# ESTIMATE THE CROP YIELD USING DATA ANALYTICS

**NALAI**

### NALAIYA THIRAN PROJECT BASED LEARNING

on

### PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND

**ENTERPRENEURSHIP A PROJECT REPORT**

### AZHAGAR R 820519205301

**KALAIMARAN R 820519205019**

### VETRIVEL S 820519205041

**SRI BHARATH K 820519205302**

**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

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO**  **1** | **TITLE**  **ABSTRACT**  **INTRODCUTION** | **PAGE NO**  **3** |
|  | **1.1 Project Overview** | **6** |
| **2** | **1.2 Purpose**  **LITERATURE SURVEY**  **2.1 Existing Problem** | **6**  **7** |
|  | **2.2 References** | **8** |
| **3** | **2.3 Problem Statement**  **IDEATION & PROPOSED SOLUTION**  **3.1 Empathy Map** | **8**  **9** |
|  | **3.2 Ideation and Brainstrom** | **10** |
|  | **3.3 Proposed Solution** | **10** |
| **4** | **3.4 Problem Solution Fit REQUIREMENT ANALYSIS**  **4.1 Functional Requirements** | **11**  **12** |
| **5** | **4.2 Non Functional Requirements PROJECT DESIGN**  **5.1 Data Flow Diagram** | **12**  **13** |
|  | **5.2 Solution and Technical Architecture** | **14** |
|  | **5.3 User Stories** | **14** |

1. **PROJECT PLANNING AND SCHEDULING**
   1. [Sprint planning and Estimation 16](#_TOC_250008)
   2. Sprint Delivery Plan 17
2. WORKING WITH DATASET & DATA VISUALIZATION
   1. [Understanding the Dataset 18](#_TOC_250007)
   2. [Loading the Dataset 19](#_TOC_250006)
3. CREATING AND EXPORTING DASHBOARD
   1. [Creating the Dashboard 22](#_TOC_250005)
   2. Export the Analytics 24
4. [ADVANTAGES AND DISADVANTAGES 25](#_TOC_250004)
5. [CONCLUSION 25](#_TOC_250003)
6. [FUTURE SCOPE 26](#_TOC_250002)
7. [APPENDIX](#_TOC_250001)
   1. [Source Code 26](#_TOC_250000)
   2. Github and Project Demo Link 26

## 1 - INTRODUCTION

* 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.

## 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

* 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.
4. 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.

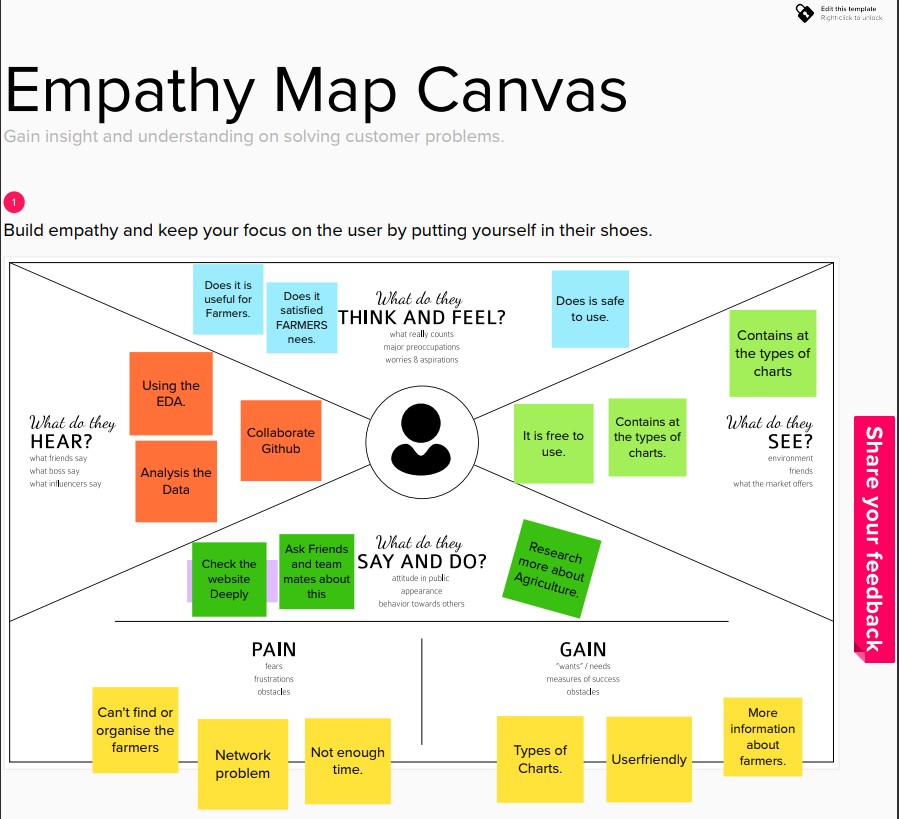
1. Crop Yield Prediction Using Data Mining Techniques : Raor ane A.A. and Kul karn i R.V., discussed few data mining techniques in their paper. They concluded that eﬃcient technique can be developed and analyzed using the appropriate data, to solve complex agricultural problems using data mining techniques. Also recommend some of the algorithms and statistical methods that give [8]good results in agriculture growth.
2. 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

## Problem Statement Deﬁnition

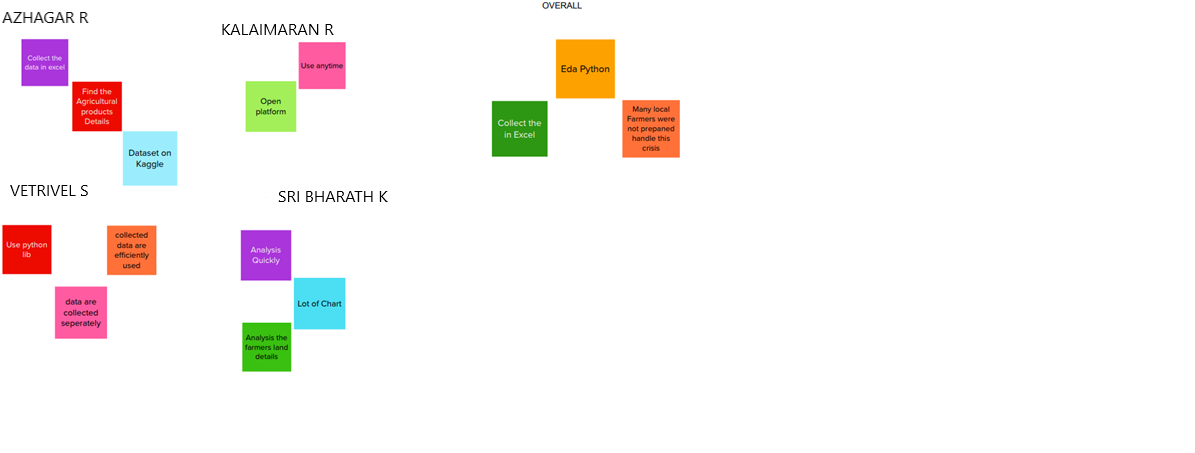
In the agriculture sector the farmers are facing diﬃculties in analysing the demand in market and soil quality analysis to achieve high crop yield through technology. The main objective of this project is to predict crop yield that will be extremely useful to farmers to plan for the harvest and sales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

