Project Report

Date	01-11-2022
Team ID	PNT2022TMID51001
Project name	Estimate The Crop Yield Using Data Analytics

1.Introduction

Agriculture is the main source of livelihood in India. It also contributes a significant figure to the Gross Domestic Product (GDP). About half of the population of India is employed in agriculture, which has the second-largest agricultural land area in the world. As a result, farmers play a crucial role in the industry that produces food for us. The weather has a major impact on agricultural yield. The amount of rainfall has an impact on rice cultivation as well. The farmers in this situation unavoidably need prompt assistance to forecast future crop productivity, and an analysis must be done to assist the farmers in maximizing crop production in their crops. A significant issue in agriculture is yield prediction. Every farmer wants to know how much of a yield to anticipate. In the past, farmer experience with a particular crop was taken into account when predicting production. The amount of data in Indian agriculture is huge. When data is transformed into information, it can be used for a variety of purposes. A web-based comprehensive business intelligence package from IBM is called Cognos. With AI capabilities, IBM Cognos Analytics with Watson advances BI by providing a more accurate, trustworthy, and comprehensive picture of an organization as well as the ability to foresee future events, make predictions, and explain why they could occur.

1.1 Project overview

The overall project is to analyze and visualize the data using various cognos tools and results are converted to story and report. Before there are certain pre requisites and initial process to be followed.

1.2 Purpose

The purpose of this project is to help Farmers in identifying the quantity of their crop yield.
Since they require accurate yield estimates for a number of reasons like crop insurance purposes , delivery estimates, planning harvest and storage requirements, etc.

2. Literature Survey

A literature review is a survey of scholarly sources on a specific topic. It provides an overview of current knowledge, allowing you to identify relevant theories, methods, and gaps in the existing research that you can later apply to your paper, thesis, or dissertation topic.

2.1 Existing problems

1] Agriculture yield prediction using predictive analytic techniques

Numerous exploratory data analyses and different predictive models were developed for this paper. Additionally, different regression models, including linear, multiple linear, and non-linear models, are examined for their ability to accurately estimate or forecast the agricultural production for a variety of crops in the states of Andhra Pradesh and Telangana.

2] An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach

This paper explains how to use a big data technique to find experiences from accurate agricultural information. In this way, effectively acquiring the valuable data pushes a framework toward significant computational hurdles in crop analysis where data is collected remotely. We plan to use the Hadoop framework for our work in order to store a massive amount of crop data for the storage purpose of enormous data availability in agriculture. Hadoop's MapReduce programming model incorporates the random forest technique.

3] Predictive Analysis to Improve Crop Yield using a Neural Network Model

In order to analyse and forecast crop production over seasons in various districts, a data-driven model that learns from historical soil as well as rainfall data has been built. This paper is focused on rice as a specific crop. The intended hybrid neural network model determines the best combinations of soil characteristics and combines them with the pattern of rainfall in a chosen area to evolve the anticipated crop production. The Time-Series technique in Supervised Learning serves as the foundation for the predictive analytic model for rainfall. Recurrent Neural Networks, another subset of machine learning, are the technology utilised to make the ultimate prediction of crop yield.

4] Machine Learning Methodologies for Paddy Yield Estimation in India: a Case Study

In this paper, we give a case study of yield estimation modelling for paddy crop based on meteorological and soil data at various spatial resolution (SR) levels, namely, at the taluk (finer SR) and district (coarser SR) levels in India. We offer a thorough examination of the yield estimation models' accuracy using various feature sets and machine learning (ML) methodologies. Additionally, we disaggregate district yield data by utilising machine learning models that were developed using yield predictions at the taluk level.

5] Crop Yield Analysis Using Machine Learning Algorithms

Two distinct Machine Learning (ML) techniques are suggested in this paper to analyse crop yield. Support Vector Regression (SVR) and Linear Regression (LR) are two techniques that are well suited for verifying the variable parameters in the prediction of continuous variables using the 140 data points that were collected.

6] A survey on Deep Learning Architectures for effective Crop Data Analytics

In several fields, deep learning has become a precise tool for both image- and non-image-based data analytics. For crop data analytics, smart farming is a significant area where deep learning techniques have been successfully employed. Research on deep learning-based models used in crop yield estimation and crop disease detection is reviewed in this paper.

7] Machine Learning based Crop Yield Prediction on Geographical and Climatic Data

The method suggested in this research uses machine learning to analyse geographic and meteorological data in order to estimate the crop yield estimate for a particular plot of land. Examples of these models include Decision Tree Regression, K-Nearest Neighbour Regression, Gaussian Process Regression, and Support Vector Regression.

8] **Weather Based Crop Prediction in India Using Big Data Analytics**

In order to help farmers increase the yield of their crops, this study will gather and analyse data on temperature, rainfall, soil, seed, crop production, humidity, and wind speed (in a few places). We first pre-process the data in a Python environment before using the MapReduce framework to continue processing and analysing the massive amount of data. Second, k-means clustering is used to analyse MapReduce outputs and produces an accurate mean result for the data.

9] A Comparative Analysis of Crop Yield Prediction using Regression

This paper focuses on estimating agricultural yield from available data using various regression models, including Linear Regression (LR), Lasso Regression (LASSO), Decision Tree Regression (DT), and Random Forest (RF) Regression, and determining which model is most effective. The goal of this paper is to provide farmers with the tools they need to forecast and see their crop's yield before they grow it, help them make the best choices, and provide fresh ideas for future research and crop yield forecasting.

10] An Experimental Analysis of Crop Yield Prediction using Modified Deep Learning Strategy

A revolutionary deep learning approach is presented in this paper to assist agricultural fields in accurately predicting crop output levels. The proposed learning scheme is referred to as the Modified Deep Learning Strategy (MDLS). This MDLS is derived from the K-Nearest Neighbour and Decision Tree Algorithms, two common learning frameworks.

2.2 References

1] S. Nagini, T. V. R. Kanth and B. V. Kiranmayee, "Agriculture yield prediction using predictive analytic techniques,"

2] S. Sahu, M. Chawla and N. Khare, "An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach,"

3] S. Kulkarni, S. N. Mandal, G. S. Sharma, M. R. Mundada and Meeradevi, "Predictive Analysis to Improve Crop Yield using a Neural Network Model,"

4] R. B. Guruprasad, K. Saurav and S. Randhawa, "Machine Learning Methodologies for Paddy Yield Estimation in India: a Case Study,"

5] F. F. Haque, A. Abdelgawad, V. P. Yanambaka and K. Yelamarthi, "Crop Yield Analysis Using Machine Learning Algorithms,"

6] N. Johnson, M. B. Santosh Kumar and T. Dhannia, "A survey on Deep Learning Architectures for effective Crop Data Analytics,"

7] S. V and A. Padyana, "Machine Learning based Crop Yield Prediction on Geographical and Climatic Data,"

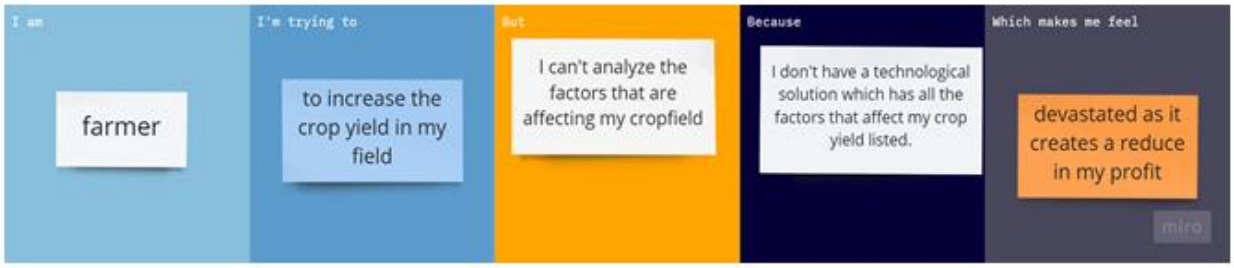
8] R. Gupta et al., "WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics,"

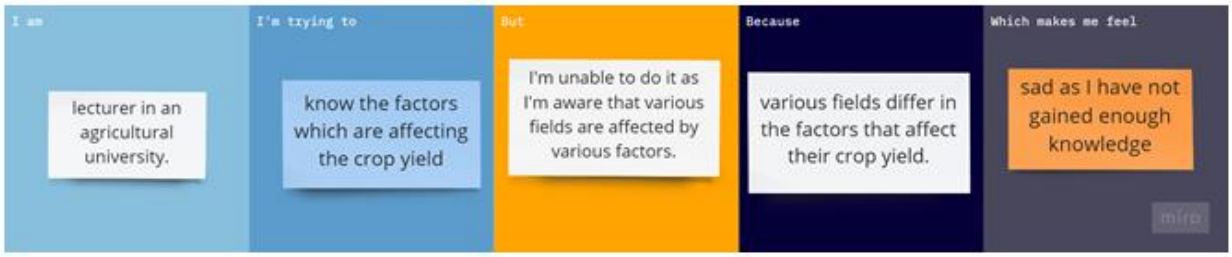
9] S. Rai, J. Nandre and B. R. Kanawade, "A Comparative Analysis of Crop Yield Prediction using Regression,"

10] P. S. Bharathi, V. Amudha, G. Ramkumar, T. J. Nagalakshmi, N. Nalini and P. Jagadeesh, "An Experimental Analysis of Crop Yield Prediction using Modified Deep Learning Strategy,"

2.3 Problem Statement Definition

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 using a IBM cognos and by going through these we will get most of the insights of Crop production in India.



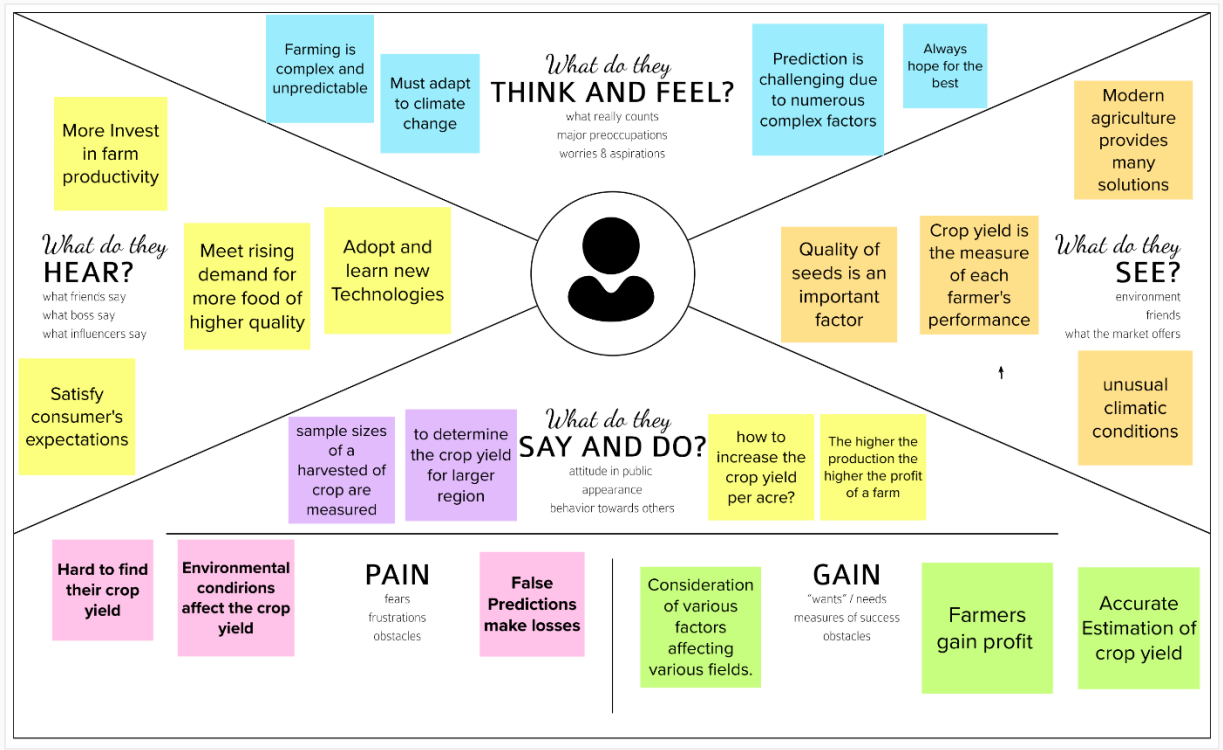


3. Ideation & Proposed Solution

In this phase we have gathered all the ideas of our team members and took an analysis in which we have discussed about the topics which were more relevant and essential. Finally we came up with the most suitable ideas to our project.

3.1 Empathy map canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors 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.



3.2 Ideation and Brain storming

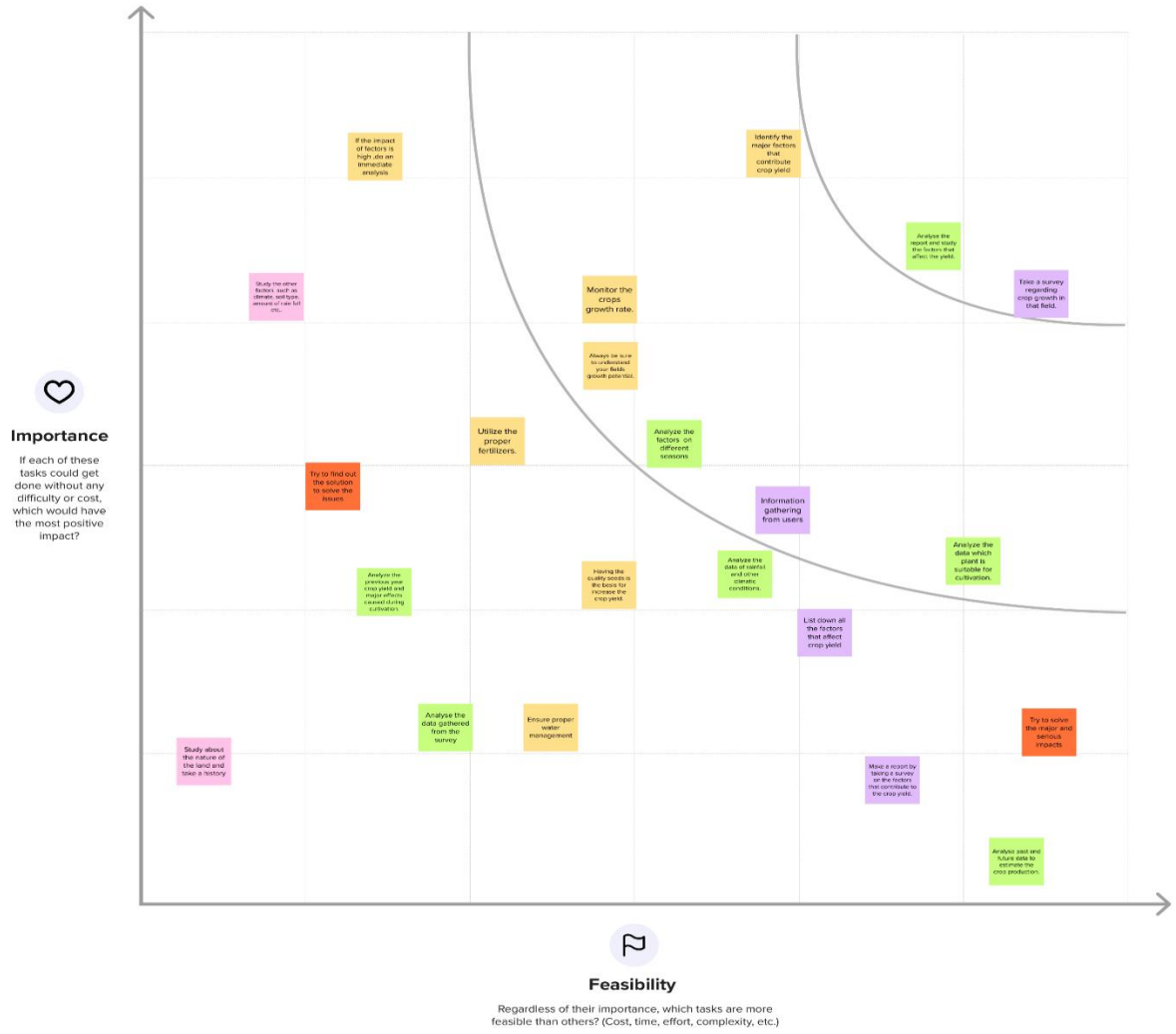
Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all the team members are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

The following are the steps followed in brain storming.

