ESTIMATE THE CROP YIELD USING DATA ANALYTICS

SUBMITTED BY
Harini S U
Nandhini T
Surya L
Maria A Rexana

TEAM ID - PNT2022TMID08893

DEPARTMENT OF

INFORMATION TECHNOLOGY

DR. MAHALINGAM COLLEGE OF ENGINEERING &TECHNOLOGY

Institution Mentor: J. Ramprasath

Industrial Mentor: Srikanth, Mahidhar, Mohammed Azhar Uddin

TABLE OF CONTENTS

1. INTRODUCTION

1.1 PROJECT REVIEW

Predicting crop yields is one of the most difficult problems in agriculture. It is crucial to decision-making at the international, regional, and local levels. Agricultural, soil, climatic, environmental, and other characteristics are used to predict crop yield.

It has become a challenging task to achieve desired targets in Agriculture 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 years and data analytics is one such trend that has penetrated into the agriculture field. To reach desired crop yield goals has become a difficult undertaking in agriculture.

Numerous elements that directly affect the yield and productivity of the crops must be taken into account. One of the crucial aspects of agricultural techniques is the forecast of crop production. Before planting seeds in their fields, farmers require knowledge about crop yield in order to increase agricultural output.

In recent years, the use of technology in agriculture has increased, andone such development is the use of data analytics.

Thus, a project that would suit the needs of a farmer and at least help them over a borderline to understand and predict or estimate thecrop yield was the main aim and was brought to life.

1.2 PURPOSE

Estimation of Crop yield has become the need of the hour and one easy tool/method that can be used is Data Analytics. The term "data analytics" describes the methods used to analyze data in order toincrease productivity and financial gain. In order to examine different behavioral patterns, data is extracted from a variety of sources, cleaned up, and classified. The methods and resources employed change depending on the group or person.

The purpose behind this project is to understand the variation incrop yield due to various parameters that can be natural or non-natural.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

With the changing of climate, agriculture faces increasing problems with extreme weather events leading to considerable yield losses of crops. Most often, crop plants are sensitiveto stresses since they were mostly selected for high yield, and not for stress tolerance. The four most important factors that influence crop yield are soil fertility, availability of water, climate, and diseases or pests. With such varying parameters, to understand or estimate the patterns with no technological involvements is very difficult. Thus, a solution that is technological and cater to the alterations and provide the predicted solution in a form that can be easily understood by end customers is essential.

2.2 REFERENCES

(i) Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture

Bini Darwin, Pamela, Dharmaraj (2021) proposed a "Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture". The methods which made use of conventional deep learning techniques have provided an average accuracy of 92.51%. This paper elucidates the diverse automation approaches for crop yield detection techniques with virtual analysis and classifier approaches. This work highlights the machine vision and deep learning models which need to be explored for improving automated precision farming expressly during this pandemic.

(ii) Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt

Aleksandra Wolanin, Gonzalo Mateo-García, You Liangzhi and Luis Guanter (2020) presented "Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt". Forecasting crop yields is becoming increasingly important under the current context in which food security needs to be ensured despite the challenges. Machine learning (ML) techniques, and deep learning (DL) methods been used. To benefit from the increased predictive performance of DL methods while maintaining the ability to interpret how the models achieve their results. The proposed methodology can be used for other crops and regions in order to facilitate application of DL models in agriculture.

(iii) Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods

ElisaKamir, FrançoisWaldner, ZviHochman (2020) designed “Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods”. In this context, using Machine-learning methods by applying it to climate and satellite image time series we can achieve reliable crop yield monitoring across years at both the pixel and the country scale. The result estimates meet the accuracy requirements for mapping the yield gap and identifying yield gap hotspots which could be targeted for further work by agricultural researchers and advisers.

(iv) Crop Production and Crop Diversity in France: A Spatial Analysis

Hermann Pythagore PierreDonfoe, Aleksandra, CécileDétang-Dessendre,EliseMaigne (2020) proposed a “Crop Production and Crop Diversity in France: A Spatial Analysis”. This paper aims to provide empirical evidence of the effect of crop diversity on crop production and spillover effect. Based on the estimation of production functions with spatial concerns on an original and rich dataset, results of the study suggest that crop diversity has a positive and significant effect on crop production. Its marginal contribution is substantial when rainfall is low in the agroecosystem. Furthermore, spatial dependence is a major issue and could be explained by topographic, climatic and agronomic constraints.

(v) Remote sensing of crop production in China by production efficiency models: models comparisons, estimates and uncertainties

FuluTao, MasayukiYokozawa, ZhaoZhang, YinlongXu, YousayHayash (2021) presented a “Remote sensing of crop production in China by production

efficiency models: models comparisons, estimates and uncertainties". By combing with remote sensing data, both the two kinds of production efficiency models are potentially useful for NPP monitoring or yield forecasting at regional or national scale. They can be used to examine the effect of climate variability on NPP, and consequently important for the researches on carbon cycle and food security. Models parameters calibration and models validation for various biomes are crucial for applications. Our study improved the soil-moisture simulation.

2.3 PROBLEM STATEMENT DEFINITION

The following instances define the problem of notice. Ram is afarmer who needs a way to understand and predict climaticconditionsbecause he can decide on the safety measures to be followed with regards to the field setup.

Raj is a farmer who needs a way to decide what to grow andwhen to grow because he is uncertain of his environmental conditions.Ranil is a grocer and crop distributor who needs to know the overall crop yield turnoverbecause he has to understand his monetary turnover for the year.

Thus, a solution that can cater to all the needs put forth is being formulated.

3. IDEATION AND PROPOSED SOLUTION

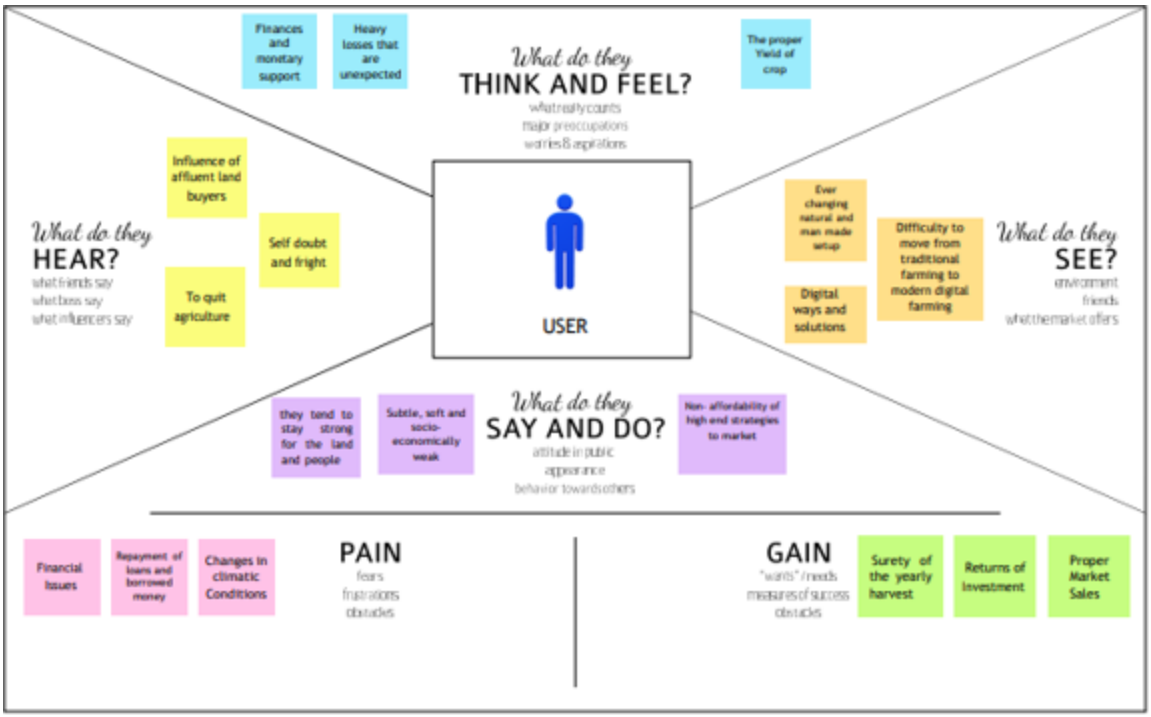
3.1 EMPATHY MAP CANVAS

A team of four members sat together to discuss on and empathizeabout the problem that people have been facing with regards to understanding and predicting the yield of crops.

As a part of what the customers or target audience felt, a conclusion was made such that they were concerned about elements like finances, monetary support, heavy unexpected losses, proper yield of crop and certain other unpredictable factors.

Under the concepts of what they see and on the basis of environment, friends and what the market offers, digital solutions for ever changing natural setup with the difficulty of moving from traditional to modern farming was sought.

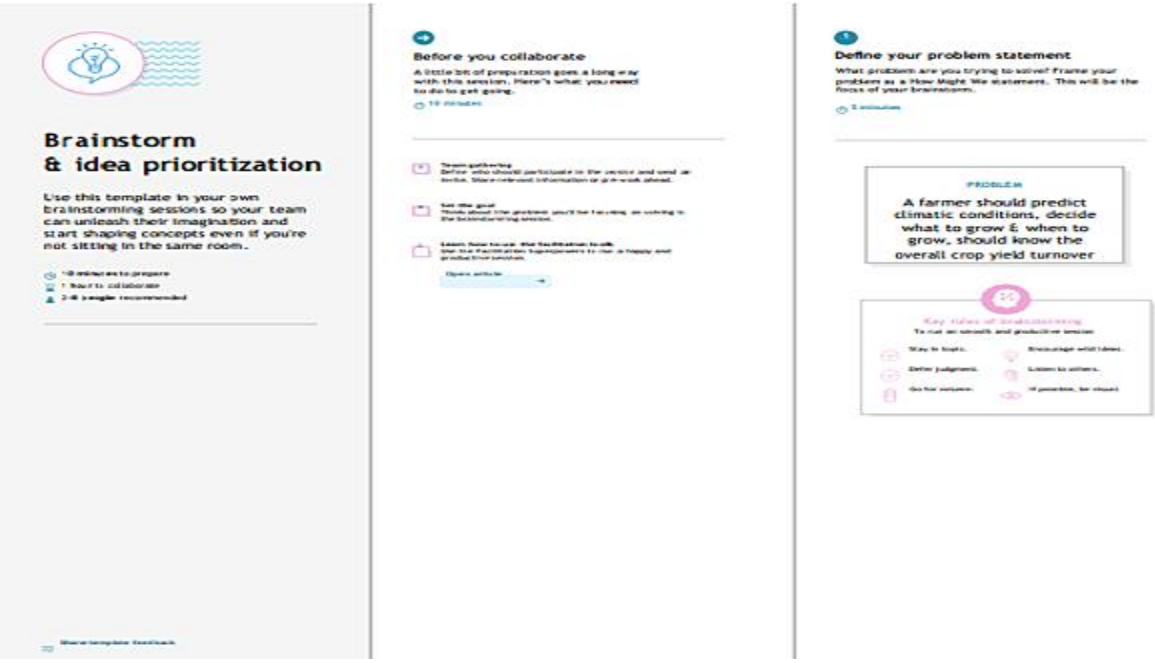
The specifics of pain and gain along with speculations of the influence of affluent land buyers, self-doubt and fright added with this comes the fear to quit agriculture under the section of what they hear.