### 3 - IDEATION AND PROPOSED SOLUTION

* 1. **Empathy Map Canvas**



### Ideation and Brainstroming

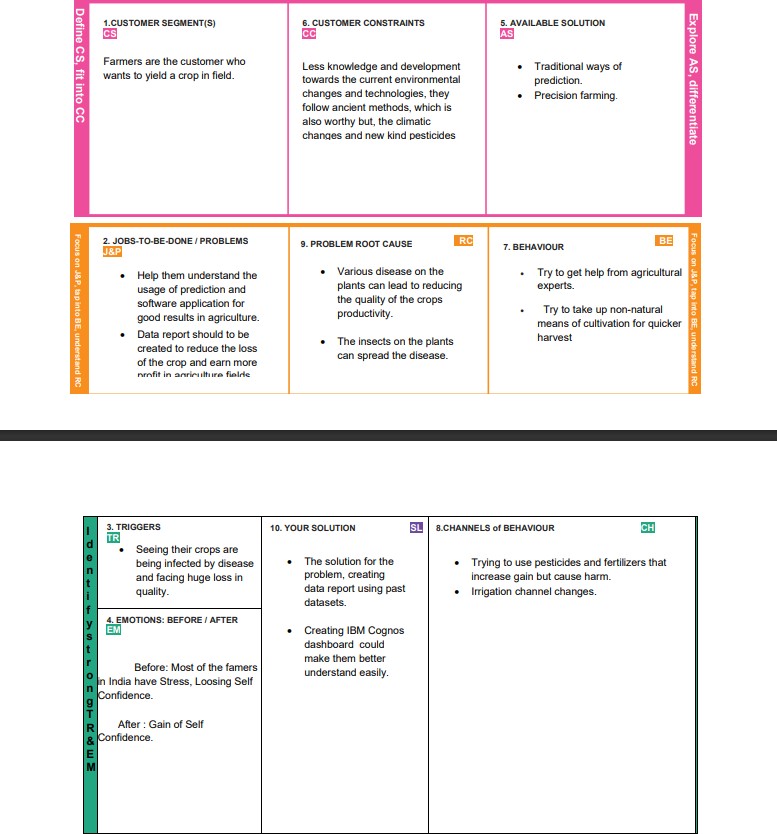
****

* 1. **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 |

* 1. **Problem Solution Fit**



## 4 - REQUIREMENT ANALYSIS

### 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 |

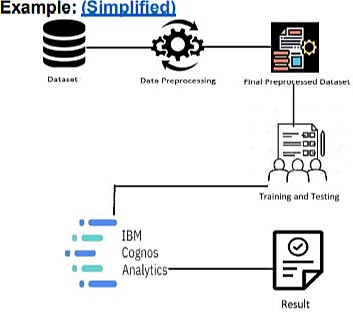
* 1. **Non Functional Requirements**

|  |  |
| --- | --- |
| **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

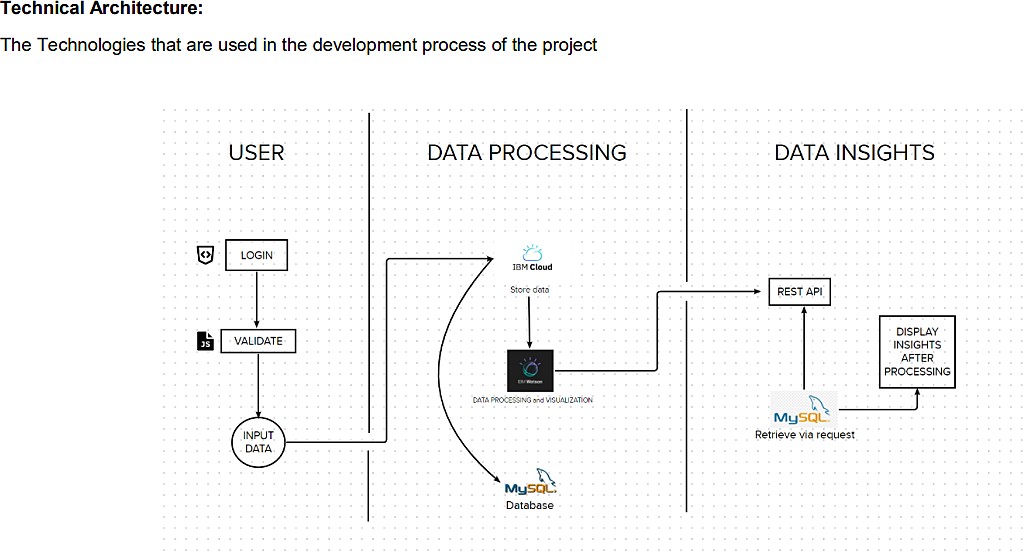
### Data Flow Diagrams

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.



### Solution and technical Architecture

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



* 1. **User Stories**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Dataset. |  |  |
|  | USN - 5 | Load the dataset to Cloud platform then Build the  required Visualizations | 10 | High |
|  |  | Showcase the Yearly usage  of Area in Crop Production. | 4 | Medium |
|  | USN - 7 | Build a visualization to show case top 10 States in Crop Yield Production by  Area | 4 | Medium |
|  |  | Build the required  Visualization to showcase the Crop Production by  State. | 4 | Medium |
|  |  | Build Visual analytics to represent the Sates with Seasonal Crop Production  using a Text representation. | 4 | Medium |
| Creating the dashboard | USN - 8 | Create the Dashboard by using the created  visualizations | 20 | High |
| Export the  Analytics | USN - 9 | Export the created  Dashboard | 20 | High |

## 6 - PROJECT PLANNING AND SCHEDULING

### Sprint Planning and Estimation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sprint** | **Functional**  **Requirements** | **user story/**  **Number** | **User Story/**  **Task** | **Story**  **points** | **Priority** |
| Sprint -1 | Registration | USN - 1 | As a user, I can  register for by entering my Agri - id card and request. | 2 | High |
| Sprint - 1 |  | USN - 2 | As a user, I can register for the  application through  Gmail | 2 | Medium |
| Sprint - 1 | Login | USN - 3 | As a user, I can Call and request or  Approach for dataset | 2 | High |
| Sprint - 1 | working with Dataset | USN - 4 | To work on the given dataset, Understand  the Dataset. | 4 | High |
|  |  | USN - 5 | Load the dataset to Cloud platform then Build the required  Visualization | 10 | High |
| 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 |
|  |  |  | Showcase the Yearly usage of Area in Crop  Production | 4 | Medium |
|  |  |  | Build a visualization to  show case top 10 | 4 | Medium |

