

# **ESTIMATE THE CROP YIELD USING DATA ANALYTICS**

## **A PROJECT REPORT**

*Submitted by*

<b>THILEEPAN .A.P</b>	<b>142219205100</b>
<b>VETRIVEL .S</b>	<b>142219205105</b>
<b>BHARATH KUMAR .R</b>	<b>142219205301</b>
<b>LOKESH .R</b>	<b>142219205501</b>

## **BACHELOR OF ENGINEERING**

*In*

## **INFORMATION TECHNOLOGY**

**IBM-Project-11969-1659363832**

**Team ID: PNT2022TMID21973**

## **NALAIYATHIRAN 2022**

### **Table of contents**

<b>S.No</b>	<b>Chapter</b>	<b>Page No</b>
<b>1.</b>	<b>INTRODUCTION</b>	<b>5</b>
1.1	PROJECT OVERVIEW	5
1.2	PURPOSE	6
<b>2.</b>	<b>LITERATURE SURVEY</b>	<b>6</b>
2.1	Existing problem	6
	2.2 References	14
	2.3 Problem Statement Definition	14
<b>3</b>	<b>IDEATION &amp; PROPOSED SOLUTION</b>	<b>15</b>

3.1	Empathy Map Canvas	15
3.2	Ideation & Brainstorming	15
3.3	Proposed Solution	16
3.4	Problem Solution fit	17
<b>4</b>	<b>REQUIREMENT ANALYSIS</b>	<b>17</b>
4.1	Functional requirement	17
4.2	Non-Functional requirements	18
<b>5</b>	<b>PROJECT DESIGN</b>	<b>18</b>
5.1	Data Flow Diagrams	19
5.2	Solution & Technical Architecture	20
5.3	User Stories	21
<b>6</b>	<b>PROJECT PLANNING &amp; SCHEDULING</b>	<b>22</b>
6.1	Sprint Planning & Estimation	22
6.2	Sprint Delivery Schedule	22

6.3	Reports from JIRA	23
<b>7</b>	<b>CODING &amp; SOLUTIONING</b>	<b>24</b>
7.1	Feature 1	24
7.2	Feature 2	27
7.3	Database Schema	29
<b>8</b>	<b>TESTING</b>	<b>30</b>
8.1	Test Cases	30
8.2	User Acceptance Testing	31
<b>9</b>	<b>RESULTS</b>	<b>31</b>
<b>10</b>	<b>ADVANTAGES &amp; DISADVANTAGES</b>	<b>46</b>
<b>11</b>	<b>CONCLUSION</b>	<b>47</b>
<b>12</b>	<b>FUTURE SCOPE</b>	<b>47</b>
<b>13</b>	<b>APPENDIX</b>	<b>48</b>
13.1	Source Code	48
13.2	GitHub & Project Demo Link	48

# **1. INTRODUCTION**

## **1.1 PROJECT OVERVIEW**

Food is the major source of energy. Every living organism on this planet needs food to stay alive and to continue all other essential life processes. Plants are the main source of food on which both humans and animals depend. We cannot imagine life without food.

With the rapidly growing population, demand for more food, loss of produced crops, and other problems in the agricultural output are the main reasons for the scarcity of food and are the biggest concern in some parts of the world facing today. This has led to an increase in the requirement of strategies that can help in the management of the crops produced. By creating a dashboard and by going through these we will get most of the insights of Crop production in India.

“Crop production is the branch of agriculture that deals with the production of crops for food and fiber.”

## **1.2 PURPOSE**

Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. As per this project we will be analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

## **2. LITERATURE SURVEY**

### **2.1 Existing problem**

in india farmers are facing an issue of getting more crop yield due to lack of insights about estimation of the crop yield. They lack in proper methodology for sowing the seeds in appropriate area, season etc.

### **2.2 References**

(1)- In “Year book of crop yields - Estimation of crop yields” by V.G.Panse says statistics of crop yields in most countries are based on periodic reports from crop

reporters. Essentially such reports represent the reporters quantitative judgment of what the yield is, based on their personal impression of the crop enquiry among farmers. In mote advanced countries such as the U.S.A. and several European countries, crop reporter's are farmers and other private individuals resident in rural areas and connected with farming, who voluntarily supply to the government agency concerned, the information called for in a mailed questionnaire. In underdeveloped countries, crop reporters are generally government officials or agents who submit reports relating to crop yields for the area under their administrative charge. These official reporters are naturally much less numerous than voluntary crop reporters in advanced countries, and

consequently, the unit or area for which a report is made is usually large, such as an

administrative sub-division of a district and possibly a village, but hardly ever a cultivator's holding or a field. Estimates of yield, derived from such reports are likely to be much more inaccurate than those for which a very large number of reporters supply data on their respective areas. Unfortunately, the most serious defect in yield estimates based on crop reporters' data, whether the reporters are official or voluntary, is that they are subjected to large and indeterminate biases. Investigations made in India have shown that the official reporters have a marked tendency to lean towards the normal, with the result that yield is underestimated in favourable seasons and overestimated in poor seasons. Voluntary crop reporter's, on other hand, might have incentives for a systematic underestimation of yield. In the U.S.A., the regression method is used as a regular routine for eliminating bias in the crop reporters estimates, the regression of these estimates on "actual" yields for past years being used for this purpose. The actual yields are provided by the revised estimates of yield late in the season, on the basis of check data from various official as well as private sources, such as marketing's, shipments, the amount of the crop processed or handled in factories, and other reasonably complete utilization information. Such data are available for a few crops like cotton, tobacco, sugar beet; but even here there can be no guarantee that the information



is e on

is really complete. For the majority of crops, including food crops which are partly consumed locally or are stored by the farmer for his own use and for feeding his livestock, such check data are not available, and indications of yield from other sources, such as an annual farm census in a few states or the quinquennial census in all states, are used as a basis for revising crop reporters estimates of yield, It is possible, however, that the data from these other sources also suffer from biases as the crop reporters estimates do. Under the conditions prevailing in most countries it is not possible to obtain worthwhile information on the total production of the majority of crops such as would permit the systematic application of the regression technique in order to improve the quality of the crop reporters' estimates of yield. Further, where such regression is feasible, it can remove only the bias for which a trend has been established. Thus, experience under a wide range of agricultural and economic conditions demonstrates the inability of subjective or personal methods of estimating yield to provide reliable results and points to the necessity of replacing such methods by those involving; (i) the selection of a representative sample of the crop for observation and (ii) using for observation the method of physical measurement of yield at harvest. (2)- In "Crop yield forecasting using data mining" by Pallavi Kamath et al, India is a heavily reliant

is o

on agriculture. Organic, economic, and seasonal factors all influence agricultural yield. Estimating agricultural production is a difficult task for our country, particularly given the current population situation. Crop production assumptions made far in advance can help farmers make the necessary planning for things like storing and marketing. Crop production prediction involves a huge amount of data, making it a perfect candidate for data mining methods. Data mining is method of accumulating previously unseen anticipated information from vast database. Data mining assists in the analysis of future patterns and character, enabling companies to make informed decisions. For a specific region, this research provides a fast inspection of agricultural yield forecast using the Random Forest approach. The process of analysing, cleaning, and modelling data to generate useful knowledge and conclusions is known as data analysis. Methods are used to convert the customer's raw data into valuable information. This research can be extended to agriculture as well. Most farmers relied on their long-term field experience with specific crops to forecast a greater yield in the coming season. Nonetheless, they do not receive a fair price for their crops. It typically occurs because of insufficient irrigation or poor crop selection, but it may also occur when crop yields are lower than expected. Due to a variety of factors, the farmers who make up the majority

is e on

do not achieve the predicted Crop yield. That data set of crop yield which consists of many components. By studying the soil and atmosphere for the specific area, by which increase crop production, optimal crop can be estimated. Advantage of this research mainly is Farmers will benefit from this forecast. To determine which crops are best for their farm based on soil type, p., and fertilizer. In this paper effort is made in order to know the region-specific crop yield analysis and it is processed by implementing by random forest algorithm. In this project have chosen dataset which in .csv format. For the training purpose 80% of data is used and remaining 20% of data is used for testing. After the successful training and testing next step is finding the accuracy of the model. We have achieved a good accuracy which means this model is good for predicting yield. We have designed the Website which consists of Four Functional Modules as shown in the 1) Crop Module: This module will provide the list of available crops. On selection of each one of it will give the detailed description of the crop. 2) Soil Module: This module will provide the list of available soils. On selection of each one of it will give the detailed description of the soil. 3) Weather Module: In this module by entering the city name the user can get the live weather forecast. Open weather app is free open source weather data. By using weather API key can fetch the current or historical weather data. (3)-

is e on

Indonesian Journal of Electrical Engineering and Computer Science [Vol .12  
,No.3,December 2018 ,pp. 1087~ 1093ISSN: 2502-4752, DOI :  
10.11591/ijeecs.v12.i3.pp1087-1093 ] BM Sagar,



Cauvery N K ays the result of p netration of t chnology into the field f agriculture. In this literature, it has been observed that analysis has been done on agriculture soils, hidden patterns discovery using data set related to climate conditions and crop yields data. The activity of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction has been surveyed and the major trends have been identified using ::data

Analytics” Agriculture forms the basis for food security and it is important. In India above 55% people depend on agriculture as per the recent information. In India ,Wheat and rice are the major grown crops along with sugar cane, oil seeds etc.. The non food items like rubber, cotton, jute etc...More than 70% of rural area depend on agriculture. 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 frame work of socio-economic fabric of India .Historical information regarding crop yield provides major input for companies in planning supply chain decision like production scheduling .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 insight on various data analytics method applied to crop yield

e

production and also signifies the important lacunae points in the proposed area of research .. Crop yield production using Big Data Analytics: In India crop yield is season dependent and majorly influenced by the biological and economic causes of an individual crop . Reporting of progressive agricultural yield in all the session is an ample task and an advantageous task for every nation with the respect to assess the over all crop yield prediction. The accurate prediction of crop yield certainly benefits the framers in choosing the right method to reduce the crop damage and gets best price for their crops. A r search group conducted a work with a objective of accurate prediction of crop yield through big data analytics to assure various crop yield influencing factors such as Area Under Cultivation (AUC) in terms of hector , Annual Rainfall (AR) rates and Food Price Index (FPI) and to develop relationship among these parameters. All the selected factors of the present study design known as AR, AUC and FPI were measured for a periods of 10 year between the years of 1990-2000. A novel method called Linear Regression (LR) applied to analyse the relationship between explanatory variables (AR, AUC, FPI) and the crop yield considered. (4)-Keerthy t et al Some of the increasing role of Remote Sensing in Crop Area Estimation. Remote Sensing may be defined as the collection of information about an object or area without being in physical contact with the

object or area. Aircraft and satellites are the most common vehicles from which remote sensing observations are made. Aerial photography is the original and most familiar form of remote sensing and is widely used for topographic mapping, engineering and environmental studies, agricultural estimation, crop disease information, military observations, and exploration for oil and minerals. The method USDA's Natural Resource Conservation Service (NRCS) has an extremely large point sample and survey known as the National Resources Inventory (NRI). The NRI collects data in all counties and parishes of the 50 states and Puerto Rico, the Virgin Islands, the District of Columbia and areas of the Pacific Basin. The NRI was conducted at five year intervals beginning in 1977 and has been conducted annually since 2000 [Boryan, 2012]. The objective of the NRI is to monitor "status, conditions and trends in soil, water and other natural resources data on non-Federal lands in the United States" [Bretit, F. and W.