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

Step-2: Brainstorm, Idea Listing and Grouping

Step-3: Idea Prioritization



3.3 Proposed solution

The largest source of income comes from agriculture, which is one of the top producing nations in the world. As part of this assignment, we will assess a few essential visualizations, create a dashboard with IBM Cognos, and use them to analyze crop yield in India to the greatest extent possible. With IBM Cognos Analytics, we can analyze the data and reach well-informed conclusions by combining reporting, modelling, analysis, exploration, dashboards, stories, and event management. A dashboard summarizes key observations and analysis about our data on one or more pages or displays, allowing us to track events or actions rapidly. In this project, a dashboard is used to view, analyze, and extract the majority of the findings.

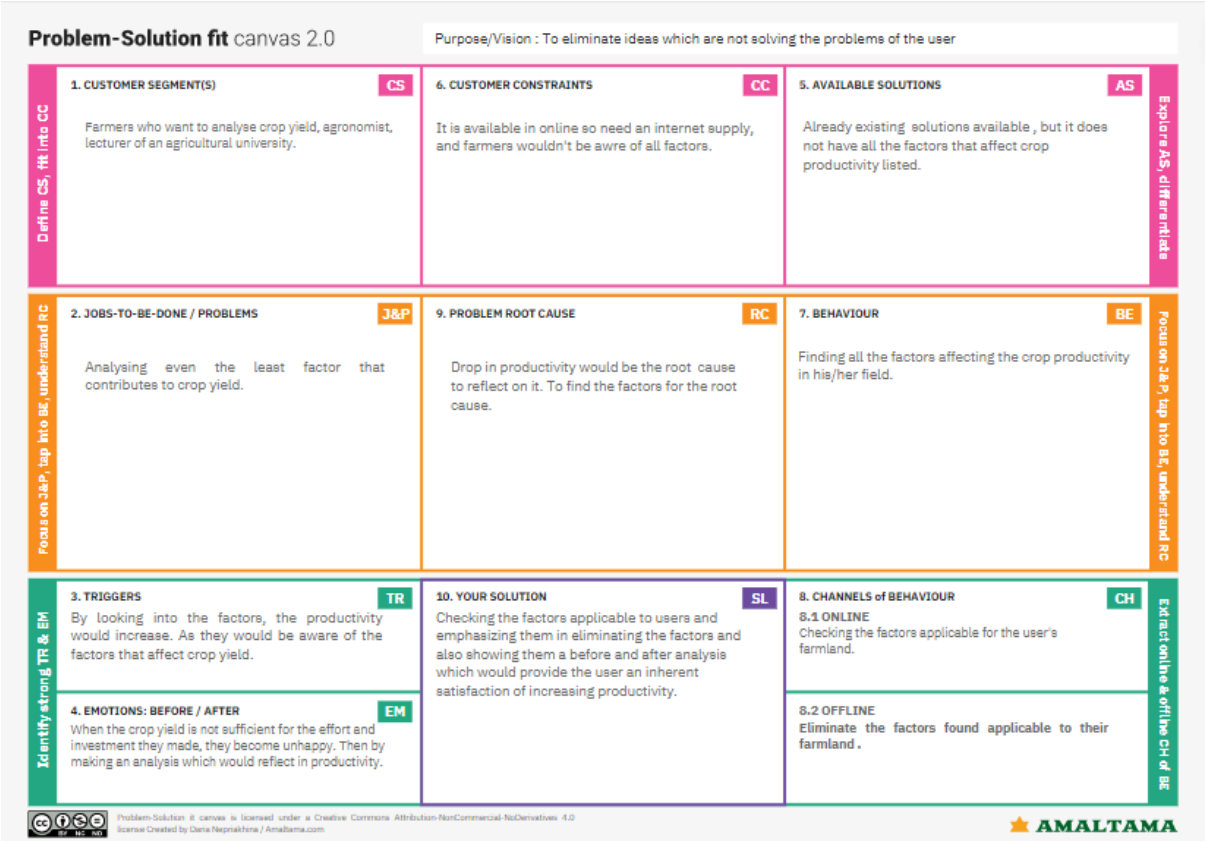
S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Crop production in India is one of the most important sources of income and India is

		one of the top countries to produce crops. As per this project we will be analyzing some important visualization, creating a dashboard using IBM Cognos and by going through these we will get most of the insights of Crop production in India.
2.	Idea / Solution description	We can comprehend the data and make wise decisions by integrating reporting, modelling, analysis, exploration, dashboards, stories, and event management with IBM Cognos Analytics. By presenting critical insights and analyses about our data on one or more pages or screens, a dashboard enables us to keep track of events or actions at a glance. In this project, we use a dashboard to view, analyse, and extract the majority of the findings.
3.	Novelty / Uniqueness	Consideration of all factors that affect crop yield.
4.	Social Impact / Customer Satisfaction	By considering all factors, the customer would gain knowledge about all the minute factors that would affect the crop yield.
5.	Business Model (Revenue Model)	By gaining knowledge about all the minute factors that would affect the crop yield, the customer will result in an increase in profit by correcting even the minute factors.
6.	Scalability of the Solution	Can be used for any type of field at any part of the world.

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. It is used to solve complex problems in a way that fits the state of

your customers. Also it sharpens your communication and marketing strategy with the right triggers and helps to Understand the existing situation in order to improve it for your target group.



4. Requirement Analysis

The data and tools which are required for our analysis is identified and gathered accordingly. Then we have analyzed the characteristics and features of the data in different ways.

4.1 Functional Requirements

This define what a product must do, what its features and functions are,

They are product features or functions that developers must implement to enable users to accomplish their tasks. Generally, functional requirements describe system behavior under specific conditions. Functional requirements can be classified according to different criteria. For example, we can group them on the basis of the functions a given feature must perform in the end product.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail

FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Collection of Data	Uploading datasets
FR-4	Data pre processing	Removing noise data and missing values
FR-5	Data visualization	Creating various charts using graphical visualization tools
FR-6	Data analyzing	Extracting information from the graphical charts

4.2 Non Functional Requirements

These are not related to the system functionality, rather define how the system performs

Here we'll just briefly describe the most typical nonfunctional requirements.

Usability: It defines how difficult it will be for a user to learn and operate the system. It can be assessed from different points of view.

Security: It ensures that the software is protected from unauthorized access to the system and its stored data.

Reliability: It defines how likely it is for the software to work without failure for a given period. It decreases because of bugs in the code, hardware failures, or problems with other system components.

Performance: It is a quality attribute that describes the responsiveness of the system to various user interactions with it. Poor performance leads to negative user experience. It also jeopardizes system safety when it's overloaded.

Availability: It is gauged by the period that the system's functionality and services are available for use with all operations.

Scalability: It describes how the system must grow without negative influence on its performance. This means serving more users, processing more data, and doing more transactions.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Easy to Use, customer can login IBM Cognos and views the data visualization charts
NFR-2	**Security**	User verification is done through mail and access limits varies for different user levels.
NFR-3	**Reliability**	Only authorised person can edit the visualization charts and there will be no controversies.
NFR-4	**Performance**	Changes are auto saved, easy to operate and create charts.
NFR-5	**Availability**	Customer can read the dataset and view the

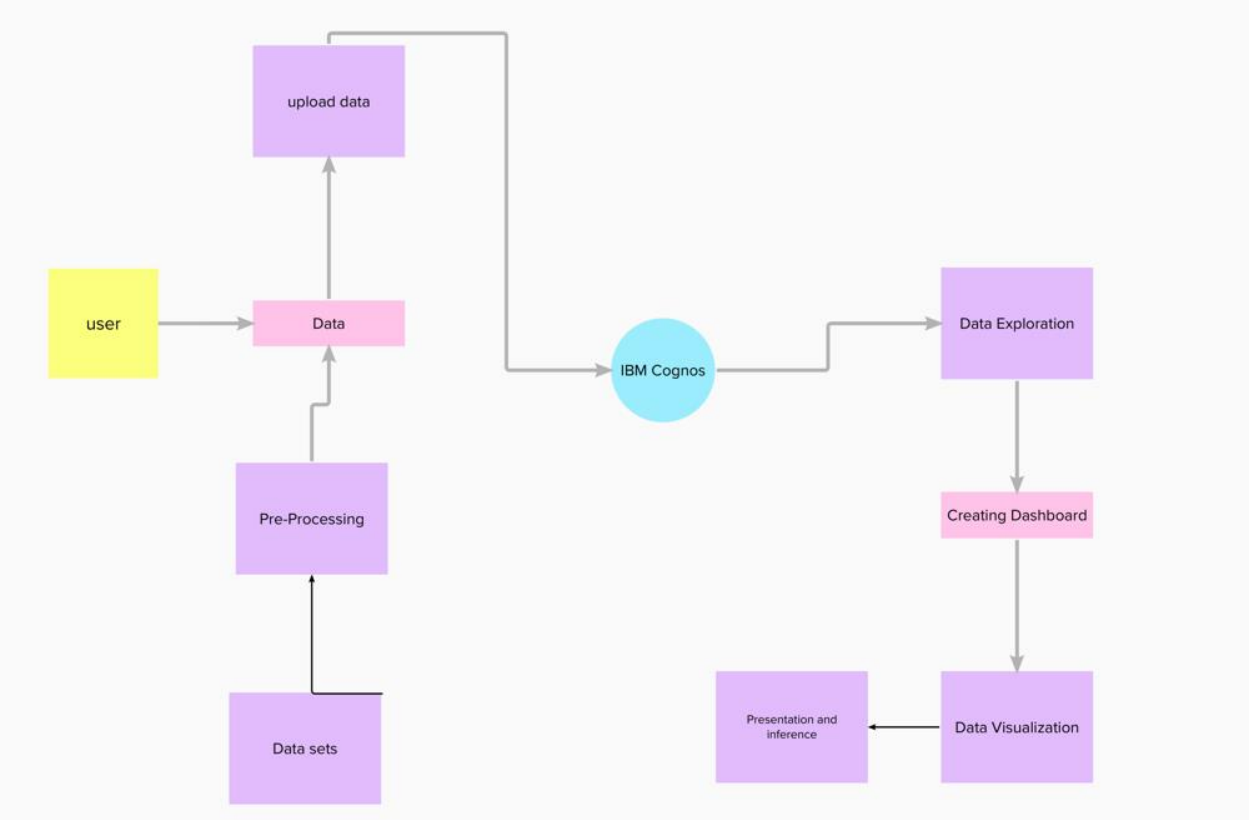
		visualization charts.
NFR-6	**Scalability**	Scalable and suitable for any kind of data set. Visualize variations with different parameters.

5. Project Design

In this phase the methodology and functionalities of the project are framed and implemented. An overall execution is mentioned.

5.1 Data flow diagram

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 Solution and Technical architecture

The Technical Architecture describes the infrastructure required to support application, operations, and requirement. In this we have stored various data of crop yield in the cloud. Then we perform data pre-processing, data exploration and data presentation. All the three process is done in IBM cognos analytics. Finally the user when login to Cognos, they can see the data visualization charts in Dashboard.

5.3 User Stories

User stories describes about the activities of user, they could perform in IBM cognos. project user stories are split into various sprint deliverables

- Create an IBM account.
- As a user one can click the url provided to view the dashboard.
- To view the dashboard they must login to IBM Cognos Analytics.
- Then it takes to the Home page, thereby choosing dashboard option, all the visualizations charts created are displayed.
- User can save the dashboard and share it through a link or e-mail.
- Other options like report and story can also be viewed.

6. Project scheduling and planning

In this project planning phase we have divided the project into different stages. Each stage is broken down into several iterations called **Sprints.** Each sprint has a different length typically one to two weeks and the team has a predefined list of items to work through in each sprint. The work items are called **user stories**.

6.1 Sprint plan

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Point	Priority	Team Members	Number of meetings conducted
Sprint-1	IBM Cloud Account Creation	1	Only after creating an IBM account, the user would be enabled to use IBM's services	2	Medium	Dharini B, Gifty Sharon K, Bhuja Shri R, Dharsana R, Kavithanjali K	2
	IBM Cognos Account Creation	2	Account Creation of IBM Cognos is necessary as it is used for visualization	5	High	Dharini B, Gifty Sharon K, Bhuja Shri R, Dharsana R, Kavithanjali K	
	Working with the Dataset	3	Understanding the Dataset	5	Medium	Dharini B, Gifty Sharon K,	
		4	Loading the Dataset	8	High	Bhuja Shri R, Dharsana R, Kavithanjali K	
	Data Visualization	5	Seasons with Average	2	High	Dharini B	

Sprint-2	Charts		Productions				5
		6	With Years Usage of Area and Production	2	High	Gifty Sharon K,	
		7	Top 10 States with Most Area	2	High	Kavithanjali K	
		8	State with Crop Production	2	High	Dharsana R,	
		9	States with The Crop Production Along with Season	2	High	Bhuja Shri R,	
	Creating the Dashboard	10	A Dashboard has to be created for the visualizations to be presented	5	High	Bhuja Shri R, Dharsana R, Kavithanjali K	
	Export the Analytics	11	The Dashboard is being exported	5	High	Dharini B, Gifty Sharon K	
Sprint-3	Creating The Report	12	A Report has to be created for the visualizations to be presented	5	High	Dharini B, Gifty Sharon K	2
	Export the Report	13	The Report is being exported	5	High	Bhuja Shri R, Dharsana R, Kavithanjali K	
Sprint-4	Creating The Story	14	A Story has to be created for the visualizations to be presented	5	High	Bhuja Shri R, Dharsana R, Kavithanjali K	2
	Export the Story	15	The Story is being exported	5	High	Dharini B, Gifty Sharon K	

6.2 Sprint delivery schedule

Project tracker:

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	12	6 Nov 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	0	8 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	10	9 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	10	9 Nov 2022

Velocity:

we have a 24-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

AV = Sprint Duration / Velocity = 12 / 20 = 0.6

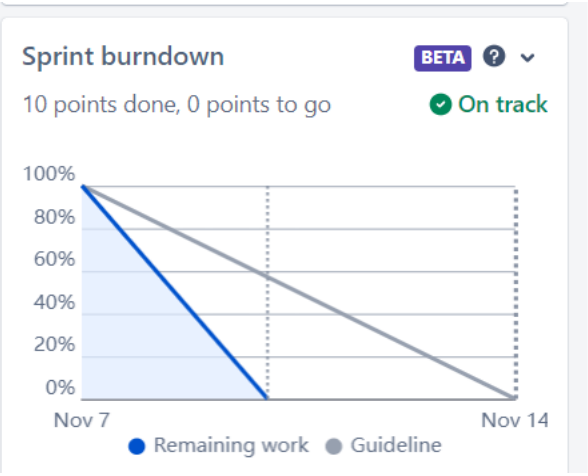
The above average velocity is for Sprint-1, Sprint-2

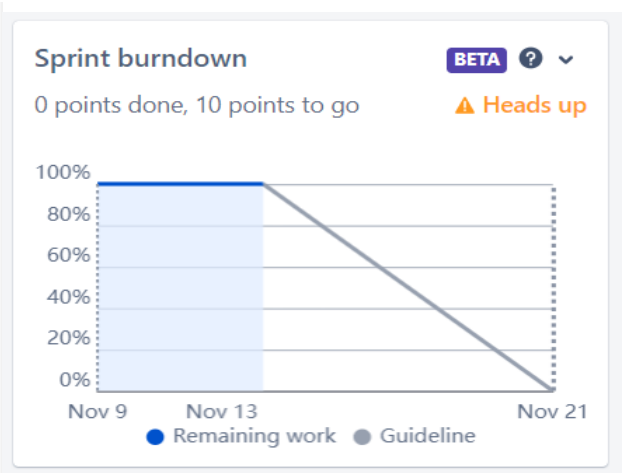
AV = Sprint Duration / Velocity = 12 / 10 = 1.2

The above average velocity is for Sprint-3, Sprint-4.

