ESTIMATE THE CROP YIELD USING DATA ANALYTICS

TEAM ID - PNT2022TMID20403

TEAM LEADER – ABISHEK G (49621911053)
TEAM MEMBER – DEVI ARUNA A (49621911006)
TEAM MEMBER – BANU PRIYA R(49621911003)
TEAM MEMBER – ARJUN D(49621911401)

ABSTRACT:

Analytics is the interpretation of data pattern that assist decision - making and performance improvement. Agriculture Data analytics in crop yield helps in analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India. IBM Cognos Analytics integrates reporting. modeling, analysis , exploration, dashboards, stories, and event management so we can understand our organisation's data , and make effective decisions. A dashboard helps us to monitor events or activities at a glance by providing key insights abd analysis about our data on one or more pages or screens. In this project, we visualize , analyse and gain most of the insights by creating a dashboard.

1. INTRODUCTION

1.1 PROJECT OVERVIEW:

In India, crop production is mostly determined by the biological and economic factors that affect a particular crop during a given season. Every nation may examine and benefit from reporting on agricultural progress through all of the seasons in terms of estimating and predicting the overall crop production. Farmers are currently under pressure to produce larger agricultural yields due to the influence of unpredictable weather fluctuations and the considerable global decline in water resources.A study was conducted to gather data on global climatic changes and the water resources that may be used to support cutting-edge and novel ways like big data analytics to retrieve the details of the previous outcomes to the forecast and estimation of crop production. According to a study, choosing and using the crop that will produce the highest crop yield under the given circumstances is helpful. Accurate crop yield forecasting helps farmers earn the greatest pricing for their crops and choose the best strategy to minimise crop damage.

1.2 PURPOSE:

Agriculture the largest economic sector, is crucial to the structure of India's socioeconomic system. Climate and economic elements, such as temperature, irrigation, cultivation, soil, rainfall, pesticides, and fertilisers, all have an impact on farming. Major input for businesses operating in this

field comes from historical data on agricultural yield. These businesses use the crop production estimation to plan supply chain decisions like production scheduling. Based on estimations of crop yield, industries like fertilisers, seeds, agrochemicals, and agricultural machinery plan production and operations like marketing. The sole method for predicting crop yield in the past was farmer experience. The integration of technology into the agricultural sector has resulted in the automation of tasks like yield estimation and crop health monitoring.

2. LITERATURE SURVEY:

1) TITLE OF PAPER: ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES AUTHOR: B Vishnu Vardhan YEAR:2015

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depending on weather conditions. Rainfall conditions also influences the rice cultivation. In this context, the farmers necessarily require 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.

2) TITLE OF PAPER: The use of satellite data for crop yield gap analysis AUTHOR: David B. Lobell YEAR:2013

ABSTRACT: Discussed the use of remote sensing technology to identify and measure the causes of yield gaps and the assess the impact on the overall crop yield. Reported very simple methodologies to measure the yield difference with respect to season, environment and the land use.

3) TITLE OF PAPER: Toward large-scale crop production forecasts for global food security AUTHOR: G. Badr L. J. Klein YEAR: 2016

ABSTRACT: Predicting crop production plays a critical role in food price forecasting and mitigating potential food shortages. Crop models may require parameters from, for example, weather, crop genotype, farm management, and soil. Sources for these data are often found in very different places. Researchers spend a significant amount of time to collect and curate them.

4) TITLE OF PAPER: Robust Model Predictive Control of Irrigation Systems With Active Uncertainty Learning and Data Analytics AUTHOR: Chao Shang , Wei-Han Che YEAR: 2019

ABSTRACT: We develop a novel data-driven robust model predictive control (DD approach for automatic control of irrigation systems. The fundamental idea is to integrate both mechanistic models, which describe dynamics in soil moisture variations, and data-driven models, which characterize uncertainty in forecast errors of evapotranspiration and

precipitation, into a holistic sysem framework.

5) TITLE: Migration-Based Online CPSCN Big Data Analysis in Data Centers AUTHOR: Xin li, Liangyuan Wangi , Zhen Lian YEAR: February 28, 2018

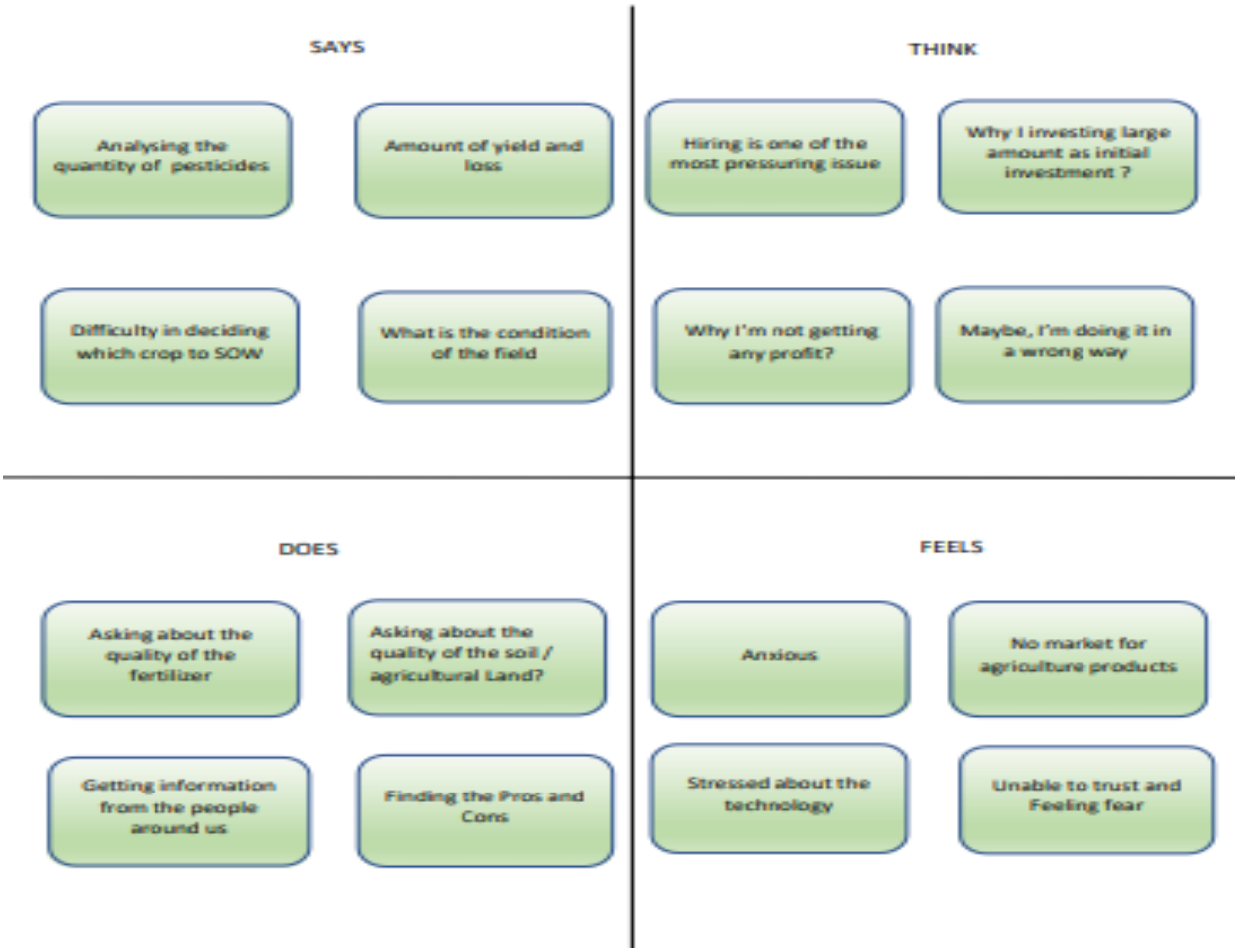
ABSTRACT: It is critical to schedule online data-intensive jobs effectively for various applications, including cyber-physical-system and social network system. It is also useful to support timely decision making and better prediction. In this paper, we investigate the online job scheduling problem with data migration for global job execution time reduction.