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

* 1. **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 |

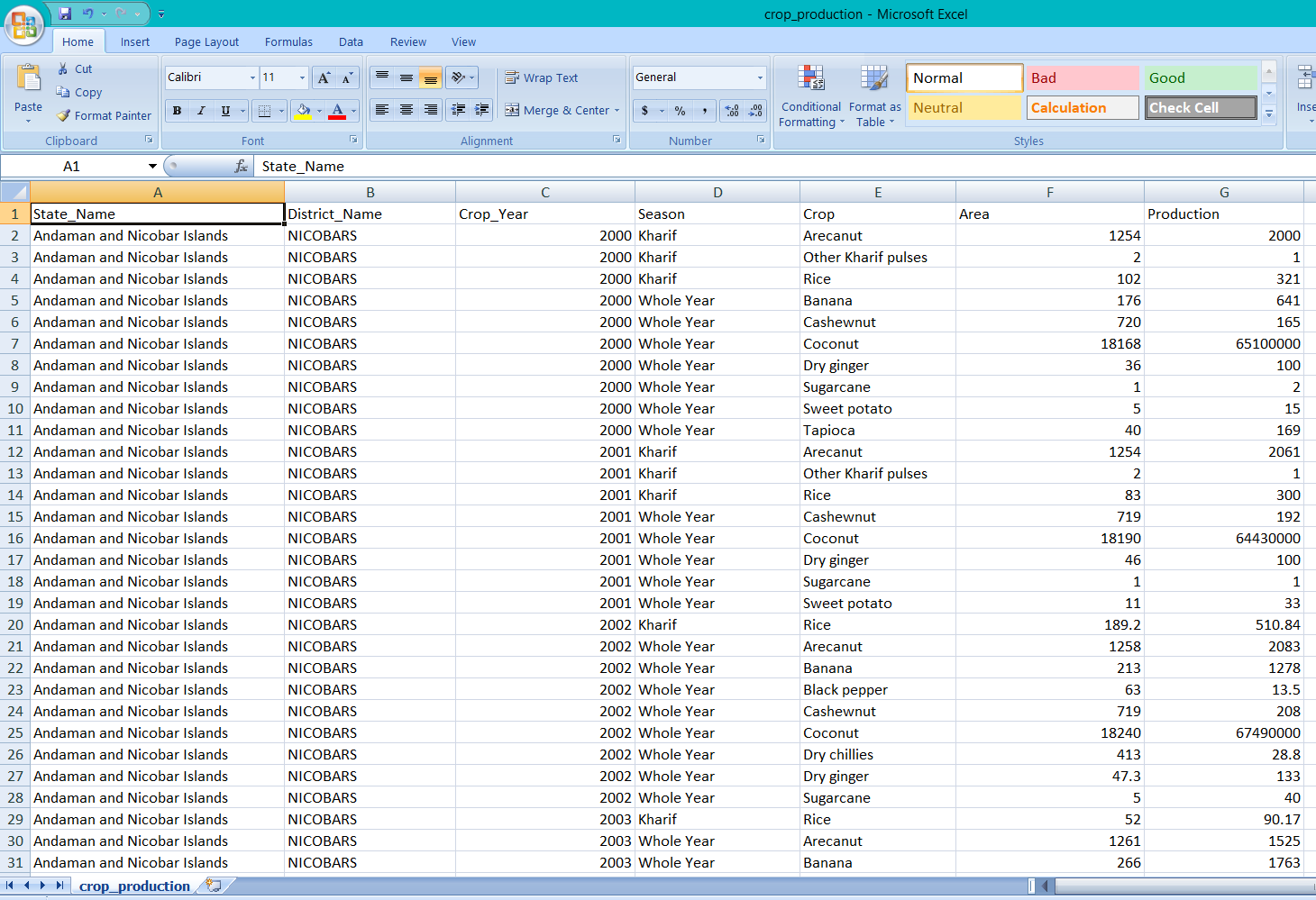
## 7 - WORKING WITH THE DATASET AND DATA VISUALIZATION

* 1. **Understanding the Dataset**

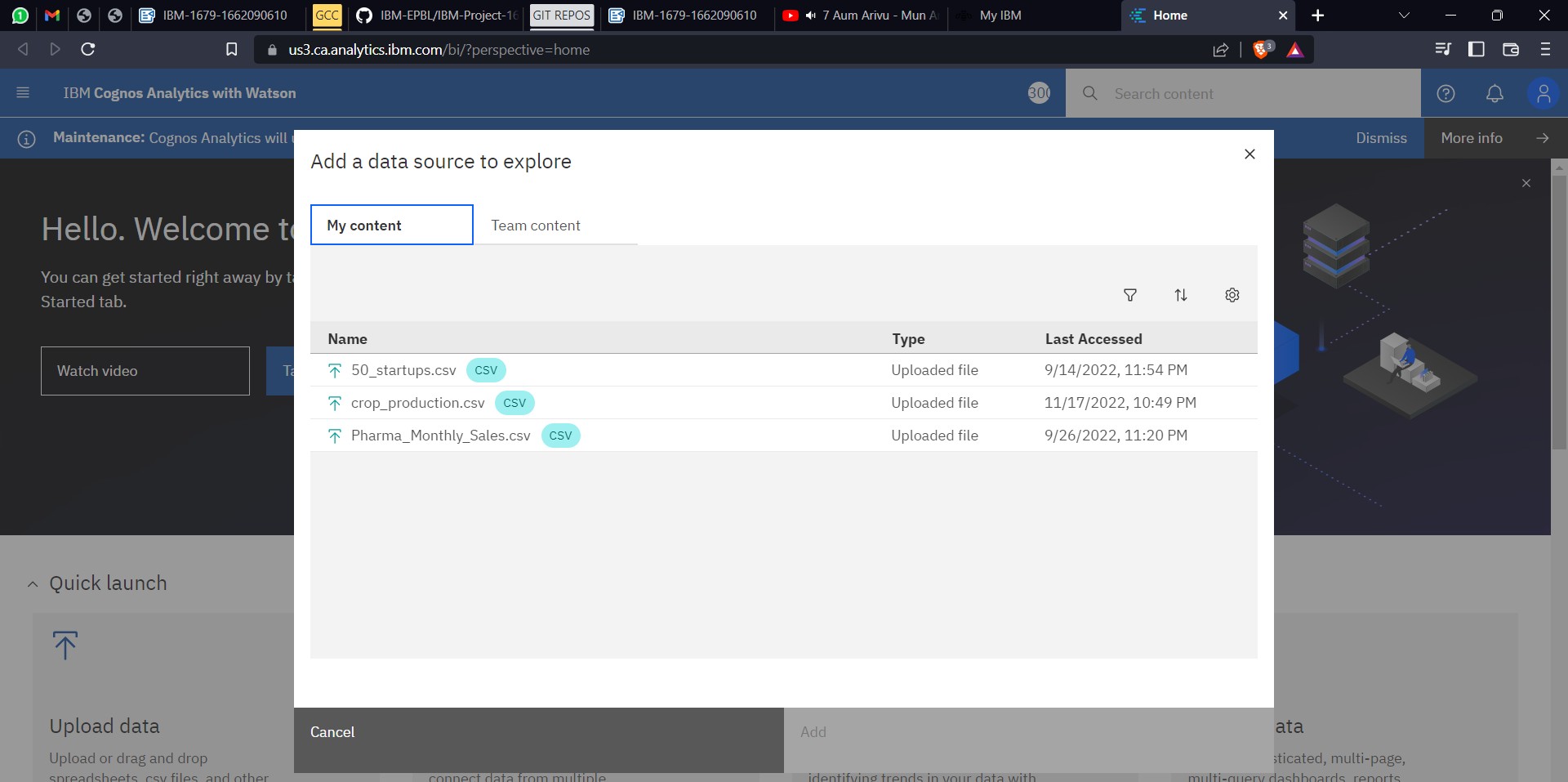
This project is based on a understanding the crop production of India.It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

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.

