

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

TEAM ID:	PNT2022TMID17362	
TEAM LEAD:	VELAVAN S	(319UCS108)
TEAM MEMBERS:		
	THARANI THARAN T	(319UCS107)
	AKTHAR KHAN	(319UCS008)
	HARISH BABU	(319UCS039)
INSTITUTION:	MAHENDRA INSTITUTE OF TECHNOLOGY	

**ESTIMATE THE CROP YIELD USING DATA
ANALYTICS**
(DOMAIN: DATA ANALYTICS)

PROJECT REPORT

November 2022

TABLE OF CONTENTS

1.		INTRODUCTION
	1.1.	PROJECT OVERVIEW
	1.2.	PURPOSE
2.		LITERATURE REVIEW
	2.1.	EXISTING PROBLEM
	2.2.	REFERENCES
	2.3.	PROBLEM STATEMENT DEFINITION
3.		IDEATION & PROPOSED SOLUTION
	3.1.	EMPATHY MAP CANVAS
	3.2.	IDEATION & BRAINSTORMING
	3.3.	PROPOSED SOLUTION
	3.4.	PROBLEM SOLUTION FIT
4.		REQUIREMENT ANALYSIS
	4.1.	FUNCTIONAL REQUIREMENT
	4.2.	NON-FUNCTIONAL REQUIREMENTS
5.		PROJECT DESIGN
	5.1.	DATA FLOW DIAGRAMS
	5.2.	SOLUTION & TECHNICAL ARCHITECTURE
	5.3.	USER STORIES
6.		PROJECT PLANNING & SCHEDULING
	6.1.	SPRINT PLANNING & ESTIMATION
	6.2.	SPRINT DELIVERY SCHEDULE
	6.3.	BURNDOWN CHART
7.		CODING & SOLUTIONING
	7.1.	DASHBOARD CREATION
	7.2.	DASHBOARD CREATION
	7.3.	REPORT CREATION
	7.4.	STORY CREATION
8.		ADVANTAGES
9.		DISADVANTAGES
10.		CONCLUSION
11.		FUTURE SCOPE
12.		APPENDIX

1. INTRODUCTION

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.

Agriculture fulfils a fundamental need, which makes it crucial for human survival. It is a well-known fact that in India, agriculture employs the bulk of the population (about 55%). There are obstacles to expanding crop production in India because of weather changes. The way modern farms and agricultural enterprises operate differs greatly from how they did a few decades ago, largely due to technological developments in the form of sensors, machinery, devices, and information technology.

PROJECT OVERVIEW

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, and one 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 the crop yield was the main aim and was brought to life.

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 analyse data in order to increase productivity and financial gain. In order to examine different behavioural 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 in crop yield due to various parameters that can be natural or non-natural.

2. LITERATURE SURVEY

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 sensitive to 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.

REFERENCES

- i. How data analytics is transforming agriculture - ScienceDirect - <https://doi.org/10.1016/j.bushor.2017.09.011>
- ii. https://www.researchgate.net/publication/329467349_Agriculture_Data_Analytics_in_Crop_Yield_Estimation_A_Critical_Review
- iii. https://www.researchgate.net/publication/359131334_Data_analytics_platforms_for_agricultural_systems_A_systematic_literature_review
- iv. N. Chergui, M. -T. Kechadi and M. McDonnell, "The Impact of Data Analytics in Digital Agriculture: A Review," 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA), 2020, pp. 1-13, doi: 10.1109/OCTA49274.2020.9151851.
- v. D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian

Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2020, doi: 10.1109/ACCESS.2020.2992480.

PROBLEM STATEMENT DEFINITION

The following instances define the problem of notice. Ram is a farmer who needs a way to understand and predict climatic conditions because 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 and when 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 turnover because 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 & PROPOSED SOLUTION

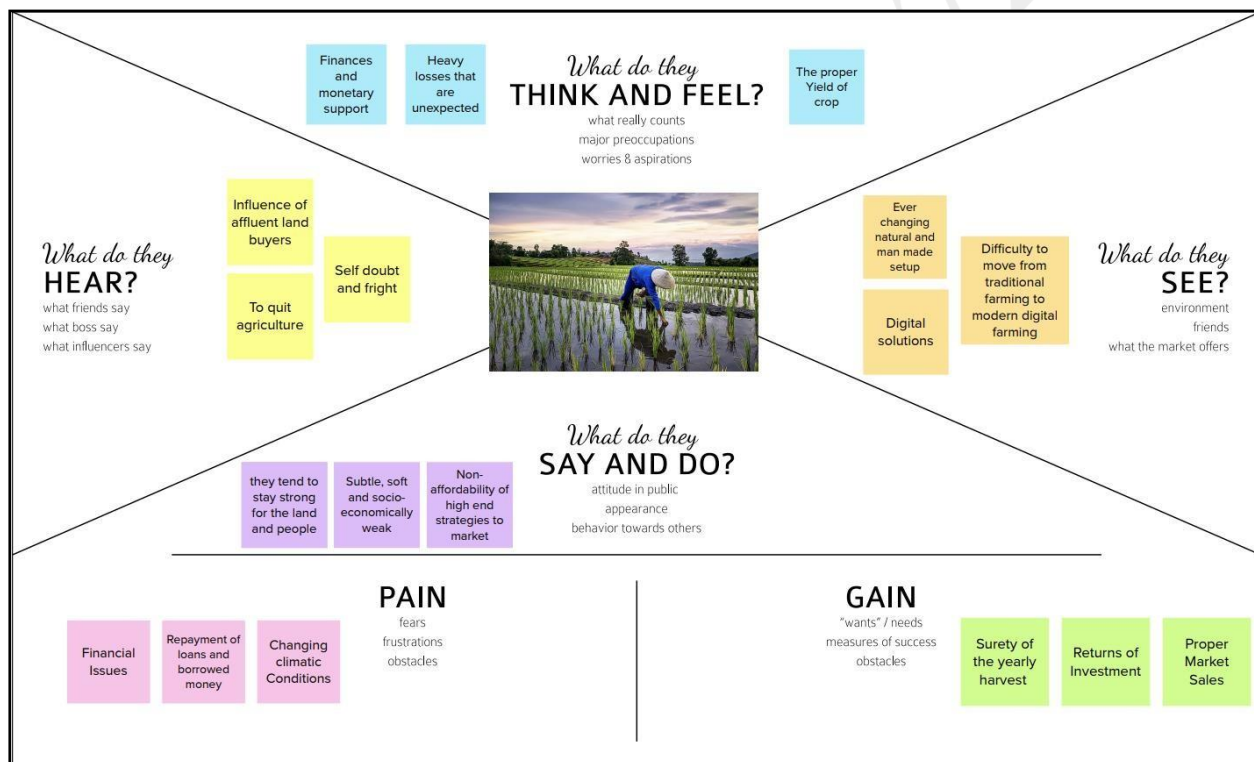
EMPATHY MAP CANVAS

A team of four members sat together to discuss on and empathize about 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.



IDEATION & BRAINSTORMING

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

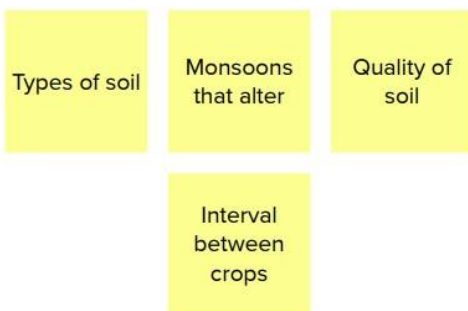
PROBLEM STATEMENT:

PROBLEM

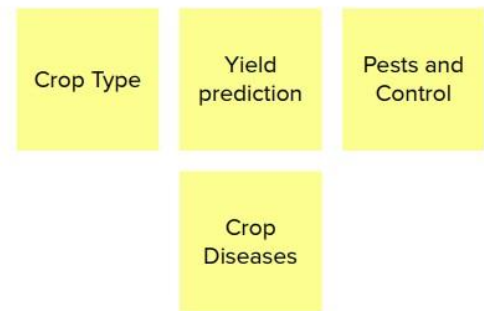
A farmer should predict climatic conditions, decide what to grow & when to grow, should know the overall crop yield turnover

BRAINSTORMING:

BAVYAA R



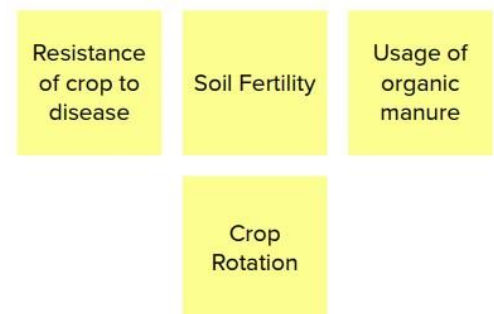
SATHYANARAYANAN K



PONRAJ S



ARUNKUMAR



GROUPING:

CROP PARAMETERS

Crop Type

Varied Crop
Growth

Crop Growth
Rate

Interval
between
crops

Crop
Diseases

Crop
Rotation

Resistance
of crop to
disease

SOIL PARAMETERS

Types of soil

Soil Fertility

Quality of
soil

ENVIRONMENTAL SETUP

Environmental
Condition

Monsoons
that alter

Irrigation
type

OTHERS

Yield
prediction

Usage of
organic
manure

Pests and
Control

PRIORITIZATION:



PROPOSED SOLUTION

A farmer should predict 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 inspite of the environmental and other parameters. 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 is chosen and proposed as the solution.

A one-stop solution for understanding and to get an insight about the previous years' data related to the harvest and cultivation. There is no other setup that's required to be installed as an adage;

Availability to all the farmers who need help and as this is a simple approach, understanding issues will not arise. A profit can be made by promoting the solution as an easily available mobile application for anyone to access and benefit out of it. Venture joints with government can be made to pull out monetary benefits. There is no issue with regards to storage of datasets and collection of data. Hence, the solution can be easily scaled to handle data needs, traffic and increased number of users.

PROBLEM SOLUTION FIT

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

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS <ul style="list-style-type: none"> • Farmers • Individuals associated with agricultural activities, cultivation, harvest and sales of the harvested goods. 	6. CUSTOMER CONSTRAINTS CC <p>What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.</p> <ul style="list-style-type: none"> • Monetary Issues • Network Issues • Lack of Awareness • Quality of soil, manure, water etc. 	5. AVAILABLE SOLUTIONS AS <p>Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking</p> <ul style="list-style-type: none"> • Traditional ways of prediction • Precision farming 	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS J&P <ul style="list-style-type: none"> • Help them understand the usage of prediction and software for better results in agriculture • Data is to be collected and awareness should be brought in order to orchestrate the above mentioned 	9. PROBLEM ROOT CAUSE RC <ul style="list-style-type: none"> • Weather conditions • Soil Conditions • Water availability • Unpredictable weather conditions • Pest issues • Manure and other usages • Crop resistance 	7. BEHAVIOUR BE <ul style="list-style-type: none"> • Try to get help from agricultural experts • Try to take up non-natural means of cultivation for quicker harvest 	

Identify strong TR & EM	3. TRIGGERS <small>What triggers customers to act? i.e. seeing their neighbor installing solar panels, reading about a more efficient solution in the news.</small> Seeing their neighbor farmers have a better yield by usage of natural or non-natural means	10. YOUR SOLUTION <small>If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behavior.</small> An interactive, visualization dashboard that precisely indicates the predictions of the needed parameters for farmers to get insights about what is needed. It is a one-stop solution and there is no extra setup that is required.	8. CHANNELS of BEHAVIOUR 8.1 ONLINE <small>What kind of actions do customers take online? Extract online channels from #7</small> None 8.2 OFFLINE <small>What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.</small> <ul style="list-style-type: none"> Trying to use pesticides and fertilizers that increase gain but cause harm Irrigation channel changes 	Identify strong TR & EM
	4. EMOTIONS: BEFORE / AFTER <small>How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design.</small> Before: insecure > strength / trust of growth: After			

4. REQUIREMENT ANALYSIS

FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Signup	Registration through Gmail Registration through IBM
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Data Collection	
FR-4	Data Processing	Data cleaning, removal of noise and obsolete data
FR-5	Visualization Tool	Graphical visualization choices