2.2 PROBLEM STATEMENT DEFINITION:

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

3. IDEATION & PROPOSED SYSTEM:

3.1 EMPATHY MAP CANVAS:



3.2 IDEATION & BRAINSTORMING:

IDEA 1:

In India, Agricultural yield is mainly depends on whether condition. A vast amount of agricultural data is created continuously. This leads to the big data era has brought with it agricultural data. To predict the future crop productivity ,an analysis needs to be made in order to help the farmers to maximize the crop yield production. By utilizing electronics devices for data collection benefits from smart technologies. So in our project using technology like Data Analytics and Machine learning we'll analyse the agriculture data and obtain the results. For such kind of data analytics in crop prediction, there are several different algorithms are available by using that we can analyze the data and predict the result. By using algorithm like K means clustering, Apriori Algorithm, Naïve Byes Algorithm and with the help of inter-relation between those, we can analyze the data and get the result. This result will be provided to farmers for a higher crop yield productivity.

IDEA 2:

Agriculture is very important for humans,it is one of the basic needs in human life. When we consider india and indians ,majority of the Indian population is in to agriculture. There are barriers to boosting agriculture yield because of change in like the factors like climatic changing, geographical conditions, economic conditions are to be considered for the impact on the production of the crops. Before planting, farmers need to know about the crop yield so that they can increase the productivity.

Crop yield prediction is important to predict the yield. By using Maching learning technology we can predict Agriculture crop yield. WEKA a java based dialect programming for less challenging assistance with information data sets, assigning design outcomes tool was applied for dataset processing and the overall methodology includes 1) Pre-Processing the dataset 2) building the predicition model utilizing WEKA 3) Analyzing the outcomes.

IDEA 3: Analyzing the yields of crop is necessary to update the policies to ensure food security. A research group conducted a study with the aim in suggesting a novel data mining method to predict the yields of crop depends on agricultural big data analytics methodologies, which were progressively contrast with conventional data mining methodologies in the process of handling data and modeling designs. Nearest neighbors modeling is one such novel data mining technique which works on the results collected based on data processing structures form the farmers and suggest a well unbiased result on the base of accuracy and prediction time in advance. Simulation models based on field experiment are valuable technologies for studying and understanding crop yield gaps, but one of the critical challenge remain with these methods is scaling up of these approach to assess the data collated between different time intervals from the broader geographical regions. Satellite retrieved data have frequently been revealed to present data sets that, by itself or in grouping with other information and model designs, can precisely determine the yields of crop in agricultural lands.

The yield maps developed shall provide an unique opportunity to overcome both spatial and temporal based scaling up challenges and thus improve the ideology of crop yield gaps prediction. First method works closely with the constructive maps representing the average crop yields, it can be used directly to accesses specific crop yield influencing factors for further studies whereas the second method use the remote sensing technology to retrieve the data for providing the useful information regarding the crop yield prediction and estimation.

IDEA 4:

The main objective is collecting agricultural dataset which can be used to analyzed for useful crop yield forecasting. To predict the crop yield with the help of data mining technique, and it is also helps the farmer to choose the most suitable crop, thereby improving the value and gain of the farming area. Firstly the data set should collect and it is subjected to preprocess for noise removing and computational methods. From that dataset, it is subjected to Feature selection for make a predictive modeling. In this proposed approach it is mainly focused on Regression Techniques. Various regression analysis should be performed and it was compared and tested. Regression analysis is a form of predictive modeling technique which investigates the association between a dependent and independent variable. This technique is used for forecasting, time series modeling and discovers the causal effect relationship between the variables. Regression analysis indicates the significant relationships between dependent variable and independent variable and it indicates the strength of impact of multiple

independent variables on a dependent variable. Yield gaps can be defined as the difference between the expected crop yields with respect to the actual crop yield and accurate, spatially unambiguous awareness and information about the yield gaps is necessary to achieve sustainable amplification of agricultural yields.

3.3 PROPOSED SOLUTION:

S.No.	Parameter	Description
1.	Problem Statement (Problemto besolved)	It is a well-known fact that in India, agriculture employs the bulk of the population (about55%). There are obstacles to expanding crop production in Indiabecause of weatherchanges. The task of achieving desired crop yield goals has grown difficult. A vast amountof agricultural data is created continuously. As a result, the big data era has brought with it agricultural data. Utilizing electronic devices for data collection benefits from

		smart technologies.
2 .	Idea / Solution description	Using technologies like data analytics and machine learning, we will analyse and mine this agricultural data in our project to obtainrelevant results that will be provided to farmers for a higher crop output in terms of productivity and efficiency Also to increase agricultural productivityand decrease the waste.
3 .	Novelty / Uniqueness	● Identifying appropriate problem or conditions and soil typesfor each cropby visually showingthe relationship betweennumerous parameters.

		• Improved and accurate display
4.	Social Impact / CustomerSatisfaction	The customer will be able to identify the perfect circumstances for growth so they may create strategies to increase crop yield and revenue.
5.	Business Model (RevenueModel)	Numerous factors, including soil fertility, climate, rainfall, and others, will have an impact on crop output, which willhave an impacton revenue. Crop yieldsand income are increased through examination of eachfactor's ideal conditions..
6.	Scalability of the Solution	Crop yieldcan be boosted by using theknowledge gained through the examination of many parameters, suchas weather patterns and soil types.

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.

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) The project's target audience are farmers.	6. CUSTOMER CONSTRAINTS High adoption costs and concerns about safety. Ignorant of agricultural applications. Use it with consideration for climate change	5. AVAILABLE SOLUTIONS AS Farmers can solve the problem by using their past experience in the crop yield By changing the fertilizer, pesticides or planting

			crops, soil.
Focus on J&P, tap into BE, understand RC	2. JOBS-TO-BE-DONE / PROBLEMS Problem: Agriculture sector is struggling to increase the productivity of crop in india. Due to variation in climate, there exist bottlenecks for increasing the crop production for farmers. Jobs-to-be-done: In the fast moving world, by using technology we can increase the crop yield production.	9. PROBLEM ROOT CAUSE Agriculture is very important for human survival. In india majority of population is into agriculture. The root casue for this problem is climate and economic	7. BEHAVIOUR A remote location with Good internet connectivity can prevent transmission speeds. Customer should have the knowledge about the technology

		factors like temperature, irrigation,soil, rainfall, pesticides,drought	which is used to predict the crop yield.
	3. TRIGGERS Seeing their neighbour using the crop yield prediction techonolgy, Create possibilities to help individuals in underdeveloped countries escape poverty. Smart farming lessens its impact on the environment.	10. YOUR SOLUTION Our solution is to predict the crop yield using data analytics, For this we are going to use some algorithms in data analytics and Machine learning by which we can	8.CHANNELS OF BEHAVIOUR ONLINE The farmer will receive the data sent through the application and sensor data. OFFLINE Farmers take control actionto

		predict the crop yield and obtain the results.	keep an eye on their land through quick response.
	4. EMOTIONS: BEFORE / AFTER EM BEFORE: Customers have to face loss in their agriculture field, due to this they may feel Anger,depression,restlessness,disbelief, Sadness. AFTER:Confidence,stability,happy,peace.		