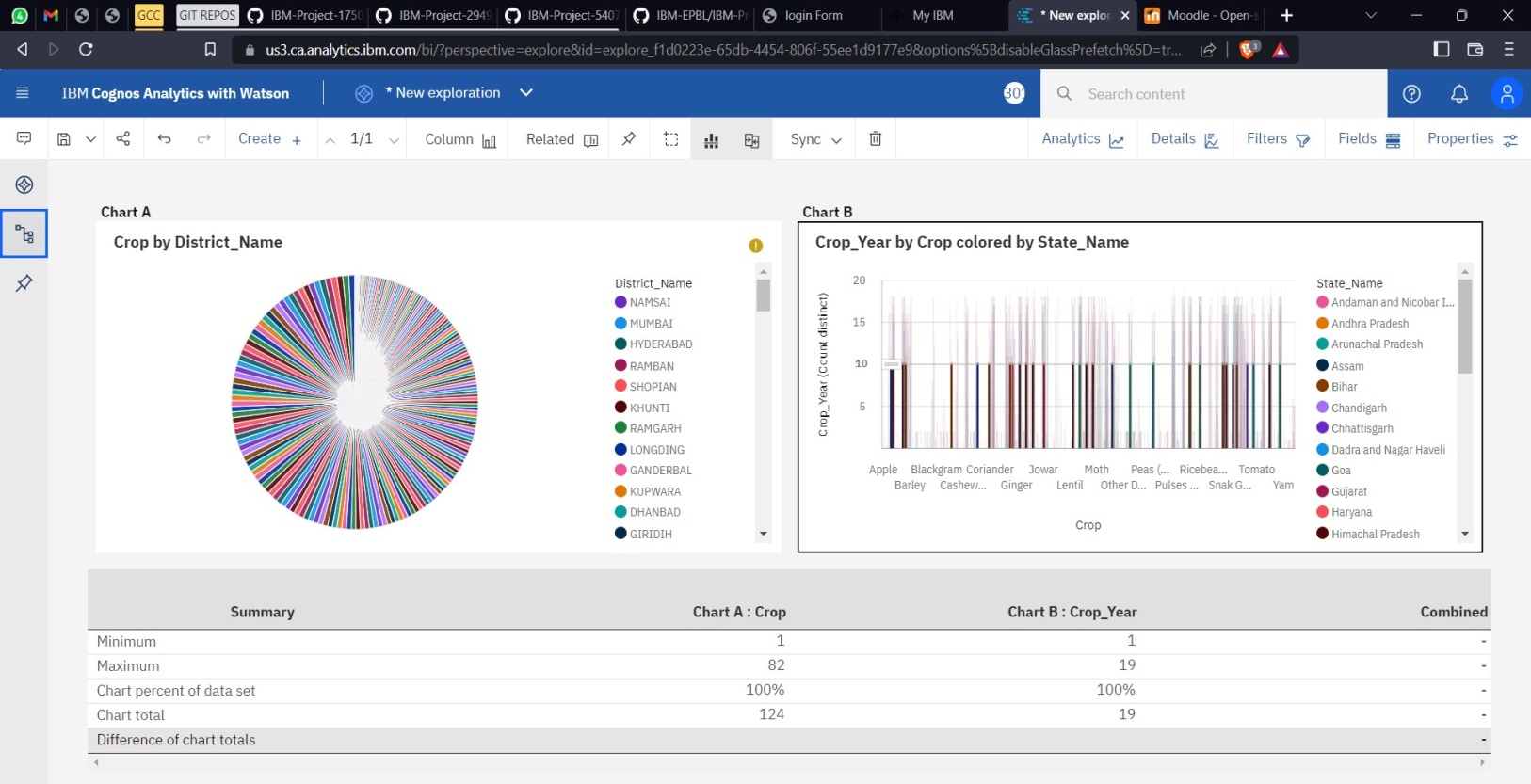
The data format is as shown in the below image

## Loading the Dataset

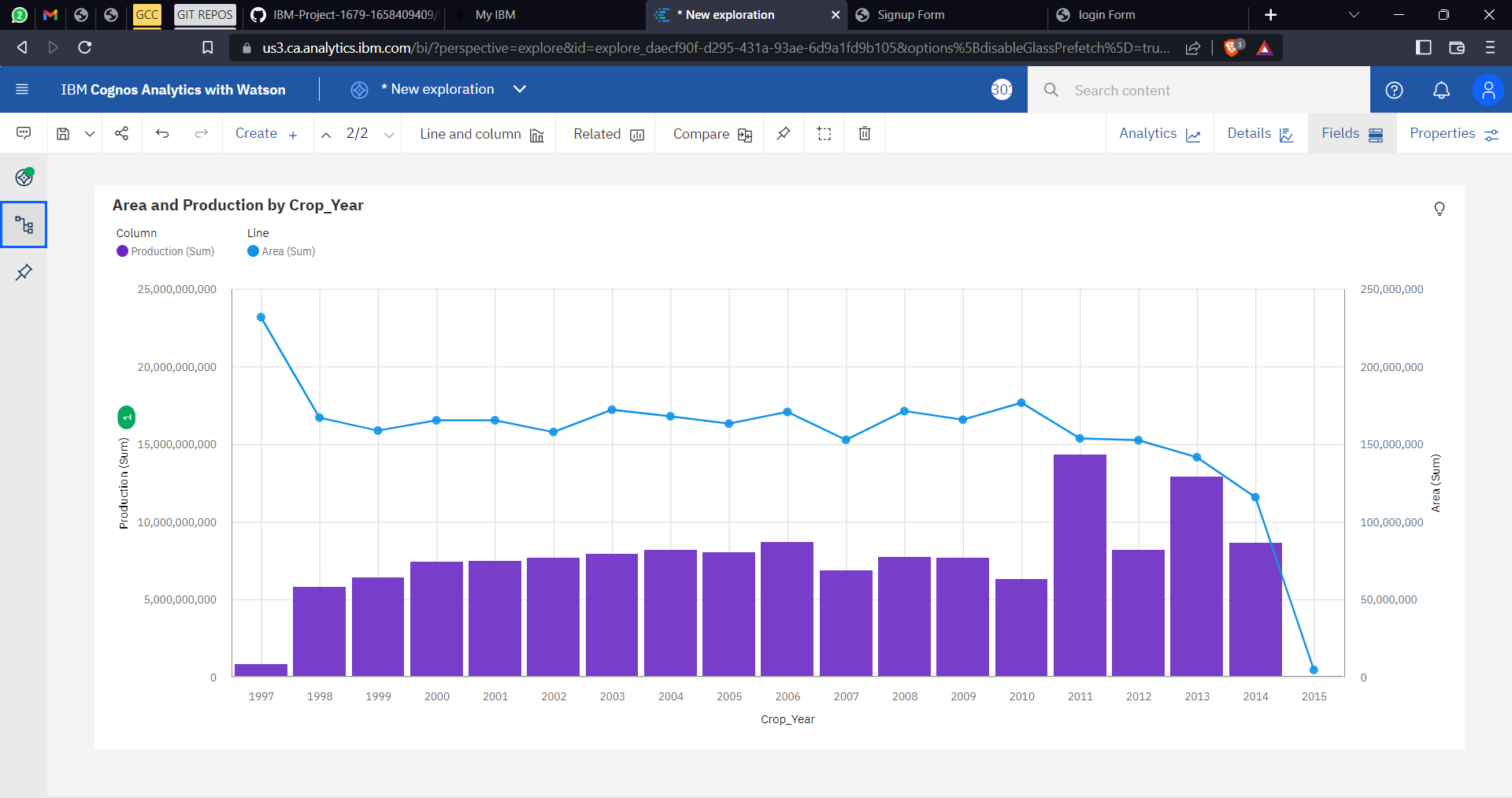


* + - Open the menu dialog box
    - Then click on upload data and then the uploaded dataset can be viewed here

