**Project Report**

|  |  |
| --- | --- |
| Date | 17 November 2022 |
| Team ID | PNT2022TMID30683 |
| Project Name | Estimate the crop yield using Data Analytics |

# 1. INTRODUCTION

**1.1 Project Overview**

India is primarily an agricultural country. Agriculture is currently the most important emerging sector in the world and a key industrial and economic pillar of our country. The field of agricultural information technology has recently undergone major changes making crop yield prediction an interesting research topic. Crop yield forecasting is a technique for estimating crop yield based on many characteristics such as temperature, rainfall, fertilizers, pesticides, and other climatic variables and parameters. Using data analytics to analyze these parameters and provide the patterns or trends tracked over the past few years when estimating yield so farmers can make the right choices when choosing crop varieties, etc. helps. Finally, to catch people's breath and use the knowledge presented, place the visualizations you create on dashboards and present them with the most appropriate and appropriate charts, graphs or maps.

**1.2 Purpose**

Analytics is the interpretation of data patterns that supports decision-making and performance improvement. Crop Yield Agricultural Data Analysis helps analyze some key visualizations and create dashboards. Examining them will give you the most insight into 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 and the 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. The information identified and collected was randomly distributed into 10 sections, one data section was used for testing and all other data sections were used for preparatory info Priya et al., (2018) proposed a random forest algorithm that considers various parameters such as rainfall, regional seasonal yields (Rabi and Kharif), temperature (maximum), and crop yield in kg. used to predict crop yields in specific regions. / tons. The study area was Tamil Nadu. The dataset was collected by the Indian government for his 15 years of rice production. In their experimental results, they proved that predictive analytics performed with random forest algorithms (supervised machine learning algorithms) can help farmers predict crop yields before they are planted in the field. This algorithm performs efficiently on large databases with high classification accuracy.

**Yield prediction using data miningtechniques:**

Raor ane AA and Kul Karn i R.V. described some data mining techniques in their article. They concluded that with the right data, efficient techniques could be developed and analyzed, and data mining techniques could be used to solve complex agricultural problems. We also recommend some of the algorithms and statistical methods that yield [8] positive results for agricultural growth.

**Yield prediction by big data analysis:**

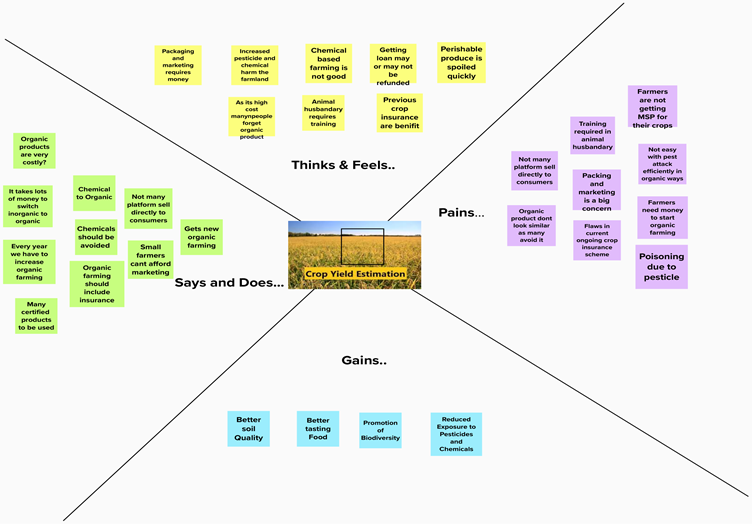
In India, crop yield is seasonal and primarily he is influenced by biological and economic causes of one crop. Reporting on progressive agricultural yields in all seasons is a huge and useful task for any country in assessing forecasts and estimates of total yields. A common problem around the world today is that farmers are stressed to achieve higher yields due to the impact of unpredictable climate change and a drastic reduction in global water resources. A study was conducted to collect data on global climate change and available water resources. This can be used to facilitate advanced and new approaches such as big data analytics to obtain information on previous outcomes for crop yield forecasting and estimation. This study imported that selecting and using the most desirable crops according to existing conditions helps to obtain higher and improved crop yields. S. Athmaja, M. Hanumanthappa, and V. Kavitha, a survey on machine learning algorithms, presents effective strategies for big data analytics. Farmers around the world have benefited somewhat from comparative knowledge through big data analytics. Machine learning algorithms using vast amounts of data will give farmers comparative knowledge to bring about changes in normal farming. I'm here.

**2.2 Problem definition**

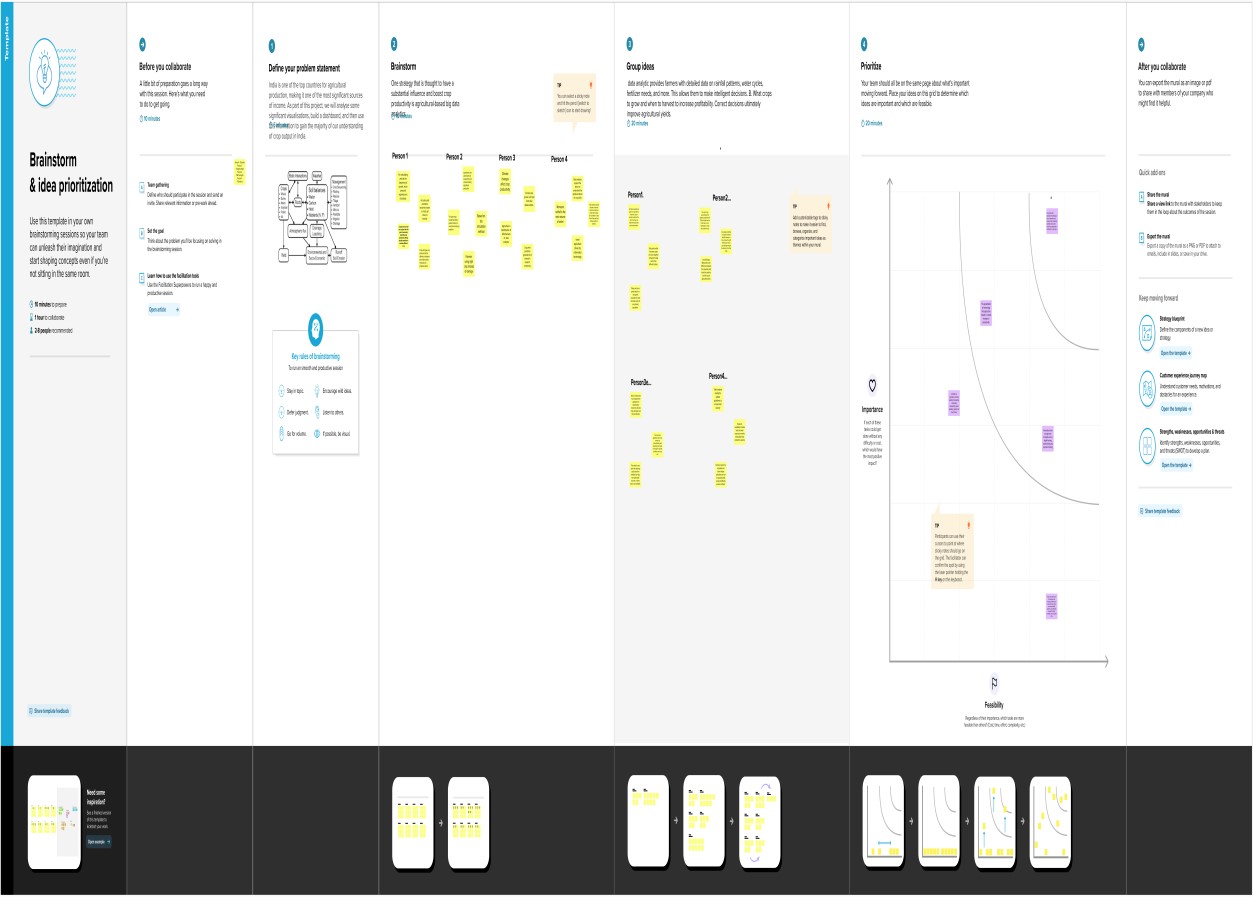
In agriculture, farmers struggle to analyze the need for market analysis and soil quality analysis to achieve high yields through technology. The main purpose of this project is to predict yields. This is very helpful for farmers to plan their harvest and sales of harvested grains.