Full r,1999]. Data collected include broad land use and land cover categories such as irrigated and not irrigated cropland. These categories are then evaluated in terms of land capability class and subclass such as prime farm land; erosion and potential erodibility. Annual data collection is conducted primarily using photo interpretation and by ancillary data sources in house [Nusser and Goebel, 1997]. It was also expected to reduce the variance by approximately a factor of three and avoid the problems of incomplete PSUs. It was also considered to be more logical from the



implementation perspective, and the new design has served to reduce survey costs. The base sample for LUCAS corresponds to an initial 1 Km grid of about 4,000,000 points for the entire area of the EU. The LUCAS master sample is a subset of the base sample from a 2 Km grid created by using all the even points of the base sample, consisting of around 1,000,000 points. The master sample is then stratified by land cover classes and sub-sampled for ground data collection. Some of the issues faced by them at the time. First, the area planted for harvest of a given crop may change throughout the growing season. Such issues as use for purposes other than grain, abandonment, extreme weather damage, or unusual economic conditions may cause this change. It's usually necessary to make estimates several times throughout the crop season even for a given crop. Certainly the most significant challenge in estimating crop production is in doing so early in the season. Area estimation can create problems especially in countries prone to drought or flooding problems. For example, terraced crops versus those planted on a steep incline. SAR was deemed to be necessary for this application data because of its robustness to cloudiness, and the fact that there is considerable rain and cloudiness in the paddy rice growing area during the growing season. Apparently, the flooding of the target crop during the growing season, with water signatures mixed with the green plants, made SAR imagery more useful in this effort than it

might be for a non-flooded crop. Overall, the results of this crop area estimation effort seemed to be very good for the 2005 crop season. This has been expanded and implemented on an operational level in the four major summer grain producing provinces of Northwest, Mpumalanga, Free State, and Gauteng, and it is reportedly providing reliable crop area estimates. Advantage of USDA They are an excellent and timely source for crop area estimation. However, they also serve a very important statistical goal of measuring what is not available from a list of farmers (known as NOL for 'not on list' for hundreds of variables. It is the combination of area and list based information through a multiple frame sampling approach that leads to complete and cost efficient results for agricultural statistics of all sorts. Advantages: It can be used to statistically combine results, where comparable to sample surveys, or more simply as a quality control for both the administrative and sample survey data. This is a way for saving money, reducing response burden producing figures for very detailed domains and allowing estimation of transition over time. The test results for the six Southwest Rostov districts indicated that all methods worked reasonably well for large fields, and there was no significant difference between the methods. The test Some of the methods used to solve. Statistical methods have made clear that the enumeration of small samples can

greatly reduce the cost of the collection of agricultural statistics while increasing their accuracy. A well designed sample, for which the data are carefully collected, can provide much cheaper statistics than a census and provide more timely information on current conditions. Combining area and list frames, known as multiple frame sampling, has some very good qualities and solves some of the problems with each of them individually. Advantage of the VITO NN tended to perform best, and, in general, better results were obtained when using VGT, rather than MODIS data. As a result, this device aids farmers in their decision-making process, allowing them to save time. Suvidha Jambekar et al. Regression analysis is applied as a predictive modelling tool to predict crop production for crop production. The regression algorithms applied were, Multivariate Adaptive Regression Splines, and then Multiple Linear Regression, Random Forest Regression. According to the results, Random Forest Regression may be used to accurately estimate wheat, and rice, and maize production. B. Devika, B. Ananthi et al. Agriculture expands yield production to meet demand to limit overlapping, and the government encourages it for crop yield forecast on Tamil Nadu dataset imports. The regression method is put to the test of yield prediction capabilities in this study. R. Vidhya et al. They observed accuracy rate improves when a dataset with more features is used. As opposed to other approaches, such as Decision trees, linear regression, random forest algorithm is shown to be superior to other prediction algorithms. The included dataset incorporates a lot more variables resulting in more precise prediction. Het al Patel, Dharmendra Patel et al. They measured performance of the classification algorithms Naive Bayes, J48, and Simple Cart. This crop prediction comparative analysis employs a large dataset. At

e e

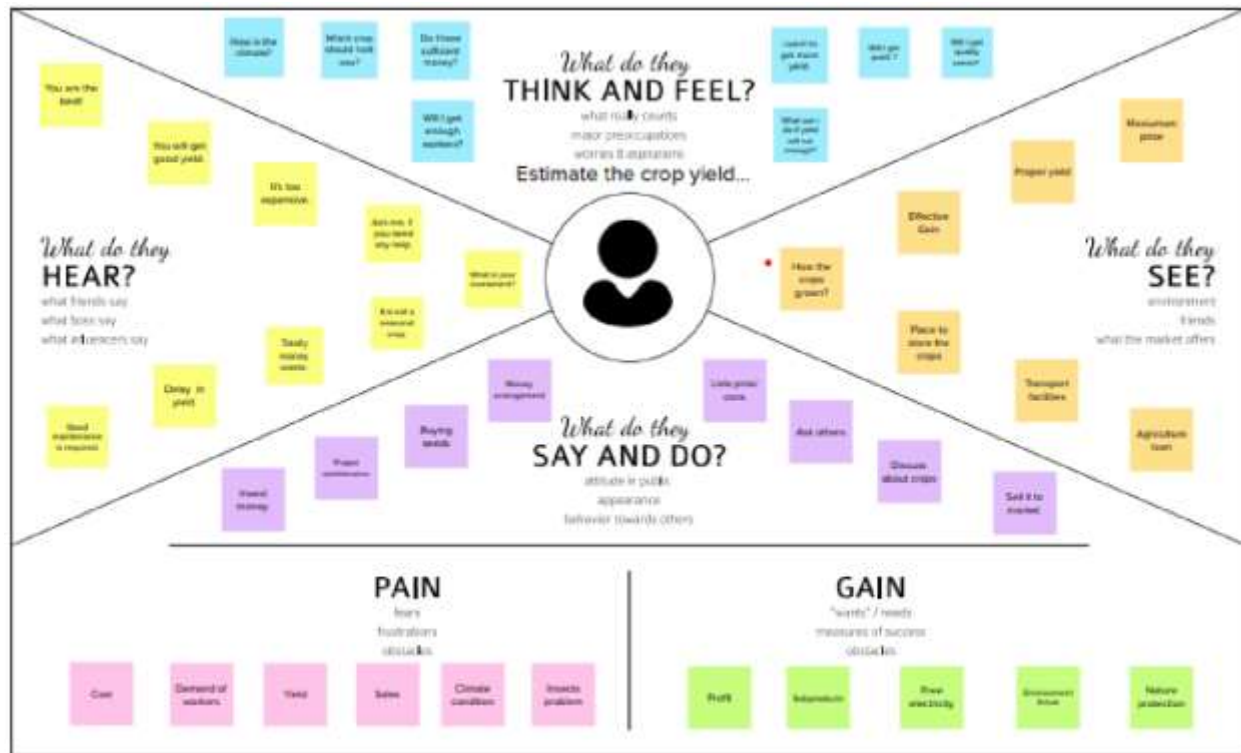
last, all the papers show the pros and cons of the estimation of crop yields. We can analyse and give the solution for the problem using “Data analytics”.

### **2.3 Problem Statement Definition**

This project deals with the estimation of the crop yield in india based on the area, season, state, etc.

## **3. IDEATION & PROPOSED SOLUTION**

### **3.1 Empathy Map Canvas**



### 3.2 Ideation & Brainstorming



### 3.3 Proposed Solution

S. No.	Parameter	Description
1	Problem Statement (Problem to be solved)	This project deals with the estimation of the crop yield in india based on the area, season, state, etc.
2	Idea / Solution description	The key research observations as follows:  ✓ The data's can be visualized using Cognos analytics in various methods.  ✓ Season with average production  ✓ With year usage of area and production  ✓ Top 10 states with most area  ✓ State with crop production  ✓ State with crop production along with season.
3	Novelty / Uniqueness	The uniqueness present in our project is we can easily identify the crop yield based on the seasonal changes and based on the area. Farmers can easily understand the crop sowing.

4	Social Impact / Customer Satisfaction	From the public perception as worst impacts of present data visualization system which does not gave any additional impacts and they have faced
		many losses.
5	Business Model (Revenue Model)	<p>This project mainly helps for business two operations:</p> <p>✓It helps farmers to understand easily so that, the organization will earn profit by getting more responses and feedback. ✓ We can digitally market because the bar, pie charts are easy to understand it reduces time for organization.</p>
6	Scalability of the Solution	We can estimate the crop yield by season, area, and yearly production. So, the farmers can easily predict the suitable crop.

### 3.4 Problem Solution fit

<p><b>1. CUSTOMER SEGMENT(S)</b> Who is your customer?</p> <ul style="list-style-type: none"> <li>❖ Farmers the main customers who will using this to get good profit.</li> <li>❖ Peoples who are entered into farming are our secondary customers.</li> </ul>	<p><b>6. CUSTOMER CONSTRAINTS</b> What constraints prevent your customers from taking action or limit their choices of solutions?</p> <ul style="list-style-type: none"> <li>❖ Budget</li> <li>❖ Land</li> <li>❖ Labor</li> <li>❖ feed quality</li> <li>❖ water, disease</li> </ul>	<p><b>5. AVAILABLE SOLUTIONS</b> 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?</p> <ul style="list-style-type: none"> <li>❖ In the past days they hears about crop sowing but it does not give correct results.</li> <li>❖ Data visualization provides them a good understanding about sowing of crops.</li> </ul>	<p>Explore AS, differentiate</p>
<p><b>2. JOBS-TO-BE-DONE / PROBLEMS</b> Which jobs-to-be-done (or problems) do you address for your customers?</p> <ul style="list-style-type: none"> <li>❖ This provides a best data visualization chart's.</li> <li>❖ It gives the complete guide about which crop to be sowed in the month and based on the area.</li> </ul>	<p><b>9. PROBLEM ROOT CAUSE</b> What is the real reason that this problem exists? What is the back story behind the need to do this job?</p> <ul style="list-style-type: none"> <li>❖ Lack of knowledge about the crop production.</li> </ul>	<p><b>7. BEHAVIOUR</b> What does your customer do to address the problem and get the job done?</p> <ul style="list-style-type: none"> <li>❖ By continuous monitoring they can easily find the problem. By data visualization charts they can get good idea.</li> </ul>	

<p><b>3. TRIGGERS</b> What triggers customers to act?</p> <ul style="list-style-type: none"> <li>❖ By seeing other farmers who are getting more crop production.</li> </ul>	<p><b>10. YOUR SOLUTION</b> 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.</p> <div style="border: 2px solid green; padding: 10px; margin: 10px 0;"> <p><b>SOLUTION:</b></p> <ul style="list-style-type: none"> <li>❖ We are currently working on an existing project of estimation of crop yield production in India.</li> <li>❖ By using this data visualization charts we get best crops for the season and area.</li> <li>❖ Really, Data visualization and dashboard helps the farmers to understand very easy as compared to traditional method and also some paragraphs.</li> </ul> </div>	<p><b>8. CHANNELS of BEHAVIOUR</b> <b>8.1 ONLINE</b> What kind of actions do customers take online?</p> <ul style="list-style-type: none"> <li>❖ Helps to create a new way for them to answer questions.</li> </ul>
<p><b>4. EMOTIONS: BEFORE / AFTER</b> How do customers feel when they face a problem or a job and afterwards?</p> <ul style="list-style-type: none"> <li>❖ At first before using the charts farmers don't know the actual sowing of crops.</li> <li>❖ After using this chart they get idea about the crop sowing and get good yield.</li> </ul>		<p><b>8.2 OFFLINE</b> What kind of actions do customers take offline?</p> <ul style="list-style-type: none"> <li>❖ They can see this data visualization in online and do the crop sowing on their respective land.</li> </ul>