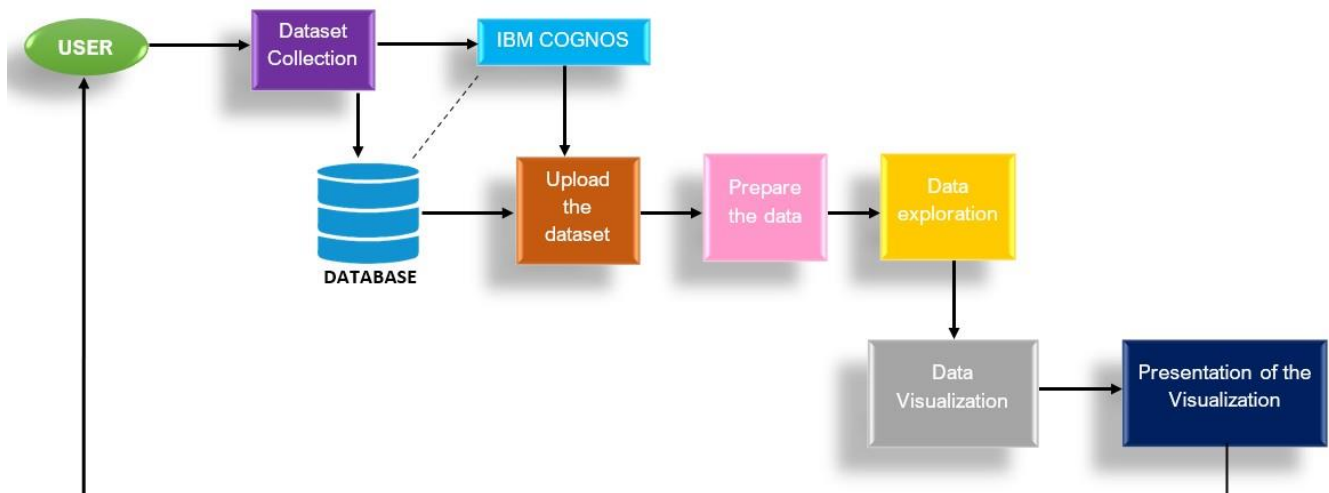
NON-FUNCTIONAL REQUIREMENTS

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

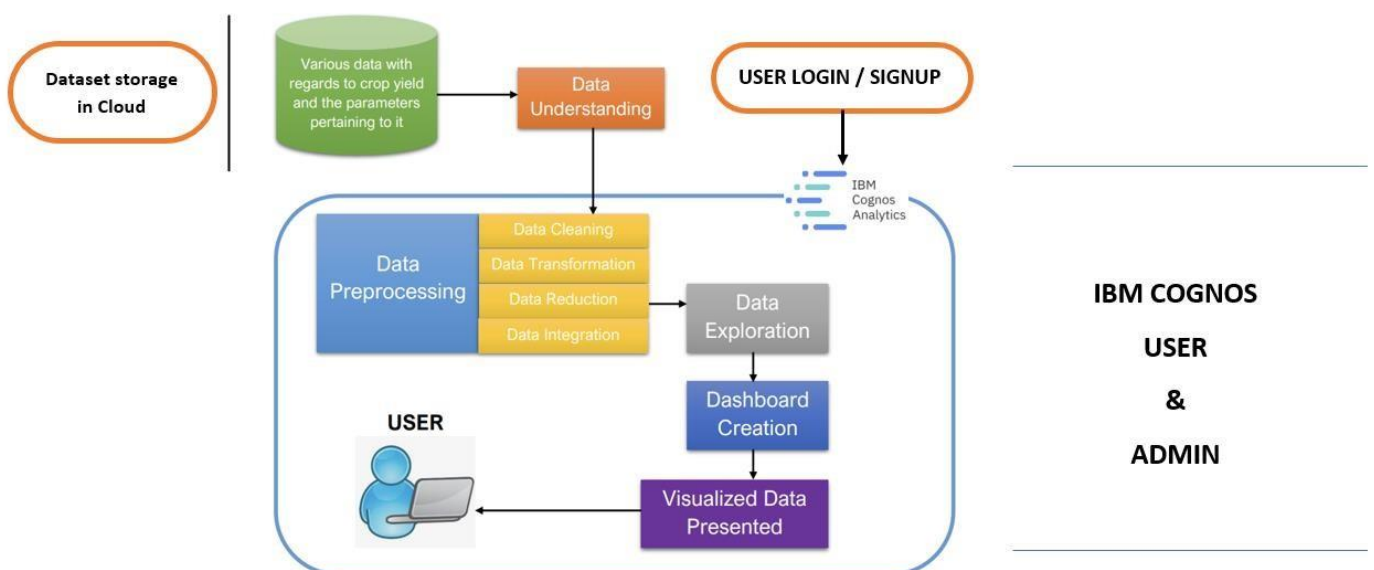
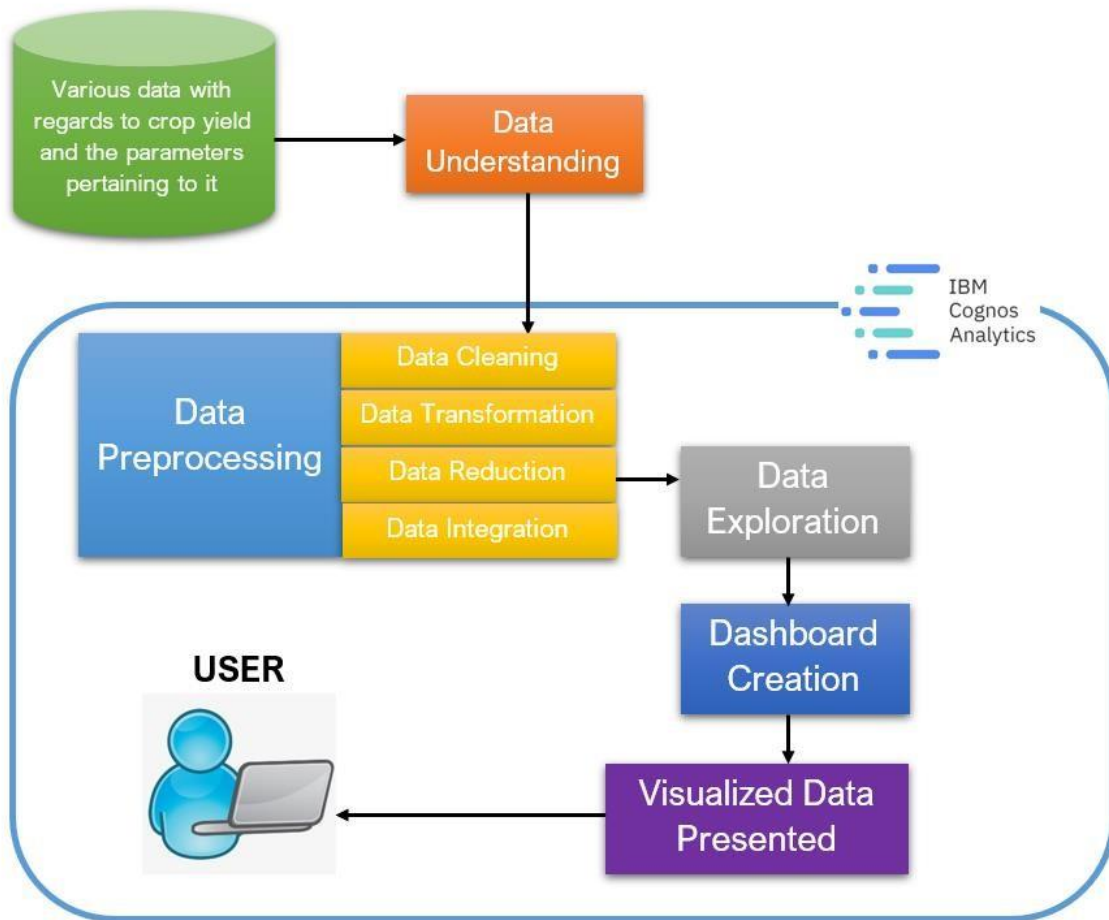
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Ease of usage along with ease in-access of tools and features
NFR-2	Security	Access to resources through two factor authentication and credentials
NFR-3	Reliability	There should be no crashes or loss of data or processes
NFR-4	Performance	High speed rendering of visualization and other readily available features
NFR-5	Availability	Should be available on demand
NFR-6	Scalability	Should be able to incorporate as many visualizations and datasets as possible

5. PROJECT DESIGN

DATA FLOW DIAGRAMS



SOLUTION & TECHNICAL ARCHITECTURE



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 Google	I can register & access the dashboard with Google Login	Low	Sprint-2
		USN-4	As a user, I can register		Medium	Sprint-1

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
			for the application through Gmail			
	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 freely use my dashboard and explore the features		High	Sprint-1
	Access of Resources	USN-7	As a user, I can use the credentials to access the resources of my application	I can securely access my resources	High	Sprint-2
Administrator	Control over the application	USN-8	I can control the users of the application		High	Sprint-2

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer	Tools	USN-9	I can perform the required tasks on the application		High	Sprint-1

6. PROJECT PLANNING & SCHEDULING

SPRINT PLANNING & 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	Arun Ponraj
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Bavyaa Sathyanarayanan
Sprint-2		USN-3	As a user, I can register for the application through Google	2	Low	Arun Bavyaa

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Low	Arun Ponraj
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Bavyaa Sathyanarayanan
Sprint-3	Dashboard	USN-6	As a user, I can freely use my dashboard and explore the features	2	High	Bavyaa Ponraj
Sprint-2		USN-7	As a user, I can use the credentials to access the resources of my application	2	High	Arun Sathyanarayanan
Sprint-3		USN-8	Performance of Data manipulations on the application	1	High	Bavyaa Arun
Sprint-3	Visualizations	USN-9	I can create dashboards with particular datasets	2	High	Ponraj Sathyanarayanan
Sprint-4		USN-10	Predictive analysis can be done	1	High	Bavyaa Ponraj
Sprint-3		USN-11	I can create stories with particular datasets	2	High	Arun Sathyanarayanan

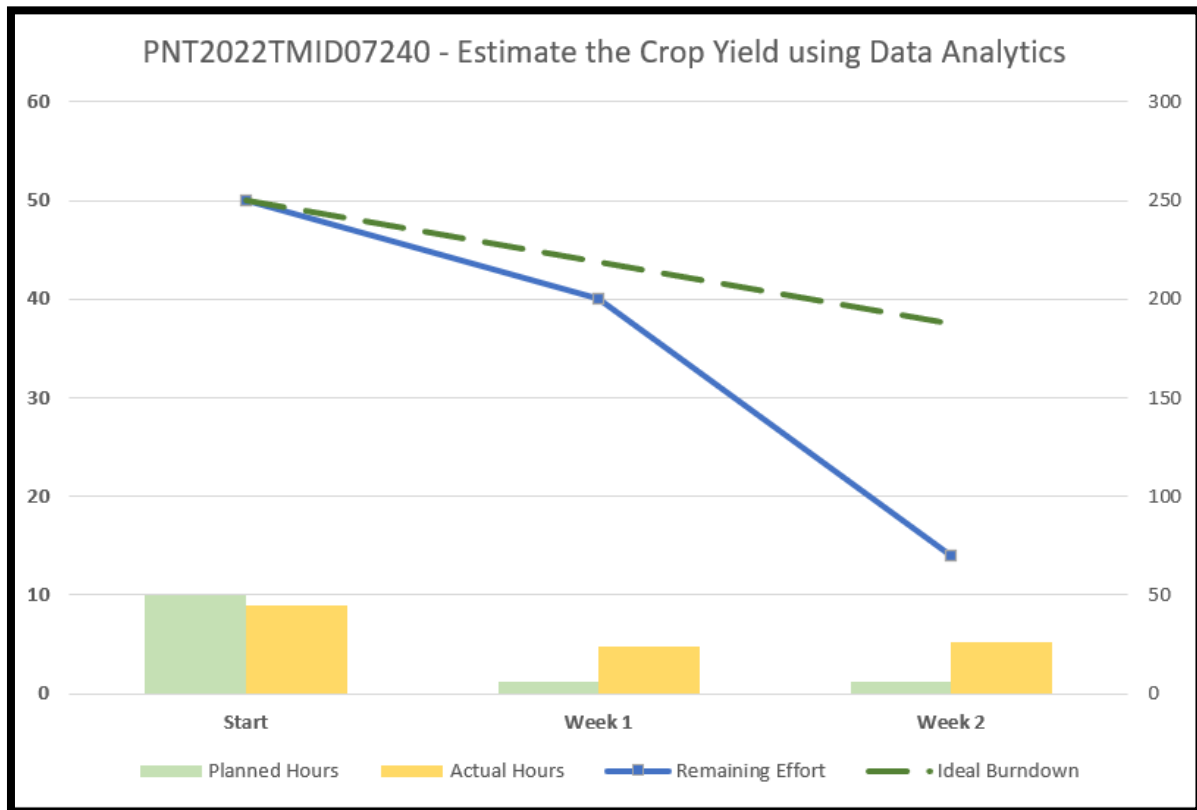
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4		USN-12	I can deliver and export reports according to the dashboards and stories created	2	High	Ponraj Sathyanarayanan