# 3. IDEATION & PROPOSED SOLUTION

**3.1 Empathy Map Canvas**



**3.2 Ideation & Brainstorming**



**3.3 Proposed Solution**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
| 1. | Problem Statement (Problem to be solved) | Poverty, farmer suicides, yield profits, and appetite loss are all on the rise. |
| 2. | Idea / Solution description | After a thorough analysis of the prior data, provide the ideal data report. assisting them in overcoming commercial and agricultural losses. |
| 3. | Novelty / Uniqueness | With the help of this technology, we can analyze and visualize the data and give farmers options for which plants or crops to grow when in order to maximiZe crop productivity. |
| 4. | Social Impact / Customer Satisfaction | It significantly affects agricultural yield by offering the ideal data visualisations. and increasing the farmers' profits. |
|  |  |  |

**3.4 Problem Solution fit**

5.AVAILABLE SOLUTIONS AS

What kind of solution for agriculture ?

Farmers are given access to adequate education. Provision of a large Land Area to Farmers lowering the price of farmer inputs for farmers. Encouragement of Age and Gender in Agriculture Encourage farmers to join cooperative societies.

6.CUSTOMER LIMITATIONS

Farmers need to limitation factors for example?

Farmers markets are well known for their significant contributions to the growth of local food systems, the encouragement of small farms, the stimulation of local economic activity, the addressing of issues relating to food access and security, and the creation of opportunities for community building.

1.CUSTOMER SEGMENTS CS

Who’s your customer

Agronomist

8.CHANNELS OF BEHAVIOR CB

8.1:ONLINE-where does this behaviour happen?

8.2:OFFLINE-what kind of action do customers take offline?

8.1: Nil

8.2: Farmers who want to boost their fields' average crop production per acre must have a well-organized irrigation system on standby.

10.YOUR SOLUTION SL

Write down your current solution first, fill in the canvas, and assess how well it corresponds with reality if you are working on an existing firm. If you are developing a new business proposition, leave the canvas blank until you have filled it in and developed a solution that satisfies the needs of the target market, addresses a problem, and takes into account the behaviour of the target market.

In light of regional conditions and extreme weather, we will advise farmers to Drought and heat can also assist farmers in increasing food production without harming ecosystems.

3.TRIGGERS TO ACT TR

What triggeres customer to act?

What prompts consumers to take action? for instance, seeing their neighbour install solar panels or reading in the news about a more effective method.

4.EMOTIONS EM

How do clients feel both during and after an issue or a job?

Use the contrast—for example, lost, insecure versus confident, in charge—in your communication strategy and design.

7.BEHAVIOR BE

How often does this related behaviour happen?

The majority of small-scale, family farmers choose agriculture as a career because they enjoy growing things. Their passion and motivation come from the hands-on, daily production activities. in terms of marketing. Many farmers, however, choose a more passive strategy, thinking that if it grows, it will sell.

9.PROBLEM ROOT CAUSE

What is the cause of every problem on the list?

 India's agriculture still faces challenges such as the adaptation to climate change disturbances, fragmented landholdings, low farm productivity, and high food price volatility.

2.PROBLEMS / PAINS PP

What problem do you solve for your customer?

Farmers deal with problems including high production costs and low profits, excessive taxes on agricultural inputs, etc. Explanation: Farmers currently deal with a variety of issues.

# 4. REQUIREMENT ANALYSIS

**4.1 Functional requirement**

|  |  |  |
| --- | --- | --- |
| **FR.NO** | **Functional Requirement (Epic)** | **Sub Requirement (story /sub - task)** |
| FR.1 | User Registration   |  | | --- | |  | | Utilizing a Form for Registration  signing up with Gmail  registering via WhatsApp  Utilizing Agri-Consultancy to register |
| FR.2 | User Confirmation   |  | | --- | |  | | Email confirmation required  Reassurance via OTP  Verification via Letter |
| FR.3 | User Profile | User Information  Farm Information |
| FR.4 | Required Data | The user’s (farmer’s) data to analyse the previous crop yield |
| FR.5 | Analysis | Clean up and analyse the data in light of a collection of previous data from multiple users (Farmer) |
| FR.6 | Estimation | Developing the ideal data module and visuals in IBM Cognos to improve crop yield estimation |

**4.2 Non-Functional requirements**

|  |  |  |
| --- | --- | --- |
| **NFR.NO** | **Non-Functional Requirements** | **Description** |
| NFR.1 | Usability | Data reports are produced based on the historical data itself. These recommendations will advise or consult on crop sowing. |
| NFR.2 | Security | The user information is protected with IBM Cognos (Data Visuals). |
| NFR.3 | Reliability | The dynamic data graphics dashboard can make the data report simple to interpret. |
| NFR.4 | Performance | Better performance among all users is made possible through interaction, and the visual advice is impressive. |
| NFR.5 | Availability | The dashboard is easily accessible and can be accessed on any smartphones, laptops, systems, etc. |
| NFR.6 | Scalability | The proposed solution's flexibility in implementation makes it very simple to boost crop yield estimation in various farms for various users. |

# 5. PROJECT DESIGN

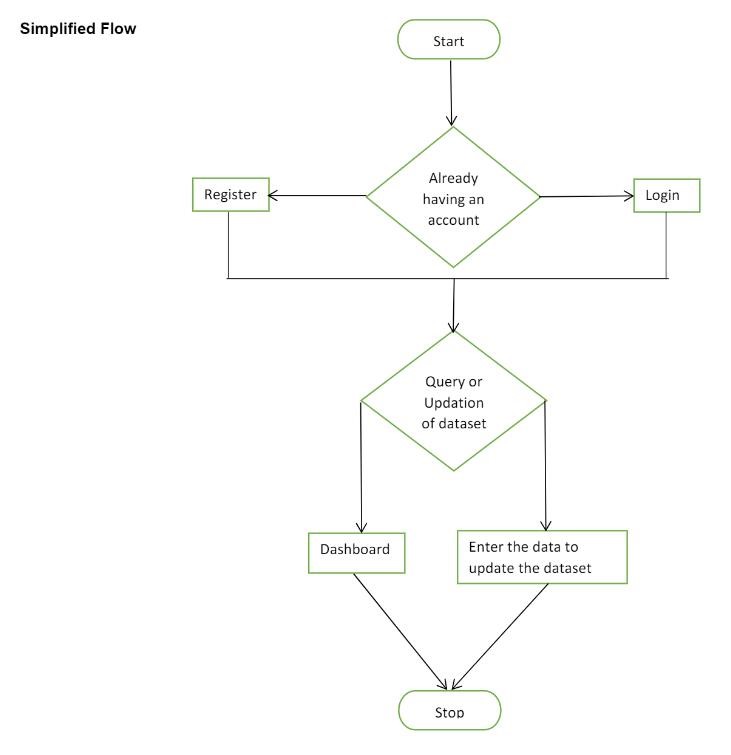
**5.1 Data Flow Diagrams**

**Analyze the Problem**

**Check the Model**

**Fit the Model to the Data**

**Experiment Design & Data Collection**



**5.2 Solution & Technical Architecture**

**Table-1 : Components & Technologies**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Components** | **Description** | **Technology** |
| 1. | User Interface | How a user interacts with a programme, such as through a chatbot or a mobile app. | HTML, CSS, MYSQL. |
| 2. | Application logic-1 | Using the application as a farmer (common user) to log in. | Java / Python. |
| 3. | Application logic-2 | Using the application's admin login information. | IBM Watson STT Service. |
| 4. | Application logic-3 | Logging into the application as a merchant. | IBM Watson Assistant. |
| 5. | Database | A database contains information regarding the crops. | MySQL, NoSQL, etc. |
| 6. | Cloud Database | To store data, IBM Watson cloud is employed. | IBM DB2, IBM Cloudant etc. |
| 7. | External API-1 | Use of an external API and its intended use. | e.g., IBM Weather API. |
| 8. | Machine Learning Model | A machine learning model's intended use. | Object Recognition Model, etc. |
| 9. | Infrastructure (Server / Cloud) | Local Server Configuration for an Application Deployed on a Local System or a Cloud.  Local Server Configuration,  Cloud Server Configuration. | Local, Cloud Foundry, Kubernetes, etc. |

**Table-2 : Application Characteristics:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Characteristics** | **Description** | **Technology** |
| 1. | Open-Source Frameworks | Describe the utilised open-source frameworks. | **Opensource framework technology** |
| 2. | Security Implementations | List every security and access control measure used, including firewalls. | e.g. SHA-256, Encryptions, IAM Controls, etc. |
| 3. | Scalable Architecture | Justify the three-tier architecture's ability to scale. | Used Technology |
| 4. | Availability | Justify the application's accessibility (using load balancers, distributed servers, etc., as examples). | Used Technology |
| 5. | Performance | Application performance (number of requests per second, use of Cache, use of CDNs, etc.) was taken into account during design. | Used Technology |

