PROJECT OVERVIEW:

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

PURPOSE:

Agriculture forms the basis for food security and hence it is important. In India, majority of the population i.e., above 55% is dependent on agriculture as per the recent information. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance. In India, wheat and rice are the major grown crops along with sugarcane, potatoes, oil seeds etc. Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture. This domain provides employment to more than 60% of the total population and has a contribution to GDP also (about 17%) [1]. In the farm output, India ranks second considering the world wide scenario. This is the widest economic sector and has an

important role regarding the framework of socio-economic fabric of India. Farming depends on various factors like climate and economic factors like temperature, irrigation, cultivation, soil, rain fall, pesticide and fertilizers

EXISTING SYSTEM:

At present we are at the immense need of another Green revolution to supply the food demand of growing population. With the decrease of available cultivable land globally and the decreased cultivable water resources, it is almost impossible to report higher crop yield. Agricultural based big data analytics is one approach, believed to have a significant role and positive impact on the increase of crop yield by providing the optimum condition for the plant growth and decreasing the yield gaps and the crop damage and wastage. With this aim the present paper reviews about the various advances, design models, software tools and algorithms applied in the prediction assessment and estimation of the crop yield. India is basically agriculture based country and approximately 70% our country economics is directly or indirectly related to the agricultural crops. The principle crop which occupies the highest (60-70%) percentage of cultivable land in the Indian soil is the paddy culture and it is the major crop especially in central and south parts of the India. Rice crop cultivation plays an imperative part in sustenance security of India, contributing over 40% to general yield generation. The enhanced yield of the rice crop depends largely on the water availability and climatic conditions

LITERATURE SURVEY:

REFERENCES

1)Rice Crop Yield Prediction using Data Mining DakshayiniPatil ,Dr.M.S.Shirdhonkar 2017 Discussed various data mining techniques utilized for prediction of rice crop yield for the state of Maharashtra, India. WEKA tool was applied in dataset processing.

2)A Survey on Crop Yield Prediction based on Agricultural Data Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R/ 2017 Presented a survey on the different algorithms applied in theassessment and rediction of crop yieldDiscussed about the mechanism of knowledge the discovery inAgricultural data mining

3) A Study on Various Data Mining Techniques for Crop Yield Prediction Yogesh Gandge, Sandhya 2017 Discussed various data mining techniques employed for predicting thecrop yield and signifies the importance of accurate data extractionmethods of big data analytics.

4) Big Data for weed control and crop protection F K Van Evert, S Fountas, D Jakovetic, V Crnojevic, I Travlos & C Kempenaar/ 2017 Outlined Big Data analytics models with numerical algorithms applied Represent the importance of reforming the mined data in the form of understandable information to the farmers. Discussed about various advances, tools and algorithms applied in transforming the data in to easily understandable information to the framers and thrown a light on success story of Netherlands in achieving the maximum crop yield and their smart forming practices.

5) The Impact of Data Analytics in Crop Management based on Weather Conditions Swarupa Rani A/ 2017 Discussed the application of mathematical model like fuzzy logicvdesigns in optimization of the crop yield, artificial neural networks in validation studies, genetic algorithms designs in accessing the fitness of the model applied, decision trees, and support vector machines to

study soil, climate conditions and water regimes related to crop growth and pest management in agriculture.

6) A Study on Crop Yield Forecasting Using Classification Techniques R.Sujatha, Dr.P.Isakki Devi/ 2016 Discuss the importance of comparing previous agricultural data with present to identify optimum condition favor enhanced crop yield. Envisaged the importance of best crop selection depending onthe season and the climatic factors which supports enhanced crop yield.

7) Prediction of Crop Yield using Regression Analysis Swarupa Rani A/ 2017 Regression analysis was carried out to find the relationship among the parameters i.e Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price Index (FPI) which influences the final crop yield and reported that the crop yield principally depends on the Annual Rainfall (AR).

8) Data requirements for reliable crop yield simulations and yield- gap analysis Patricio Grassinia, Lenny G.J. van Bussel, Hugo de Groote,Martin K. van Ittersumb, Kenneth G. Cassman/ 2015 Presented a case study (Nebraska - USA and at a national scale for Argentina and Kenya) on the application of an explicit rationale design approach in identifying the data sources which simulates Crop (maize) yield and also helps in quantifying the maize yield gaps.

9) Prediction of crop yield using big data Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua Wang Juyun/ 2015 Developed a novel model i.e Nearest neighbors modeling tocalculate and predict the yield of crop depends on the available Big data sets.

10) The use of satellite data for crop yield gap analysis David B. Lobell/ 2013 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.

11) Yield gap analysis with local to global relevance-A review Martin K. van Ittersuma, Kenneth G. Cassmanb, Patricio Grassinib, Joost Wolfa, Pablo Tittonell, Zvi Hochmand Discussed about the various methodsused on quantifying the yield gaps at local-toglobal ratio. Reported few standard operation methods, employed inquantify the crop yield potential on the data collected from the farmers ofwestern Kenya, Nebraska (USA) and Victoria (Australia). Study recommended for the use of accurate and current yield data, with calibrated and validated cropmodels and up scaling methods in the prediction of crop yield.

12) A tool for analysing vegetable crops data from a greenhouse using data mining techniques. Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,Dominguez Limaico, (2017). algorithm, which uses a decision tree based on the data entropy is used and results are visualized graphically.

13) Using hybrid support vector regression to predict agricultural output. Jheng, T.-Z., Li, T.-H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agriculture tural output. Hybrid SVR models are used for prediction

14) Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: Manjunatha, M., Parkavi, A. (2018). A Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: A Survey. Classified using fuzzy logic, decision trees, Multiple Linear Regression and Random Forest algorithm to predict the crop yield.

15) Agricultural production output prediction using Supervised Machine Learning techniques. Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques" Decision Tree Learning-ID3 (Iterative Dichotomiser 3) and K-Nearest Neighbors Regression algorithms are used for prediction.

PROBLEM STATEMENT:

Doing a crop estimate is an important practice because maize is the staple food for our country as well as Africa. Approximately 200 million Africans consume maize on a daily basis. Knowing how much maize is available in the market will determine how much you and I will need to pay for it at the supermarket. Therefore, crop estimates contribute to the price setting mechanisms of the market.

There are many other products which prices are also indirectly determined by the maize price such as livestock. Knowing how much maize will be available at harvest time will also allow us to manage and budget accordingly. Perhaps in a poor season we will be prompted to hold back on exports and rather keep reserves for our own countries consumption. In a good year we may have the freedom to push exports or perhaps dedicate a larger percentage into animal production. These are all decisions that the crop estimates aid us in making.