SPRINT DELIVERY SCHEDULE

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	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

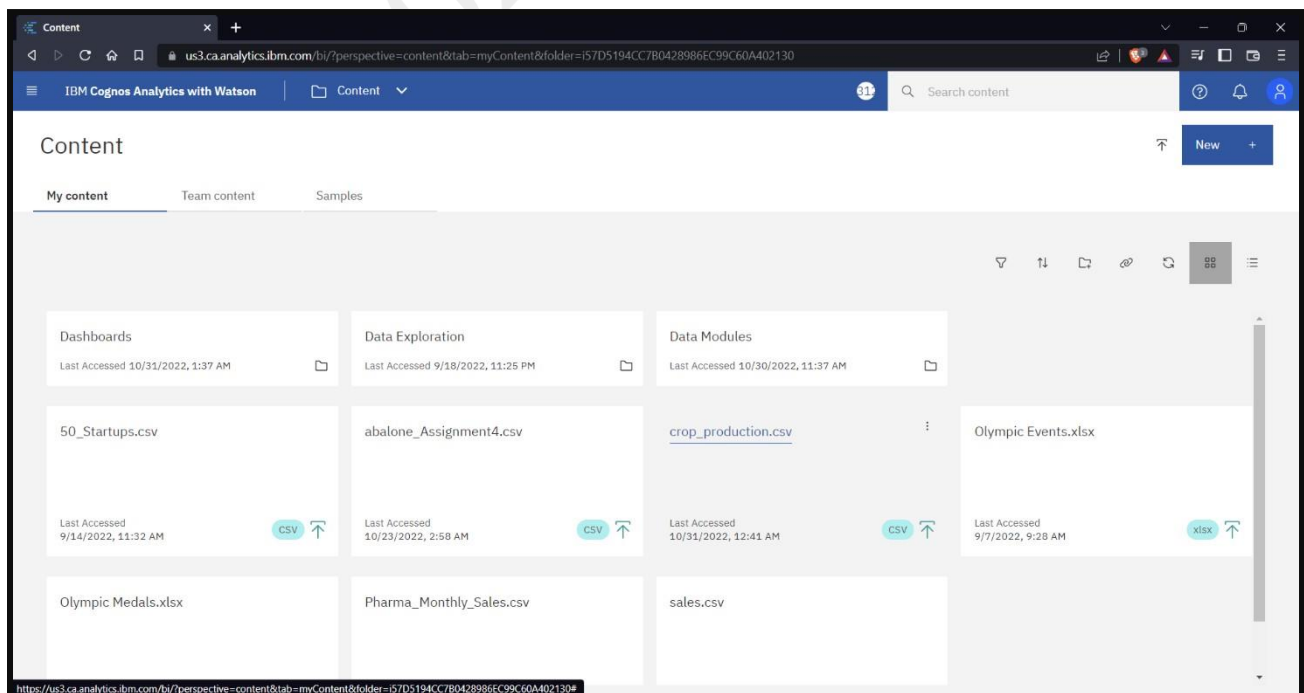
6.3. BURNDOWN CHART

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.