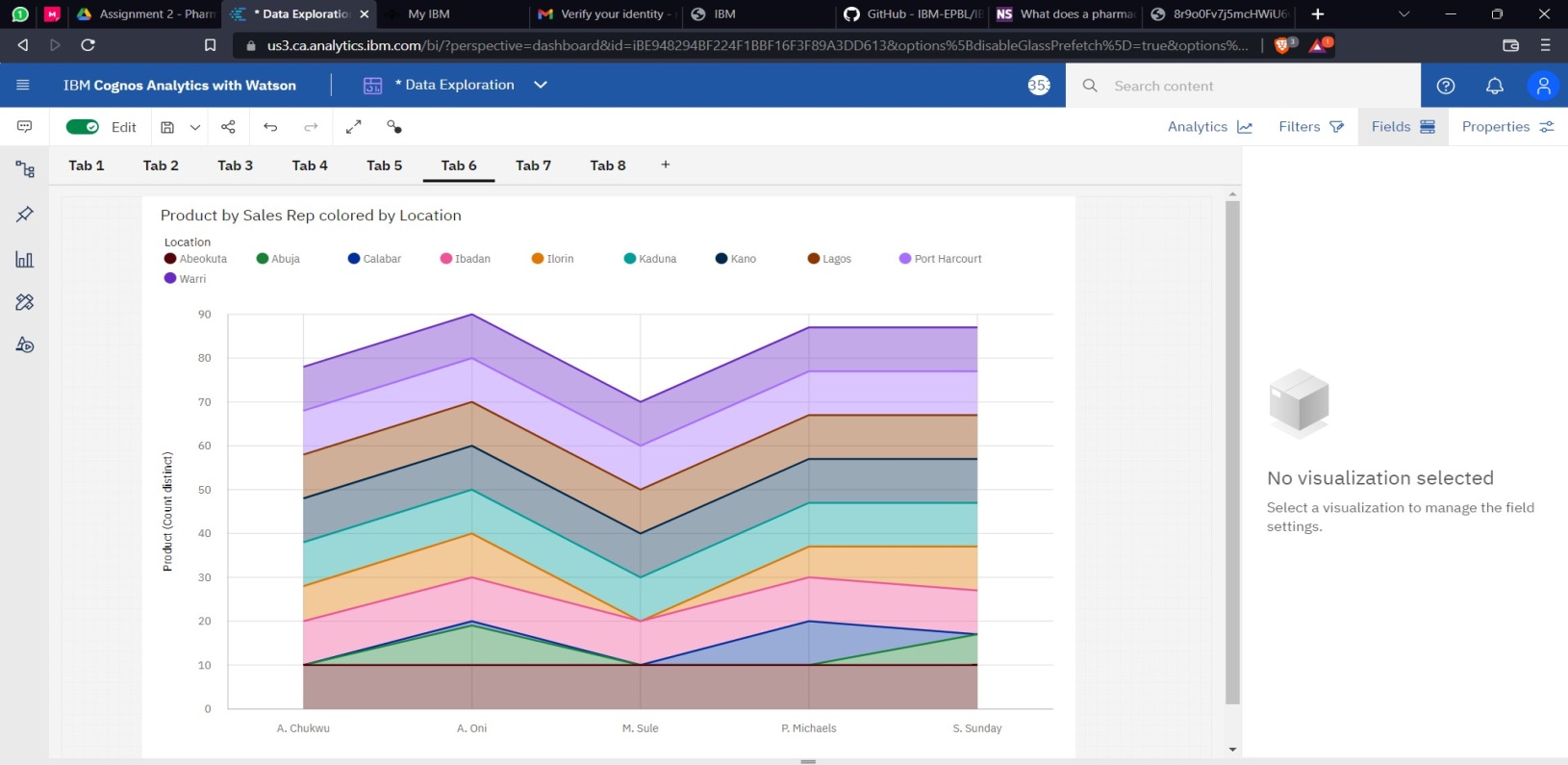
## Visualization charts



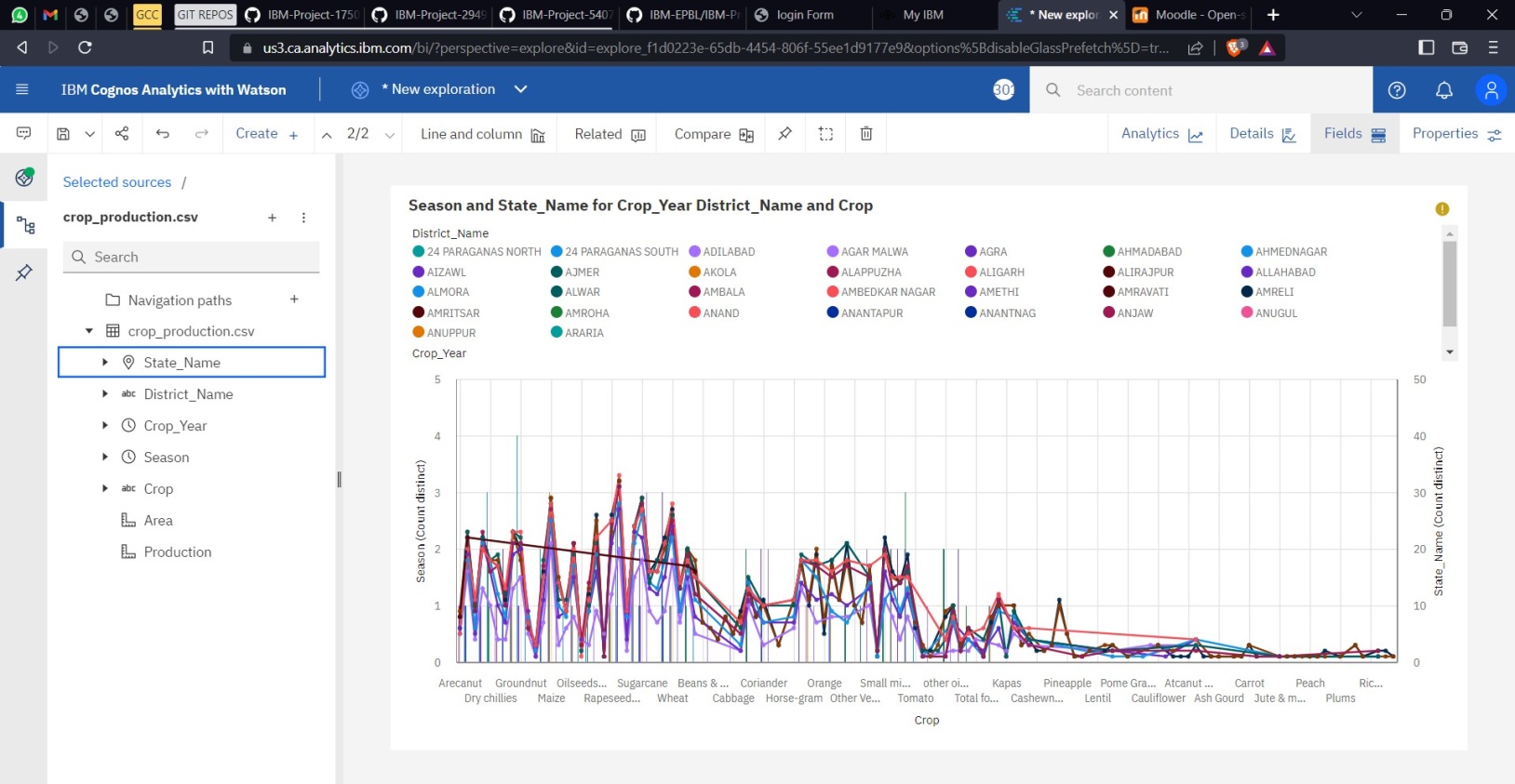
* CHART A - There are two charts displaying the Crop production by district name.
* CHART B - It shows the crop growth year according to the states where they grow.



