Estimate the Crop Yield using Data Analytics

1. INTRODUCTION:

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. Rainfall conditions also influences the rice cultivation. In this context, the farmers necessarily requires a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is enormous in Indian agriculture. The data when become information is highly useful for many purposes.The estimation of production of crop helps these companies in planning supply chain decision like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield [2]. Farmers experience was the only way for prediction of crop yield in the past days. Technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc. Crop yield prediction has generated a lot interest in the research community and also for agriculture related organizations. Crop yield prediction helps the farmers in various ways by providing the record of previous crop

yield. This is helpful to government in framing policies related to crops such as crop insurance policies, supply chain operation policies.

Project Overview:

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 seasons is an ample task and an advantageous task for every nation with respect to assesses the overall crop yield prediction and estimation. At present a common issue worldwide is, farmers are stressed in producing higher crop yield due to the influence of unpredictable climatic changes and significant reduction of water resource worldwide. A study was carried out to collect the data on world climatic changes and the available water resources which can be used to encourage advanced and novel approaches such as big data analytics to retrieve the information of the previous results to the crop yield prediction and estimation. Study imported that the selection and usage of the most desirable crop according to the existing conditions, support to achieve the higher and enhanced crop yield.

Purpose:

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.

2. LITERATURE SURVEY:

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. For example, low precipitation or temperature extremes can drastically diminish rice yield. Growing better strategies to foresee yield efficiency in a mixture of climatic conditions can help to understand the role of different principle factors that influence the rice crop yield. Big data analytic

methods related to the rice crop yield prediction and estimation will certainly support the farmers to understand the optimum condition of the significant factors for the rice crop yield, hence can achieve higher crop yield.

Existing Problem:

The environmental factors affecting crop yields can be classified into abiotic and biotic constraints. Actually, these factors are more intensified with global warming which leads to climate change. Abiotic stresses adversely affect growth, productivity and trigger a series of morphological, physiological, biochemical and molecular changes in plants. The abiotic constraints include soil properties (soil components, pH, physicochemical and biological properties), and climatic stresses (drought, cold, flood, heat stress, etc.). On the other hand, biotic factors include beneficial organisms (pollinators, decomposers and natural enemies), pests (arthropods, pathogens, weeds, vertebrate pests) and anthropogenic evolution.

References:

[1] Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. A Survey on Crop Yield Prediction based on Agricultural Data, International Journal of Innovative Research in Science, Engineering and Technology. 2017.

[2] Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki. Analysis of agriculture data using datamining techniques: application of big data. Journal of Big data. 2017.

[3] Majumdar J, Ankalaki S. Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves. International Conference on Computational Science and Engineering. 2016.

[4] D Ramesh, B Vishnu Vardhan. Data Mining Techniques and Applications to Agricultural Yield Data. International Journal of Advanced Research in Computer and Communication Engineering. 2013.

[5] Swarupa Rani. The Impact of Data Analytics in Crop Management based on Weather Conditions. International Journal of Engineering Technology Science and Research. 2017.

[6] F K Van Evert, S Fountas, D Jakovetic, V Crnojevic, I Travlos, C Kempenaar. Big Data for weed control and crop protection. John Wiley & Sons Ltd on behalf of European Weed Research Society, 2017.

[7] Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua. Prediction of crop yield using Big Data. 8th International Symposium on Computational Intelligence and Design. 2015.

[8] Dakshayini Patil, M .S, Shirdhonkar. Rice Crop Yield Prediction using Data Mining Techniques: An Overview. International Journal of Advanced Research in Computer Science and Software Engineering, 2017.

[9] Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. A Survey on Crop Yield Prediction based on Agricultural Data, International Journal of Innovative Research in Science, Engineering and Technology. 2017.

[10] Yogesh Gandge, Sandhya. A Study on Various Data Mining Techniques for Crop Yield Prediction, International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, IEEE, 2017

[11] R. Sujatha, P.Isakki Devi. A Study on Crop Yield Forecasting Using Classification Techniques, IEEE, 2016.

[12] V. Sellam and E. Poovammal. Prediction of Crop Yield using Regression Analysis, Indian Journal of Science and Technology, 2016.

[13] Patricio Grassinia, Lenny G.J. van Bussel, Justin Van Warta, Joost Wolf, Lieven Claessens, d, Haishun Yanga, Hendrik Boogaarde, Hugo de Groote, Martin K. van Ittersumb, Kenneth G. Cassman. How good is good enough,Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Research. 2015;.

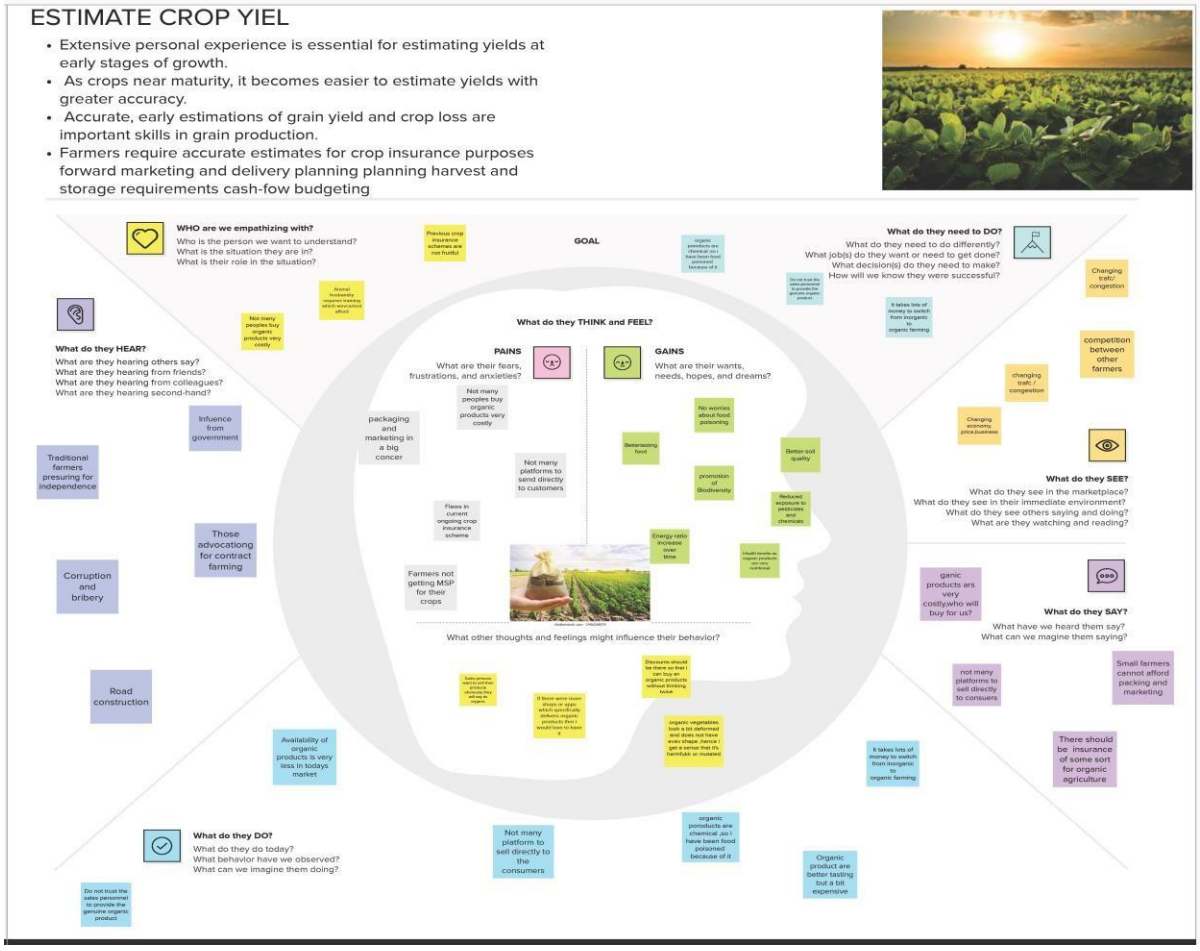
[14] David B. Lobell, The use of satellite data for crop yield gap analysis, Field Crops Research-143, 2013.

[15] Martin K. van Ittersuma, Kenneth G. Cassmanb, Patricio Grassinib, Joost Wolfa, Pablo Tittonell, Zvi Hochmand. Yield gap analysis with local to global relevance.

3.IDEATION & PROPOSED SOLUTION:

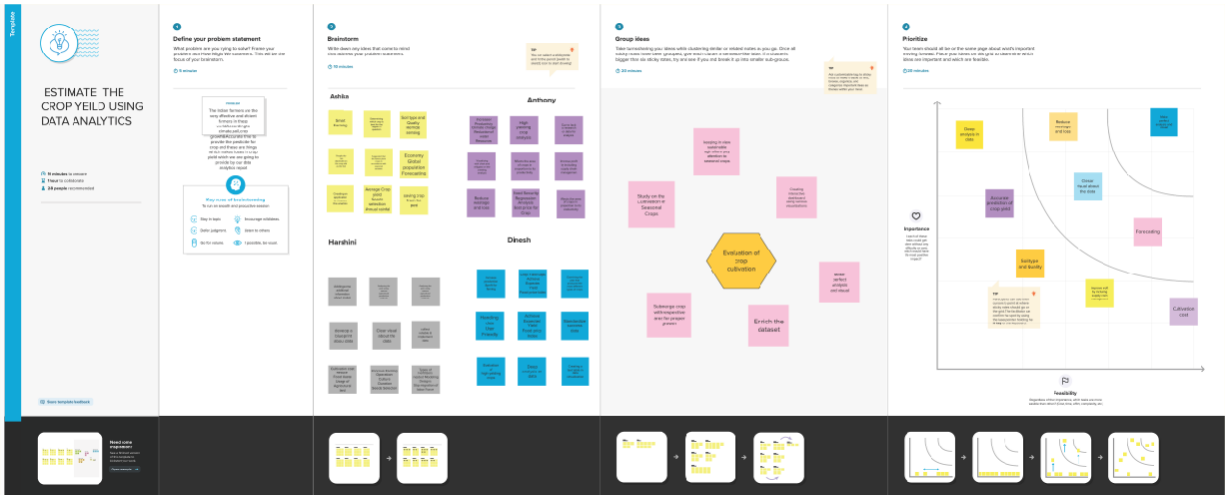
Empathy Map Canvas:

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.



Ideation & Brainstorming:

Ideation is a creative process where designers generate ideas in sessions (e.g., brainstorming, worst possible idea). It is the third stage in the Design Thinking process. Participants gather with open minds to produce as many ideas as they can to address a problem statement in a facilitated, judgment-free environment.



Proposed Solution:

- Problem Statement:
  - The frequent changes in climate conditions are affecting more in crop production.
  - Most of the forecasts are seasonal and are available around 1-2 months before the crop harvesting.
  - Farmers are benefited if recommendation and forecast of crop are available before sowing of crop.
- Idea/Solution description:
  - Utilization of data analysis associated with IBM Cognos results in estimation of crop yield with its activities such as including the climatic data, soil data into account.
  - Contribution of this project is to improve the agricultural productivity and provide the crop recommendation to the farmers.
  - By performing Weather indices-based Regression Analysis ,Seasonal analysis ,Time series analysis.