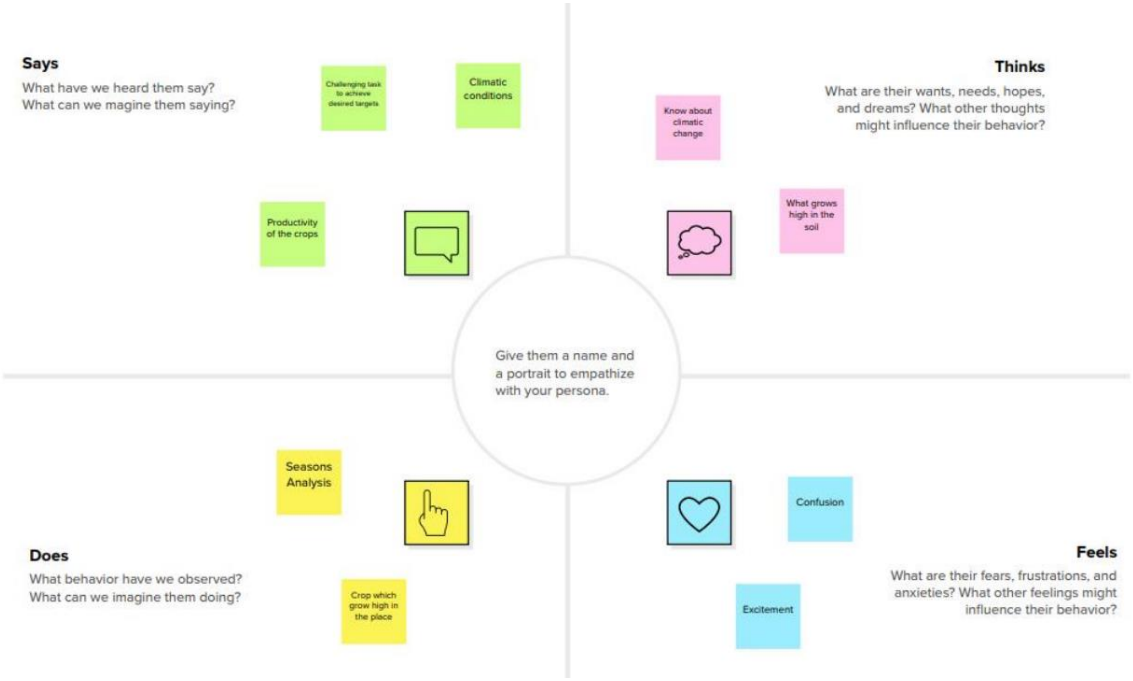
Many of these reasons for doing crop estimates are similar for the individual farmer. Farmers will do personal crop estimates in order to budget and plan for the coming season. The tonnes that you expect to harvest have a direct influence on your potential profits. Farmers who run mixed operations where livestock is a part of the business structure will be able to plan and strategise for optimum profits by deciding where the grain will make the most impact. Many farmers sell maize silage and doing a crop estimate will allow them to market their silage according to a dry grain crop estimate. Doing a maize estimate in each field is a good practice as it allows you to assess your performance throughout the planting season. It also assists you in analysing your fertilisation programme.

IDEATION AND PROPOSED SOLUTION:

empathy map canvas

An empathy map is **a collaborative tool teams can use to gain a deeper insight into their customers**. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. The empathy map was originally created by Dave Gray and has gained much popularity within the agile community.

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviours and attitudes. It is a useful tool to helps teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges

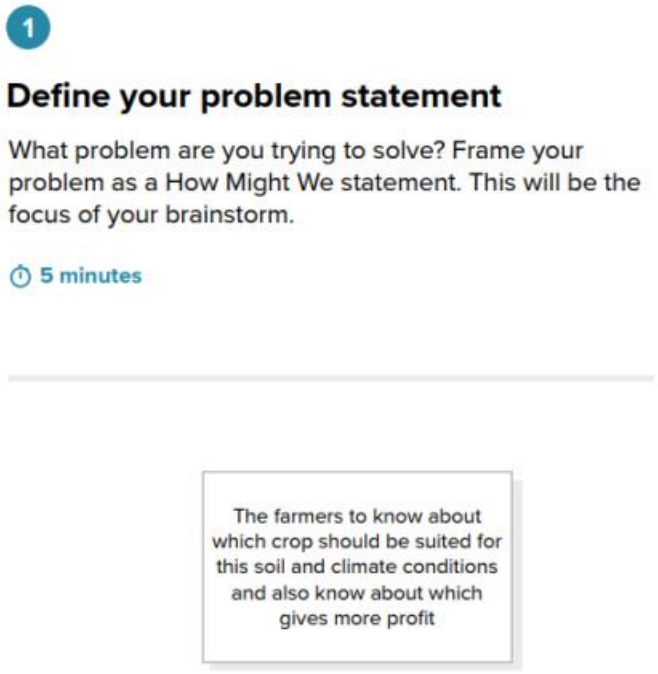


CUSTOMER JOURNEY MAP:

Brainstorm & Idea Prioritization:

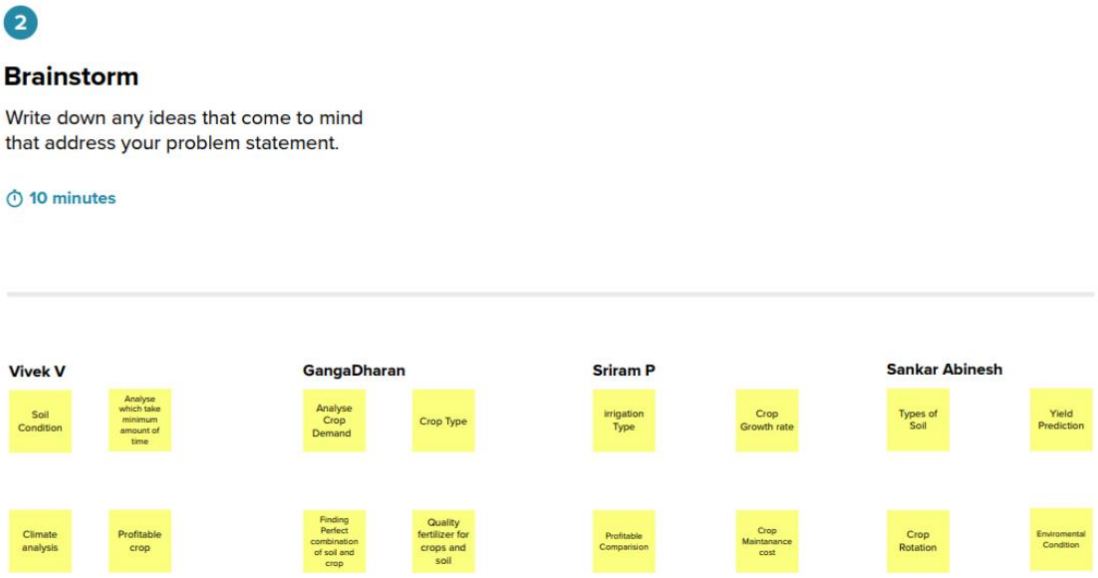
Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich number of creative solutions. Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room. Step-1: Team Gathering, Collaboration and Select the Problem Statement

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



00

Step-2: Brainstorm, Idea Listing and Grouping

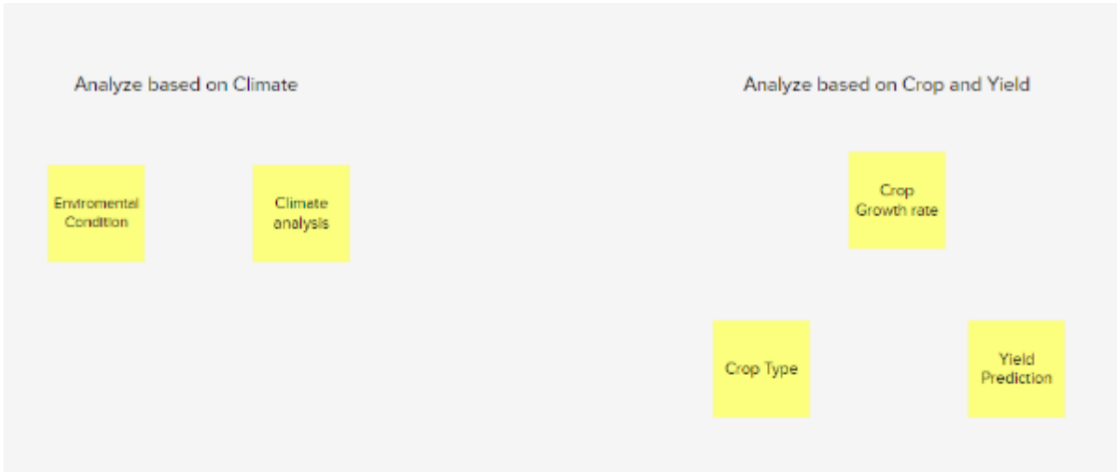


3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes



Step-3: Idea Prioritization

PROPOSED SOLUTION

Proposed Solution Template:

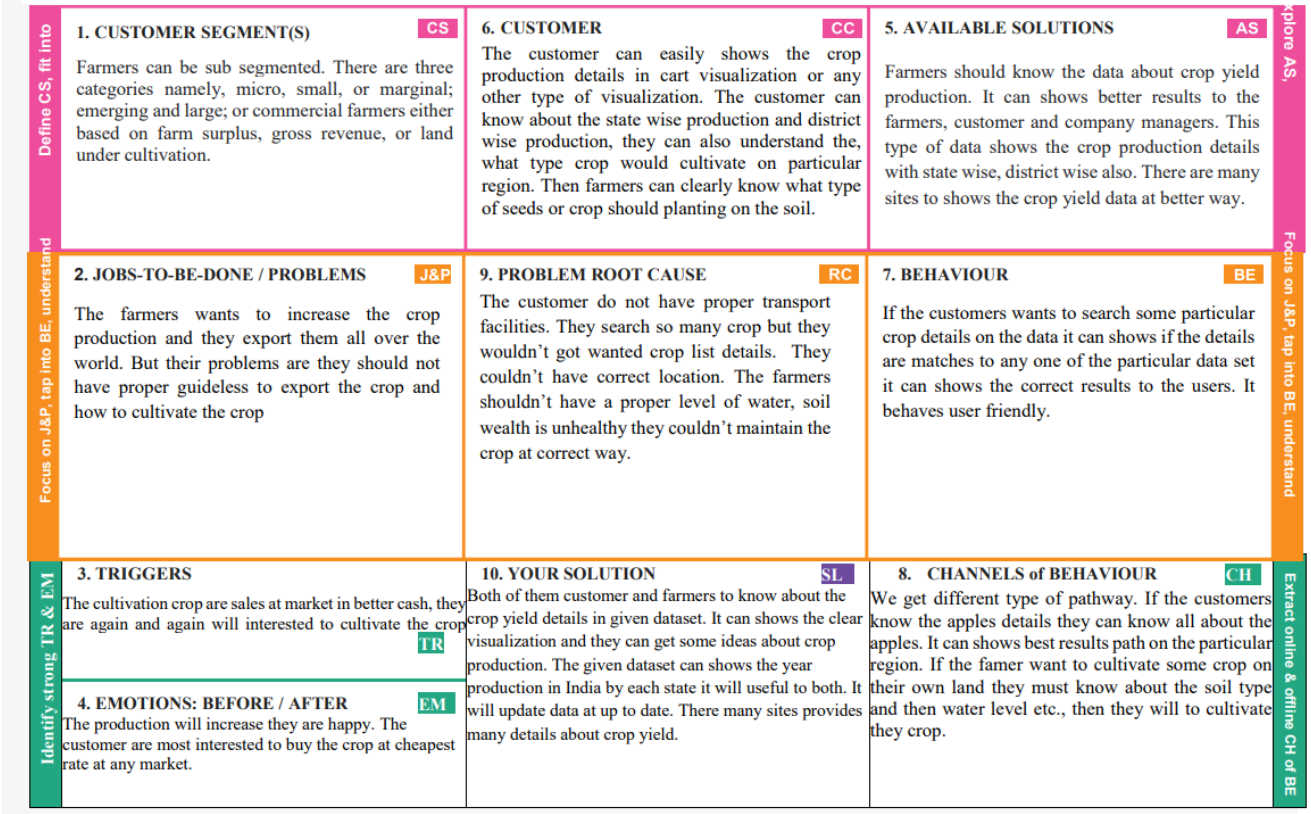
Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The farmers to know about which crop should be suited for this soil and climate conditions and also know about which gives more profit.
2.	Idea / Solution description	The Soil is analysed and give the data which crop is good for this soil and more profitable. Climatic analysis is taken to save the crop from any disaster or cause by any environmental change. Also analyse the more profitable crop in less time.
3.	Novelty / Uniqueness	New features will be extracted from given data by analysing. With these new features, more information can be gained and better decision will be taken to increase profit for farmers.
4.	Social Impact / Customer Satisfaction	• Perception of profit about particular crop • Perception of Climatic changes to save from losses • Perception of Fast-growing crops • Perception of Crop production in particular location
5.	Business Model (Revenue Model)	• Dashboard will be created to view Crop production in required location. • Better decision will be made by farmers.
6.	Scalability of the Solution	This solution can be used by every farmer. This solution can be processed with less memory and quickly. The solution can be used as open source so everyone can use it.

PROBLEM SOLUTION FIT

PROBLEM -SOLUTION FIT

Estimate the crop yield using data analytics.



FUNCTIONAL REQUIREMENTS:

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Account creation	Create an account in the dashboard
FR-4	Input credentials	Uploading the dataset and analysing the dashboard
FR-5	Processing Methods	Using IBM Cognos Analtyics Dashboard using Prediction algorithm to find them
FR-6	Output Credentials	Using the dashboard and algorithm they know about the crop yield estimate

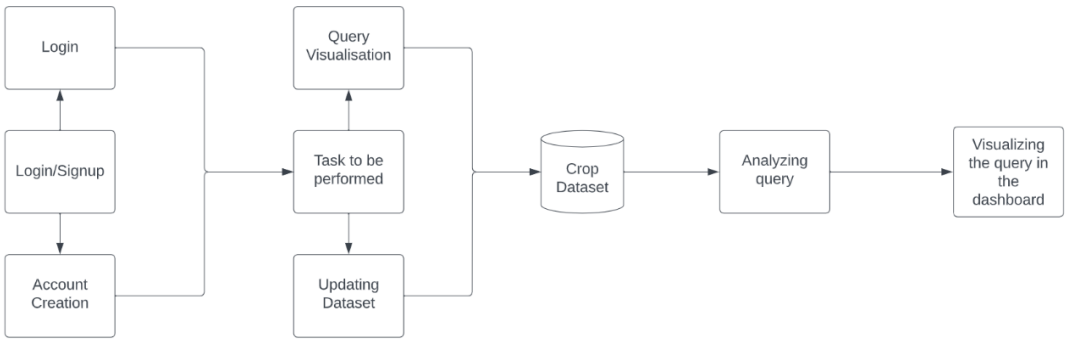
Non-functional Requirements:

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

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The user can be able to access with the system user friendly. The system is built with a simple manner and algorithm.
NFR-2	**Security**	Access permission for the particular system information it can only change by the system admin. The user must be having a high security measures.
NFR-3	**Reliability**	The database update process must roll back all the related update when any update fails. The dataset will not be modified by anyone only the admin can modify the dataset.
NFR-4	**Performance**	The performance of the dashboard is very easy and flexible to the user. It can be perform smooth.
NFR-5	**Availability**	New module deployment must not impact front page, dashboard and check out the pages availability and then show some visualization of them.
NFR-6	**Scalability**	It can be support many users to access at a time. The dashboard is scalable for the crop when their farmers and customers is used to analysis.

