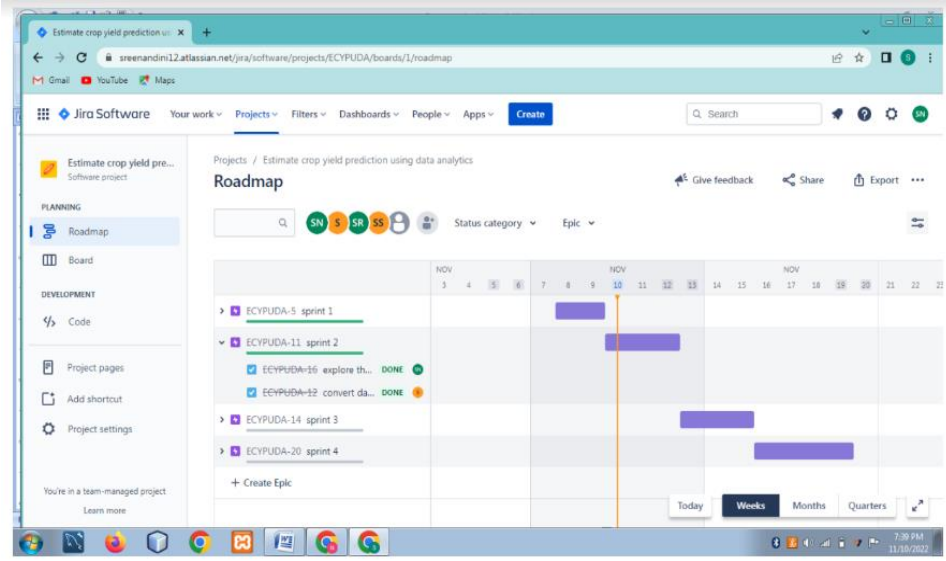

sreenandini12.atlassian.net/jira/software/projects/ECYPUDA/boards/1?issueParent=10004%2C10010%2C10013%2C10019
Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / Estimate crop yield prediction using data analytics
ECYPUDA board
Epic
Clear filters
GROUP BY
None
TO DO 7 OF 7 ISSUES
convert datas into required format
SPRINT 2
ECYPUDA-12
explore the data which is uploaded
SPRINT 2
ECYPUDA-16
creating the data visualization
IN PROGRESS
DONE 2 OF 2 ISSUES
loading dataset
SPRINT 1
ECYPUDA-7
understanding the datatset
SPRINT 1
ECYPUDA-10
+ Create issue


sreenandini12.atlassian.net/jira/software/projects/ECYPUDA/boards/1?issueParent=10004%2C10010%2C10013%2C10019
Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / Estimate crop yield prediction using data analytics
ECYPUDA board
Epic
Clear filters
GROUP BY
None
TO DO 7 OF 7 ISSUES
convert datas into required format
SPRINT 2
ECYPUDA-12
explore the data which is uploaded
SPRINT 2
ECYPUDA-16
creating the data visualization
IN PROGRESS
DONE 2 OF 2 ISSUES
loading dataset
SPRINT 1
ECYPUDA-7
understanding the datatset
SPRINT 1
ECYPUDA-10
+ Create issue

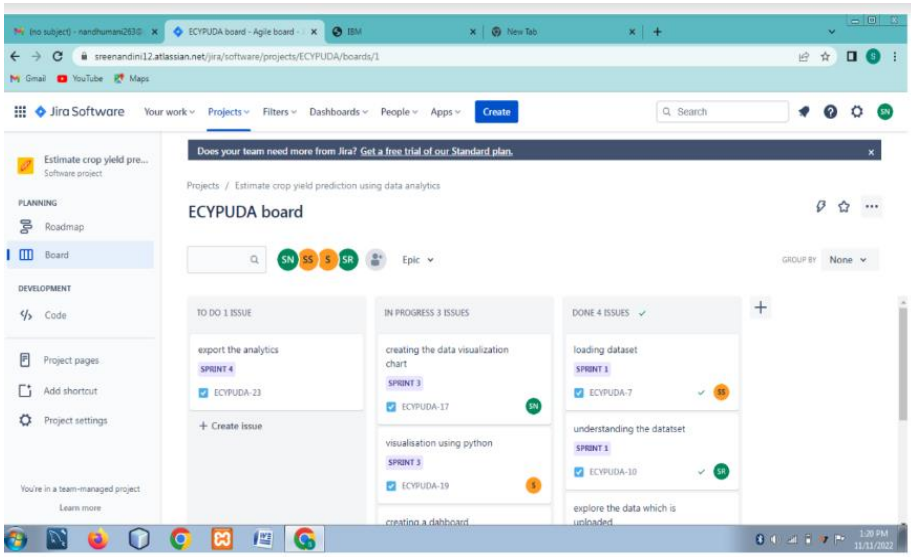
B) SPRINT 2:

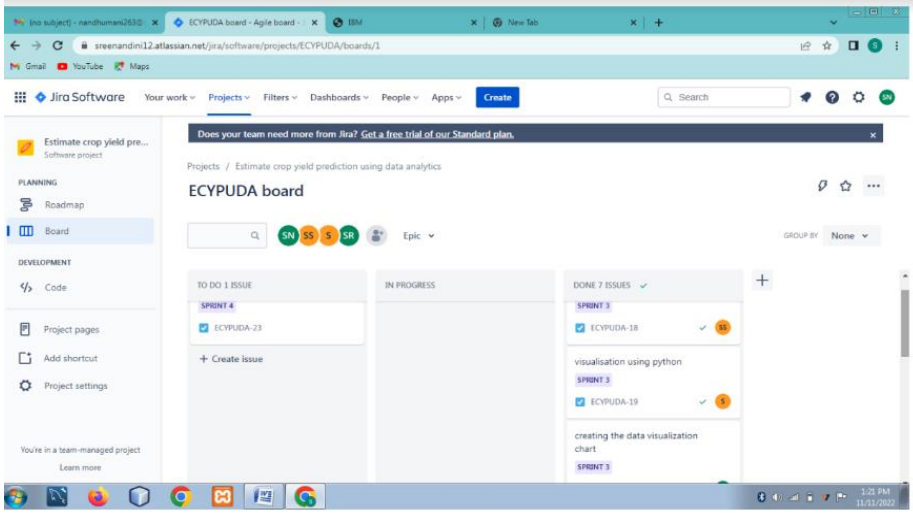

ECYPUDA board - Agile board
sreenandini12.atlassian.net/jira/software/projects/ECYPUDA/boards/1
Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
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 crop yield prediction using data analytics
ECYPUDA board
Epic
GROUP BY
None
TO DO 5 ISSUES
creating the data visualization chart
SPRINT 3
ECYPUDA-17
creating a dahboard
SPRINT 3
ECYPUDA-18
IN PROGRESS 2 ISSUES
convert datas into required format
SPRINT 2
ECYPUDA-12
explore the data which is uploaded
SPRINT 2
ECYPUDA-16
DONE 2 ISSUES
loading dataset
SPRINT 1
ECYPUDA-7
understanding the datatset
SPRINT 1
ECYPUDA-10