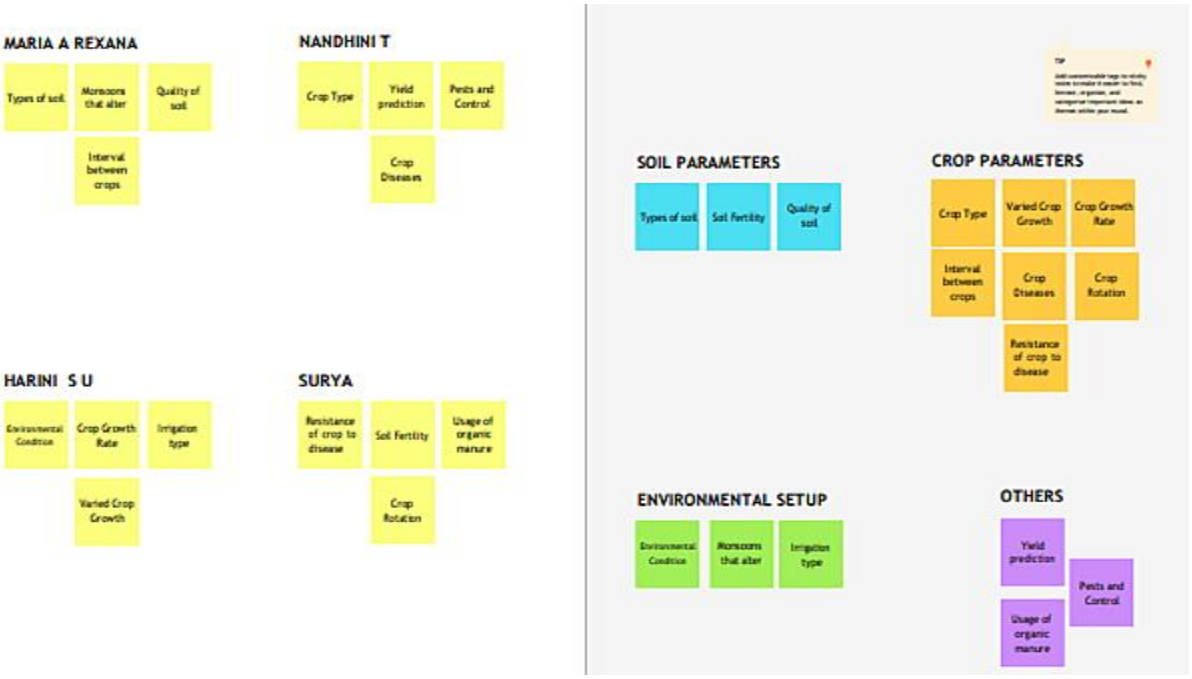
3.2 IDEATION AND BRAINSTROMING

Ideation and the process of brainstorming was done by initially tracing the problem and defining it. This was followed by pushing inindividual ideas about the problem and then grouping it in together under common groundsand making a graph out of the priority provided.

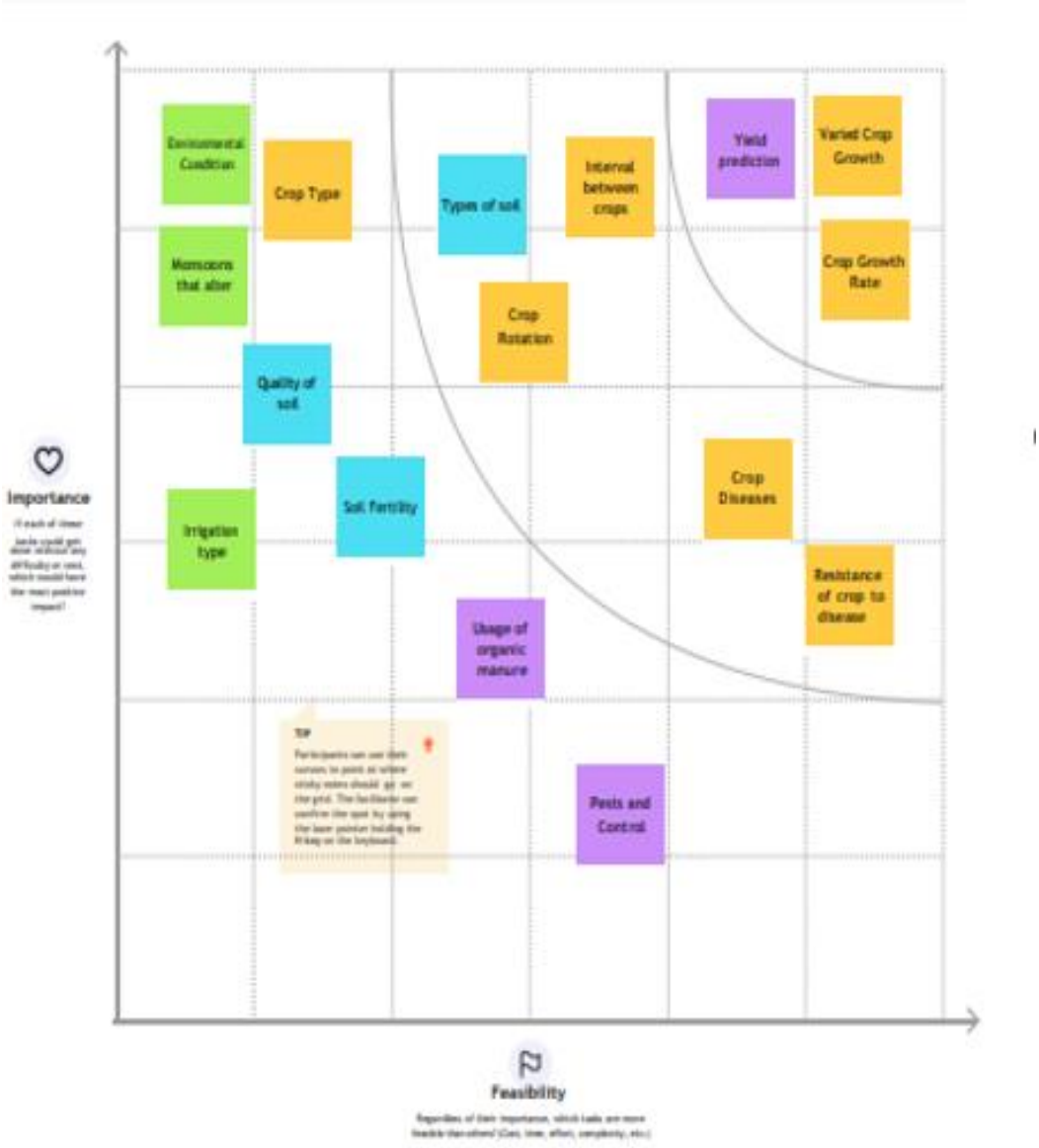
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization

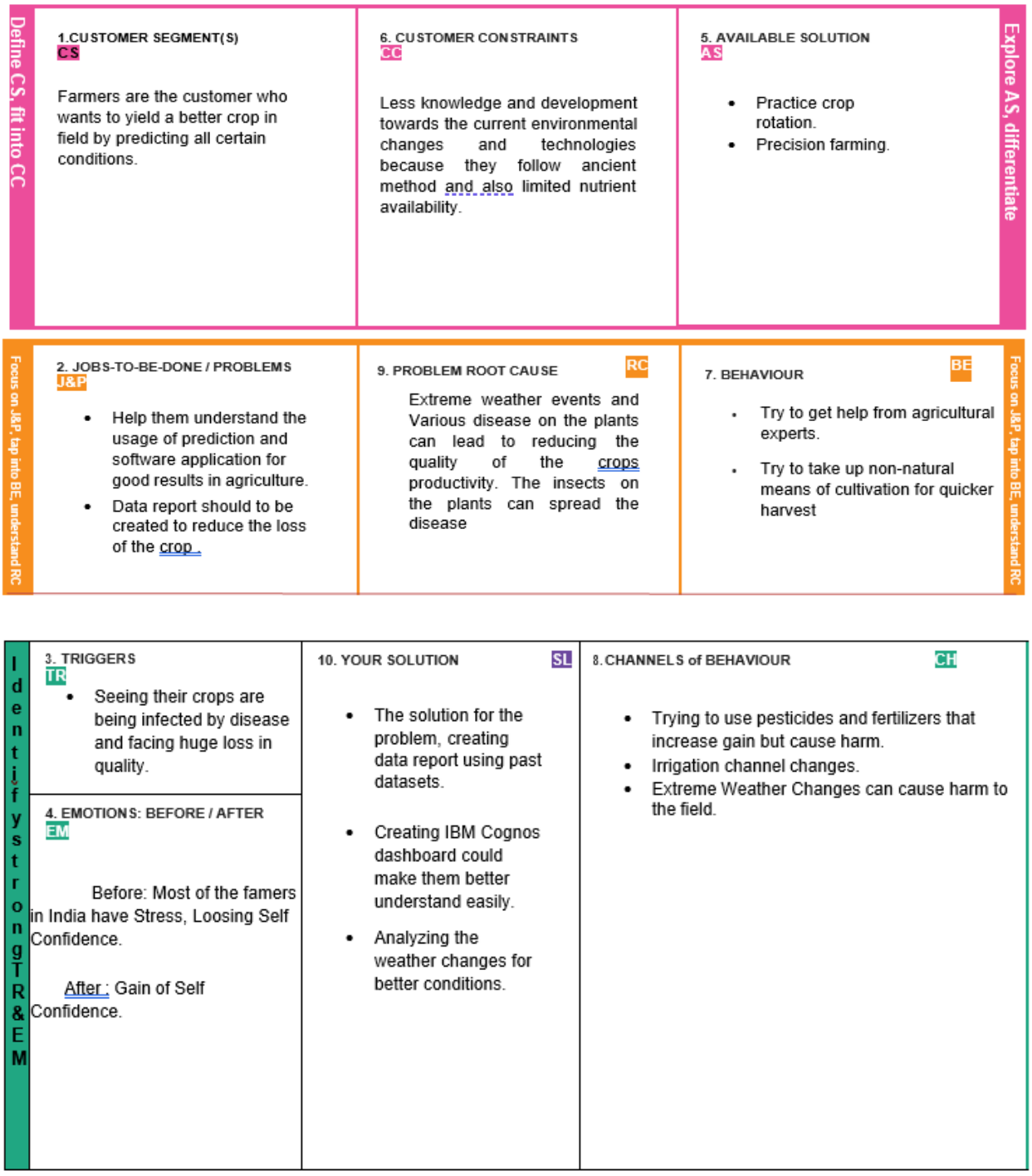


3.3 PROPOSED SOLUTION

S.NO.	PARAMETER	DESCRIPTION
1.	Problem Statement (Problem to be solved)	The farmers are facing difficulties in analyzing the demand in market and soil quality analysisto achieve high crop yield through technology. A farmer shouldpredict climatic conditions, decide what to grow & when to grow, should know the overall crop yield turnover and must be able to be sure of the crop yield in spite of the environmental andother parameters.
2.	Idea / Solution description	Analysis of important visualization using the previous years' data, creating a dashboard and by going the datasets to obtain most of the insights of Crop production in India. This helps them out to overcome loss in farmingand business.
3.	Novelty / Uniqueness	we can analyze, visualize data and give the farmers the option to choose which plant/crop to cultivate in which period of time/season. A one-stop solution for understanding and to get an insight about the previousyears' data related to theharvest and cultivation.
4.	Social Impact/ CustomerSatisfaction	Availability to all the farmers who needhelp andas this is a simpleapproach, understanding issueswill not arise.
5.	Business Model (RevenueModel)	We can increase/enhance crop production and other raw materials. Also,Increase in productivity will result in increase of Revenue for the farmers.