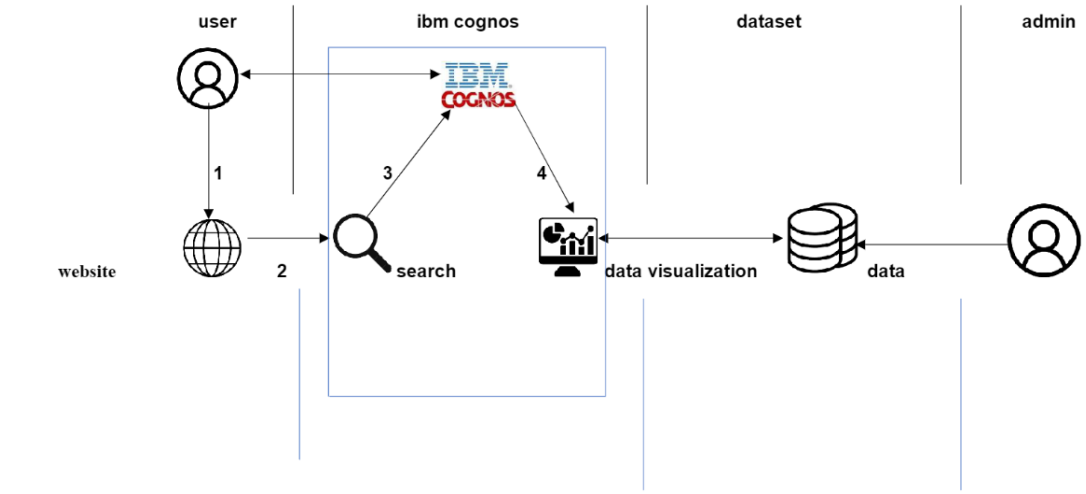
PROJECT DESIGN

DATA FLOW DIAGRAM



ARCHITECTURE DIAGRAM

Technical Architecture:



6.PROJECT PLANNING & SCHEDULING

Project Planning Phase
Milestone and Activity List

Date	25 October 2022
Team ID	PNT2022TMID20275
Project Name	Estimate the Crop Yield using Data Analytics

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring to technical papers, research publications etc.	29 Aug-3 Sept
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	5 Sept- 10 Sept
Ideation	List them by organizing the brainstorming session and prioritize the top 3 ideas based on feasibility & importance.	12 Sept -17 Sept
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	19 Sept -24 Sept
Problem Solution Fit	Prepare problem - solution fit document.	26 Sept – 1 Oct
Solution Architecture	Prepare a solution architecture document.	26 Sept – 1 Oct

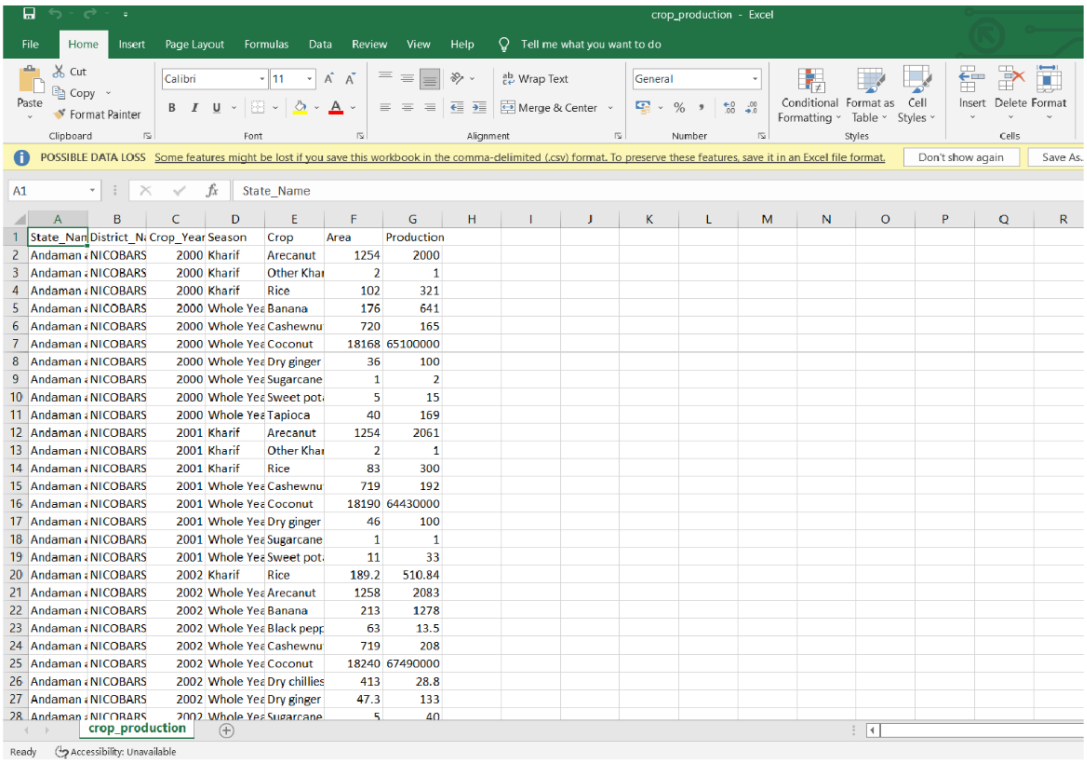
UNDERSTANDING THE DATASETS:

Understanding The Dataset This project is based on a understanding the crop production of India .

Download the dataset from the below link. It has 2,46,092 data points (rows) and 6 features(columns) describing each crop production related details.

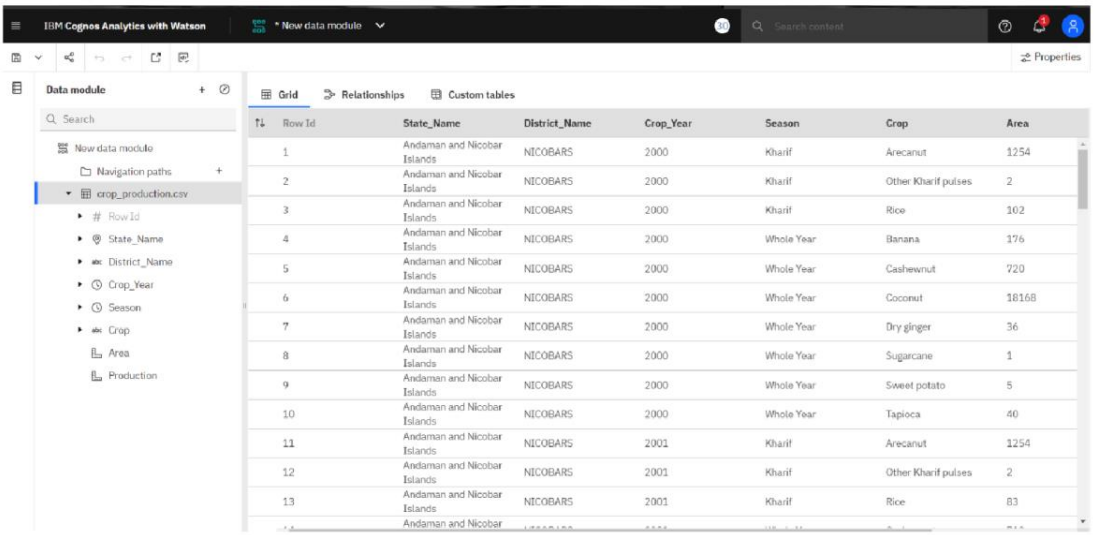
Dataset Link :Dataset Let’s understand the data we’re working with and give a brief overview of what each feature represents or should represent

1. State Name - All the Indian State names.
2. District Name -Different District names.
3. Crop Year- contains the crop years.
4. Season – Different seasons for crop production
5. Area- Total number of areas covered. 6. Production- production of crops.