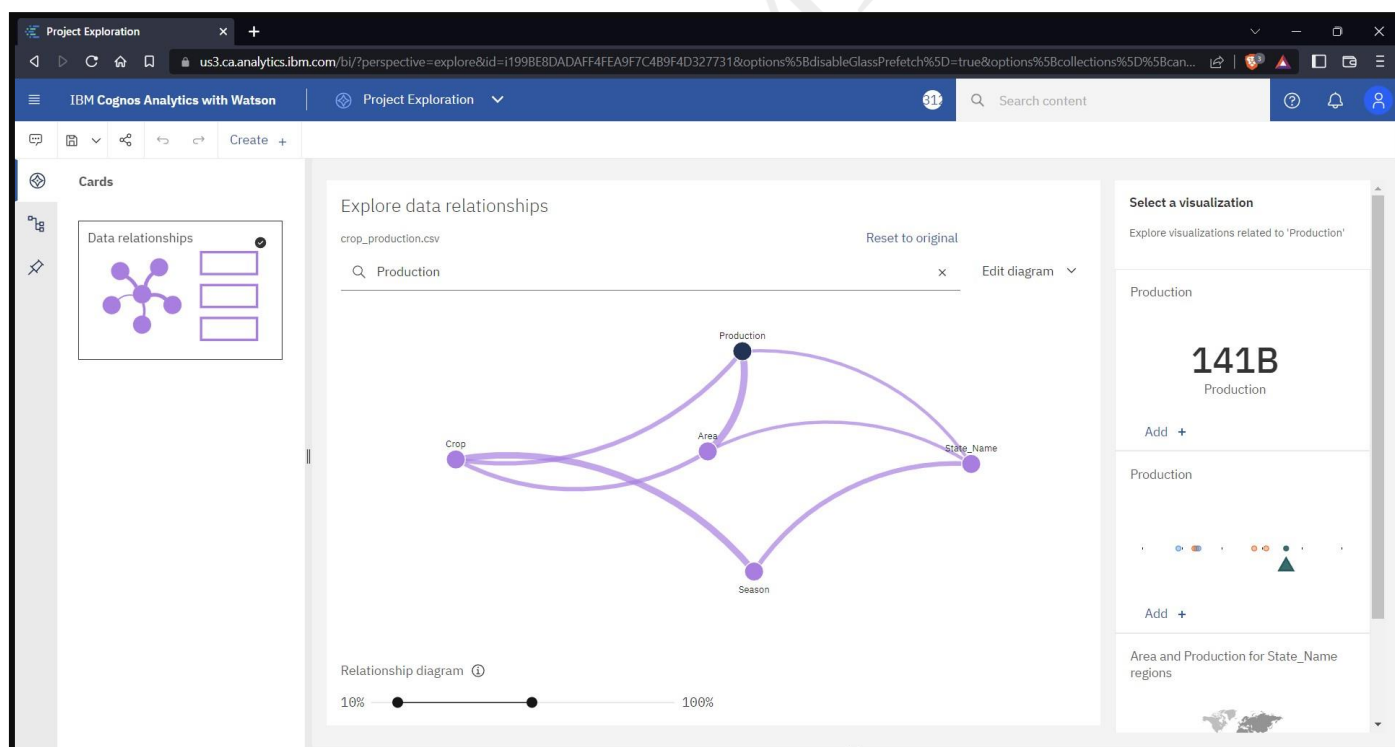
7. CODING & SOLUTIONING

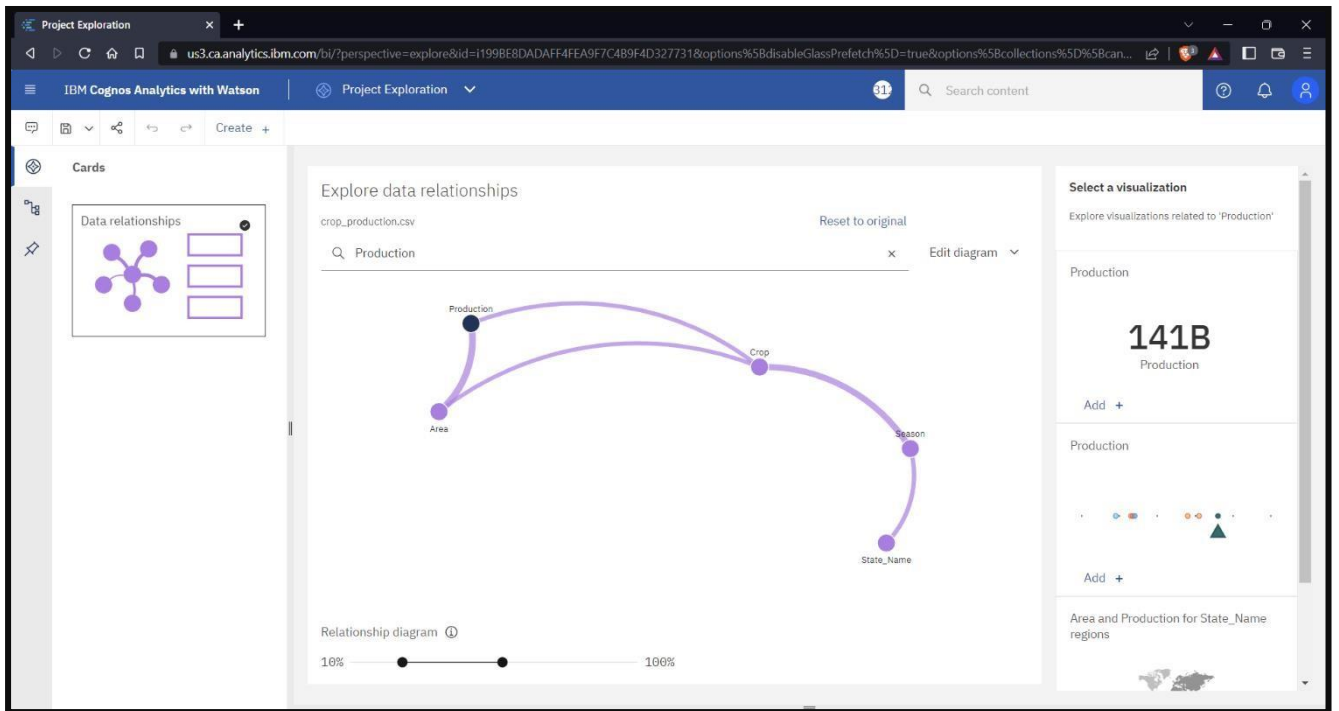
DATA COLLECTION AND PREPARATION



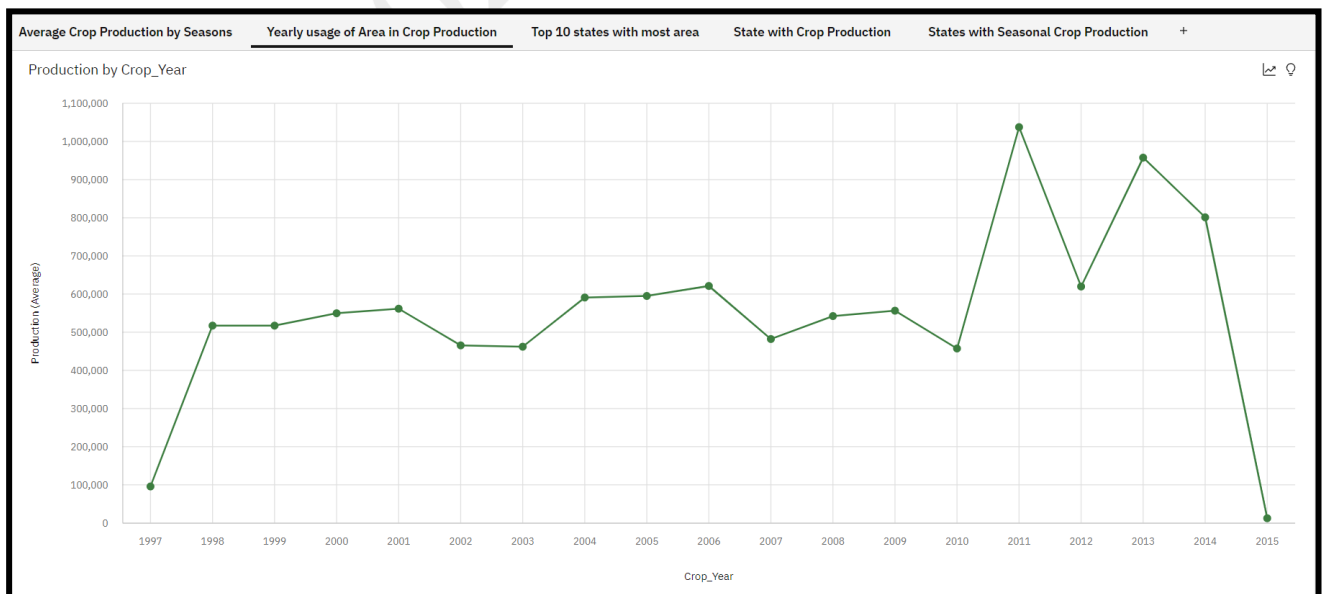
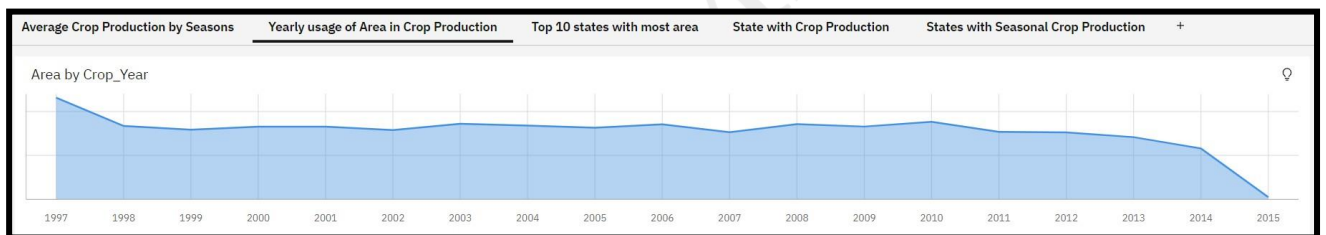
The screenshot displays the IBM Cognos Analytics web application. The top navigation bar includes the 'Project Data Module' title, a search bar, and various utility icons. The main content area is titled 'Data module' and features a left-hand navigation pane with a tree view showing the hierarchy: 'Project Data Module' > 'Navigation paths' > 'crop_production.csv'. The main pane shows a table with 8 columns: 'Row Id', 'State_Name', 'District_Name', 'Crop_Year', 'Season', 'Crop', and 'Area'. The table contains 11 rows of data, showing crop production details for various states and districts in India. The 'Season' column has values like 'Rabi', 'Whole Year', 'Kharif', and 'Post Monsoon'. The 'Area' column shows values in hectares, with some cells containing large numbers like 3540 and 650800.

