ESTIMATE THE CROP YIELD USING DATA ANALYTICS

SUGUMAR R, SANJAY R, NADEESHWARAN A, SHARUK B

1. INTRODUCTION:

1.1 PROJECT OVERVIEW:

Agriculture is the backbone of Indian Economy In India, majority of the farmers are not getting the expected crap 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 firmer'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.

1.2 PURPOSE:

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. LITERATURE SURVEY:

2.1 EXISTING PROBLEM:

Crop Predictions are important for governments, farmers ,and traders - a more accurate prediction of future supply allows for better policies,better profits,and better pricing.

The concept of using satellites to predict crop yield is very simple:Take a picture of a field,look at the crops growing, measure the area of field and predict the yield.However, there are substantial challenges to this,including:

- Clouds
- Scale-global
- Size-small farm to large agri-business
- Colour-it's all green
- Phenology-Identification of relevant crops
- Area v Yield - Giving the area does not give the yield.

2.2. REFERENCES:

- https:/ ieeexplore.ieee.org/document/8697806
- https:/ ieeexplore.ieee.org/document/7918789
- https:/ www.researchgate.net/publication/329467349_Agriculture_Data_Analytics_in_Crop_Yield_Estimation_A_Critical_Review
- https:/ ieeexplore.ieee.org/document/8746001/references#references

2.3 PROBLEM STATEMENT DEFINITION:

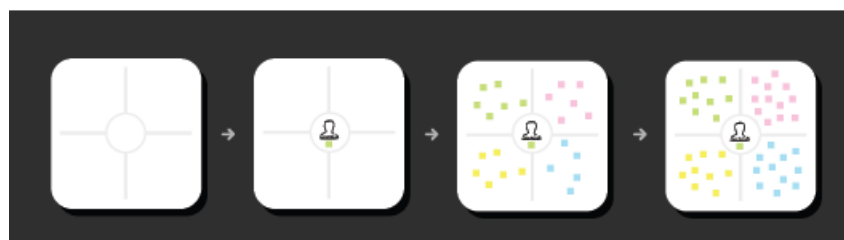| Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	A Farmer	Harvest my crops	I am not sure about the best time to harvest the crops	Harvesting in a wrong season can reduce my yield	Disappointment
PS-2	A Farmer	Get more yield when harvesting	Some birds, insects and animals damages the crops they eat the crops which reduces my production	they eat the crops which reduces my production	Frustrated
PS-3	A Farmer	Get proper production	I don't know exactly which crop to plant in my field based on my locality	I got low production last year when I planted a crop that doesn't suit my locality/soil nature	Dispirited

3.IDEATION & PROPOSED SOLUTION:
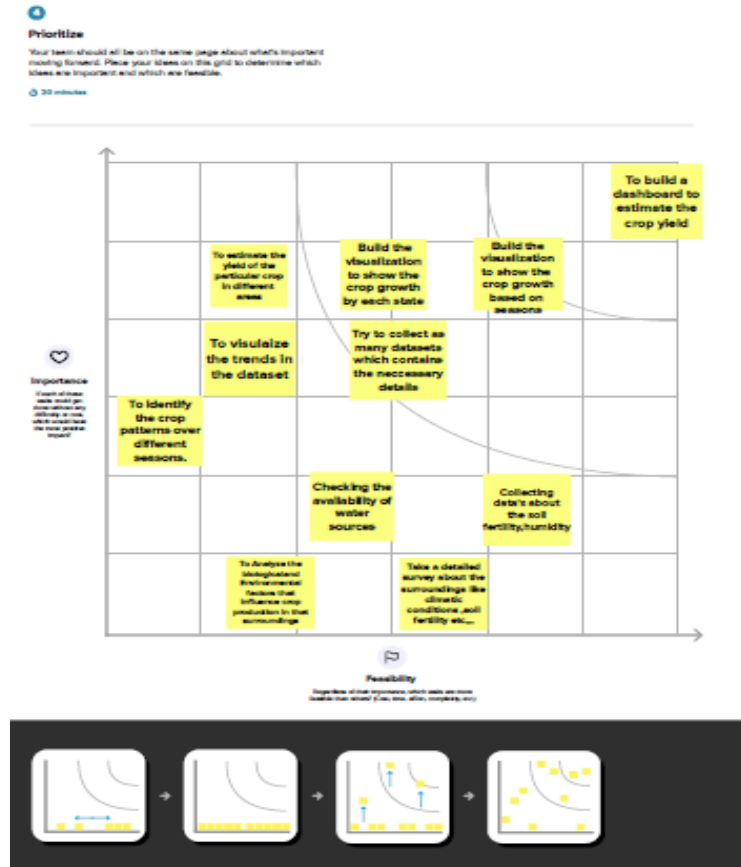
3.1 EMPATHY MAP CANVAS:

Build empathy

The information you add here should be representative of the observations and research you've done about your users.

Says

what is the condition of the field

The chemicals should be avoided

The organic products are very costly

Thinks

less man power

Extreme weather factors like heavy storms, droughts are disrupting the crops

knowledge about the climatic factors

I didn't get any profit

Does

They will ask about the quality of fertilizer

Check for products

List pros and cons

Feels

Climate changes

unhealthy soil

water scarcity

3.2 IDEATION & BRAINSTORMING:

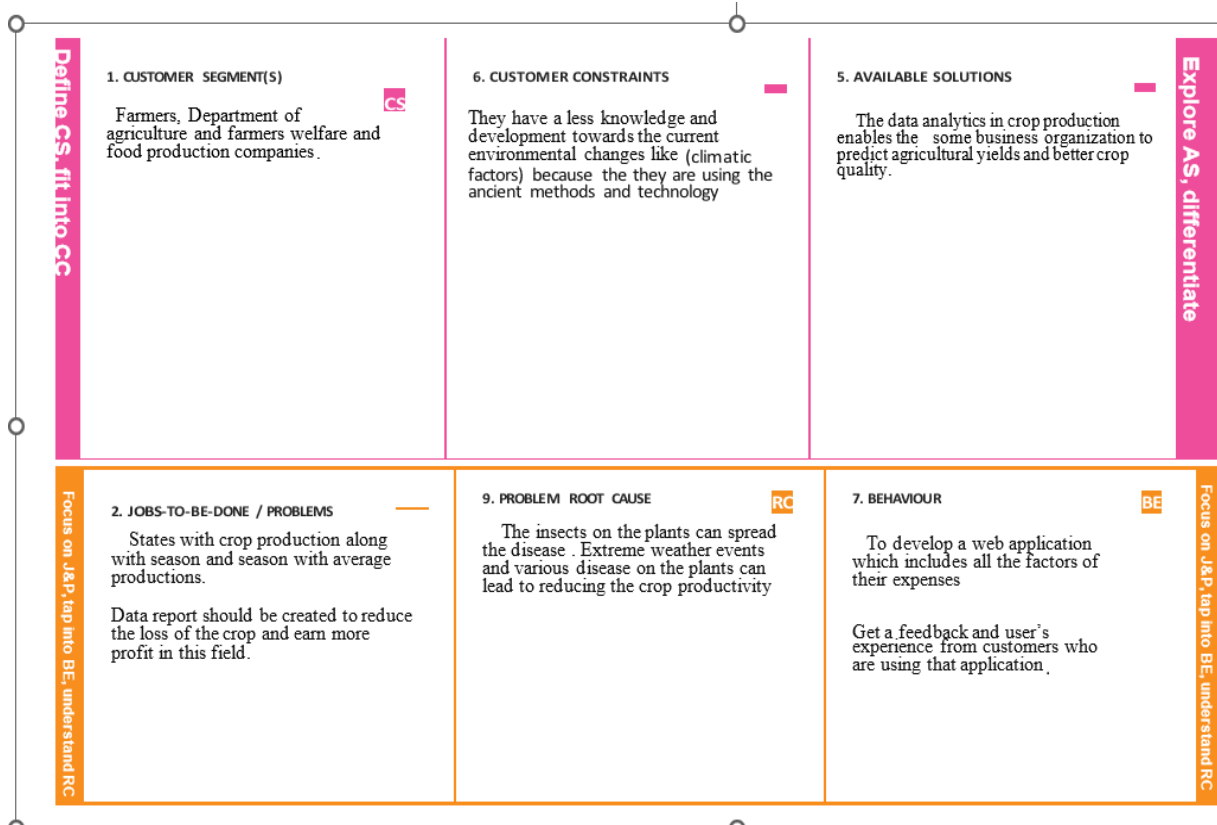**TOP THREE IDEAS**:

3.2.1   Understand how the data analytics platform will support the overall business strategy ofthe organization.

3.2.2   Develop an analytics vision and set target maturity levels for core processes.

3.2.3   Increasing Innovation, Productivity & Reduce wastage.

3.3 PROPOSED SOLUTION:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The main objective of this project is to predict crop-yield which can be extremely useful to farmers in planning for harvesting and cultivating the crops.
2.	Idea / Solution description	The farmer should get information about the climatic factors because it will save the crop from natural disaster.
3.	Novelty / Uniqueness	By visualizing the data, the farmers can able to choose the crop which will give more profit in the desired season
4.	Social Impact / Customer Satisfaction	Knowledge about the climatic factors to save the farmers from losses and crop damage.
5.	Business Model (Revenue Model)	The farmers will take better decision with the help of dashboard which shows the visualization of the crop production.
6.	Scalability of the Solution	Based on the data analytics report they can cultivate the crops according to some factors like climatic conditions, soil fertility level, etc…

3.4 PROBLEM SOLUTION FIT:

1. CUSTOMER SEGMENT(S) `CS`

Farmers, Department of agriculture and farmers welfare and food production companies .

6. CUSTOMER CONSTRAINTS —

They have a less knowledge and development towards the current environmental changes like (climatic factors) because the they are using the ancient methods and technology

5. AVAILABLE SOLUTIONS —

The data analytics in crop production enables the some business organization to predict agricultural yields and better crop quality.

2. JOBS-TO-BE-DONE / PROBLEMS —

States with crop production along with season and season with average productions.

Data report should be created to reduce the loss of the crop and earn more profit in this field.

9. PROBLEM ROOT CAUSE `RC`

The insects on the plants can spread the disease . Extreme weather events and various disease on the plants can lead to reducing the crop productivity

7. BEHAVIOUR `BE`

To develop a web application which includes all the factors of their expenses

Get a feedback and user's experience from customers who are using that application .

3. TRIGGERS `TR`

Observing by other farmers who are

producing the more crops
By supplying the water at an optimal level it will reduce the loss and gains the crop production.

10. YOUR SOLUTION `SL`

Both of them customer and farmers to know about the crop yield details in the given dataset and it can shows the clear visualization and they can get ; som ideas about the crop production

8. CHANNELS of BEHAVIOUR `CH`

• Trying to use pesticides and fertilizers that increase

• Farmers need to check the crop yield frequently

4. EMOTIONS: BEFORE / AFTER `EM`

• They don't have the knowledge of the crop sowing before using the visualization.

• They had some knowledge after saw the visualization charts.

3.5 SOLUTION ARCHITECTURE
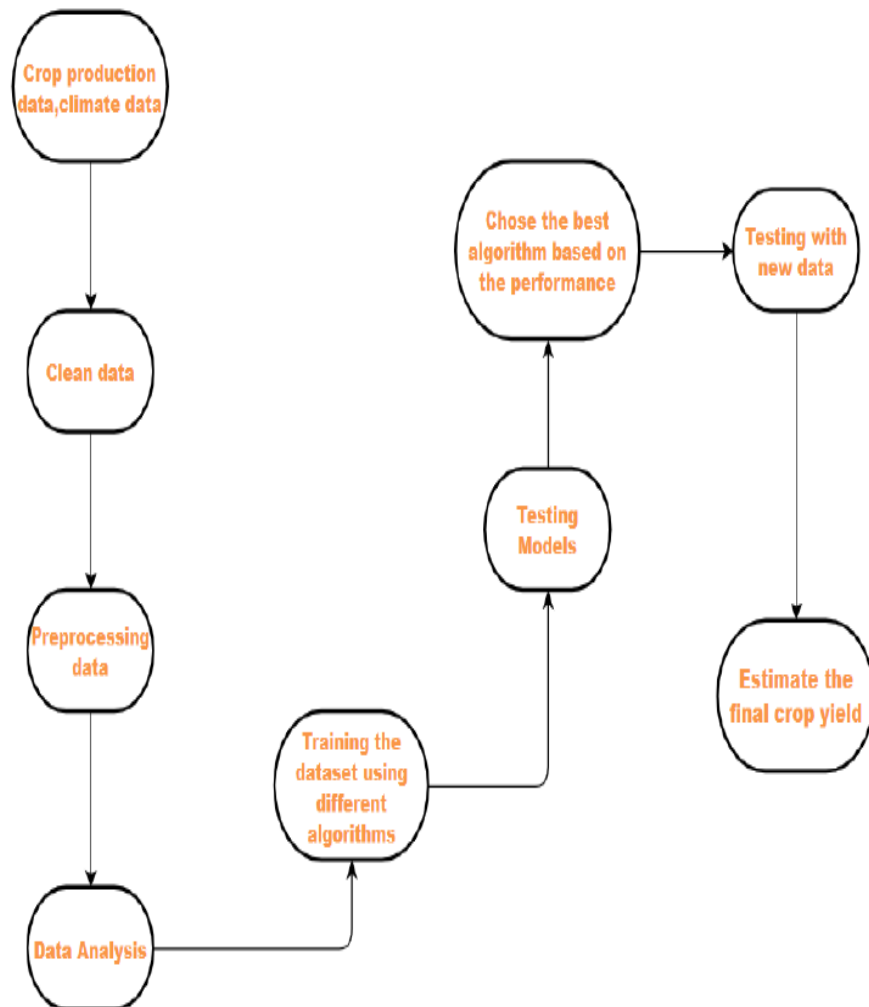
## 4 REQUIREMENT ANALYSIS:

4.1 FUNCTIONAL REQUIREMENTS:

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Data requirement	The user data is required to analyse the historical data and data visualization on the crops yield.
FR-4	Data Storage	A private cloud storage is required with enhanced security level and with data encryption facility.
FR-5	Data Analysis	Analyse the datasets and choose only the required columns and clean the dataset by removing duplicates records.
FR-6	Estimation	A clear visualization chart and appropriate techniques to improve the accuracy of prediction.

4.2 NON-FUNCTIONAL REQUIREMENTS:

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The data reports which are provided are based on historical data and these recommendations will help the farmers in choosing the right time and season to cultivate and harvest crops and to gain more profit.
NFR-2	Security	The user data and the crop information is protected with IBM Cognos.
NFR-3	Reliability	The dashboard and the visualization charts are made simple so that the data reports are simple to interpret.Recovery speed is also a important factor.
NFR-4	Performance	The System is fast enough to handle large datasets with advanced visualization techniques.
NFR-5	Availability	The dashboard is user friendly and it is designed in such a way that it is easily accessible and can be accessed on smartphones,laptops,desktops with an active internet connection.
NFR-6	Scalability	The dashboard and the visualization charts should be able to provide data reports for larger datasets with accuracy.