**5.3 User Stories**

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Milestone | Activities | Statues |
| 1. | Solution Requirement   |  | | --- | |  | | Creating the IBM cognos for creating dashboard and data visualization charts. | Completed |
| 2. | Project Objectives | Prepare the project objectives. | Completed |
| 3. | Project Flow | Prepare the Project Flow | Completed |
| 4. | IBM Cloud Account | Creating IBM Cloud account. | Completed |
| 5. | IBM Cognos Analytics | Creating IBM Cognos account. | Completed |
| 6. | Working with the Dataset | * Understanding Dataset. * Loading the Dataset. | Completed |
| 7. | Data visualization charts | * Season with average productions. * With years usage of area and production. * Top 10 States with most area. * State with crop production. * Stat with the crop production along with season. | Completed |
| 8. | Creating the dashboard | Creating the Dashboard. | Completed |
| 9. | Export the analytics | Export the analytics. | Completed |
| 10. | Ideation Phase | * Literature Survey on the selected project & information Gathering. * Empathy map. * Ideation. | Completed |
| 11. | Project Design Phase-I | * Proposed Solution * Problem Solution fit * Solution Architecture | Completed |
| 12. | Project Design Phase-II | * Customer Journey * Functional Requirement * Data Flow Diagram * Technology Architecture | Completed |
| 13. | Project Planning Phase | * Prepare Milestone & Activity List * Sprint delivery Plan | Completed |
| 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 | Completed |

# 6. PROJECT PLANNING & SCHEDULING

**6.1 Sprint Planning & Estimation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sprint | Functional Requirement  (Epic) | User Story Number | User Story & task | Story Points | Priority | Team Members |
| Sprint-1 | Registration | USN-1 | I can sign up for the application as a user by providing my email address, a password, and a password confirmation. | 2 | High | Vinotha  Yogakiruthiga  Poongodi  Sumithra |
| Sprint-1 |  | USN-2 | After registering for the application, I as a user will receive a confirmation email. | 1 | High | Vinotha  Sumithra |
| Sprint-2 |  | USN-3 | I can sign up for the application as a user using Google. | 2 | Low | Vinotha  Poongodi |
| Sprint-1 |  | USN-4 | I can sign up for the application as a user using Gmail. | 2 | Low | Vinotha  Yogakiruthiga |
| Sprint-1 | Login | USN-5 | I can access the application as a user by providing my email address and password. | 1 | High | Poongodi  Yogakiruthiga |
| Sprint-3 | Dashboard | USN-6 | I am free to use my dashboard and explore the features as a user. | 2 | High | Sumithra  Poongodi  Yogakiruthiga  Vinotha |
| Sprint-2 |  | USN-7 | I can access using the credentials as a user the assets I’m applying for. | 2 | High | Yogakiruthiga  Sumithra |
| Sprint-3 |  | USN-8 | Data manipulation operations carried out by the application. | 1 | High | Vinotha  Poongodi |
| Sprint-3 | Visualization | USN-9 | Can use certain datasets to generate dashboards. | 2 | Medium | Vinotha  Sumithra |
| Sprint-4 |  | USN-10 | One can perform predictive analysis. | 1 | High | Sumithra  Poongodi |
| Sprint | Functional Requirement  (Epic) | User Story Number | User Story & task | Story Points | Priority | Team Members |
| Sprint-3 |  | USN-11 | With certain datasets, I can produce stories. | 2 | High | Vinotha  Yogakiruthiga |
| Sprint-4 |  | USN-12 | Can export and send reports in accordance with the built-in dashboards and stories. | 2 | High | Yogakiruthiga  Poongodi  Vinotha |

**6.2 Sprint Delivery Schedule**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Total**  **Points** | **Story** | **Duration** | **Sprint Start Date** | **Sprint End**  **(Planned)** | **Date** | **Story Points**  **Completed (as on**  **Planned End Date)** | **Sprint Release**  **(Actual)** | **Date** |
| 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 & DATA VISUALISATION

**7.1 Understanding the dataset**

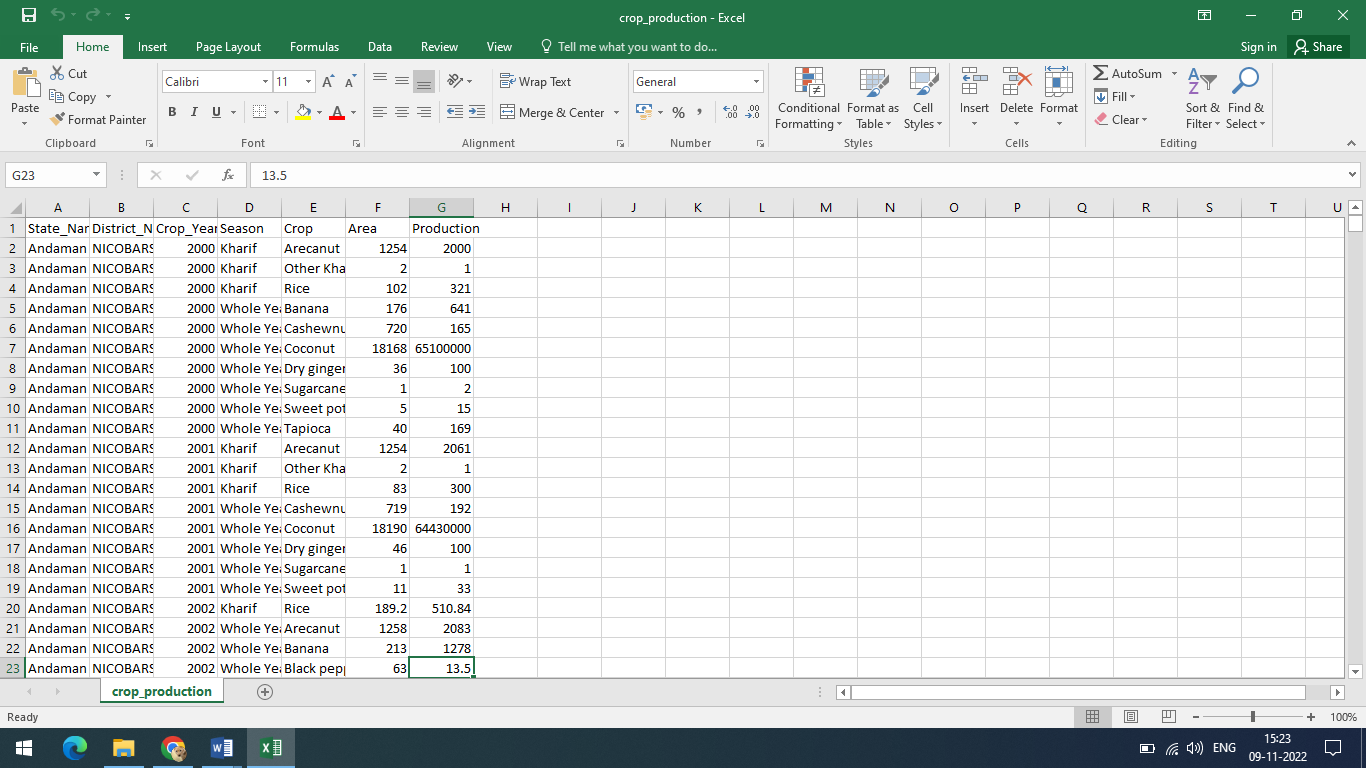
This project is based on a understanding the crop production of India. It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