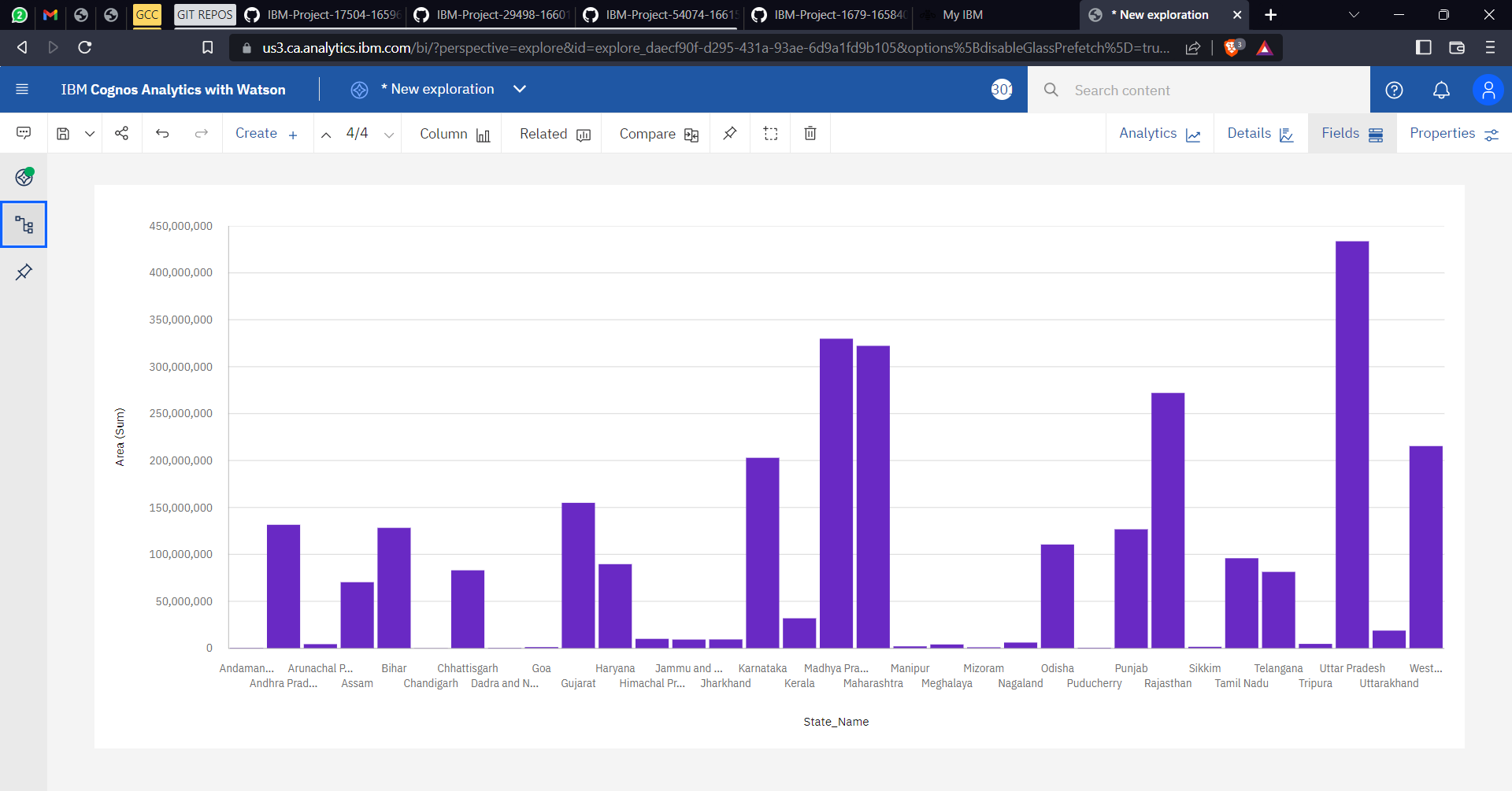
* It shows the Area and Production of the crop by the year