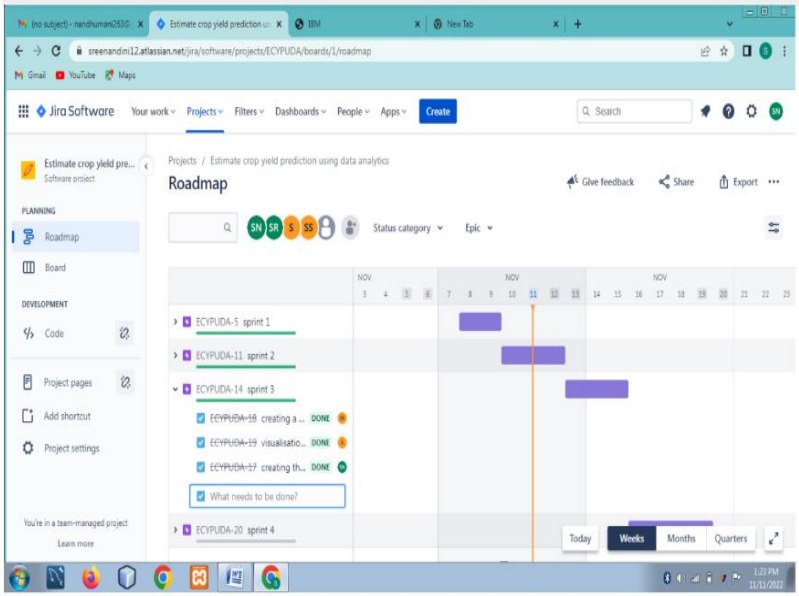

ECYPUDA board - Agile board
sreenandini12.atlassian.net/jira/software/projects/ECYPUDA/boards/1
Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
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 crop yield prediction using data analytics
ECYPUDA board
Epic
GROUP BY
None
TO DO 5 ISSUES
visualisation using python
SPRINT 3
ECYPUDA-19
creating the report
SPRINT 4
ECYPUDA-21
IN PROGRESS
DONE 4 ISSUES
explore the data which is uploaded
SPRINT 2
ECYPUDA-16
convert datas into required format
SPRINT 2
ECYPUDA-12


Jira Software
Roadmap
ECYPUDA-5 sprint 1
ECYPUDA-11 sprint 2
ECYPUDA-14 sprint 3
ECYPUDA-20 sprint 4
Create Epic

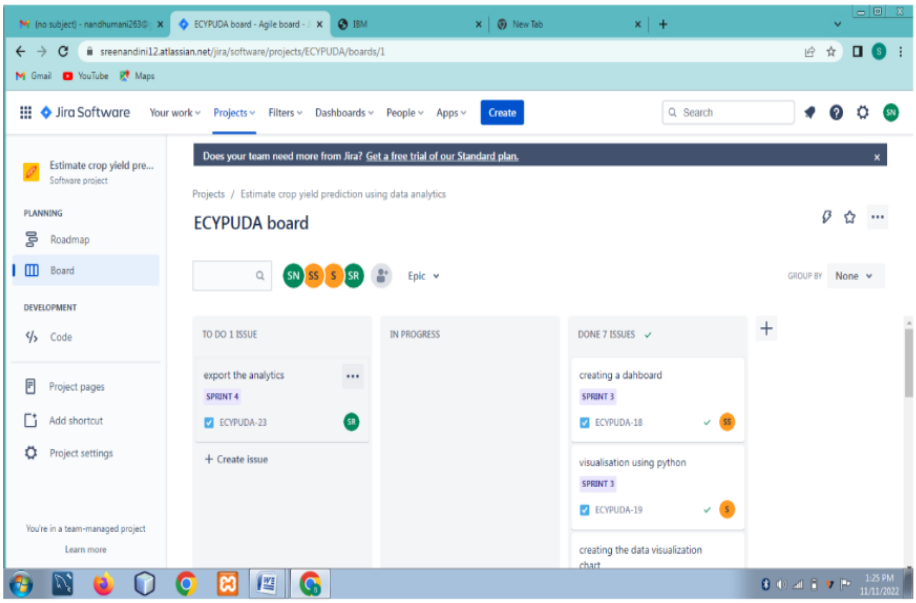
C) SPRINT 3:

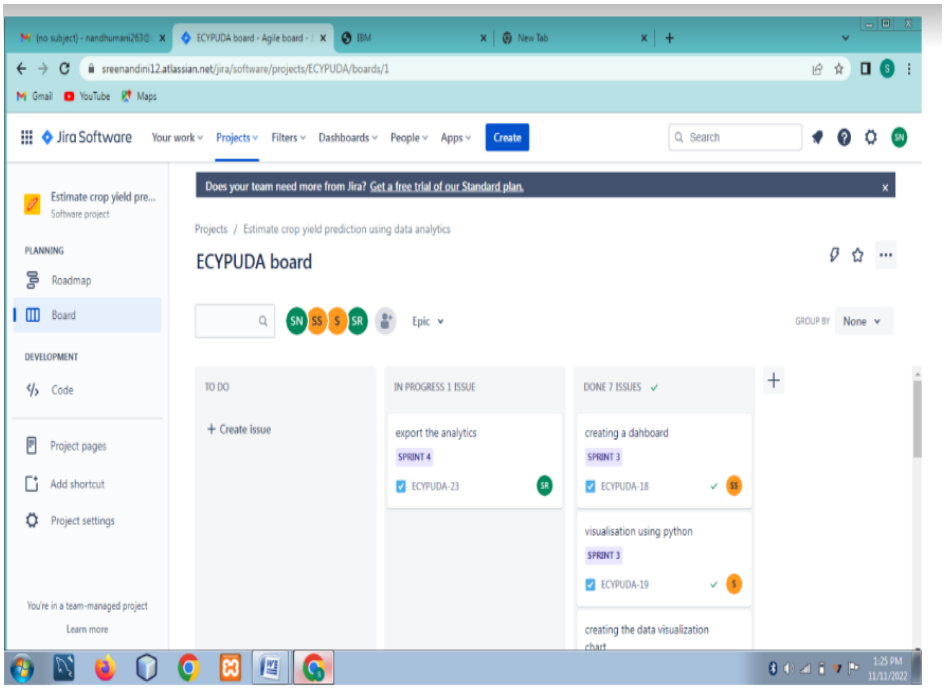

Jira Software
ECYPUDA board
TO DO 1 ISSUE
export the analytics
IN PROGRESS 3 ISSUES
creating the data visualization chart
visualisation using python
DONE 4 ISSUES
loading dataset
understanding the datatset
explore the data which is