# Dataset Link: [Dataset](https://www.kaggle.com/datasets/abhinand05/crop-production-in-india)

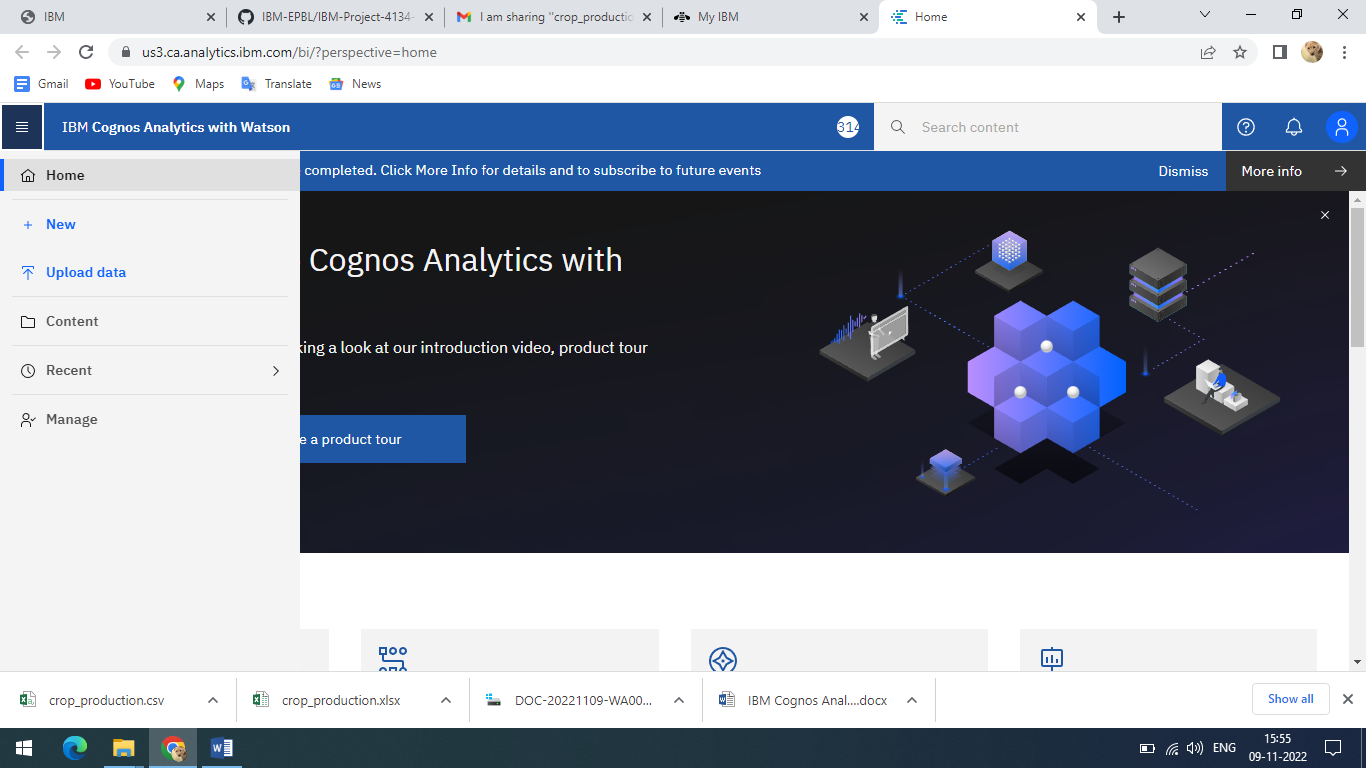
Let’s understand the data we’re working with and give a brief overview of what each feature represents or should represent 1. State Name - All the Indian State names.

1. District Name -Different District names.
2. Crop Year- contains the crop years.
3. Season – Different seasons for crop production.
4. Area- Total number of areas covered.
5. Production- production of crops.

The data format is as shown in the below image:



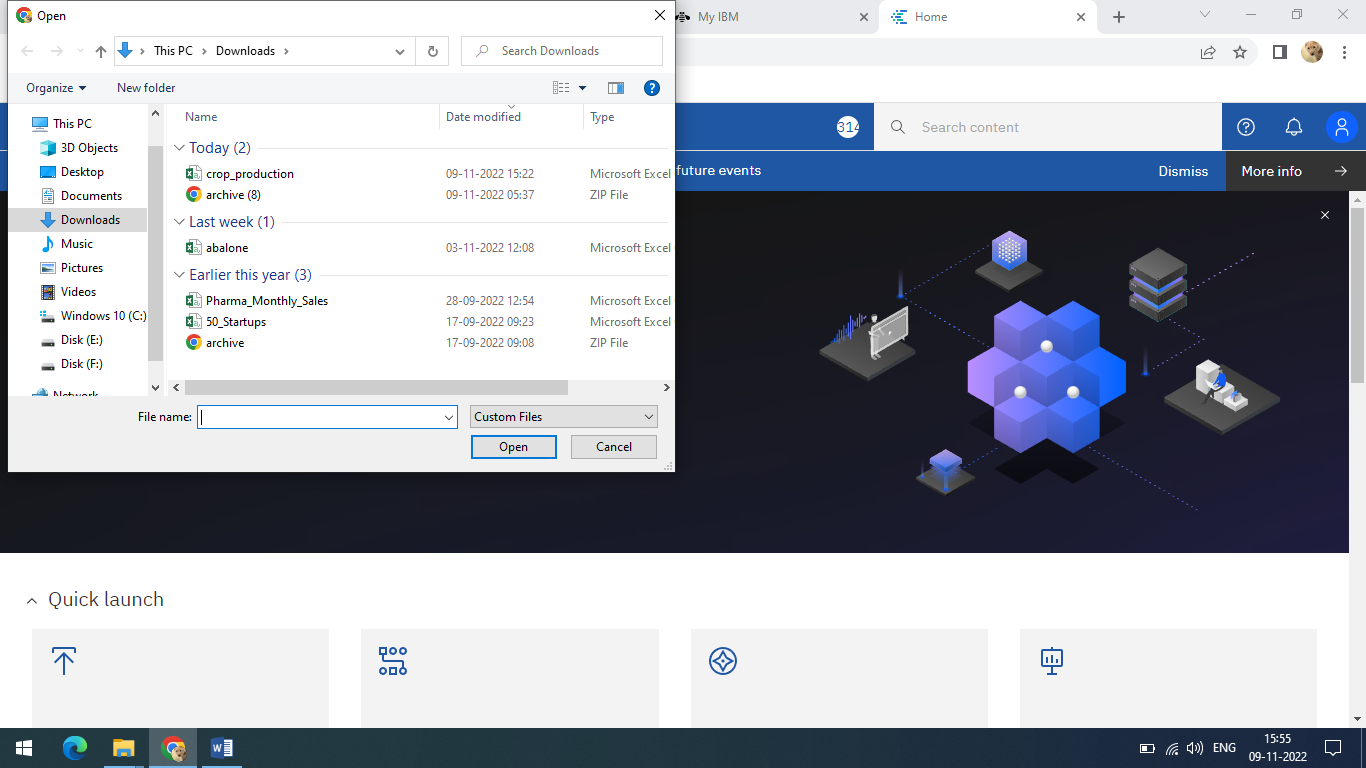
**7.2 Loading the dataset**

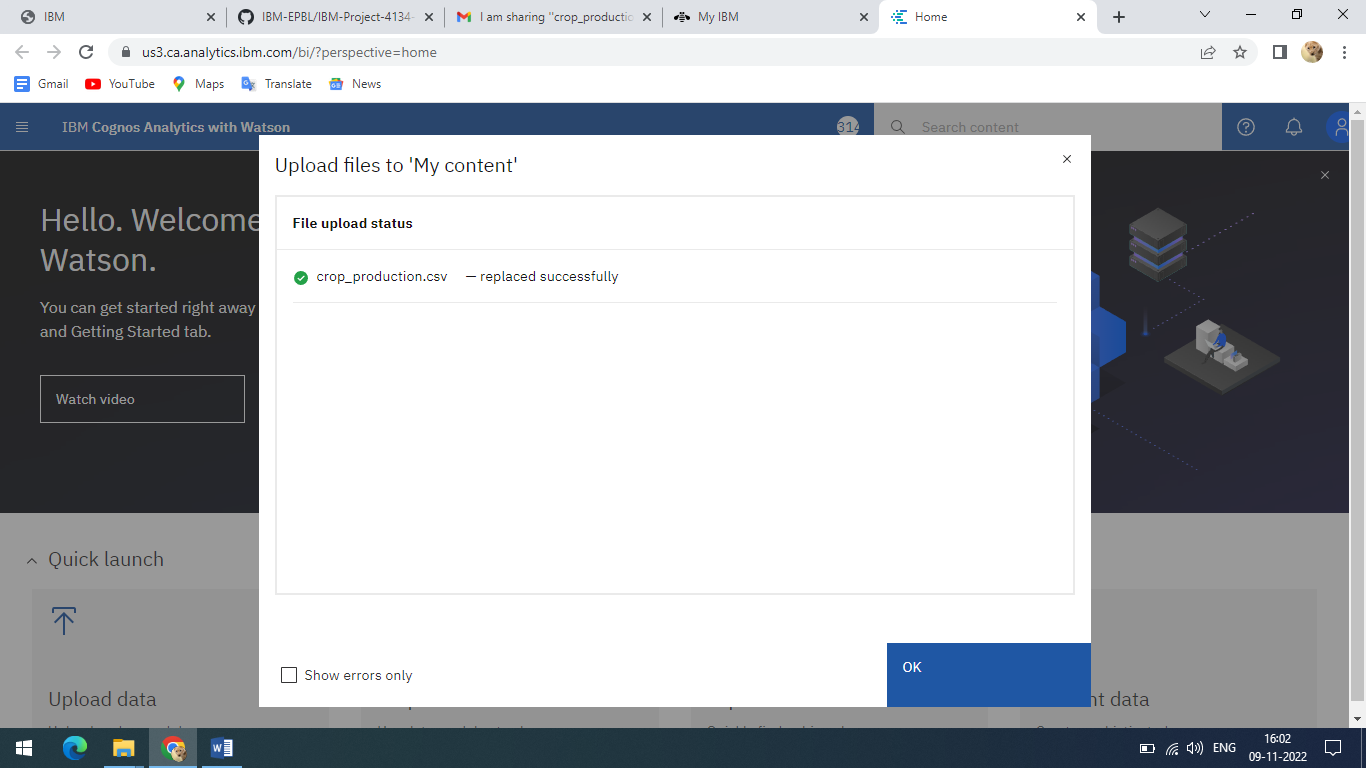


* + - Click the open menu in the top left corner.