* It displayes the product (crop) by sales

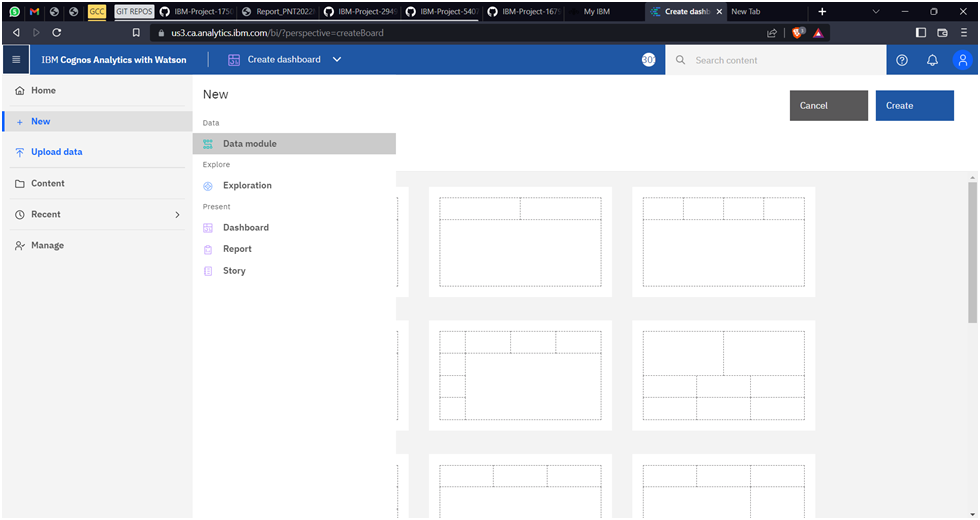


* The season and the statename for the year with good growth of the crop



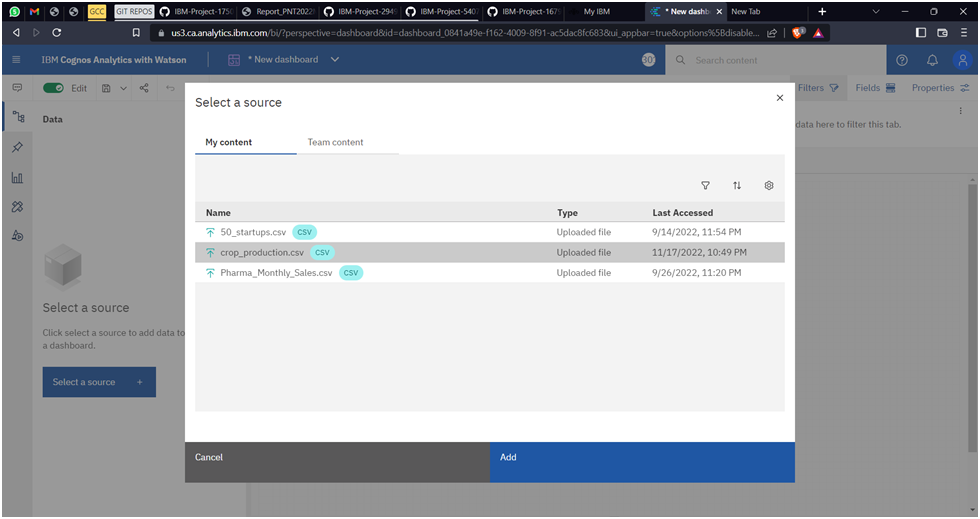
## 8 - CREATING THE DASHBOARD AND EXPORT THE ANALYTICS

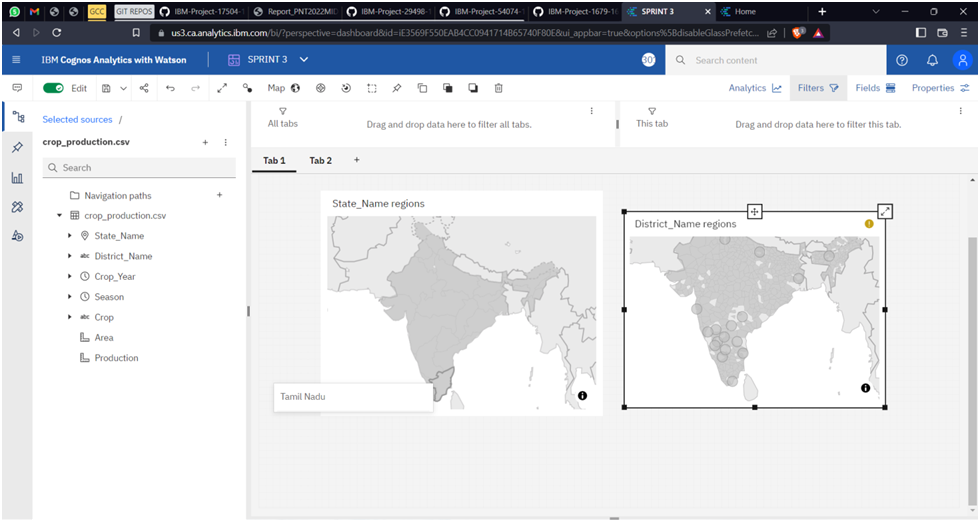
### Creating the Dashboard

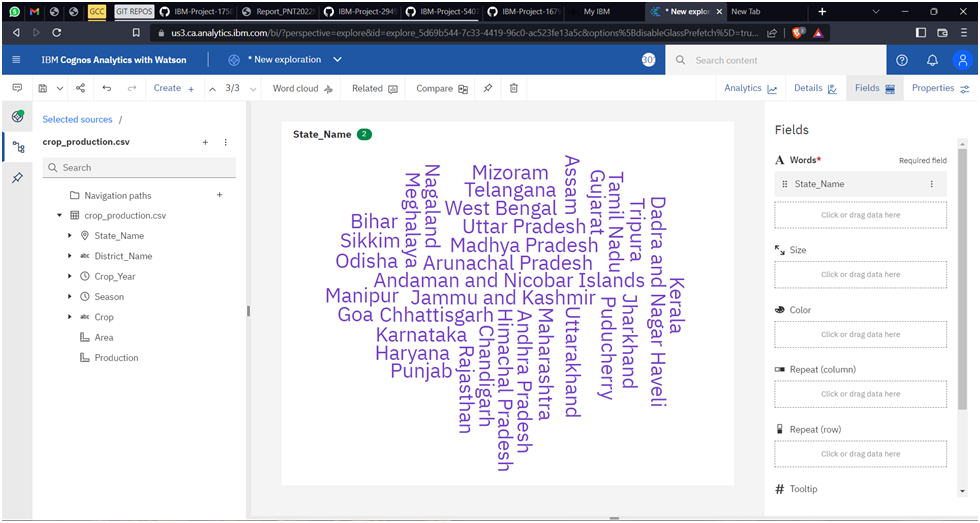
\* Login into IBM COGNOS and then open menu and click NEWàDASHBOARD

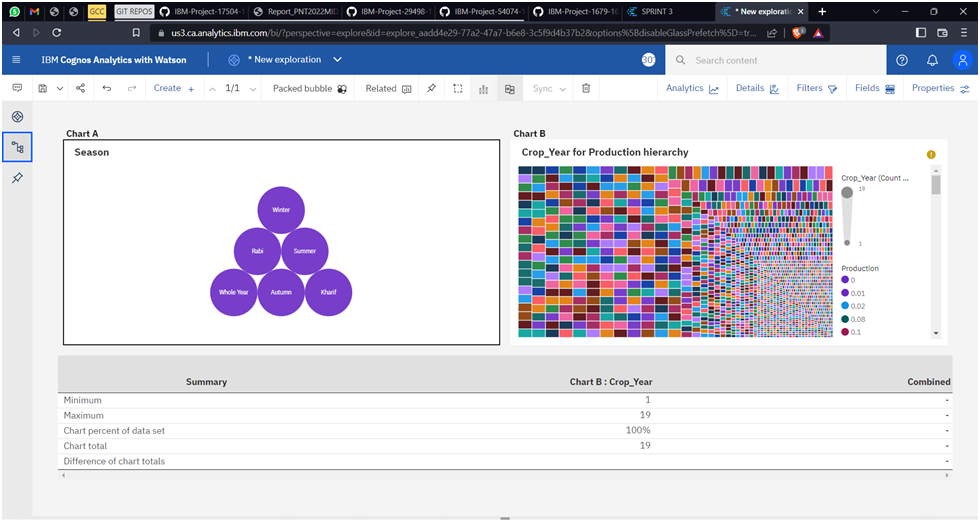
\* After clicking on new dashboard select the type of template required

**\*** Then select the **crop\_production.csv** ﬁle dataset under my content (previously uploaded for visualization and Exploration) and click add