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time. Burnt down charts of sprint3,4

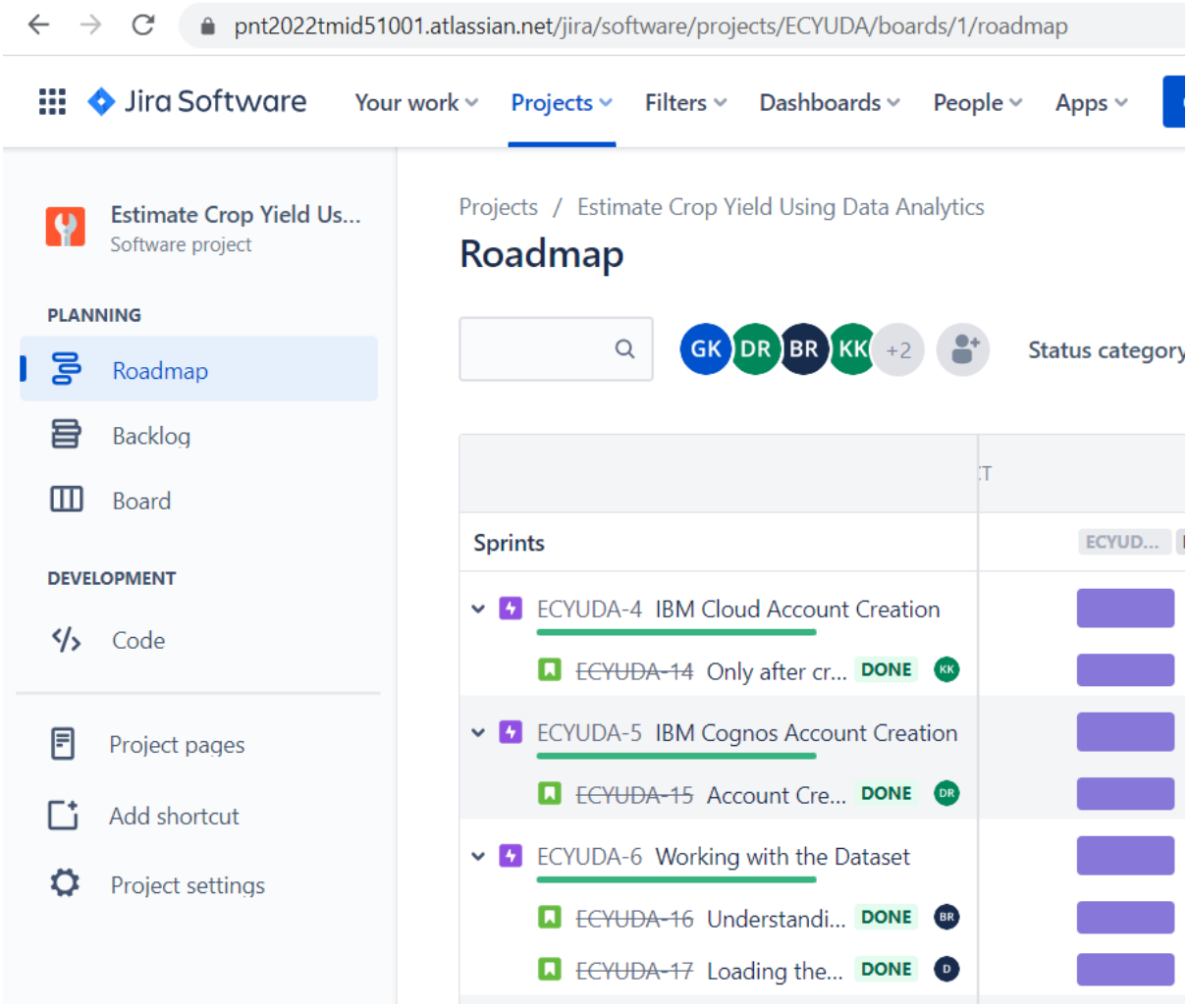




6.3 Reports from Jira

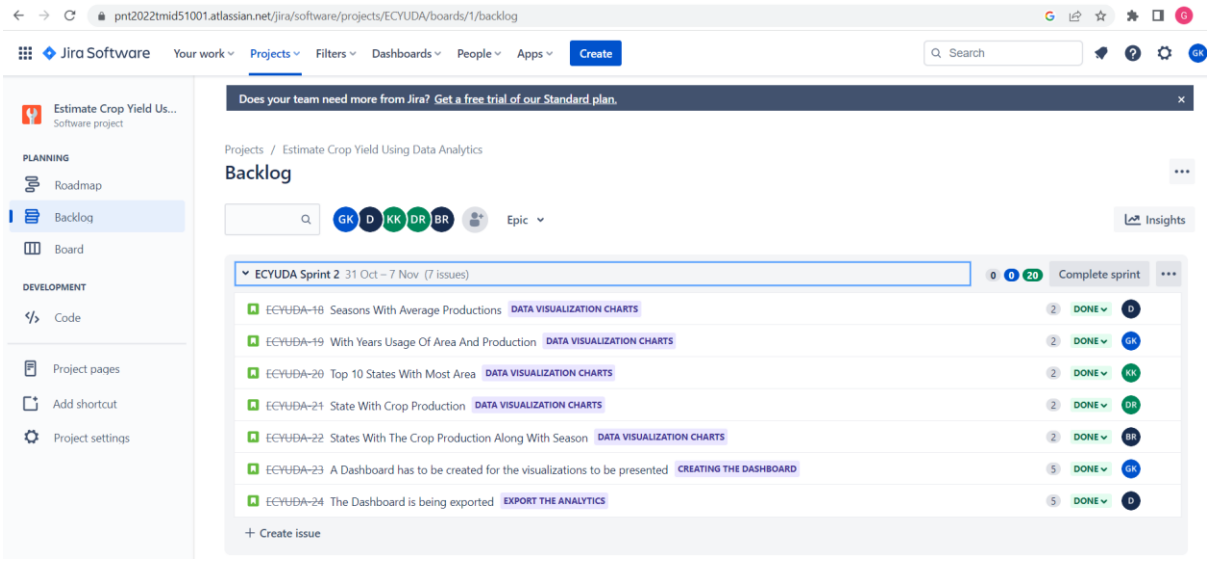
Sprint-1:

In Sprint-1, the epics listed are IBM Cloud Account Creation, IBM Cognos Account Creation and Working with the dataset. In Working with the dataset, the stories listed within are Understanding the dataset and Loading the dataset. Since all of the stories and epics are completed, they are moved from To Do to Completed. And thereby click on complete sprint thereby completing the sprint and moving on to the next sprint.

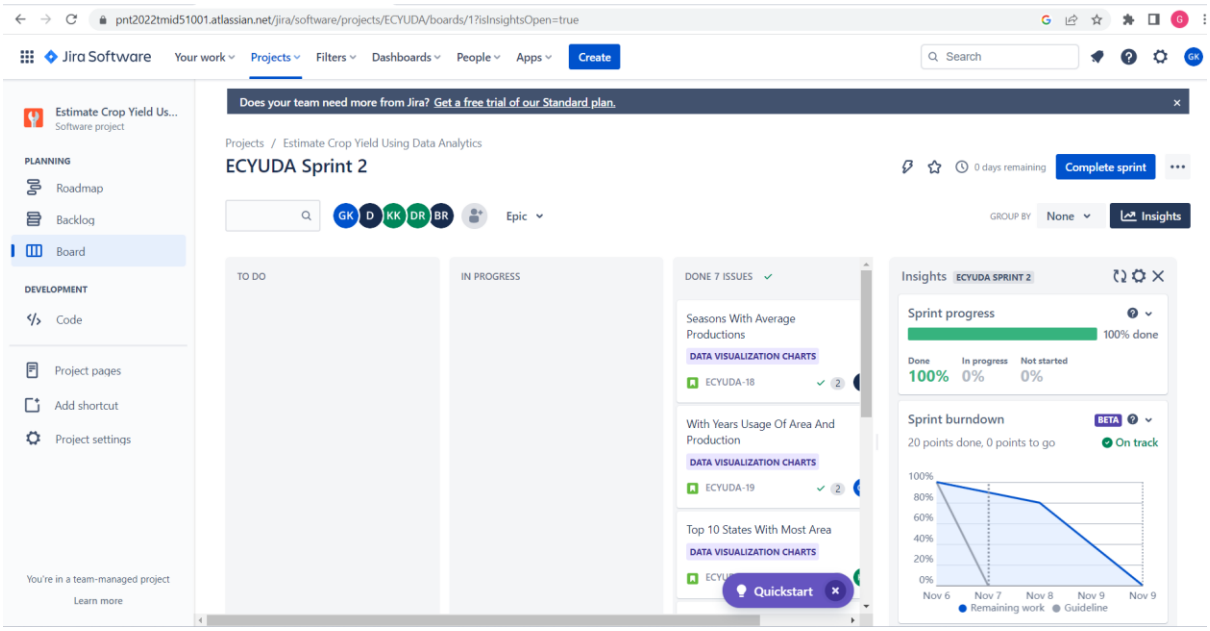


Sprint -2:

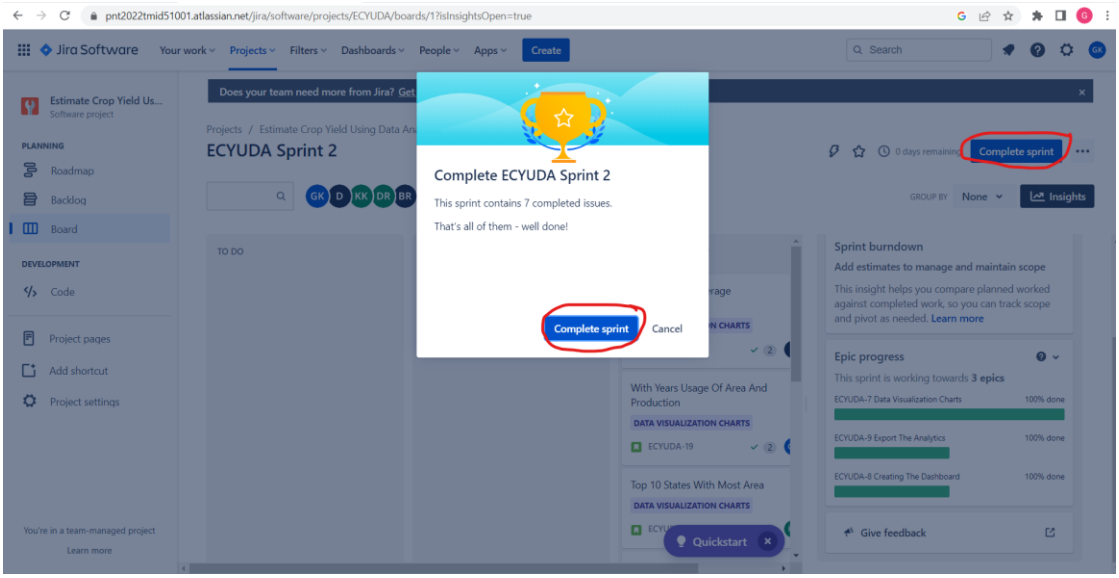
In Sprint-2, the epics were Data Visualization Charts, Creating the Dashboard and exporting the analytics. And in Data Visualization Charts there are multiple visualizations like Seasons with Average Production, With Years Usage of Area and Production, Top 10 States with most area, State with crop production, States with the crop production along with season as the stories inside the epic Data Visualization Charts.



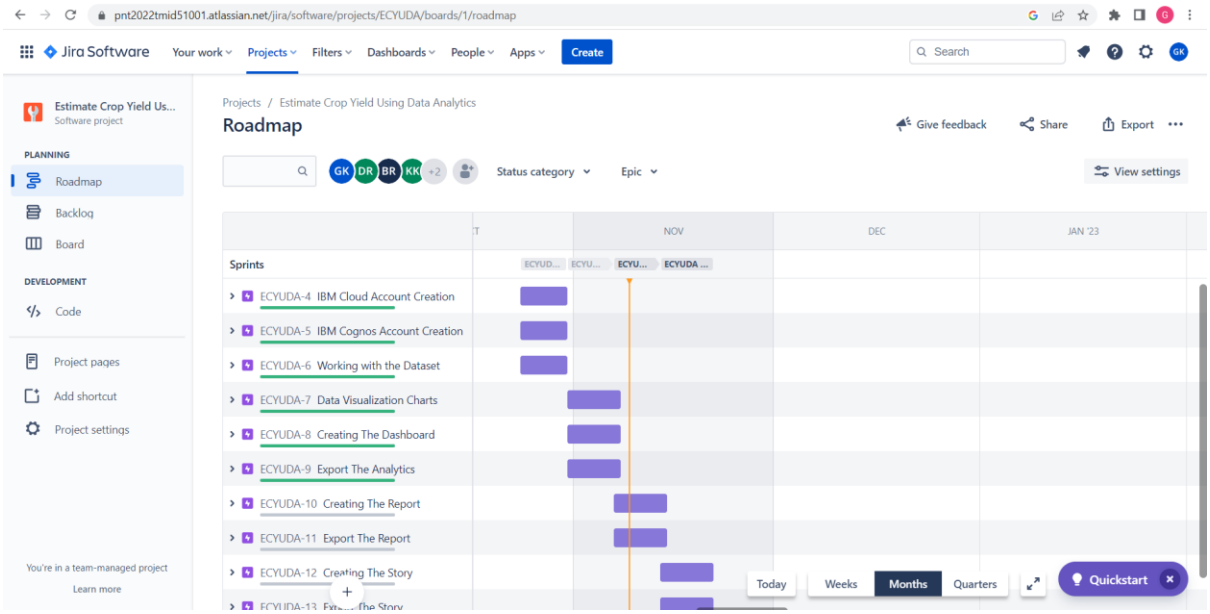
And in the Kanban board we are to do the move all the tiles from To Do to Done. Since all the epics are completed, the sprint progress is shown and the sprint burndown is also shown in it.



By clicking on the complete Sprint in the Screen the following popup appears and then click “Complete Sprint” in it.

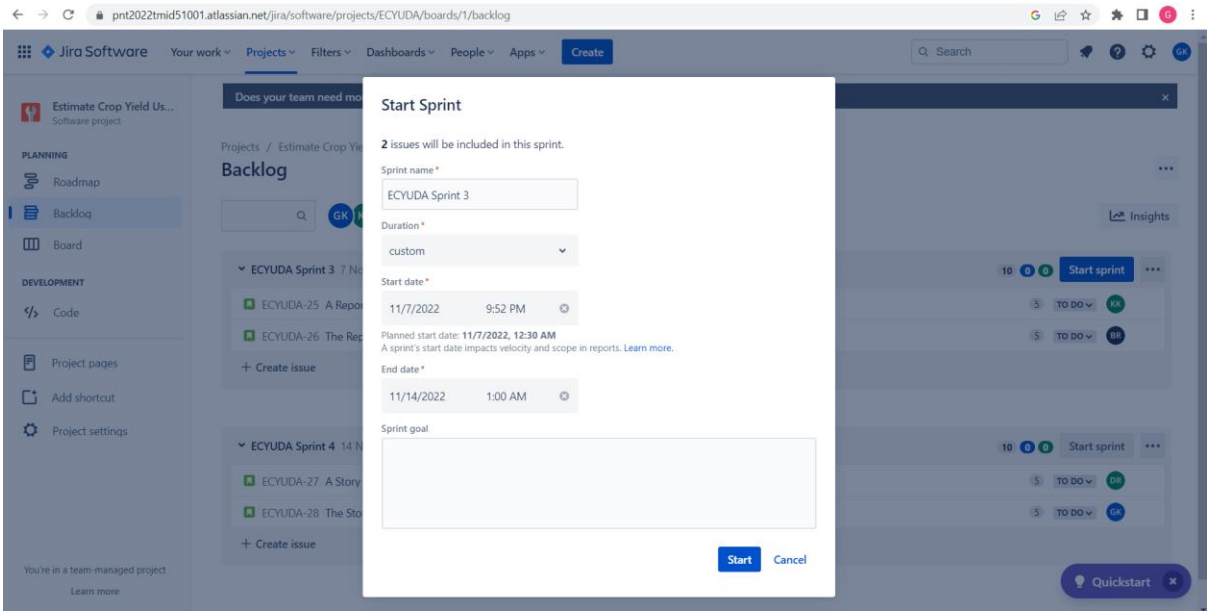


And then on choosing the roadmap, we'll be knowing the number of epics completed.