* + - Select the **Upload Data** in the menu and select the Dataset that you want upload.

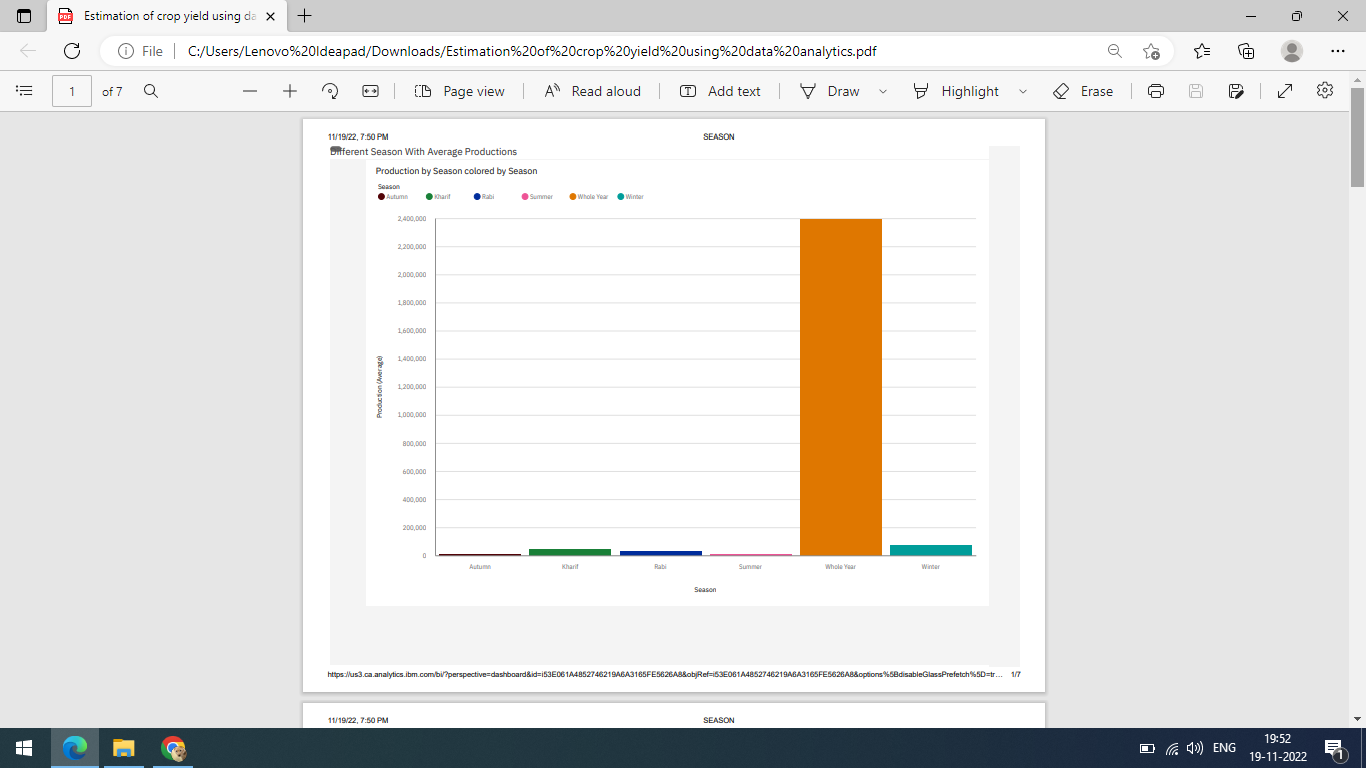
* + - Once the Dataset is Uploaded it will be displayed in content.