3.4 PROBLEM SOLUTION FIT

There were multiple segments considered under the Problem Solution Fit and is illustrated below.



4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

Following are the functional requirements of the proposed solution.

Functional requirement	Description
Registration	User can create an account byregistering through form
Google authentication	Add newuser with a pre-existing google account
Login	Easier login for already existing users
Take in the required data	Input data is taken from the user for data prediction
Estimation / Prediction	A prediction of crop yield is done basedon the user input
Analysis	An analysis is done on the given data togain useful insights on the crop yield

4.2 NON-FUNCTIONAL REQUIREMENTS

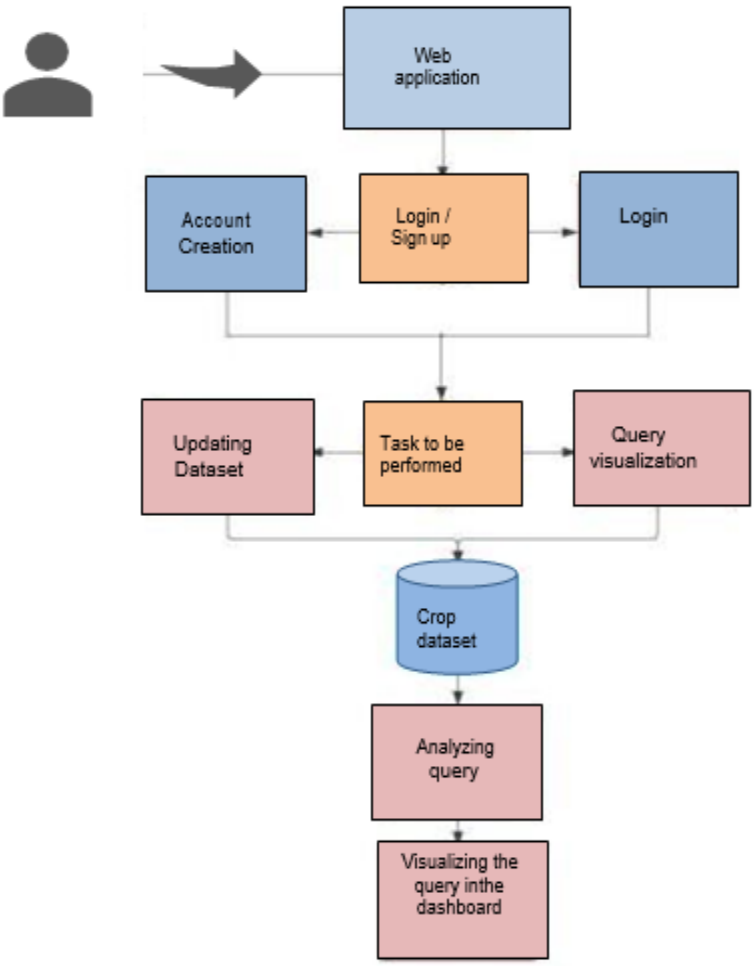
Following are the non-functional requirements of the proposed solution.

Non-functional requirement	Description
Performance	The software should provide usgood performance
Reliability	The UI should be user friendly and easily understandable
Availability	It should be available for access at anytime from anywhere
Scalability	The software should be scalable forlarger datasets
Security	The user login and registration is insecured manner

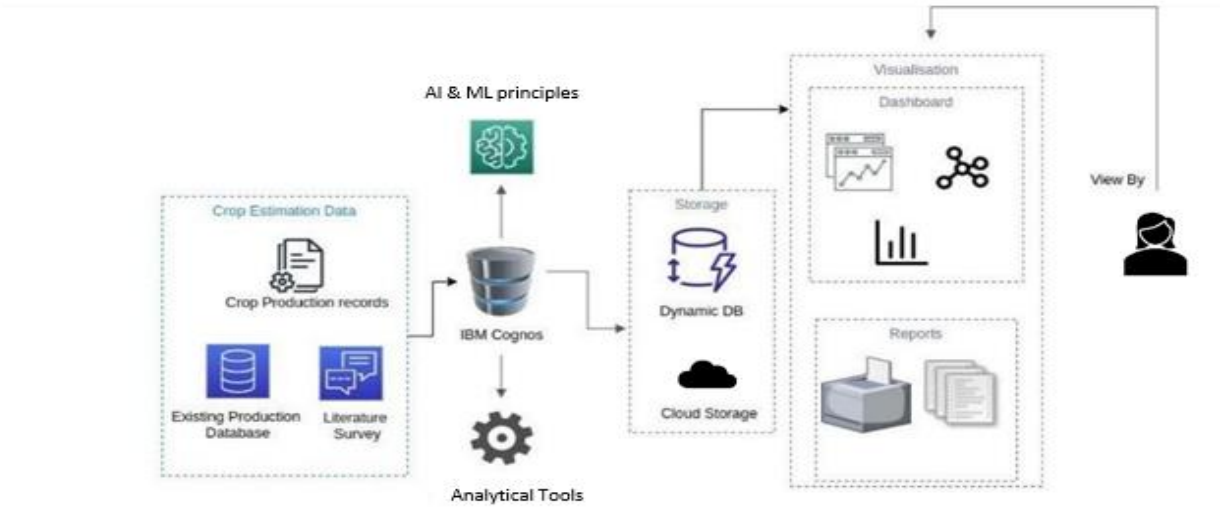
5. PROJECT DESIGN

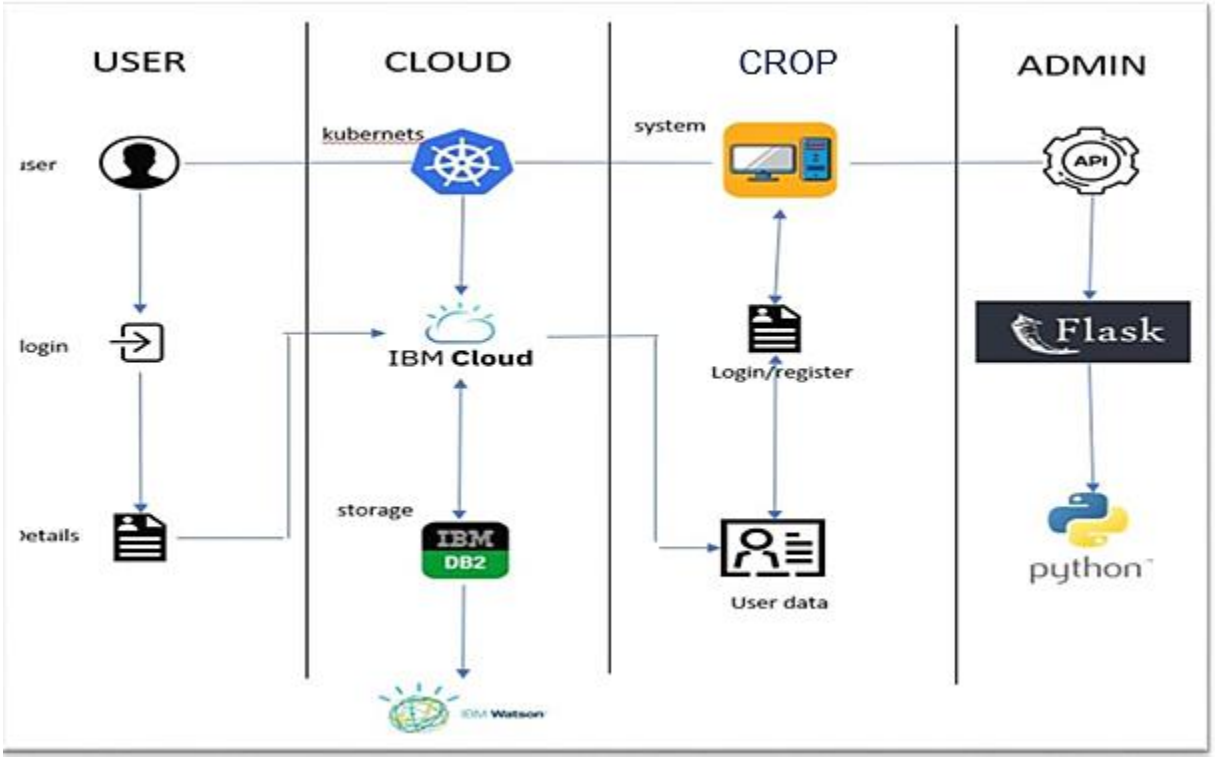
5.1 DATA FLOW DIAGRAMS

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.



5.2 SOLUTION AND TECHNICAL ARCHITECTURE


AI & ML principles
Visualisation
Dashboard
View By
Crop Estimation Data
Crop Production records
Existing Production Database
Literature Survey
IBM Cognos
Storage
Dynamic DB
Cloud Storage
Reports
Analytical Tools


USER
CLOUD
CROP
ADMIN
kubernets
system
API
login
IBM Cloud
Login/register
Flask
storage
IBM DB2
User data
python

5.3 USER STORIES

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 & accessthe dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for theapplication through Gmail	I can register and login with my email	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the applicationby entering email& password	I can use my mailidif I forget my password o reset it.	High	Sprint-1
	Dashboard	USN-6	As a user I can accessthe dashboard toview the required information	I can edit my profile	High	Sprint 2

6. PROJECT PLANNING AND SCHEDULE

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	I can sign up for the application as a user by providing my email address, a password, and a password confirmation.	2	High	Surya, Maria Rexana,Nandhini, Harini
Sprint-1		USN-2	After registering for the application, I as a user willreceive a confirmation email.	1	High	Surya, Maria Rexana,Nandhini, Harini
Sprint-2		USN-3	I can sign up for the application as a user using Google.	2	Low	Surya, MariaRexana, Nandhini, Harini
Sprint-1		USN-4	I can sign up for the application as a user usingGmail.	2	Low	Surya, Maria Rexana,Nandhini, Harini
Sprint-1	Login	USN-5	I can access the application as a user byproviding my email address and password.	1	High	Surya, Maria Rexana,Nandhini, Harini
Sprint-3	Dashboard	USN-6	I am free to use my dashboard and explore the features as a user.	2	High	Surya, MariaRexana, Nandhini, Harini

Sprint	Functional Requirement (Epic)	User Story Number	User Story & task	Story Points	Priority	Team Members

