ESTIMATE THE CROP YIELD USING DATA ANALYTICS

NALAIYA THIRAN PROJECT BASED LEARNING

on

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTERPRENEURSHIP

A PROJECT REPORT

GOKUL M	**727619BIT018**
SUSMITHA V G	**727619BIT004**
RACHNA K	**727619BIT024**
VARUNVIKASH K	**727619BIT028**

in partial fulfillment for the award of the degree

of

Bachelor of Technology

in

Information Technology

Dr. Mahalingam College of Engineering and Technology

Pollachi - 642003

An Autonomous Institution

Affiliated to Anna University, Chennai - 600 025

NOVEMBER 2022

INTERNAL MENTOR

Dr. J. RAMPRASATH AP/IT,
ASSISTANT PROFESSOR,
Department of Information Technology,
Dr. Mahalingam College of Engineering and Technology.

INDUSTRY MENTORS

Srikanth,
Mahidhar,
Mohammed Azhar Uddin

ABSTRACT

Estimate the crop yield using data analytics is an technique of approaching the futuristic yield and the results are processed and extracted from the previous year's yield with the same due of climatic changes, water, seasonal approach, the area of yield etc,.The impact of climate change in India, most of the agricultural crops are being badly affected in terms of their performance over a period of the last two decades.

Predicting the crop yield in advance of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and thus help them to make the appropriate decisions.

Preparing for the event of business an farmer can literally estimate the quantity of yield he/she can get from tha area of crop investing and thus can calculate the sum of banknotes he can earn.

TABLE OF CONTENTS

1 - INTRODUCTION

1.1 Project Overview

India is a predominantly agricultural nation. Agriculture is currently the most significant emerging sector in the actual world and the key industry and economic pillar of our nation. The discipline of agricultural information technology has recently undergone significant changes that have made crop yield prediction an interesting research topic. Crop yield prediction is a technique for estimating crop yield using many characteristics, including temperature, rainfall, fertilisers, insecticides, and other climatic variables and parameters.Using data analytics to analyse those parameters and provide the patterns or trends that has been followed over the past years in estimating the yield can help farmers to make right chioce in the selection of crop varieties,etc.,To make people gasp and use the knowledge represented we finally put the visualisations made in a dashboard and represent it with the most suitable and appropriate charts or graphs or maps.

1.2 Purpose

Analytics is the interpretation of data pattern that assist decision- making and performance improvement. Agriculture Data analytics in crop yield helps in analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

2 - LITERATURE SURVEY

2.1 Existing problem

Crop Yield Prediction Using Machine Learning : A research group investigated the utilization of various information mining methods which will foresee rice crop yield for the data collected from the state of Maharashtra, India. A total of 27 regions of Maharashtra were selected for the assessment and the data was collected related to the principle rice crop yield influencing parameters such as different atmospheric conditions and various harvest parameters i.e Precipitation rate, minimum, average, maximum and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield for the season between June to November referred as Kharif, for the years 1998 to 2002 from the open source, Indian Administration records. WEKA a Java based dialect programming for less challenging assistance with information data sets, assigning design outcomes tool was applied for dataset processing andthe overall methodology of the study includes,

(1) pre-processing of dataset

(2) Building the prediction model utilizing WEKA and

(3) Analyzing the outcomes. Cross validation study is carried out to scrutinize how a predictable information mining method will execute on an ambiguous dataset. Study applied 10-fold higher cross validation study design to assess the data subsets for screening and testing. Identified and collected information was randomly distributed into 10 sections where in one data section was used for testing while all other data sections were utilized for the preparation information. Study reported that themethod applied was supportive in the precise estimation of rice crop yield for the state.

[1] In the year (2018) has proposed a random Forest Algorithm for predicting the crop yield of particular area considering various parameters such as rainfall, seasonal crop (Rabi and Kharif) district-wise, temperature (max.), crop production in terms of Kgs/tonnes. Area for doing research was Tamil Nadu. Dataset record were collected from Indian Government over 15years for rice production. They proved in experimental results that prediction analysis done using Random Forest Algorithm – a supervised machine learning algorithm will help farmer to

predict the yield of the crop before cultivating onto the agricultural field. This algorithm run efficiently on large databases with high classification accuracy.

[2] Crop Yield Prediction Using Data Mining Techniques : Raor ane A.A. and Kul karn i R.V., discussed few data mining techniques in their paper. They concluded that efficient technique can be developed and analyzed using the appropriate data, to solve complex agricultural problems using data mining techniques. Also recommend some of the algorithms and statistical methods that give [8]good results in agriculture growth.
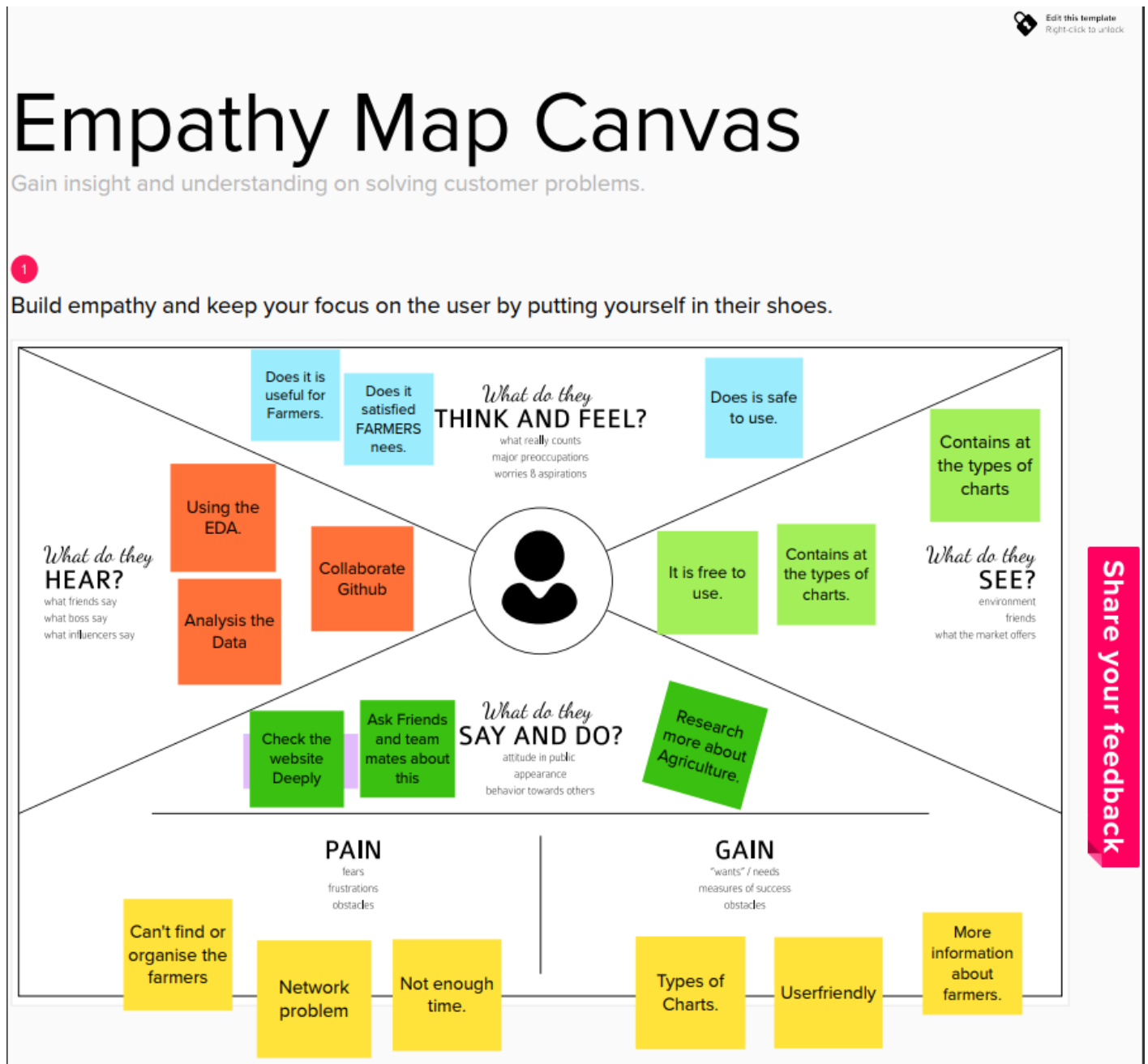
[3] Crop yield prediction using Big Data Analytics: In India crop yield is season dependent and majorly influenced by the biological and economic causes of an individual crop. Reporting of progressive agricultural yield in all the seasons is an ample task and an advantageous task for every nation with 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 dueto 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 suchas 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 S. Athmaja, M. Hanumanthappa, and V. Kavitha,a survey of machine learning algorithms has presented effective strategies by for big data analytics. All over the world the agriculturalpeoples gained some advantages through the comparative knowledge from big data analysis, with machine learning algorithm by using huge data the agricultural peoples get some comparative knowledge and changes in regular agriculture