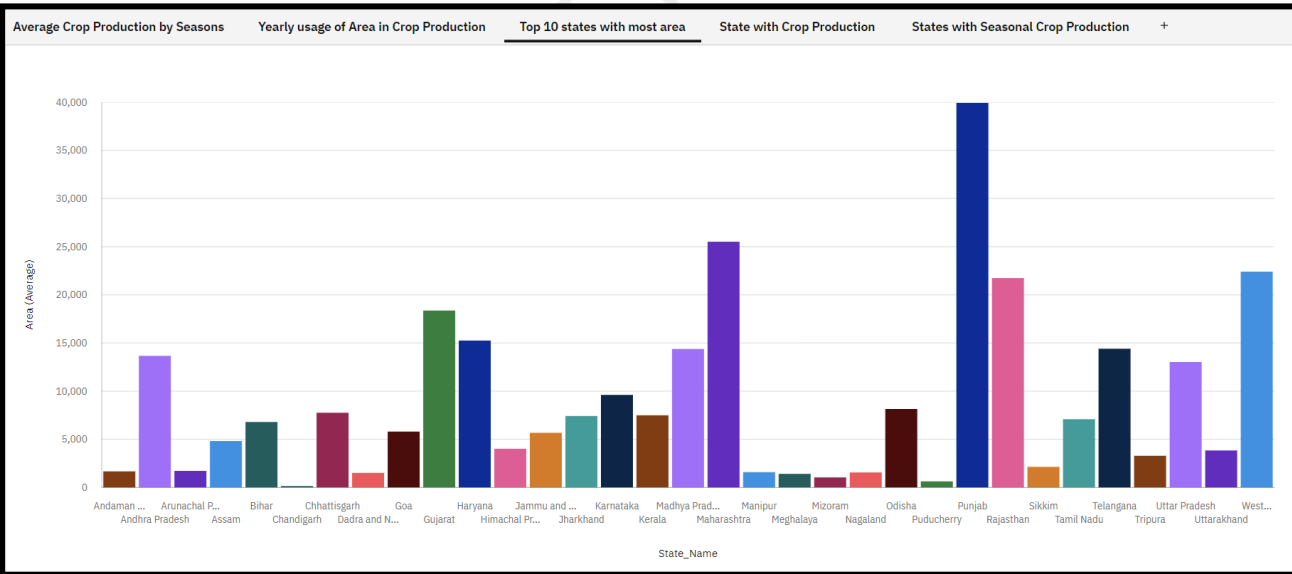
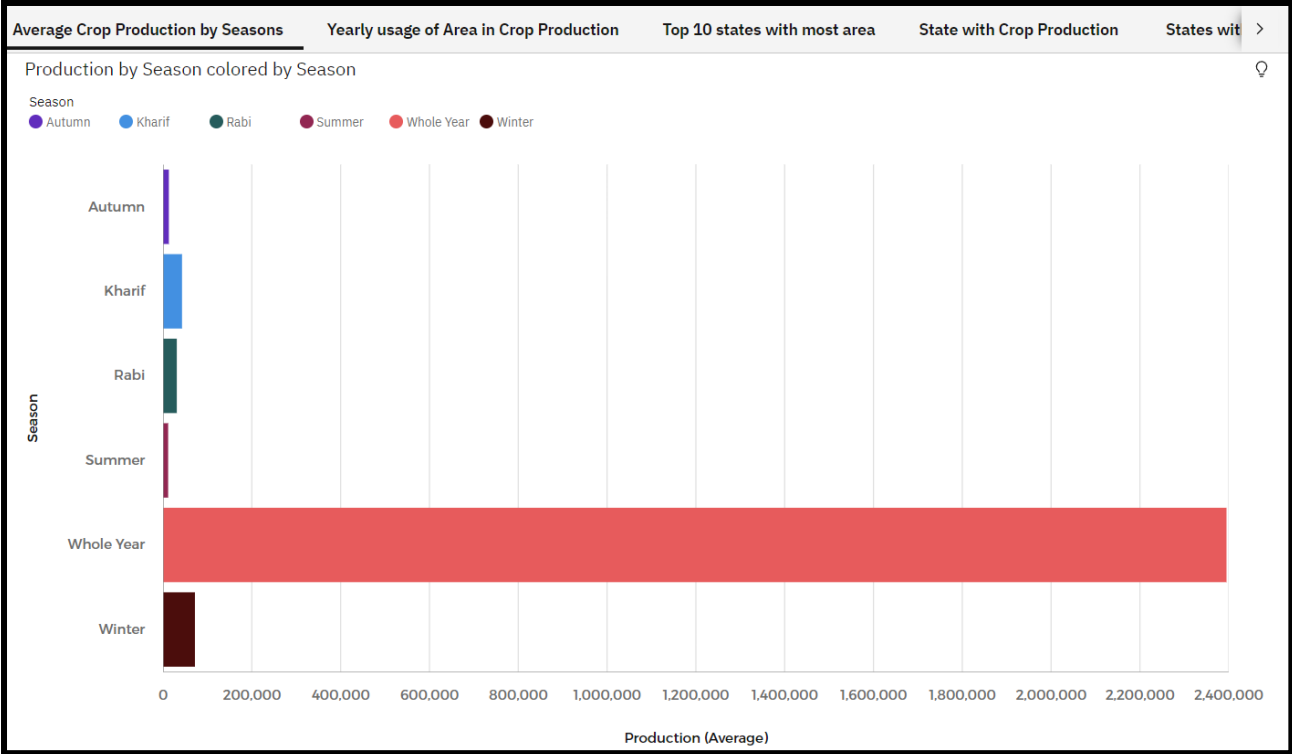
Row Id	State_Name	District_Name	Crop_Year	Season	Crop	Area
198	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	Rabi	Sweet potato	32
199	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	Rabi	Turmeric	10
200	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	Rabi	Urad	34
201	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	Whole Year	Banana	360
202	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	Whole Year	Coconut	3540
203	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	Whole Year	Tapioca	22.5
204	Andhra Pradesh	ANANTAPUR	1997	Kharif	Arhar/Tur	21400
205	Andhra Pradesh	ANANTAPUR	1997	Kharif	Bajra	1400
206	Andhra Pradesh	ANANTAPUR	1997	Kharif	Castor seed	1000
207	Andhra Pradesh	ANANTAPUR	1997	Kharif	Cotton(lint)	7300
208	Andhra Pradesh	ANANTAPUR	1997	Kharif	Dry chillies	3700
209	Andhra Pradesh	ANANTAPUR	1997	Kharif	Groundnut	650800
210	Andhra Pradesh	ANANTAPUR	1997	Kharif	Horse-gram	3300
211	Andhra Pradesh	ANANTAPUR	1997	Kharif	Jowar	10100