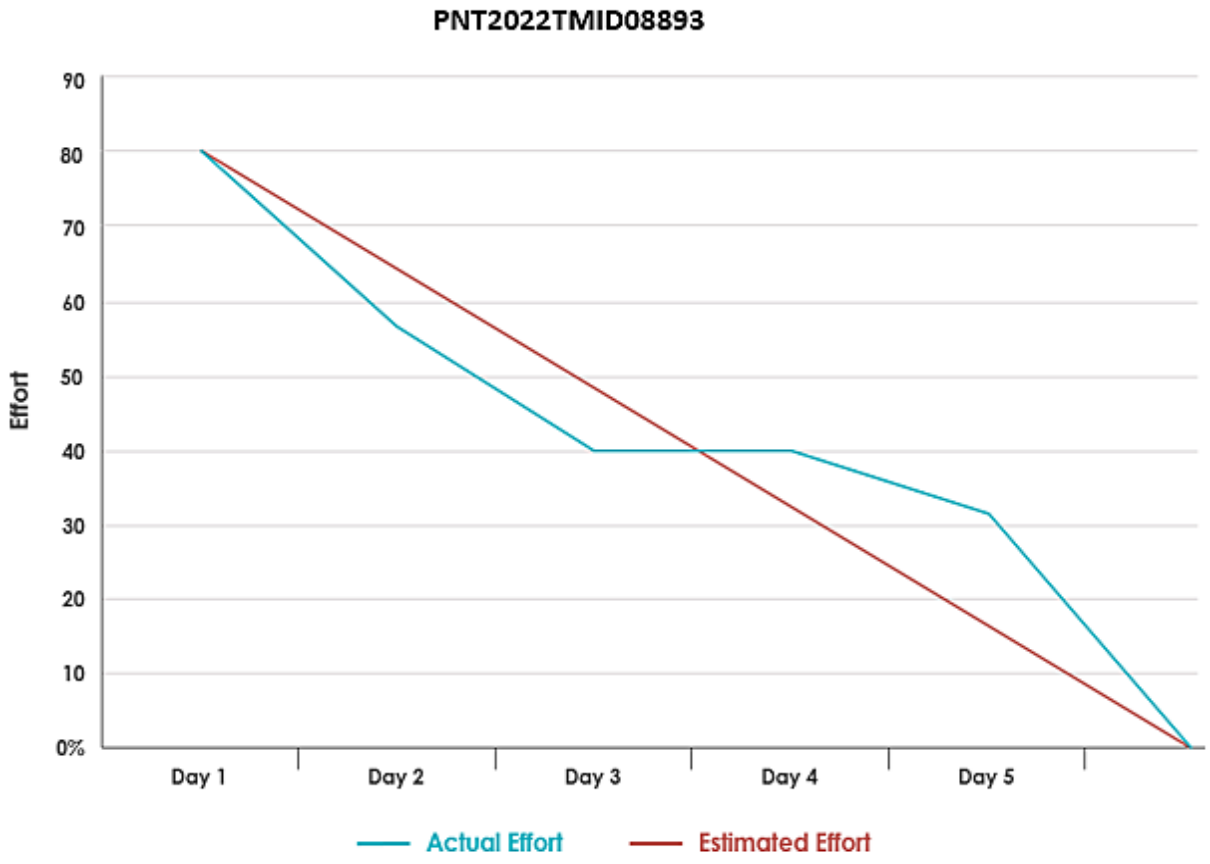
Sprint-2		USN-7	I can access using the credentials as a user theassets I'm applying for.	2	High	Surya, Maria Rexana,Nandhini, Harini
Sprint-3		USN-8	Data manipulation operations carried out by the application.	1	High	Surya, Maria Rexana
Sprint-3	Visualization	USN-9	Can use certain datasets to generate dashboards.	2	Medium	Nandhini, Harini
Sprint-4		USN-10	One can perform predictive analysis.	1	High	Maria Rexana,Nandhini
Sprint-3		USN-11	With certain datasets, I can produce stories.	2	High	Maria Rexana, Nandhini,Harini
Sprint-4		USN-12	Can export and send reportsin accordance withthe built-in dashboards and stories.	2	High	Surya, Maria Rexana,Nandhini, Harini

6.2 SPRINT DELIVERY SCHEDULE

Sprint	**Total Story Point**	**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	25 Oct 2022	02 Nov 2022	20	19 Nov 2022
Sprint-2	20	6 days	02 Nov 2022	09 Nov 2022	20	19 Nov 2022
Sprint-3	20	6 days	09 Nov 2022	15 Nov 2022	20	19 Nov 2022
Sprint-4	20	6 days	15 Nov 2022	19 Nov2022	20	19 Nov 2022

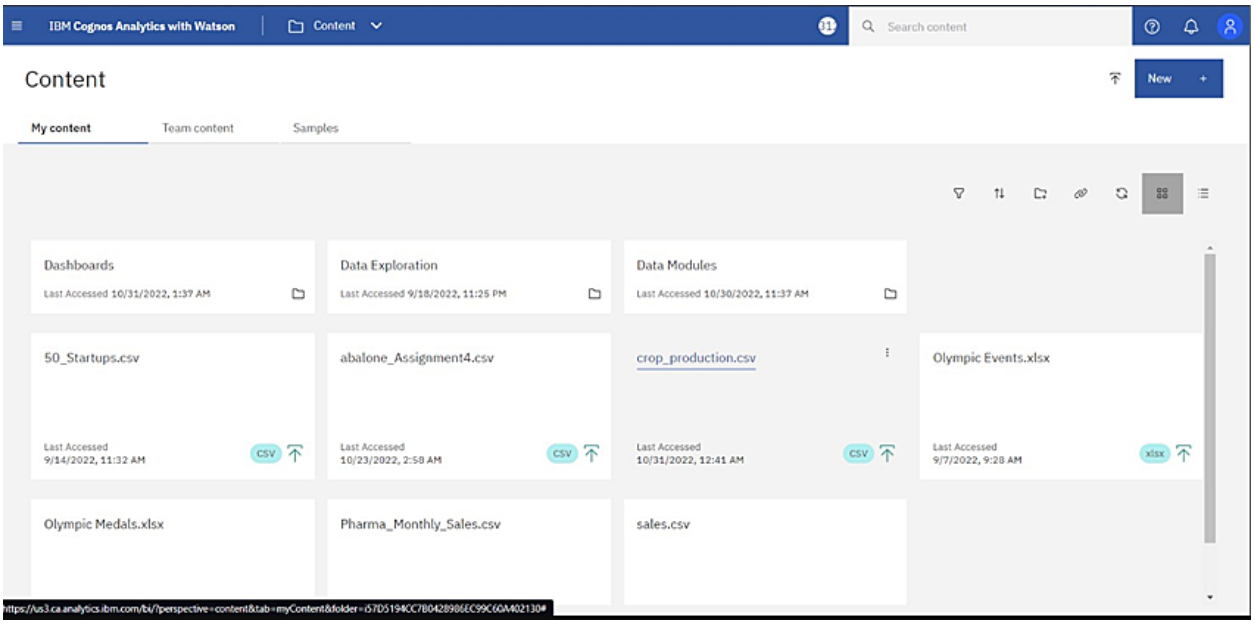
6.3 BURNDOWN CHART

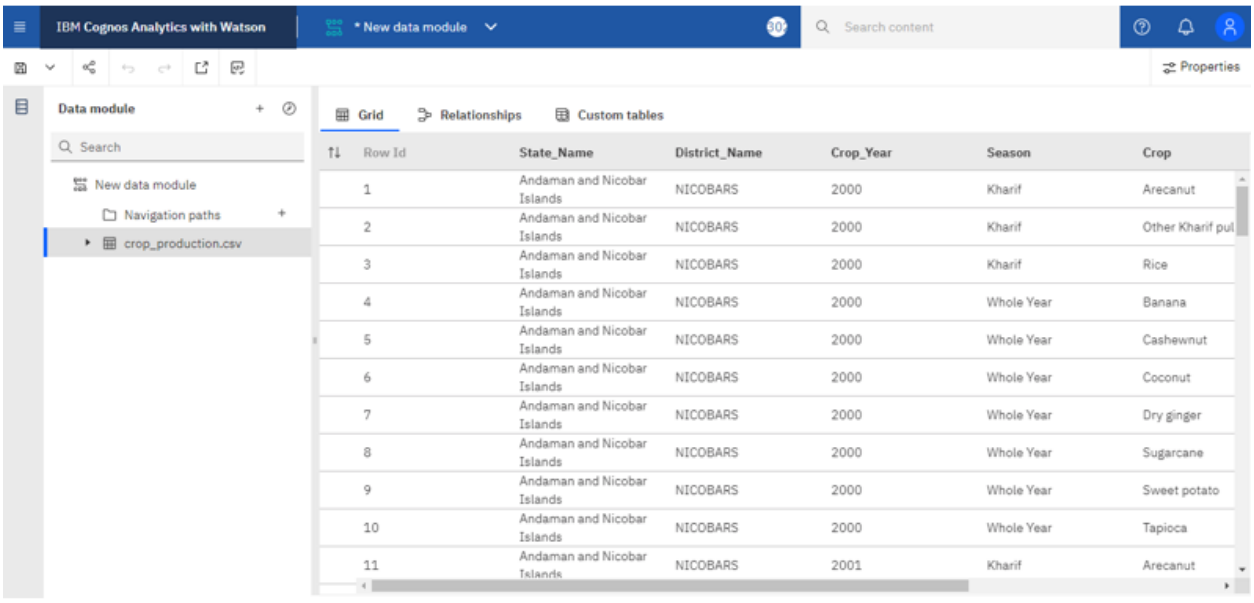
A burn down chart plots the amount of work remaining to perform against the amount of time. Inagile software development approaches like Scrum, it is frequently employed. Burn down charts, however, can be used for any project that makes observable progress over time.