5 PROJECT DESIGN:

5.1 DATA FLOW DIAGRAMS:
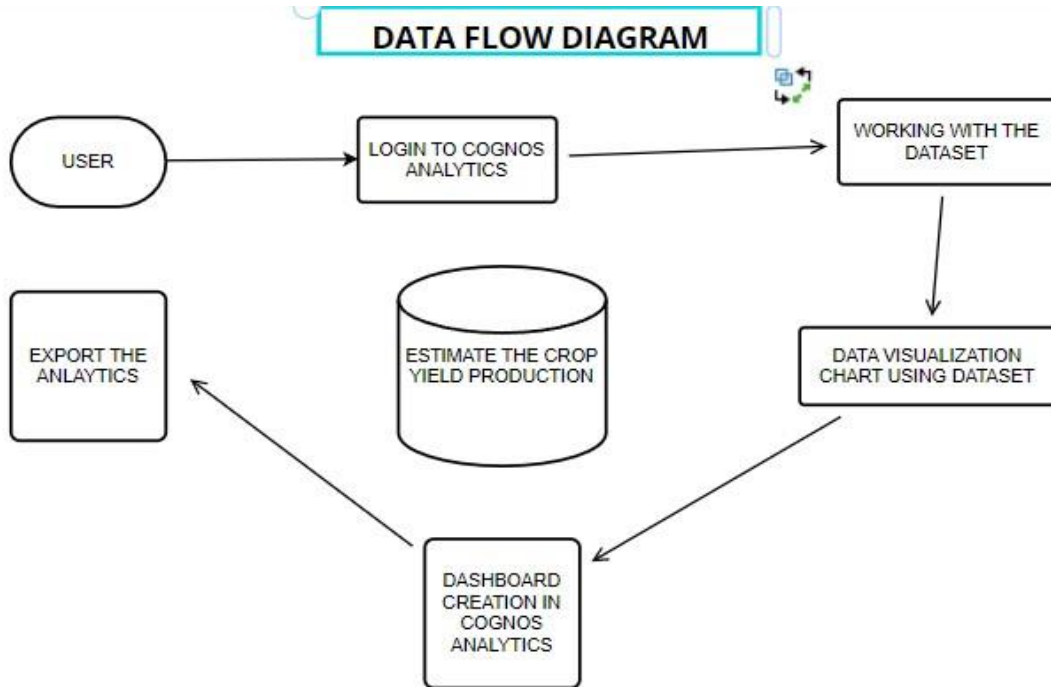
User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	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	As a user I can access the dashboard to view the required information about the crop yield.	User Profile can be edited.	High	Sprint -2
Customer (Web user)		USN-7	As a user I can register/login the web application and can access	I can login to the website and access my account	Medium	Sprint -1

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
			the dashboard and view the required information			
Customer Care Executive		USN-8	Provides support regarding the application and communicates with the user if the user has any queries.	Authentication is provided for accounts and access is denied for accounts without permission.	High	Sprint-2
Administrator		USN-9	As a user I can update the datasets and request to improve the dashboard.		High	Sprint-1

5.2 SOLUTION AND TECHNICAL

ARCHITECTURE:TECHNICAL ARCHITECTURE:

DATASET COLLECTION

UPLOAD DATASET

TRAIN AND TEST DATASET

DATA PREPROCESSING

DASHBOARD CREATION

DATA VISUALIZATION

EXPLORE DATA
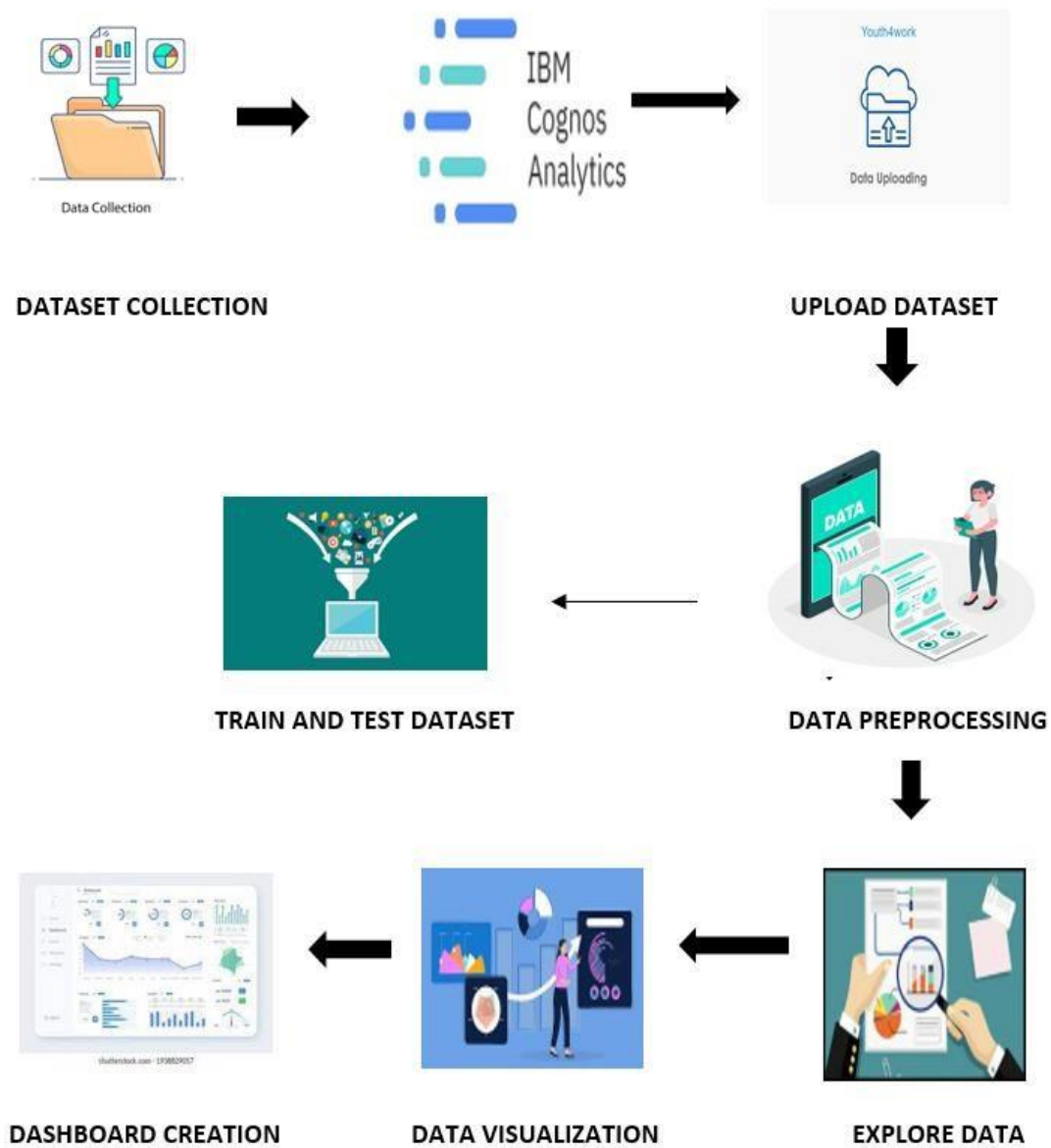
Table-1 :

Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application Web UI	HTML, CSS
2.	Weather data	Past Weather data about the surrounding to predict the yield.	IBM Watson service
3.	Crop yield data	Data about the amount of yield produced by crops on that particular area.	IBM Watson Assistant
4.	Cloud Database	Uploading of dataset and dataset storage	IBM DB2.
5.	File Storage	File storage requirements	Local Filesystem
6.	External API-1	To provide current and forecasted conditions, seasonal and sub-seasonal forecasts, lifestyle indices, severe weather and historical weather data for analysis.	IBM Weather API.

Table-2:

Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	No open-source frameworks used	Python
2.	Security Implementations	Data is stored in cloud with high security	SHA-256, Encryptions, IAM Controls, OWASP etc.
3.	Scalable Architecture	Large datasets can also be uploaded for visualization and many users can use the model for visualization at same time.	IBM Cloud, IBM Cognos
4.	Availability	Soil recommendations, weather forecasts and contaminated zones recommendations are available in application.	IBM Watson Assistant
5.	Performance	Large requests can be handled at the same time.	Cognos analytics.

5.3 CUSTOMER JOURNEY

Document an existing experience

Narrow your focus to a specific scenario or process within an existing product or service. In the **Steps** row, document the step-by-step process someone typically experiences, then add detail to each of the other rows.

> **TIP**
> As you add steps to the experience, move each these "Five Es" the left or right depending on the scenario you are documenting.