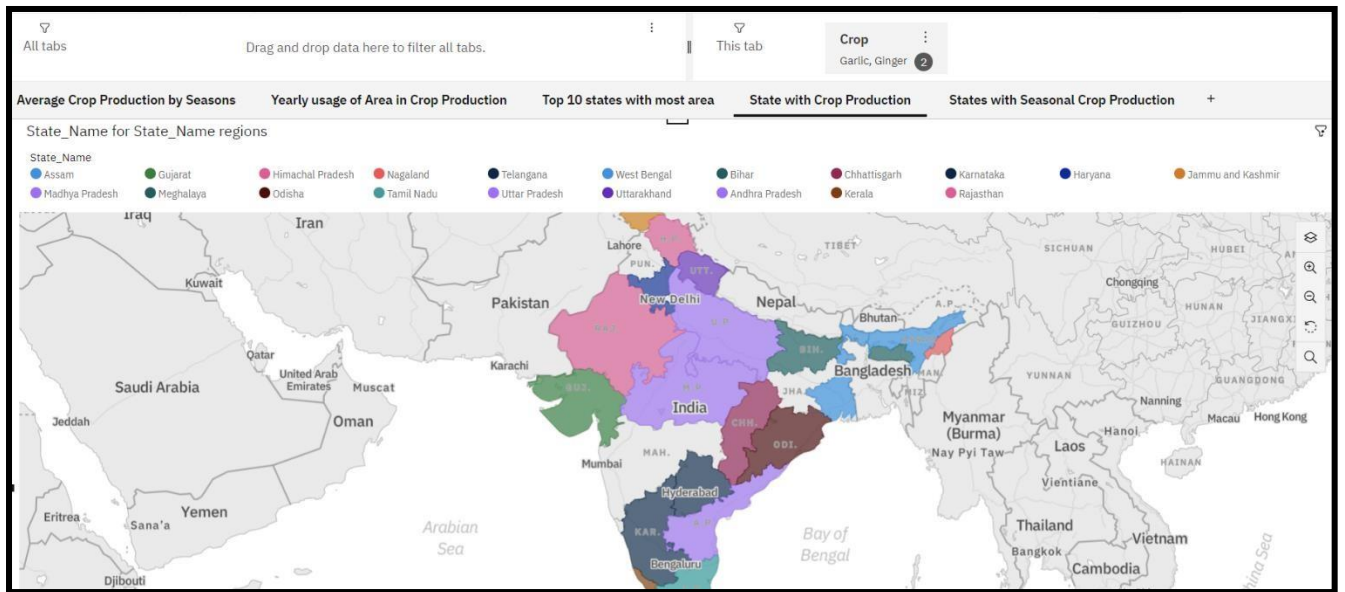




DASHBOARD CREATION





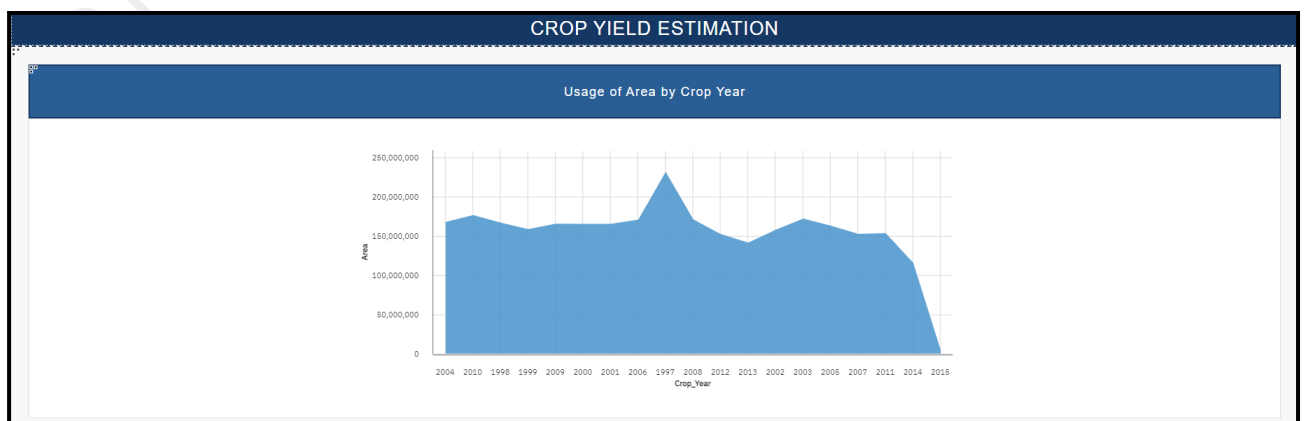


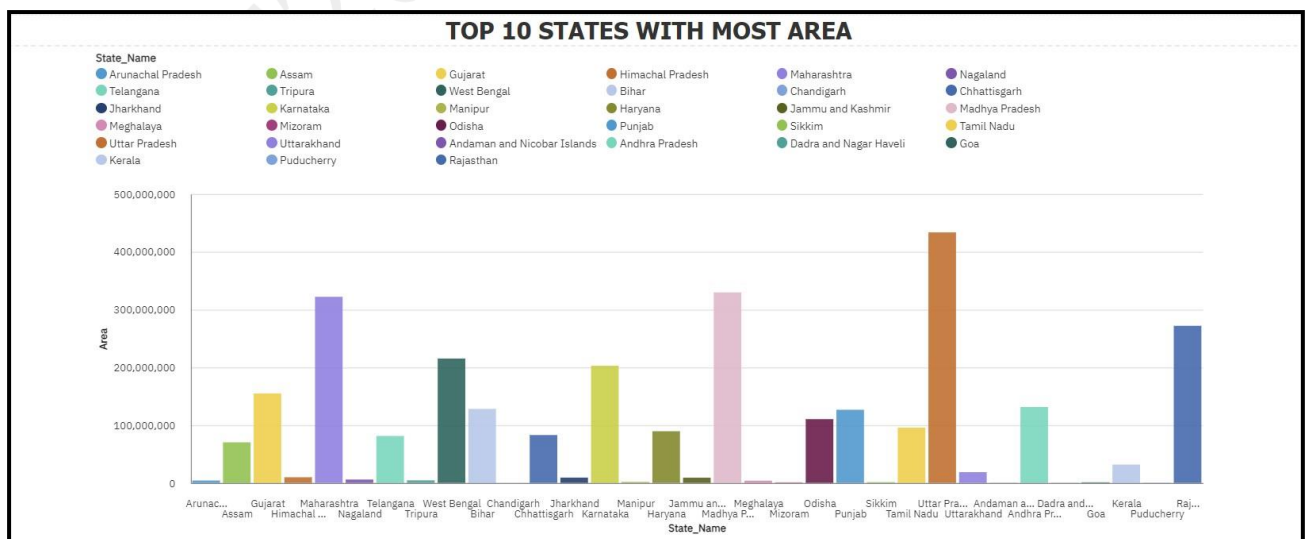
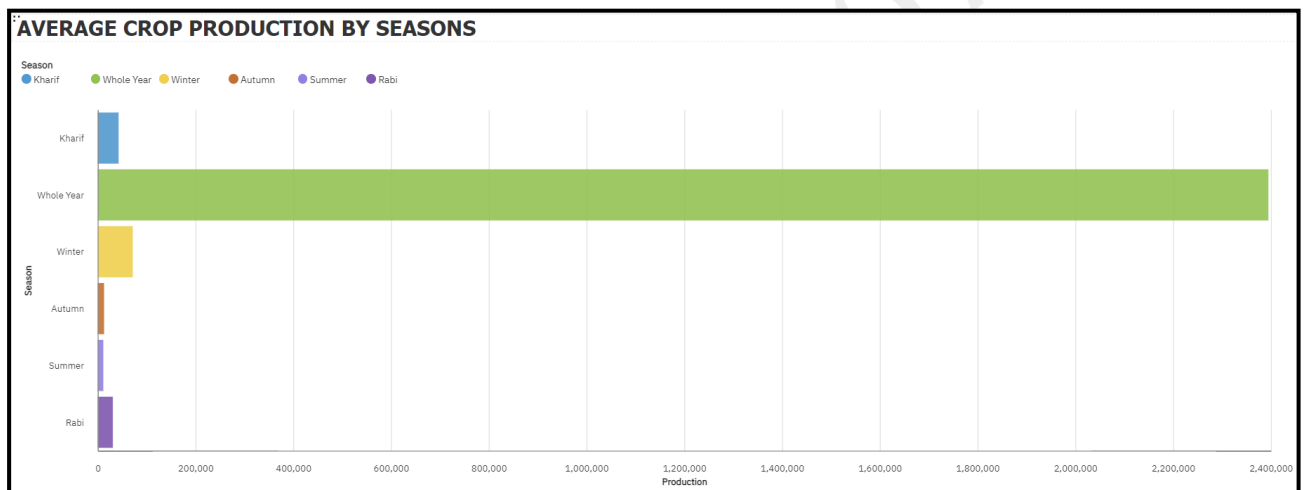
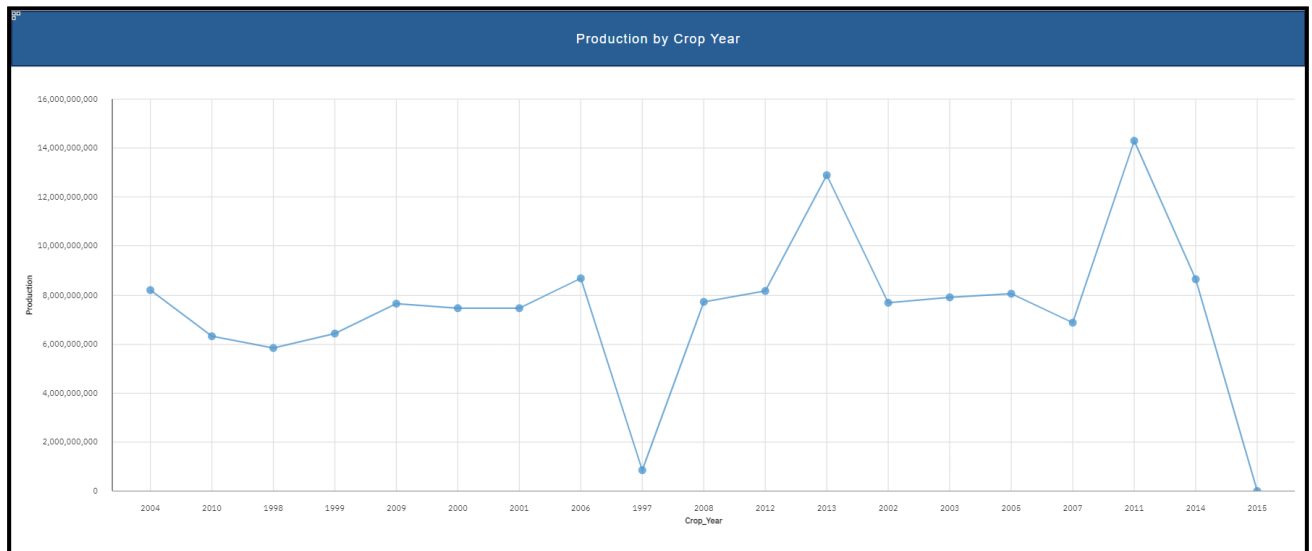
Average Crop Production by Seasons Yearly usage of Area in Crop Production Top 10 states with most area **State with Crop Production** States with Seasonal Crop Production +