## 4. REQUIREMENT ANALYSIS

### 4.1 Functional requirements



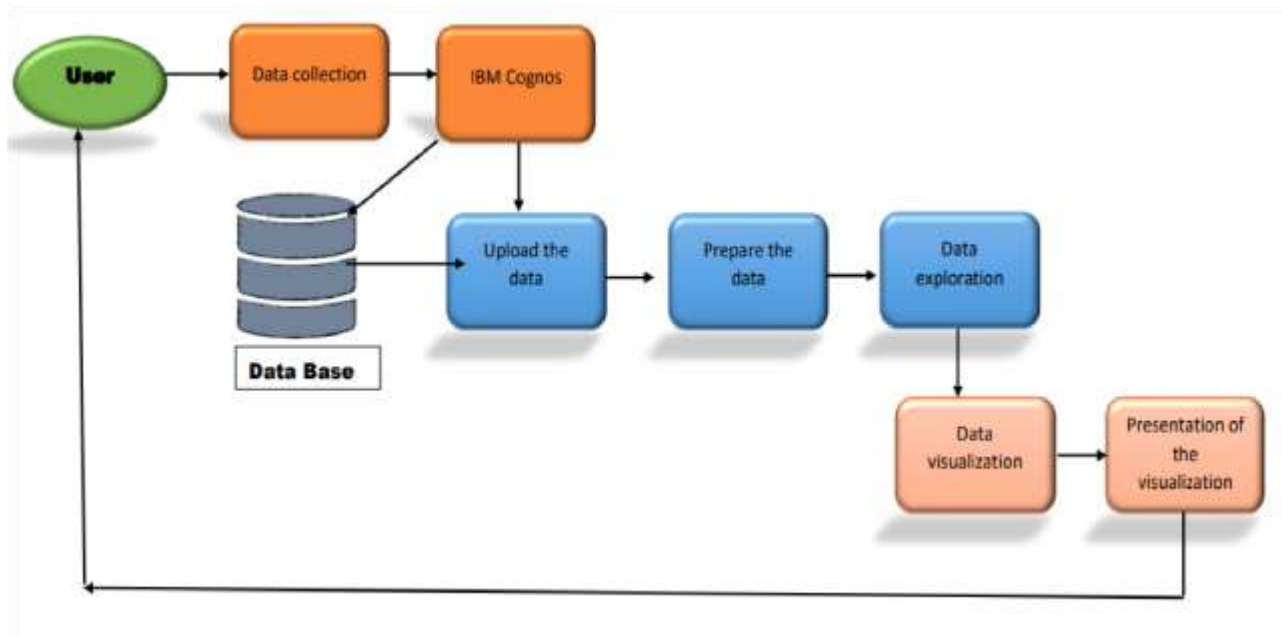
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Gmail Registration through IBM Cognos
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Data collection	Data collection through various resources.
FR-4	Data processing	Data can be optimised for the required needs.
FR-5	Data visualization	Data can be visualized through Cognos

## 4.2 Non-Functional requirements

FR No.	Non-Functional Requirement	Description
NFR-1	<b>Usability</b>	Its user friendly. Everybody who didn't know can easily create.
NFR-2	<b>Security</b>	Secured one. This has verification process.
NFR-3	<b>Reliability</b>	100% reliable due to its security reasons.
NFR-4	<b>Performance</b>	It has a good performance than other ways like by traditional methods.
NFR-5	<b>Availability</b>	Available on everyone's PC and Laptop and It has optimized for the user mobile.
NFR-6	<b>Scalability</b>	Able to adopt for any visualization.

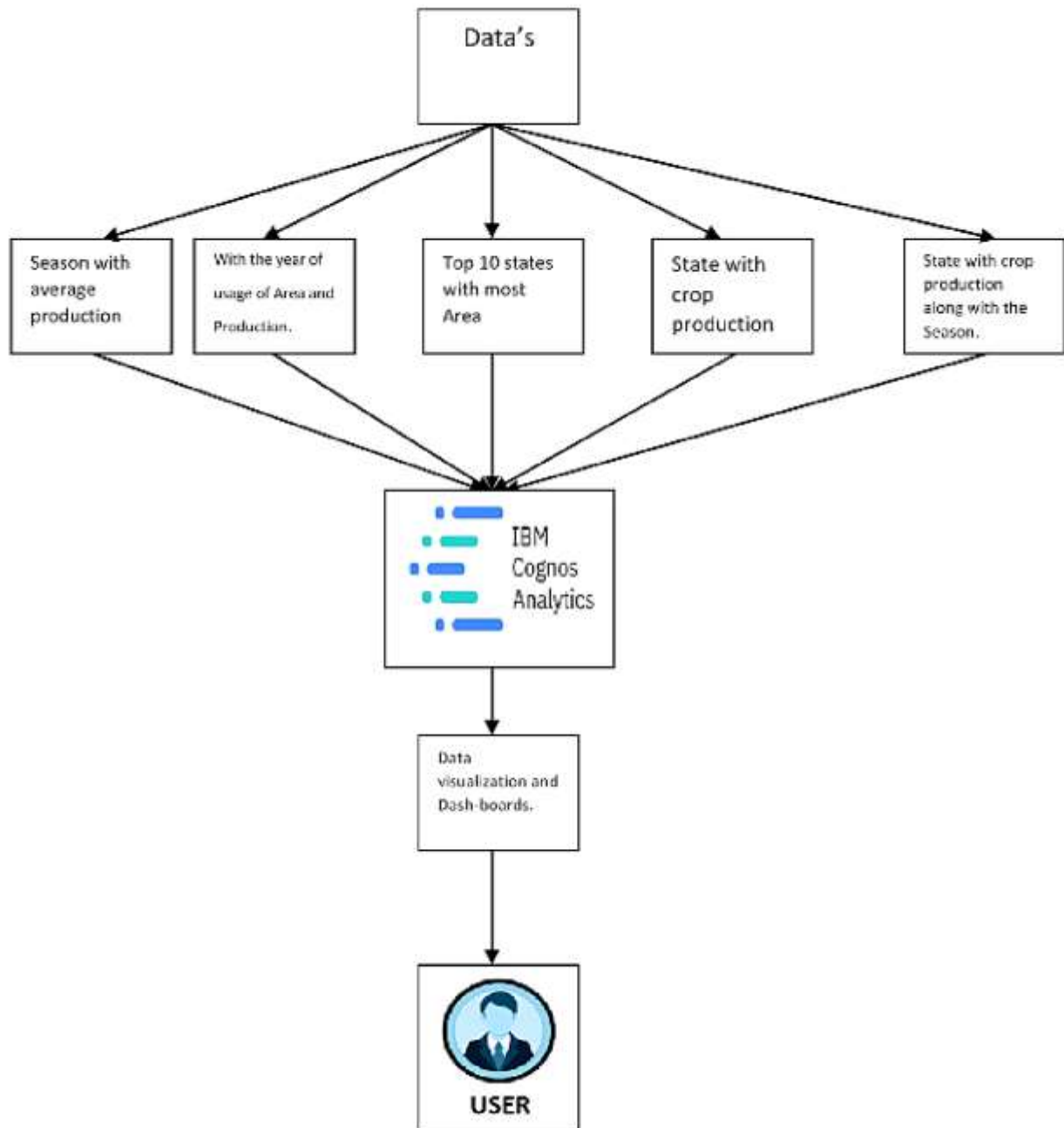
## 5. PROJECT DESIGN

### 5.1 Data Flow Diagrams

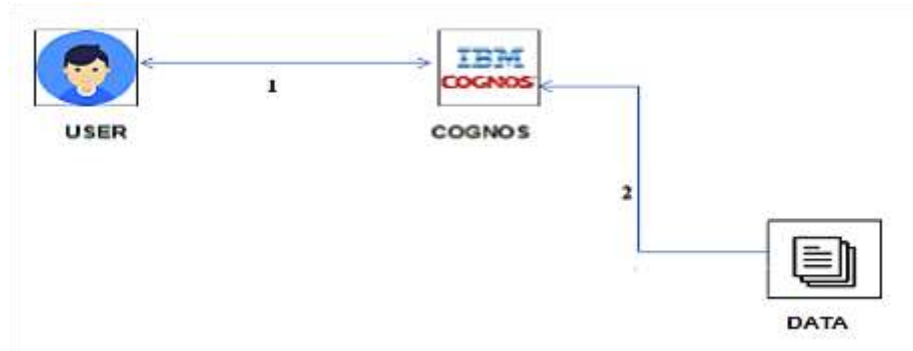




## 5.2 Solution & Technical Architecture



**Solution architecture**



## Technical Architecture

### 5.3 User Stories

User journey Estimate the crop yield using data analytics					
1 Phases	Collecting data		Data cleansing and preprocessing		Model building
2 Steps	Identify the data to be collected	Collect the data from the sensors	Identify the data to be cleaned	Clean the data	Build the model
3 Feelings		Happy to have the data	Happy to have the data	Happy to have the data	Happy to have the data
		Happy to have the data	Happy to have the data	Happy to have the data	Happy to have the data
4 Pain points	Data is not accurate		Data is not clean		Data is not useful
5 Opportunities	Data is not accurate		Data is not clean		Data is not useful