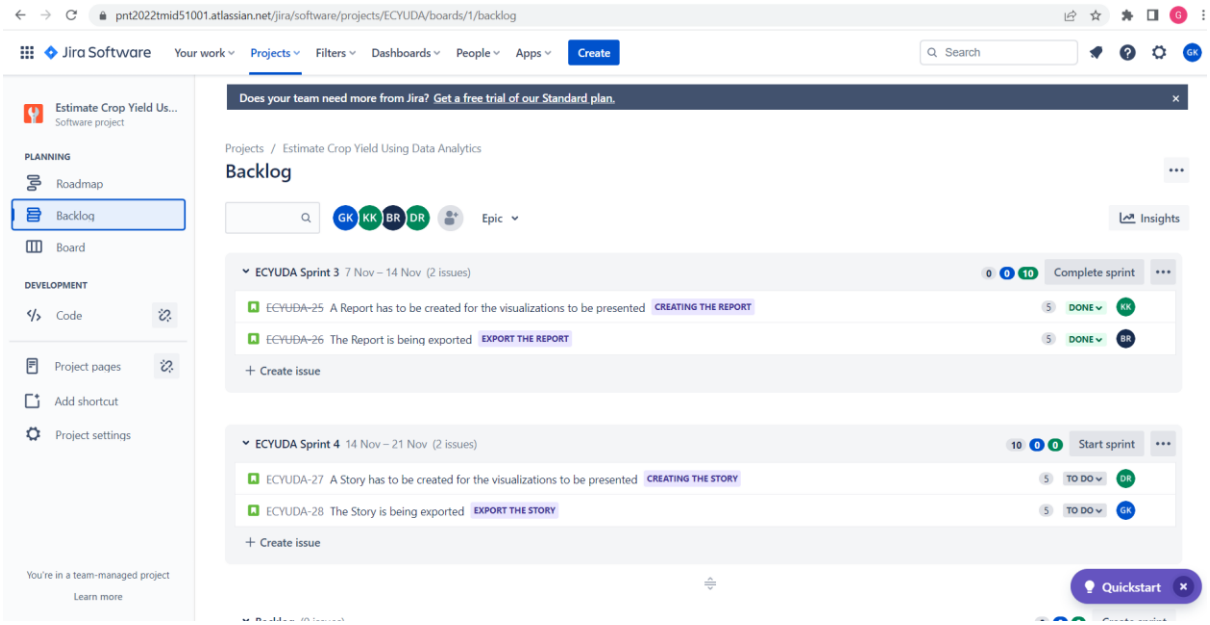


Sprint-3:

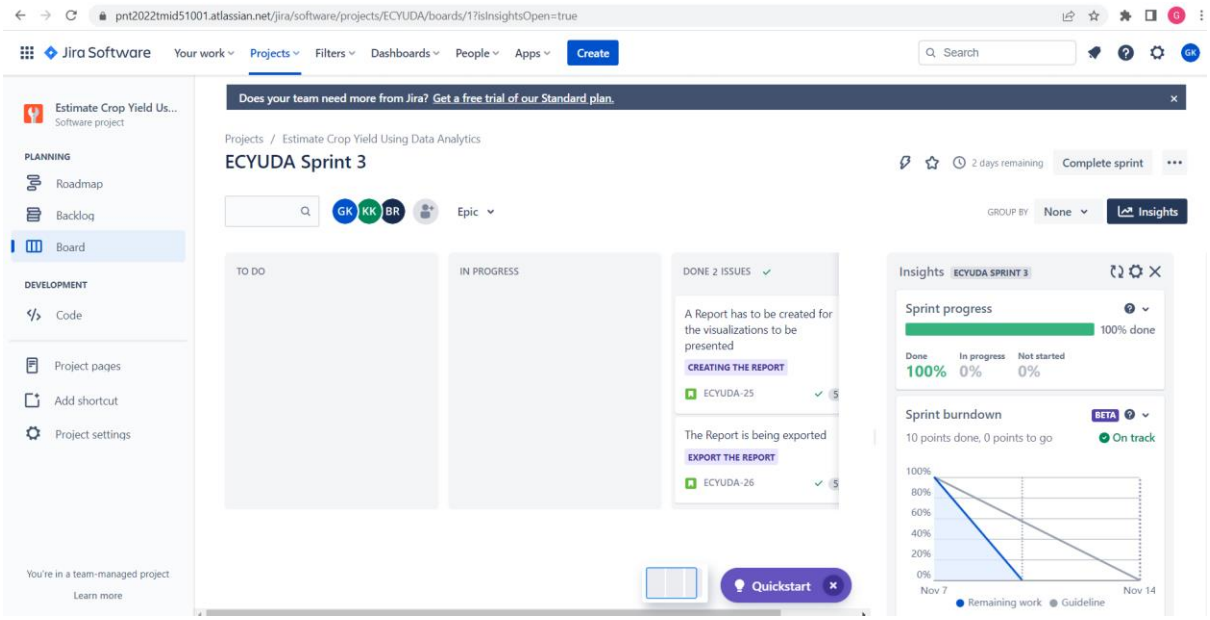
The Sprint-3 is named as ECYUDA Sprint-3 where ECYUDA is the short form of the project name provided. ECYUDA stands for Estimate the Crop Yield Using Data Analytics. It would contain the information of the start date and the supposedly end date when choose a duration of one week.



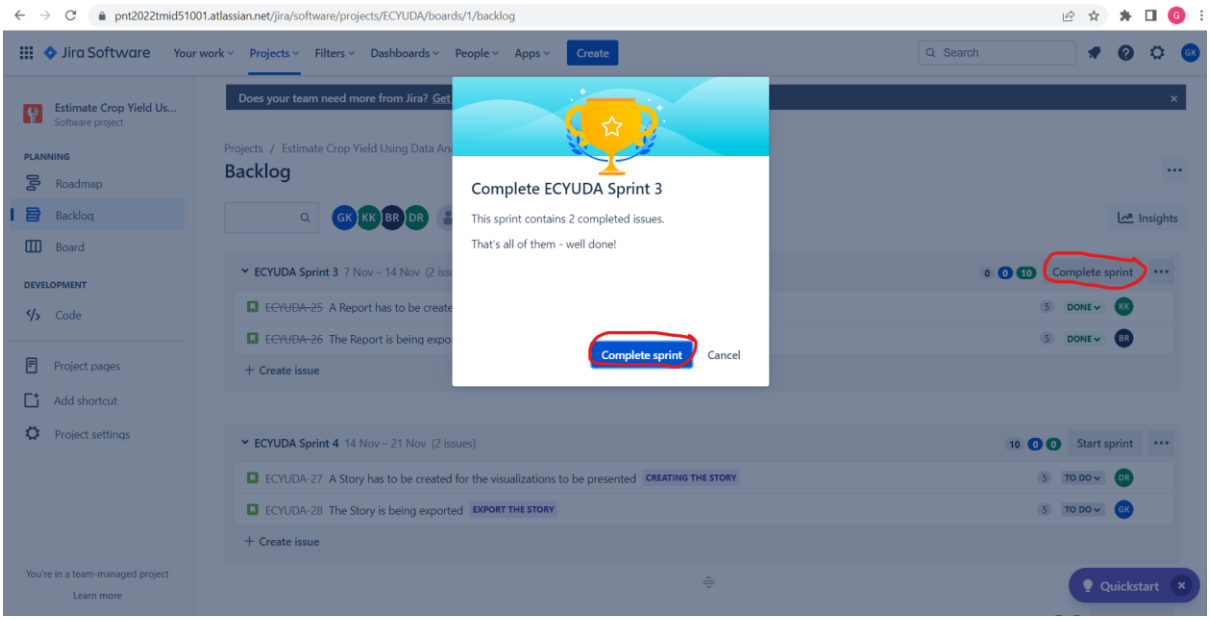
On Clicking on Start, the sprint-3 would start. And it consists of two epics namely, Creating the report and Export the Report.



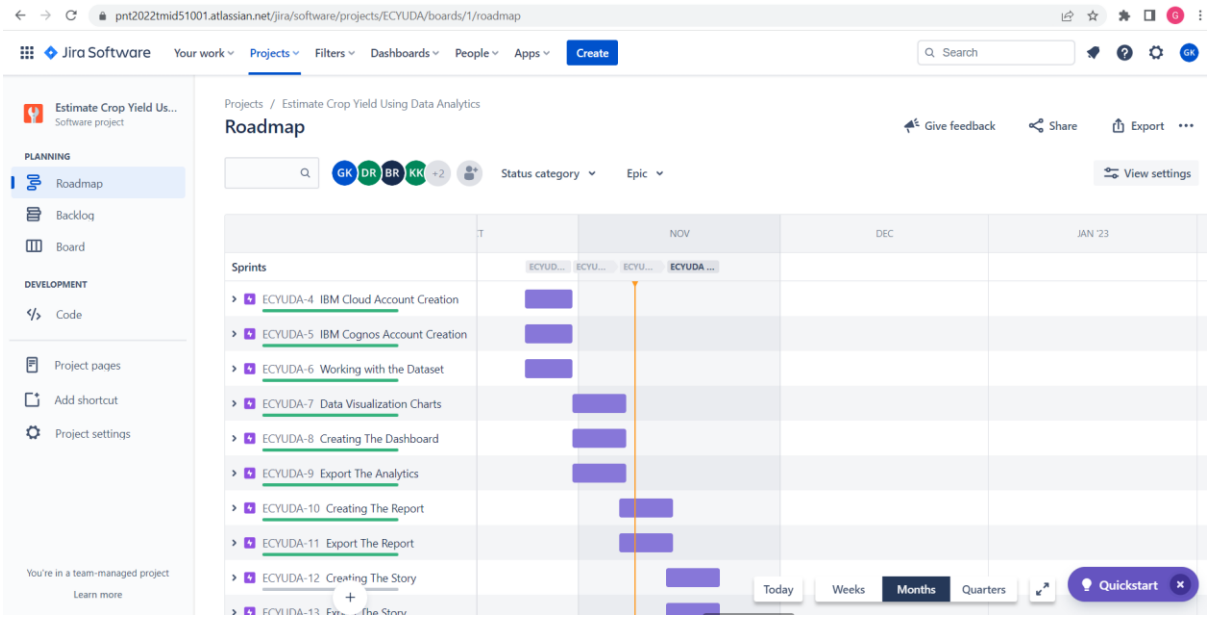
On completing the epics, it is being moved from To Do to Done in the Kanban Board and we also get to see the burndown chart for week-3.



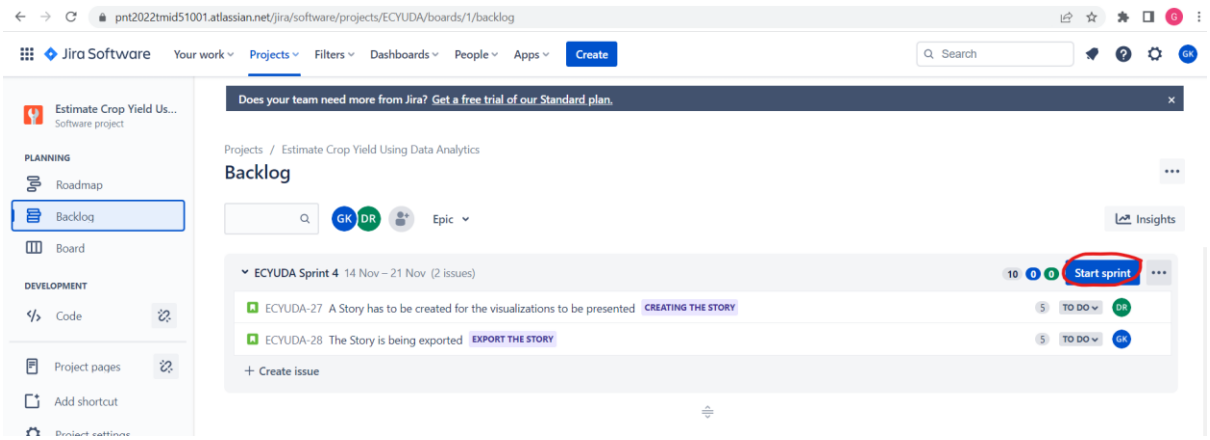
And then after placing the epics in done in Kanban Board, we can click on Complete Sprint .



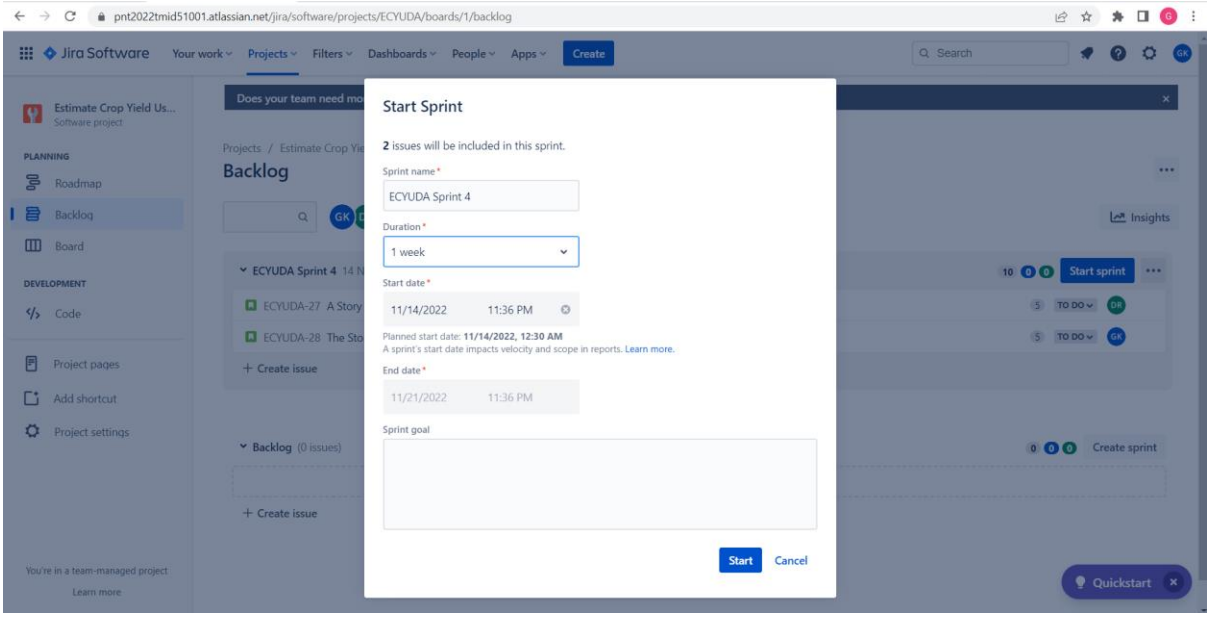
And the roadmap till sprint-3 is like the below



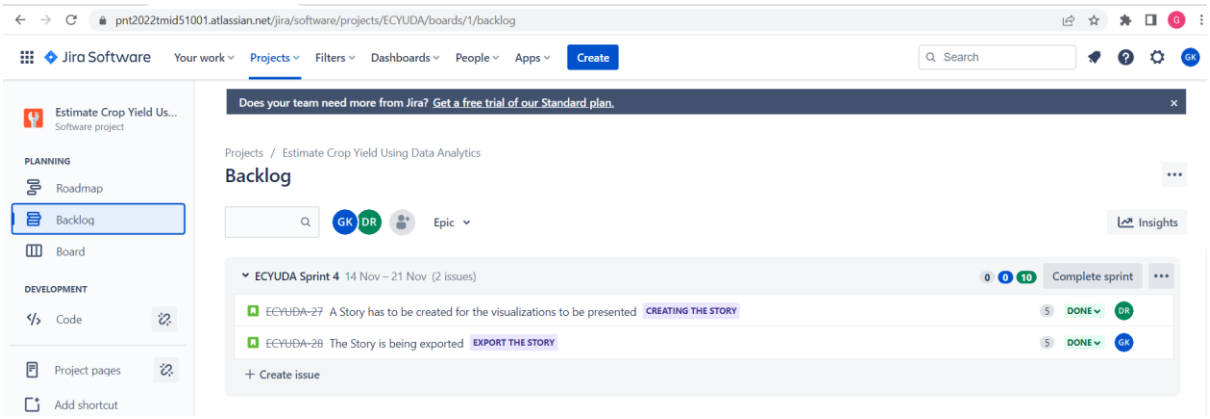
And then after completing sprint-3, we have to start sprint-4 by clicking on “Start Sprint”