- Novelty / Uniqueness:
    - Improve operational efficiency and increase productivity and profitability.
    - Draw analytical insights on expenses, inventory and crop growth.
    - Prediction of the outcome is the remarkable advantage of our project.

Problem Solution fit:

- CUSTOMER SEGMENT(CS):

Farmer Crop production will become more difficult with weather change and environmental degradation.

- JOB-TO-BE-DONE / PROBLEM :
    - Analyse the growth of plants
    - To improve the crop productivity
    - Improve the profit
- TRIGGERS
    - Before: Fear and doubt fullness about the software rather than hope.
    - After: Trust.
- AVAILABLE SOLUTION
    - Improve the Quality of the product
    - Increased Productivity

4. REQUIREMENT ANALYSIS:

Functional requirement:

Functional requirements are product features or functions that developers must implement to enable users to accomplish their tasks.

- **User Installation -** User can install the app from Google play store or from the website.
- **User Registration-**User can create an account by registering through form or gmail
- **User Confirmation-**Confirmation via Email, OTP or SMS.
- **User Login-**User should login the app with the user name and password or easier login for already existing users.
- **Data Collection-**The previous year crop yield data set and the farm yield methodology,User data of the farmer ,Details of the Seasons and the Regions.
- **Data Analysis-**The analysis is done on the given data to gain useful insights on the crop yield and also the data according to the set of past data of the multiple farmers.
- **Estimation-**Creating the perfect data module through attractive stories, dashboard and reports to increase the understandability of data.

Non-Functional requirements:

Nonfunctional requirements not related to the system functionality, rather define the system should perform.

- **Usability-**The given datasets , analyzing is done and report is created. Accordingly sowing of crops are recommended. The features of this dashboard is all the data which is needed will be displayed in one which is easily understandable.

- **Security**-Only recognized users can access the resource. Not all the users are allowed to access any particular or whole data from dashboard.
- **Reliability**-Using the interactive data visual dashboards we can easily understand the data reports and the added/erased dashboard will be retrieved soon.
- **Performance**-Interaction makes better performance between all user and impressing by the visual advise.
- **Availability**-The dashboard is easily available and accessible in smartphones and PC's and laptops.
- **Scalability**-The flexibility of the methodology to implement the proposed solution is very easy that can make increase in the estimation of crop yield in the differ farms for different user. It has the high database storage so that the data stored can be viewed and retrieved anytime, anywhere.

5. PROJECT DESIGN:

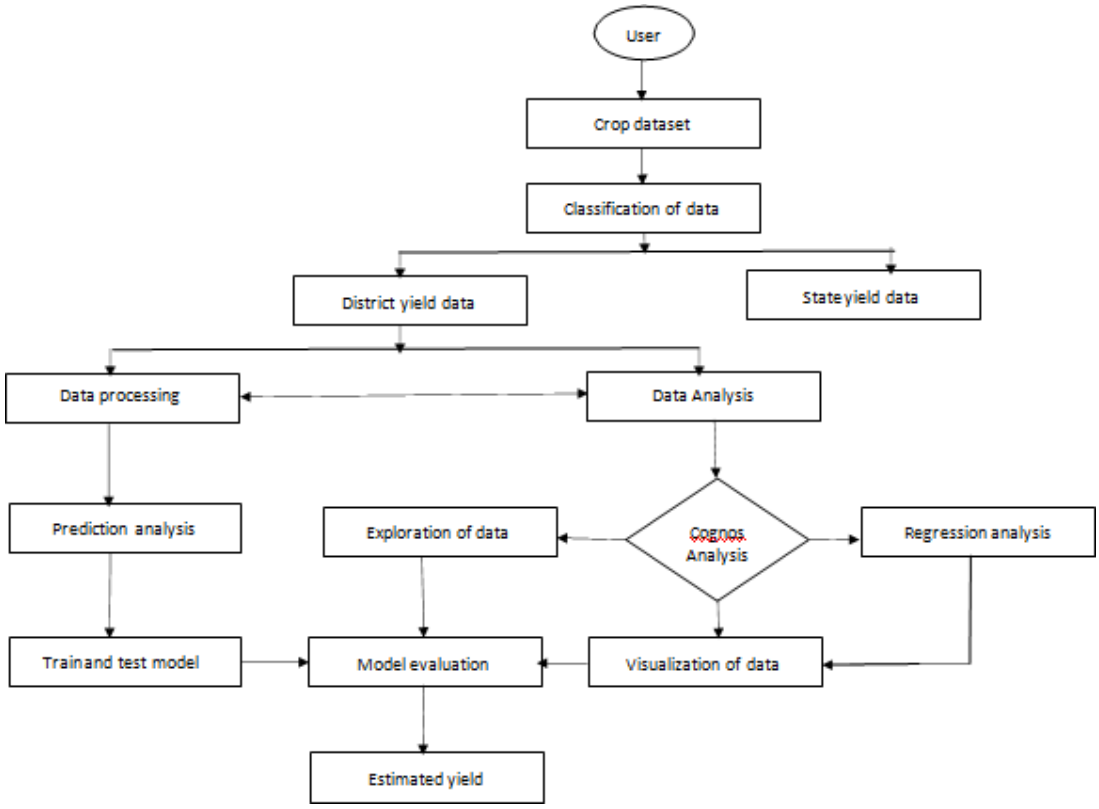
Data Flow Diagrams:

- 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.

- Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled.

- They can be used to analyze an existing system or model a new one. Like all the best diagrams and charts, a DFD can often visually say things that would be hard to explain in words and they work for both technical and nontechnical audiences, from developer to CEO.

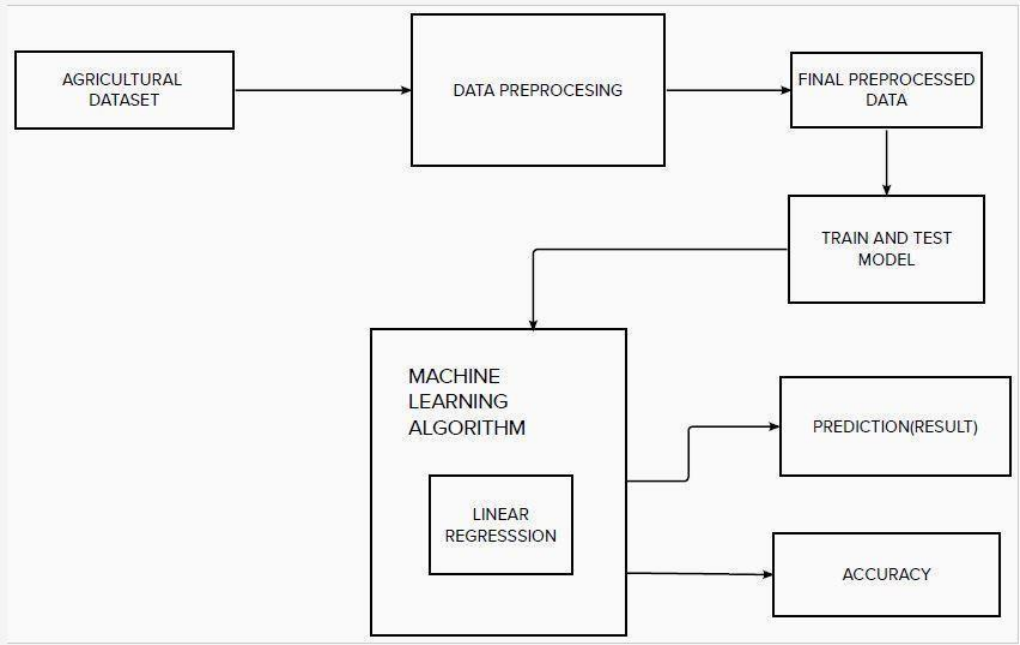


Solution & Technical Architecture:

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

production in India.

- The data is fetched from the user and data is analyse , cleaned , pre-processed and so on thus the data report is been created.
- Using IBM Cognos , the Data visual are been generated according to the data report which we have created using the user data.
- This can create huge change in felid , crop yield estimating and profit to the farme



User Stories:

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user and Laptop users)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Face book	I can register & access the dashboard with Face book Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	Can use the methods provided in the dashboard		Medium	Sprint-1
	Invest	USN-7	With help of desired results obtained from application, making profit or loss	Gain or Loss	High	Sprint-2
Administrator	Updating data		Collecting the data and storing it	Checking and updating dataset	High	Sprint-1
Customer (Web User)	Accessing the resources	USN -8	Using my own credentials for accessing the data	These resources cannot be accessed by others but only me	High	Sprint -1
	Satellite Visioning	USN -9	Having a view with geographic data		Medium	Sprint-2
Customer tools	Tools	USN -10	Analysis is performed by tools like cognos analytics	Ease of accessing the tools	High	Sprint 2

6. PROJECT PLANNING & SCHEDULING:

Sprint Planning & Estimation:

S.no	Milestone	Activities	Start Date	End Date
1	Solution Requirement	Creating the IBM Cognos for creating dashboard anddata visualization charts.	22-Aug-2022	24-Aug-2022
2	Project Objectives	Prepare the project objectives.	22-Aug-2022	24-Aug-2022
3	Project Flow	Prepare the project flow.	22-Aug-2022	24-Aug-2022
4	IBM Cloud Account	Creating IBM cloud account.	22-Aug-2022	24-Aug-2022
5	IBM Cognos Analytics	Creating IBM cognos account.	22-Aug-2022	24-Aug-2022
6	Working with the Dataset	Understanding The Dataset Loading The Dataset.	24-oct-2022	19-Nov-2022
7	Data Visulization Charts	➢ Seasons With Average Productions ➢ With Years Usage of Area And Production ➢ Top 10 States with Most Area ➢ States With the Crop Production Along with Season	24-oct-2022	19-Nov-2022
8	Creating the	Creating The Dashboard	24-oct-	19-Nov-

	Dashboard		2022	2022
9	Export the Analytics	Export The Analytics	24-oct-2022	19-Nov-2022
10	Ideation Phase	➢ Literature Survey On The Selected Project & Information Gathering Prepare ➢ Empathy Map ➢ Ideation	22-Aug-2022	17-Sept-2022
11	Project Design Phase - I	➢ Proposed Solution ➢ Problem Solution Fit ➢ Solution Architecture	22-Aug-2022	17-Sept-2022
12	Project Design Phase - II	➢ Customer Journey ➢ Functional Requirement ➢ Data Flow Diagrams ➢ Technology Architecture	22-Aug-2022	01-Oct-2022
13	Project Planning Phase	➢ Prepare Milestone & Activity List ➢ Sprint Delivery Plan	17-Oct-2022	22-Oct-2022
14	Project Development Phase	➢ Project Development - Delivery of Sprint-1 ➢ Project Development - Delivery of Sprint-2 ➢ Project Development - Delivery of Sprint-3 ➢ Project Development - Delivery of Sprint-4	24-Aug-2022	19-Nov-2022