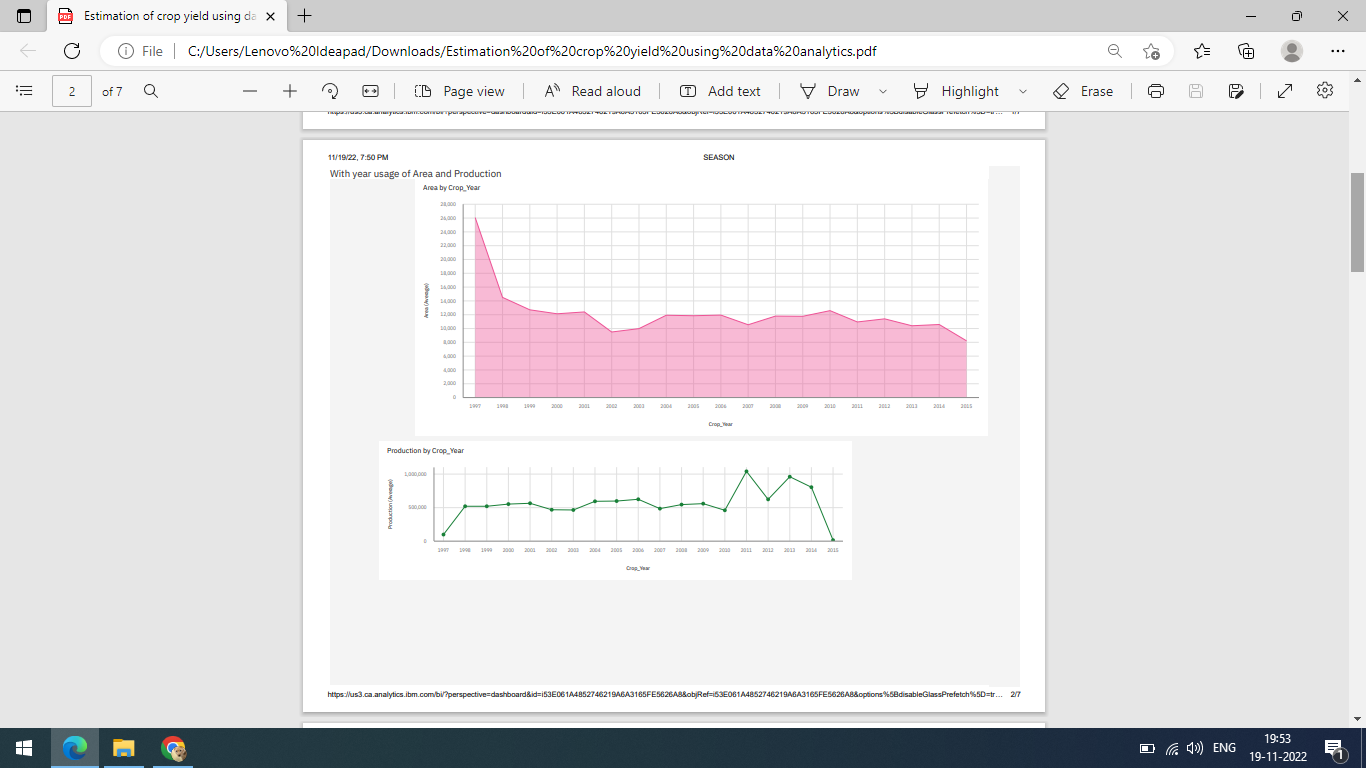


**7.3 Data Visualization charts**

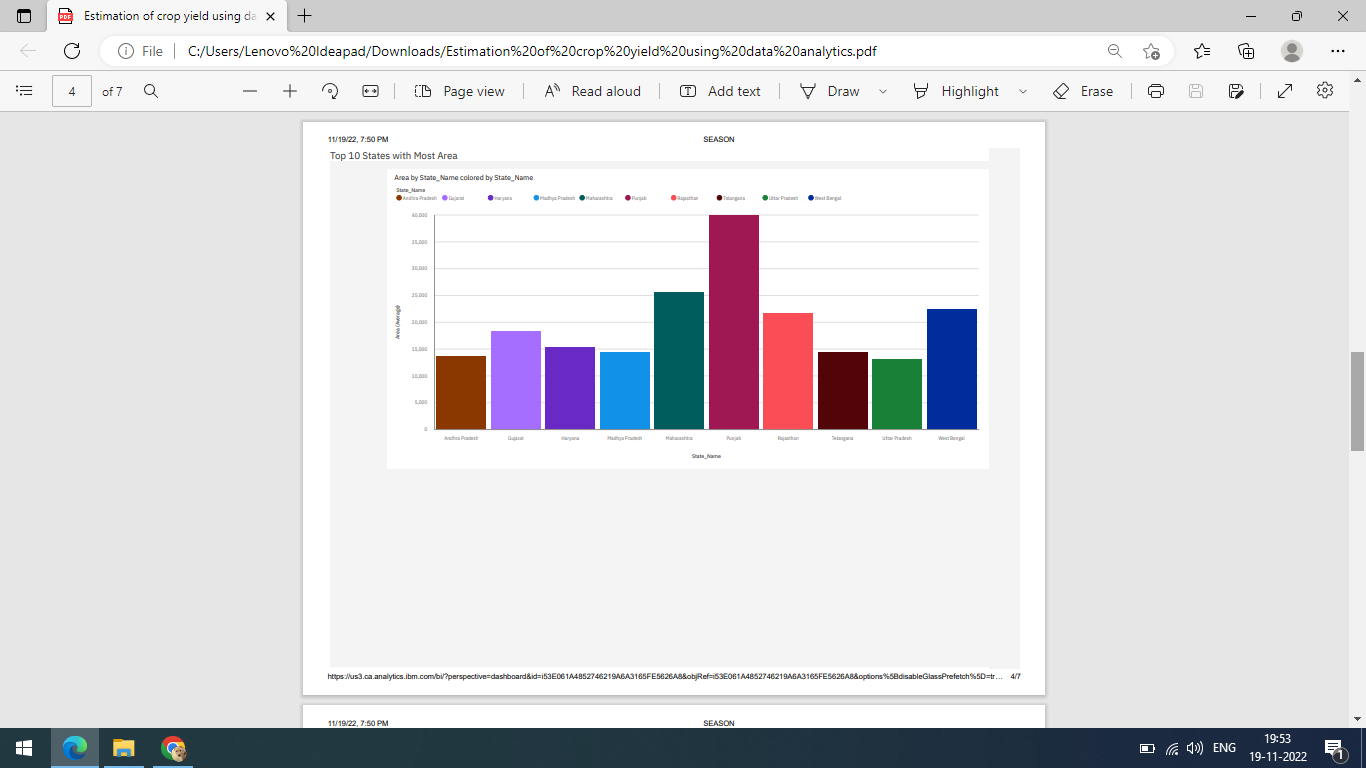
* **Seasons with average productions**



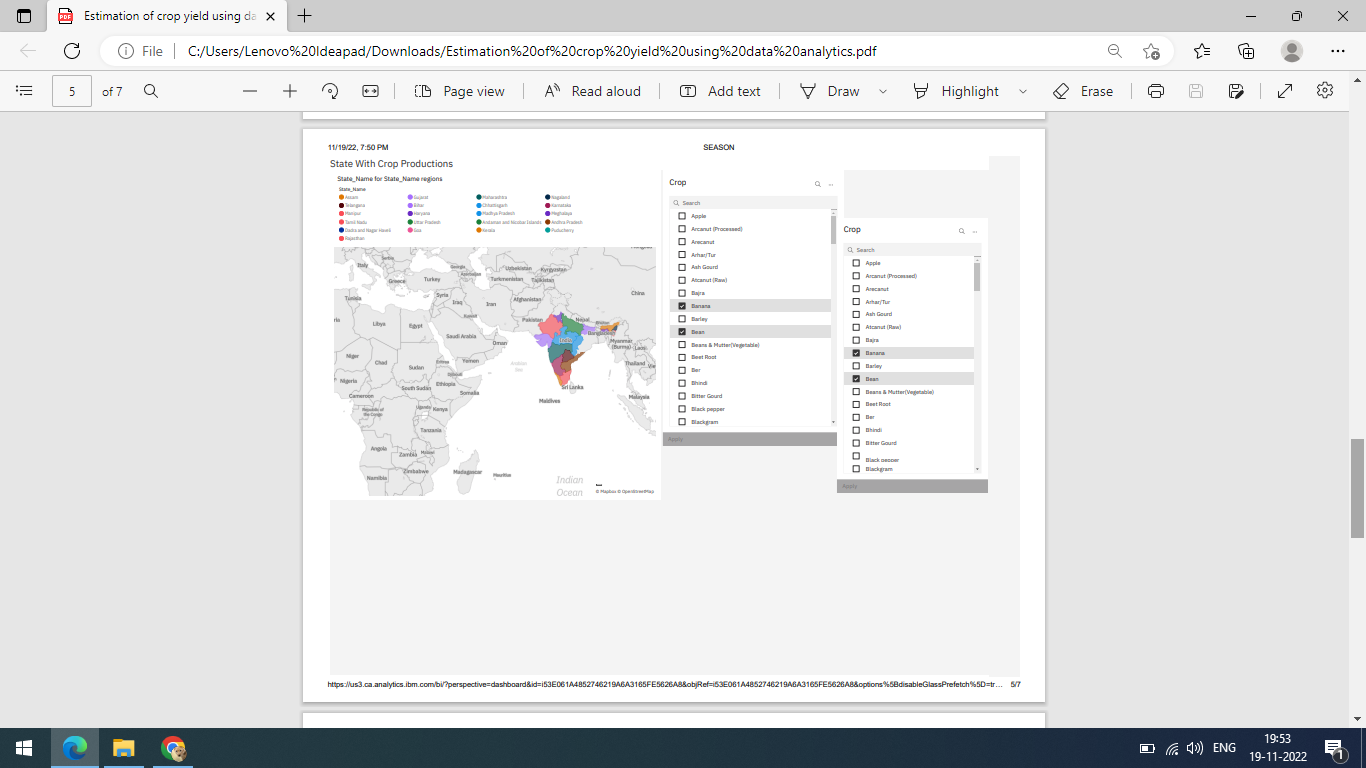
* **With years usage of Area and Production**