SCENARIO: Browsing, booking, attending, and rating a local city tour	**Entice** — How does someone initially become aware of this process?	**Enter** — What do people experience as they begin the process?	**Engage** — In the core moments in the process, what happens?	**Exit** — What do people typically experience as the process finishes?	**Extend** — What happens after the experience is over?
Steps — What does the person (or group) typically experience?	The user has to login to the website; On boarding; The user has registering for the new user; The user id and psw are must required to login	Lot of confusion to how to use the dashboard; different types of charts; Seeing a lot of information displayed	Comparison of different charts; Provide me with the user friendly; Getting an idea on the factors which leads to the better yield	In order to determine the satisfaction of users,feedback forms are provided	A real time view
Interactions — What interactions do they have at each step along the way? • People: Who do they see or talk to? • Places: Where are they? • Things: What digital touchpoints or physical objects would they use?	Interacting with the online helpers for farmers; Interacting with the agricultural experts	Interacting with the other farmers that how did they started using it; Login using your registered id and psw	They are able to use the insights for better decisions of the crop yield; Analysis can be performed and profit can be made by users; Farmers are surveyed and the results are analyzed	Aware of farming knowledge; Digitized farming experience	Any user can use this website; Concept for time saving
Goals & motivations — At each step, what is a person's primary goal or motivation? ("Help me..." or "Help me avoid...")	Help me to estimate the crop yield; Help me to get a healthy crop; Assist me for prediction	Help me for better customer experience; assist me for user friendly; Help me to use and access the necessary sources	Help me to make new ideas for creating the dashboard; Help me to track the profit analysis; Help me to create an interactive dashboard	Help me with sharing thoughts	Help me for high productivity; Help me to avoid user inconvenience
Positive moments — What steps does a typical person find enjoyable, productive, fun, motivating, delightful, or exciting?	Excited about using the new technology in farming; Eager to see all visible outcome in the field	It is interactive for using; Feeling that it is very easy to proceed	It is very useful and informative; It provides the crop protection details	It is an update feature of farming; proper irrigation prediction	profit will be increased
Negative moments — What steps does a typical person find frustrating, confusing, angering, costly, or time-consuming?	The user has fear of loss because it is new technology; Unable to access through the network	Feeling bad to analyze; unable to use it faster	Confusion in finding the right decision; Problems in understanding the visualization	Consumes a loyalty of customer	Engaging to collecting good cultivation studies
Areas of opportunity — How might we make each step better? What ideas do we have? What have others suggested?	Creating awareness among the people about the precision agriculture; Increasing the advertisment	The security can be increased; Providing the step by step guideliness	updating the database; updating the features and policies	user interface can be improved	Improve crop production; Improve the quality of crop

16

6.1 SPRINT PLANNING AND ESTIMATION:

(I) PROJECT TRACKER:

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

(II)VELOCITY:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \text{Sprint Duration} / \text{Velocity} = 20 / 20 = 1$$

(I) BURDOWN CHART:

     A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

BURNDOWN CHART
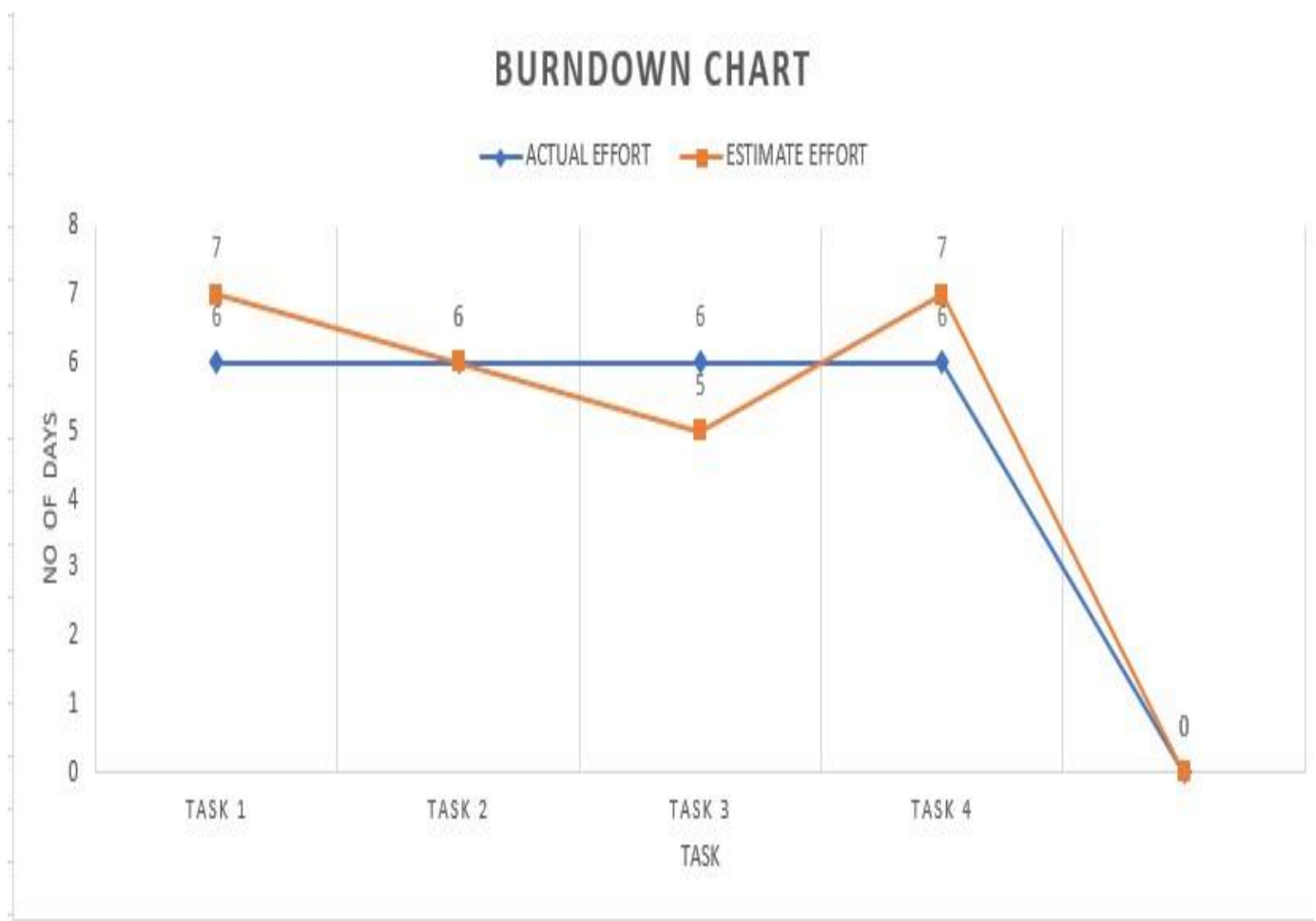
6.2 SPRINT DELIVERY SCHEDULE:

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	3	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R
		USN-2	As a user, I can register for the application through Gmail	2	Medium	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R
		USN-3	As a user, I will receive confirmation email once I have registered for the application	3	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R
	Login	USN-4	As a user, I can log into the application by entering email & password	4	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R
	Working with the Dataset	USN-5	To work on the given dataset, Understand the Dataset.	4	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R
		USN-6	Load the dataset to Cloud platform then Build the required Visualizations.	4	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-2	Data Visualization Charts	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	Nadeeshwaran.A Sanjay R
			Build a Visualization to showcase the Yearly usage of Area in Crop Production.	4	Medium	Sharuk B Sugumar R
			Build a visualization to show case top 10 States in Crop Yield Production by Area.	4	Medium	Nadeeshwaran.A Sugumar R
			Build the required Visualization to showcase the Crop Production by State.	4	Medium	Nadeeshwaran.A
			Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation.	4	Medium	Sharuk B Sanjay R
Sprint-3	Creating The dashboard	USN-8	Create the Dashboard by using the created visualizations.	20	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R
Sprint-4	Export The Analytics	USN-9	Export the created Dashboard	20	High	Nadeeshwaran.A Sanjay R Sharuk B Sugumar R