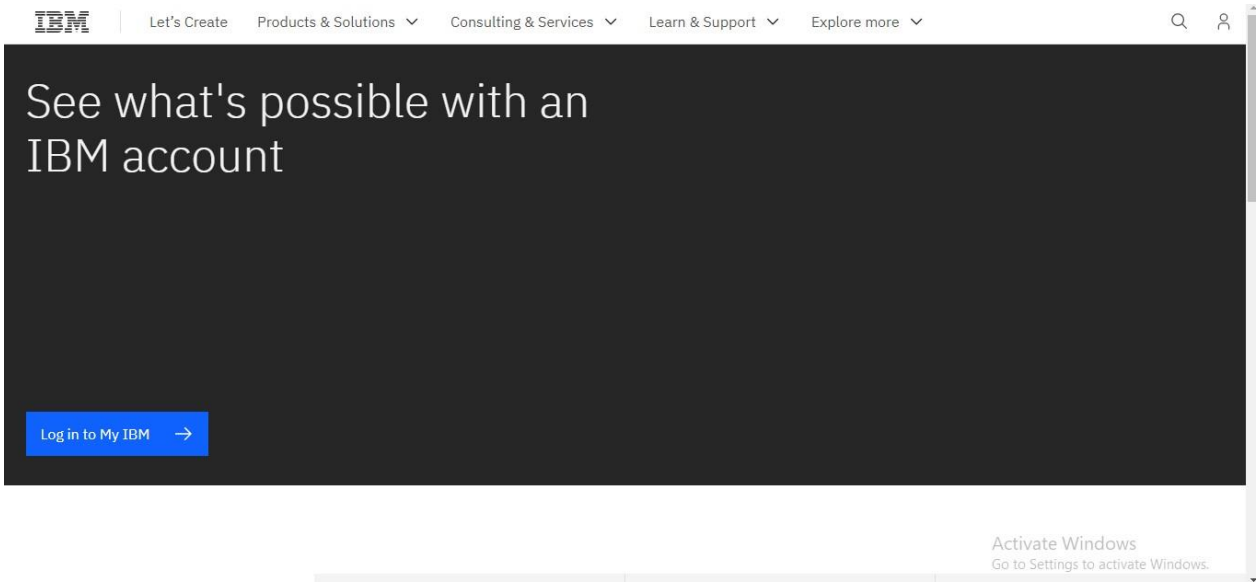
Sprint Delivery Schedule:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Ashika,Dinesh
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Harshini,Anthony Thomas Yeswin.
Sprint-2		USN-3	As a user, I can register for the application through Google	2	Low	Ashika, Harshini
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Low	Dinesh, Anthony Thomas Yeswin
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Ashika,Dinesh
Sprint- 3	Dashboard	USN-6	As a user, I can freely use my dashboard and explore the features	2	High	Harshini,Anthony Thomas Yeswin.
Sprint- 2		USN-7	As a user, I can use the credentials to access the resources of my application	2	High	Ashika, Harshini
Sprint- 3		USN-8	Performance of Data manipulations on the application	1	High	Dinesh, Anthony Thomas Yeswin
Sprint- 3	Visualizations	USN-9	I can create dashboards with particular datasets	2	High	Ashika,Dinesh
Sprint- 4		USN-10	Predictive analysis can be done	1	High	Harshini,Anthony Thomas Yeswin
Sprint- 3		USN-11	I can create stories with particular datasets	2	High	Ashika,Dinesh
Sprint- 4		USN-12	I can deliver and export reports according to the dashboards and stories created	2	High	Harshini,Anthony Thomas Yeswin

7. WORKING WITH THE DATASET & DATA VISUALISATION:

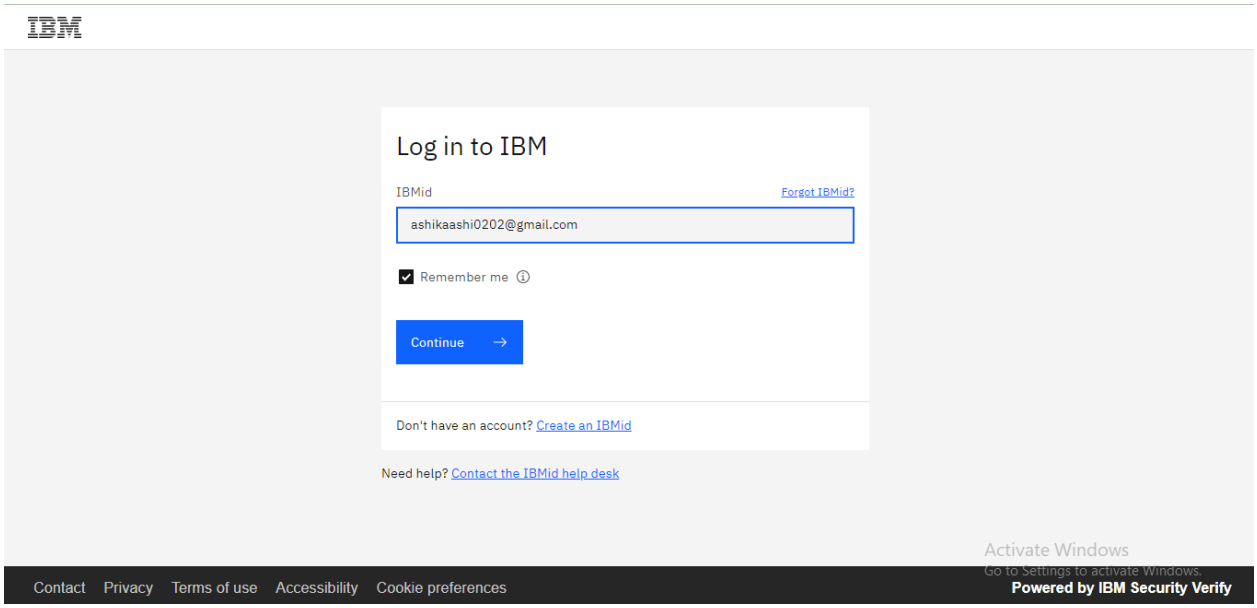
Registration:

- The registration to the IBM cogons analytics web page for the student user.
- Register with validly Email ID.
- Enter the user name and password through the register pages.
- Enter Next button.
- Enter the Additional Information through the IBM cogons analytics.
- Finally verify the Email.
- Thus the IBM cogons account is registered.



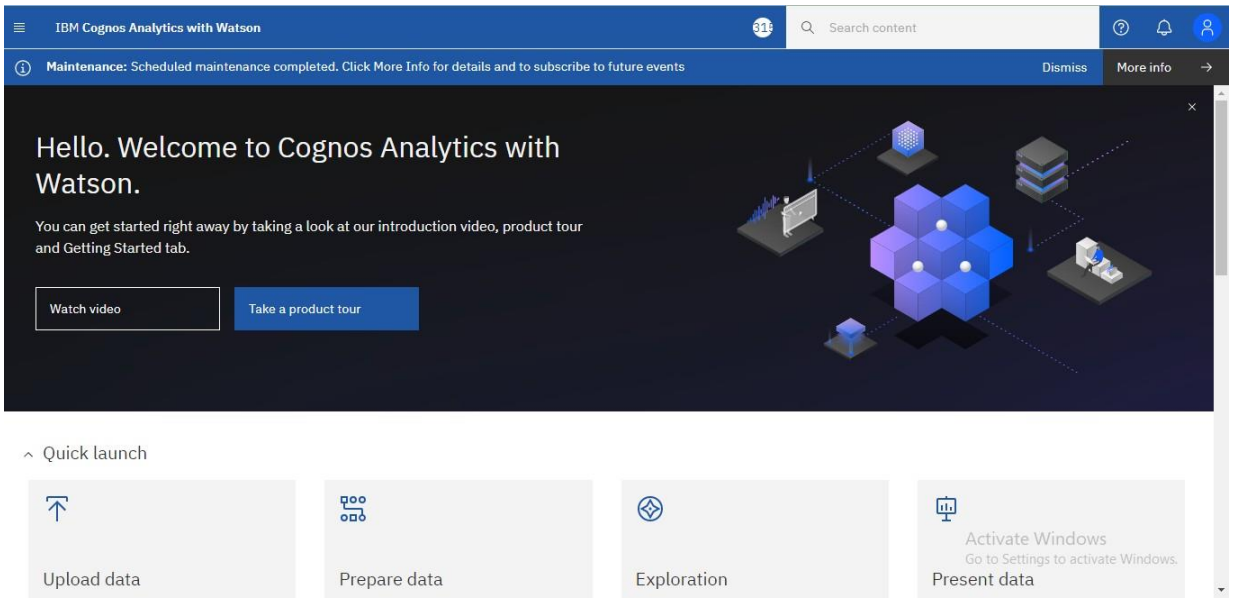
Login page:

- After complete the registration process and click login page.
- Enter the registration email ID on the login page and then click continue.
- Enter the password and click continue.



Home page:

- After complete and then click launch button.
- Then the particular page will be display.



Working with 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.