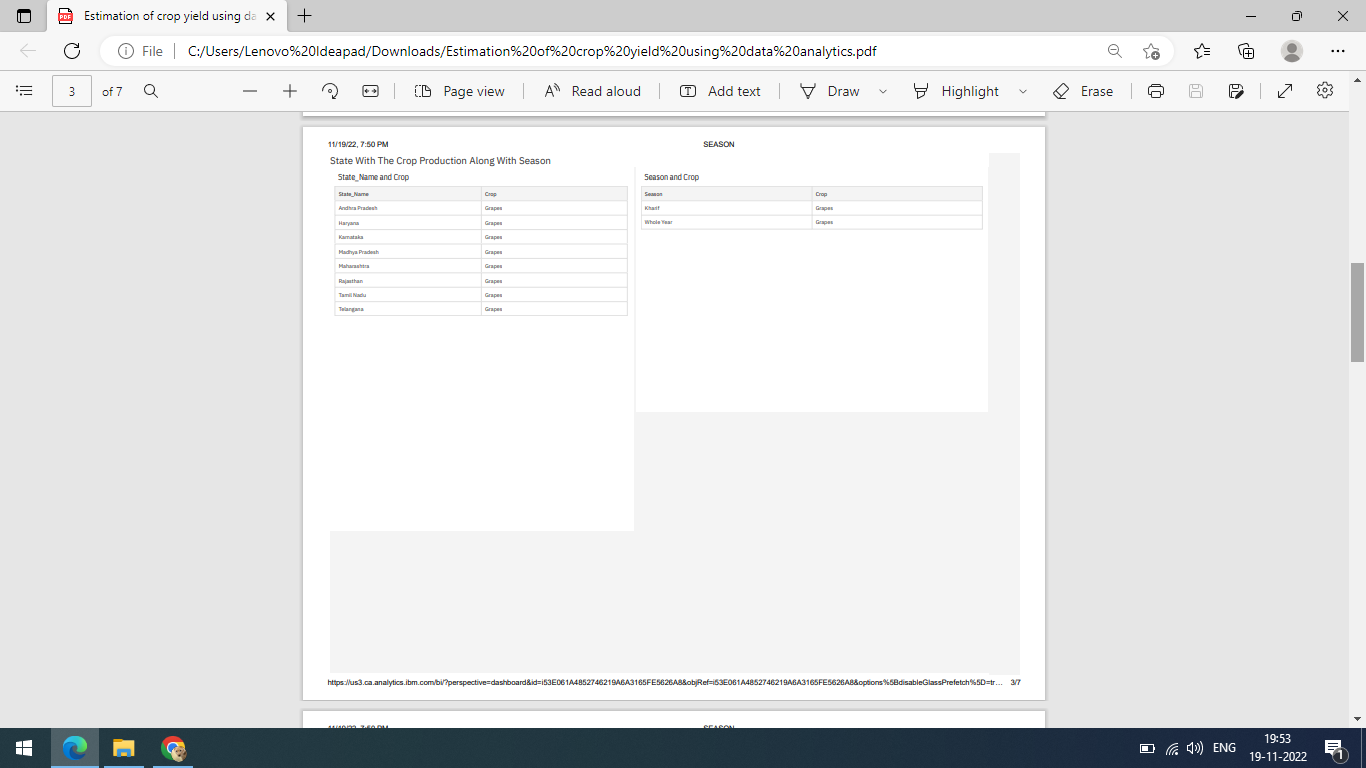
* **Top 10 States with most area**



* **State with crop production**

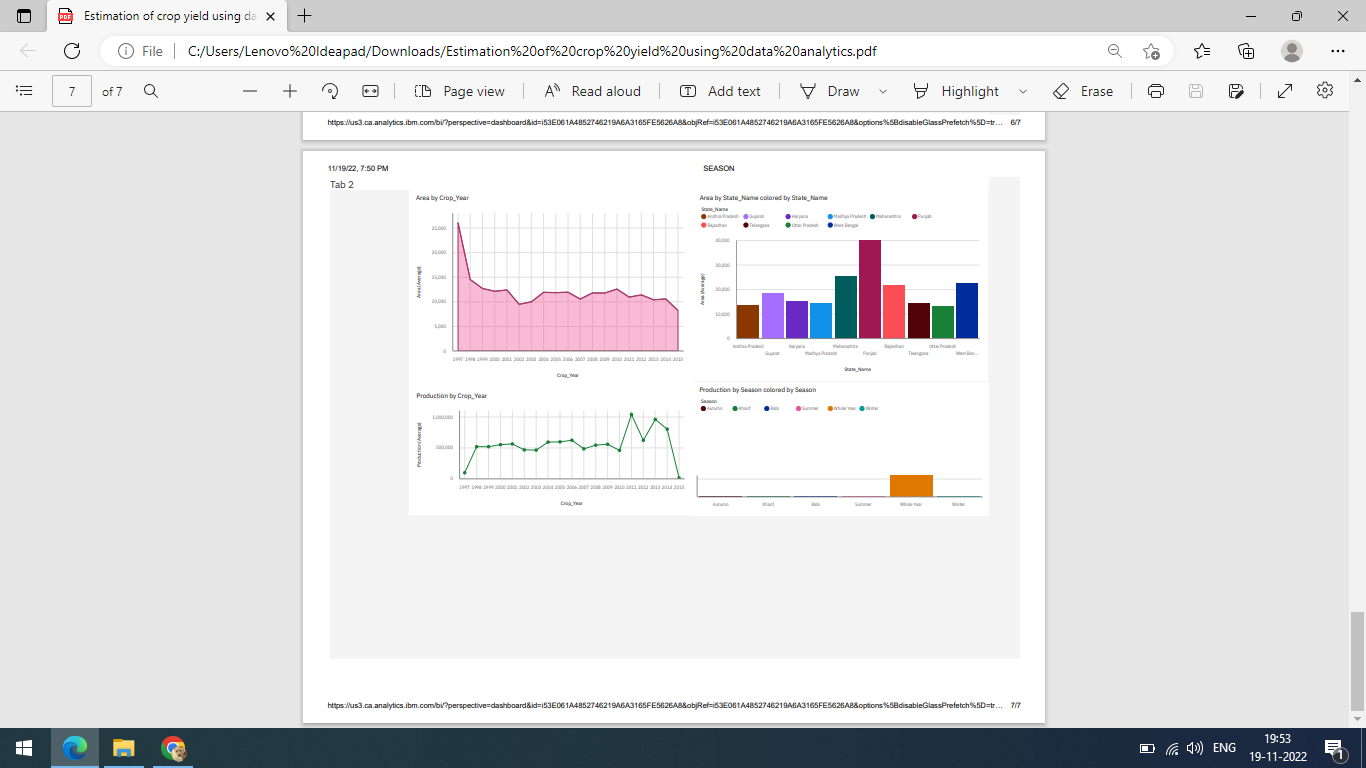


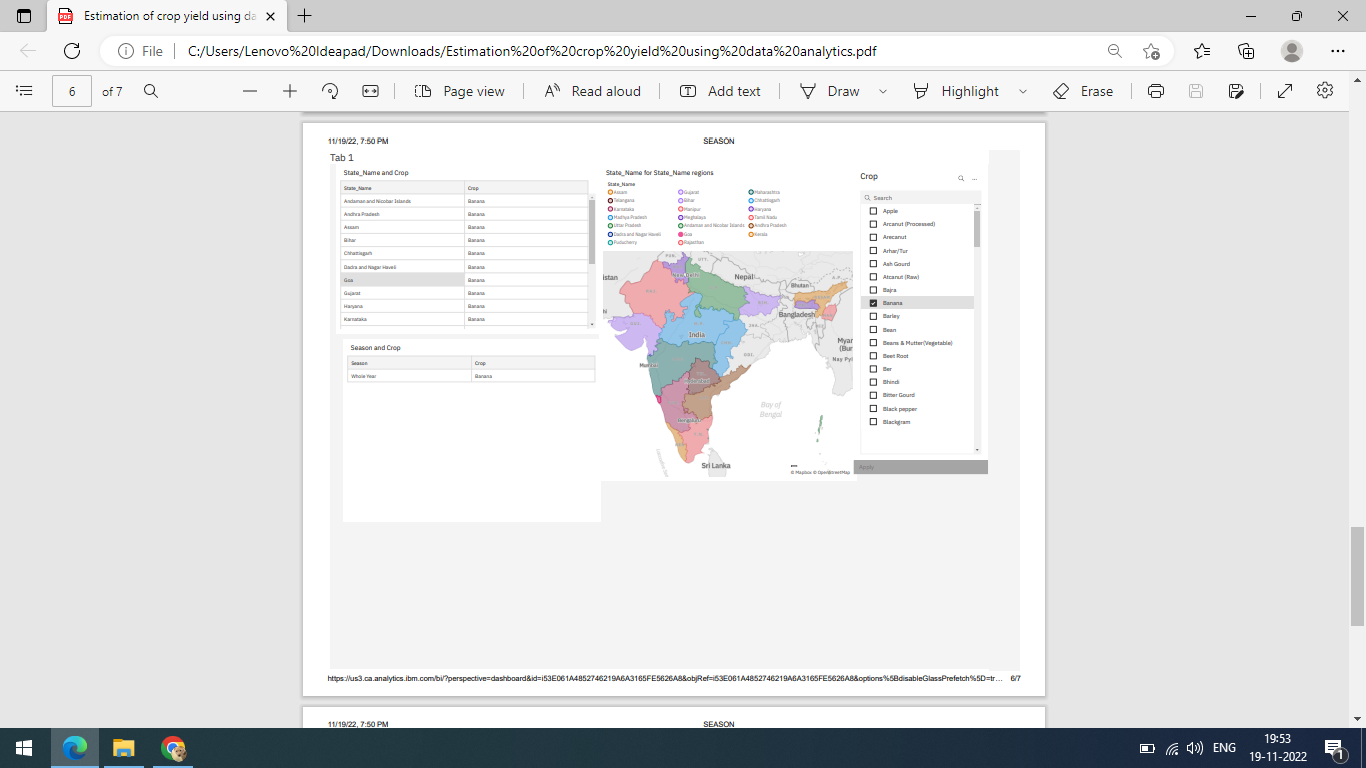
* **States with the crop production along with season (Text Table)**