## 6.PROJECT PLANNING & SCHEDULING

### 6.1 Sprint Planning & Estimation

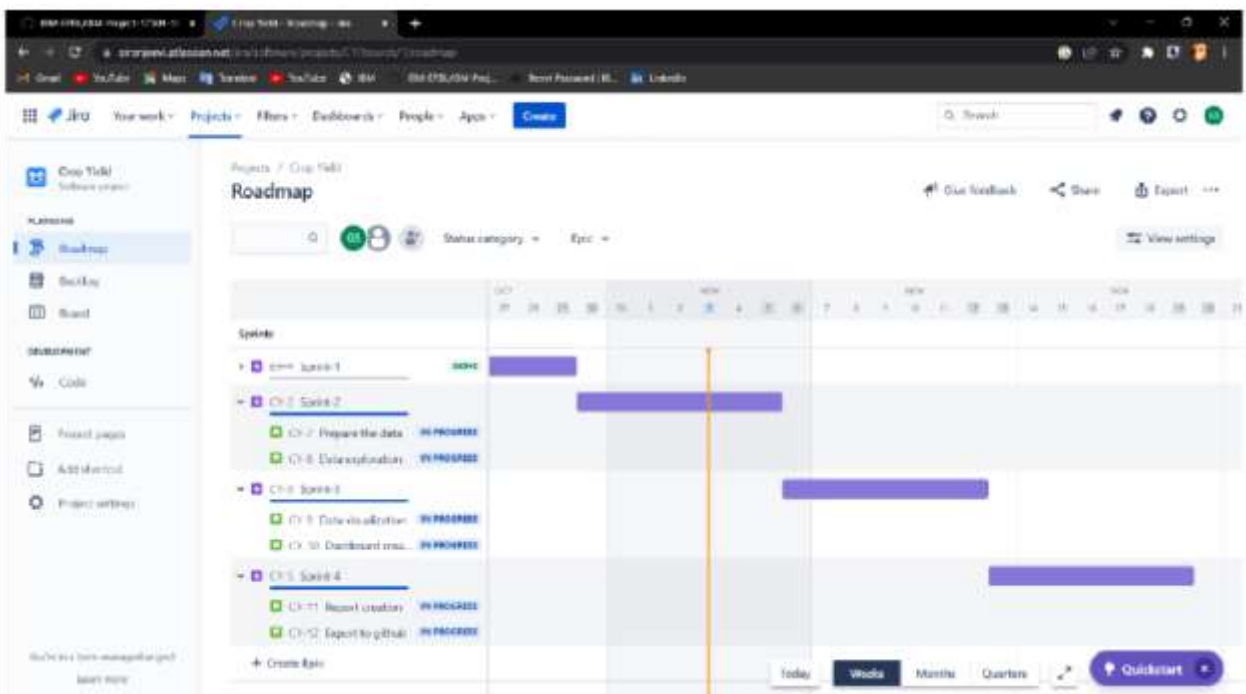
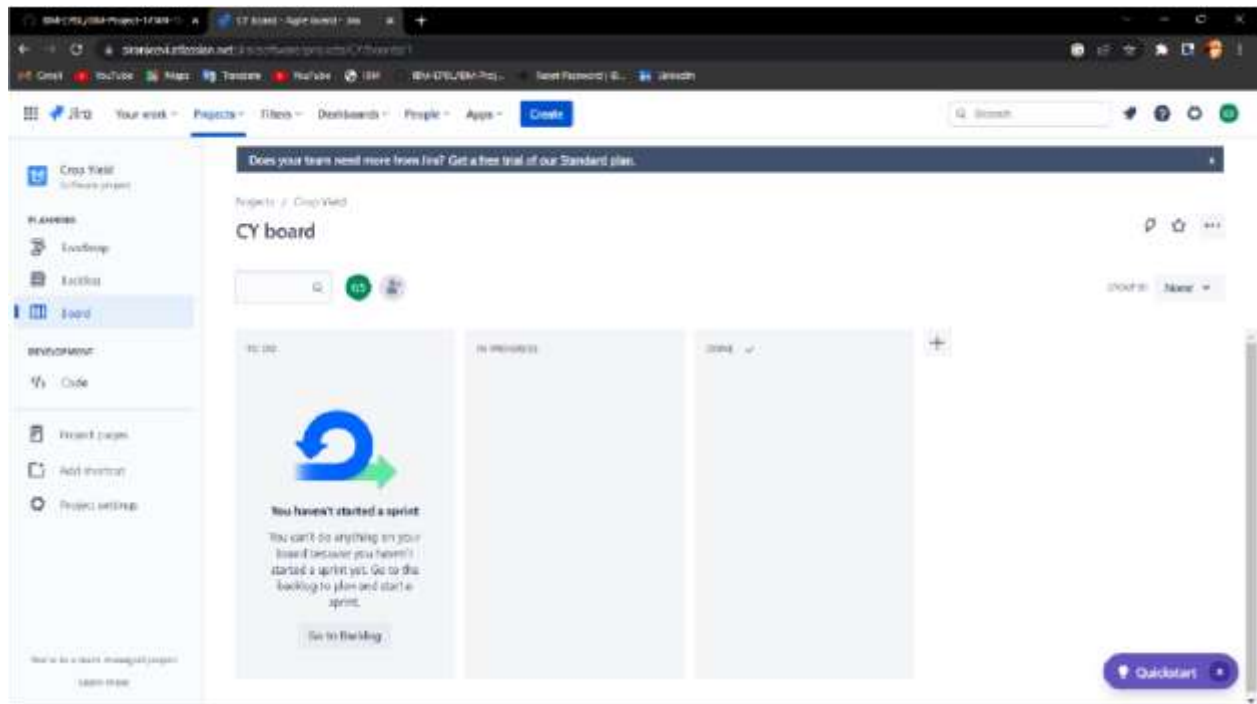
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022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

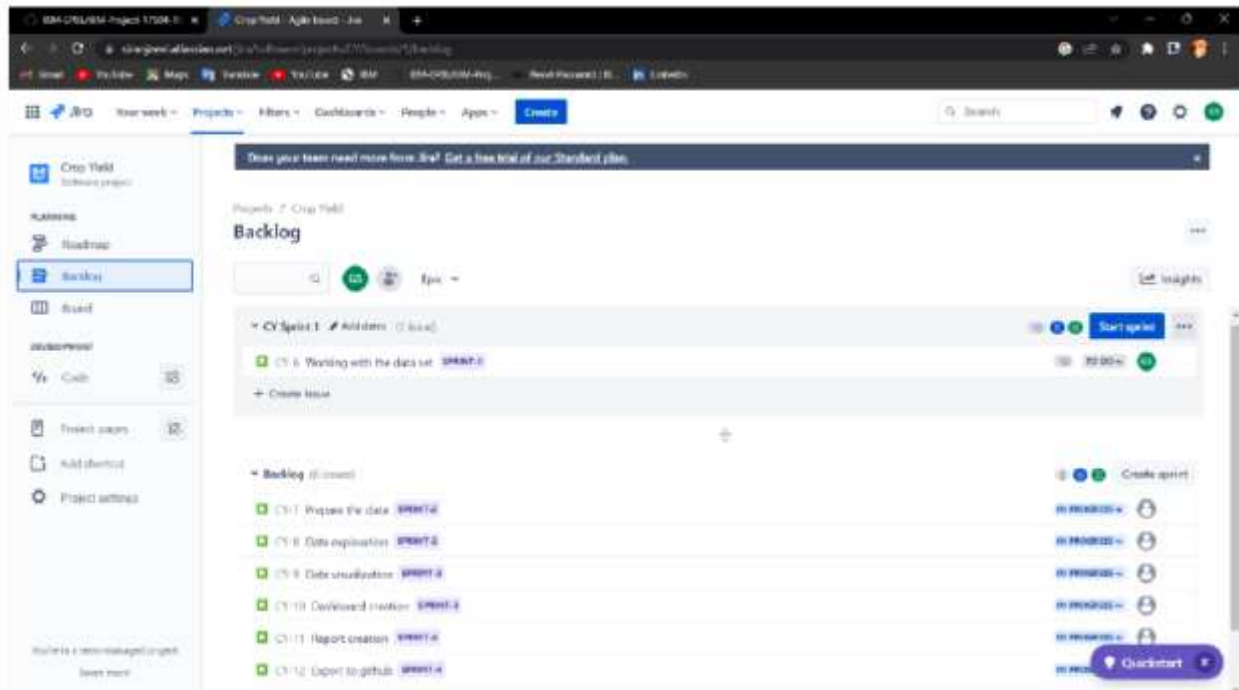
Total sprint points=80

Total sprint=4

Average velocity=80/40=20

## 6.2 Reports from JIRA





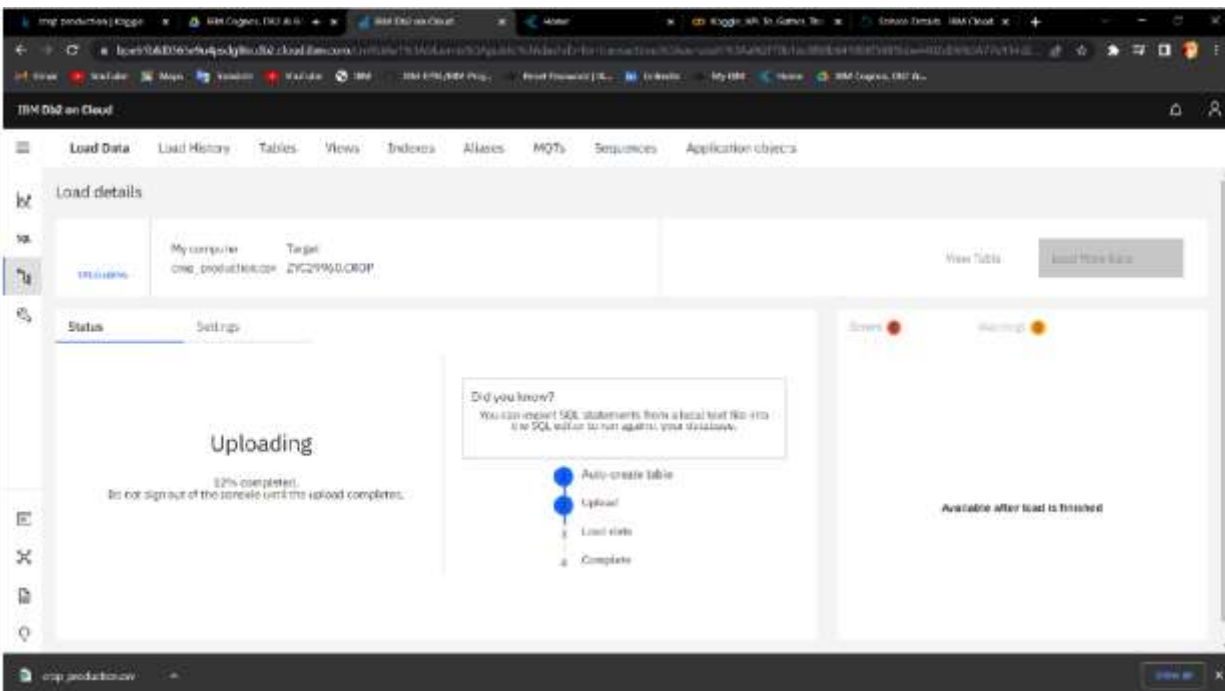
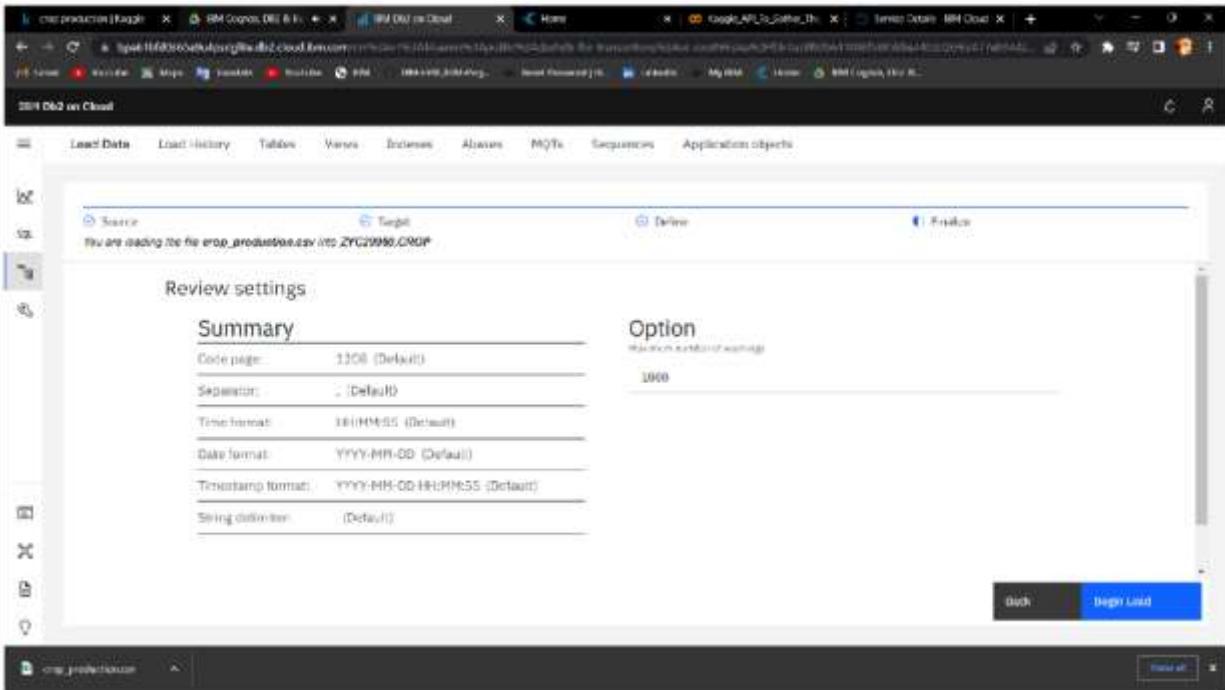
## 7.CODING & SOLUTIONING