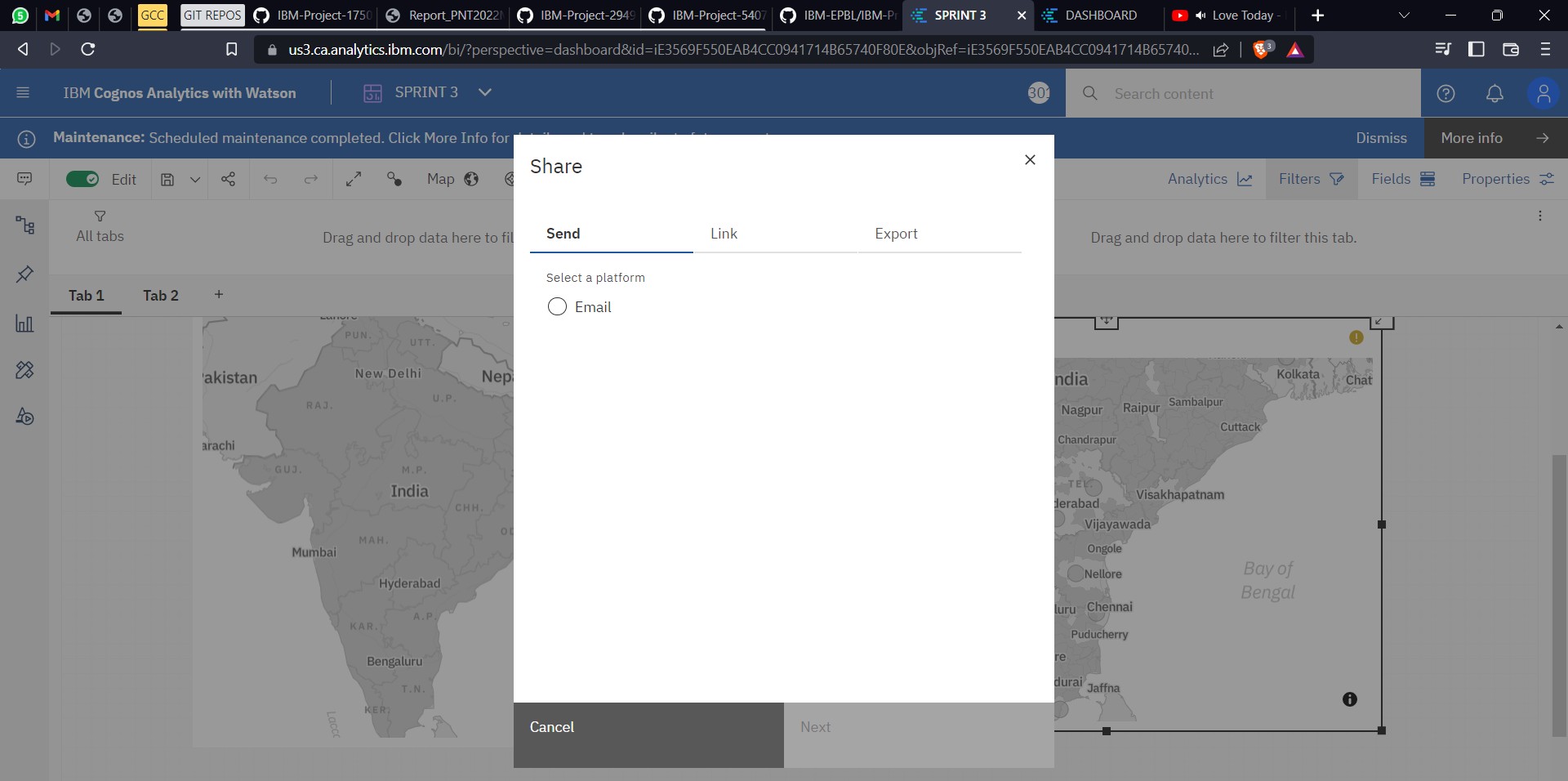
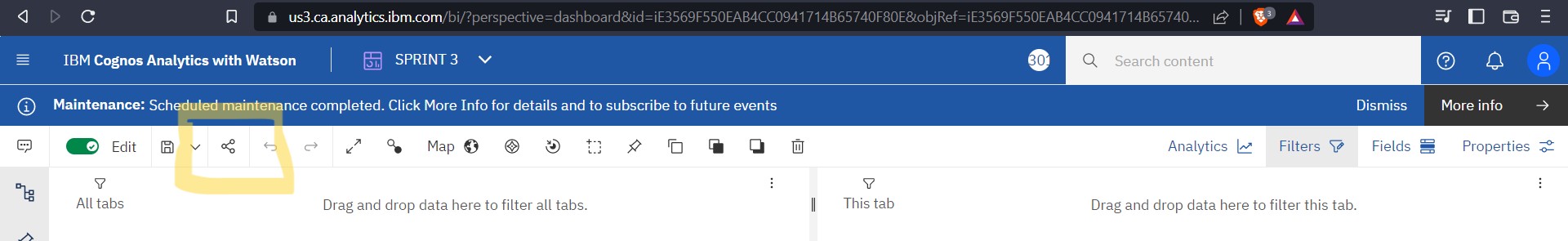
* Rename the tab title as sprint 3 / Dashboard
* Copy the Each Visualization and paste it in the tab and arrange the chart





### 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

### Source code

[**My Dashboard Sprint for crop yield**](https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=iBE948294BF224F1BBF16F3F89A3DD613&objRef=iBE948294BF224F1BBF16F3F89A3DD613&options%5BdisableGlassPrefetch%5D=true&options%5Bcollections%5D%5BcanvasExtension%5D%5Bid%5D=com.ibm.bi.dashboard.canvasExtension&options%5Bcollections%5D%5BfeatureExtension%5D%5Bid%5D=com.ibm.bi.dashboard.core-features&options%5Bcollections%5D%5Bbuttons%5D%5Bid%5D=com.ibm.bi.dashboard.buttons&options%5Bcollections%5D%5Bwidget%5D%5Bid%5D=com.ibm.bi.dashboard.widgets&options%5Bcollections%5D%5BcontentFeatureExtension%5D%5Bid%5D=com.ibm.bi.dashboard.content-features&options%5Bcollections%5D%5BsaveServices%5D%5Bid%5D=com.ibm.bi.dashboard.saveServices&options%5Bcollections%5D%5Btemplates%5D%5Bid%5D=com.ibm.bi.dashboard.templates&options%5Bcollections%5D%5BvisualizationExtension%5D%5Bid%5D=com.ibm.bi.dashboard.visualizationExtensionCA&options%5Bcollections%5D%5BboardModel%5D%5Bid%5D=com.ibm.bi.dashboard.boardModelExtension&options%5Bcollections%5D%5BcontentTypes%5D%5Bid%5D=com.ibm.bi.dashboard.contentTypes&options%5Bcollections%5D%5BserviceExtension%5D%5Bid%5D=com.ibm.bi.dashboard.serviceExtension&options%5Bcollections%5D%5BlayoutExtension%5D%5Bid%5D=com.ibm.bi.dashboard.layoutExtension&options%5Bcollections%5D%5BcolorSetExtensions%5D%5Bid%5D=com.ibm.bi.dashboard.colorSetExtensions&options%5Bconfig%5D%5Bproduct%5D=CA&options%5Bconfig%5D%5BeditPropertiesLabel%5D=true&options%5Bconfig%5D%5BenableCustomVisualizations%5D=true&options%5Bconfig%5D%5BassetTags%5D%5B%5D=dashboard&options%5Bconfig%5D%5BfilterDock%5D=true&options%5Bconfig%5D%5BshowMembers%5D=true&options%5Bconfig%5D%5Bupgrades%5D=dashboard-core%2Fjs%2Fdashboard%2Fupgrades&options%5Bconfig%5D%5BassetType%5D=exploration&options%5Bconfig%5D%5BgeoService%5D=CA&options%5Bconfig%5D%5BsmartTitle%5D=true&options%5Bconfig%5D%5BnavigationGroupAction%5D=true&options%5Bconfig%5D%5BenableDataQuality%5D=false&options%5Bconfig%5D%5BmemberCalculation%5D=false&isAuthoringMode=false&boardId=iBE948294BF224F1BBF16F3F89A3DD613)

### Github LInk and project demo Link

<https://github.com/IBM-EPBL/IBM-Project-48815-1660813323/>

<https://drive.google.com/file/d/1G3ZoJOlCNaSMNV3zHWl2raGLdTxMGIyQ/view?usp=drivesdk>