# 8. CREATING THE DASHBOARD AND EXPORT THE ANALYTICS

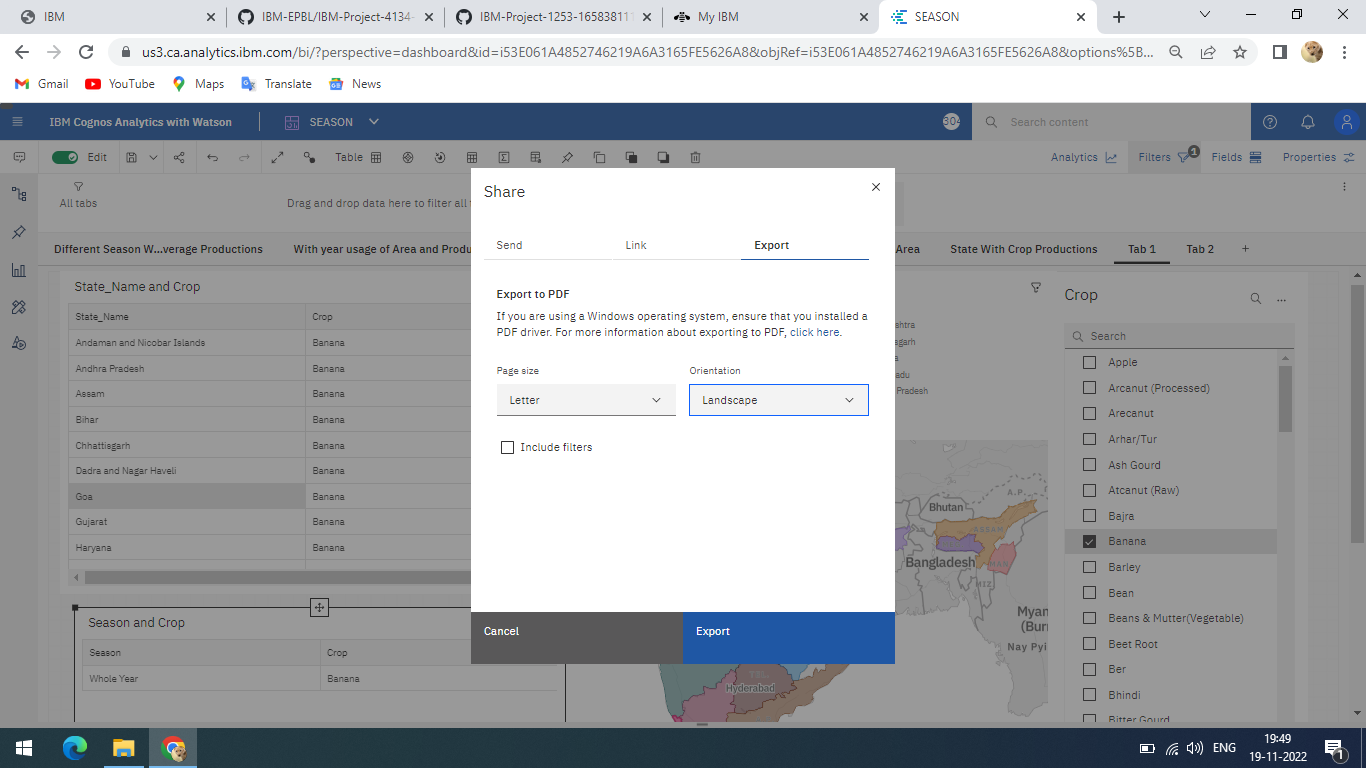
**8.1 Creating the Dashboard**



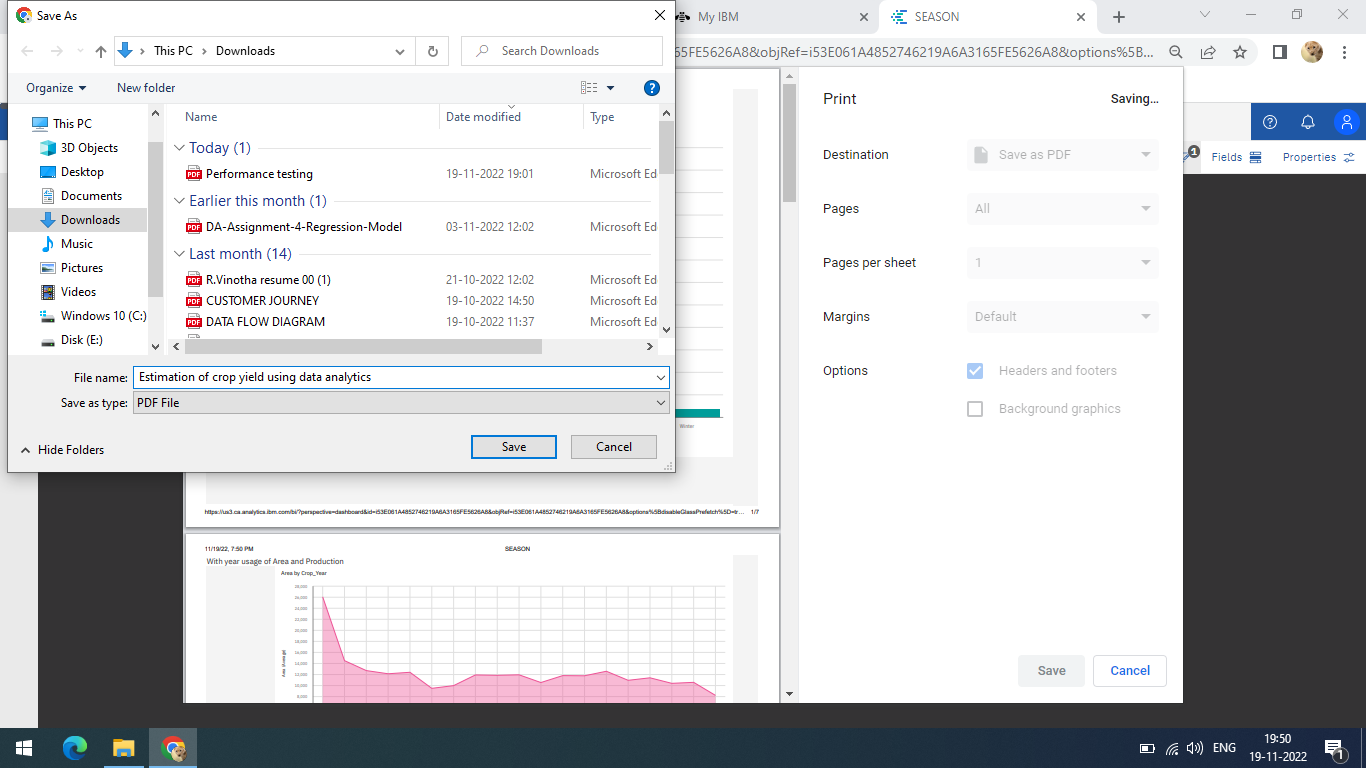


**8.2 Export the Analytics**

* **Click on the share icon**



* You can share using Email or Link or Export as pdf.
* Click the Export tab in the Share dialog box.
* You can change the page size and Orientation setting then click **Export**.



# 9. Advantages and Disadvantages Advantages:

# The resulting dashboard makes it easy to analyze and understand trends in land cultivation patterns and seasonal behavior in different regions.Anyone (literate or not) can easily infer the knowledge presented in various charts, graphs, or maps without prior skill or knowledge of the tools used for analysis. To help farmers make the right decisions in the future. Cons:

# Only visible factors were included in the analysis, so not all factors affecting crop yield are considered in the analysis.

# 10. CONCLUSION

Agricultural productivity has increased slightly as a result of the introduction of technology.

New ideas such as digital farming, smart farming and precision farming.

made possible by innovation. Analytics dashboards point out that agricultural productivity analysis and hidden pattern detection are related to the use of datasets

Seasons and harvests were carried out. Observed and analyzed using IBM Cognos

For different cultivated crops, regions and products in different states and districts, including:

1) Average production season. These analyzes tell you which seasons have high average production and which have low production. 2) Distribution of production by harvest year. This study tells us which years have high and low production.

3) Region-based production. These analyzes allow you to identify the states and counties where your selected crops are grown.

4) Production volume by region. This allows you to estimate yields and determine how much land you need to plant. After creating the dashboard, I looked at which states, in which years, and how much square footage was produced.

# 11. FUTURE SCOPE

Agriculture is a means of survival for people because they need food that can only be obtained directly or indirectly from agriculture. As the population grows, it is important to analyze agricultural production each year. You can know the right time, place and crops to grow, taking into account all the factors that affect crop production.

In summary, research in agriculture is still in its early stages in terms of leveraging IT trends such as data analytics. Food is a basic human need, so all you need is to get it.

In the near future, population growth will require optimal use of resources for maximum yields. Findings indicate that techniques for crop yield analysis need to be improved. There are many research areas in this research area.

# 12. APPENDIX

IBMCognosAnalytics: <https://login.ibm.com/authsvc/mtfim/sps/authsvc?PolicyId=urn:ibm:security:authentication:asf:basicldapuser&Target=https%3A%2F%2Flogin.ibm.com%2Foidc%2Fendpoint%2Fdefault%2Fauthorize%3FqsId%3D05e634c6-c305-4c6d-b3c6-f21c7b9635b4%26client_id%3DMyIBMLondonProdCI>

GitHub Link:

<https://github.com/IBM-EPBL/IBM-Project-4134-1658720742>

Project Demo Link: <https://vidmails.com/v/80KXcAJoi4>