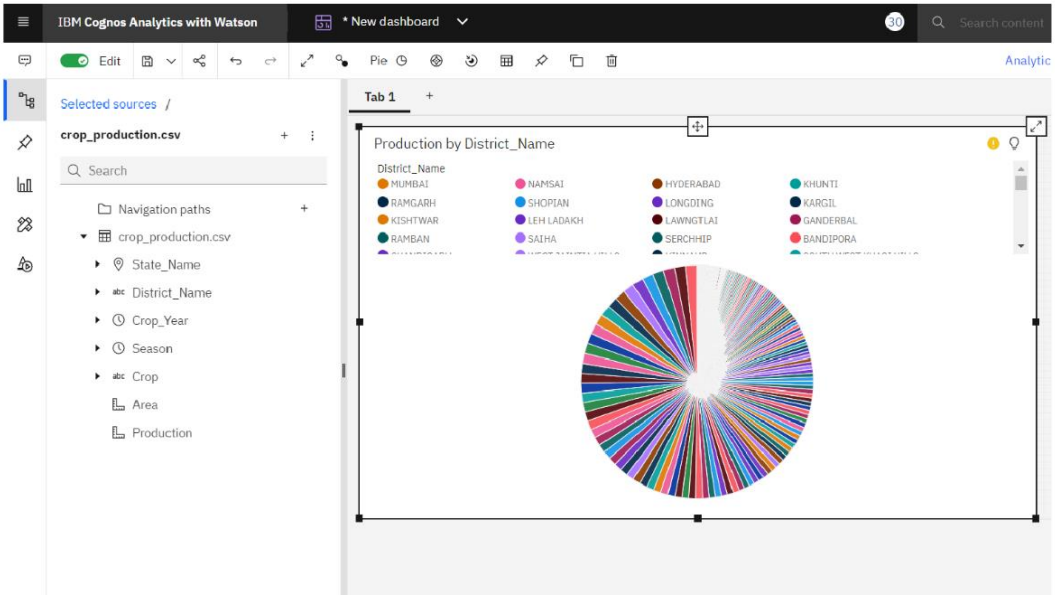


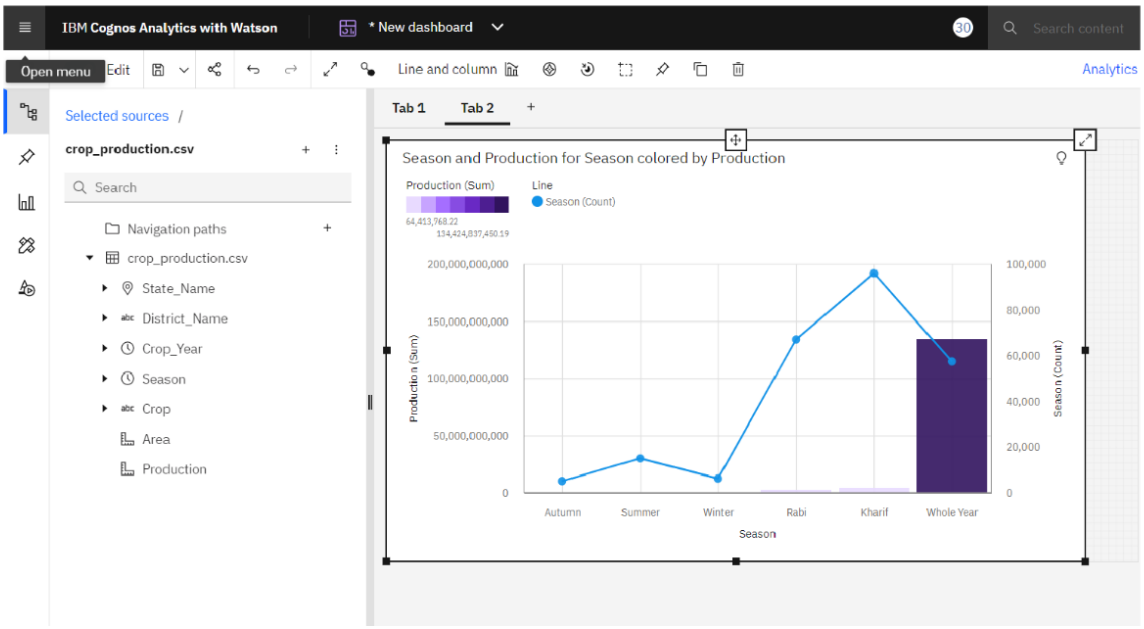
crop_production - Excel

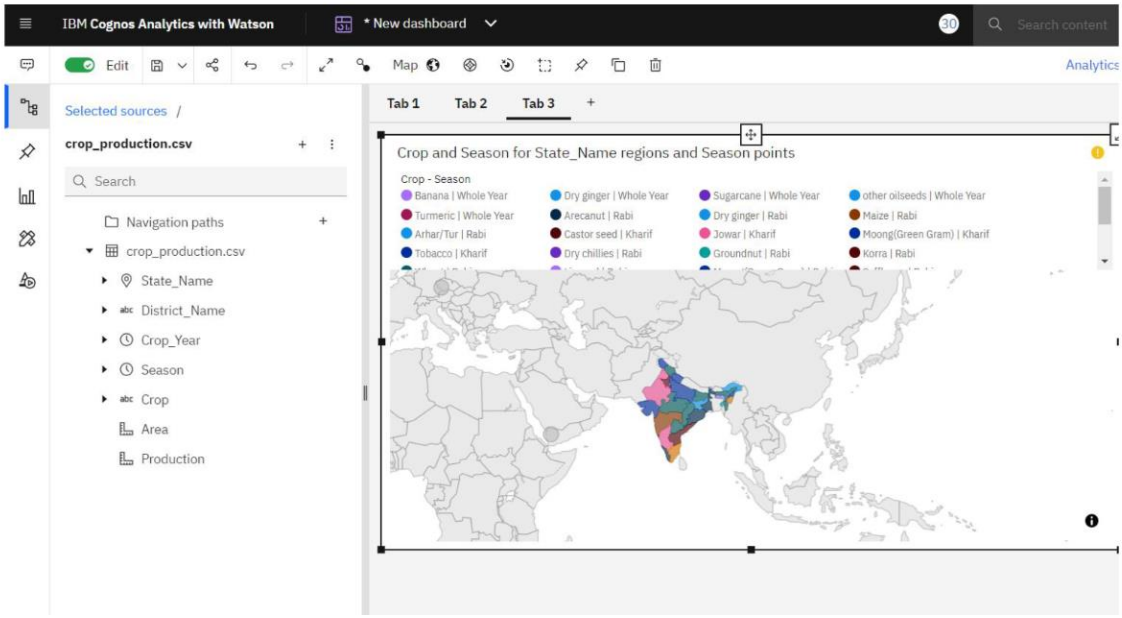
State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254	2000
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176	641
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720	165
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168	65100000
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36	100
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1	2
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5	15
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40	169
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254	2061
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83	300
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Cashewnut	719	192
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Coconut	18190	64430000
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Dry ginger	46	100
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sugarcane	1	1
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sweet potato	11	33
Andaman and Nicobar Islands	NICOBARS	2002	Kharif	Rice	189.2	510.84
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Arecanut	1258	2083
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Banana	213	1278
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Black pepper	63	13.5
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Cashewnut	719	208
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Coconut	18240	67490000
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Dry chillies	413	28.8
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Dry ginger	47.3	133
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Sugarcane	5	40