4. REQUIREMENT ANALYSIS:

4.1 **Functional requirements:**

Functional requirement	**Description**
Registration	Registers a user throughregistrati on form
Google authentication	New user can get added even byusing a google account
Login	Lets the registered user to login to the portal
Take in the required data	This takes in the required data from the user
Estimation	A prediction of crop yield is donebased on the current data
Analysis	An analysis is done on the given datatogain useful insights

Non-functional requirements:

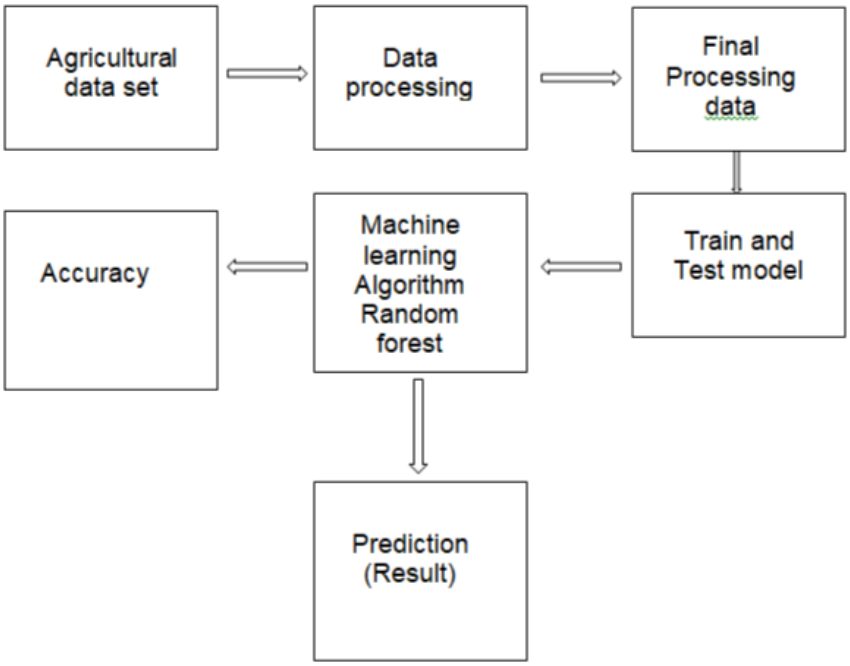
Non-functional requirement	Description
Performance	The software should provide usgood performance
Reliability	The UI should be user friendly andeasily understandable
Availability	It should be available for access atany time from anywhere
Scalability	The software should be flexible andother developers must be able to improve its capabilities

5. PROJECT DESIGN:

5.1 DATA FLOW DIAGRAM:

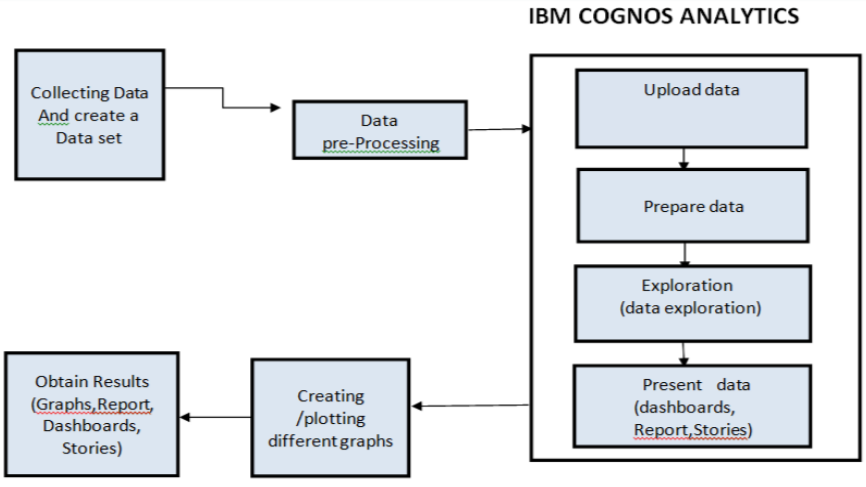
Project flow describes a preset sequence of activities required to plan, produce, deliver and maintain project product, along with information, materials, and resources required by the project. Project flow is a convenient way to define and plan projects.

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



5.2 SOLUTION AND TECHNOLOGY ARCHITECTURE:

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



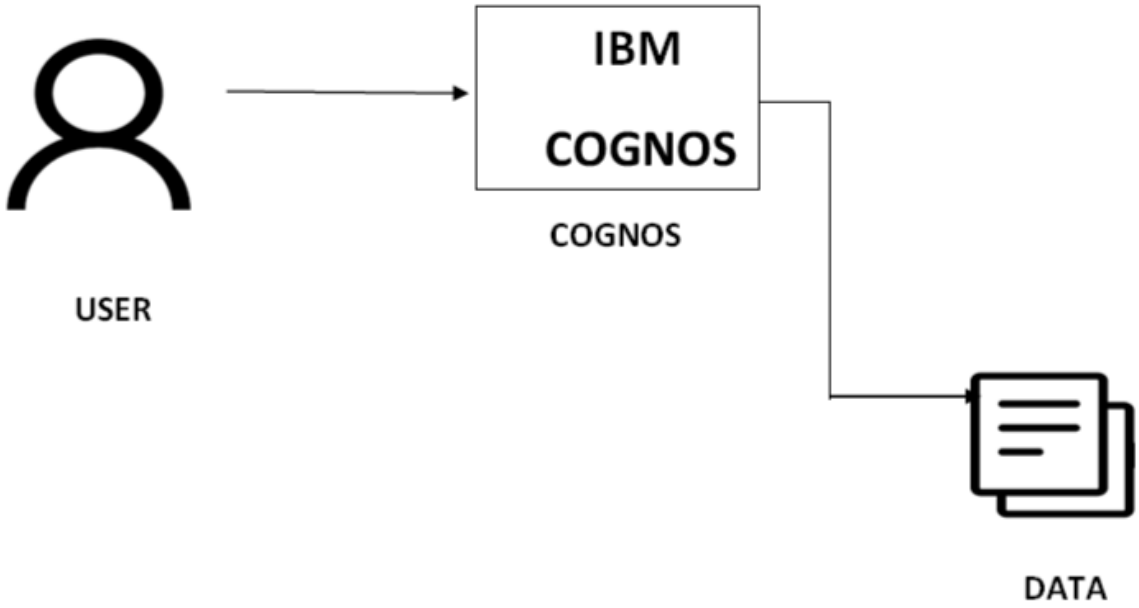


Table-1 : Technologies

S.No	Components	Description	Technology
1.	Tool used	The tool whichis used to predict	IBM Cognos analytics
2.	Application login1	Installation of software/create account	Ibm cloud, ibm cognosanalyti cs
3.	Application logic-1	Using the application's adminlogin information.	IBM Watson STT Service.
4.	Application logic-2	Logging into the applicationasa	IBM Watson Assistant.

		merchant.	
5.	Database	A database contains information regarding the crops.	Spreadsheet.
6.	Cloud Database	To store data, IBM Watson cloudis employed.	IBM Cloud etc.
7.	External API-1	Use of an external API and its intended use.	e.g., IBM Weather API.
8.	Machine Learning Model	A machine learning model's intended use.	Object Recognition Model, etc.
9.	Infrastructure (Server /Cloud)	Local Server Configuration for an Application Deployed on a Local System or a Cloud. LocalServer Configuration, Cloud Server Configuration.	Local, Cloud Foundry, Kubernetes, etc.