2.2 Problem Statement Definition
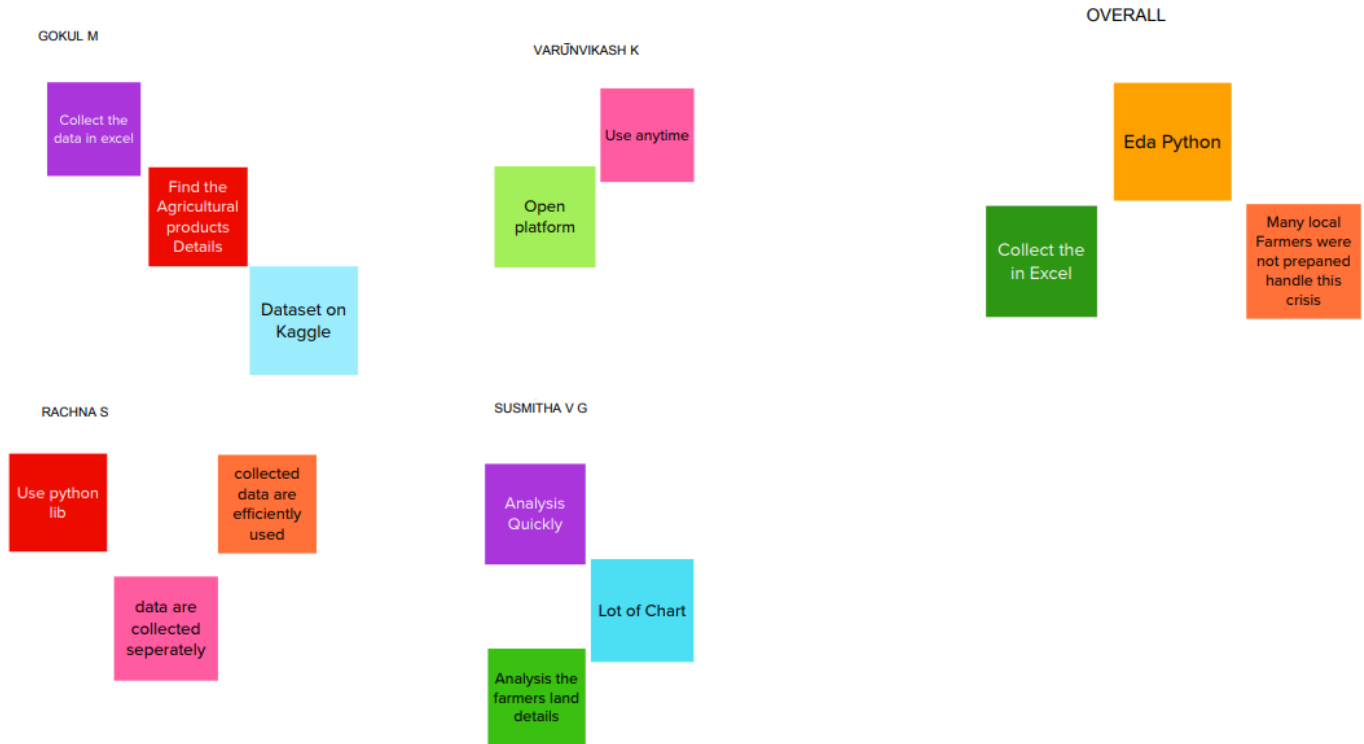
In the agriculture sector the farmers are facing difficulties in analysing the demand in market and soil quality analysis to achieve high crop yield through technology. The main objective of this project is to predict crop yield that will be extremely useful to farmers to plan for the harvest and sales of harvested grain.

3 - IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation and Brainstroming

GOKUL M

Collect the data in excel

Find the Agricultural products Details

Dataset on Kaggle

VARUNVIKASH K

Use anytime

Open platform

OVERALL

Eda Python

Collect the in Excel

Many local Farmers were not prepaned handle this crisis

RACHNA S

Use python lib

collected data are efficiently used

data are collected seperately

SUSMITHA V G

Analysis Quickly

Lot of Chart

Analysis the farmers land details

3.3 Proposed Solution

S.NO	PARAMETER	DESCRIPTION
1	Problem Statement (Problem to be solved)	The main objective of this project is to predict crop yield that will be extremely useful to farmers to plan for the harvest and sales of harvested grain.
2	Idea / Solution description	Provide perfect data report after deep analysis of the past data. Helping them out to overcome loss in farming and business.
3	Novelty / Uniqueness	With this solution we can analyse, visualize data and give the farmers the option to choose which plant/crop to cultivate in which period of time/season to earn more profit from the crop yield
4	Social Impact / Customer Satisfaction	Perfect data visuals create a large impact in the crop yield. And hence farmers will be able to gain more profit

5	Business Model (Revenue Model)	We can increase/enhance crop production and other raw materials. Also, Increase in productivity will result in increase of Revenue for the farmers.
6	Scalability of the Solution	With the data visual reports, farmers will be able to cultivate crop according to the area, climate, soil and other features that impact the crop yield and hence enhancing the productivity

3.4 Problem Solution Fit

Define CS, fit into CC

1.CUSTOMER SEGMENT(S)
CS

Farmers are the customer who wants to yield a crop in field.

6. CUSTOMER CONSTRAINTS
CC

Less knowledge and development towards the current environmental changes and technologies, they follow ancient methods, which is also worthy but, the climatic changes and new kind pesticides

5. AVAILABLE SOLUTION
AS

- Traditional ways of prediction.
- Precision farming.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS
J&P

- Help them understand the usage of prediction and software application for good results in agriculture.
- Data report should to be created to reduce the loss of the crop and earn more profit in agriculture fields

9. PROBLEM ROOT CAUSE   RC

- Various disease on the plants can lead to reducing the quality of the crops productivity.
- The insects on the plants can spread the disease.

7. BEHAVIOUR   BE

- Try to get help from agricultural experts.
- Try to take up non-natural means of cultivation for quicker harvest

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS
TR
- Seeing their crops are being infected by disease and facing huge loss in quality.

4. EMOTIONS: BEFORE / AFTER
EM

Before: Most of the famers in India have Stress, Loosing Self Confidence.

After : Gain of Self Confidence.

10. YOUR SOLUTION   SL

- The solution for the problem, creating data report using past datasets.
- Creating IBM Cognos dashboard could make them better understand easily.

8.CHANNELS of BEHAVIOUR   CH

- Trying to use pesticides and fertilizers that increase gain but cause harm.
- Irrigation channel changes.

4 - REQUIREMENT ANALYSIS

4.1 Functional Requirements

| REQUIREMENT	DESCRIPTION
Registration	User can create an account by registering through form
Google authentication	Add new user with a pre-existing google account
Login	Easier login for already existing users
Take in the required data	Input data is taken from the user for data prediction
Estimation/ Prediction	A prediction of crop yield is done based on the user input
Analysis	An analysis is done on the given data to gain useful insights on the crop yield

4.2 Non Functional Requirements

REQUIREMENT	DESCRIPTION
Performance	The software should provide us good performance
Reliability	The UI should be user friendly and easily understandable
Availability	It should be available for access at any time from anywhere
Scability	The software should be scalable for larger datasets
Security	The user login and registration is in secured manner