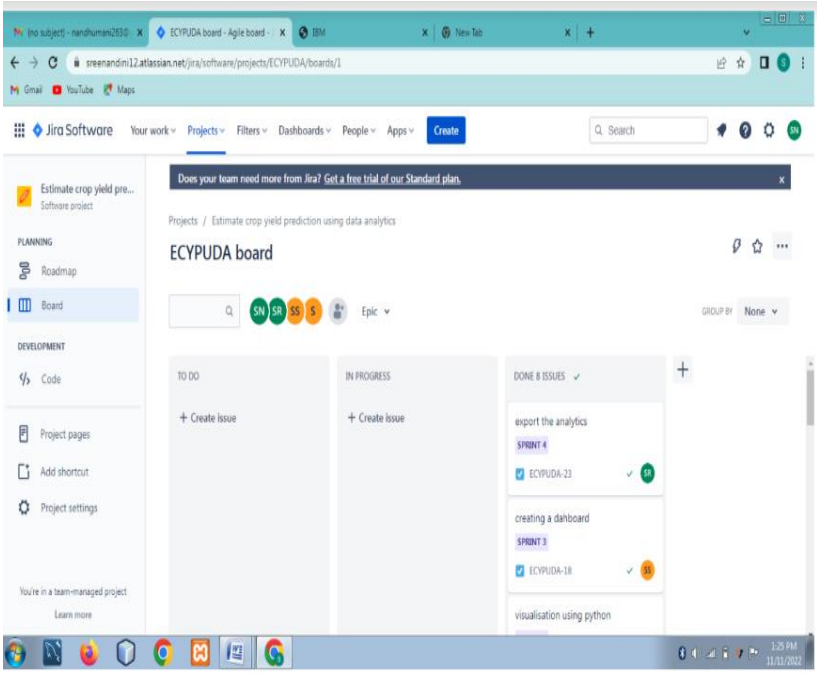

Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Does your team need more from Jira? Get a free trial of our Standard plan.
Projects / Estimate crop yield prediction using data analytics
ECYPUDA board
Epic
GROUP BY None
TO DO 1 ISSUE
SPRINT 4
ECYPUDA-23
+ Create issue
IN PROGRESS
DONE 7 ISSUES
SPRINT 3
ECYPUDA-18
visualisation using python
SPRINT 3
ECYPUDA-19
creating the data visualization
chart
SPRINT 3

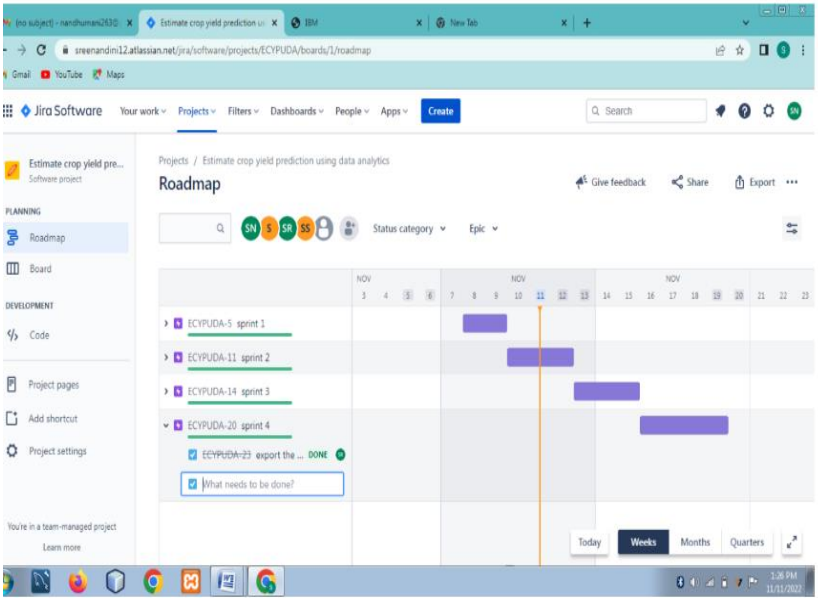

Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / Estimate crop yield prediction using data analytics
Roadmap
Give feedback
Share
Export
Status category
Epic
NOV
ECYPUDA-5 sprint 1
ECYPUDA-11 sprint 2
ECYPUDA-14 sprint 3
ECYPUDA-18 creating a ... DONE
ECYPUDA-19 visualisatio... DONE
ECYPUDA-17 creating th... DONE
What needs to be done?
ECYPUDA-20 sprint 4
Today
Weeks
Months
Quarters

D) SPRINT 4:


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
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 crop yield prediction using data analytics
ECYPUDA board
Epic
GROUP BY
None
TO DO 1 ISSUE
export the analytics
SPRINT 4
ECYPUDA-23
+ Create issue
IN PROGRESS
DONE 7 ISSUES
creating a dahboard
SPRINT 3
ECYPUDA-18
visualisation using python
SPRINT 3
ECYPUDA-19
creating the data visualization


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
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 crop yield prediction using data analytics
ECYPUDA board
Epic
GROUP BY
None
TO DO
+ Create issue
IN PROGRESS 1 ISSUE
export the analytics
SPRINT 4
ECYPUDA-23
DONE 7 ISSUES
creating a dahboard
SPRINT 3
ECYPUDA-18
visualisation using python
SPRINT 3
ECYPUDA-19
creating the data visualization


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
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 crop yield prediction using data analytics
ECYPUDA board
Epic
GROUP BY
None
TO DO
+ Create issue
IN PROGRESS
+ Create issue
DONE 8 ISSUES
export the analytics
SPRINT 4
ECYPUDA-23
creating a dahboard
SPRINT 3
ECYPUDA-18
visualisation using python

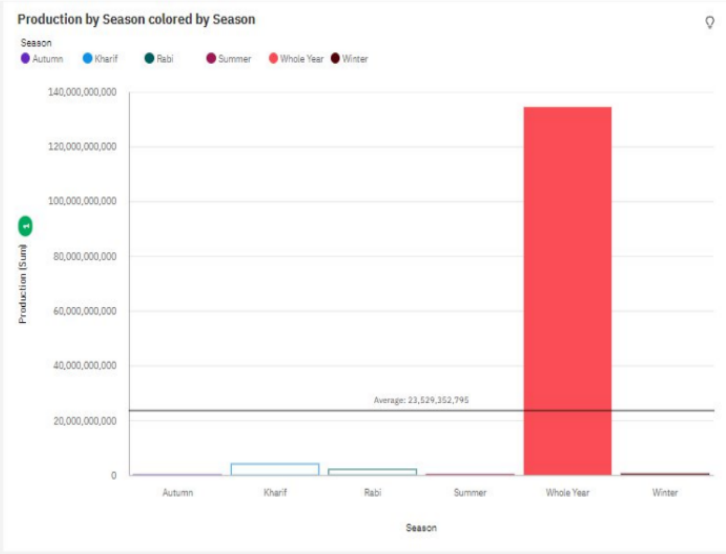

Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
Estimate crop yield pre...
Software project
PLANNING
Roadmap
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / Estimate crop yield prediction using data analytics
Roadmap
Give feedback
Share
Export
Status category
Epic
NOV
ECYPUDA-5 sprint 1
ECYPUDA-11 sprint 2
ECYPUDA-14 sprint 3
ECYPUDA-20 sprint 4
ECYPUDA-23 export the ... DONE
What needs to be done?
Today
Weeks
Months
Quarters