5.3 USER STORIES:

User StoryNumber	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'swhich is uploaded in the IBM cognos
USN-5	Creating the datavisualization chart
USN-6	Creating a dashboard
USN-7	Export the analytics

6. PROJECT PLANNING AND SCHEDULING:

6.1 SPRINT PLANNING AND ESTIMATION:

Sprint	Functional Requirement (Epic)	User Story Number	User Story/ Task	Story Points	Priority	Team Members
Sprint-1	Working with the da	USN-1	Understanding the data set .	10	Medium	Abishek

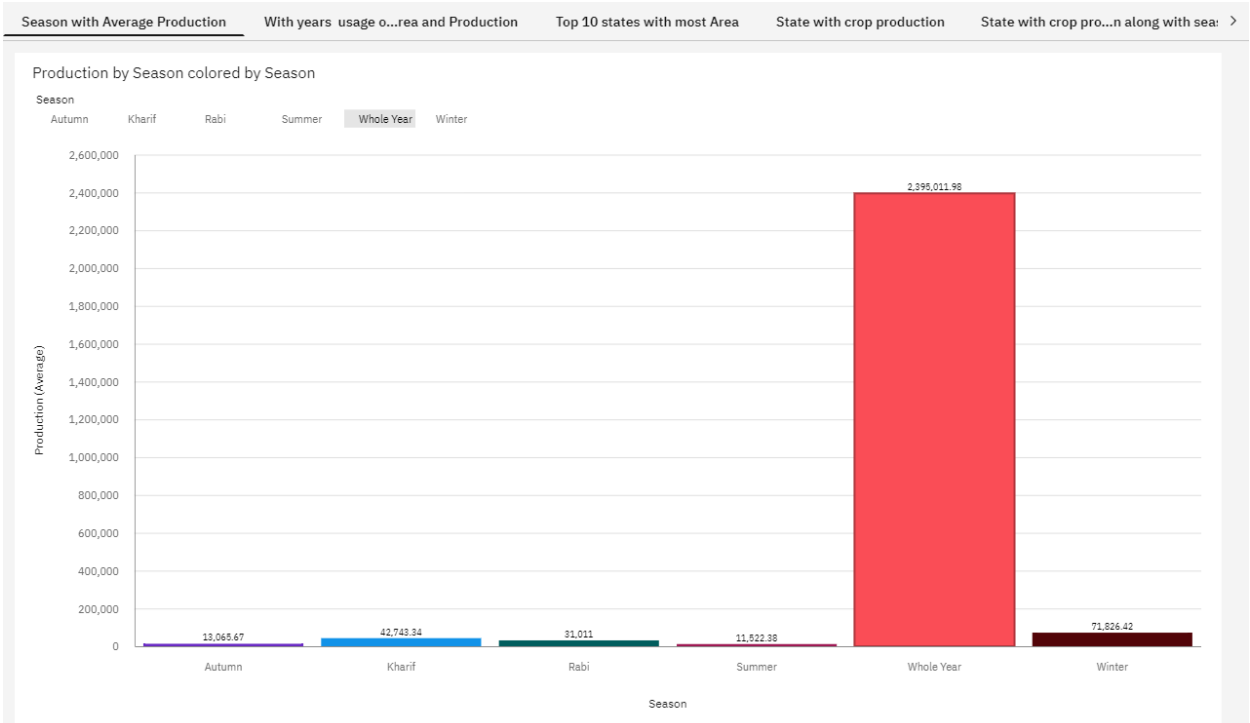
	ta set					
Sprint-1	Working with the data set	USN-2	Loading the data set.	10	High	Abishek
Sprint-2	Prepare the data	USN-3	Convert the data into required format	10	Medium	Banu priya Arjun
Sprint-2	Data exploration	USN-4	Explore the data's which is uploaded in the IBM cognos	10	Medium	Banu priya Arjun
Sprint-3	Data visualization	USN-5	Creating the data visualization chart	10	High	Banu priya Arjun

Sprint-3	Dashboard	USN-6	Creating a dashboard	10	High	Devi Aruna
Sprint-3	Visualization	USN-7	Estimation of accuracy using random forest algorithm	10	High	Abishek
Sprint-4	Export	USN-8	Export the analytics	10	High	Abishek

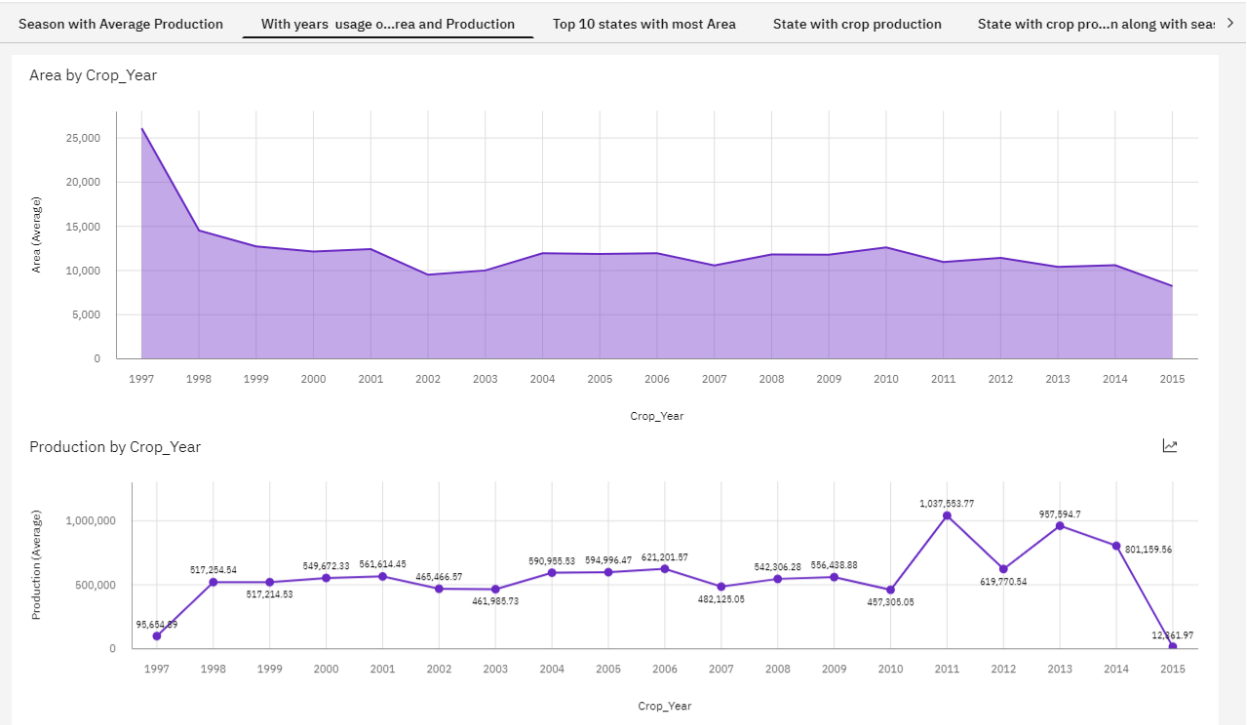
7.RESULT:

A)DATA EXPLORATION:

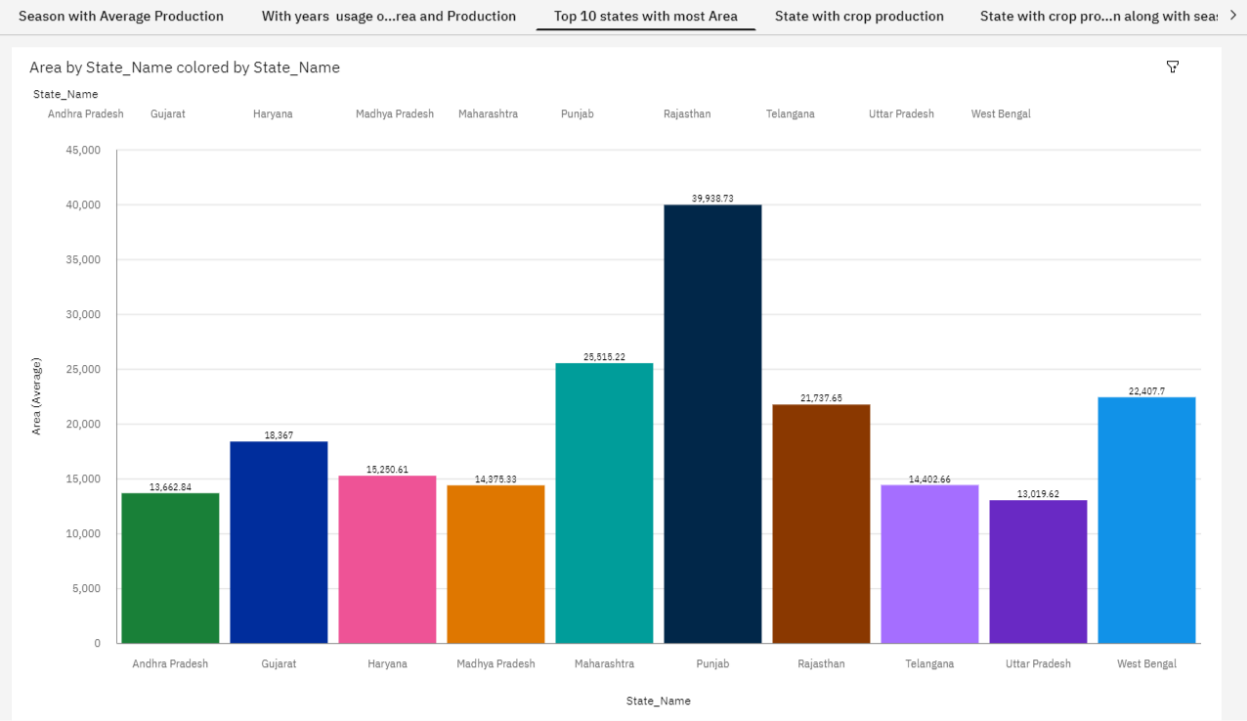
1)Season with average production:



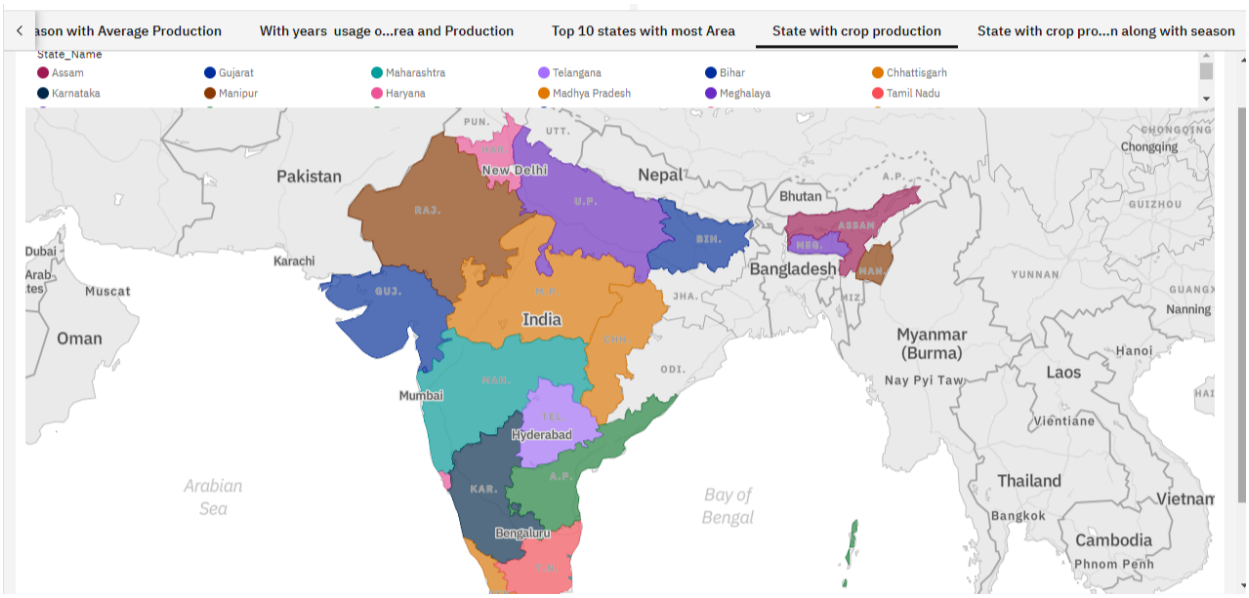
2)with years usage of area and production:



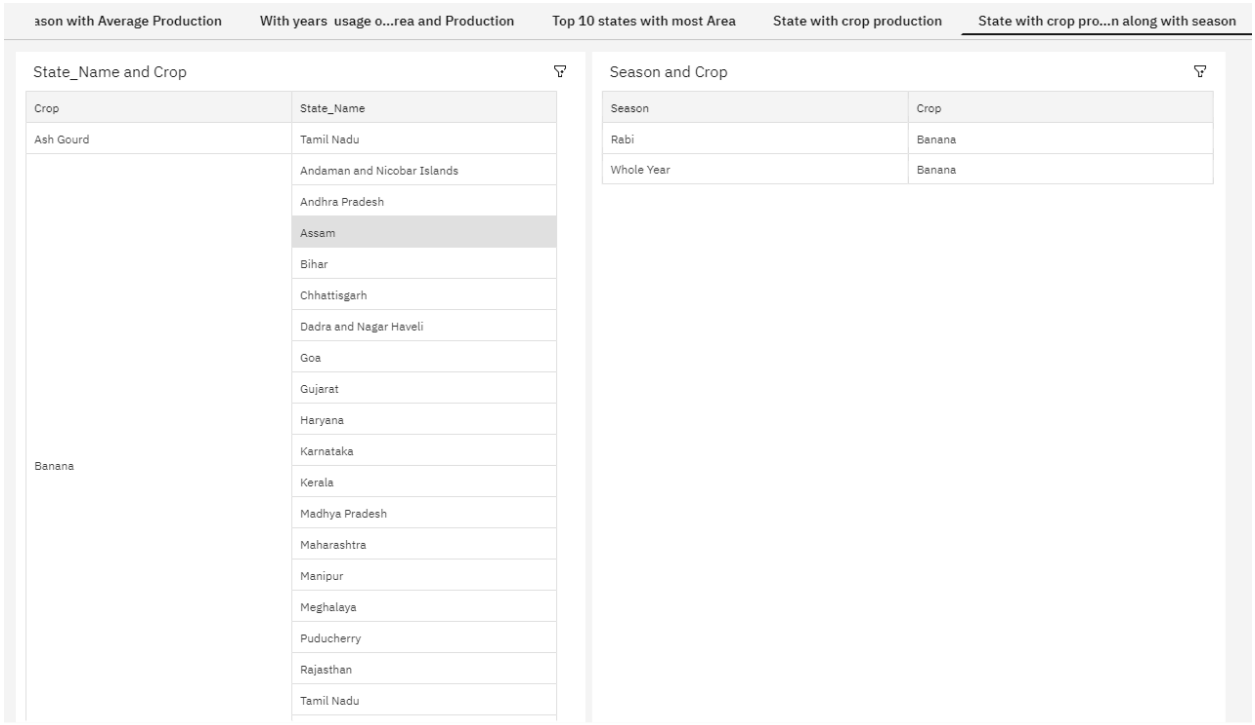
3)Top 10 states with the most area:



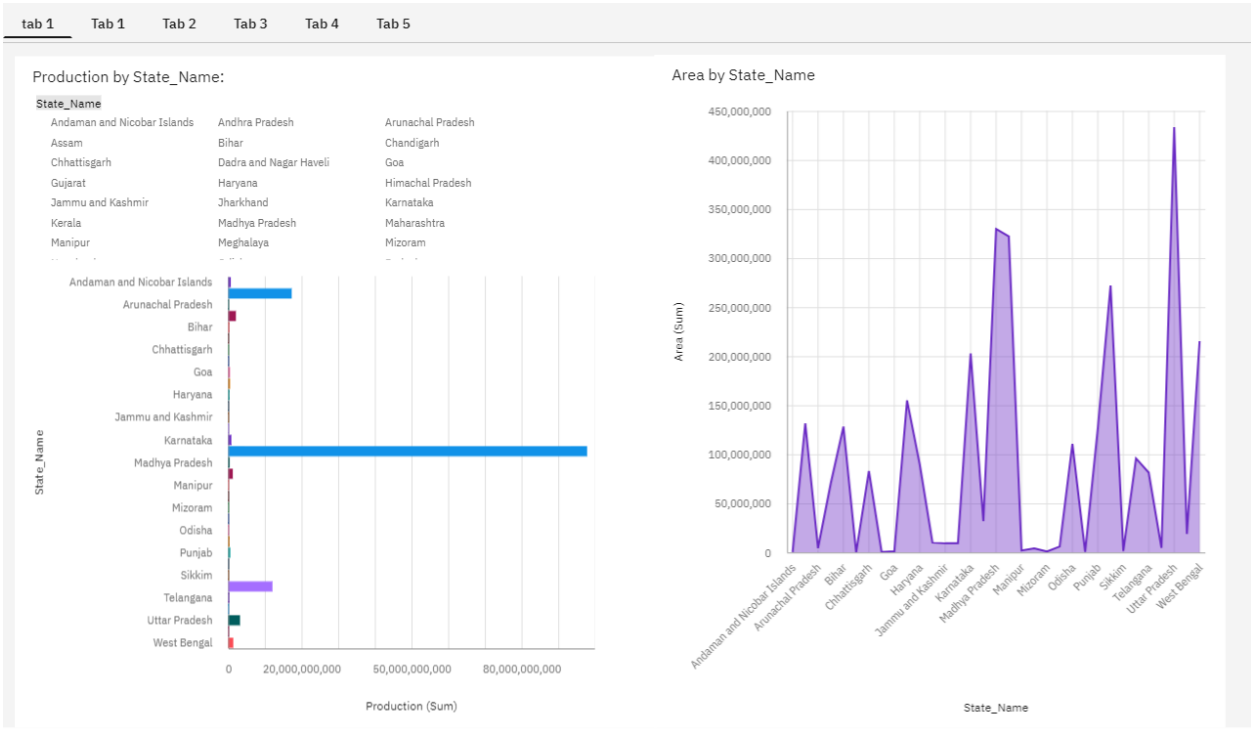
4)State with crop production:

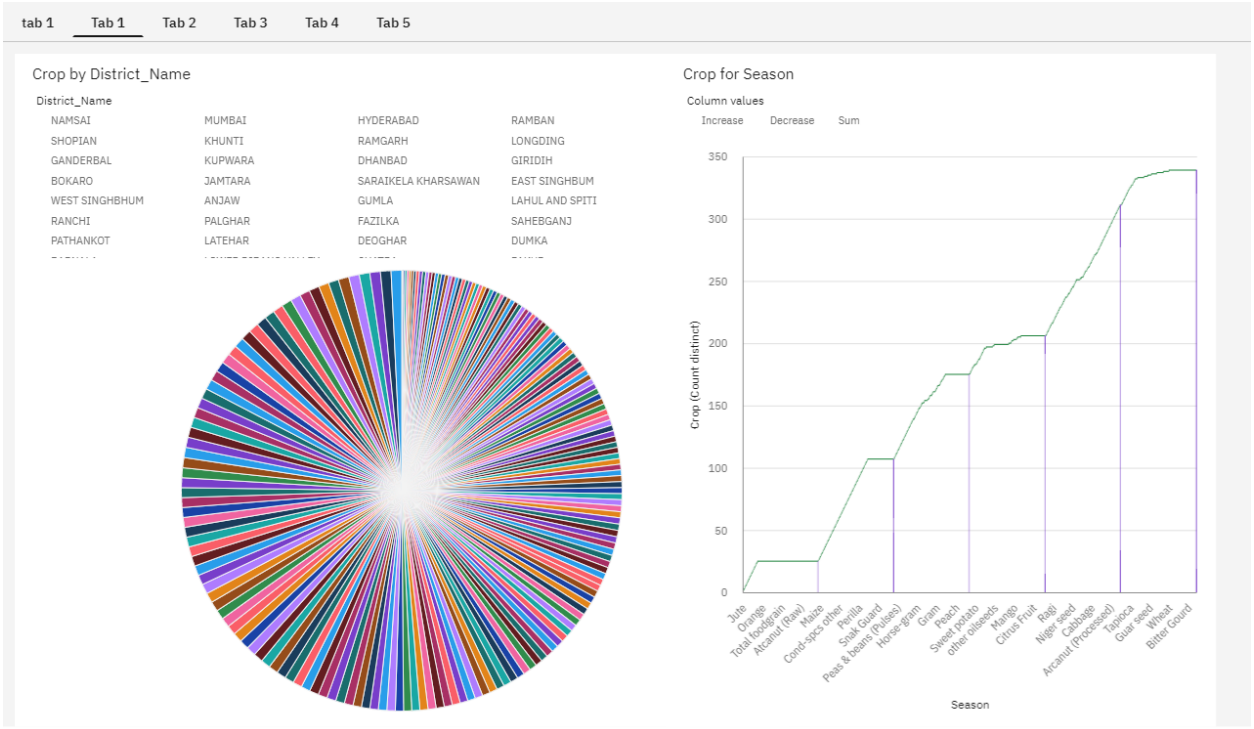


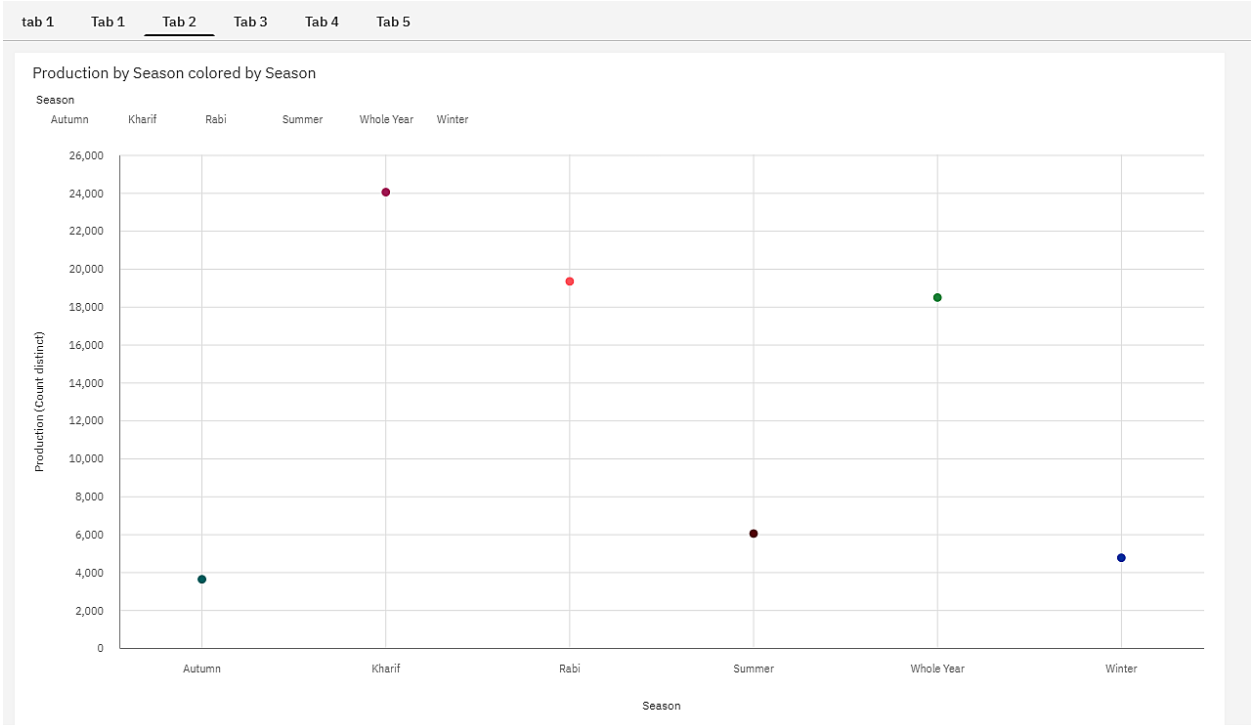
5)State with crop production along with season:

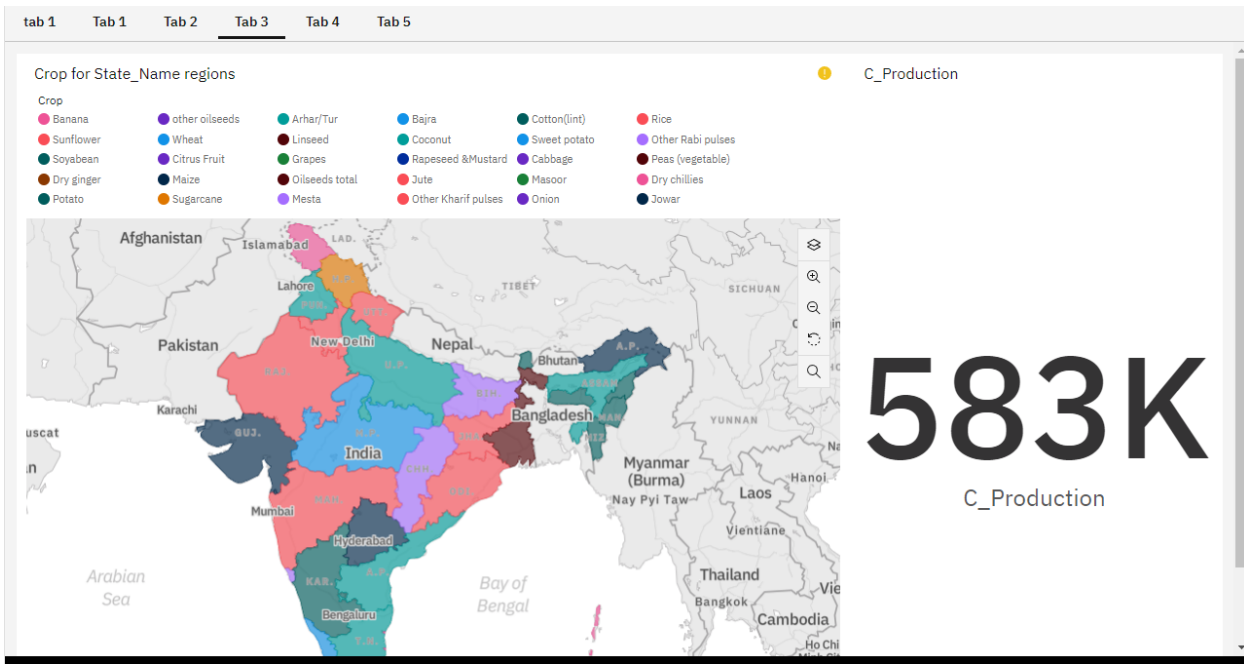


B)DASHBOARD:


tab 1
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Production by State_Name:
State_Name
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
State_Name
Production (Sum)
0
20,000,000,000
50,000,000,000
80,000,000,000
Area by State_Name
Area (Sum)
State_Name