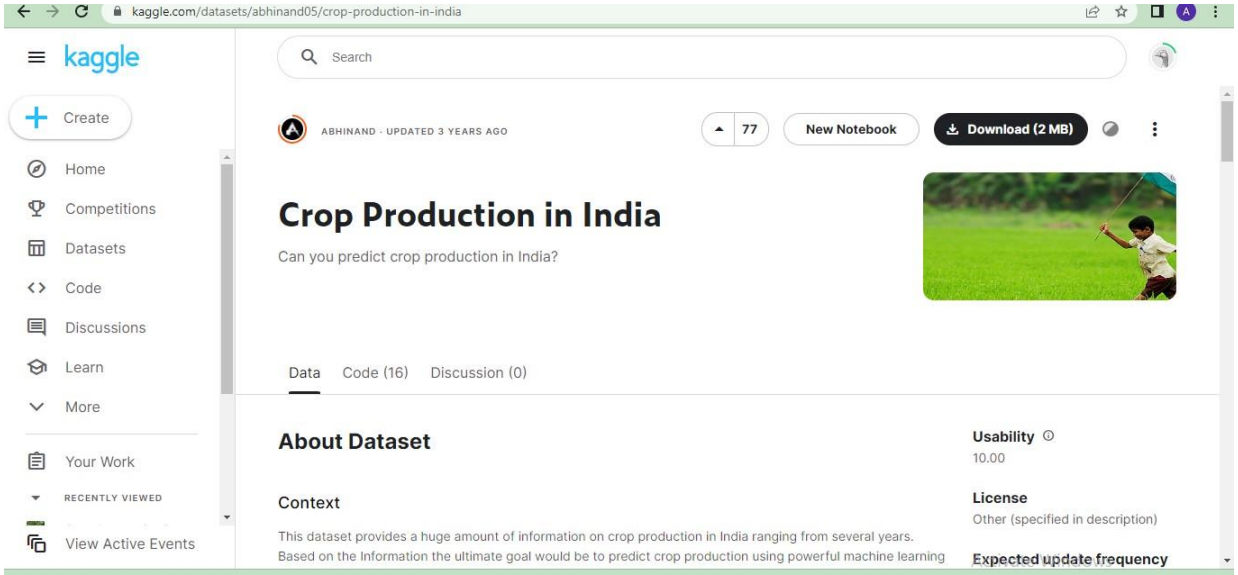
Dataset Link : Dataset

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

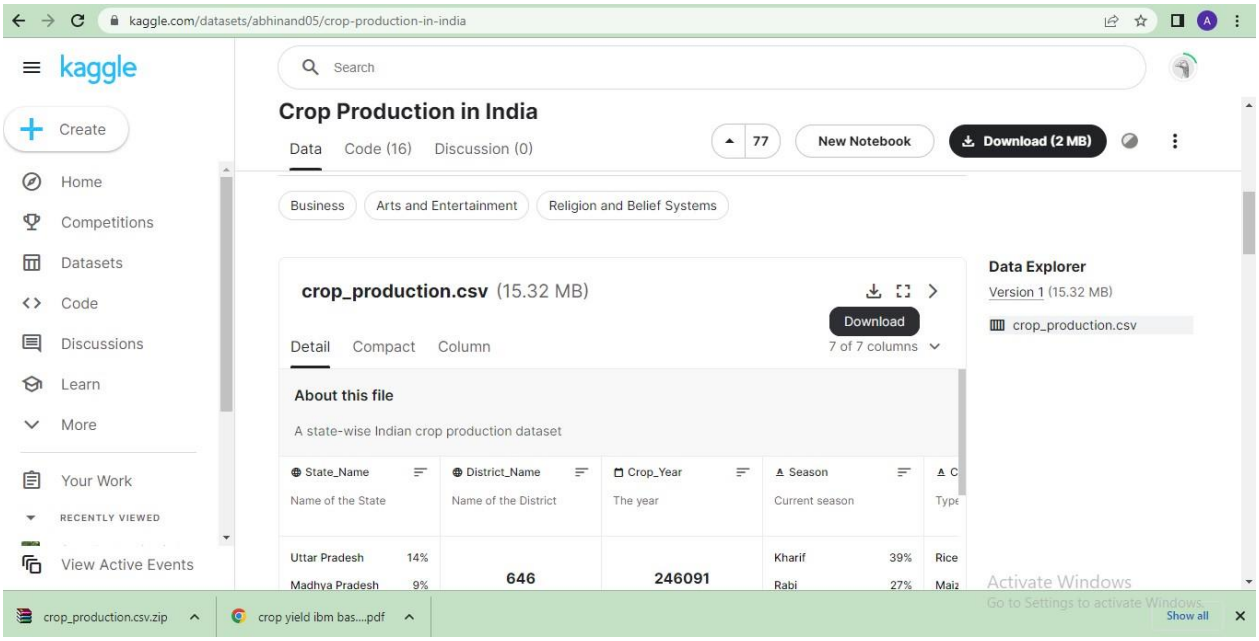
- State Name - All the Indian State names.
- District Name -Different District names.
- Crop Year- contains the crop years.
- Season – Different seasons for crop production.
- Area- Total number of areas covered.
- Production- production of crops.

Downloading the dataset:

- Click on the given dataset.
- Thus opens the below dataset window.



- Scroll on the bottom of the window and click on the download option provided.

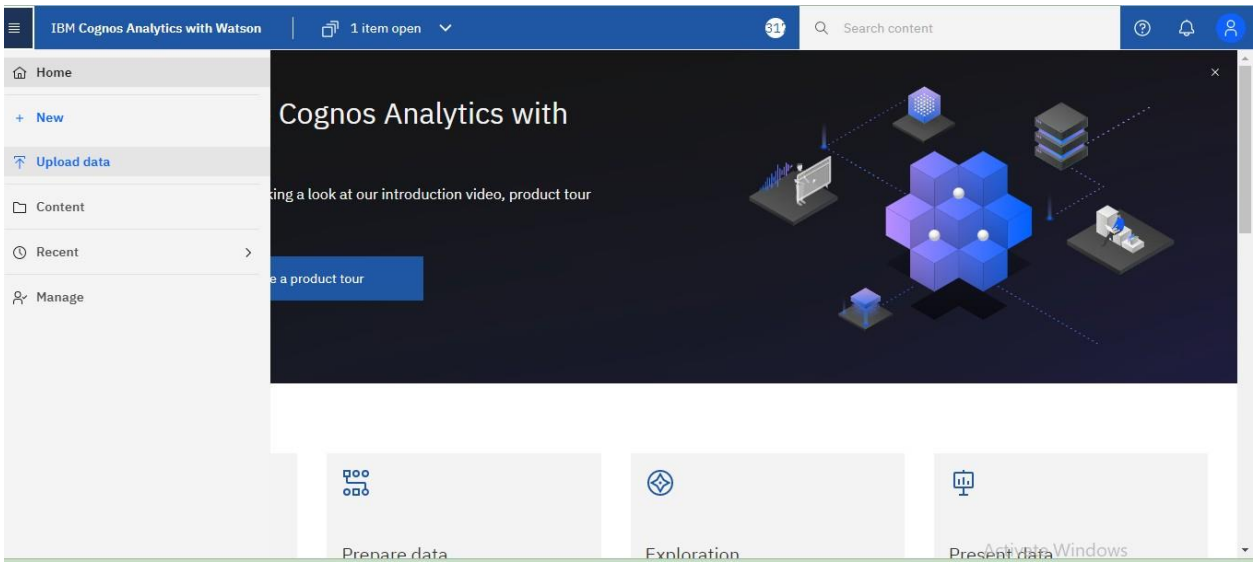


- Thus the dataset is downloaded.

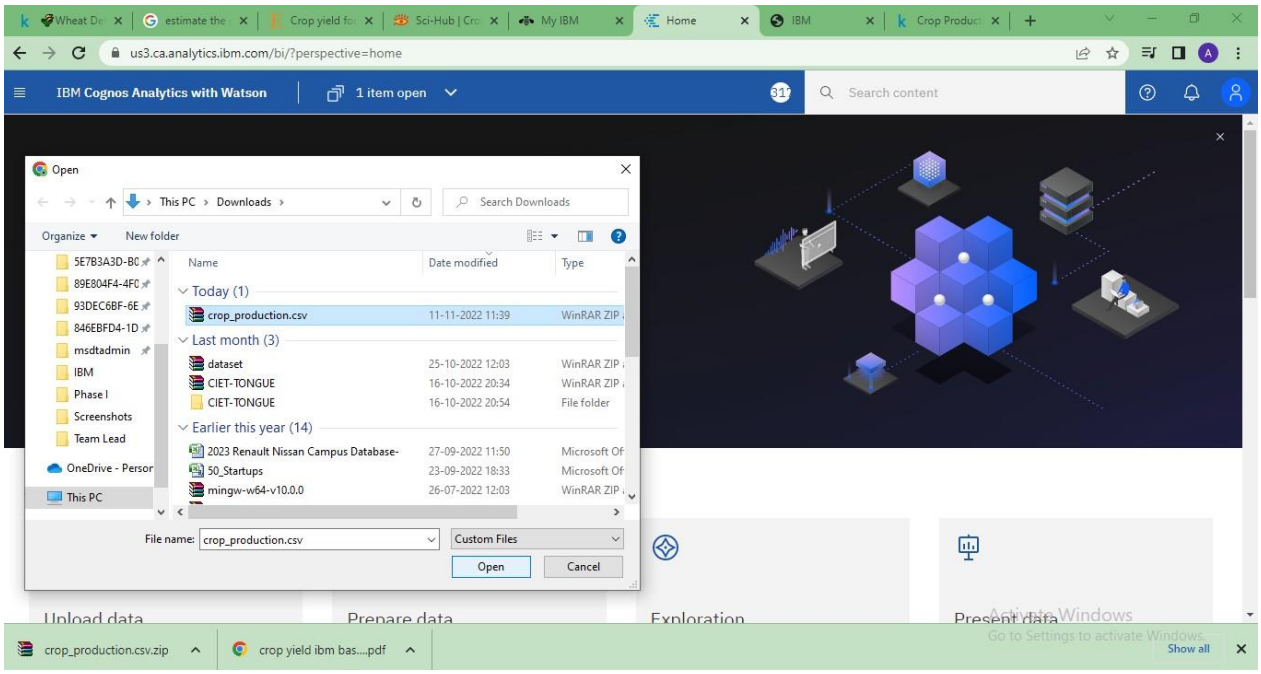
Loading the Dataset:

Before we can build a view and analyze our data, we must first connect the data to IBM Cognos.Cognos supports connecting to a wide variety of data, stored in a variety of places.The data might be stored on our computer in a spreadsheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in our enterprise.

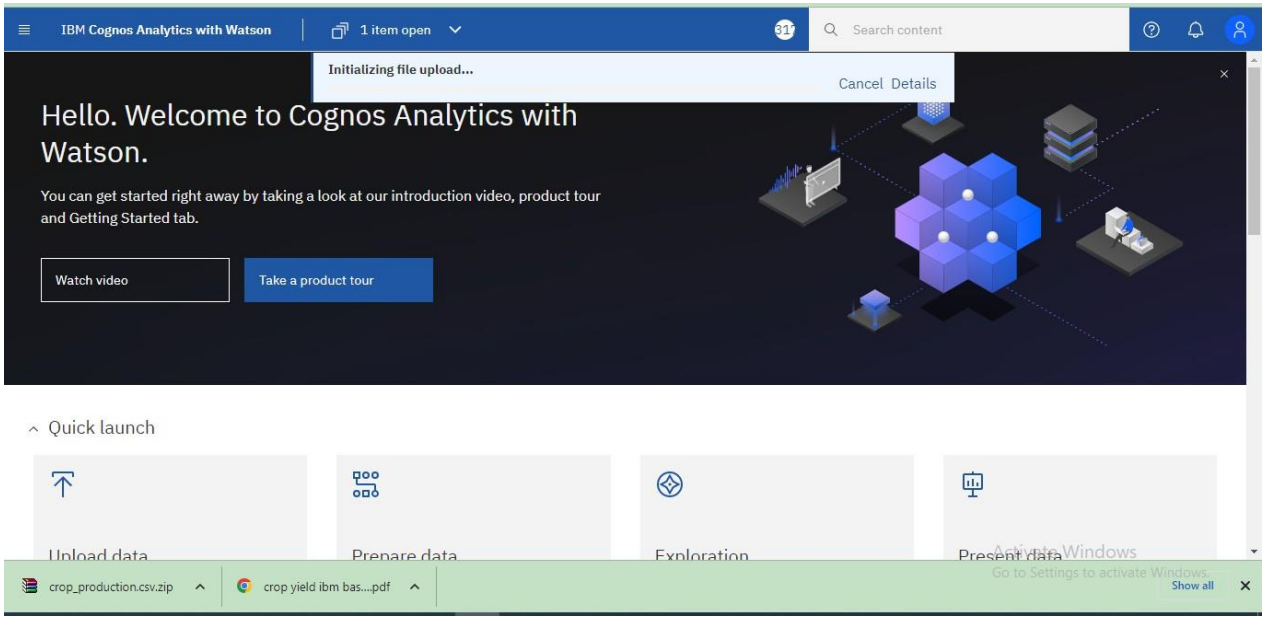
- Click on the menu icon in the left corner of the window.
- Therefore opens the menu.
- Click on the upload option from the menu.