7. CODING AND SOLUTIONING:

7.1 FEARTURE 1:

A) DATA EXPLORATION:

SEASONS WITH AVERAGE PRODUCTION:



2)WITH YEARS USAGE OF AREA AND PRODUCTION:

DATA VISUALIZATION CHARTS

WITH YEARS USAGE OF AREA AND PRODUCTION:

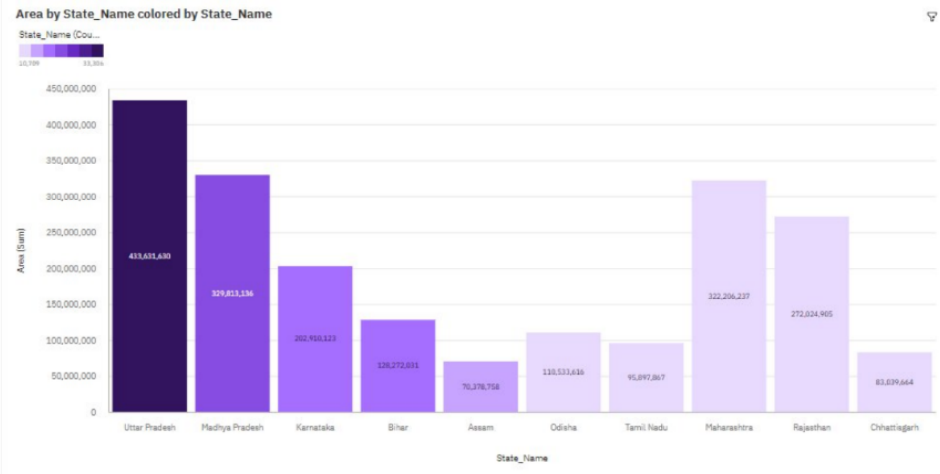
Area and Production for Crop_Year

	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	##########
2014	115,757,541.56	##########
2015	4,601,298	6,935,064.7
Summary	##########	##########

2013
Production: 12,903,588,632.88

3) TOP 10 STATES WITH THE MOST AREA

TOP 10 STATES WITH MOST AREA:



4) STATES WITH CROP PRODUTION WITH SEASONS:

STATES WITH CROP PRODUCTION WITH SEASONS:

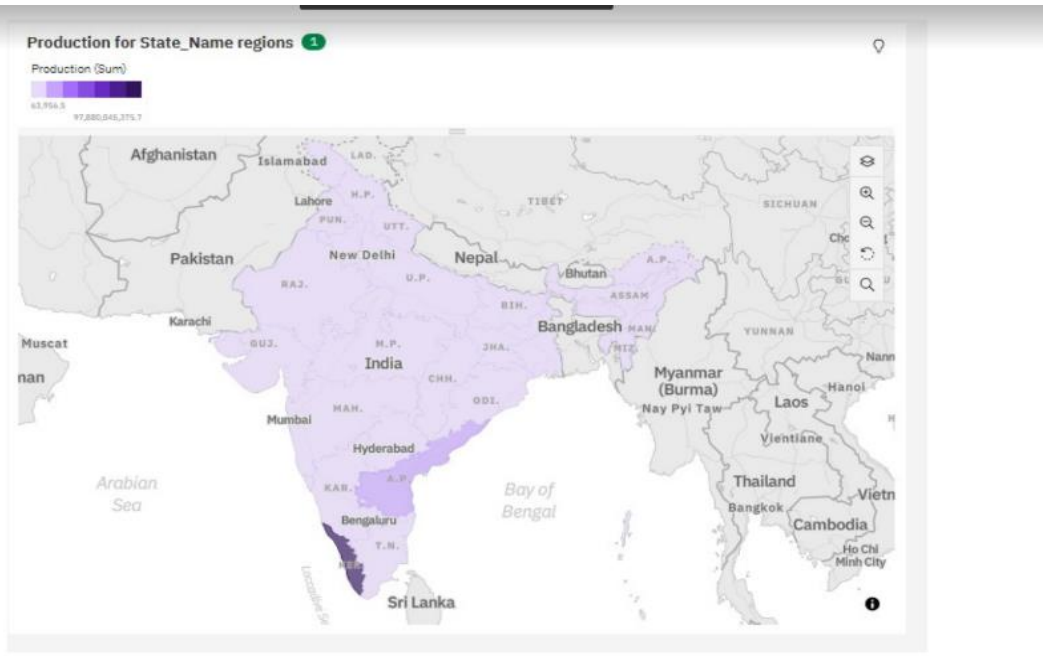
Production for Season, State_Name and Crop 1

Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Andaman and Ni...	Urad	(no value)	(no value)	475	(no value)	(no value)	(no value)	475
	other oilseeds	(no value)	(no value)	(no value)	(no value)	248.31	(no value)	248.31
	Summary	**25,248.95**	**214,239.06**	**11,598.92**	**(no value)**	**717,972,153.06**	**(no value)**	**718,223,239.99**
Andhra Pradesh	Arecanut	(no value)	620	(no value)	(no value)	3,164	(no value)	3,784
	Arhar/Tur	(no value)	1,527,520	22,435	(no value)	(no value)	(no value)	1,549,955
	Bajra	(no value)	1,192,328	73,381	(no value)	(no value)	(no value)	1,265,709
	Banana	(no value)	(no value)	(no value)	(no value)	15,041,631	(no value)	15,041,631
	Beans & Mutter(...	(no value)	(no value)	(no value)	(no value)	70,041	(no value)	70,041
	Bhindi	(no value)	(no value)	(no value)	(no value)	199,605	(no value)	199,605
	Bottle Gourd	(no value)	(no value)	(no value)	(no value)	0	(no value)	0
	Brinjal	(no value)	303,726	248,154	(no value)	463,200	(no value)	1,015,080
	Cabbage	(no value)	34,615	37,504	(no value)	0	(no value)	72,119
	Cashewnut	(no value)	110,702	(no value)	(no value)	408,463	(no value)	519,165
	Castor seed	(no value)	362,476	6,934	(no value)	25,783	(no value)	395,193
	Citrus Fruit	(no value)	(no value)	(no value)	(no value)	456,274	(no value)	456,274
	Coconut	(no value)	(no value)	(no value)	(no value)	16,806,002,260	(no value)	16,806,002,260
	Coriander	(no value)	1,135	29,915	(no value)	135,326	(no value)	166,376

Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Assam	Peas & beans (P...	(no value)	(no value)	220,368	(no value)	(no value)	(no value)	220,368
	Pineapple			1,027,864	(no value)	433,460	(no value)	1,461,324
	Potato				(no value)	11,021,521	(no value)	11,021,521
	Rapeseed &Must...	(no value)	(no value)	2,495,723	(no value)	(no value)	(no value)	2,495,723
	Rice	7,008,303	(no value)	(no value)	13,677,048	(no value)	51,115,739	71,801,090
	Sesamum	(no value)	131,973	(no value)	(no value)	8,257	(no value)	140,230
	Small millets	(no value)	69,650	(no value)	(no value)	(no value)	(no value)	69,650
	Sugarcane	(no value)	1,287,451	(no value)	(no value)	17,558,543	(no value)	18,845,994
	Sweet potato	(no value)	(no value)	115,095	(no value)	386,214	(no value)	501,309
	Tapioca	(no value)	117,319	(no value)	(no value)	180,478	(no value)	297,797
	Tobacco	(no value)	612	(no value)	(no value)	5,743	(no value)	6,355
	Turmeric	(no value)	(no value)	(no value)	(no value)	180,527	(no value)	180,527
	Urad	(no value)	(no value)	338,637	(no value)	(no value)	(no value)	338,637
	Wheat	(no value)	(no value)	1,218,370	(no value)	(no value)	(no value)	1,218,370
	other misc. pulses	(no value)	(no value)	6,672	(no value)	(no value)	(no value)	6,672
	Summary	**8,230,196**	**15,782,850**	**8,404,649**	**17,171,024**	**1,996,667,293**	**65,495,747**	**2,111,751,759**
	Arhar/Tur	(no value)	817,139	(no value)	(no value)	(no value)	(no value)	817,139
	Bajra	(no value)	51,189	(no value)	(no value)	(no value)	(no value)	51,189

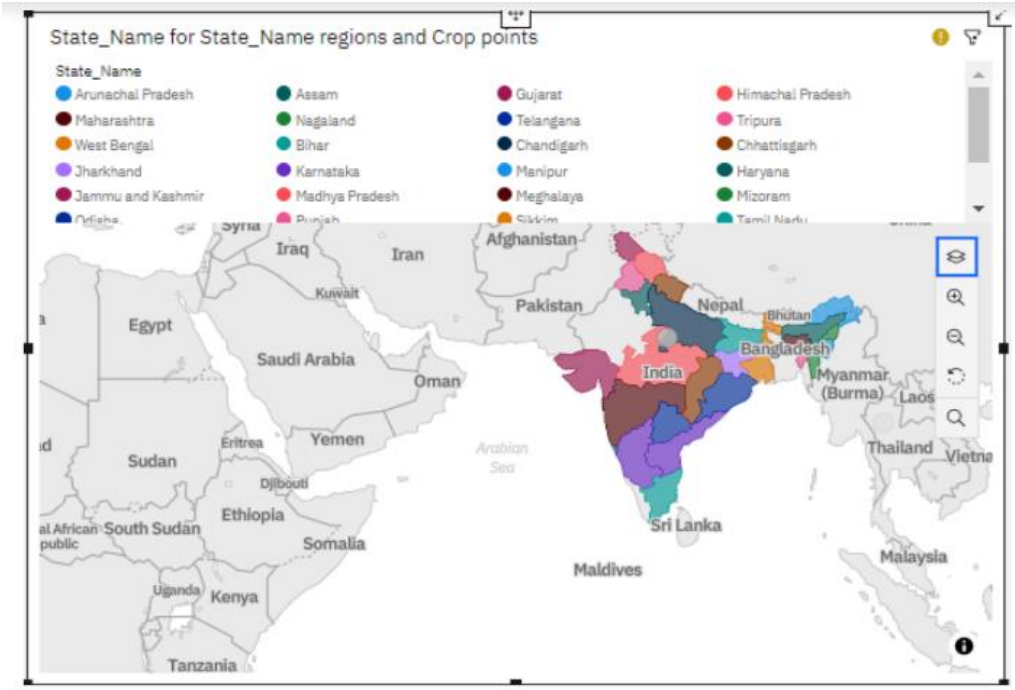
Rabi | Assam | Peas & beans (Pulses)
Production: 220,368

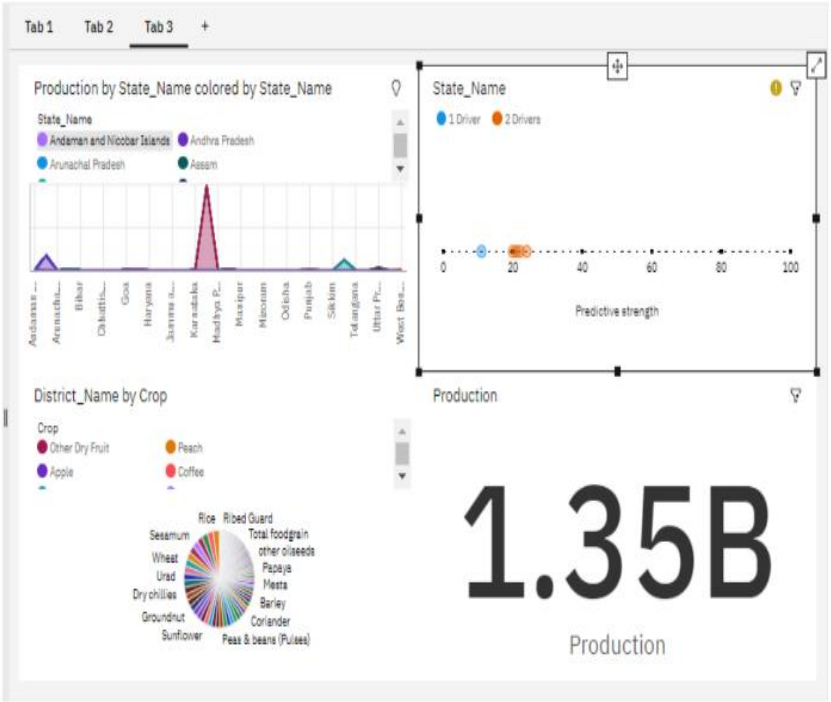
5) STATES WITH CROP PRODUCTION:

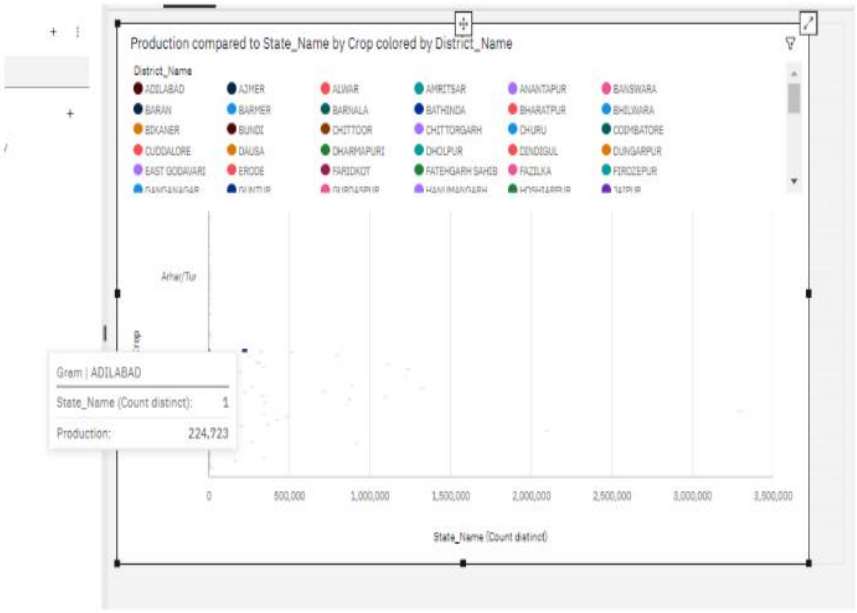


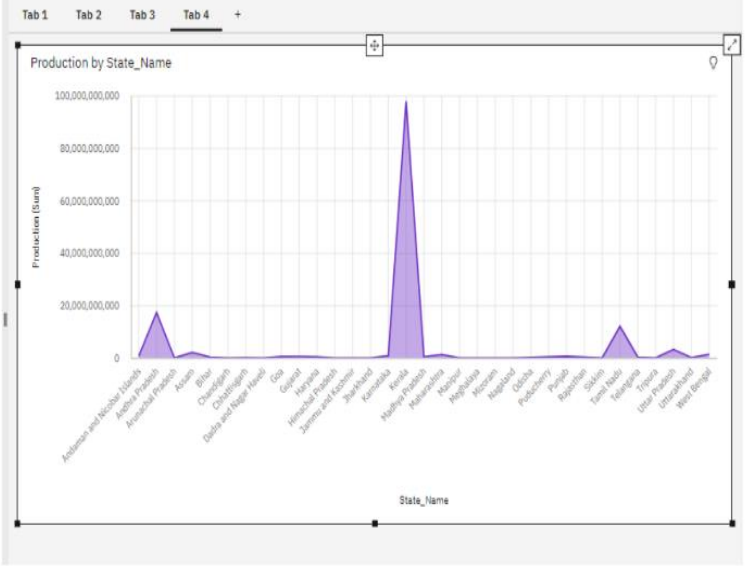
7.2 FEATURE 2:

B) CREATING THE DASHBOARD:






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

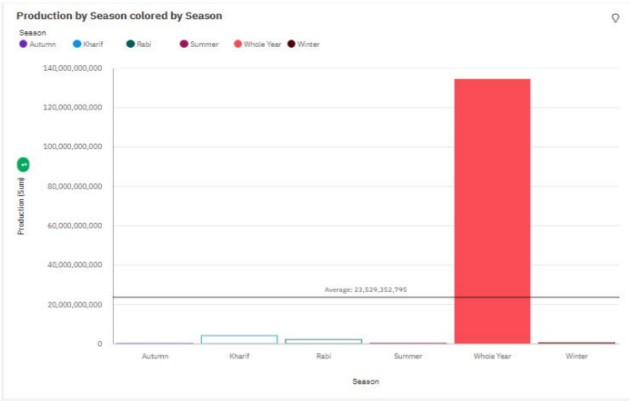
7.3 CODING:

https://github.com/IBM-EPBL/IBM-Project-20106-1659712640/blob/main/project%20development%20phase/Sprint%203/Esimate_crop_yield_prediction.ipynb

8.RESULTS:

A) DATA EXPLORATION:

SEASONS WITH AVERAGE PRODUCTION:



DATA VISUALIZATION CHARTS

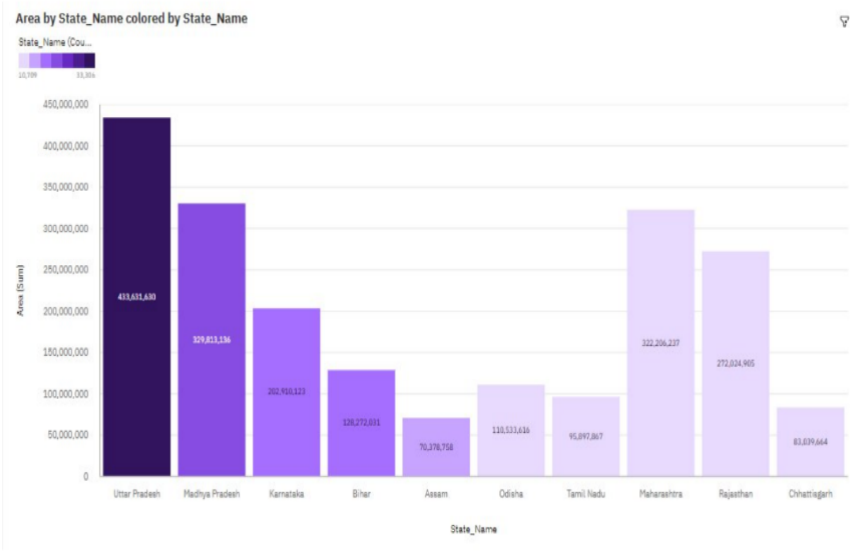
WITH YEARS USAGE OF AREA AND PRODUCTION:

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	##########
2014	115,757,541.56	##########
2015	4,601,298	6,935,064.7
Summary	##########	##########

2013
Production: 12,903,588,632.88

TOP 10 STATES WITH MOST AREA:



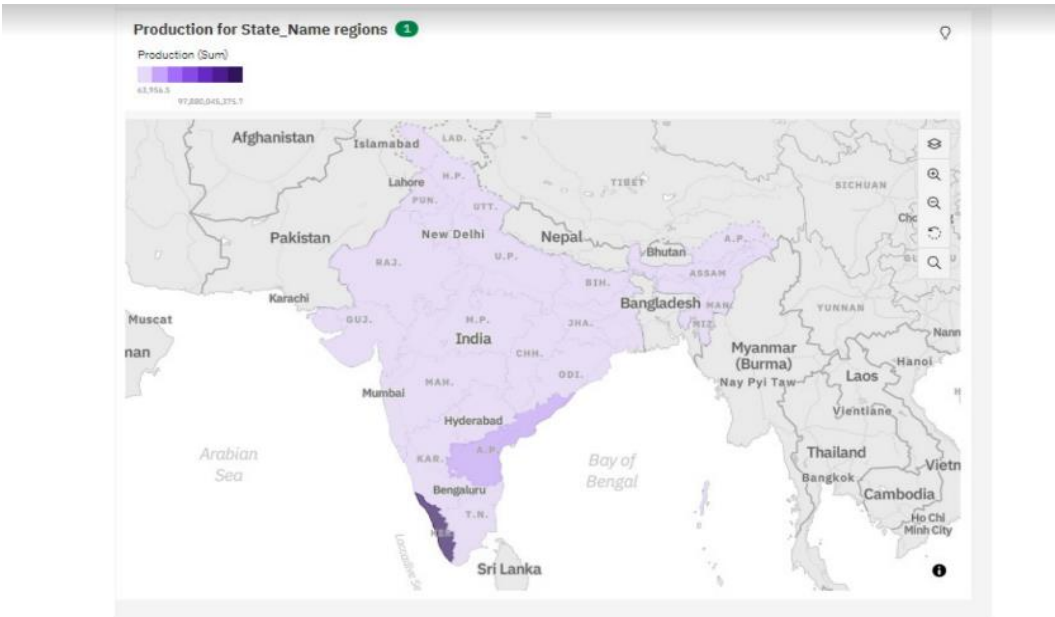
STATES WITH CROP PRODUCTION WITH SEASONS:

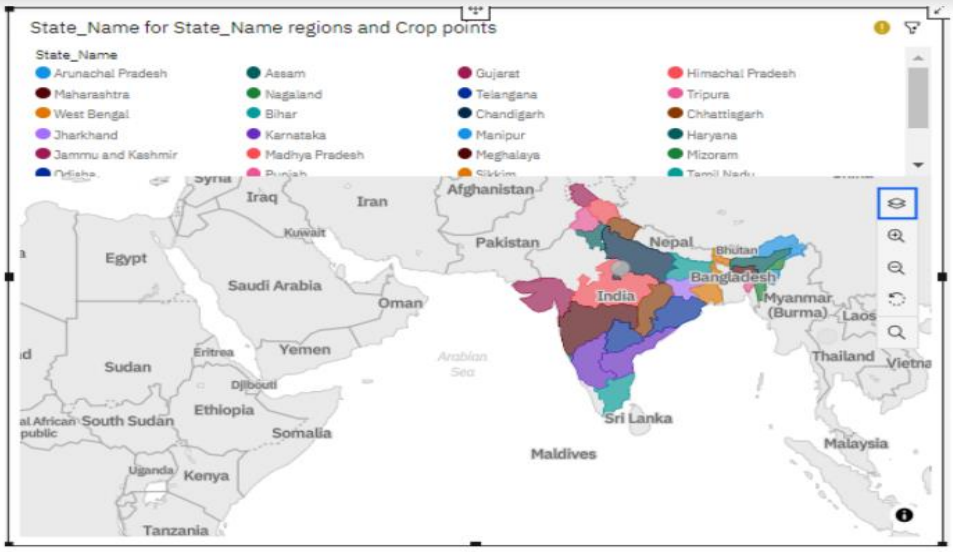
Production for Season, State_Name and Crop 1

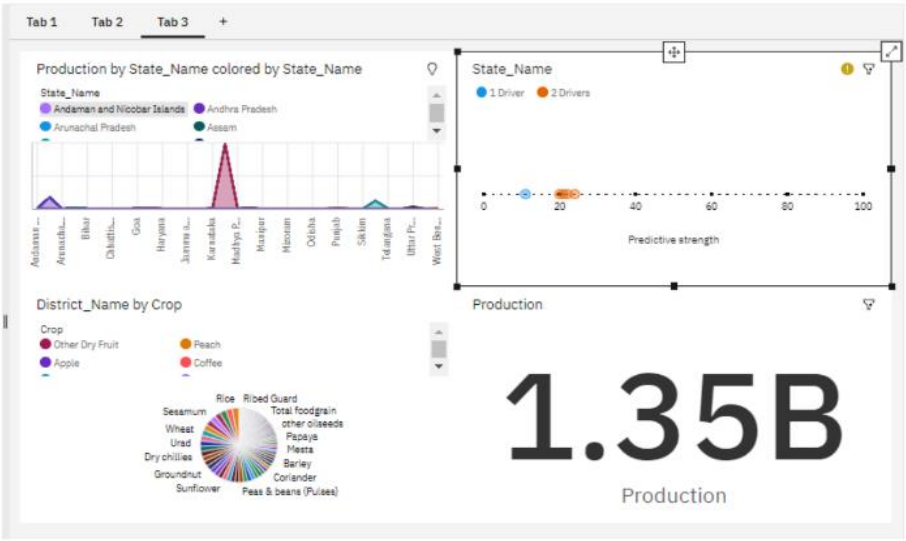
Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Andaman and Ni...	Urad	(no value)	(no value)	475	(no value)	(no value)	(no value)	**475**
	other oilseeds	(no value)	(no value)	(no value)	(no value)	248.31	(no value)	**248.31**
	Summary	**25,248.95**	**214,239.06**	**11,598.92**	**(no value)**	**717,972,153.06**	**(no value)**	**718,223,239.99**
Andhra Pradesh	Arecanut	(no value)	620	(no value)	(no value)	3,164	(no value)	**3,784**
	Arhar/Tur	(no value)	1,527,520	22,435	(no value)	(no value)	(no value)	**1,549,955**
	Bajra	(no value)	1,192,328	73,381	(no value)	(no value)	(no value)	**1,265,709**
	Banana	(no value)	(no value)	(no value)	(no value)	15,041,631	(no value)	**15,041,631**
	Beans & Mutter(...	(no value)	(no value)	(no value)	(no value)	70,041	(no value)	**70,041**
	Bhindi	(no value)	(no value)	(no value)	(no value)	199,605	(no value)	**199,605**
	Bottle Gourd	(no value)	(no value)	(no value)	(no value)	0	(no value)	**0**
	Brinjal	(no value)	303,726	248,154	(no value)	463,200	(no value)	**1,015,080**
	Cabbage	(no value)	34,615	37,504	(no value)	0	(no value)	**72,119**
	Cashewnut	(no value)	110,702	(no value)	(no value)	408,463	(no value)	**519,165**
	Castor seed	(no value)	362,476	6,934	(no value)	25,783	(no value)	**395,193**
	Citrus Fruit	(no value)	(no value)	(no value)	(no value)	456,274	(no value)	**456,274**
	Coconut	(no value)	(no value)	(no value)	(no value)	16,806,002,260	(no value)	**16,806,002,260**
	Coriander	(no value)	1,135	29,915	(no value)	135,326	(no value)	**166,376**