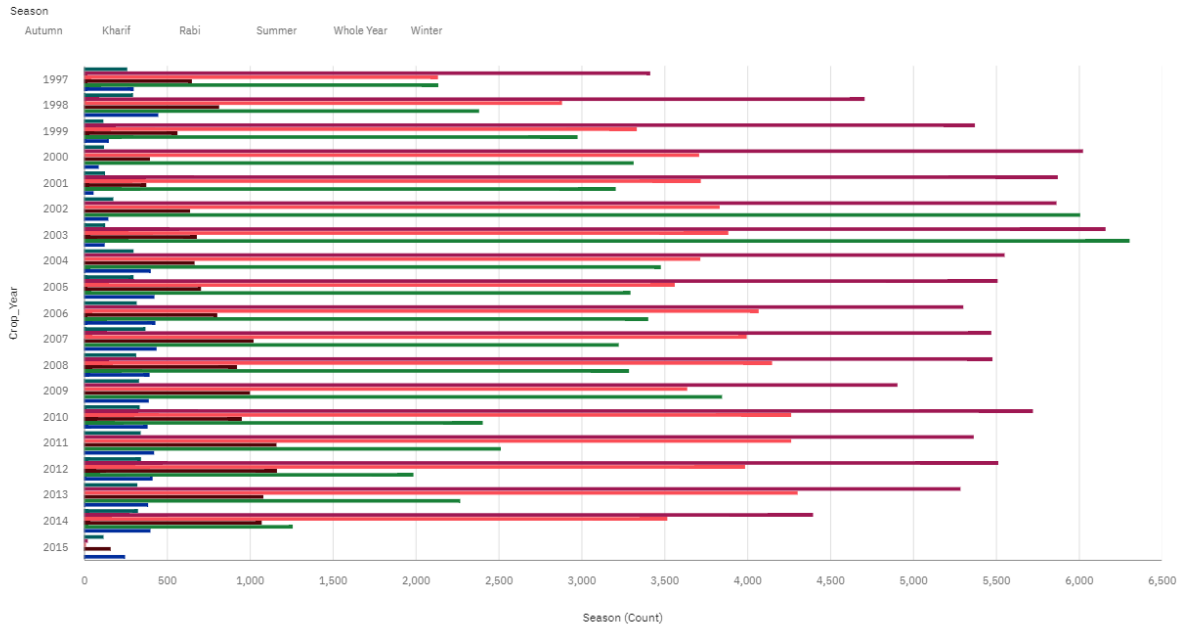

tab 1
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Crop by District_Name
District_Name
NAMSAI
MUMBAI
HYDERABAD
RAMBAN
SHOPIAN
KHUNTI
RAMGARH
LONGDING
GANDERBAL
KUPWARA
DHANBAD
GIRIDIH
BOKARO
JAMTARA
SARAIKELA KHARSAWAN
EAST SINGHBUM
WEST SINGHBHUM
ANJAW
GUMLA
LAHUL AND SPITI
RANCHI
PALGHAR
FAZILKA
SAHEBGANJ
PATHANKOT
LATEHAR
DEOGHAR
DUMKA
Crop for Season
Column values
Increase
Decrease
Sum
Crop (Count distinct)
Season


tab 1
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Production by Season colored by Season
Season
Autumn
Kharif
Rabi
Summer
Whole Year
Winter
Production (Count distinct)
26,000
24,000
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Autumn
Kharif
Rabi
Summer
Whole Year
Winter
Season


tab 1
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Crop for State_Name regions
Crop
Banana
other oilseeds
Arhar/Tur
Bajra
Cotton(lint)
Rice
Sunflower
Wheat
Linseed
Coconut
Sweet potato
Other Rabi pulses
Soyabean
Citrus Fruit
Grapes
Rapeseed &Mustard
Cabbage
Peas (vegetable)
Dry ginger
Maize
Oilseeds total
Jute
Masoor
Dry chillies
Potato
Sugarcane
Mesta
Other Kharif pulses
Onion
Jowar
Afghanistan
Islamabad
Lahore
Pakistan
New Delhi
Nepal
Bhutan
Bangladesh
India
Karachi
Mumbai
Hyderabad
Bengaluru
Myanmar
(Burma)
Nay Pyi Taw
Laos
Hanoi
Vientiane
Thailand
Bangkok
Cambodia
Arabian
Sea
Bay of
Bengal
TIBET
SICHUAN
YUNNAN
C_Production
583K
C_Production

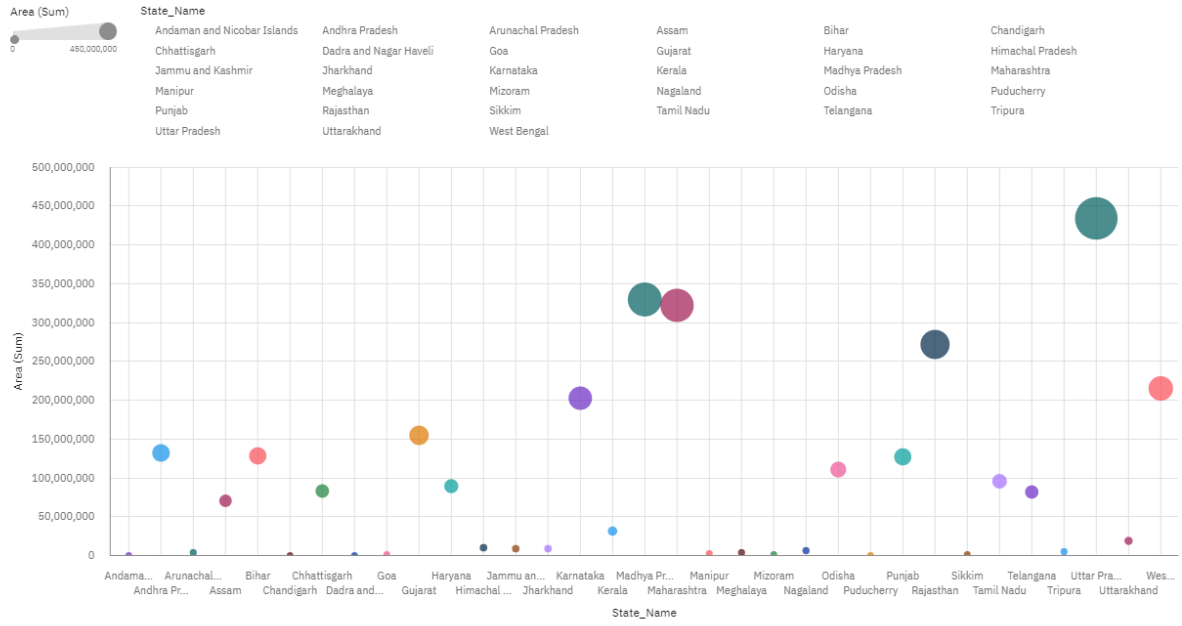
tab 1 | Tab 1 | Tab 2 | Tab 3 | Tab 4 | Tab 5

Season by Crop_Year colored by Season



tab 1 | Tab 1 | Tab 2 | Tab 3 | Tab 4 | Tab 5

State_Name by Area colored by State_Name and sized by Area



8.BENEFITS:

- Predicting crop productivity under different climatic circumstances can assist farmers and other partners in making fundamental decisions about agronomy and product choice.
- The value and profit of farming can be improved by using this model to choose the best crops for the region and their yield.
- This will make it easier for the state's policymakers to set the budget.
- If a crop's productivity is seen to be dropping, early implementation of the plans can be planned. As a result, the state will avoid a product scarcity.
- aids in the formulation of government policies.
- Yield information aids the farmer in deciding how much to plant the following year.
- Monitors the growth of healthy crops.
- Helps the farmer in Seed Selection, Pest Management, Irrigation Scheduling, etc,...

CHALLENGES:

Challenges are a big factor that will likely have a negative impact on the current endeavour. Some among the difficulties in predicting crop yield are: Choosing the right dataset, tweaking the parameters after choosing the dataset, and makes a project more productive to achieve the desired outcomes.
Model training must take into account lower computing efficiency and power.
An increase in error rate as a result of the environment changing dynamically.

CONCLUSION:

As a result of penetration of technology into agriculture field,there is a marginal improvement in the production. The innovation have led to new concepts like digital agriculture,smart farming, precision agriculture,etc In the literature it has been observedr analysis has been done on agriculture productivity, hidden patterns discovery using data set related to seasons and crop yields data. We have noticed and made analysis about different crop cultivated,area and production in different states and districts using IBMCognos some of them are 1)Seasons with average production.In this we come to know in which season average production is more and in which

season average production is less.2)Production by crop year, in this analysis it is come to know in which years the production is high and low. like this different data exploration has been doned.Finally created the dashboard and make analysis that in which state and in which year with crop,are and what extend the production will be are analyzed.