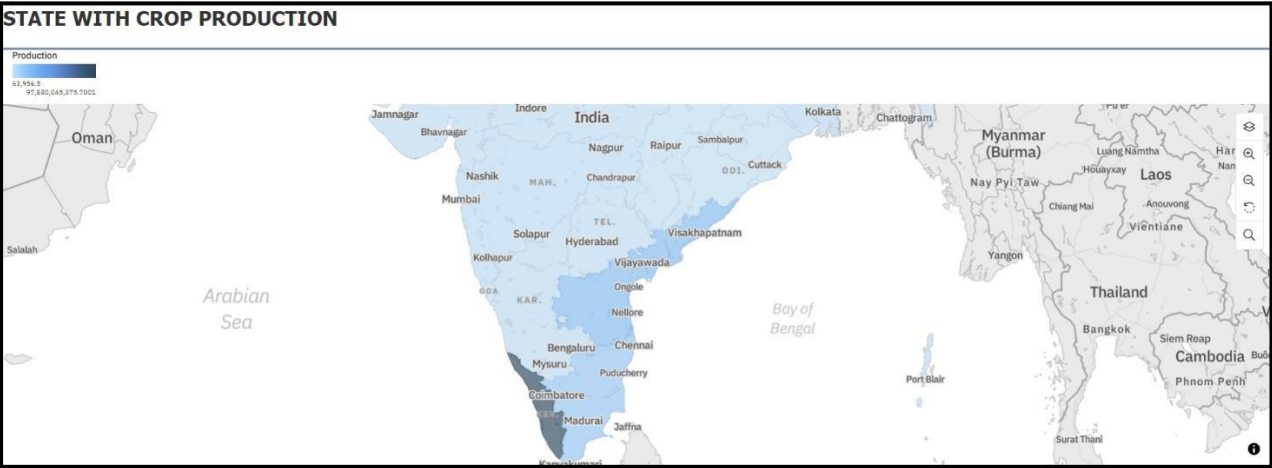
Season, State_Name and Crop

Crop	State_Name	Season
Apple	Tamil Nadu	Whole Year
Arcanaut (Processed)	Karnataka	Whole Year
Arecanut	Andaman and Nicobar Islands	Kharif
		Rabi
		Whole Year
	Andhra Pradesh	Kharif
		Whole Year
	Assam	Rabi
	Goa	Whole Year
	Karnataka	
	Kerala	
	Meghalaya	Rabi
		Whole Year
	Puducherry	Kharif

REPORT CREATION







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

STORY CREATION

Estimation of Crop Yield

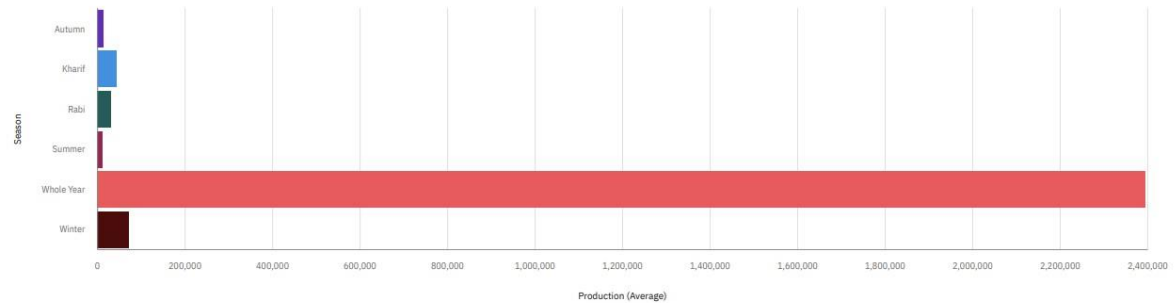
Using Data Analytics

PNT2022TMD07240

Average Crop Production By Seasons

Production by Season colored by Season

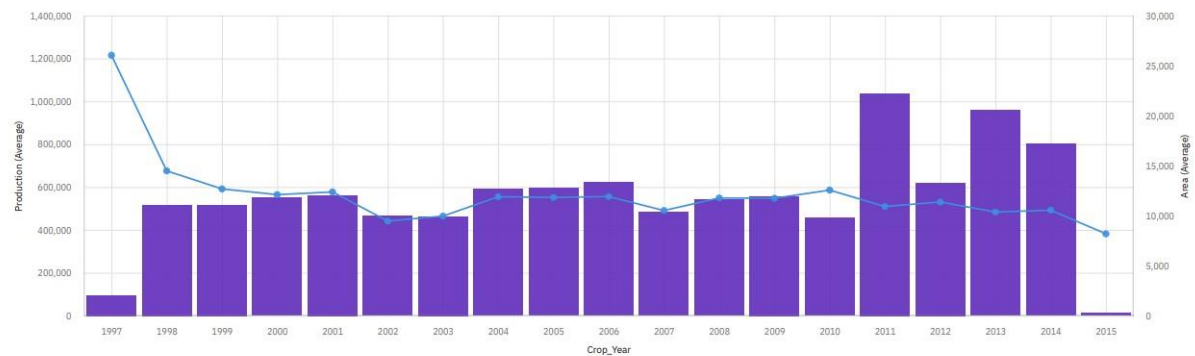
Season
● Autumn ● Kharif ● Rabi ● Summer ● Whole Year ● Winter



Yearly Usage of Area in Crop Production

Area and Production by Crop_Year

Column ● Production (Average) Line ● Area (Average)



Top 10 states with most Area

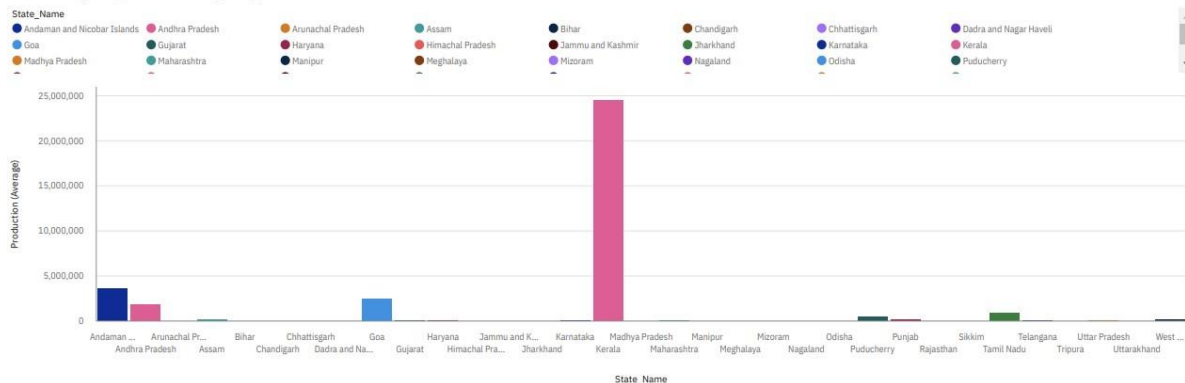
State_Name colored by Area sized by Area

Area (Sum) 12,532 433,831,890
Area (Sum) 12,532 433,831,890



State with Crop Production

Production by State_Name colored by State_Name



States with Seasonal Crop Production

Season, State_Name and Crop_Year

Crop_Year	State_Name	Season
1997	Andhra Pradesh	Kharif
		Rabi
	Arunachal Pradesh	Kharif
		Rabi
		Whole Year
	Assam	Autumn
		Kharif
		Rabi
		Summer
		Whole Year
	Bihar	Winter
		Autumn
		Kharif
		Rabi

8. ADVANTAGES

Productivity is boosted by technology. Agriculture has seen a significant boost in productivity as a result of technology; farmers can now do 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 number of 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 analyse and improve the crop yield and represent in the form of a Graphs through data visualization technique.

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

11. FUTURE SCOPE

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

12. APPENDIX

Link to the GitHub Repository:

<https://github.com/IBM-EPBL/IBM-Project-29026-1660120250>

Project Demo Video Link:

<https://drive.google.com/file/d/19ovHnY7Mzbli1AvswtFcupITbv4ntCSF/view?usp=sharing>

PNT2022TMID07240