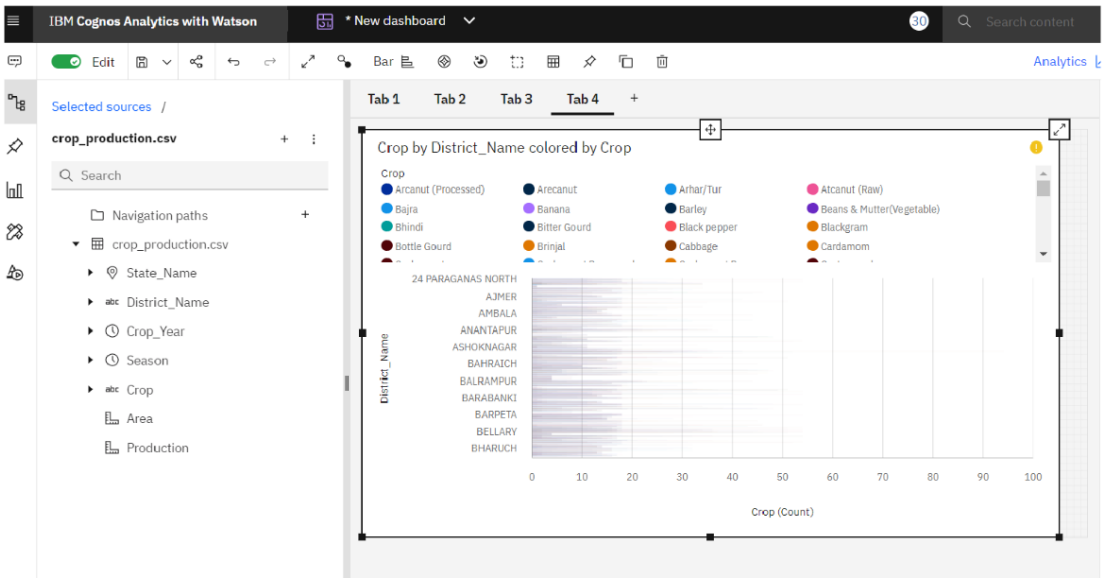


Data Exploration:

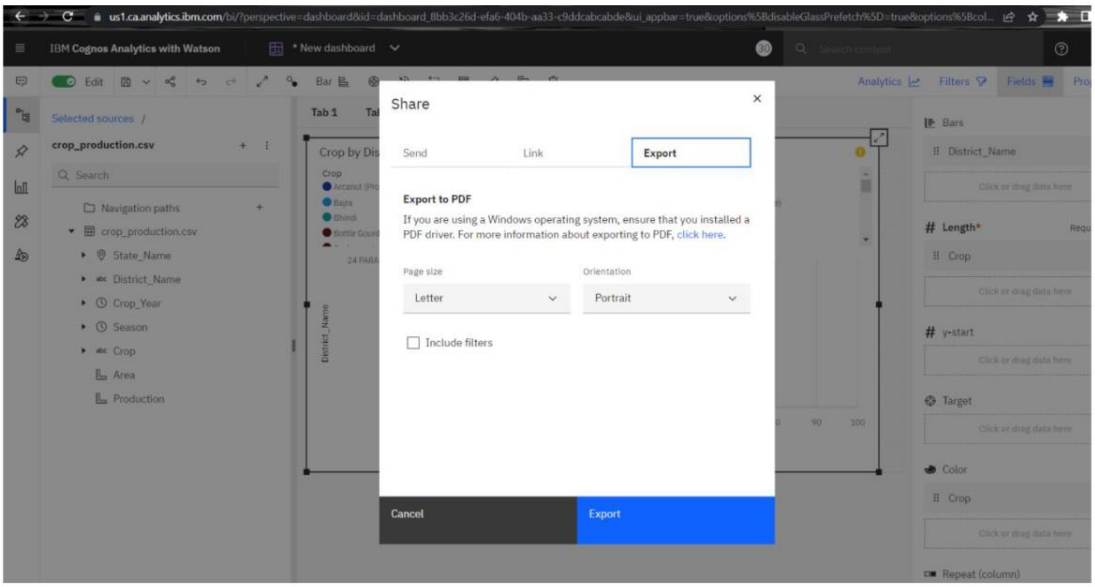



IBM Cognos Analytics with Watson
* New dashboard
Search content
Open menu
Edit
Line and column
Analytics
Selected sources /
crop_production.csv
Search
Navigation paths
crop_production.csv
State_Name
District_Name
Crop_Year
Season
Crop
Area
Production
Tab 1
Tab 2
Season and Production for Season colored by Production
Production (Sum)
64,413,768.22
134,424,837,450.19
Line
Season (Count)
200,000,000,000
150,000,000,000
100,000,000,000
50,000,000,000
0
Production (Sum)
100,000
80,000
60,000
40,000
20,000
0
Season (Count)
Autumn
Summer
Winter
Rabi
Kharif
Whole Year
Season


IBM Cognos Analytics with Watson
* New dashboard
Search content
Edit
Map
Analytics
Selected sources /
crop_production.csv
Search
Navigation paths
crop_production.csv
State_Name
District_Name
Crop_Year
Season
Crop
Area
Production
Tab 1
Tab 2
Tab 3
Crop and Season for State_Name regions and Season points
Crop - Season
Banana | Whole Year
Dry ginger | Whole Year
Sugarcane | Whole Year
other oilseeds | Whole Year
Turmeric | Whole Year
Arecanut | Rabi
Dry ginger | Rabi
Maize | Rabi
Arhar/Tur | Rabi
Castor seed | Kharif
Jowar | Kharif
Moong(Green Gram) | Kharif
Tobacco | Kharif
Dry chillies | Rabi
Groundnut | Rabi
Korra | Rabi

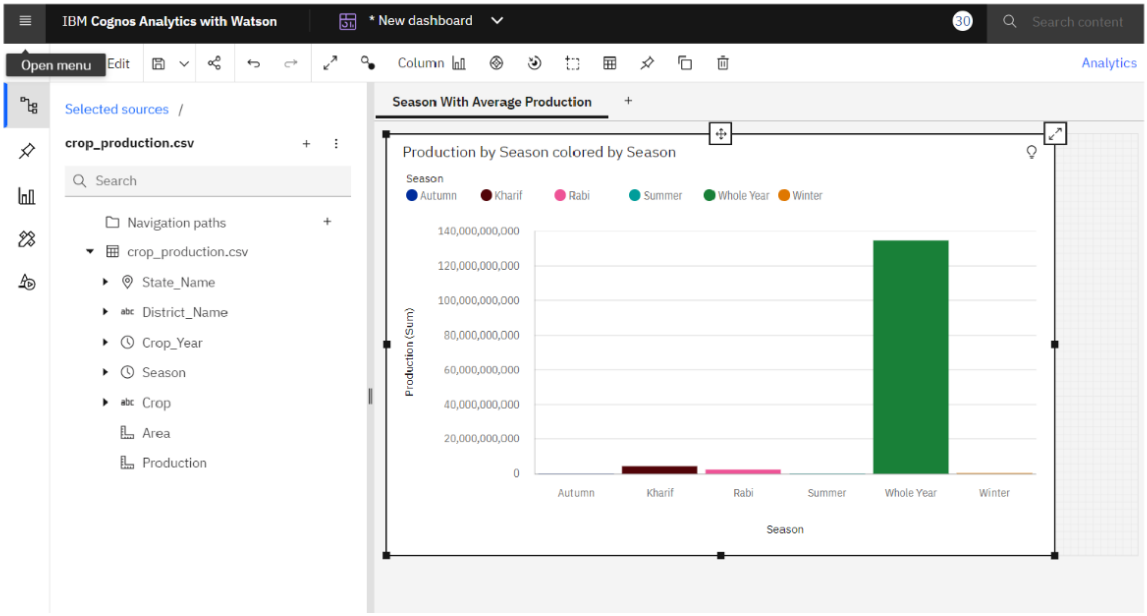


Export Dataset:

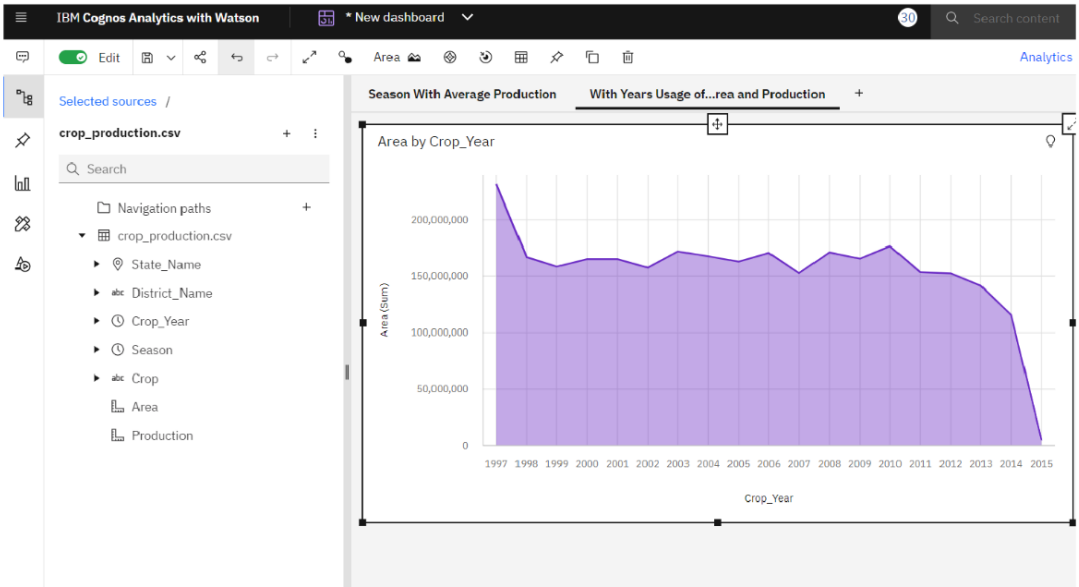


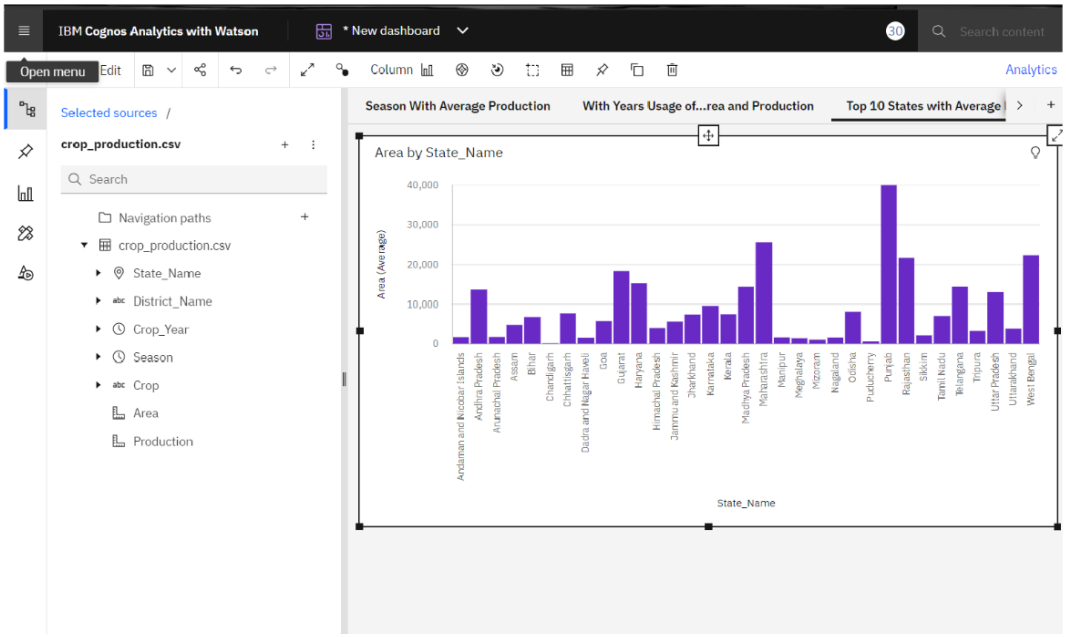
Dashboard Creation and Data Visualization

Season with Average Production:

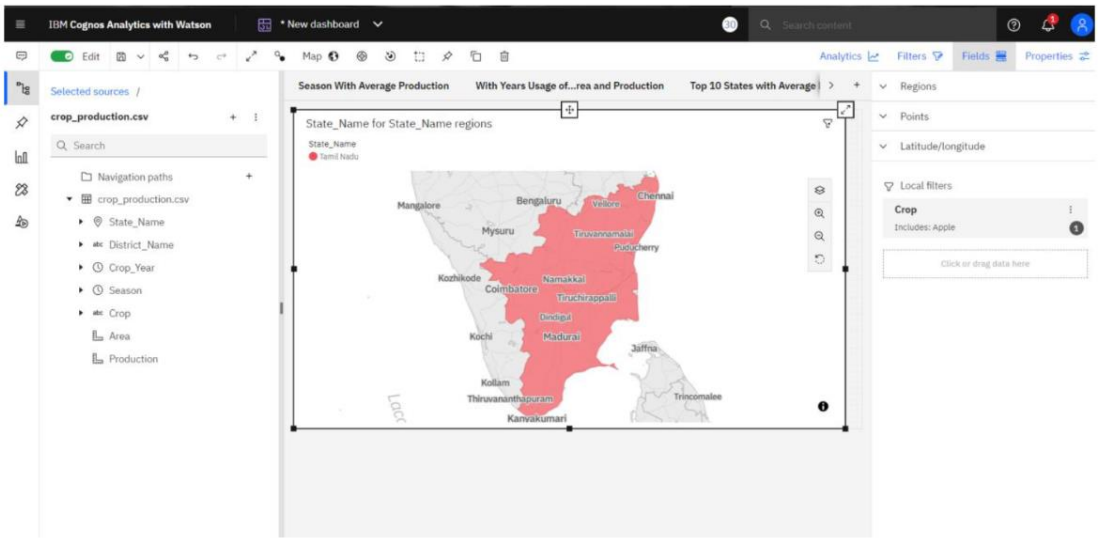


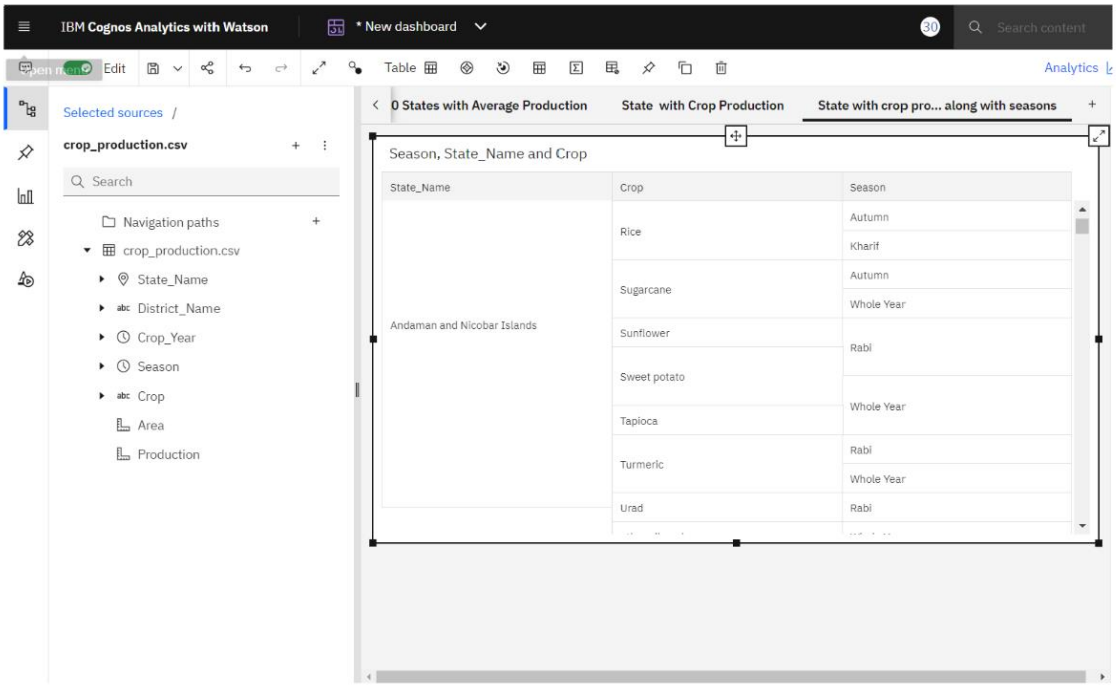
With Years Usage of Area and Production:

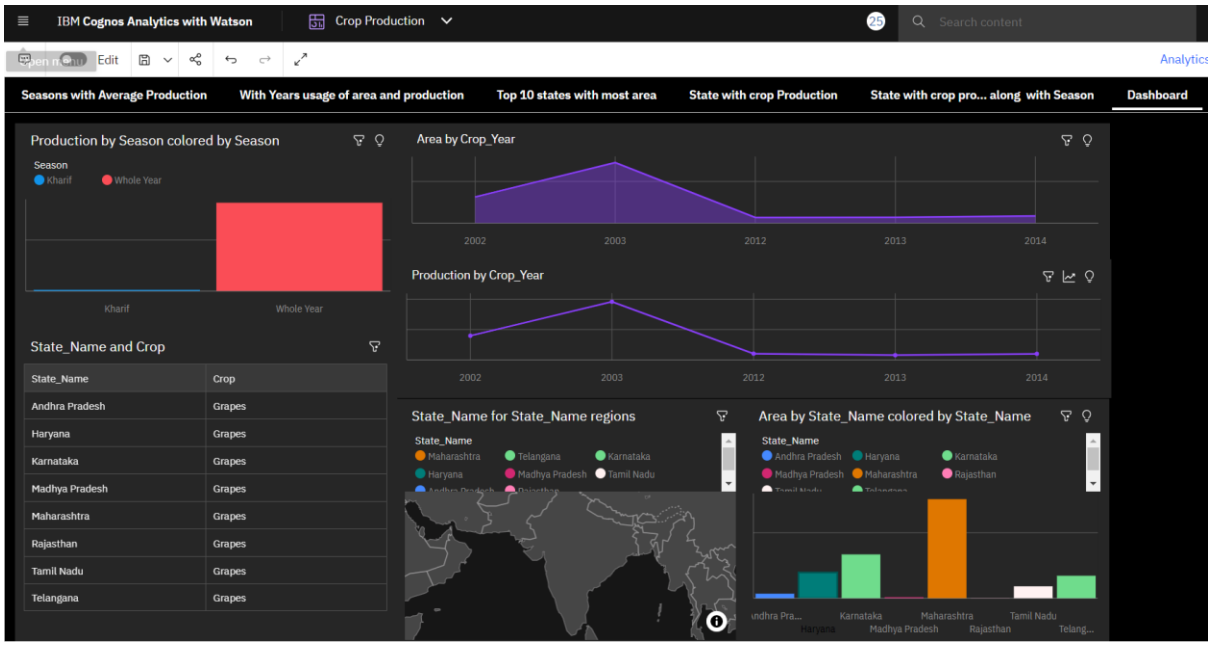




State with Crop Production:







Advantages of crop yield prediction

Crop yield prediction is also used by farmers to make decisions about when to plant and harvest crops based on soil moisture content, pest infestations, and other factors such as weather conditions and fertilizer requirements.

The quantity and quality of crops produced in the world is a major factor affecting food security.

This is especially important in developing countries where agriculture is still an important part of the economy. To predict crop yield, we need to understand how much sunlight plants receive and how much water they need.

Plant growth depends on these two factors, but many other factors affect plant growth such as temperature, humidity, and soil type.

Yield data can also be used to monitor progress toward global goals set by governments, non-governmental organizations, and other stakeholders.

In addition, yield data is used to identify strengths and weaknesses in farming practices and make recommendations based on this information.

Disadvantages:

Although these crops are known for their high yielding property, the high yield crops also have some disadvantages:

- The high yield crops require more water and fertilizers as compared to the normal varieties of crops.
- They require frequent weeding.
- Continuous use of pesticides.
- The high yield crops, when compared with the traditional varieties are generally more susceptible to diseases.

CONCLUSION

As a result of penetration of technology into agriculture field, there is a marginal improvement in the productivity. The innovations have led to new concepts like digital agriculture, smart farming, precision agriculture etc. In the literature, it has been observed that analysis has been done on agriculture soils, hidden patterns discovery using data set related to climatic conditions and crop yields data. The activities of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield Indonesian J Elec Eng & Comp Sci ISSN: 2502-4752 ρ Agriculture Data Analytics in Crop Yield Estimation: A Critical Review (B M Sagar) 1093 prediction etc. In this

survey, the specific activity, crop yield prediction has been surveyed and the major trends have been identified. The rice crop yield prediction has been done in the state of Maharashtra using data mining techniques in one of the works [8]. The analysis has been done using machine learning framework WEKA. In the work carried out in [9], various algorithms applied in the assessment crop yield and mechanism for knowledge discovery has been discussed. The challenges and opportunities in the field of Big Data analytics in agriculture has been discussed in [6] with a case study of Netherlands. Fuzzy logic designs have been used in optimizing the crop yields and the same has been explained in the research work in [5]. A case study of Nebraska - USA and at a national scale for Argentina and Kenya has been done and presented in [14]. The remote sensing technology for identification and measurement of the causes of yield gaps and their impact on final crop yield is presented in [15]. 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 this research area.

Future scope:

1. Predict appropriate crop and maximum yield in the climate change.
2. Create an android app.
3. Collection of data, Analysis of it and modification of the algorithm.
4. IOT application in agriculture, automation in production line and man free agriculture
which is the future of the world ,this is the first step of it.
5. Find the percentage yield to happen from the match given percentage in terms of % error
s