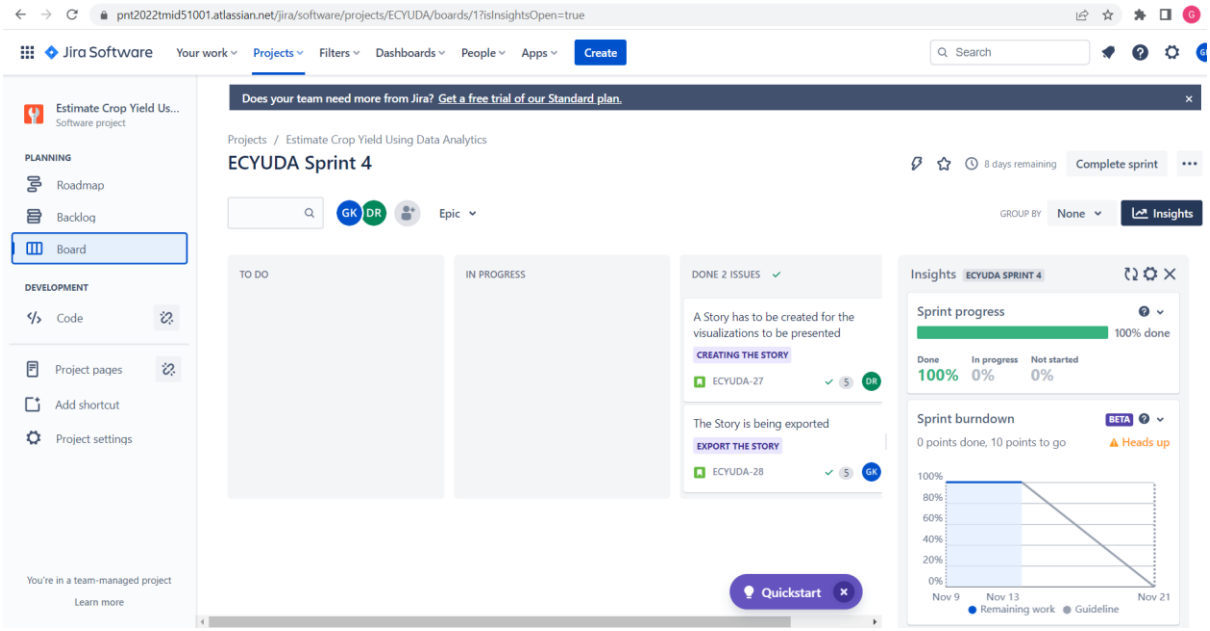
On clicking start sprint, we get



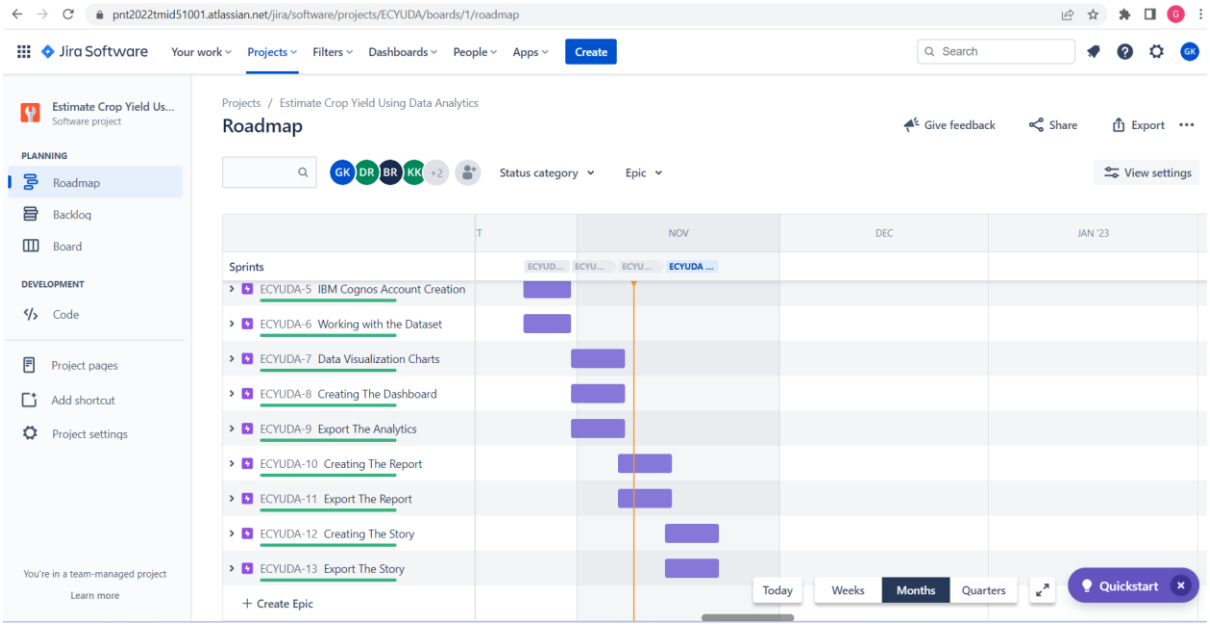
It has the sprint name, duration, start date and end date and followed by this we have to click on Start.



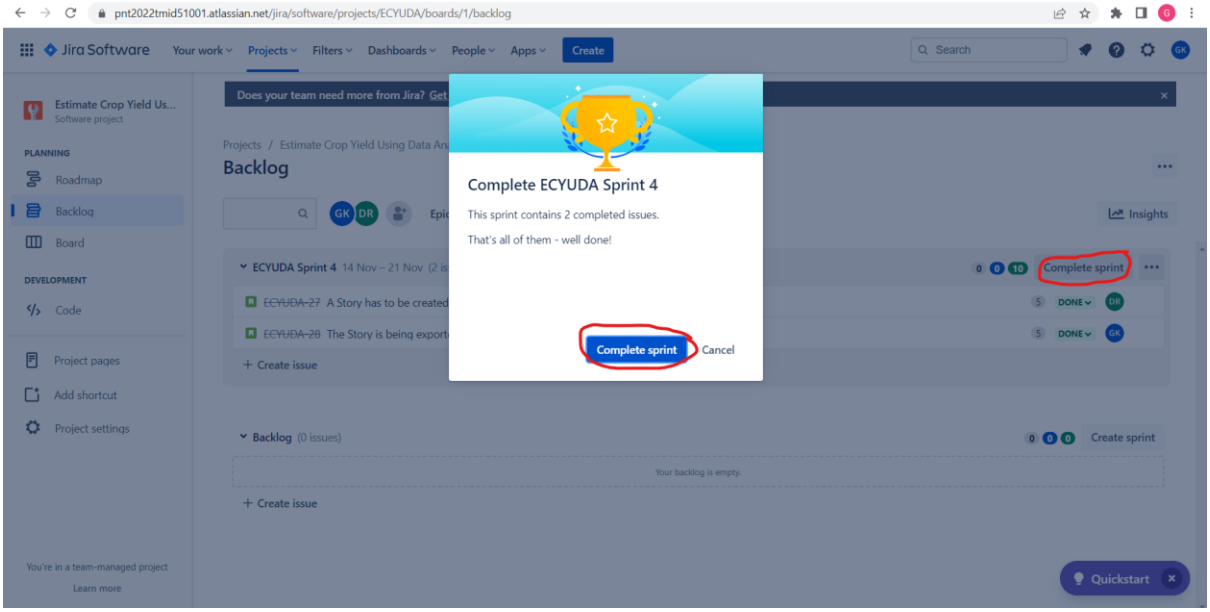
Since we have completed the two epics of sprint-4 namely, Creating the Story and Exporting the Story. In the Kanban Board, we move the epic from To Do to Done and along with it we get the burndown chart.



The roadmap for the entire project becomes



Since we have completed sprint-4, we are prompted to click the “Complete Sprint”.



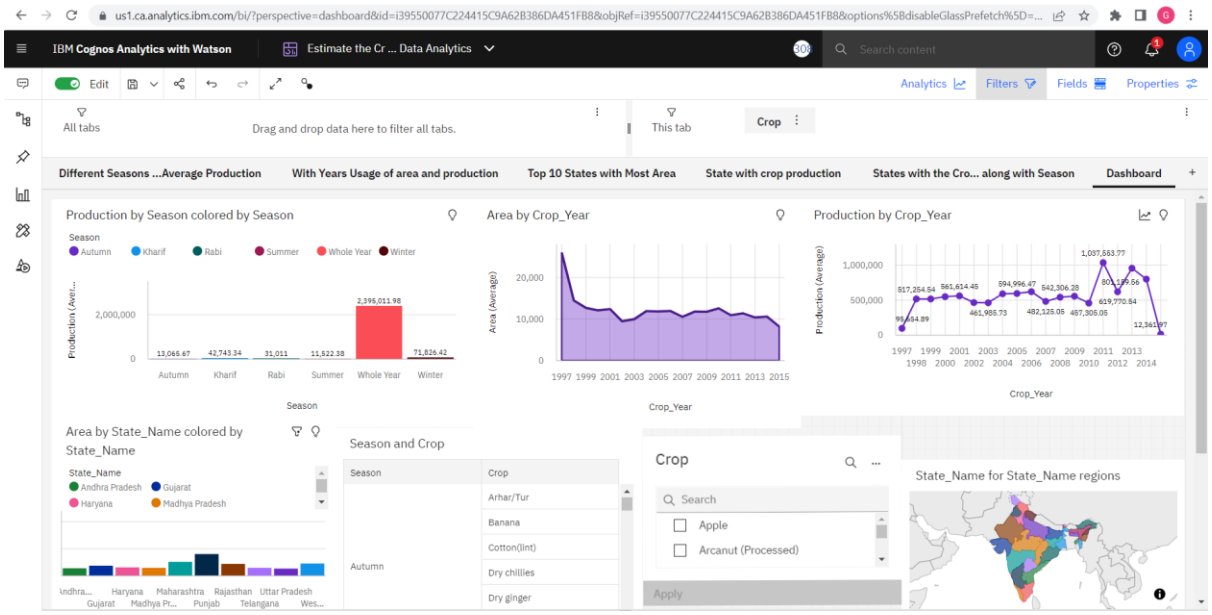
The below is the roadmap in terms of the entire project in terms of weeks.

7. Coding and Solutioning

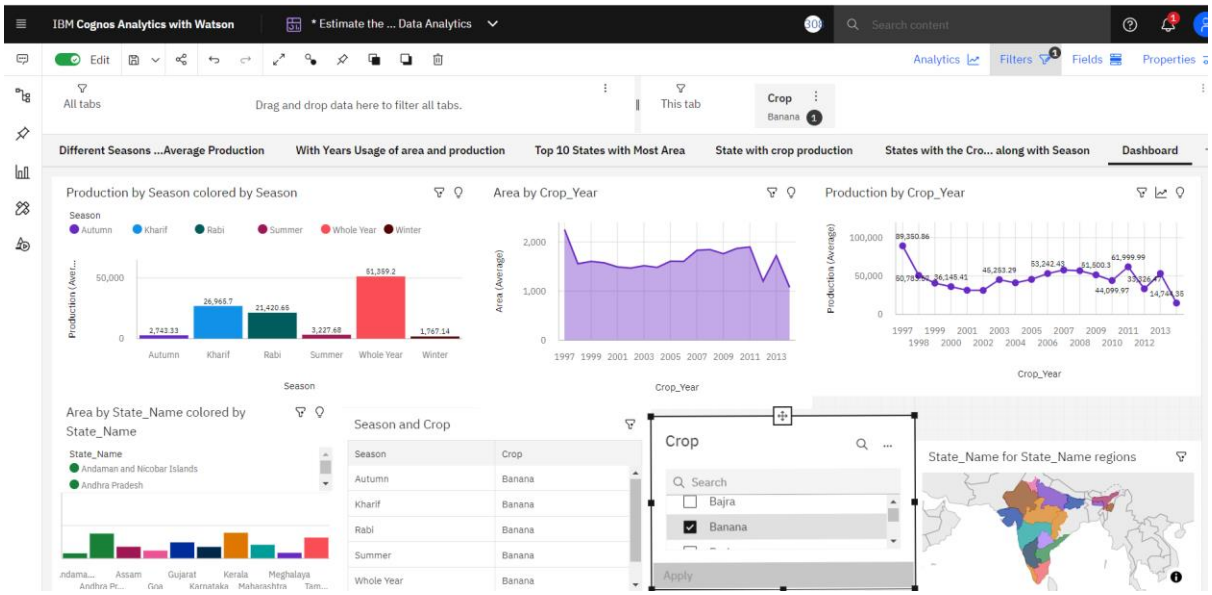
7.1 Feature 1 - Dashboard

The dashboard comprises visualizations for Seasons With Average Productions, With Years Usage Of Area And Production, Top 10 States With Most Area, State With Crop Production, States With The Crop Production Along With Season which are been detailed explained in the data visualization charts.

The below image is without the application of canvas,

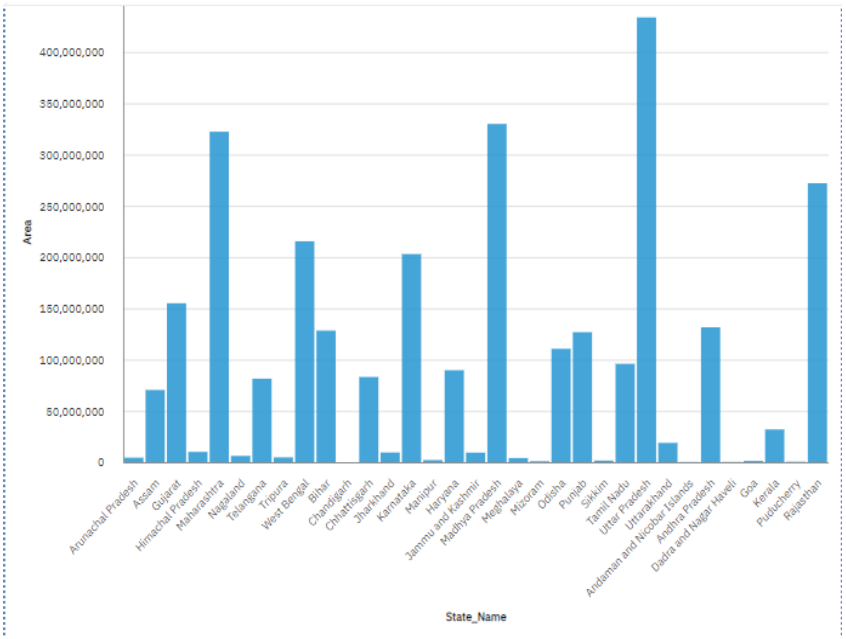


The below image is with the application of canvas and choosing banana as the crop, the visualizations become like