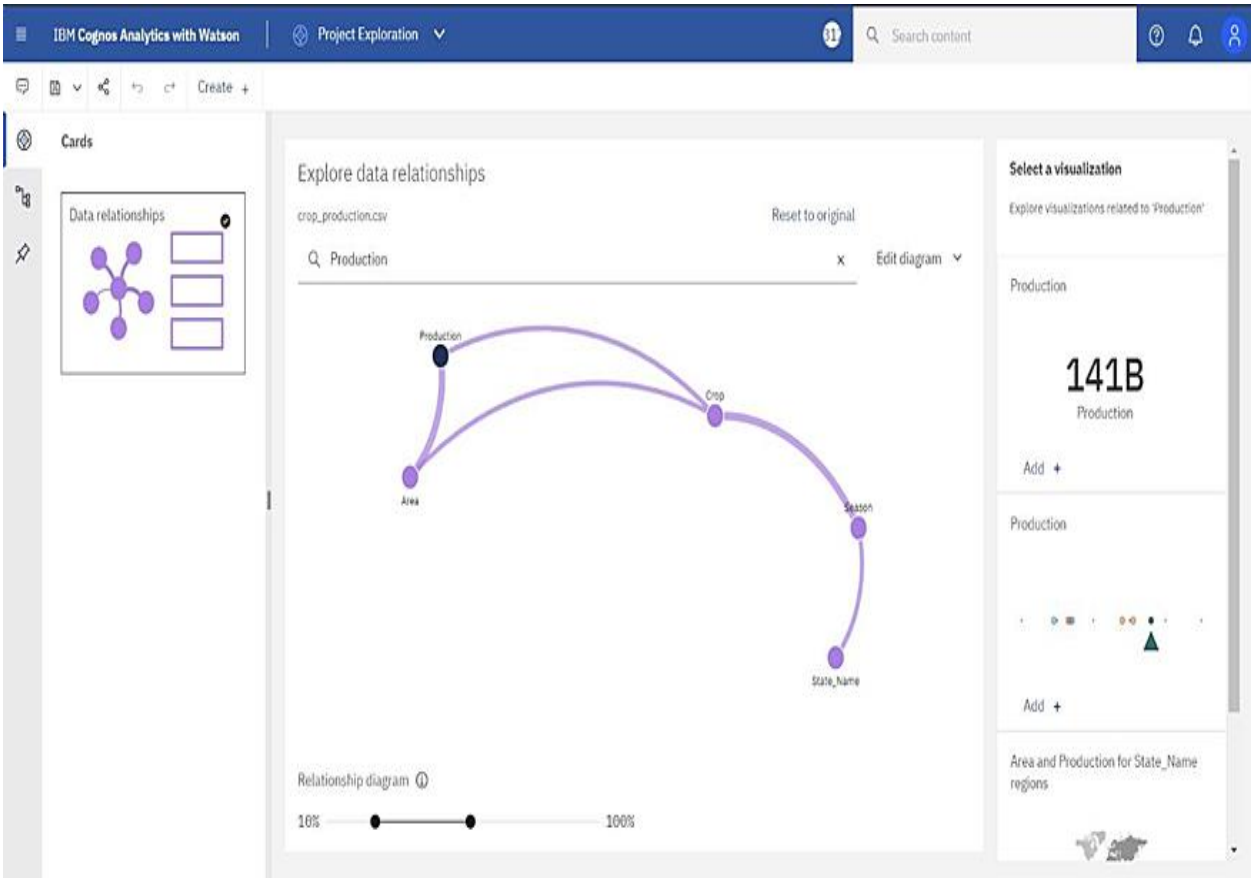


7. CODING AND SOLUTIONING

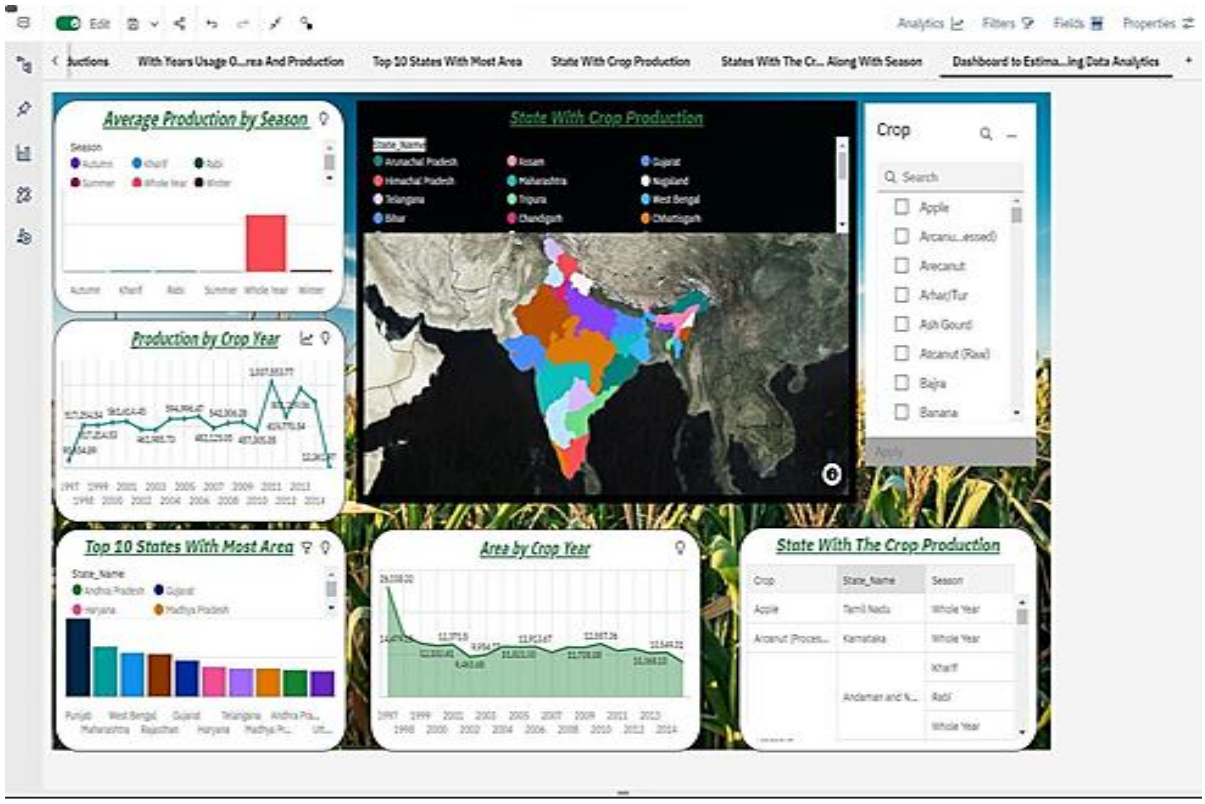
7.1 DATA COLLECTION AND PREPARATION

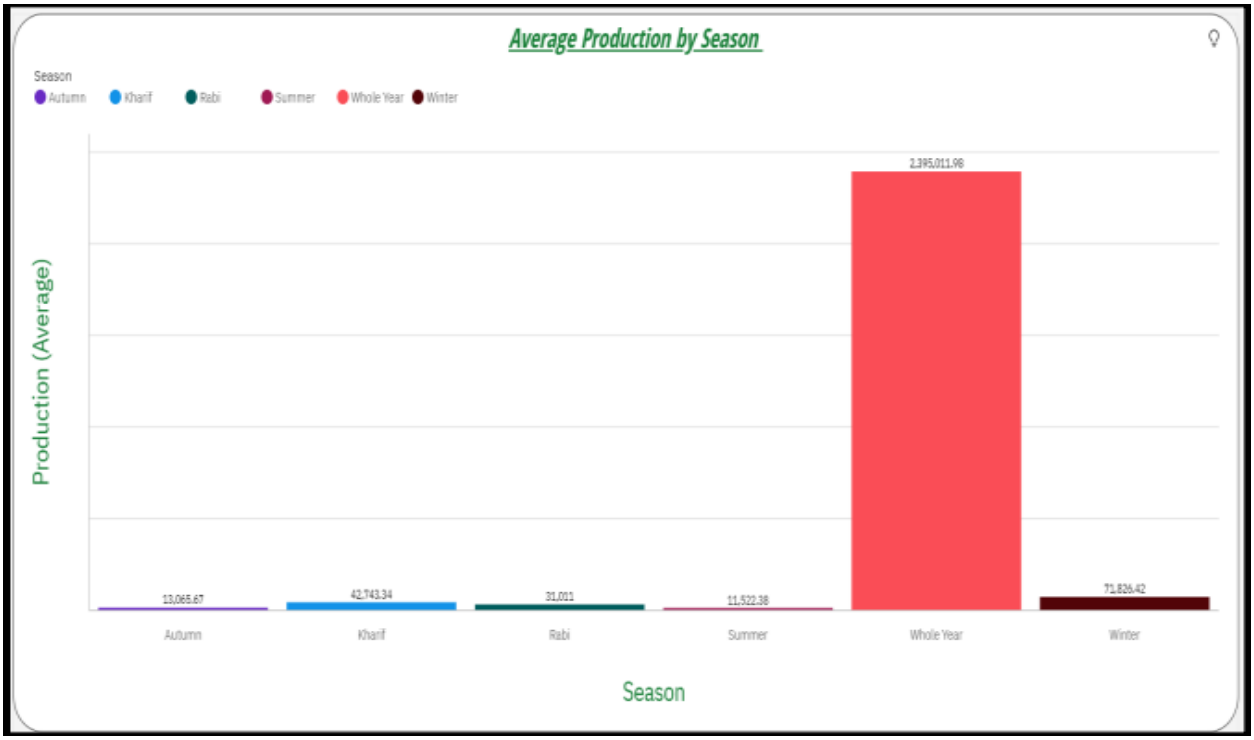


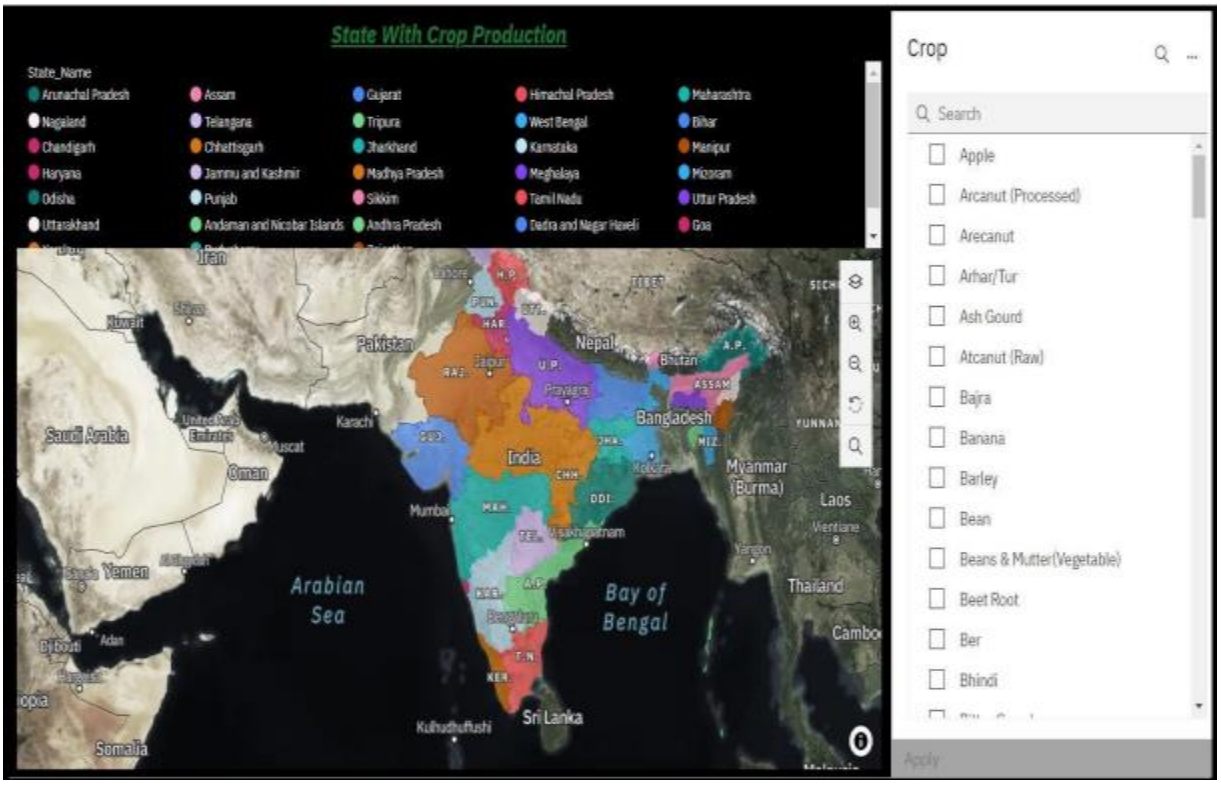


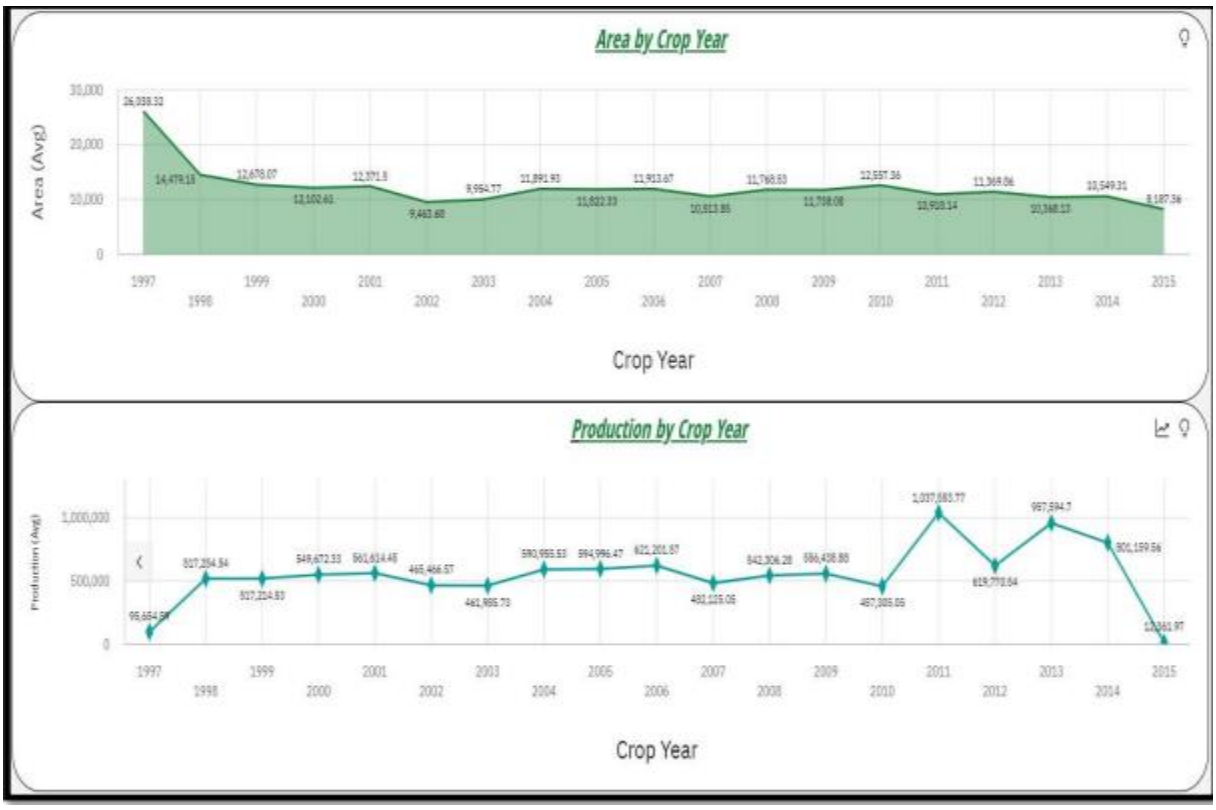


7.2 DASHBOARD CREATION

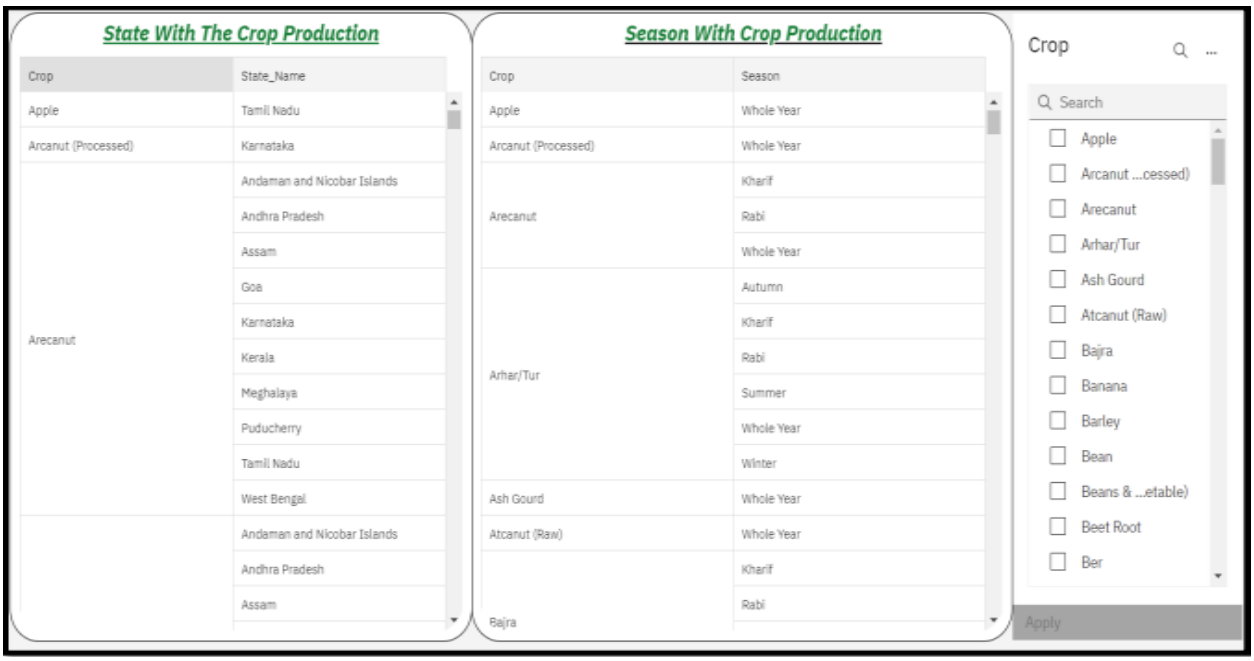

Edit
Analytics
Filters
Fields
Properties
With Years Usage O...rea And Production
Top 10 States With Most Area
State With Crop Production
States With The Cr... Along With Season
Dashboard to Estima...ing Data Analytics
Average Production by Season
Season
Autumn
Kharif
Rabi
Summer
Whole Year
Winter
State With Crop Production
State_Name
Arunachal Pradesh
Assam
Gujarat
Himachal Pradesh
Maharashtra
Nagaland
Telangana
Tripura