5 - PROJECT DESIGN

5.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditonal visual represenation of the information flows within a system. A neat and clear DFD can depcit 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.
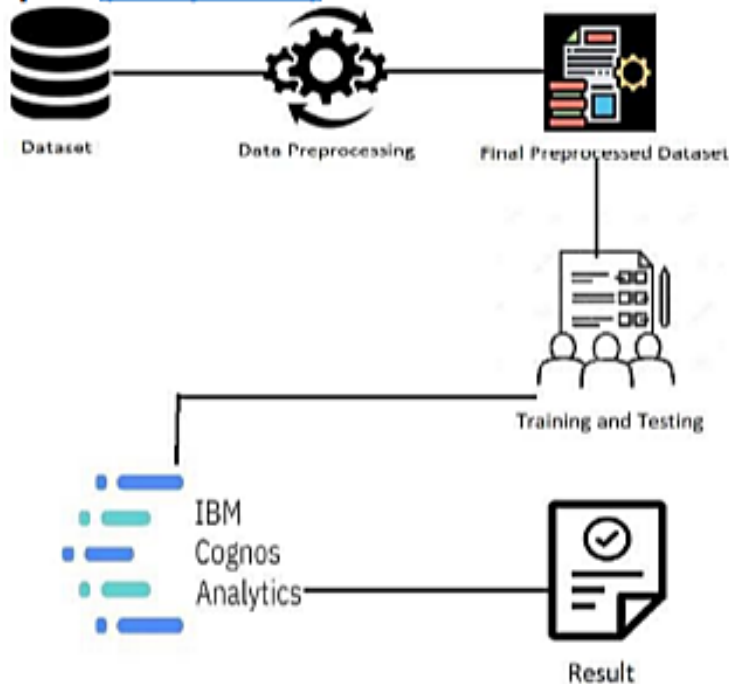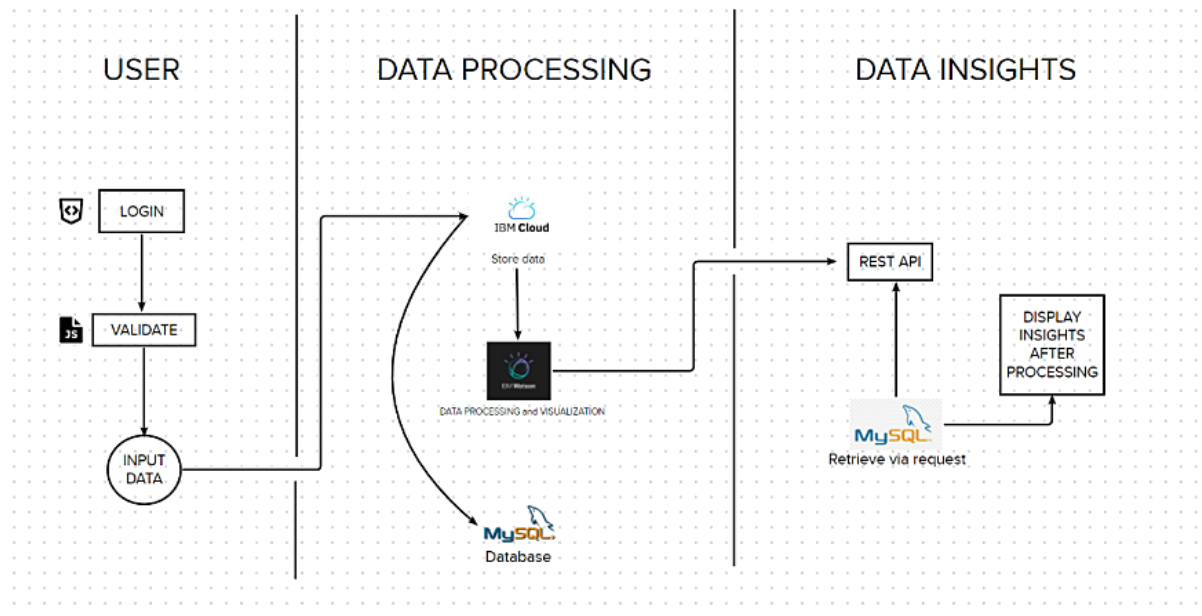
5.2 Solution and technical Architecture

It shows the technologies that are used in the work flow of the problem.

Technical Architecture:

The Technologies that are used in the development process of the project

5.3 User Stories

Functional requirement	User Story/Number	User Story/task	Story Points	Priority	Team Members
Registration	USN - 1`	As a user, I can register for by entering my Agri - id card and request..	2	High	GOKUL, SUSMITHA
	USN - 2	As a user, I can register for the application through Gmail	2	Medium	RACHNA, VARUN
Login	USN - 3	As a user, I can Call and request or Approach for dataset	2	High	GOKUL, VARUN
Working with dataset	USN - 4	To work on the given dataset, Understand the	4	High	SUSMITHA, RACHNA

		Dataset.			
	USN - 5	Load the dataset to Cloud platform then Build the required Visualizations	10	High	GOKUL, RACHNA
Data Visualization Chart	USN - 6	Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations. Build a Visualization to showcase Average Crop Production by Seasons.	4	Medium	VARUN, SUSMITHA
		Showcase the Yearly usage of Area in Crop Production.	4	Medium	SUSMITHA, GOKUL
	USN - 7	Build a visualization to show case top 10 States in Crop Yield Production by Area	4	Medium	RACHNA, VARUN
		Build the required Visualization to showcase the Crop Production by State.	4	Medium	RACHNA, GOKUL
		Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation.	4	Medium	SUSMITHA, VARUN
Creating the dashboard	USN - 8	Create the Dashboard by using the created visualizations	20	High	GOKUL, SUSMITHA
Export the Analytics	USN - 9	Export the created Dashboard	20	High	RACHNA, VARUN

6 - PROJECT PLANNING AND SCHEDULING

6.1 Sprint Planning and Estimation

Sprint	Functional Requirements	user story/ Number	User Story/ Task	Story points	Priority	Team Members
Sprint -1	Registration	USN - 1	As a user, I can register for by entering my Agri - id card and request.	2	High	GOKUL, SUSMITHA
Sprint - 1		USN - 2	As a user, I can register for the application through Gmail	2	Medium	RACHNA, VARUN
Sprint - 1	Login	USN - 3	As a user, I can Call and request or Approach for dataset	2	High	GOKUL, VARUN
Sprint - 1	working with Dataset	USN - 4	To work on the given dataset, Understand the Dataset.	4	High	SUSMITHA, RACHNA
		USN - 5	Load the dataset to Cloud platform then Build the required Visualization	10	High	GOKUL, RACHNA
Sprint - 2	Data Visualization chart	USN - 6	Using the Crop production in Indian dataset, create various graphs and charts to highlight the insights and visualizations. Build a Visualization to showcase Average Crop Production by Seasons.	4	Medium	VARUN, SUSMITHA
			Showcase the Yearly usage of Area in Crop Production	4	Medium	SUSMITHA, GOKUL
			Build a visualization to show case top 10	4	Medium	RACHNA, VARUN

			States in Crop Yield Production by Area			
			Build the required Visualization to showcase the Crop Production by State.	4	Medium	RACHNA, GOKUL
			Build Visual analytics to represent the Sates with Seasonal Crop Production using a Text representation	4	Medium	SUSMITHA, VARUN
Sprint - 3	Creating the dashboard	USN - 8	Create the Dashboard by using the created visualizations.	20	High	GOKUL, SUSMITHA
Sprint - 4	Export the Analytics	USN - 9	Export the created Dashboard	20	High	RACHNA, VARUN

6.2 Sprint Delivery Schedule

Sprint	Total Story points	Duration	Sprint Start Date	Sprint End Date	Story Points Completed	Sprint Release (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

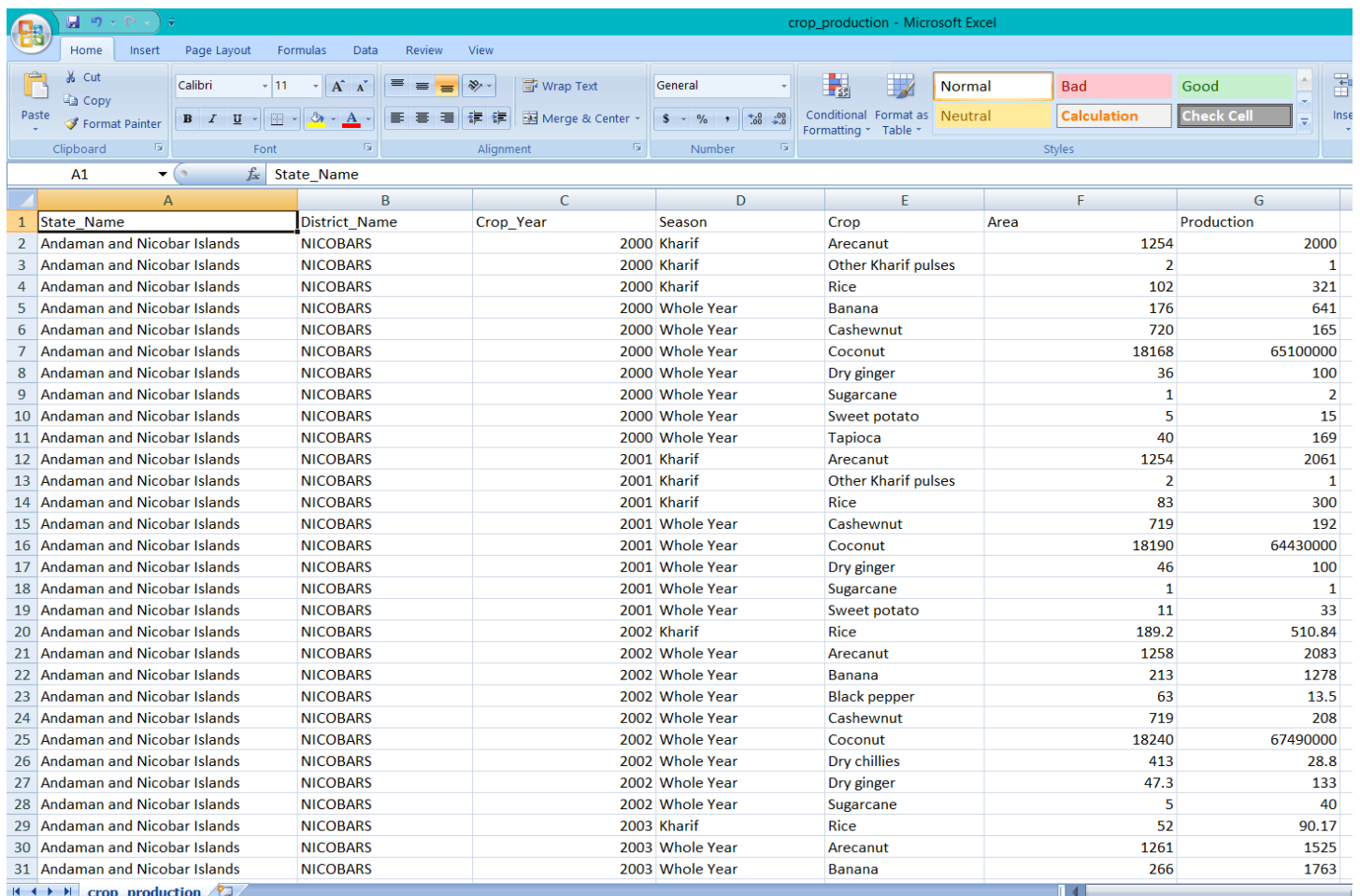# 7 - WORKING WITH THE DATASET AND DATA VISUALIZATION

7.1 Understanding the Dataset

This project is based on a understanding the crop production of India.It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

Let's understand the data we're working with and give a brief overview of what each feature represents or should represent
1. State Name - All the Indian State names
2. District Name -Different District names
3. Crop Year- contains the crop years
4. Season – Different seasons for crop production
5. Area- Total number of areas covered
6. Production- production of crops.