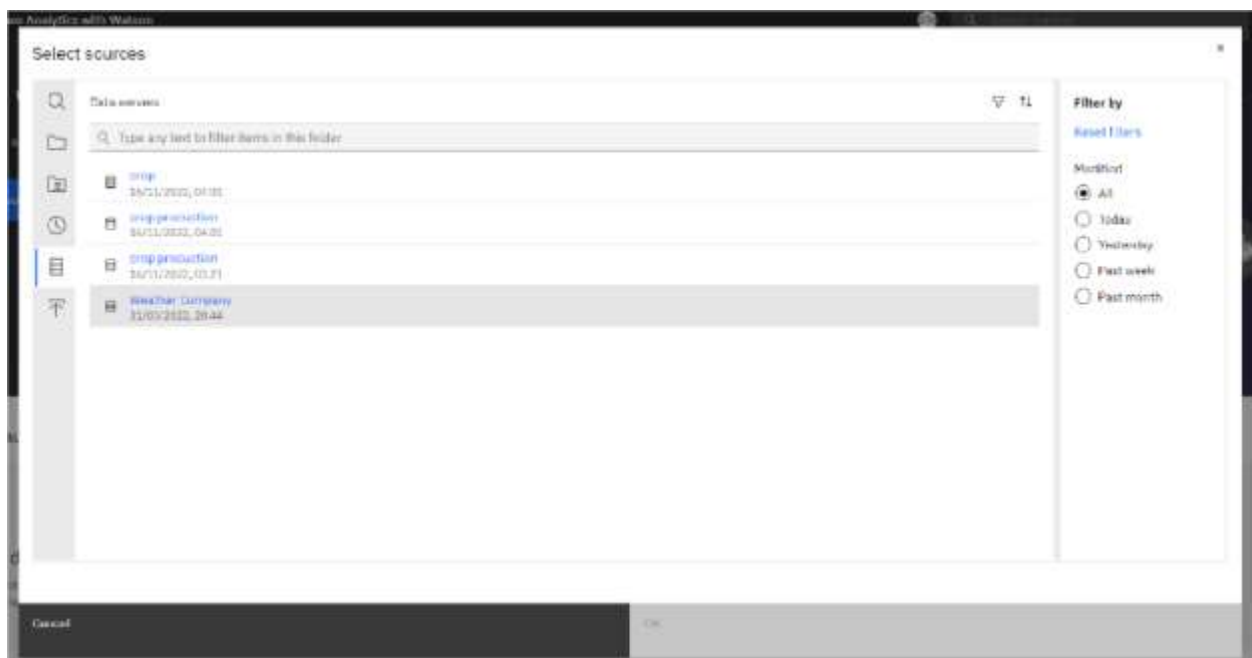
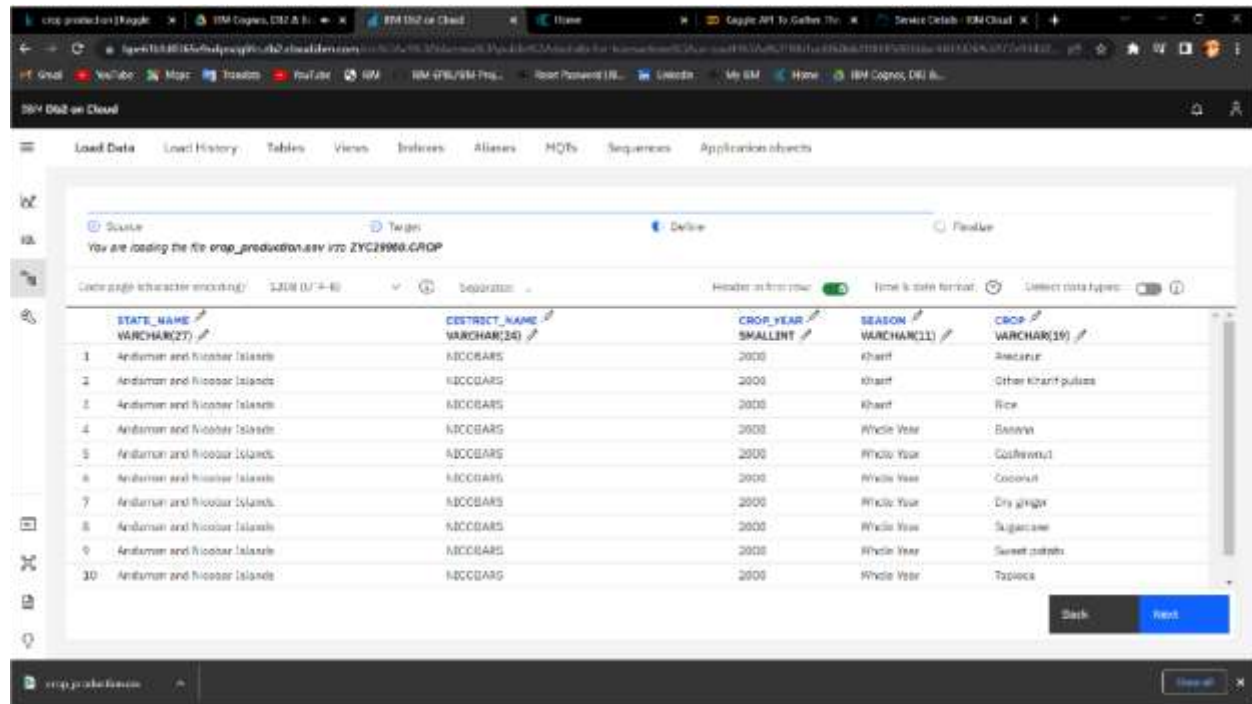
### 7.1 External feature 1

Fetching data from kaggle and IBM DB2 connections









## 7.2 External feature 2

## HTML CODING FOR WEBPAGE

```
<html>
```

```
<head>
```

```
<title> Crop yield estimation </title>
```

```
</head>
```

```
<body>
```

```
<body>
```

```
<body style="background-color:powder blue;"> <h1 style="background-  
color:DodgerBlue;"><p
```

```
style="color:white;">&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;
```

```
&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;
```

```
&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;
```

```
0;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;&#160;<marquee behavior="scroll"
```

```
direction="left">Estimate The Crop Yield Using Data
```

```
Analytics!!!...</marquee>
```

```
</p></h1>
```

```
</h1>
```



<p1 style="font-family:Verdana;"> <br> <h3 style="backgroundcolor:DodgerBlue;"><p style="color:white;">Estimate The Crop Yield Using Data

Analytics</p>

</p></h3>

</p>

<br> Crop production in India is one of the most important sources of income and India is one of the top countries to produce crops. <br>

As per this project we will be analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

<p>Let us look upon an attractive dashboard to know the crop production using IBM COGNOS Analytics!</p>

<p><iframe

src="https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=

.

my\_folders%2Fproject&closeWindowOnLastView=true&ui\_appbar=false&ui\_navbar=false&shareMode=embedded&action=view&mode=dashboard&subView=model0000018460369f2b\_00000000" width="1480" height="720" frameborder="0" gesture="media" allow="encryptedmedia" allowfullscreen=""></iframe></p>

<p style="font-family:veranda;">

</p><br>

<p style="color:DodgerBlue;">Please use the below link to see our Git repository..</p>

<a href="https://github.com/IBM-EPBL/IBM-Project-17504-1659672666">Github

</a>

</p>

<iframe

src="https://us1.ca.analytics.ibm.com/bi/?perspective=story&pathRef=.my\_folders%2FPROJECTFINAL&closeWindowOnLastView=true&ui\_appbar



```
=false&ui_navbar=false&shareMode=embedded&action=view&am  
p;sceneId=model0000018489af65cb_00000002&sceneTime=0"  
width="1480" height="920" frameborder="0" gesture="media"  
allow="encryptedmedia" allowfullscreen=""></iframe>
```

```
</body>
```

```
</html>
```

### 7.3 Database Schema

The Schema which is used in our project is **ZYC29950.CROP**



## **8.2 User Acceptance Testing**

### **Purpose of Document**

The purpose of this document is to briefly explain the test coverage and open issues of the Estimate the crop yield using Data Analytics project at the time of the release to User Acceptance Testing (UAT).

### **Defect Analysis**

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	3	1	1	1	6
Duplicate	1	0	0	0	1
External	0	2	0	0	2
Fixed	1	2	0	2	5
Not Reproduced	0	0	1	0	1
Skipped	0	0	0	0	0
Won't Fix	0	1	0	0	1
Totals	5	6	2	3	16

Section	Total Cases	Not Tested	Fail	Pass
Home	4	0	0	4
Dashboard	2	0	0	2
Report	1	0	0	1
Story	4	0	0	4
Contact Us	4	0	0	4

## 9. RESULTS

The Dataset is collected from external API(Kaggle) and it was prepared for the data visualization charts.

The screenshot shows a Jupyter Notebook interface with the following content:

```
[1]: !pip install -q kaggle
```

```
[2]: !mkdir ~/kaggle & creating a kaggle directory
```

```
[3]: !cp kaggle.json ~/kaggle/ & copying your file to folder
```

```
[7]: !kaggle datasets download -d gaelrandjevi/crop-production
```

Warning: Your Kaggle API key is readable by other users on this system! To fix this, you can run "chmod 600 ~/root/.kaggle/kaggle.json" (crop-production.zip) Skipping, trust more recently modified local copy (you --have to force download)

```
[8]: !kaggle datasets download -d gaelrandjevi/crop-production
```

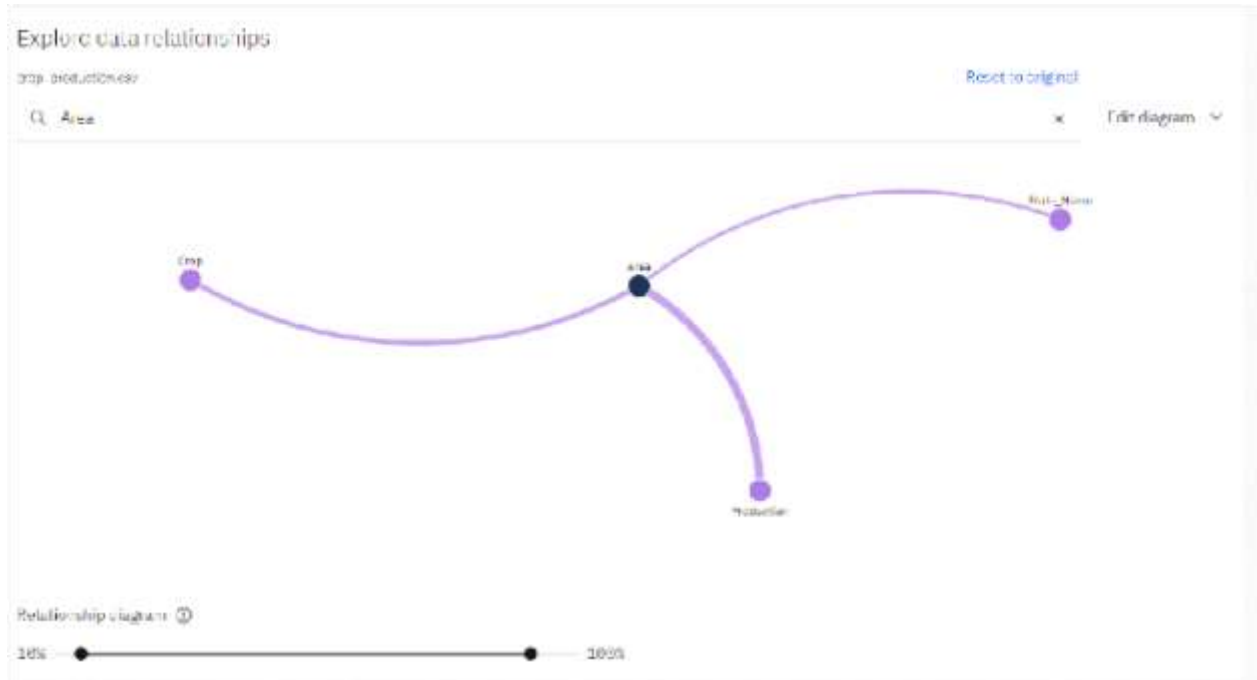
Warning: Your Kaggle API key is readable by other users on this system! To fix this, you can run "chmod 600 ~/root/.kaggle/kaggle.json" (crop-production.zip) Skipping, trust more recently modified local copy (you --have to force download)