West Bengal
Bihar
Chandigarh
Chhattisgarh
Crop
Search
Apple
Arcanu...essed)
Arecanut
Arhar/Tur
Ash Gourd
Atcanut (Raw)
Bajra
Banana
Apply
Production by Crop Year
Top 10 States With Most Area
State_Name
Andhra Pradesh
Gujarat
Haryana
Madhya Pradesh
Punjab
West Bengal
Gujarat
Telangana
Andhra Pra...
Maharashtra
Rajasthan
Haryana
Madhya Pr...
Area by Crop Year
State With The Crop Production
Crop
State_Name
Season
Apple
Tamil Nadu
Whole Year
Arcanut (Proces...
Karnataka
Whole Year
Kharif
Andaman and N...
Rabi
Whole Year

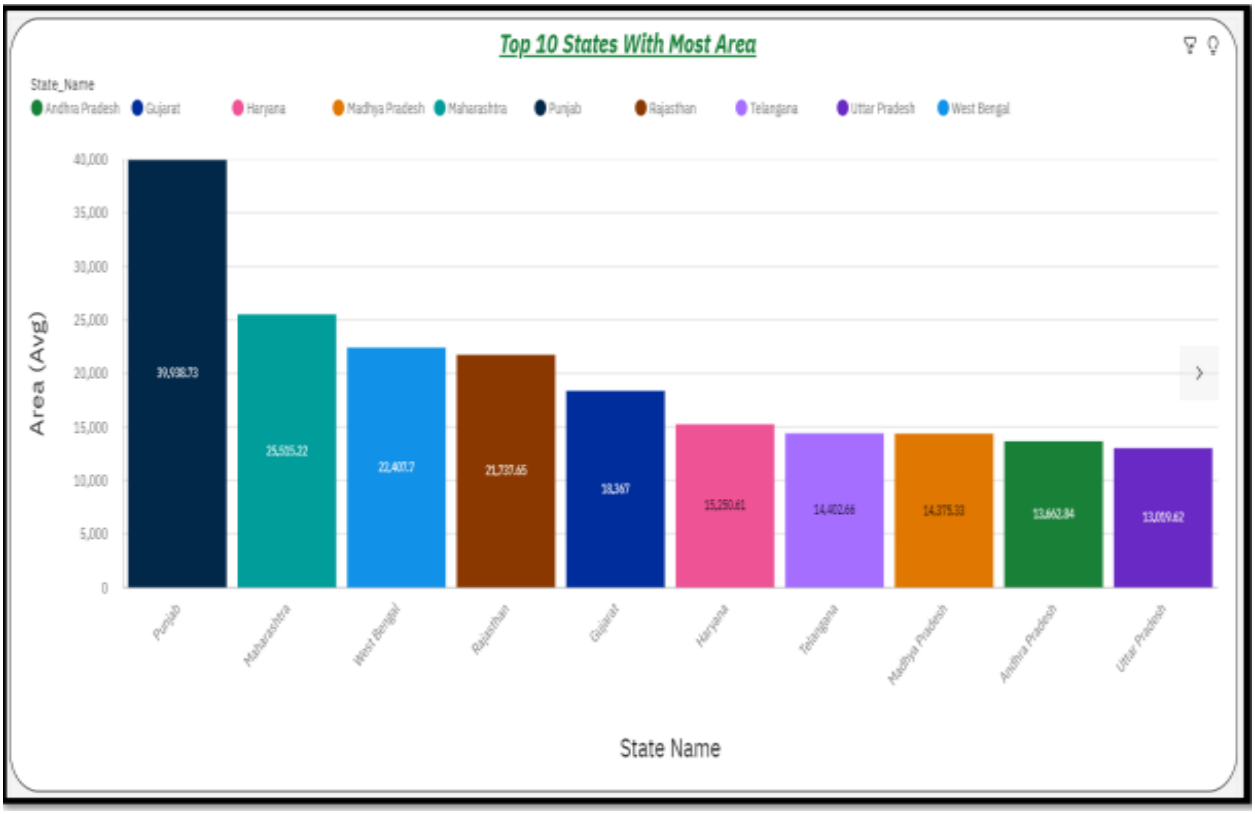

Average Production by Season
Season
Autumn
Kharif
Rabi
Summer
Whole Year
Winter
Production (Average)
2,395,011.98
13,065.67
42,743.34
31,001
11,522.38
71,826.42
Autumn
Kharif
Rabi
Summer
Whole Year
Winter
Season