The data format is as shown in the below image

|	A	B	C	D	E	F	G
1	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254	2000
3	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
4	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321
5	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176	641
6	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720	165
7	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168	65100000
8	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36	100
9	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1	2
10	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5	15
11	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40	169
12	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254	2061
13	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2	1
14	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83	300
15	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Cashewnut	719	192
16	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Coconut	18190	64430000
17	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Dry ginger	46	100
18	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sugarcane	1	1
19	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sweet potato	11	33
20	Andaman and Nicobar Islands	NICOBARS	2002	Kharif	Rice	189.2	510.84
21	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Arecanut	1258	2083
22	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Banana	213	1278
23	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Black pepper	63	13.5
24	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Cashewnut	719	208
25	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Coconut	18240	67490000
26	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Dry chillies	413	28.8
27	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Dry ginger	47.3	133
28	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Sugarcane	5	40
29	Andaman and Nicobar Islands	NICOBARS	2003	Kharif	Rice	52	90.17
30	Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Arecanut	1261	1525
31	Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Banana	266	1763

7.2 Loading the Dataset
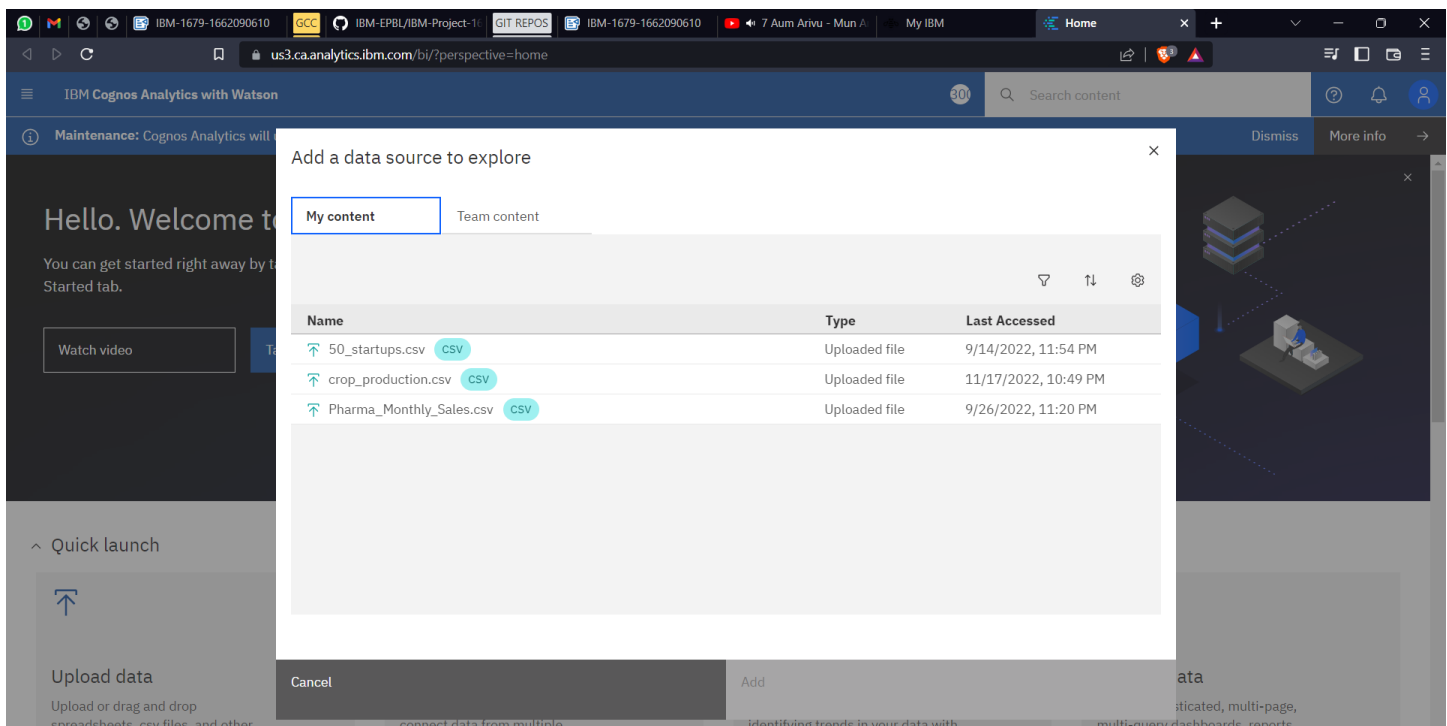
- Open the menu dialog box
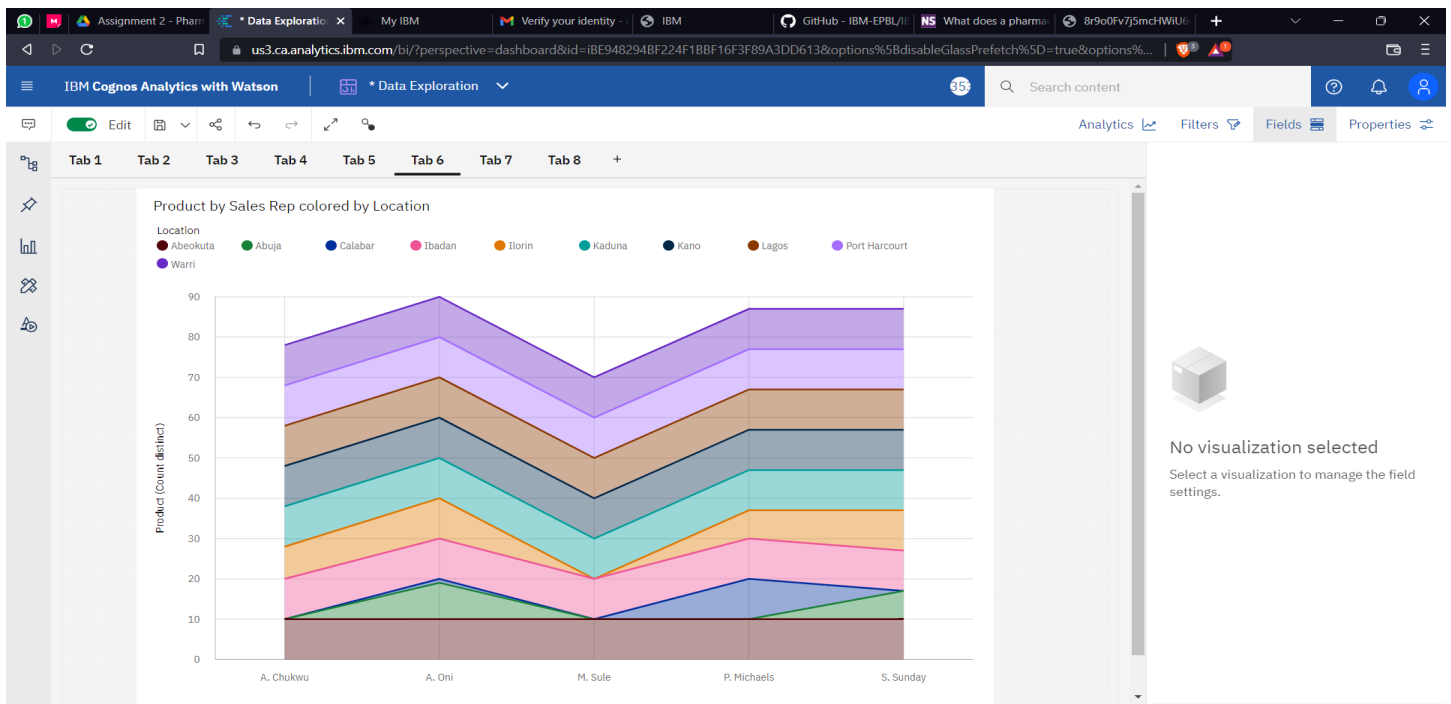- Then click on upload data and then the uploaded dataset can be viewed here
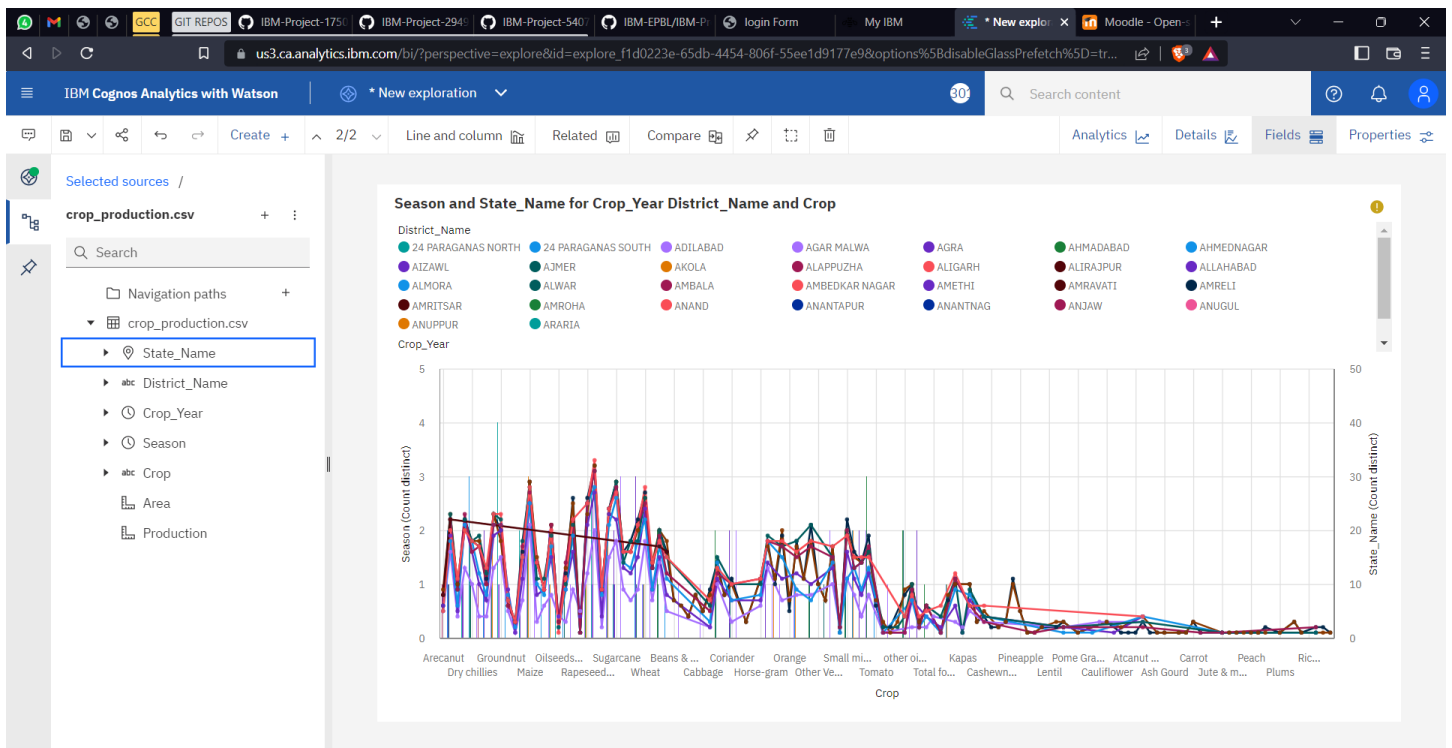
7.3 Visualization charts
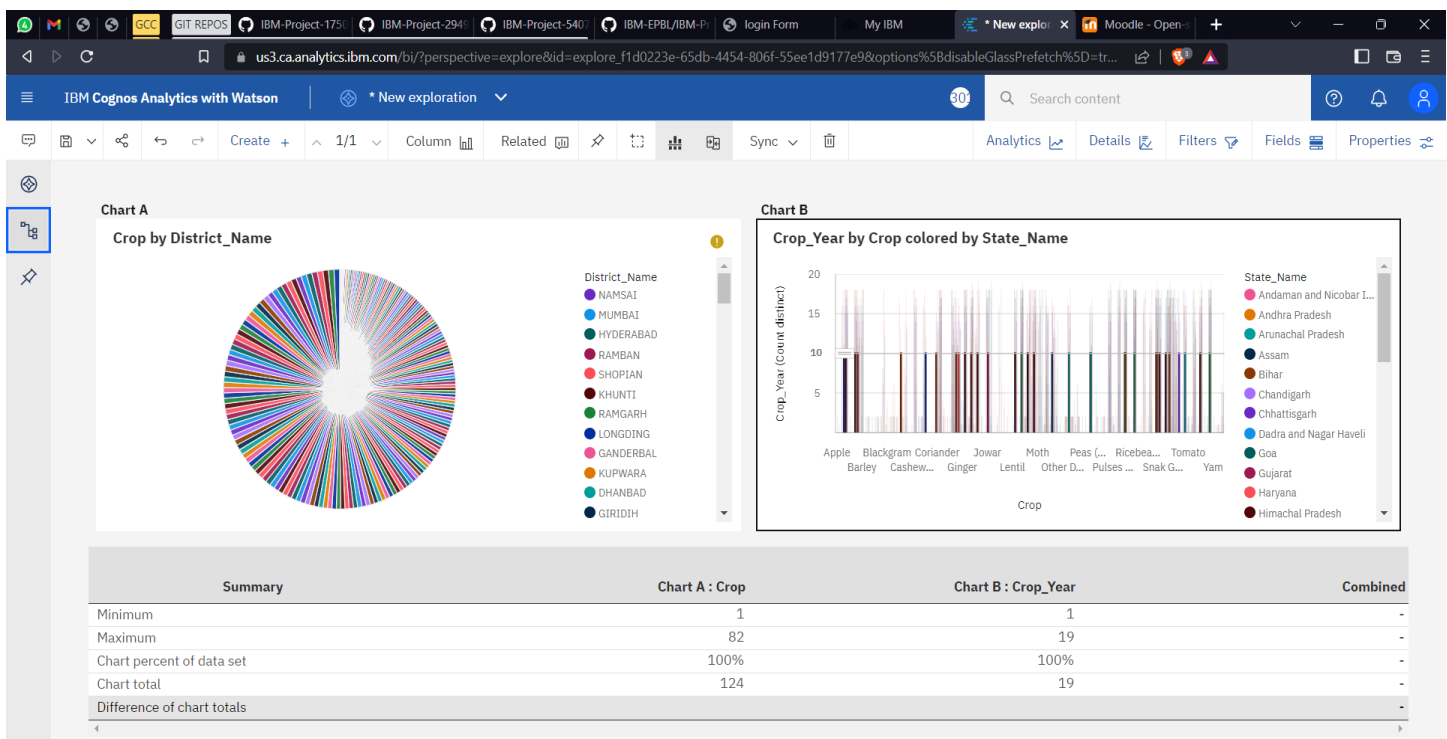
- CHART A - There are two charts displaying the Crop production by district name.
- CHART B - It shows the crop growth year according to the states where they grow.
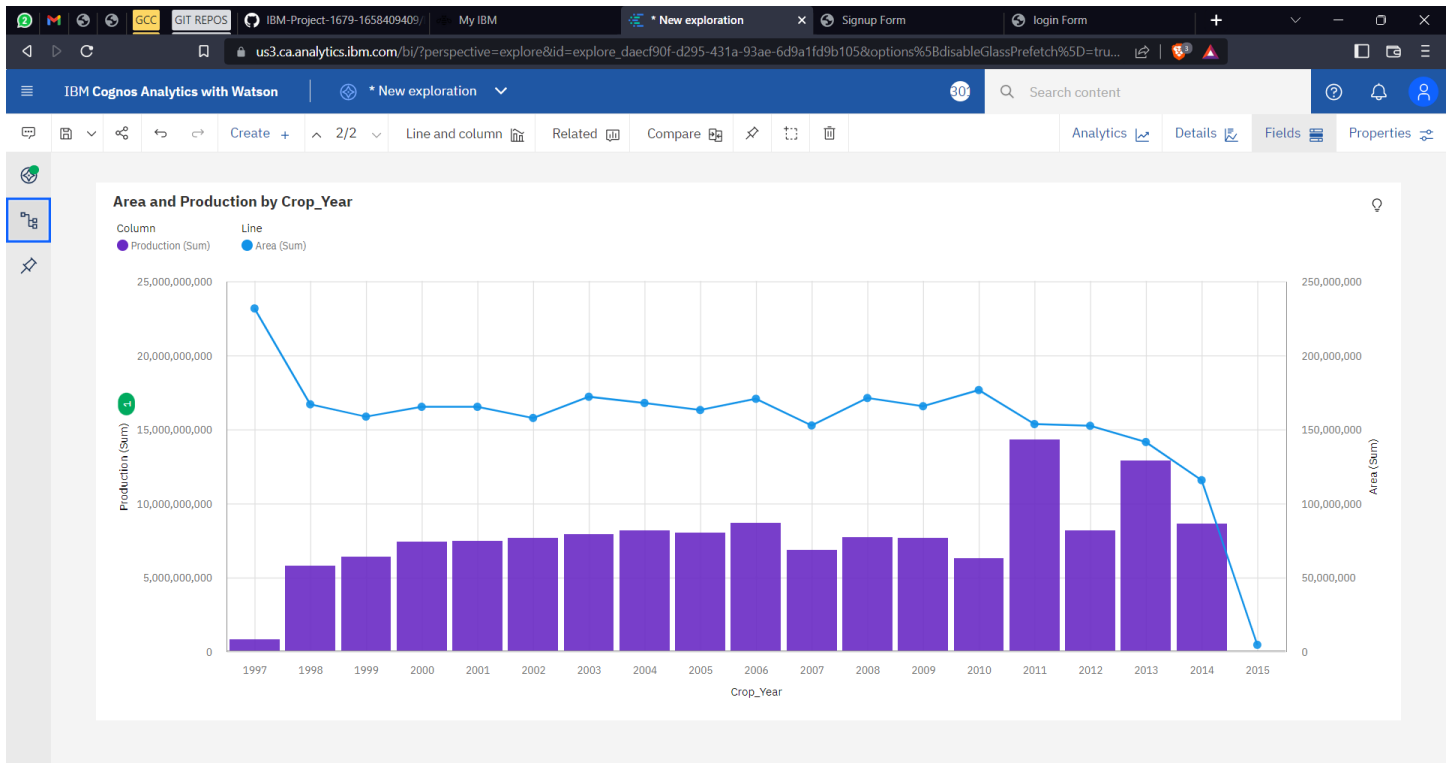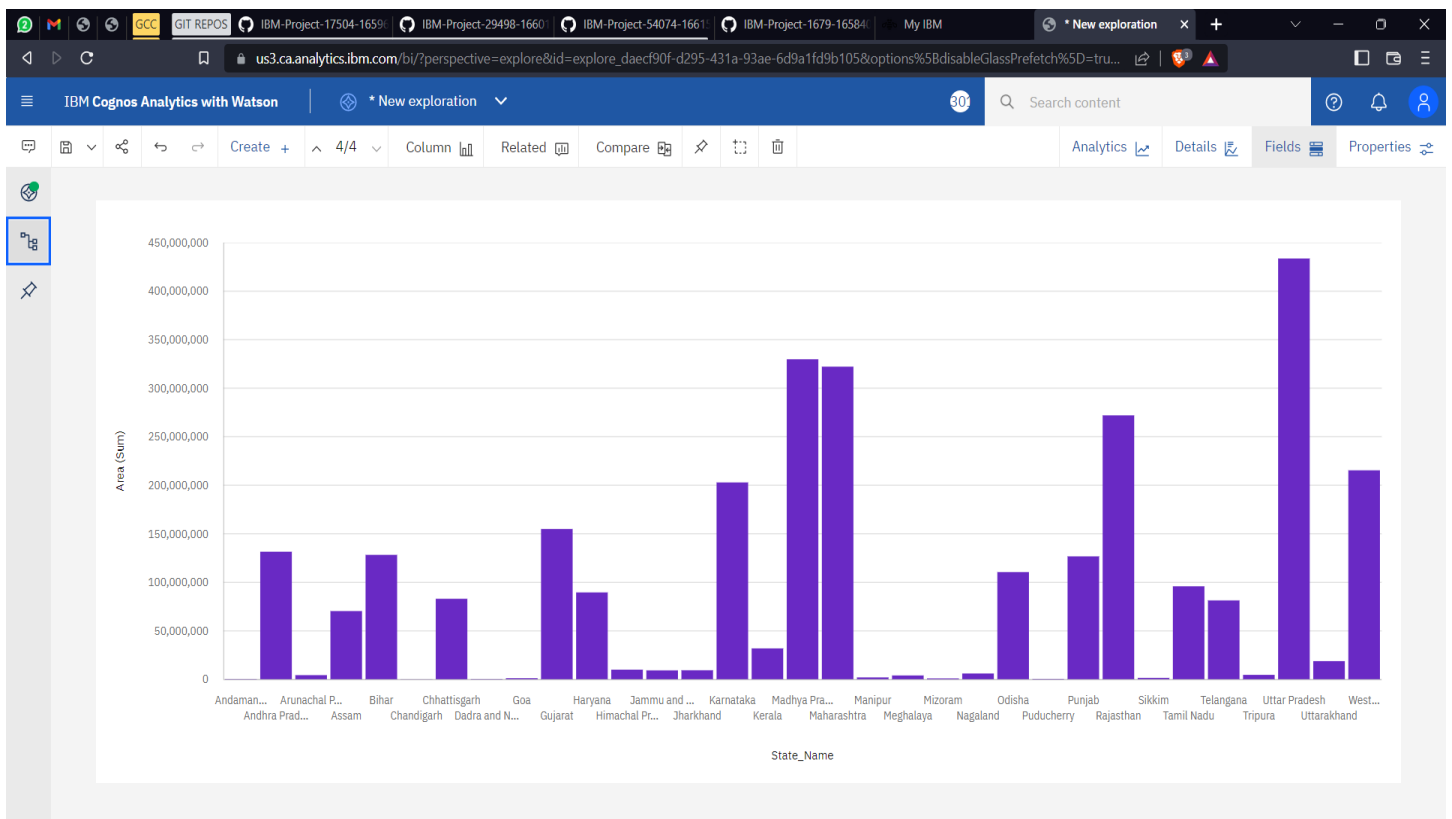