The file explorer on the left shows the following files:

- sample\_data
- crop\_production.zip
- crop\_production.csv
- kaggle.json

33

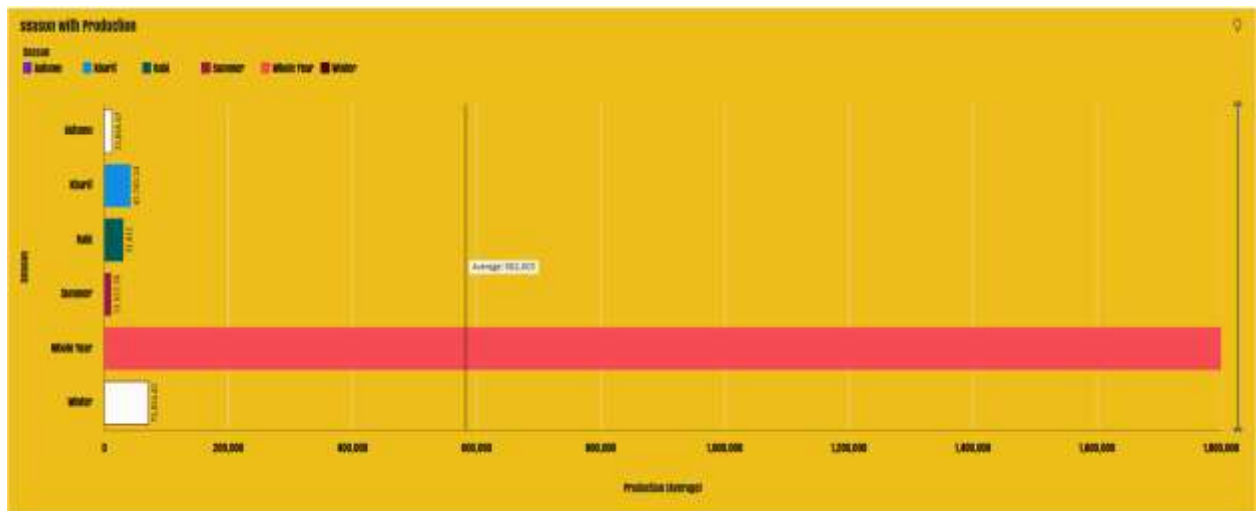


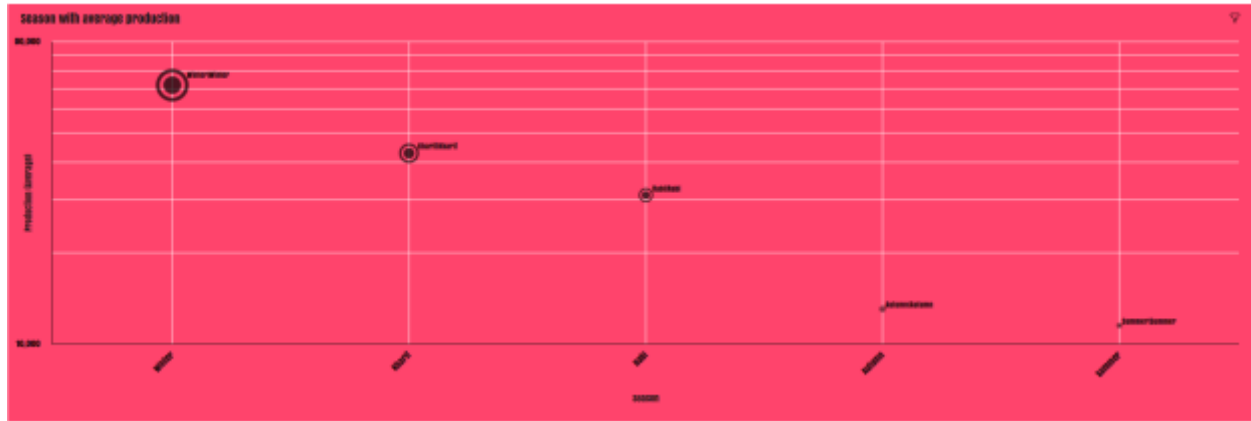




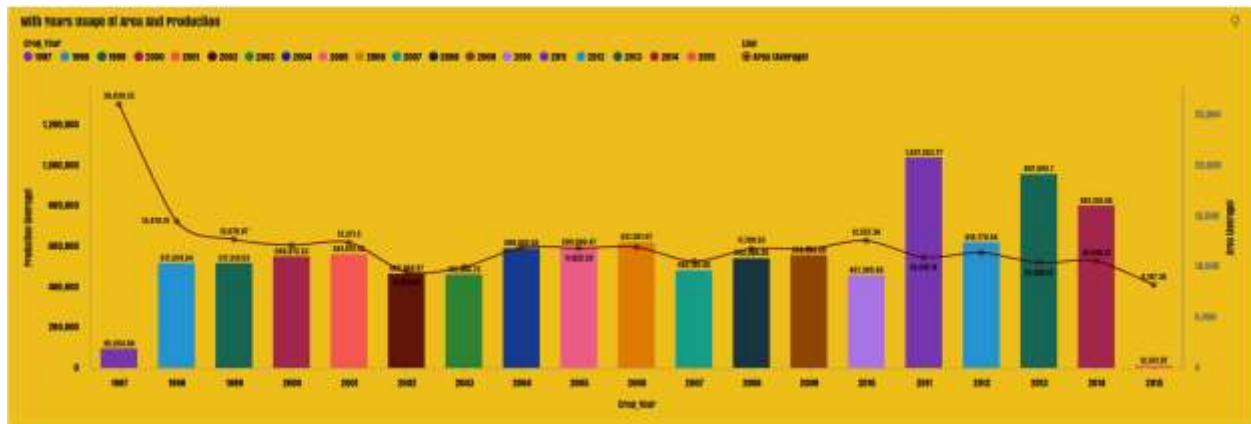
## Data visualization charts

### 1. Season with average production





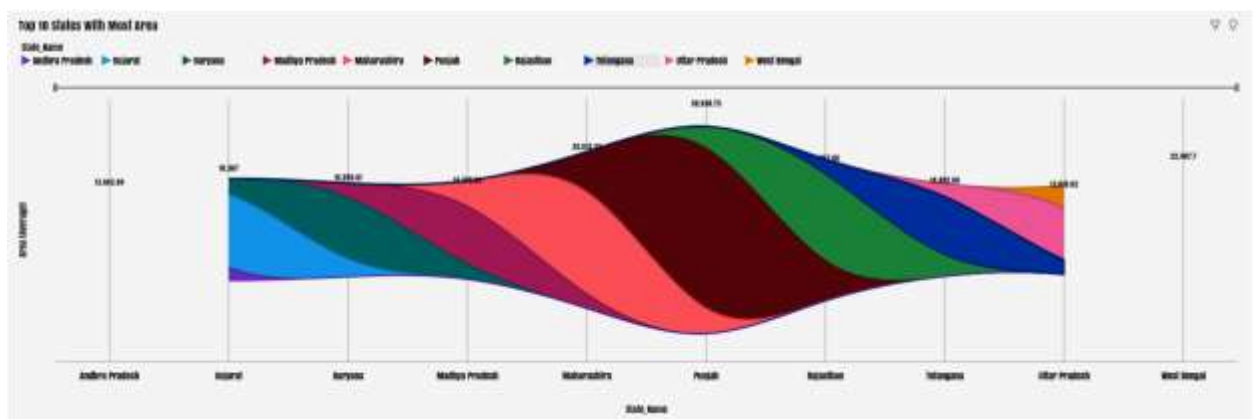
## 2. With year usage of area and production



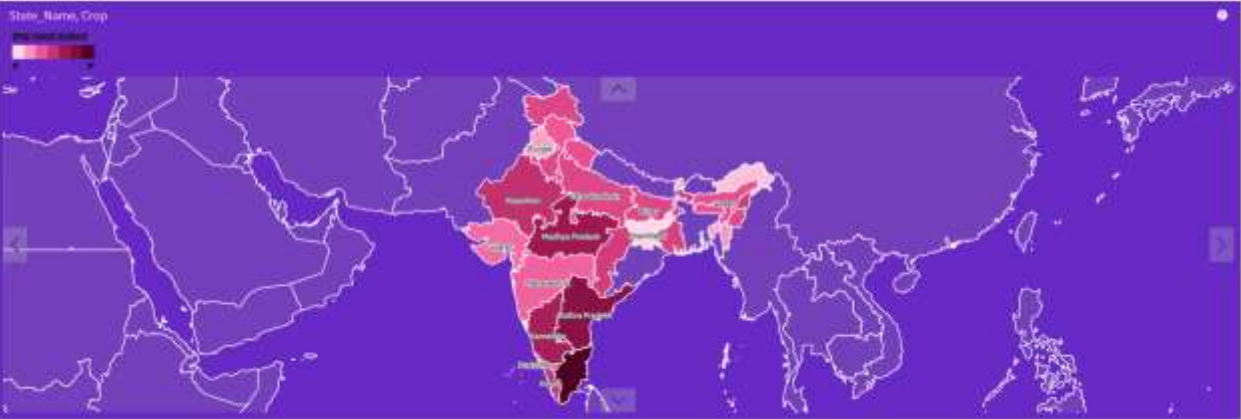
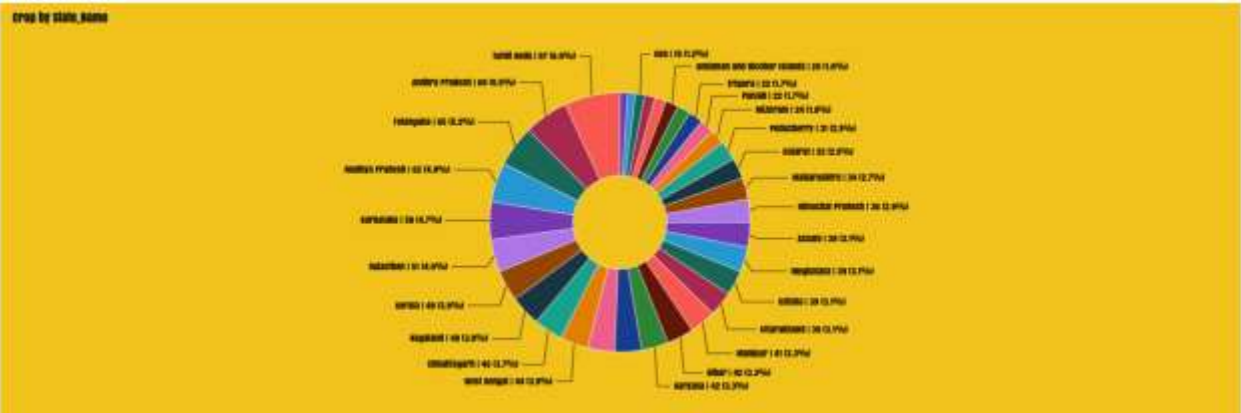