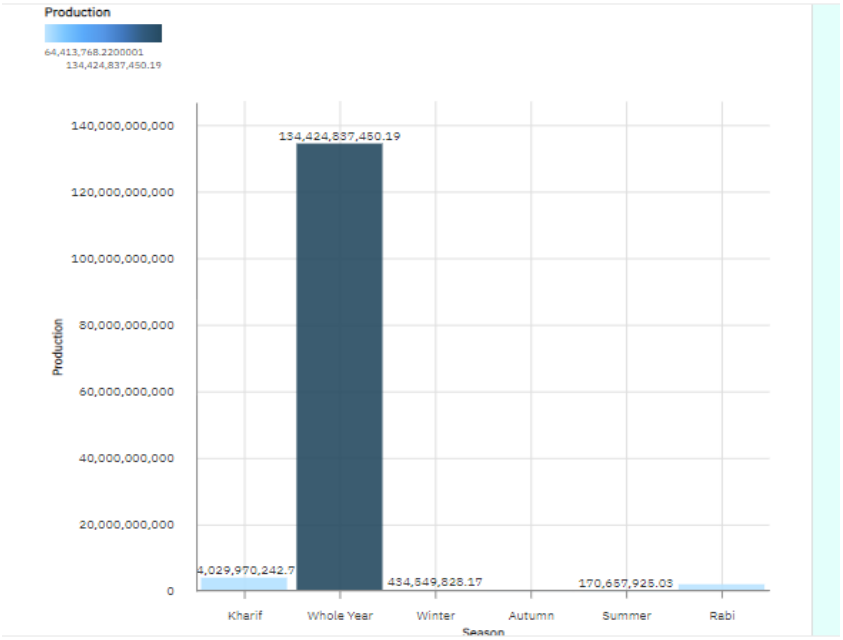


7.2 Feature 2 - Report

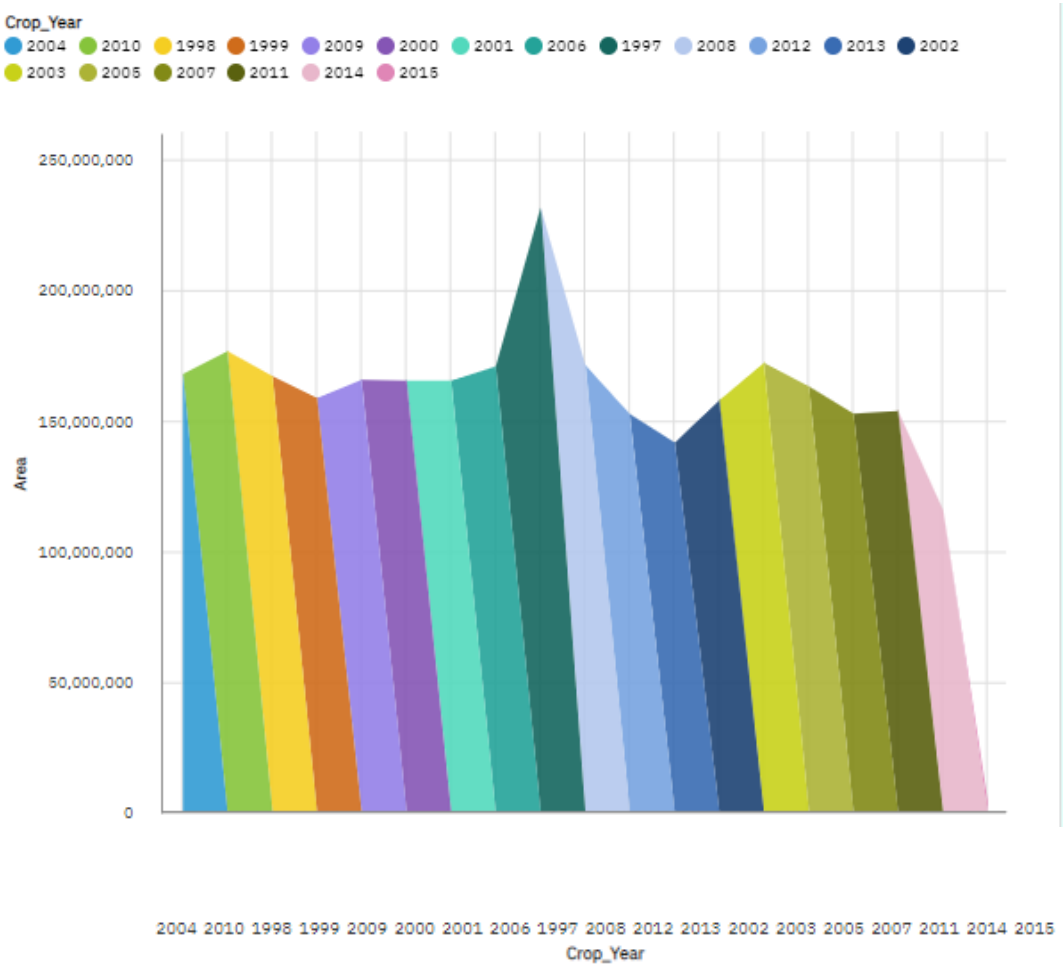
States with area of production :



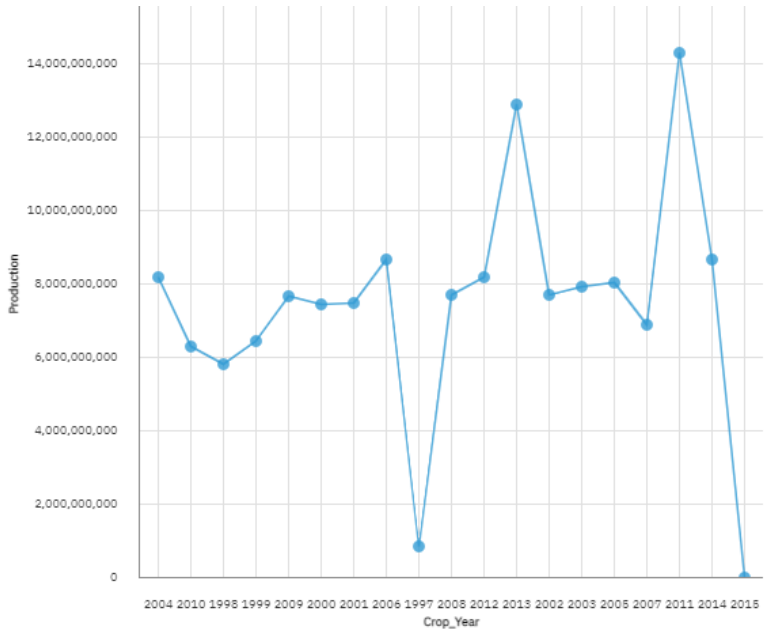
Seasons with productions :



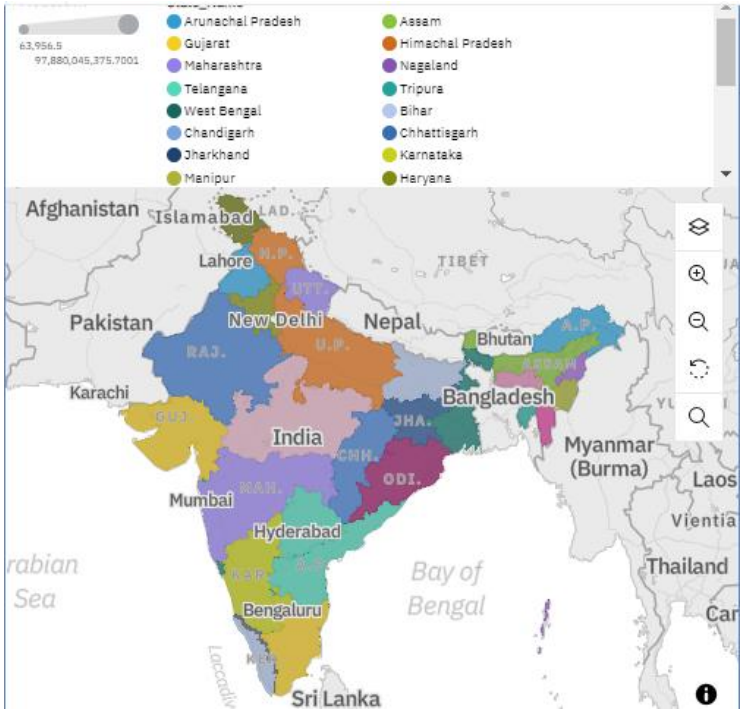
With years usage of area :



With years usage of area and production :



States with crop production :



State with crop production along with seasons :

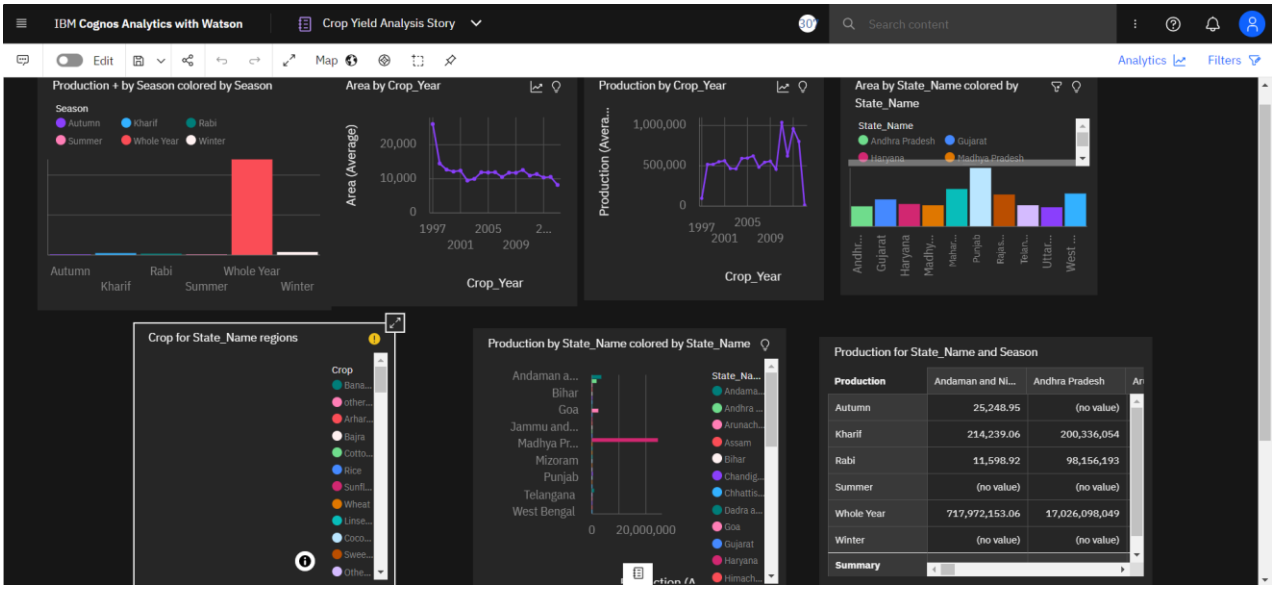
Production	Rabi	Autumn	Summer	Whole Year	Kharif	Winter
Andhra Pradesh	98,156,193			17,026,098,049	200,336,054	
Assam	8,404,649	8,230,196	17,171,024	1,996,667,293	15,782,850	65,495,747
Bihar	103,402,165	23,622,936	15,679,752	124,903,630	31,029,656.66	67,845,457
Chhattisgarh	9,228,963.01		1,494	1,946,226	89,775,225	
Gujarat	51,737,400		12,268,700	283,969,858	176,315,379	
Haryana	173,272,098			119,408,311	88,593,481	
Himachal Pradesh	6,723,873			947,088.1	10,134,207.5	
Jammu and Kashmir	4,124,447.7			96,323.2	9,070,244.8	
Jharkhand	1,609,846.66	1,923,687.84		91,369.88	32,220.12	7,120,617.25
Karnataka	43,860,390.2		19,552,520	624,988,842.21	175,028,059.27	
Kerala		3,442,449.33	2,093,790.53	97,869,331,257.1701	120,672.95	5,057,205.72
Madhya Pradesh	191,648,533.66			90,948,638.07	166,243,566.93	
Maharashtra	70,912,135.5	19,695	2,957,812	653,287,157	536,463,806.7	
Manipur	18,910	244,230	28,600	1,111,195	3,619,792	208,190
Meghalaya	725,401	525,139	358,352	4,378,228	4,723,682	1,401,694
Mizoram	67,405.53			406,907	1,187,227.3	

Report link:

https://us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2Fdata%2Bmodules%2Freport&action=run&format=HTML&prompt=false

7.3 Feature 3 - Story

All visualizations are showcased on the same screen like the following. Since it is a story it is played in a video format. The visualizations are represented as follows

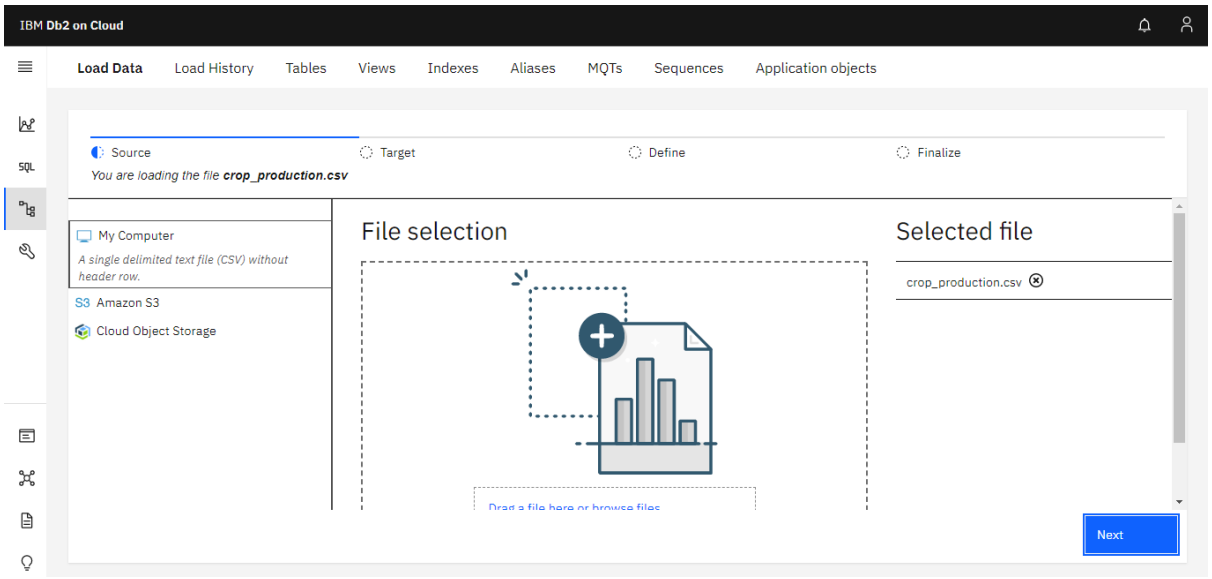


And the story is in the form of a video which is present in the Sprint-4 folder itself.

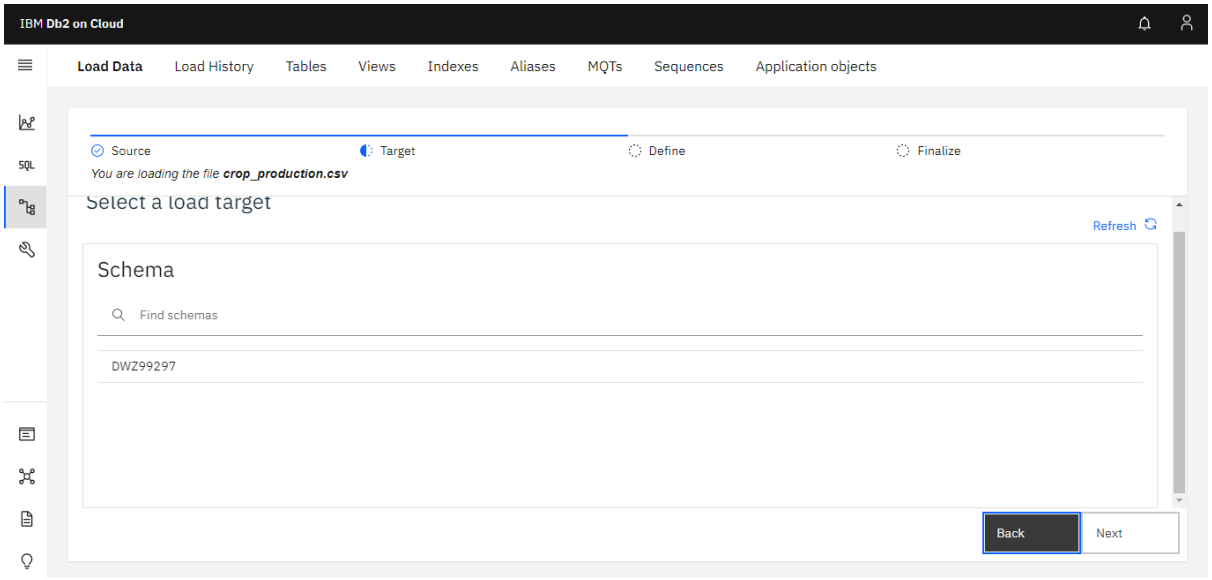
Story Link:
https://us1.ca.analytics.ibm.com/bi/?perspective=story&pathRef=.my_folders%2FMy%2BAnalytics%2BStory%2FCrop%2BYield%2BAnalysis%2BStory&action=view&sceneId=model000001845cf13b42_00000000&sceneTime=10150