- It shows the Area and Production of the crop by the year
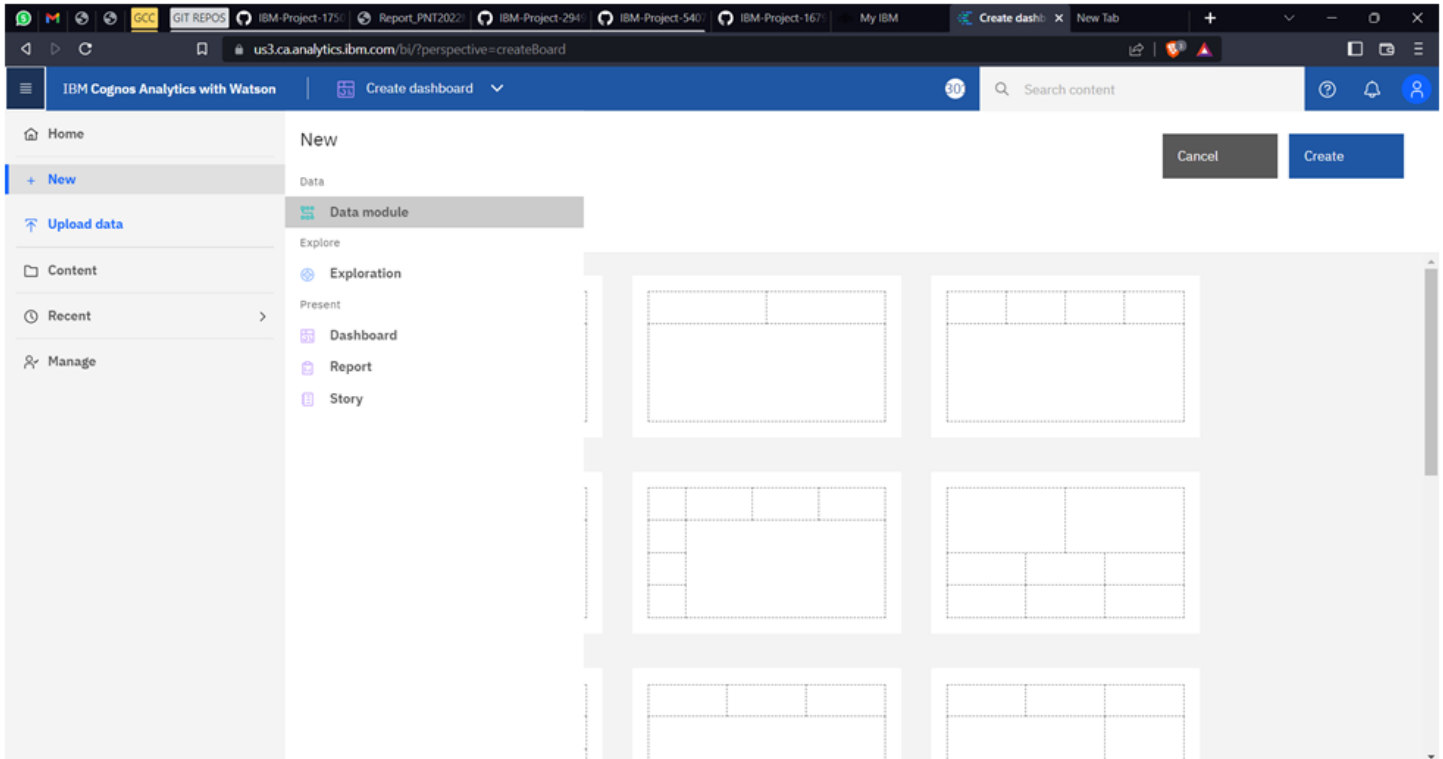