### 3. Top 10 states with most area



### 4. State with production



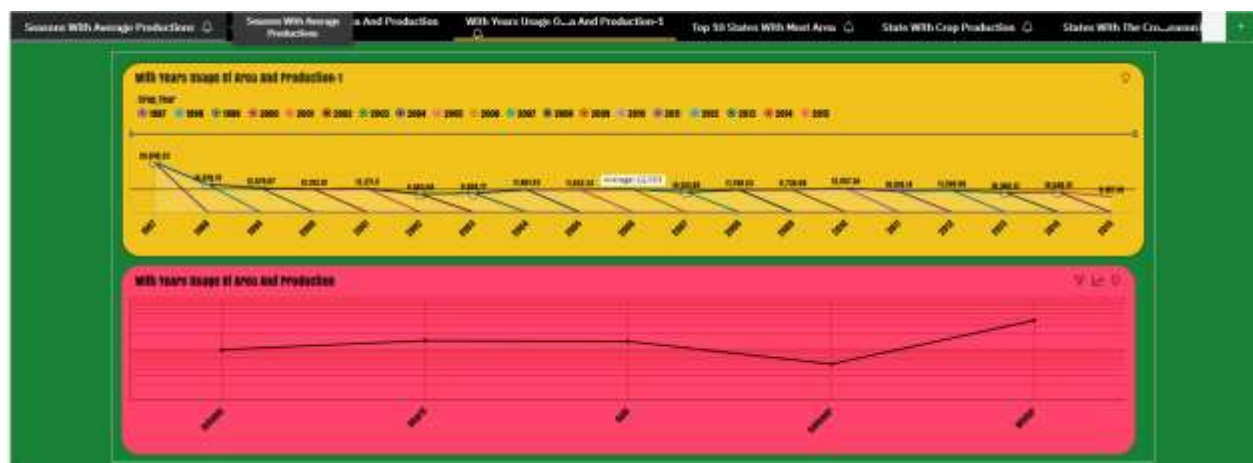
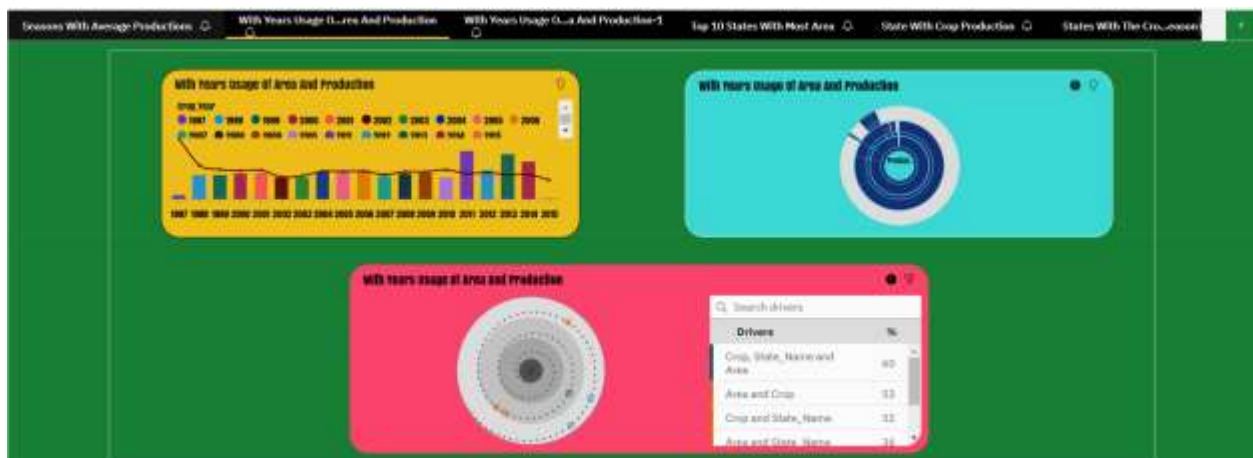


## 5.States with crop production along with season

State_Name and Crop	
Crop	State_Name
Apple	Tamil Nadu
Orange (Processed)	Karnataka
	Andaman and Nicobar Islands
	Andhra Pradesh
	Assam
	Bihar
	Karnataka
	Kerala
	Meghalaya
	Puducherry
	Tamil Nadu
	West Bengal
	Andaman and Nicobar Islands
	Andhra Pradesh
Orange	

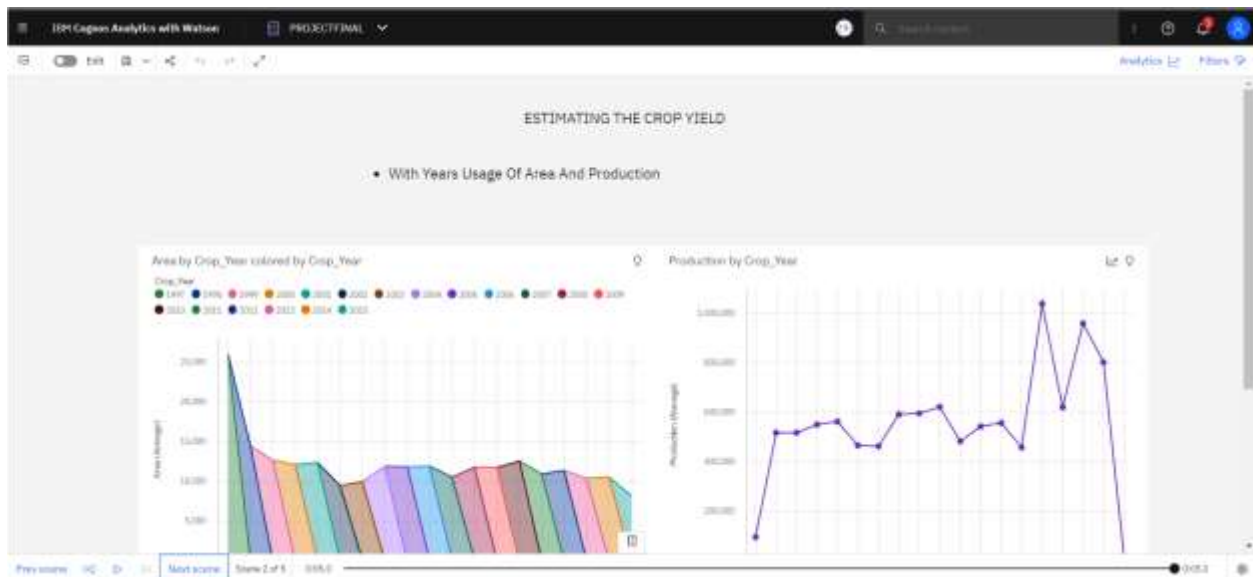
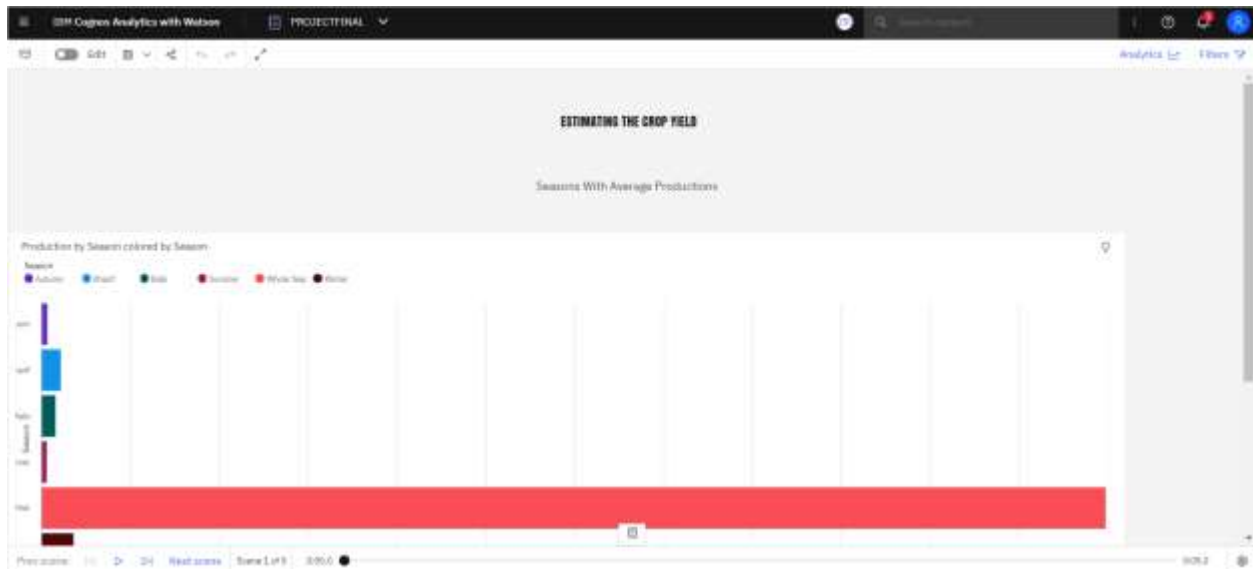
Season and State_Name	
Season	State_Name
Autumn	Odisha
Winter	Andhra Pradesh
	Assam
	Bihar
	Chandigarh
	Chhattisgarh
	Goa and Nagar Haveli
	Gujarat
	Haryana
	Himachal Pradesh
	Jharkhand
	Karnataka
	Madhya Pradesh
	Maharashtra