6.3 REPORTS FROM JIRA:

(I) BEFORE START OF THE SPRINT:
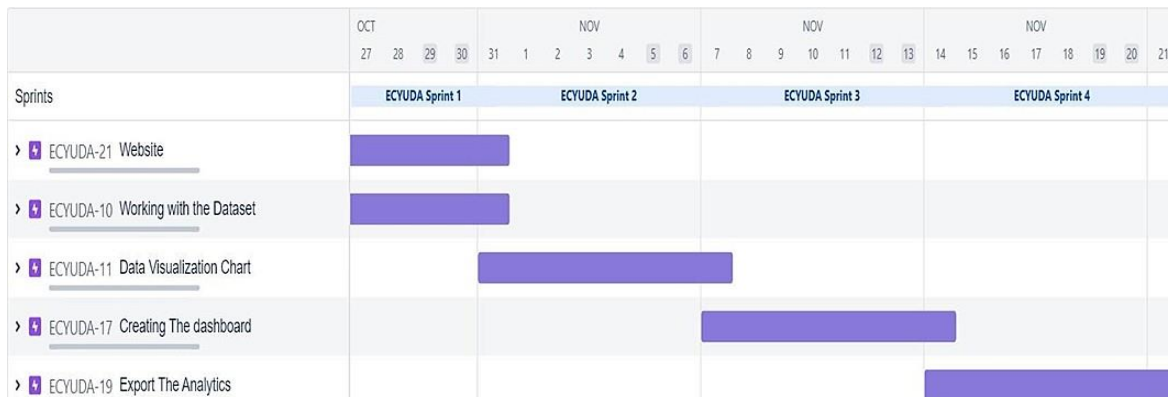
(II)   SPRINT 2
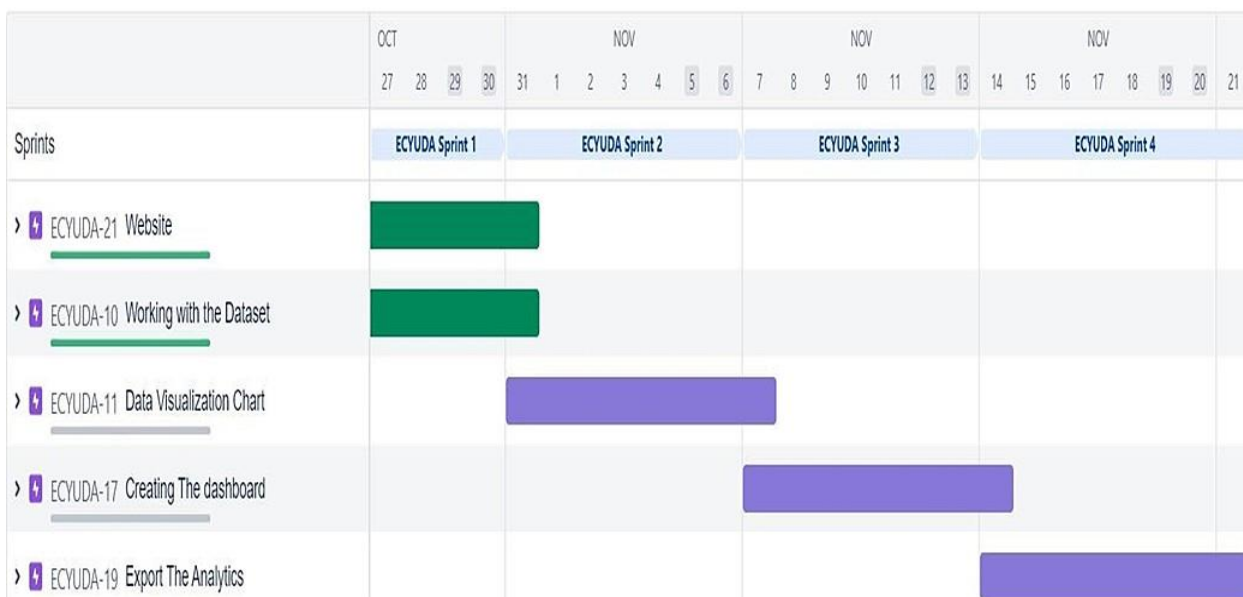
(III) SPRINT 3:
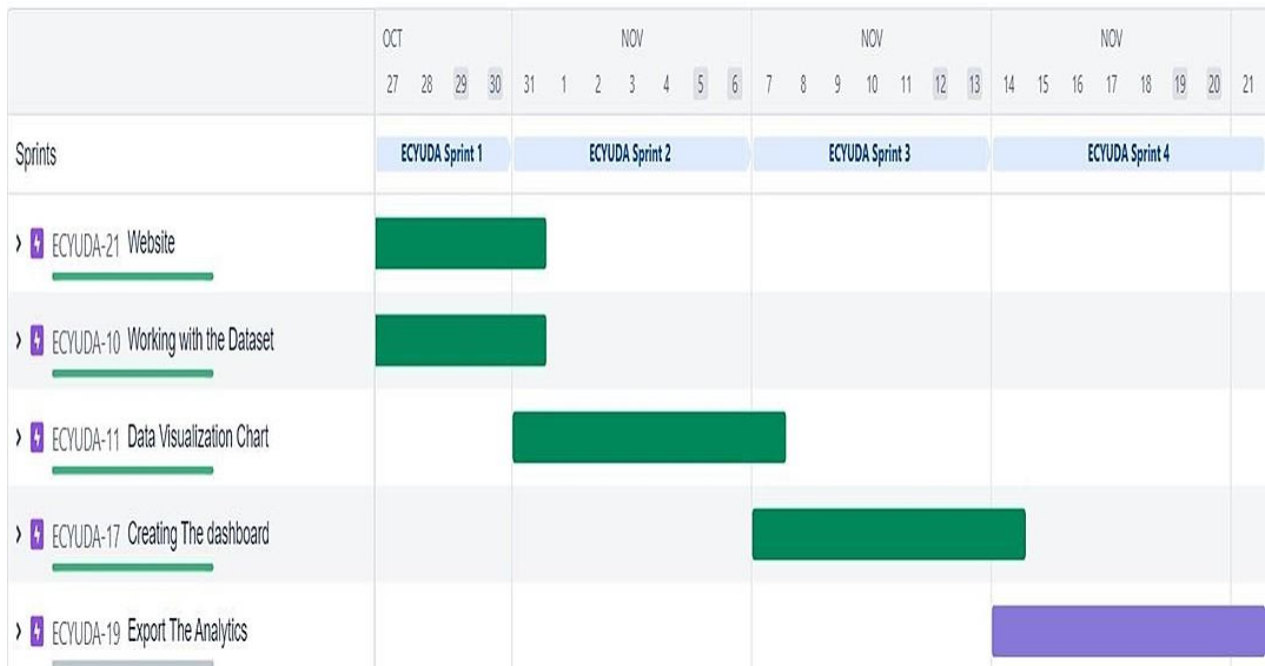
(IV) SPRINT 4:
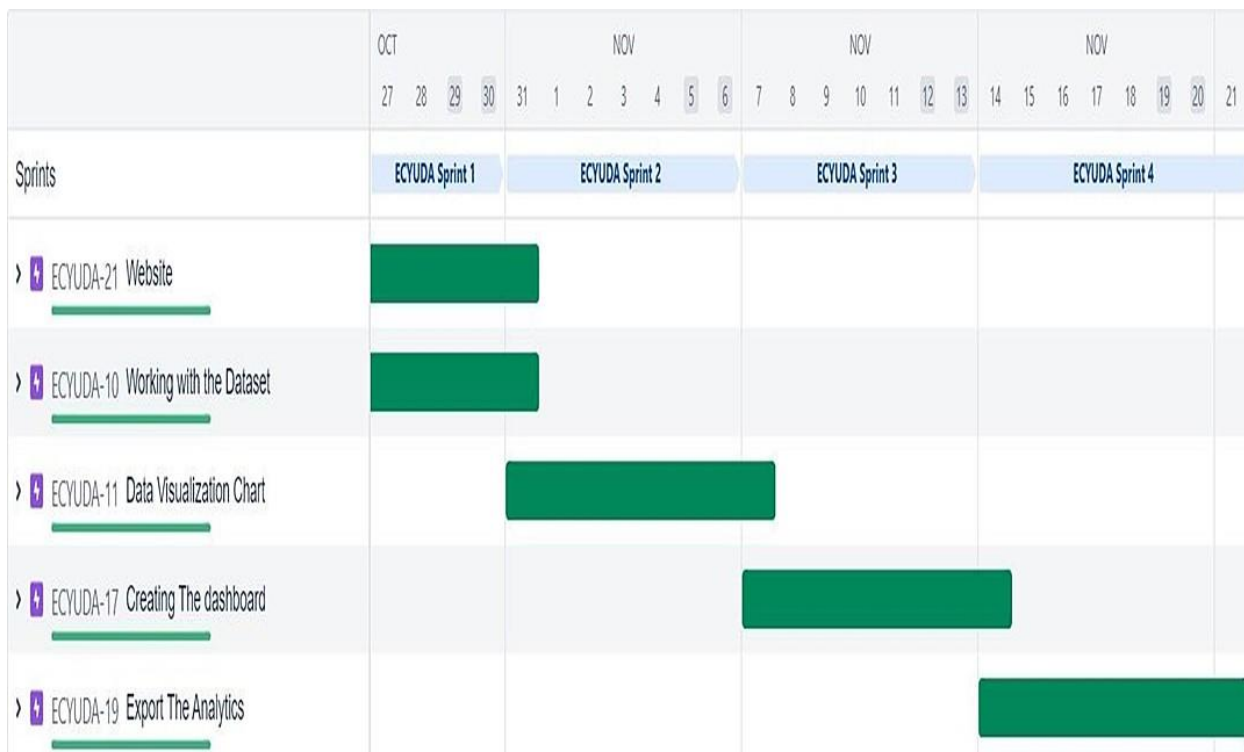
7. CODING & SOLUTIONING:

7.1 FEATURE 1:

DASHBOARD DESIGN:

The dashboard is created using IBM cognos tool which efficiently visualises a given data. The design is incorporated along with login page and provides excellent insights on various data regarding crops.

```html
<!DOCTYPE html>
<html>
<head>
        <title>Crop Production</title>
        <style>
                button{
                        width: 308px;
        height: 39px;
        background-color: lawngreen;
        border-radius: 10px;
                }
        </style>
</head>
<body style="background-image: url(b1.jpg)">
  <center><h1>Crop Production</h1></center>

 <h3 style="margin-top: 62px;margin-left: 8px;"><button><a href="Season_Av.html" style="text-decoration:
none;">Season with Average Production</a></button> </h3>
  <h3 style="margin-top: 65px; margin-left: 225px;
"><button><a href="With_years.html" style="text-decoration: none;">SeasonWith year Usage of Areaa
Production</a> </button></h3>
  <h3 style="
  margin-top: 70px;
  margin-left: 498px;
"><button><a href="Top10.html" style="text-decoration: none;">SeasonTop 10 State with Most
Area</a></button></h3>
  <h3 style="
  margin-top: 72px;
  margin-left: 700px;
"><button><a href="StateWith.html" style="text-decoration: none;">SeasonState with Crop
Production</a></button></h3>
   <h3 style="
  margin-top: 78px;
  margin-left: 1000px;
"><button><a href="State.html" style="text-decoration: none;">SeasonState with Crop Production Along With
Season</a></button></h3>
</body>
</html>
```
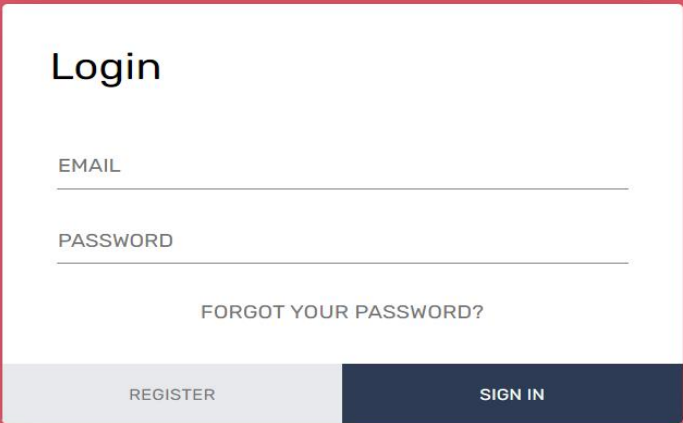
7.2 FEATURE 1:

```html
<!DOCTYPE html>
<html>
<head>
        <title></title>
</head>
<body  style="background-image: url(cr3.jpg)">
   <center><h1 style="
   color: white;
">Season With Average Productions </h1></center>
   <center><img src="1.png" style="height: 445px;width: 800px;margin-top: 28px;"></center>
</body>
</html>
```

FEATURE 2:

```html
<!DOCTYPE html>
<html>
<head>
      <title></title>
</head>
<body style="background-image: url(cr3.jpg)">
      <center><h1 style="
   color: white;
">With Year sage of Area and Production </h1></center>
      <center><img src="2.png" style="height: 445px;width: 800px;margin-top: 28px;"></center>
  <img src="">
</body>
</html>
```

FEATURE 3:

```html
<!DOCTYPE html>
<html>
<head>
        <title></title>
</head>
<body style="background-image: url(cr3.jpg)">
   <center><h1 style="
   color: white;
">Top 10 States With Most Area</h1></center>
   <center><img src="3.png" style="height: 445px;width: 800px;margin-top: 28px;"></center>
   <img src="">
</body>
</html>
```

FEATURE 4

```
<!DOCTYPE html>
<html>
<head>
        <title></title>
</head>
<body style="background-image: url(cr3.jpg)">
   <center><h1 style="
   color: white;
">State With Crop Production Along with Season</h1></center>
   <center><img src="5.png" style="height: 445px;width: 800px;margin-top: 28px;"></center>
   <img src="">
</body>
</html>
```

FEATURE 5

```
<!DOCTYPE html>
<html>
<head>
        <title></title>
</head>
<body style="background-image: url(cr3.jpg)">
   <center><h1 style="
   color: white;
">State With Crop Production </h1></center>
   <center><img src="4.png" style="height: 445px;width: 800px;margin-top: 28px;"></center>
   <img src="">
</body>
</html>
```

8 RESULT AND SCREENSHOT

LOGIN

Login
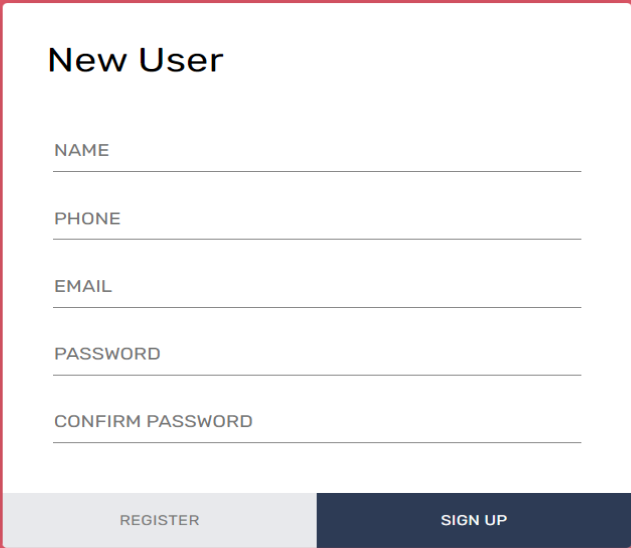
EMAIL

PASSWORD

FORGOT YOUR PASSWORD?