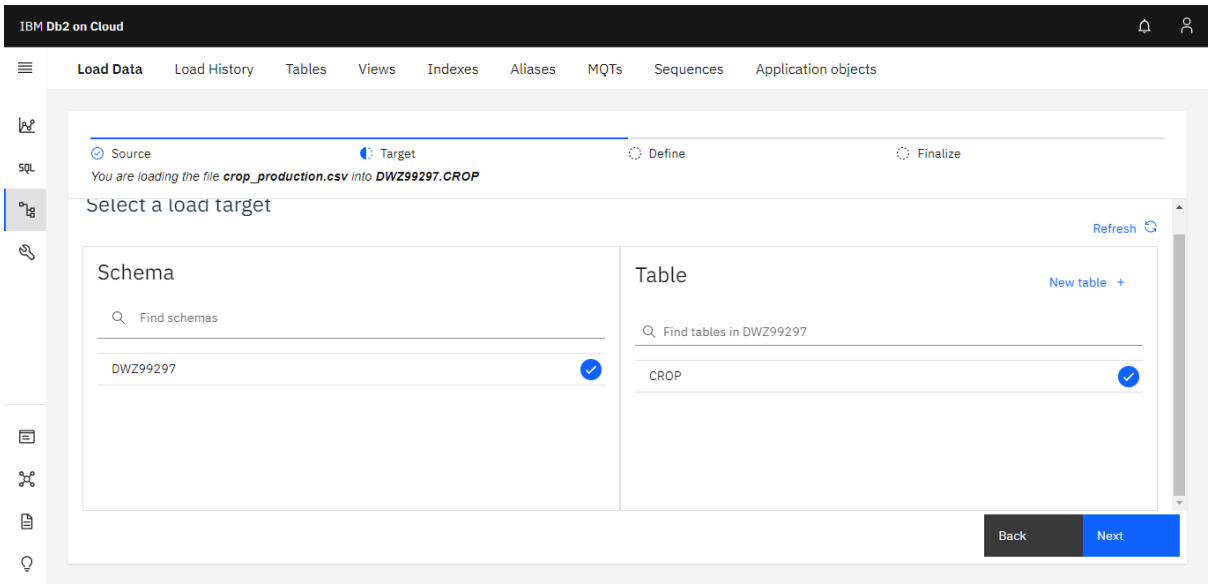
7.4 Database Schema

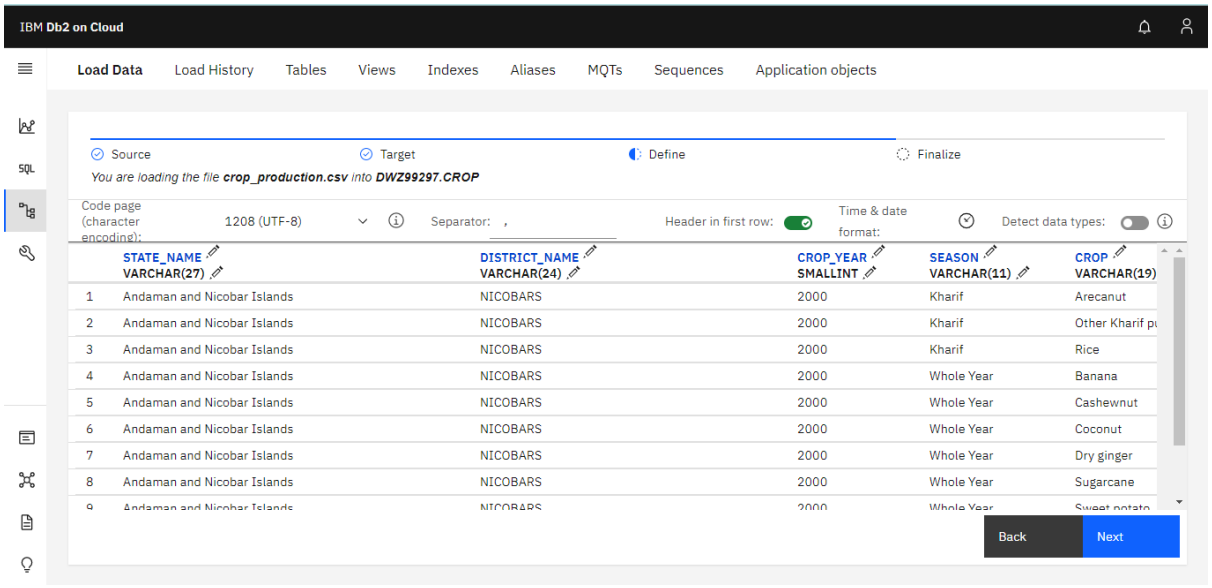
There are given options to load the dataset from my computer as csv file , Amazon S3 as bucket here we are loading the dataset from my computer option and then click on next button.



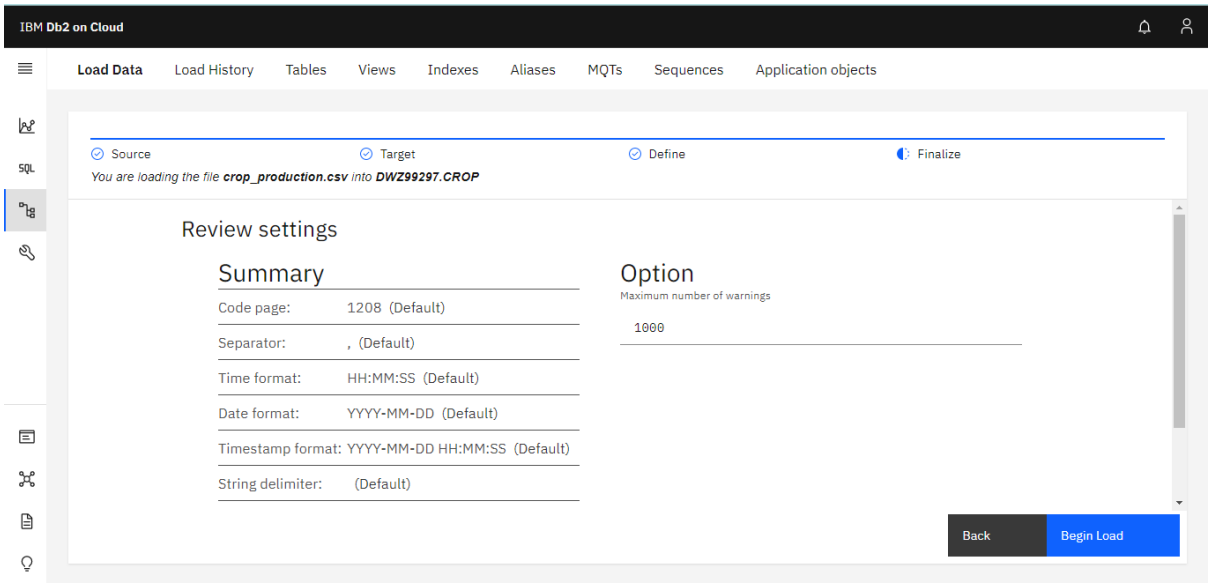
Then we are to select the target load by clicking on the schema, then we get options to create table. And in the define the dataset is defined.


IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
Source
Target
Define
Finalize
You are loading the file crop_production.csv
Select a load target
Refresh
Schema
Find schemas
DWZ99297
Back
Next

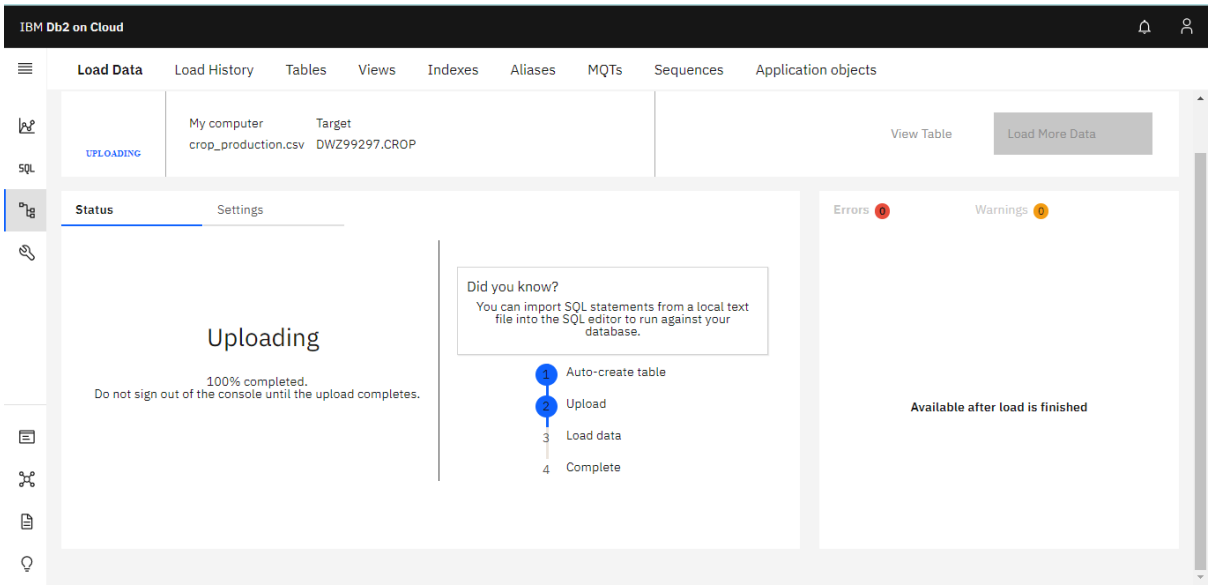

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
Source
Target
Define
Finalize
You are loading the file crop_production.csv into DWZ99297.CROP
Select a load target
Refresh
Schema
Find schemas
DWZ99297
Table
New table +
Find tables in DWZ99297
CROP
Back
Next



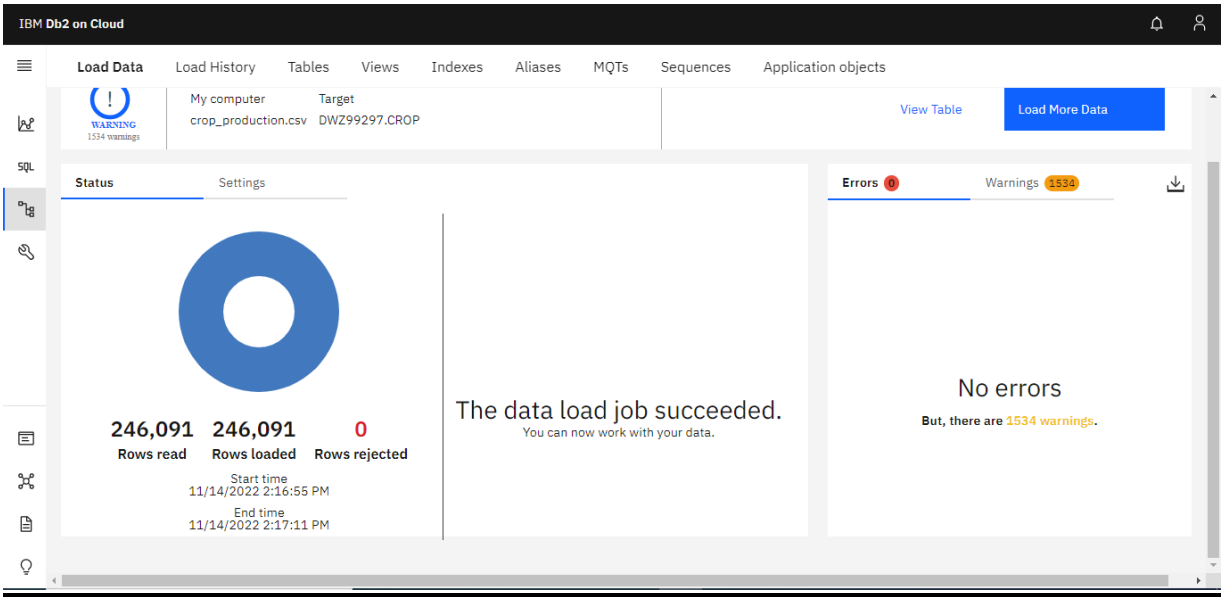
And in the finalize section, we are to change the maximum number of warnings to 10000 to prevent the warnings have exceeded error and click on begin load.



Then the processes of Auto-create table, Upload, Load Data and then it is completed.



Now the loading job is successful and the rows are read and loaded successfully.



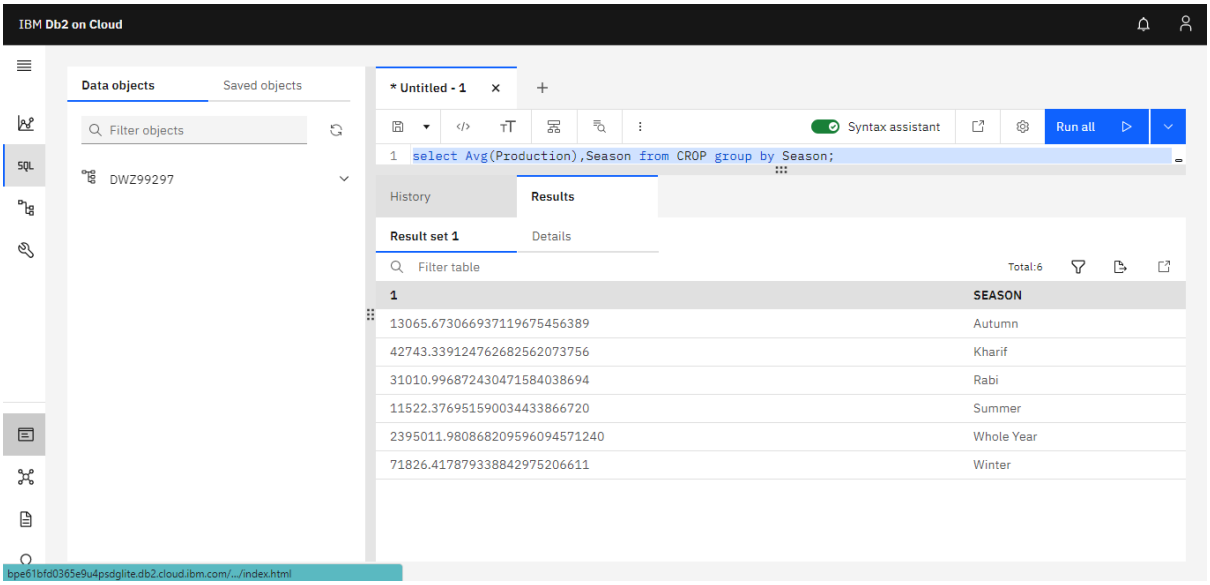
8. Testing

8.1 Test cases

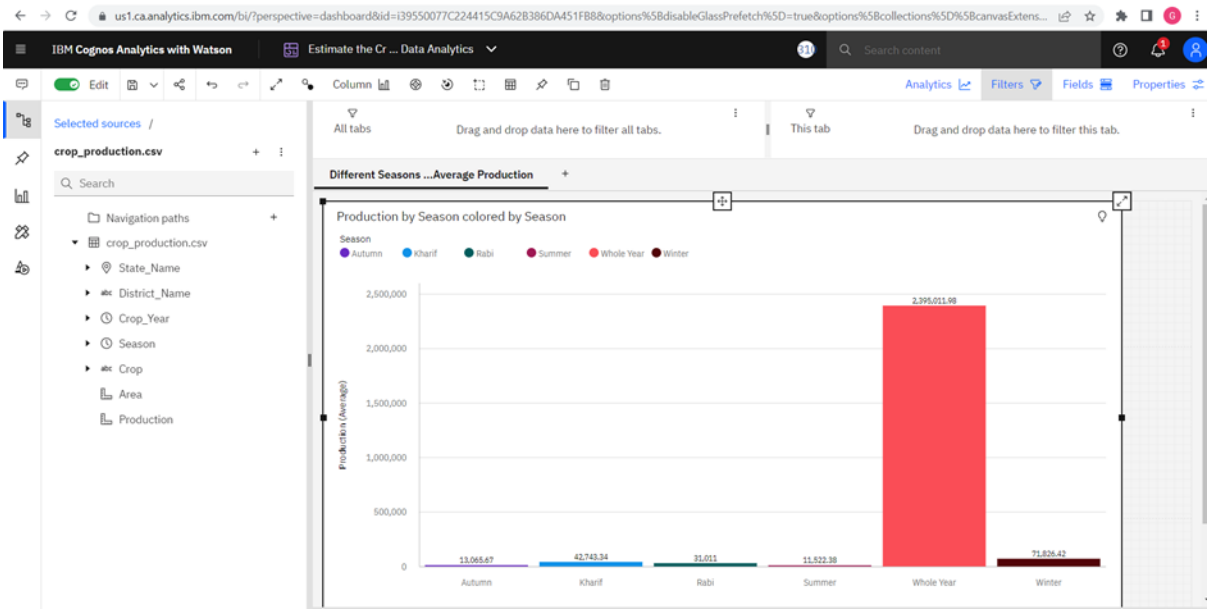
- Print Engine
- Client Application
- Security
- Outsource Shipping
- Exception Reporting
- Final Report Output
- Version Control

8.2 User Acceptance Testing

User acceptance testing (UAT), also known as application testing or end-user testing, is a stage of the software development process when the target user group tests the product in the real world. Here we verify the result achieved is correct or wrong.
We are taking a query, we are taking seasons with average production.