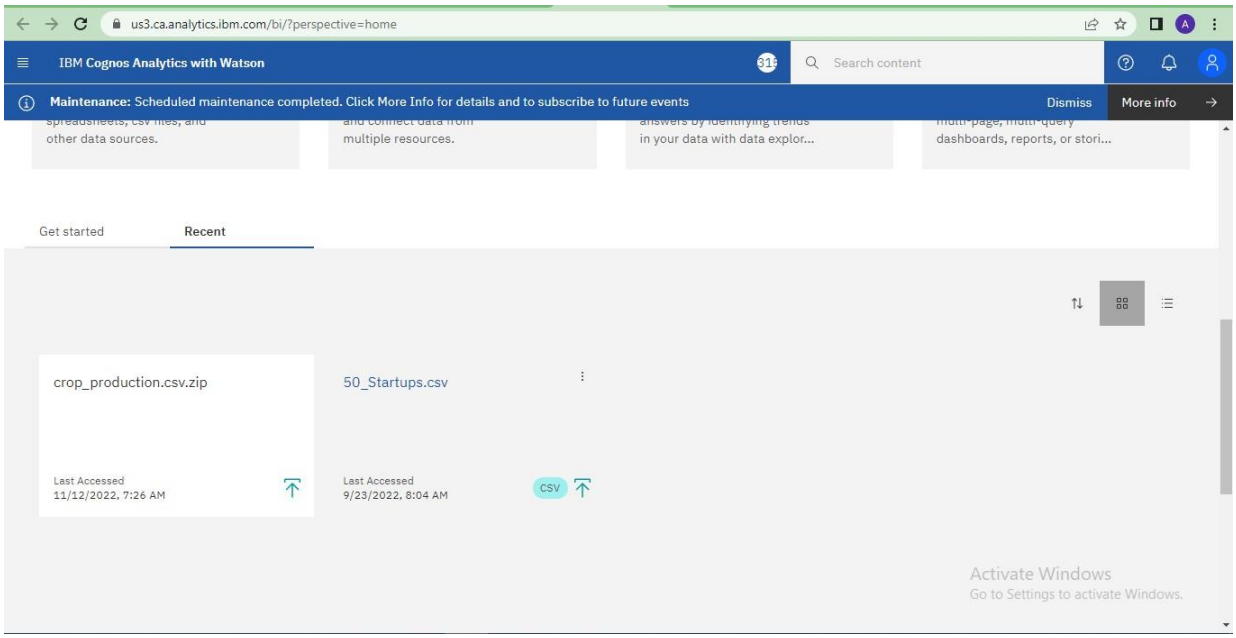
- Upload the dataset from the file.



- Thus the dataset inialized to upload.



- After initialization completed the dataset is uploaded successfully.



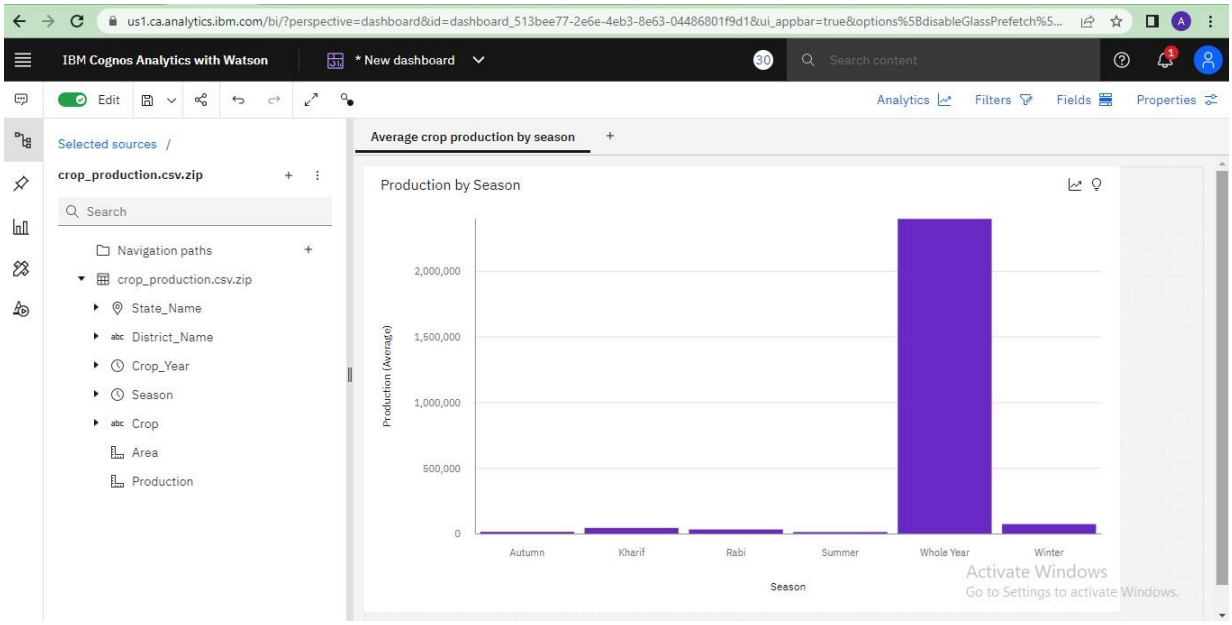
Data Visualization:

Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations. *Build a Visualization to showcase

- Login to IBM Cognos then in open menu click New - Dashboard.
- Select a template for our dashboard and click create.
- Then select the dataset crop_production.csv dataset under my content tab and click add In Add a source.
- Rename the tab title with suitable name.
- Select the visualizations chart whatever we want.

Average Crop Production by Season:

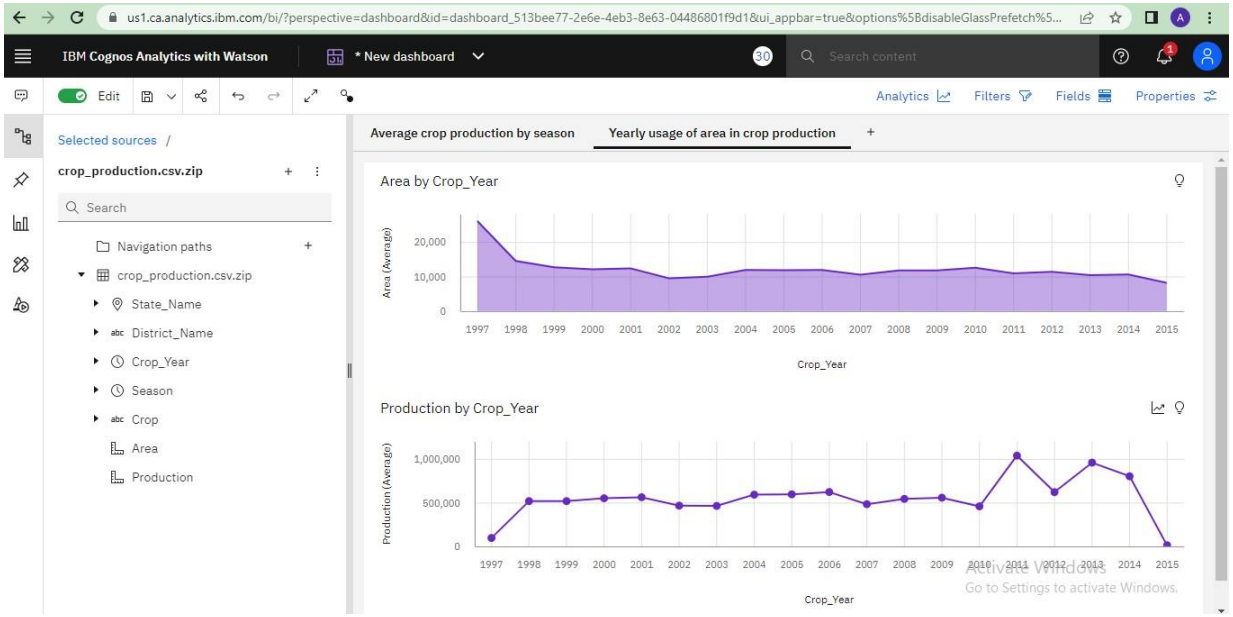
- We can select the bar chart for visualization the average crop production by season.
- Place the season in bar.
- Place the production in length.
- In color we place the seasons.



Year usage of Area in Crop Production:

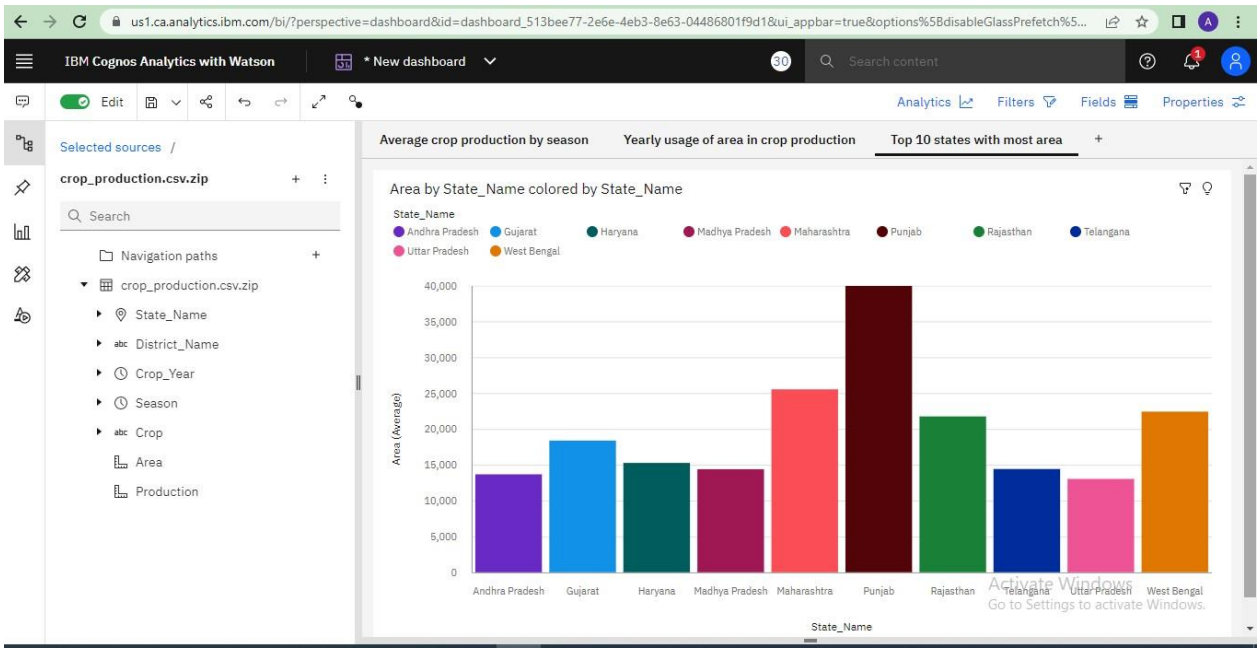
- First we can create the area by production.
  - We can choose the line and column chart for visualization.
  - In x-axis we put crop_year.
  - In y-axis we put area.