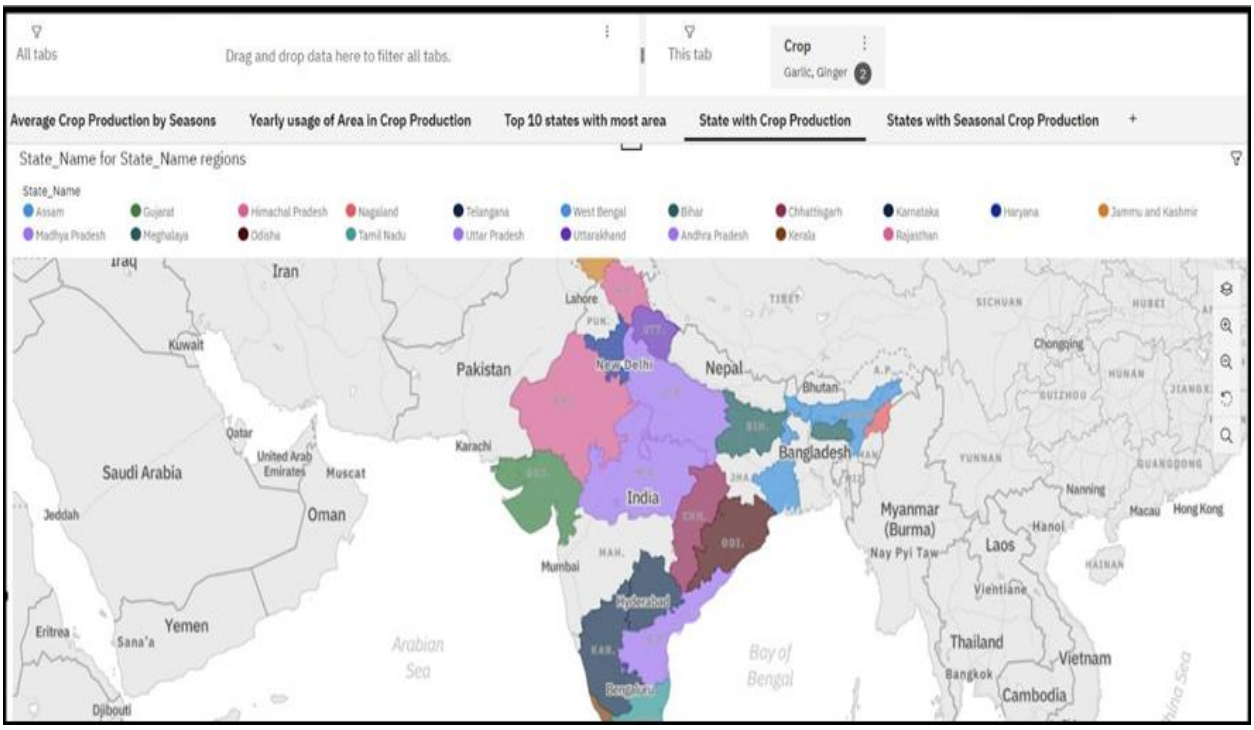

State With Crop Production
State_Name
Arunachal Pradesh
Assam
Gujarat
Himachal Pradesh
Maharashtra
Nagaland
Telangana
Tripura
West Bengal
Bihar
Chandigarh
Chhattisgarh
Jharkhand
Karnataka
Manipur
Haryana
Jammu and Kashmir
Madhya Pradesh
Meghalaya
Mizoram
Odisha
Punjab
Sikkim
Tamil Nadu
Uttar Pradesh
Uttarakhand
Andaman and Nicobar Islands
Andhra Pradesh
Dadra and Nagar Haveli
Goa
Iran
Kuwait
Pakistan
Nepal
Bhutan
Bangladesh
Saudi Arabia
Karachi
Oman
India
Myanmar (Burma)
Laos
Mumbai
Yemen
Arabian Sea
Bay of Bengal
Thailand
Djibouti
Adan
Somalia
Sri Lanka
Kulhudhuffushi
Crop
Search
Apple
Arcanut (Processed)
Arecanut
Arhar/Tur
Ash Gourd
Atcanut (Raw)
Bajra
Banana
Barley
Bean
Beans & Mutter(Vegetable)
Beet Root
Ber
Bhindi
Apply

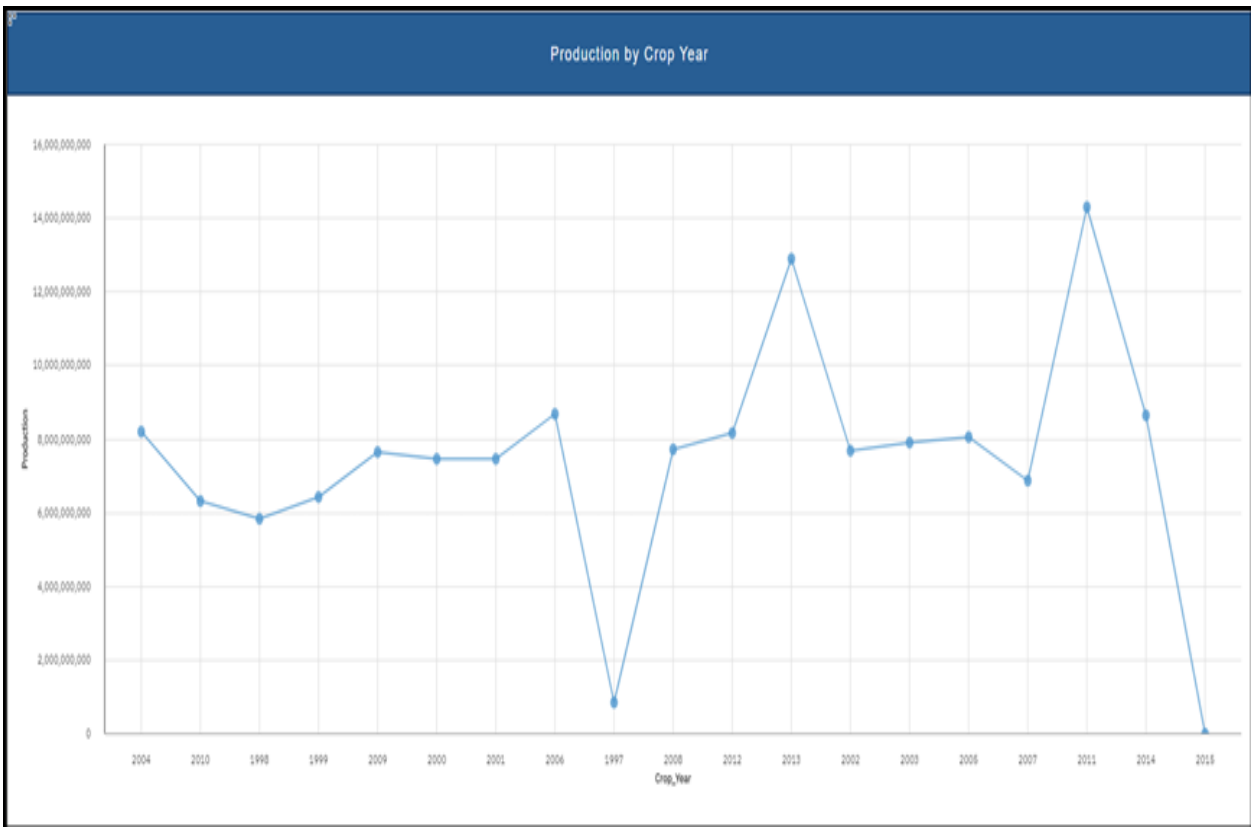

Area by Crop Year
Area (Avg)
Crop Year
Production by Crop Year
Production (Avg)
Crop Year

7.3 REPORT CREATION

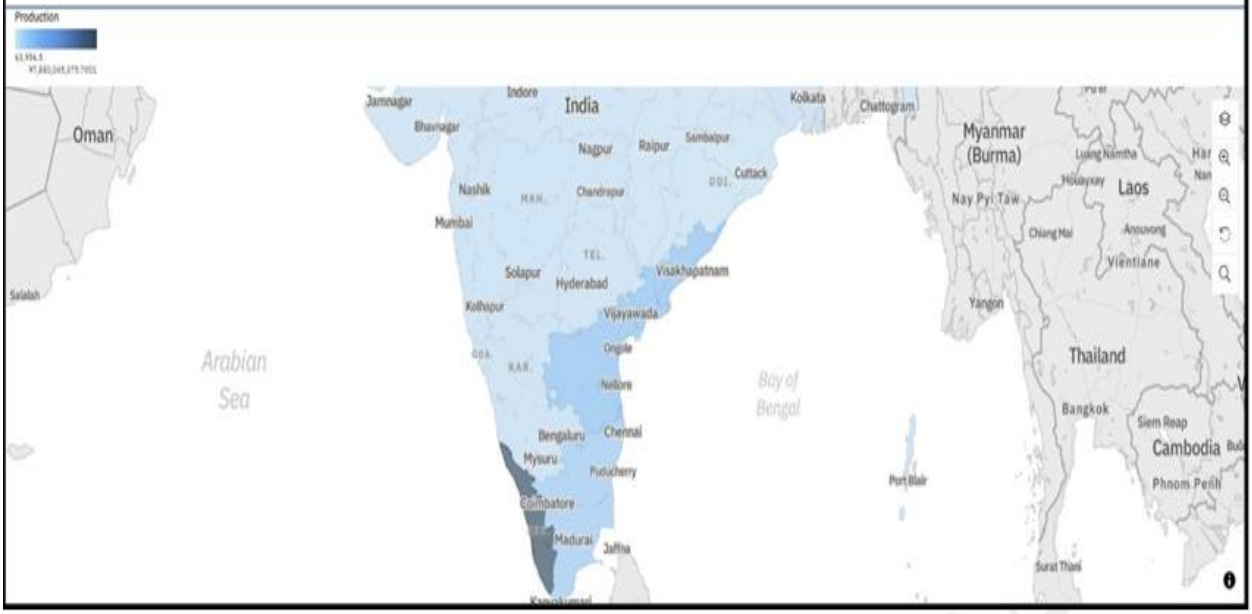





All tabs
Drag and drop data here to filter all tabs.
This tab
Crop
Garlic, Ginger
2
Average Crop Production by Seasons
Yearly usage of Area in Crop Production
Top 10 states with most area
State with Crop Production
States with Seasonal Crop Production
+
State_Name for State_Name regions
State_Name
Assam
Gujarat
Himachal Pradesh
Nagaland
Telangana
West Bengal
Bihar
Chhattisgarh
Karnataka
Haryana
Jammu and Kashmir
Madhya Pradesh
Meghalaya
Odisha
Tamil Nadu
Uttar Pradesh
Uttarakhand
Andhra Pradesh
Kerala
Rajasthan
Iran
Lahore
Kuwait
Pakistan
New Delhi
Nepal
Bhutan
Bangladesh
Qatar
United Arab Emirates
Muscat
Karachi
Saudi Arabia
India
Jeddah
Oman
Myanmar (Burma)
Nay Pyi Taw
Laos
Hanoi
Nanning
Macau
Hong Kong
Mumbai
Hyderabad
Vientiane
Eritrea
Yemen
Sana'a
Arabian Sea
Bay of Bengal
Thailand
Bangkok
Vietnam
Cambodia
Djibouti


Production by Crop Year
16,000,000,000
14,000,000,000
12,000,000,000
10,000,000,000
8,000,000,000
6,000,000,000
4,000,000,000
2,000,000,000
0
Production
2004
2010
1998
1999
2009
2000
2001
2006
1997
2008
2012
2013
2002
2003
2005
2007
2011
2014
2015
Crop_Year