- It displayes the product (crop) by sales

- The season and the statename for the year with good growth of the crop

8 - CREATING THE DASHBOARD AND EXPORT THE ANALYTICS

8.1 Creating the Dashboard

* Login into IBM COGNOS and then open menu and click NEWàDASHBOARD
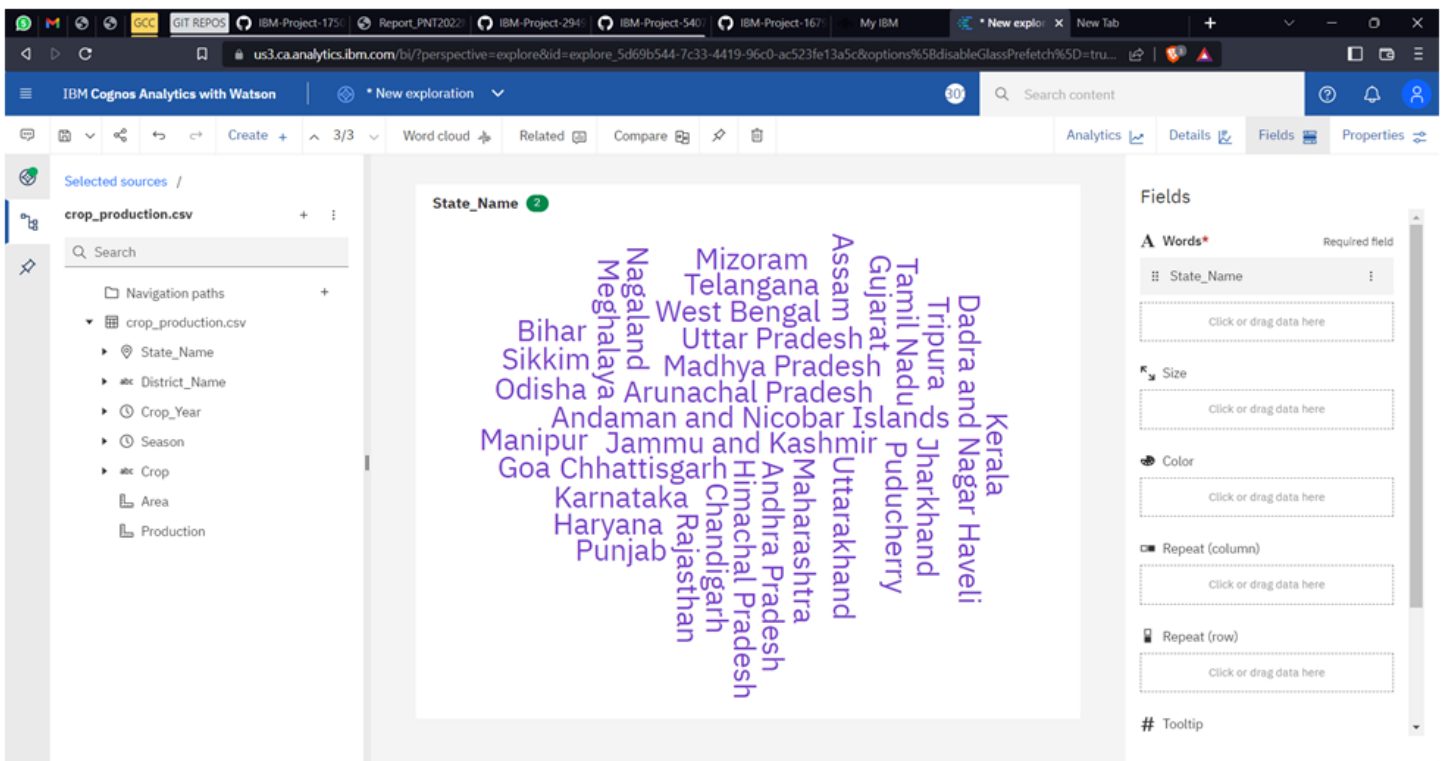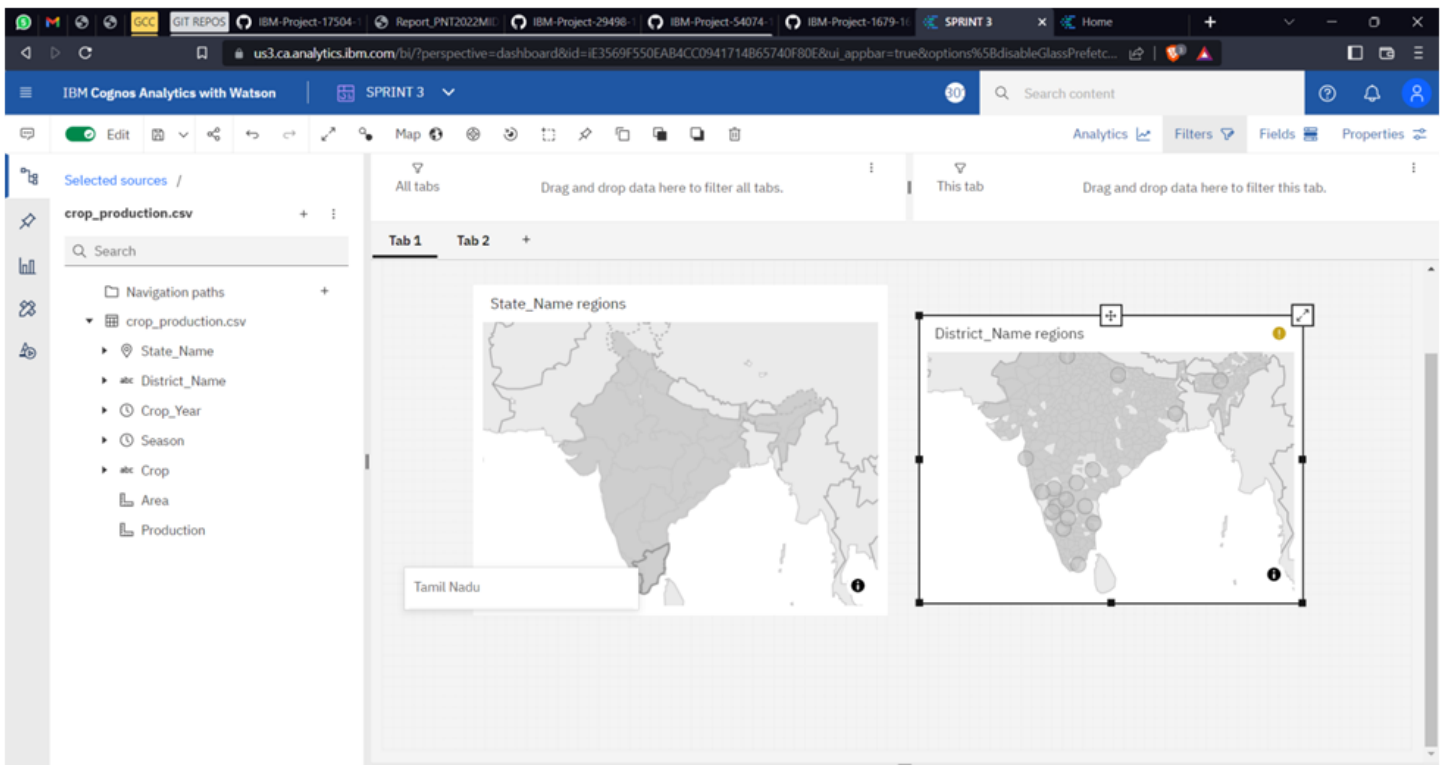