REGISTER   SIGN IN
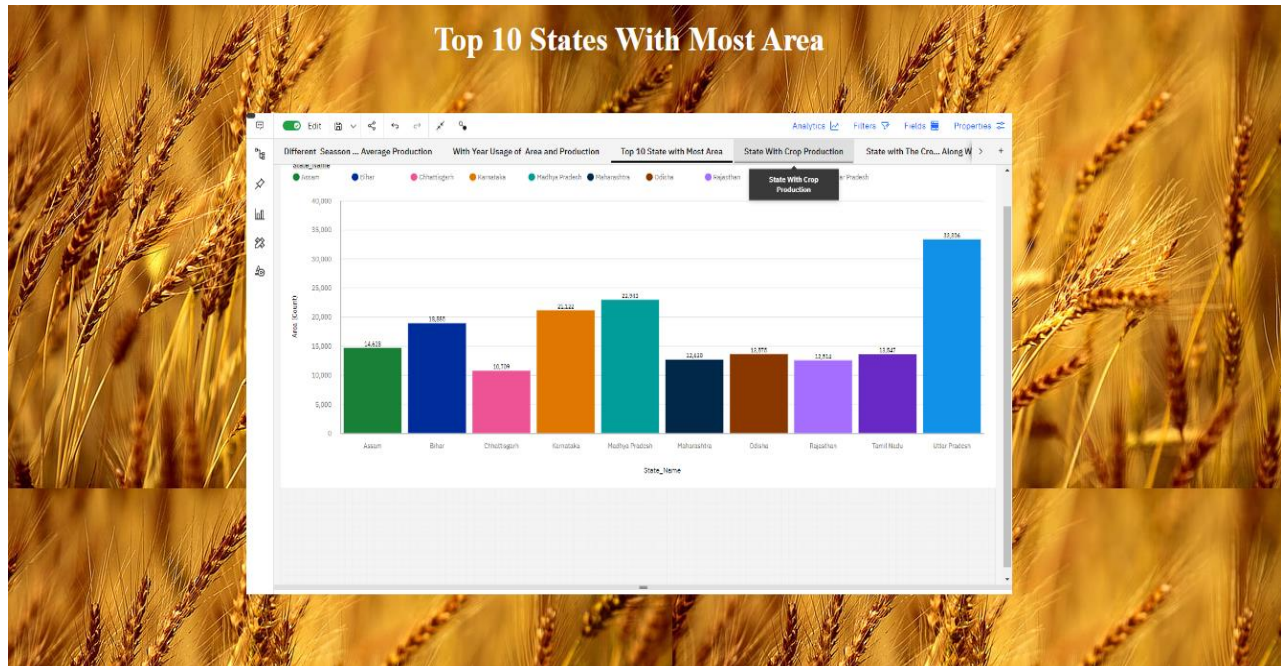
REGISTERATION

New User

NAME

PHONE

EMAIL

PASSWORD

CONFIRM PASSWORD

REGISTER   SIGN UP

DASHBOARD

Crop Production

- Season with Average Production
- SeasonWith year Usage of Areaa Production
- SeasonTop 10 State with Most Area
- SeasonState with Crop Production
- SeasonState with Crop Production Along With Season

SEASON WITH AVERAGE PRODUCTION

Season With Average Productions
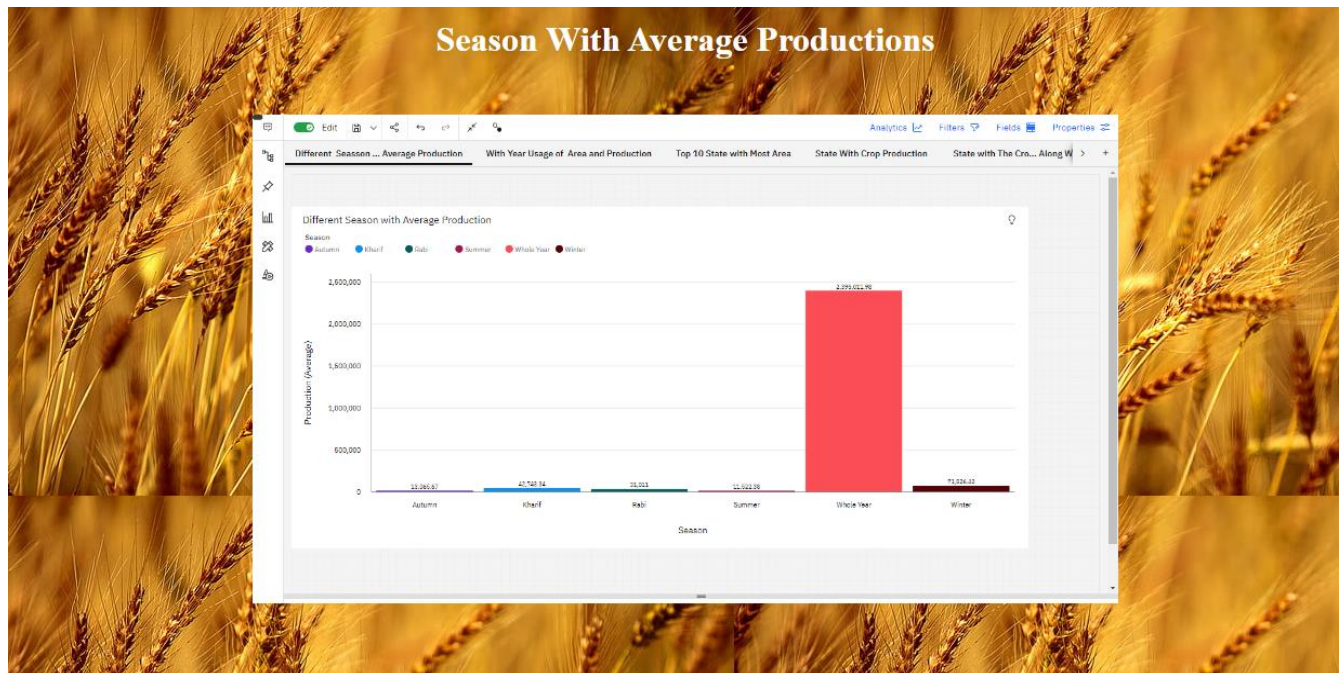
TOP 10 STATE WITH MOST AREA