Now we cross verify the results with cognos analytics by presenting the data in chart

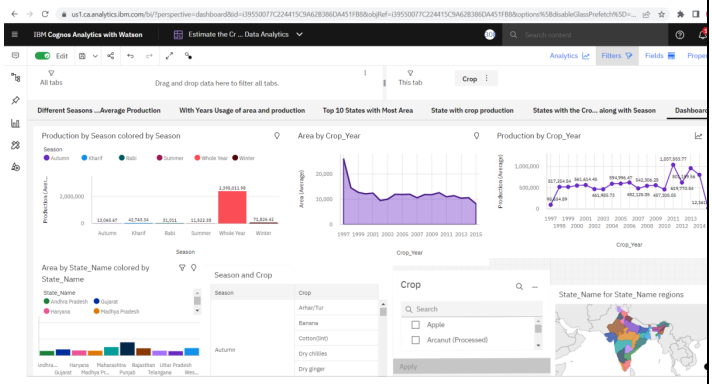
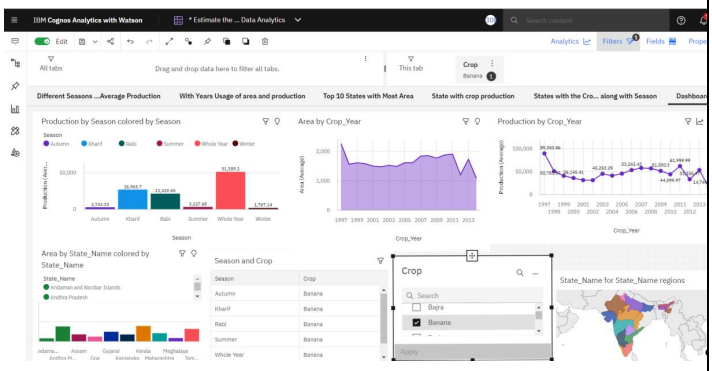


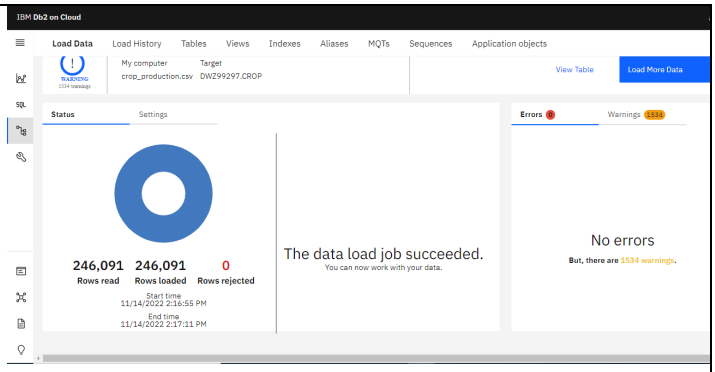
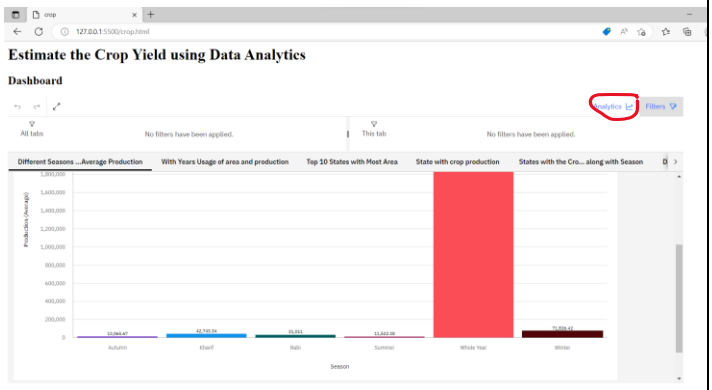
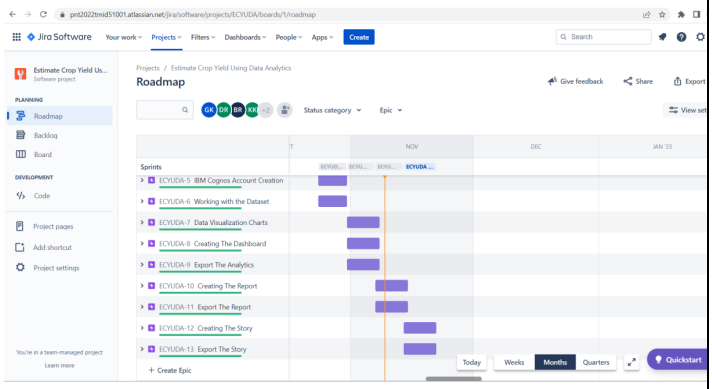
And the values turn out to be the same and so it is concluded that the output of visualisations and query's must be equal

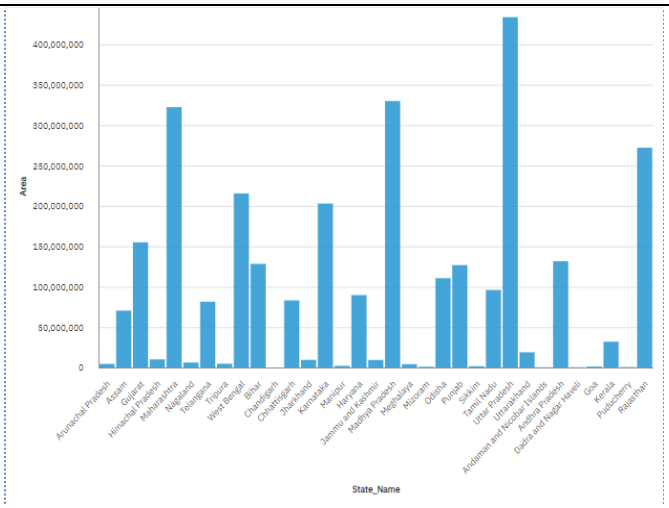
9.RESULTS

9.1 Performance Metrics:

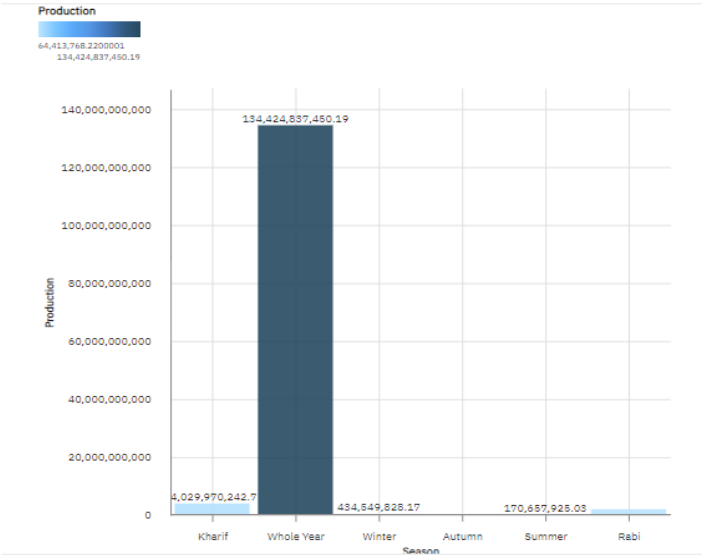
There are several parameters like Dashboard Design, Data Responsiveness, Amount Data to Rendered (DB2 Metrics), Utilization of Data Filters, Effective User Story, Descriptive Reports which contributes to the performance metrics.

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs – 5 
2.	Data Responsiveness	A canvas is provided with the dashboard when a particular crop is selected, the visualizations change accordingly. In this case, we have chose banana and the visualizations change accordingly as follows 

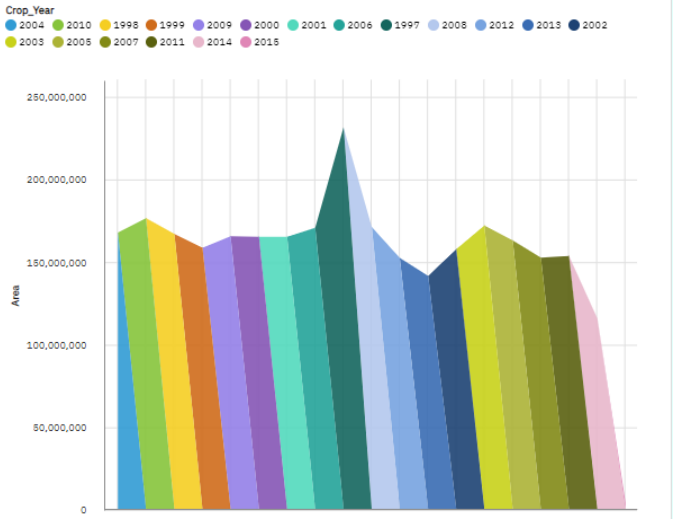
3.	Amount Data to Rendered (DB2 Metrics)	
4.	Utilization of Data Filters	Filters are provided with every visualization and it can be changed accordingly. 
5.	Effective User Story	No of Scene Added – 15 user story scenes were added and are completed successfully. 
6.	Descriptive Reports	No of Visualizations / Graphs – 6 States with area of production:



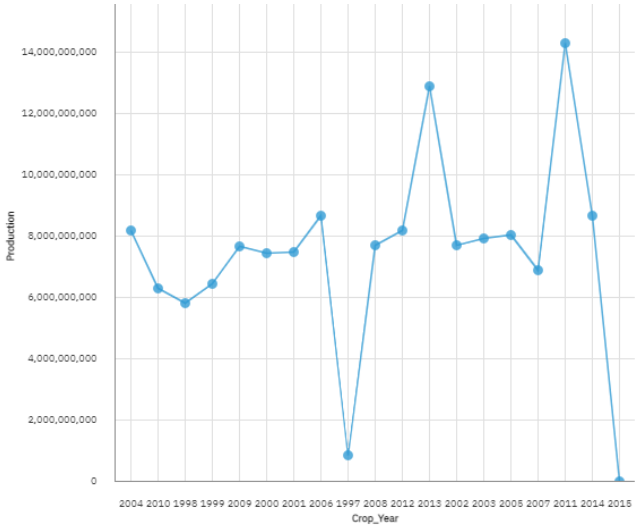
Seasons with productions:



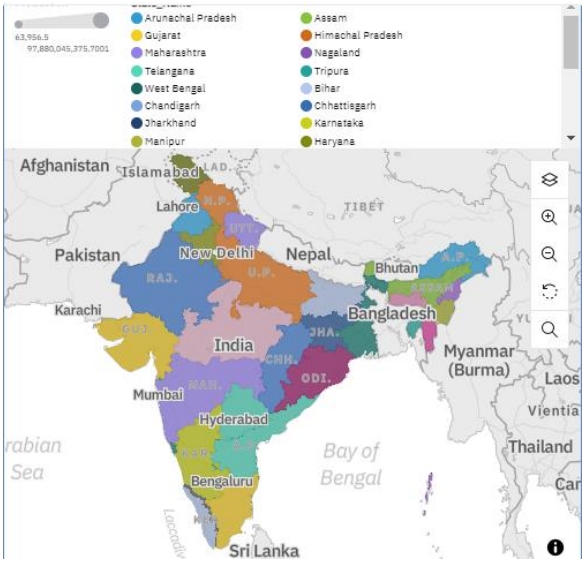
With years usage of area:



With years usage of production:



States with crop production:



		State with crop production along with seasons:

Production	Rabi	Autumn	Summer	Whole Year	Kharif	Winter
Andhra Pradesh	98,156,193			17,026,098,049	200,336,054	
Assam	8,404,649	8,230,196	17,171,024	1,996,667,293	15,782,850	65,495,
Bihar	103,402,165	23,622,936	15,679,752	124,903,630	31,029,656.66	67,845,
Chhattisgarh	9,228,963.01		1,494	1,946,226	89,775,225	
Gujarat	51,737,400		12,268,700	283,969,858	176,315,379	
Haryana	173,272,098			119,408,311	88,593,481	
Himachal Pradesh	6,723,873			947,088.1	10,134,207.5	
Jammu and Kashmir	4,124,447.7			96,323.2	9,070,244.8	
Jharkhand	1,609,846.66	1,923,687.84		91,369.88	32,220.12	7,120,617
Karnataka	43,860,390.2		19,552,520	624,988,842.21	175,028,059.27	
Kerala		3,442,449.33	2,093,790.53	97,869,331,257.1701	120,672.95	5,057,205
Madhya Pradesh	191,648,533.66			90,948,638.07	166,243,566.93	
Maharashtra	70,912,135.5	19,695	2,957,812	653,287,157	536,463,806.7	
Manipur	18,910	244,230	28,600	1,111,195	3,619,792	208,
Meghalaya	725,401	525,139	358,352	4,378,228	4,723,682	1,401,
Mizoram	67,405.53			406,907	1,187,227.3	

10. Advantages

- Able to use filters efficiently.
- Able to share the dashboard, report and story to anybody.
- Less time-consuming process.
- Less expensive.
- Accurate measurements can be made.
- Able to predict easily using visualizations.

Disadvantages

- Internet is required.
- If data inadequate, provide wrong results.
- If data incorrect would result in improper results generation.

11. Conclusion

Estimation of crop yield is a very important aspect. Using data analytics and with the provided data we are performing the study in IBM Cognos. With the help of IBM Cognos, we are able to upload the dataset, preprocess it, and start creating visualizations with it. And with the visualizations, we generate dashboard, report and story. And it being embedded in a website and the output is being cross verified with checking with IBM DB2 database. And thus, various other factors can be studied and multiple factors can be analyzed. For any of the given filters the subsequent output is generated.

12. Future Scope

Addition of multiple factors to improve our understanding of the factors which are provided for data yield. Addition of an attractive user interface and try to automate some basic

tasks. Include a chat bot that may help us when we are stuck. And also adding features that roadmap and introduce all the options.

14. Appendix

Source Code :

```
<html>

  <head>
    <title>crop</title>
  </head>
  <body>
    <h1>Estimate the Crop Yield using Data Analytics</h1>
    <h2>Dashboard</h2>
    <iframe
src="https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&amp;pathRef=.my_folders%2FFinal%2BDeliverables%2FEstimate%2Bthe%2BCrop%2BYield%2Busing%2BData%2BAnalytics&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;action=view&amp;mode=dashboard&amp;subView=model000001844e26c7a9_00000002" width="1400" height="600" frameborder="0" gesture="media" allow="encrypted-media" allowfullscreen=""></iframe>
    <h2>Report</h2>
    <iframe
src="https://us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2Fdata%2Bmodules%2Freport&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;action=run&amp;format=HTML&amp;prompt=false" width="320" height="200" frameborder="0" gesture="media" allow="encrypted-media" allowfullscreen=""></iframe>
    <h2>Story</h2>
    <iframe
src="https://us1.ca.analytics.ibm.com/bi/?perspective=story&amp;pathRef=.my_folders%2FMy%2BAnalytics%2BStory%2FCrop%2BYield%2BAnalysis%2BStory&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;action=view&amp;sceneId=model000001845cf13b42_00000000&amp;sceneTime=7550" width="1400" height="600" frameborder="0" gesture="media" allow="encrypted-media" allowfullscreen=""></iframe>
</html>
```

GitHub Link: https://github.com/IBM-EPBL/IBM-Project-25917-1659977265

Video Link: https://drive.google.com/file/d/1yPsSF-wRZTmugrROLg3rrwzEH9JlLMzS/view?usp=sharing

YouTube Link: https://youtu.be/D6bTiizvezw