- Second we can create the production by crop_year.
  - We can choose the line chart for visualization.
  - In x-axis we put crop_year.
  - In y-axis we put production.



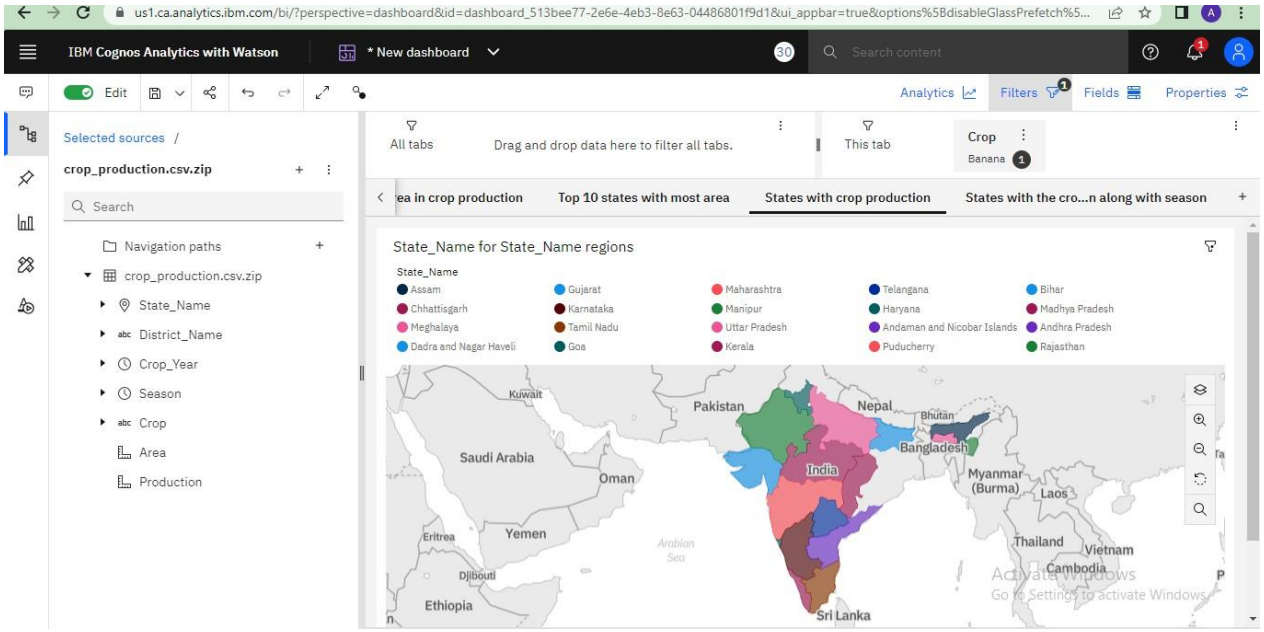
Top 10 states with most area:

- We can choose bar chart for the visualization.
- In bar we place the state_name.
- In length we place the area.
- In color we place the state_name.



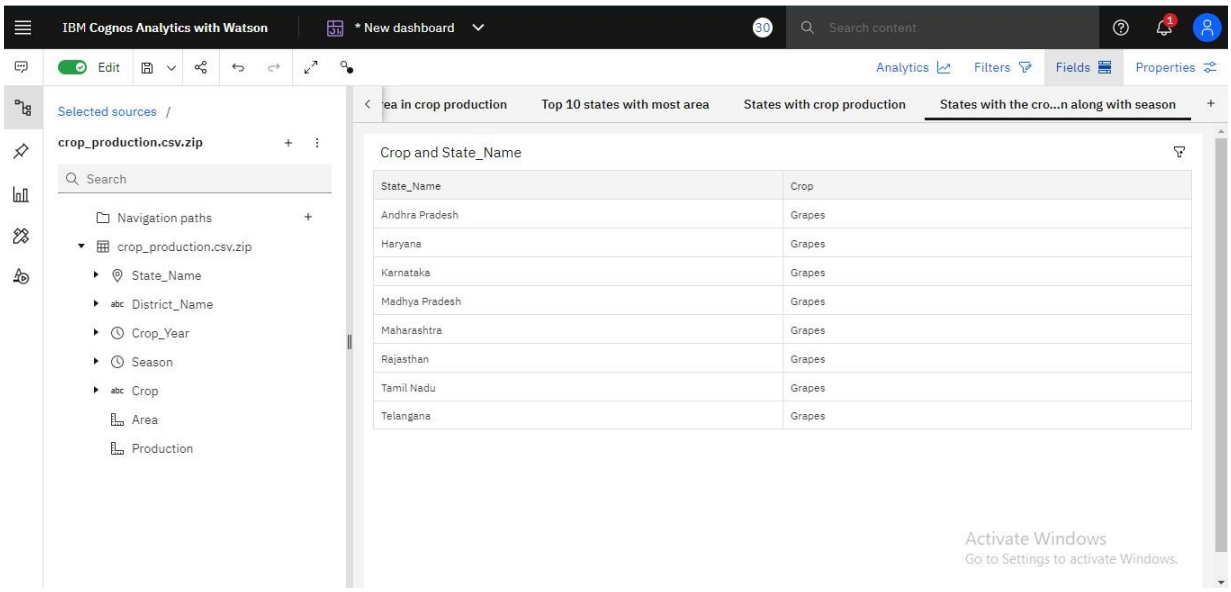
State with crop production:

- We choose the map chart visualization.
- In location we place the state_name.
- In color we place the state_name.
- We place crop production in filters.
- It will be able to see the production of India.



State with the crop production along with season:

- We choose the table chart In column we place the state_name.
- We use filters
- Then create another table for the season.
- Place season in column.



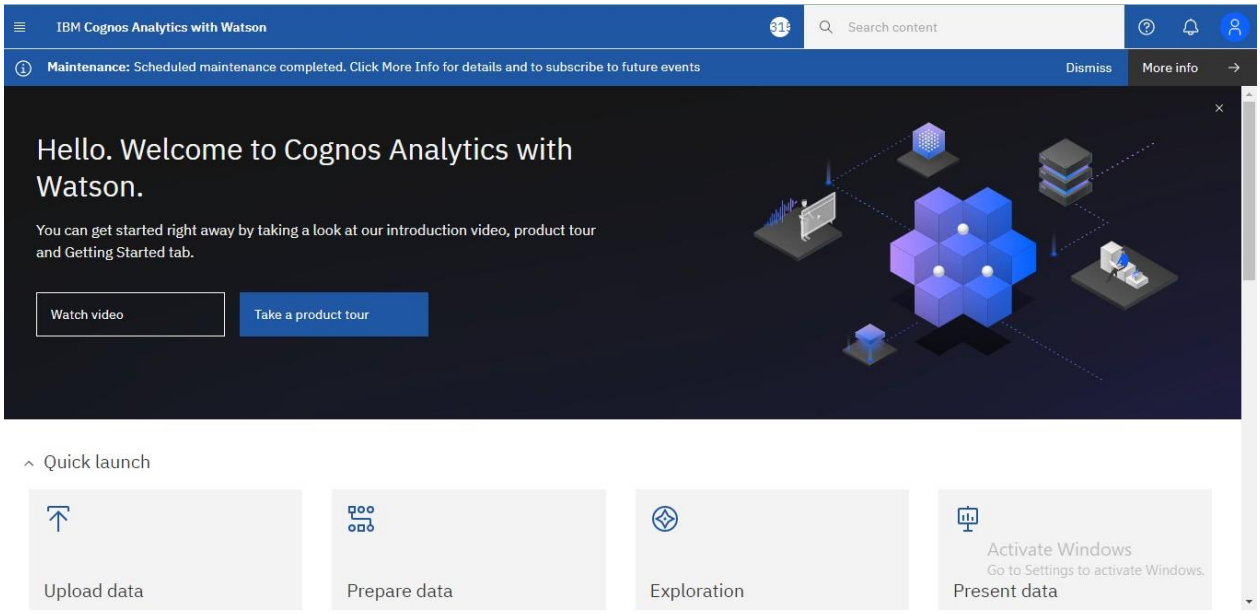
8. Creating Dashboard and visualization:

Creating Dashboard:

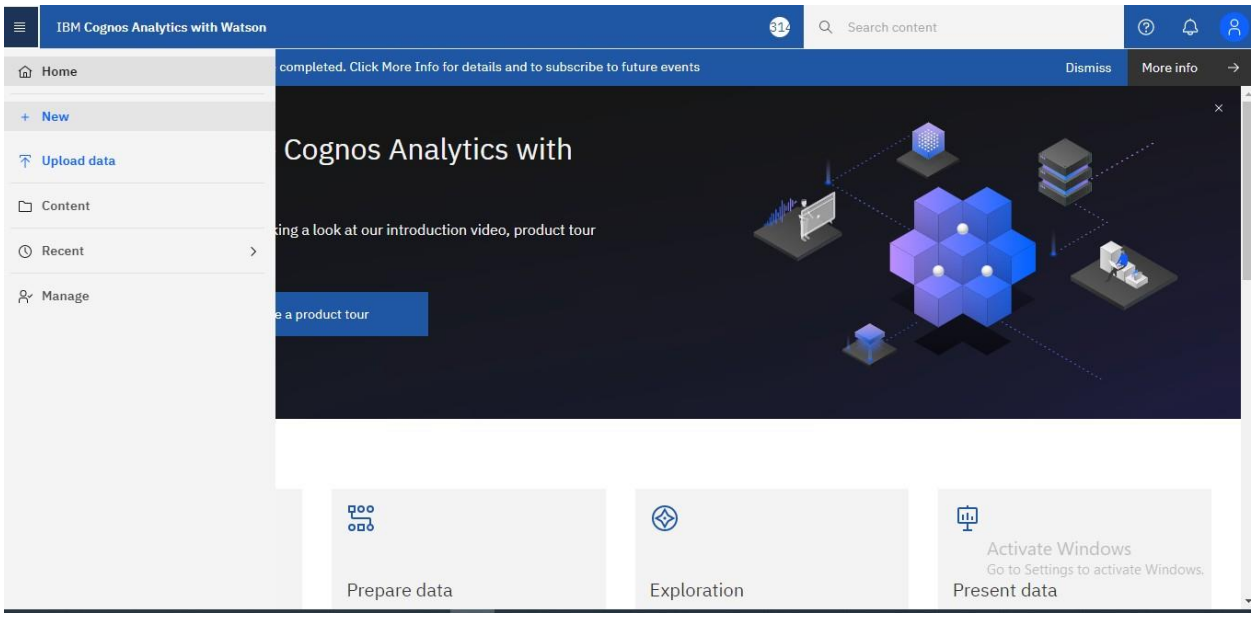
- A dashboard is a way of displaying various types of visual data in one place.
- Usually, a dashboard is intended to convey different, but related information in an easy-to-digest form.