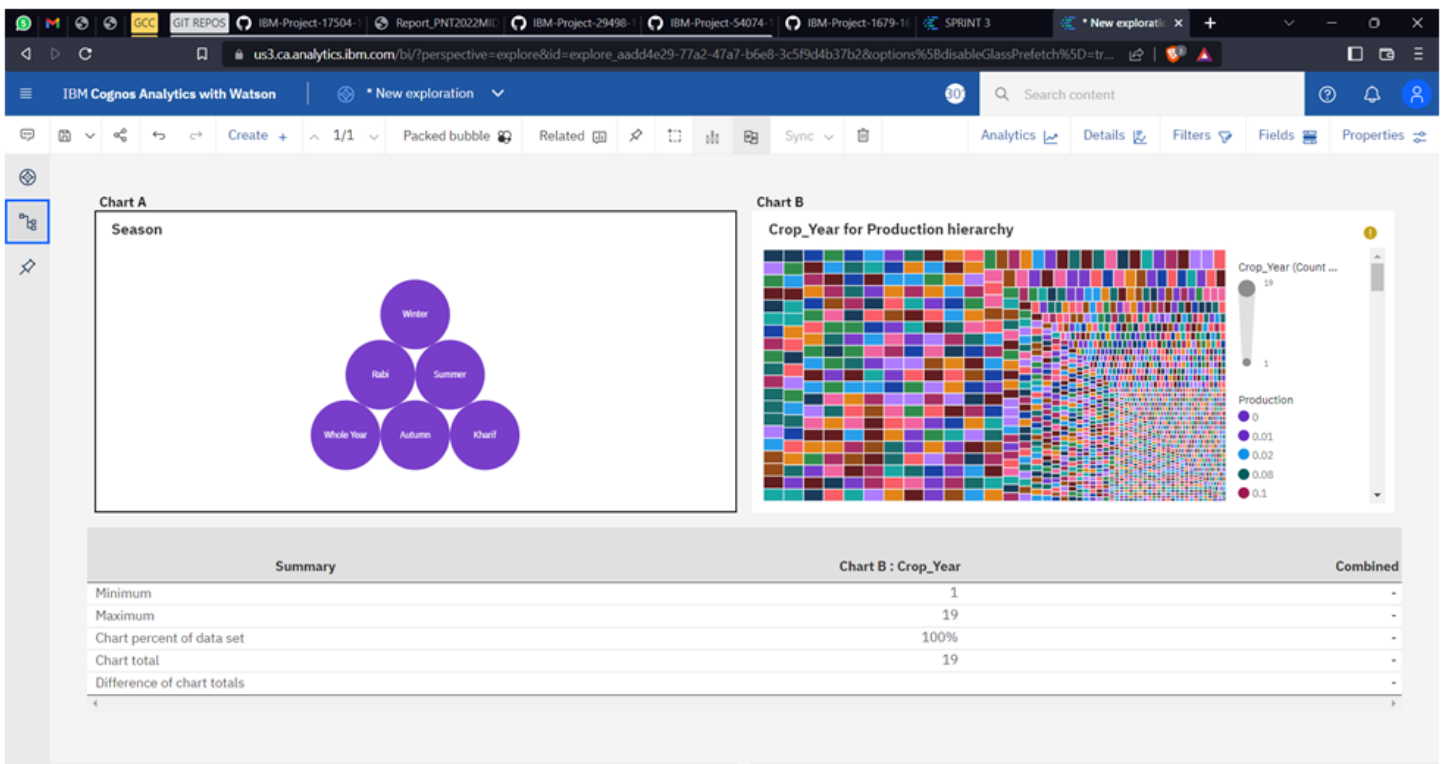* After clicking on new dashboard select the type of template required

* Then select the **crop_production.csv** file dataset under my content (previously uploaded for visualization and Exploration) and click add
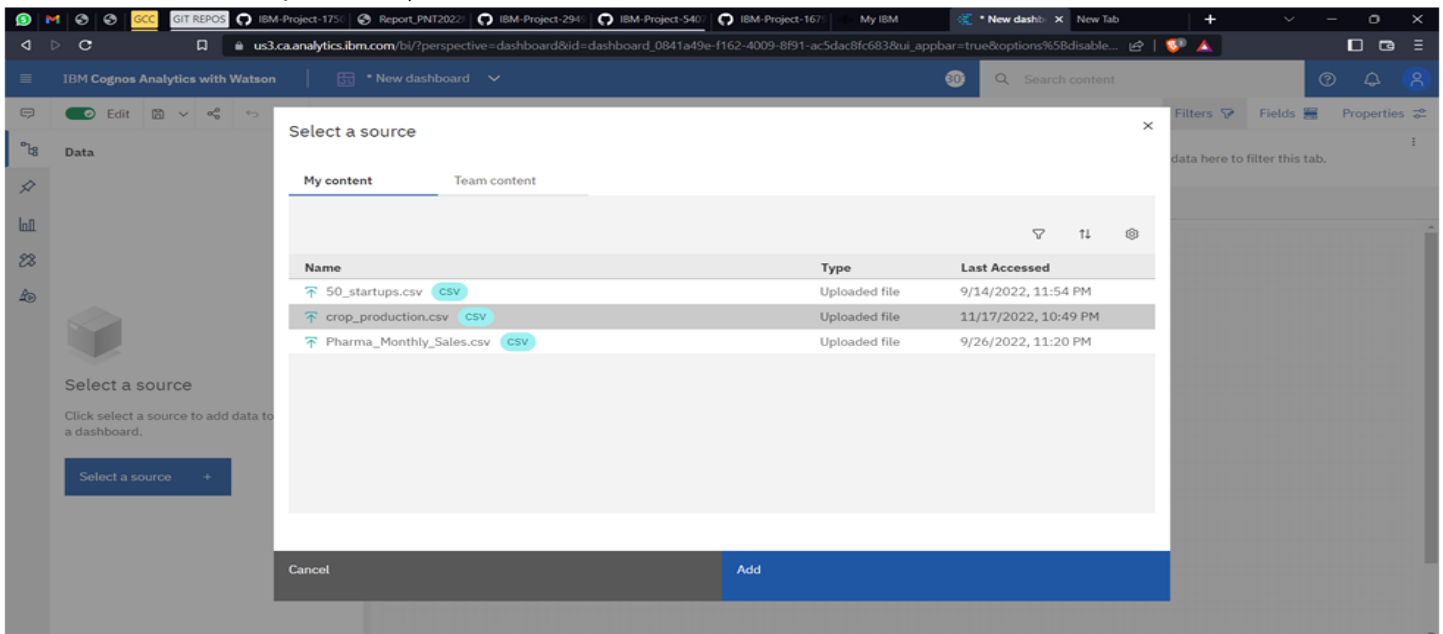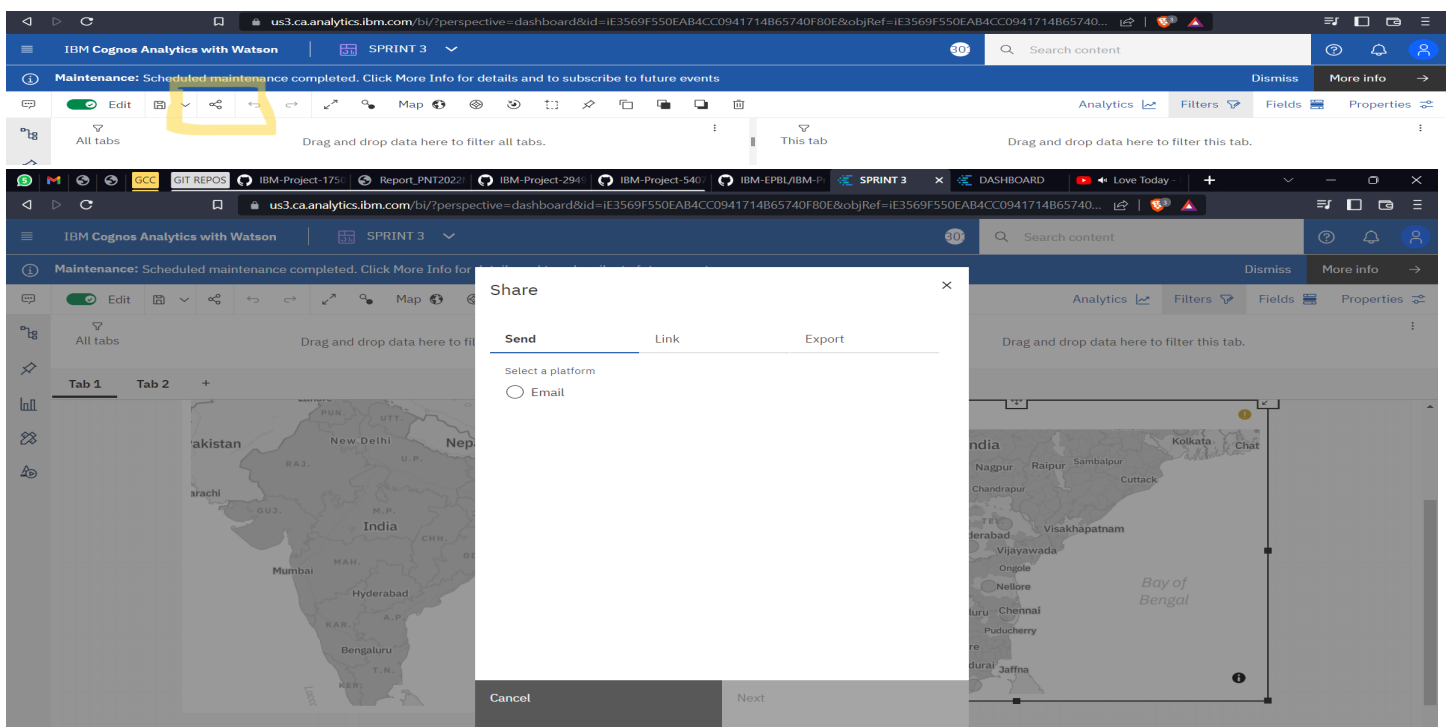
* Rename the tab title as sprint 3 / Dashboard
* Copy the Each Visualization and paste it in the tab and arrange the chart

8.2 Exporting the Analytics

- Login into IBM COGNOS and then open Menu
- Then open your dashboard and click on the share icon
- We can use E-mail or Generate an link or Export that as an pdf

9 - ADVANTAGES AND 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

1) Seasons with average productions. We learn from these analytics which seasons have higher average production and which have lower production.

2)Production split up per crop year. We learn from this study which years have high and low production.

3) District-based production. With the help of these analytics, we may identify the states and districts that farm the chosen crops.

4) 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

12.1 Source code

My Dashboard Sprint for crop yield

12.2 Github LInk and project demo Link

https://github.com/IBM-EPBL/IBM-Project-1679-1658409409