## Dashboard creation

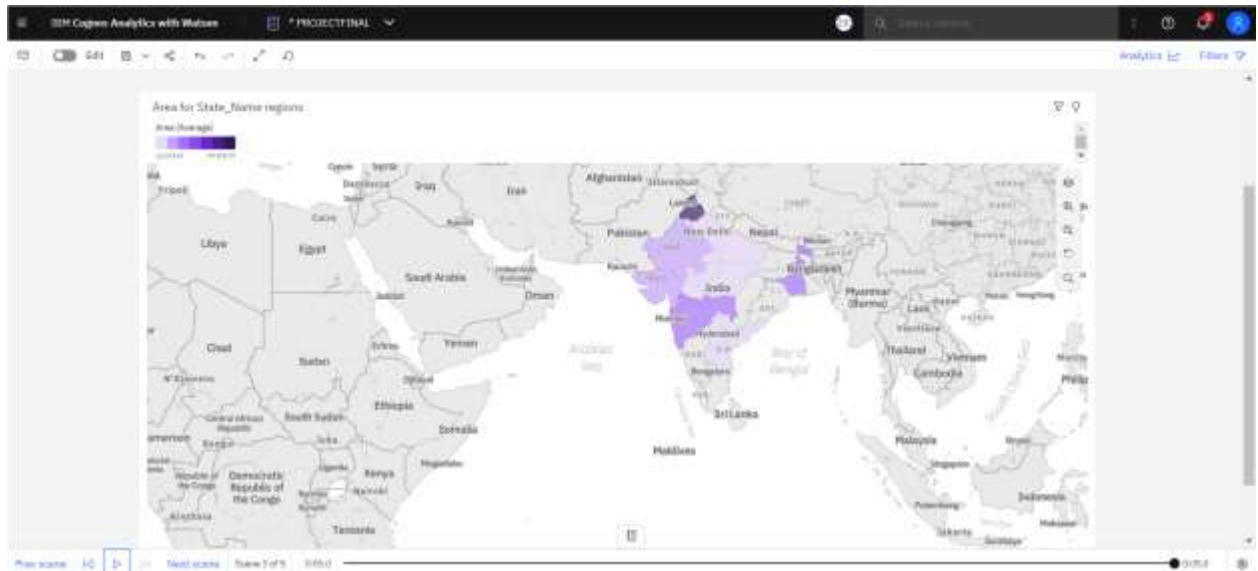




## Creating a story



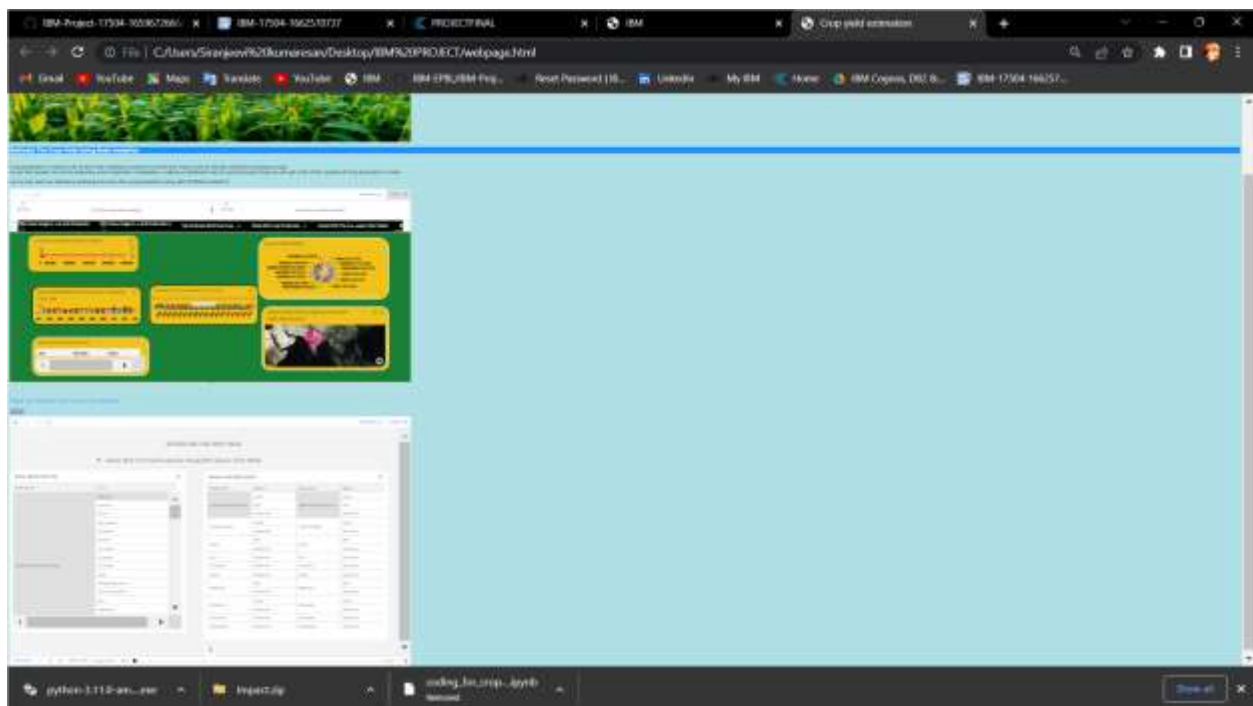
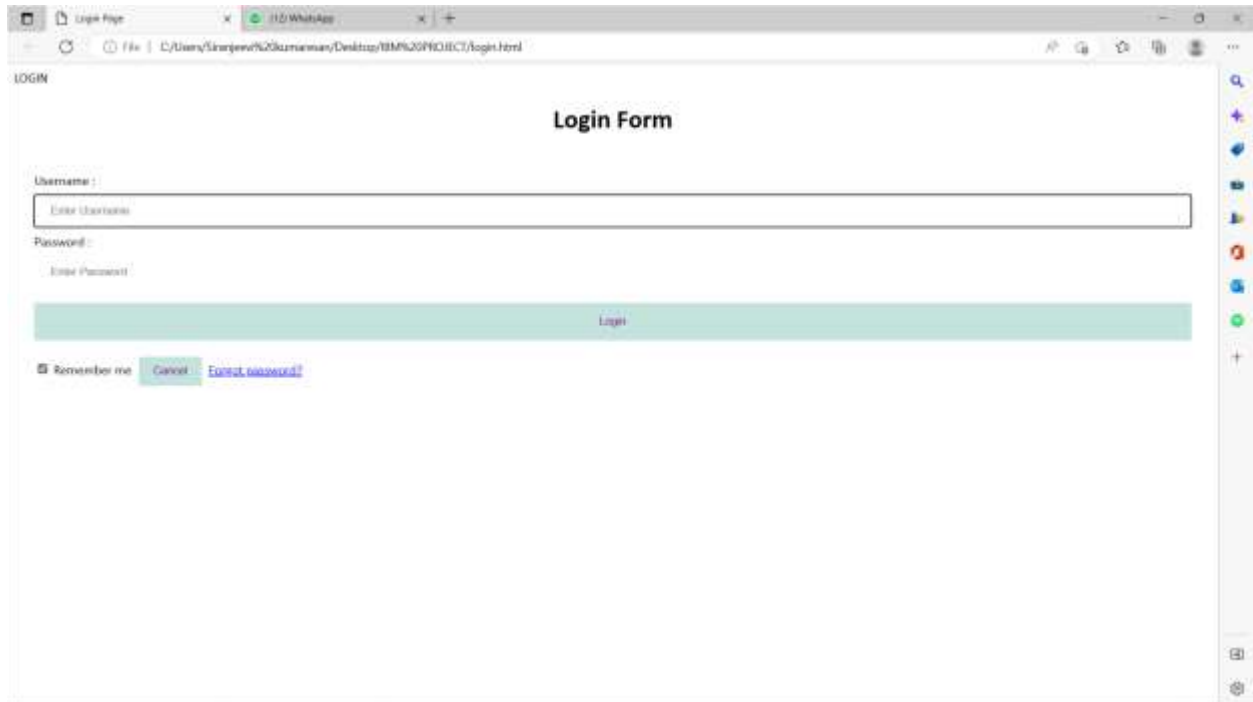


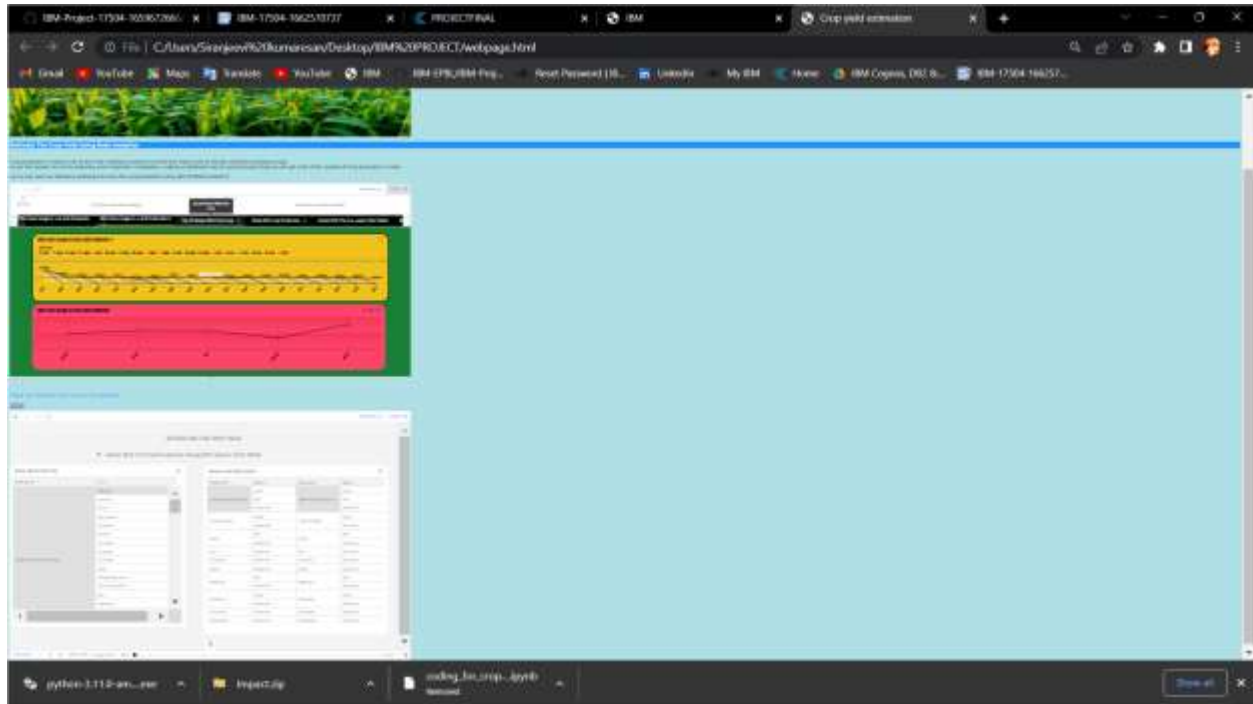


Product and Product Name	Product Name	Product Name	Product Name
Centaurus	Whole Year	Whole Year	Whole Year
Centaur	Half	Half	Half
Dry chives	Whole Year	Whole Year	Whole Year
Dry ginger	Whole Year	Whole Year	Whole Year
Groundnut	Whole Year	Whole Year	Whole Year
Maize	Whole Year	Whole Year	Whole Year
Maize (Green Grain)	Whole Year	Whole Year	Whole Year
Other Wheat Products	Whole Year	Whole Year	Whole Year
Rice	Whole Year	Whole Year	Whole Year
Sugarcane	Whole Year	Whole Year	Whole Year
Sunflower	Whole Year	Whole Year	Whole Year
Sweet potato	Whole Year	Whole Year	Whole Year
Tapioca	Whole Year	Whole Year	Whole Year
Turmeric	Whole Year	Whole Year	Whole Year
Wheat	Whole Year	Whole Year	Whole Year

Website :







## 10 ADVANTAGES & DISADVANTAGES

### 10.1 Advantages

- Among the various methods present for this problem Data visualization

charts helps us to understand the problem easier.

- Pictorial representation so easy to analyse.
- Saves times for the farmers because it gives most of the insights.
- Easily portable to any servers.

- Also helps new persons to sow crops in seasonal base.
- More profitable.

## **10.2 Disadvantages**

- Helps only for the literates because of the usage app is not easier.
- Common language is required to study the charts.
- But this does not give the exact production reports.

## **11. CONCLUSION**

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 analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

## **12. FUTURE SCOPE**

- Farmers get most of the insights about crop production in India.
- As it works for all the land fields the India will get higher production and get

more profits.

### **13. APPENDIX**

Github Link: <https://github.com/IBM-EPBL/IBM-Project-11969-1659363832>

Project Demo link: [https://drive.google.com/file/d/1tO\\_--W8ybNZYo\\_f5z-jy4VZWtklvGr-f/view?usp=drivesdk](https://drive.google.com/file/d/1tO_--W8ybNZYo_f5z-jy4VZWtklvGr-f/view?usp=drivesdk)