Login to IBM Cognos:

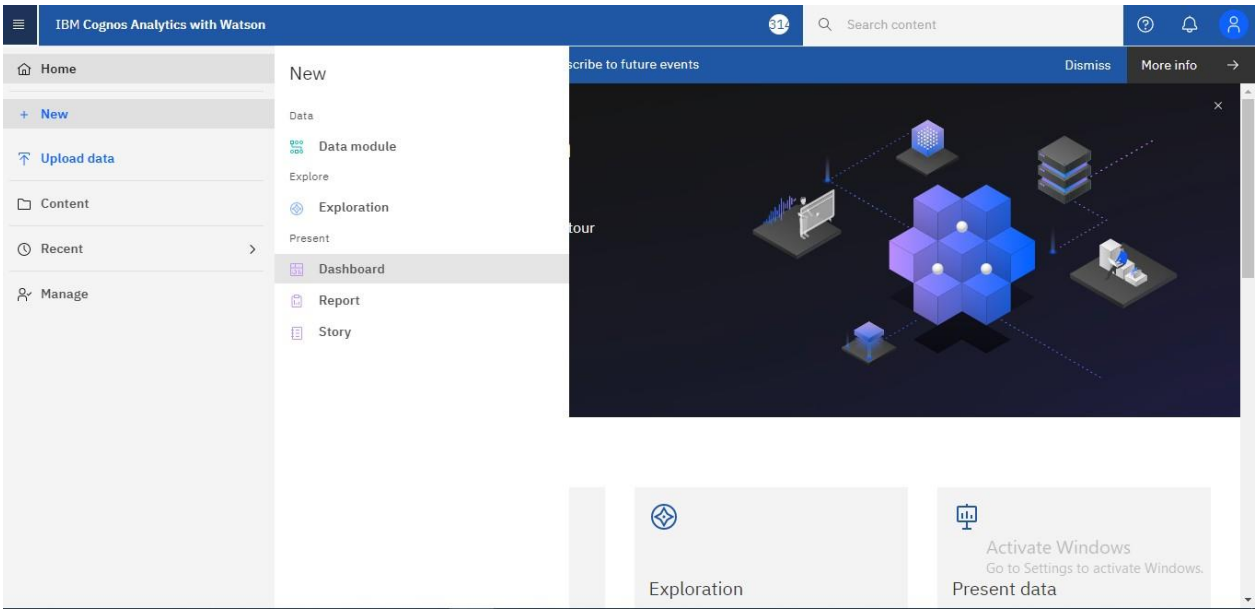
➢ Open to IBM cognos account.



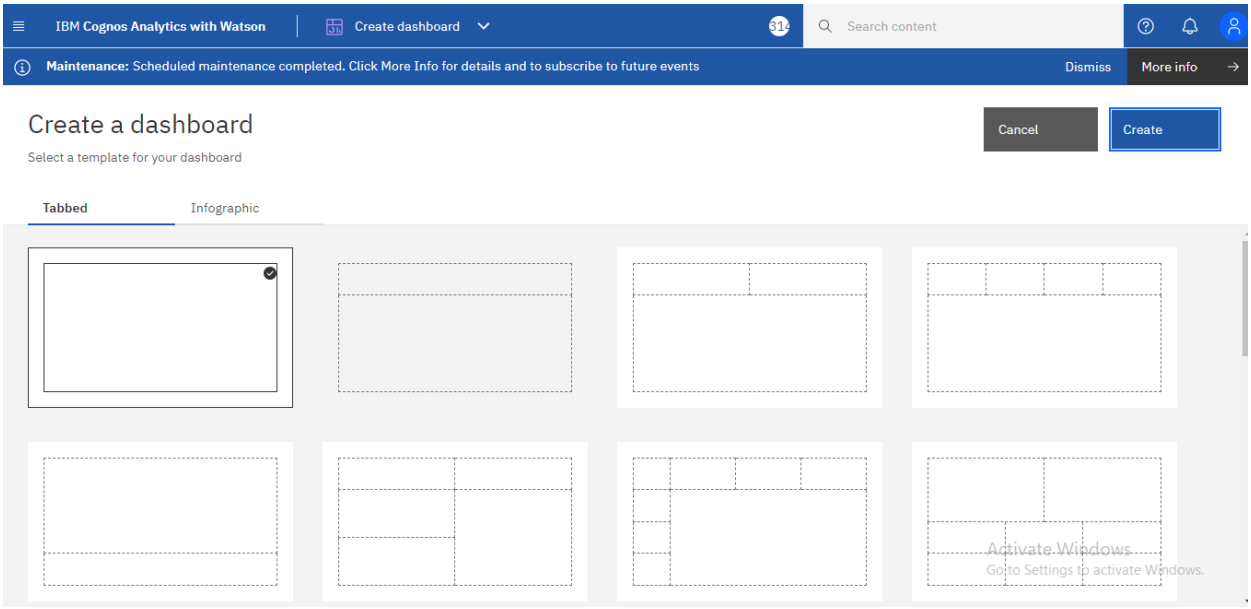
➢ Click on the menu and select new option.



- Select Dashboard option.

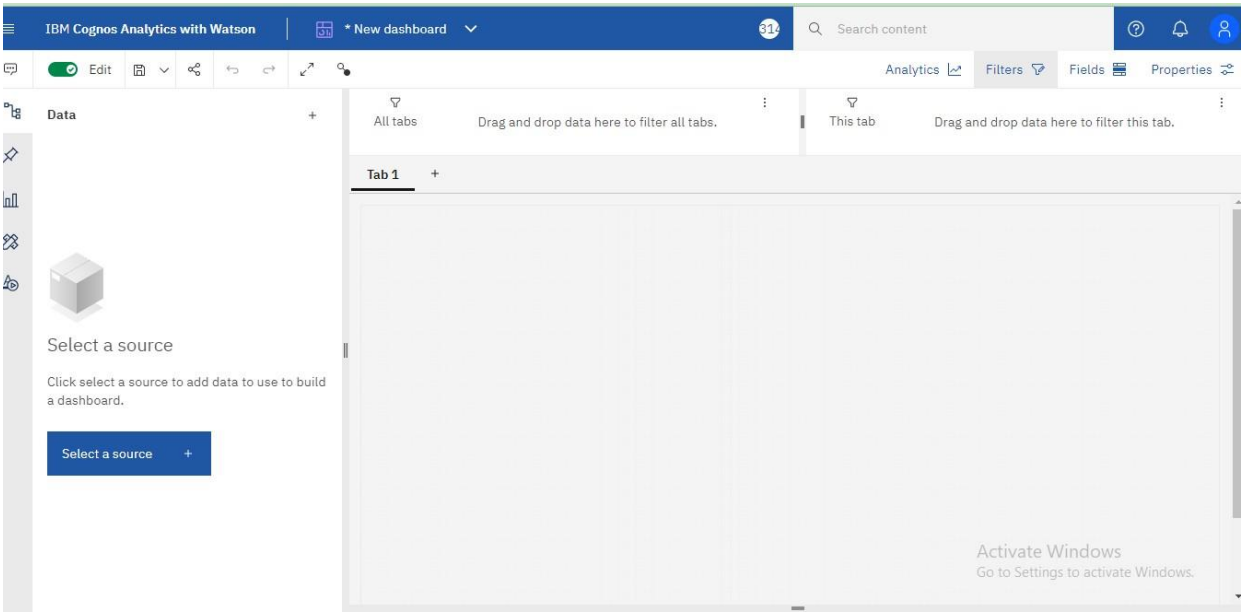


- Select the table from the option and click on the create option.

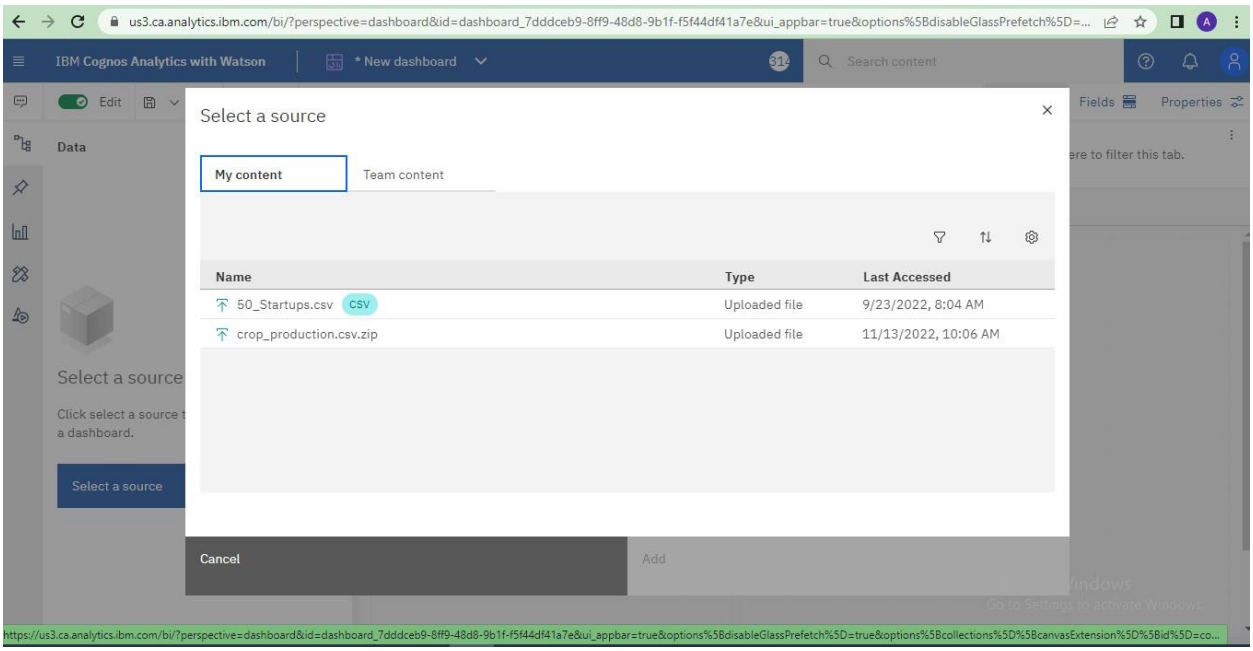


➢ Thus the dashboard is created.