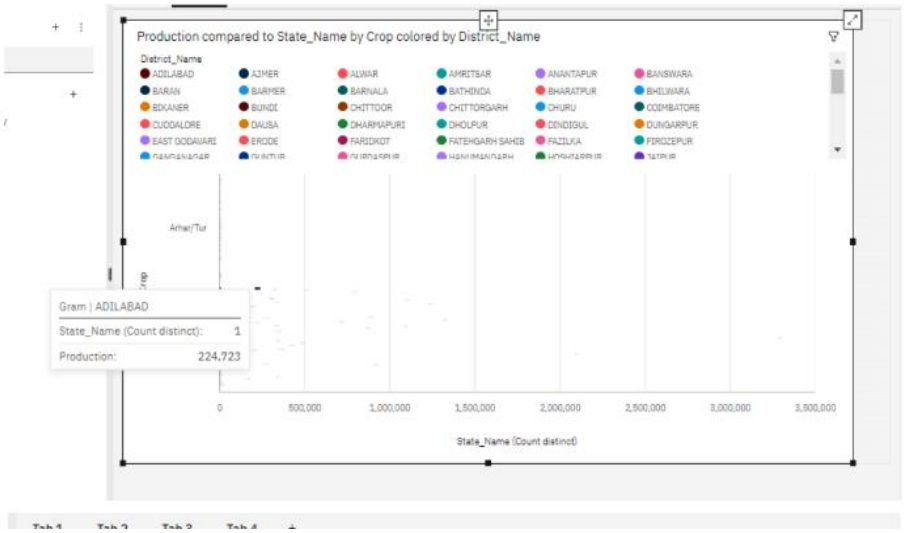
Production		Autumn	Kharif	Rabi	Summer	Whole Year	Winter	Summary
Assam	Peas & beans (P...	(no value)	(no value)	220,368	(no value)	(no value)	(no value)	**220,368**
	Pineapple			1,027,864	(no value)	433,460	(no value)	**1,461,324**
	Potato			(no value)	(no value)	11,021,521	(no value)	**11,021,521**
	Rapeseed &Must...	(no value)	(no value)	2,495,723	(no value)	(no value)	(no value)	**2,495,723**
	Rice	7,008,303	(no value)	(no value)	13,677,048	(no value)	51,115,739	**71,801,090**
	Sesamum	(no value)	131,973	(no value)	(no value)	8,257	(no value)	**140,230**
	Small millets	(no value)	69,650	(no value)	(no value)	(no value)	(no value)	**69,650**
	Sugarcane	(no value)	1,287,451	(no value)	(no value)	17,558,543	(no value)	**18,845,994**
	Sweet potato	(no value)	(no value)	115,095	(no value)	386,214	(no value)	**501,309**
	Tapioca	(no value)	117,319	(no value)	(no value)	180,478	(no value)	**297,797**
	Tobacco	(no value)	612	(no value)	(no value)	5,743	(no value)	**6,355**
	Turmeric	(no value)	(no value)	(no value)	(no value)	180,527	(no value)	**180,527**
	Urad	(no value)	(no value)	338,637	(no value)	(no value)	(no value)	**338,637**
	Wheat	(no value)	(no value)	1,218,370	(no value)	(no value)	(no value)	**1,218,370**
	other misc. pulses	(no value)	(no value)	6,672	(no value)	(no value)	(no value)	**6,672**
	Summary	**8,230,196**	**15,782,850**	**8,404,649**	**17,171,024**	**1,996,667,293**	**65,495,747**	**2,111,751,759**
	Arhar/Tur	(no value)	817,139	(no value)	(no value)	(no value)	(no value)	**817,139**
	Bajra	(no value)	51,189	(no value)	(no value)	(no value)	(no value)	**51,189**

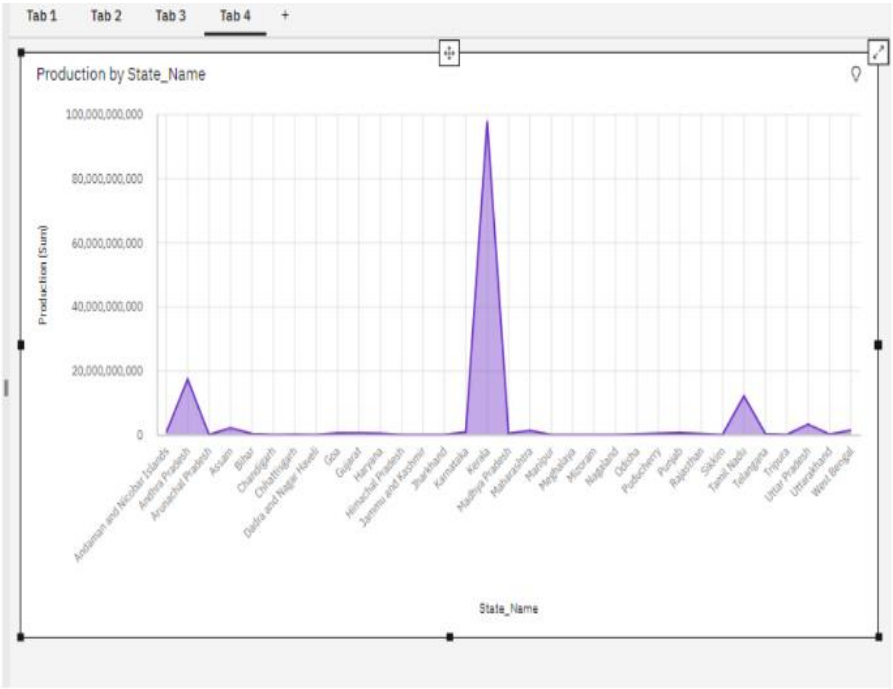
Rabi | Assam | Peas & beans (Pulses)
Production: 220,368




State_Name for State_Name regions and Crop points
State_Name
Arunachal Pradesh
Assam
Gujarat
Himachal Pradesh
Maharashtra
Nagaland
Telangana
Tripura
West Bengal
Bihar
Chandigarh
Chhattisgarh
Jharkhand
Karnataka
Manipur
Haryana
Jammu and Kashmir
Madhya Pradesh
Meghalaya
Mizoram
Iraq
Iran
Afghanistan
Kuwait
Pakistan
Nepal
Bhutan
Egypt
Saudi Arabia
Bangladesh
India
Oman
Myanmar (Burma)
Laos
Yemen
Eritrea
Thailand
Vietna
Sudan
Djibouti
Arabian Sea
Ethiopia
South Sudan
Sri Lanka
Somalia
Maldives
Malaysia
Uganda
Kenya
Tanzania


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
Sesamum
Total foodgrain
other oilseeds
Wheat
Papaya
Urad
Mesta
Dry chillies
Barley
Groundnut
Coriander
Sunflower
Peas & beans (Pulses)
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

8. 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.

9.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.

10. FUTURE SCOPE:

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

11.APPENDIX:

- 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

• https://www.agriculture.com/technology/crop-management/trimble-claas-to-develop-next-gen-precision-system-for-claas-ag-equipment

• https://www.ijert.org/crop-prediction-using-machine-learning-approaches