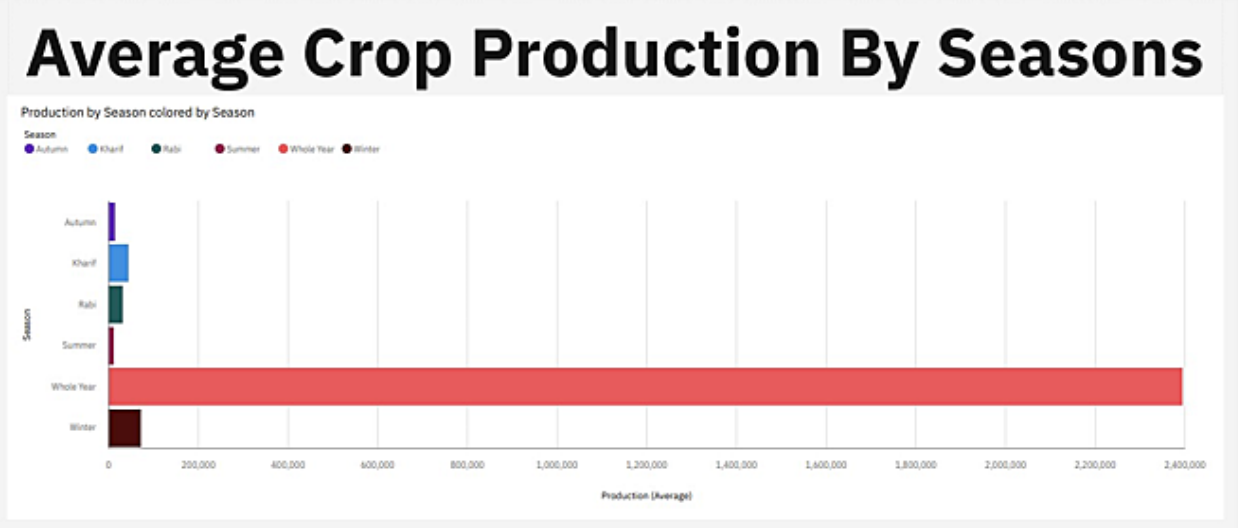
STATE WITH CROP PRODUCTION

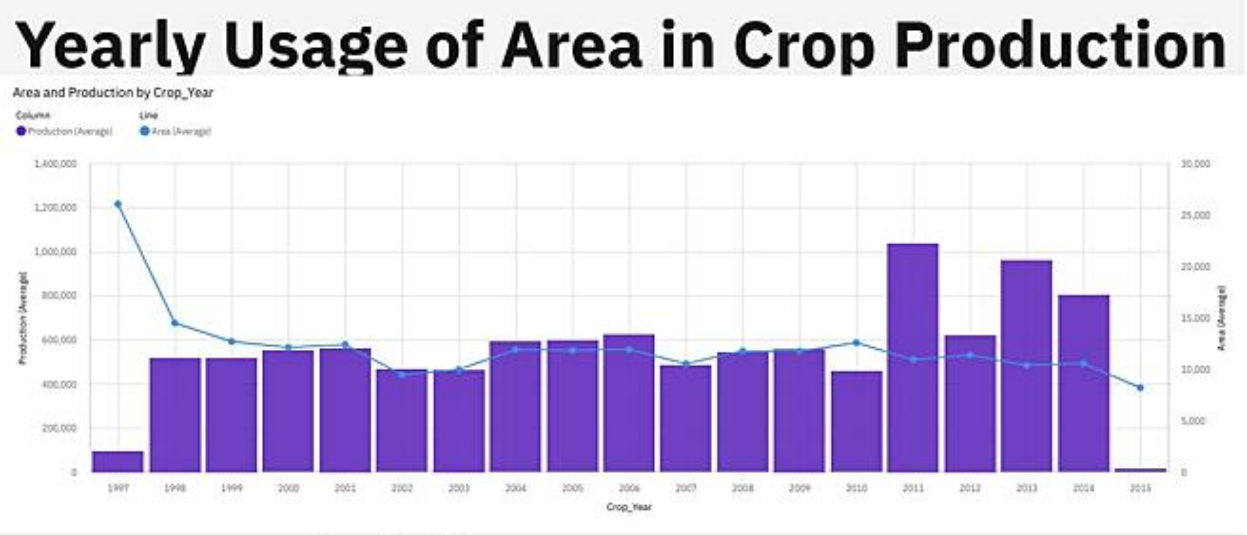


STATES WITH SEASONAL CROP PRODUCTION

Crop	State_Name	Season
Other Kharif pulses	Andaman and Nicobar Islands	Kharif
Rice	Andaman and Nicobar Islands	Kharif
Cashewnut	Andaman and Nicobar Islands	Whole Year
Horse-gram	Andhra Pradesh	Kharif
Tobacco	Andhra Pradesh	Kharif
Ragi	Andhra Pradesh	Rabi
Onion	Andhra Pradesh	Whole Year
other misc. pulses	Andhra Pradesh	Kharif
Sweet potato	Andhra Pradesh	Whole Year
Turmeric	Andhra Pradesh	Whole Year
Soyabean	Andhra Pradesh	Kharif
Beans & Mutter(Vegetable)	Andhra Pradesh	Whole Year
Bhindi	Andhra Pradesh	Whole Year
Grapes	Andhra Pradesh	Whole Year
Cowpea(Lobia)	Andhra Pradesh	Kharif
Arecanut	Andhra Pradesh	Kharif
Coriander	Andhra Pradesh	Kharif
Linseed	Andhra Pradesh	Kharif
Sapota	Andhra Pradesh	Kharif
Tomato	Andhra Pradesh	Kharif

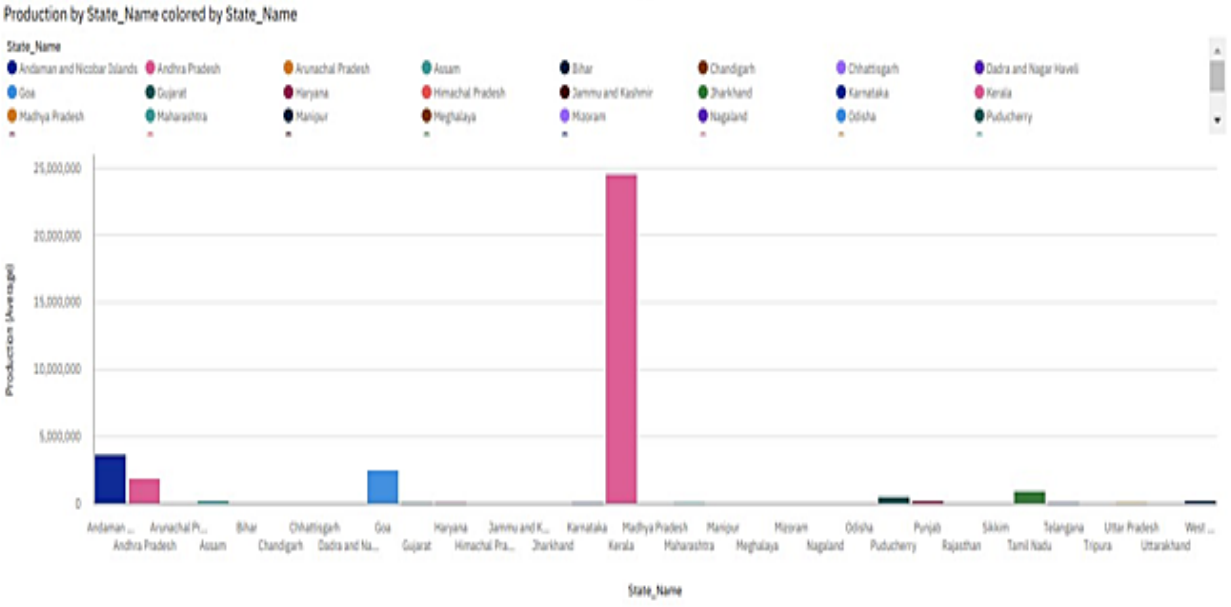
7.4 STORY CREATION





Top 10 states with most Area

State with Crop Production



8. ADVANTAGES

Productivity is boosted by technology. Agriculture has seen a significant boost in productivity as a result of technology; farmers can nowdo more work with less effort and in less time.

Technology saves money. Using current agricultural technology can help farmers save money. With the help of modern technologies, farmers may work more efficiently, with less effort, and in less time.

With modern technology, work that formerly required a big numberof people and a lengthy period of time may now be accomplished swiftly and cheaply. Farmers are not compelled to pay a separate price for their services.

9. DISADVANTAGES

High costs of maintenance. One of the downsides of agriculture technology is its high maintenance costs. The hefty maintenance costs of the technology make it tough for small enterprises and farmers to handle.

Farmers find it difficult to keep up with technology since they cannot afford the high maintenance costs of contemporary technical gadgets and machines. Farmers Who Are Undereducated are illiterate, and understanding how to use current technologies in farming is challenging.

10. CONCLUSION

Agriculture yield data is used to analyze and improve the crop yield and represent in the form of a Graphs through data visualization technique.

The visualization methods presented include interactive charts toenable our data users to drill down and focus on more detailed views ofthese data displays.

11. FUTURE SCOPE

In the future, we expect to extend the same as an even more easilyaccessible mobile application and further enhancements on the user experience is aimed to be implemented.

12. APPENDIX

Source Code

```
<!DOCTYPE html>

<html lang="en">

<head>

<title>Estimate The Crop Yield Using Data Analytics</title>

<meta charset="utf-8">
```

```
 <meta name="viewport" content="width=device-width, initial-scale=1">
<linkhref="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/css/bootstrap.min.css" rel="stylesheet">
<scriptsrc="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/js/bootstrap.bundle.min.js"></script>
<linkhref="https://fonts.googleapis.com/css2?family=Merienda:wght@700&display=swap" rel="stylesheet">
<linkrel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">
<style>
body {
        background-align: center;
        height: auto;
        }
        a: link, a:visited {
                background-color: #004d26;
                color: white;
                padding: 15px 25px;
                text-align: center;
                text-decoration: none;
```

```
        display: inline-block;
        border: white;
        border-width:4px;
        border-style: dashed;
        }
a:hover, a:active {
        background-color: DarkBlue;
        font-size: 140%;
        }
h2{
        font-size:35px;
        color: white;
        position: absolute;
        top:20px;
        right: 490px;
        }
h4{
        font-size:30px;
        color: white;
```

```
        }
    p{
        font-size:25px;
        color: white;
        }
</style>
</head>
<body background="3.jpg">
<div class="container-md">
    <div class='p-5'>
     <div class="p-5">
    <br>
    <h2>ESTIMATE THE CROP YIELD USING DATA ANALYTICS</h2><br><br>
    <ahref="https://drive.google.com/file/d/1o1RRP0imauJZ1QOhknU-
    XYigSxaksq9N/view?usp=sharing">View Dashboard</a><br><br>
    <h4>
    <strong>The Dashboard is based on 5 visualizations.<br>
    </h4>
    <p>
```

They are :

1.Average Crop Production by Seasons

2.Yearly usage of Area in Crop Production

3.Top 10 States in Crop Yield Production by Area

4.Crop Production by State

5.States with Seasonal Crop Production using a Text representation

</p>

</body>

</html>

Link to Github Repository

https://github.com/IBM-EPBL/IBM-Project-1707-1658410286

Project Demo Video link

https://drive.google.com/open?id=1vwbsclLIYHFNMdmiE5pIuuOPAIIjIwnO&authuser=thangarajnk19755%40gmail.com&usp=drive_fs