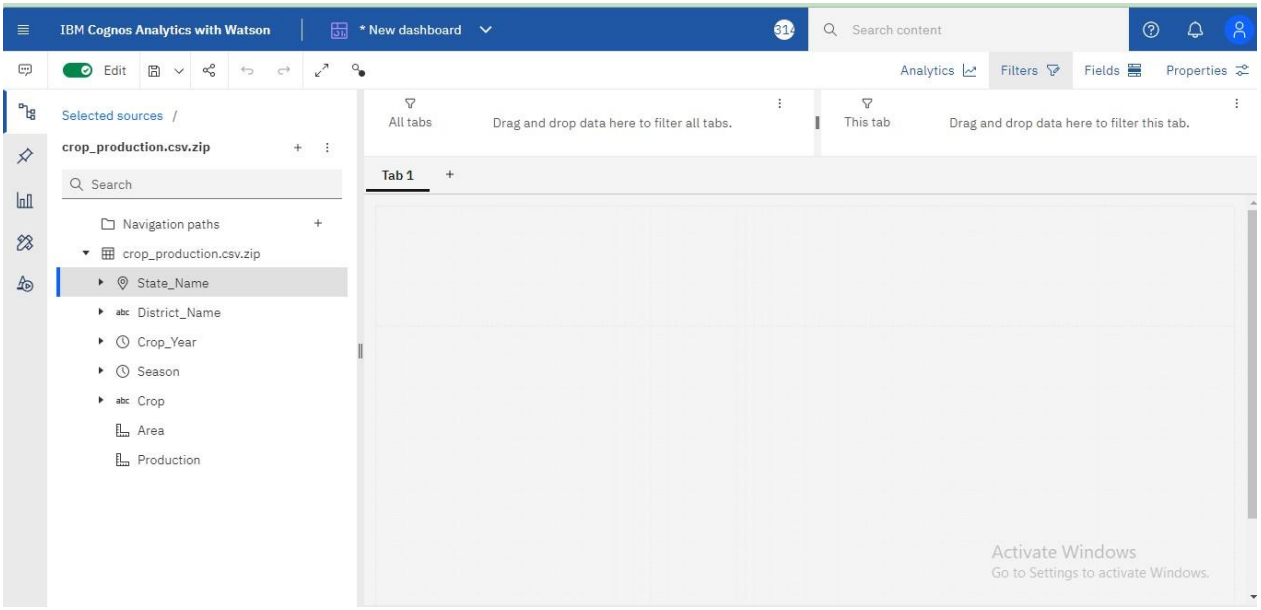
➢ Click on the select source option.



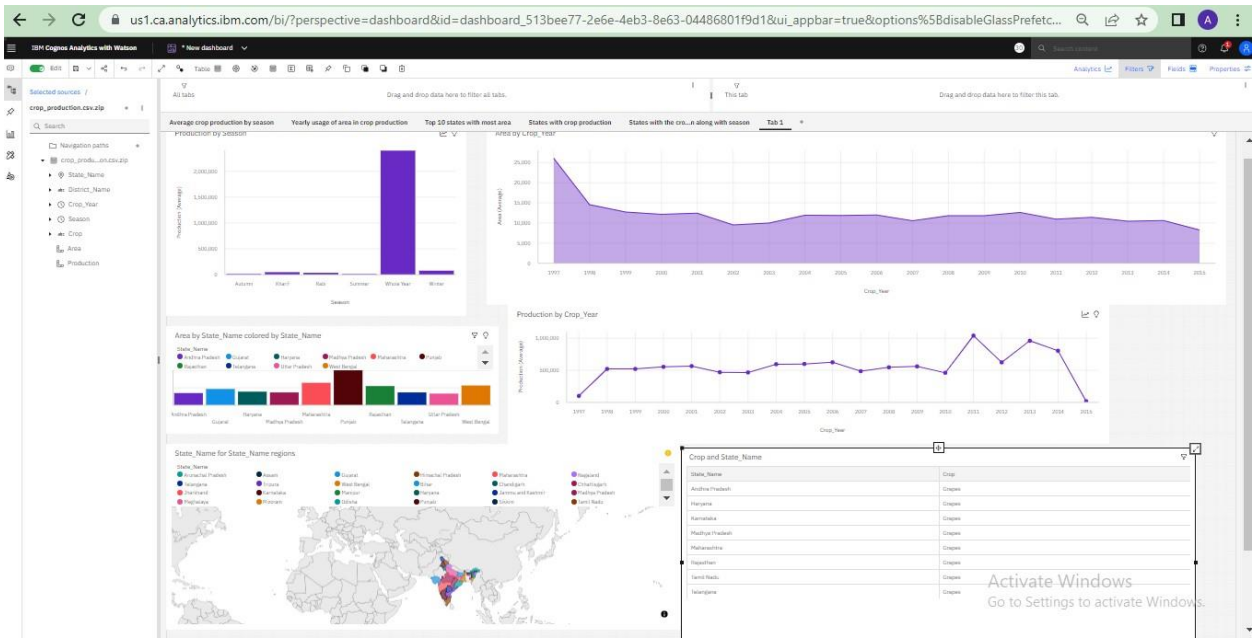
➢ Select the content option and select the dataset.



➢ Thus the dataset is uploaded successfully.



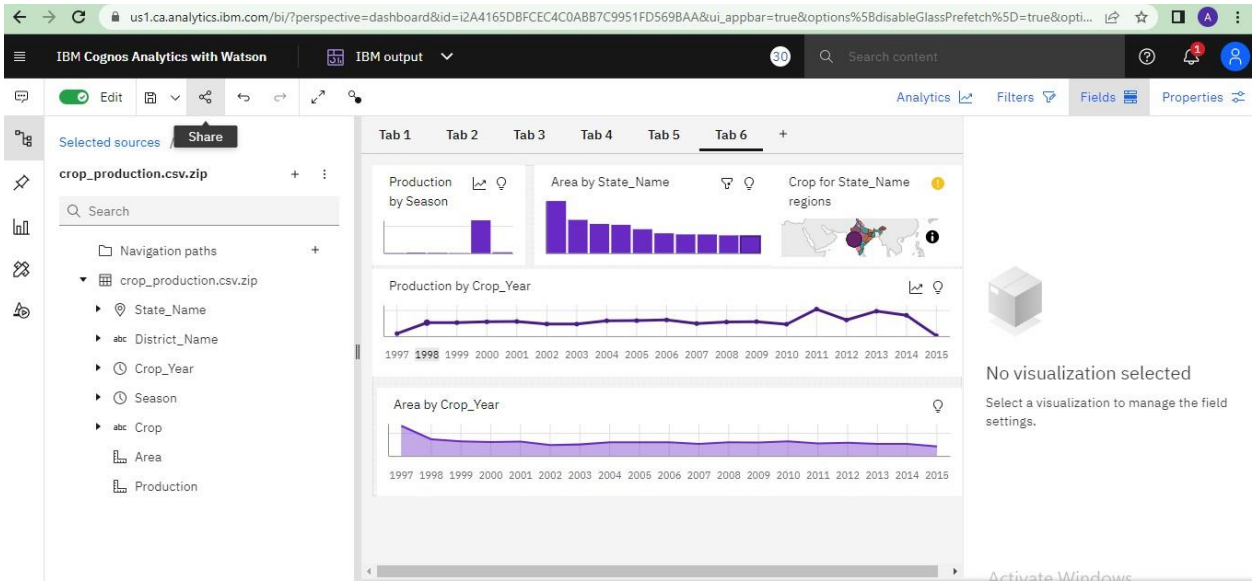
➢ Create the dashboard and collect the contents together.



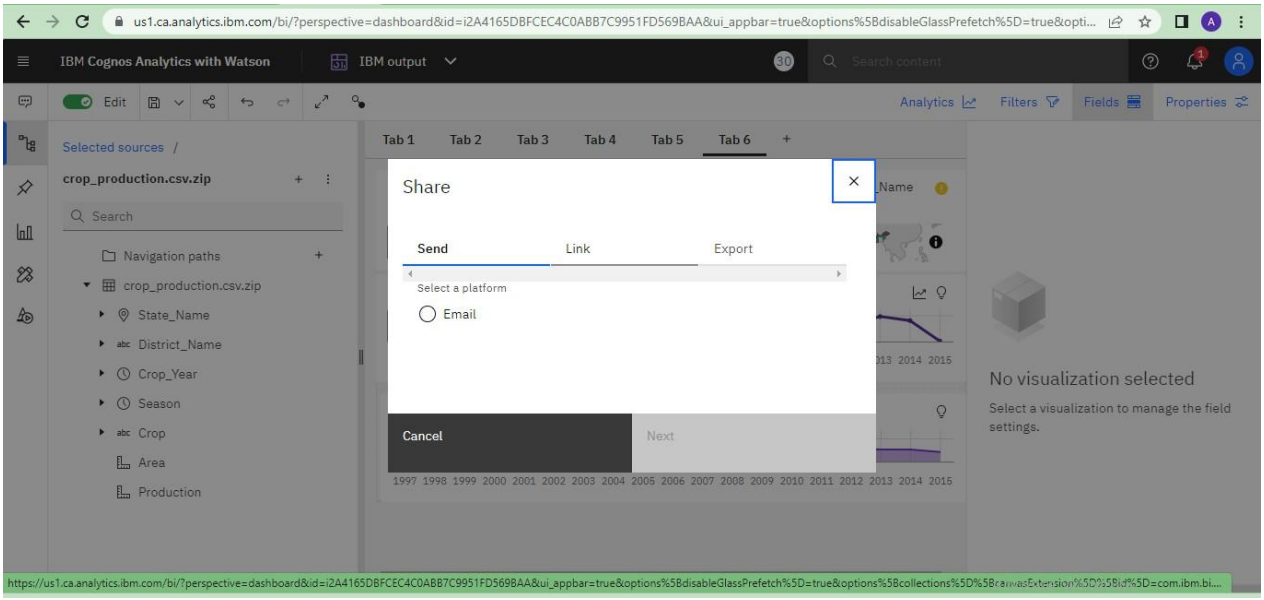
Export Analysis:

Link Generation:

- After creating the dashboard,click on the share option from the icons given below.



- Therefore opens the share window,click on the link option from the menu.



- There generates the link,just copy and paste the generated link.

9. ADVANTAGES & DISADVANTAGES:

Advantages:

- One can easily analyse and understand trends in croppping pattern, seasonal behaviour of land in various areas with the created dashboard.
- With no prior skills and knowledge about the tools that we use for analysis,anyone (literate or illiterate)can easily infer the knowlegde that we represent in various charts or graphs or maps.
- So that it would be helpful to farmers to make appropriate decisions in the future.

Disadvantages:

- Not all factors influencing the crop yield are being considered for the analyis as we have only taken visible factors into account for the analysis.

10. CONCLUSION:

The productivity of agriculture has slightly increased as a result of technology's introduction. New ideas like digital agriculture, smart farming, precision agriculture, etc. have been made possible by the innovations. From the analysis dashboard, it has been noted that analyses of agricultural productivity and the detection of hidden patterns utilising data sets related to seasons and crop yields have been conducted.Using IBM Cognos, we have observed and conducted analysis on various crops grown, area, and productions in various states and districts, including

- Seasons with average productions. We learn from these analytics which seasons have higher average production and which have lower production.
- Production split up per crop year. We learn from this study which years have high and low production.
- District-based production. With the help of these analytics, we may identify the states and districts that farm the chosen crops.
- Production by area.This will allow us to estimate the yield and determine how much land needs to be planted. After creating the dashboard, study was done to determine which state, which year, and how much crop area will be produced.

11. FUTURE SCOPE:

- Farming is the means of survival as humans require food that is obtained only through farming directly or indirectly.

- With the growing human population, it is critcal to analyse the production in farming every year.
- So, that we can know the right time,right place and right crop to be cultivated considering all the factors that influence the crop production.
- 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.

12. APPENDIX:

Source Code: IBM Cognos Dashboard

GitHub Link: IBM-EPBL/**IBM-Project-54792-1662456891**

Project Demo link: https://drive.google.com/file/d/1uIvLCPYwFZFLtR3q-ajtrT55JGB4d_0B/view?usp=drivesdk