STATE WITH CROP PRODUCTION

28
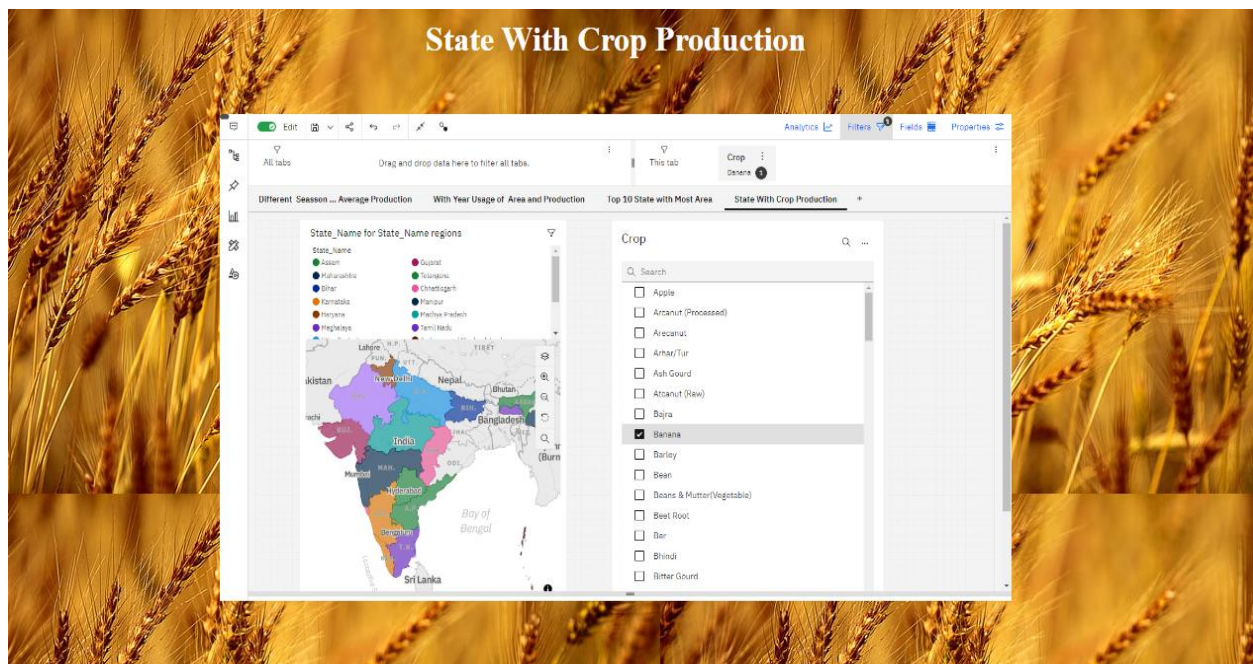
WITH YEAR STAGE OF AREA AND PRODUCTION
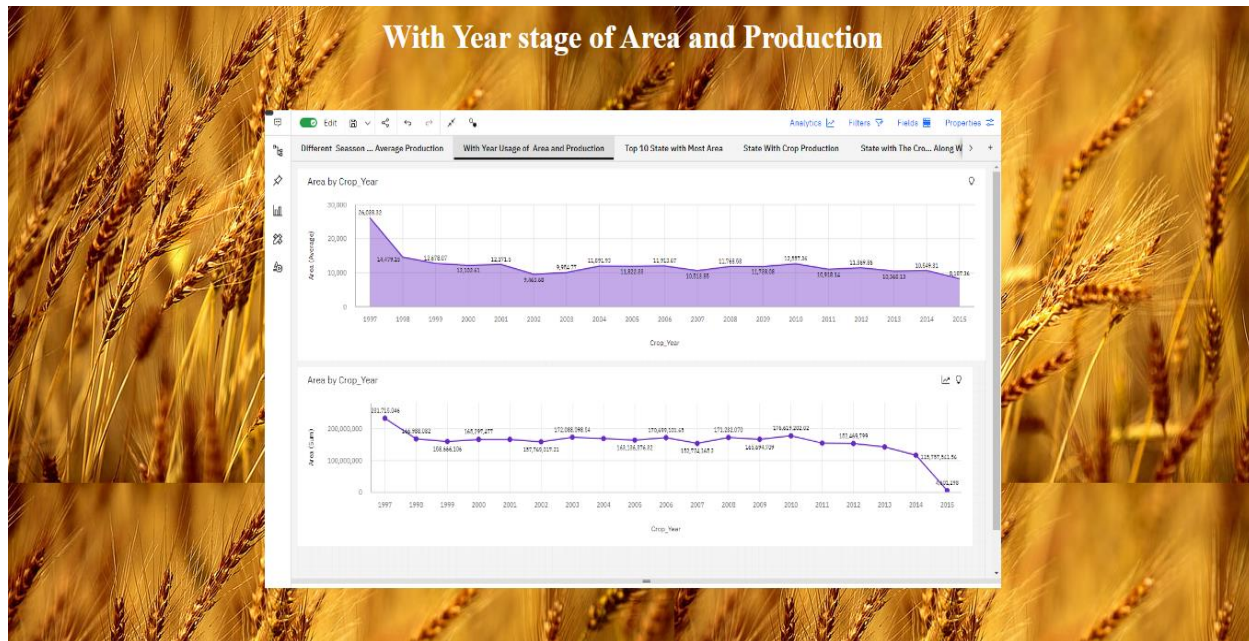
STATE WITH CROP PRODUCTION ALONG WITH SEASON
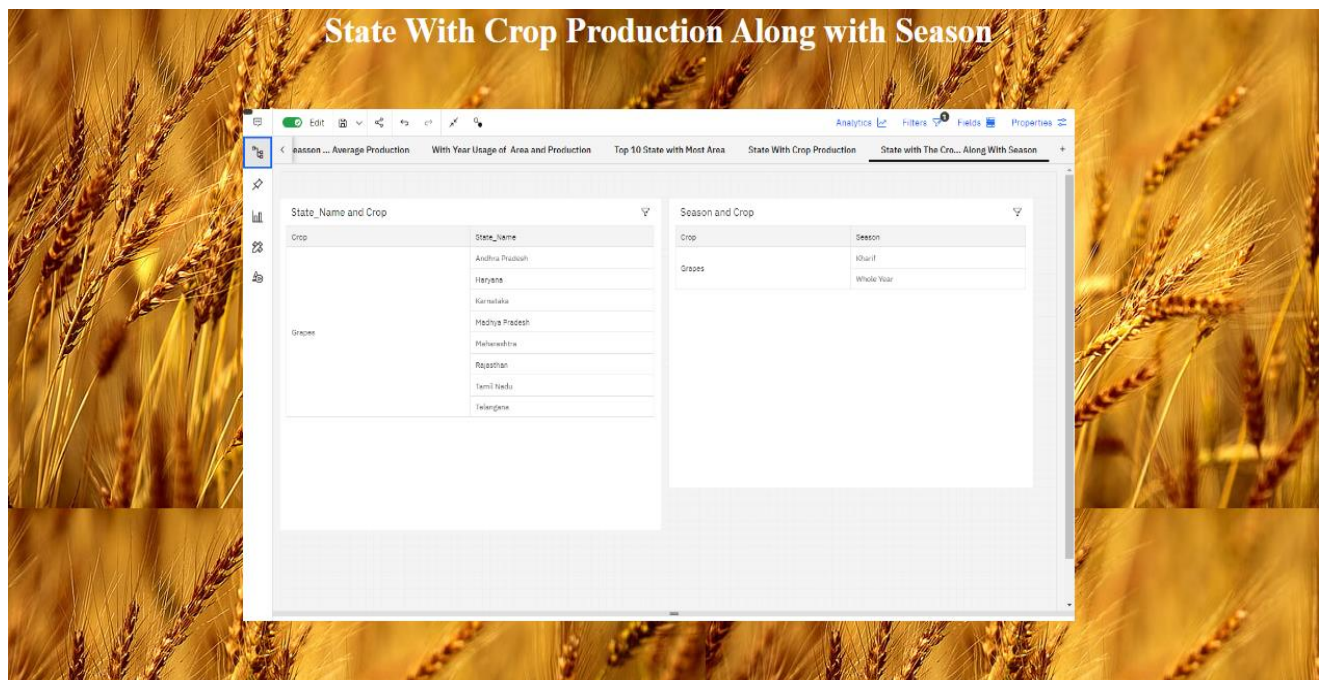
ADVANTAGES & DISADVANTAGES:

ADVANTAGES:

The advantage of this Crop yield estimation it is relatively less time-consuming and inexpensive. We can able to know the average productions of the crop, the amount of crop produced in different year and in different districts and in different area. And it is also used by farmers to make decisions about when to plant and harvest crops based on soil moisture content and weather conditions.

DISADVANTAGES:

The disadvantage of the system is number of data used for the estimation are less. This method is highly subjective, as the information are collected from the farmers' knowledge and experience.

CONCLUSION:

The proposed "ESTIMATE THE CROP YIELD USING DATA ANALYTICS" is used to predict the crop yield using the attributes such as State_Name,District_Name,Crop_Year,Season,Crop,Area and Production.The proposed model is build with IBM Cognos Watson.As a result of penetration of technology into agricultural field,there is a marginal improvement in the productivity.The innovation have led to new concepts like digital agriculture, smart farming, precision agriculture etc.It has been observed that analysis has been done on crop, hidden pattern discovery using dataset related to season, area, production data.The activities of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction etc. In this survey, the specific activity, crop yield prediction has been surveyed and the major trends have been identified. It can be concluded that the research in the field of agriculture with

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

FUTURE SCOPE:

The dashboard creation,visualization have taken lots of procedures and steps.The aim of the future work is to analyze the target attribute by reducing the number of procedures and steps.To improve the accuracy of the analysis algorithm selection procedure need to be optimized.

As a future work, the results of the analysis can be improved, using the large number of crop datasets and more weather parameters. This can be also implemented in machine learning model to build in a strong yield prediction model and analysis of all the crops with different climatic conditions and different areas.

9 APPENDIX

9.1GITHUB

GITHUB:
https://github.com/IBM-EPBL/IBM-Project-29452-1660125642