Estimate the Crop Yield using Data Analytics

TEAM ID : PNT2022TMID10801

Team Leader : Dhivagar D

Team member : Abinesh Kumar V

Team member : Balaji Ragavendra Raj R

Team member : Gopinath S

1. ABSTRACT

Analytics is the study of data trends in order to improve performance and make decisions. Agriculture Data analytics in agricultural yield is useful for analysing specific important visualisations and creating a dashboard. We can gather the majority of information regarding crop yield in India by looking at these. We can understand the data in our organisation and make sound decisions by combining IBM Cognos Analytics with reporting, modelling, analysis, exploration, dashboards, storytelling, and event management. A dashboard allows us to keep track of events or activities at a glance by displaying crucial insights and analysis about our data on one or more pages or displays. The majority of the findings in this project are seen, analysed, and extracted using a dashboard.

2. INTRODUCTION

Agriculture is the backbone of the Indian economy. The majority of farmers in India are not obtaining the expected agricultural production for a variety of reasons. Weather has a significant influence on agricultural productivity. The amount of rainfall also has an effect on rice farming. Farmers in this circumstance would undoubtedly want early aid in forecasting future crop productivity, and an analysis will be performed to assist farmers in optimising crop production in their crops. Yield prediction is a key topic in agriculture. Every farmer wants to know how much output to expect. Previously, farmer expertise with a certain crop was included when estimating productivity. In Indian agriculture, there is a massive amount of data.Data may be turned into information and utilised for a number of reasons. Cognos Business Intelligence is an IBM web-based comprehensive business intelligence suite. It provides a package of tools for analytics, score carding, reporting, and event and data tracking. The software is made up of several parts that are designed to meet the various information needs of a business. IBM Cognos, for example, includes IBM Cognos Framework Manager, IBM Cognos, Cube Designer, and IBM Cognos Transformer. Cognos Analysis Studio enables business users to receive prompt responses to commercially relevant questions.

3. LITERATURE SURVEY

We are currently in desperate need of another Green Revolution to meet the food demands of a burgeoning population. Higher agricultural yields are nearly difficult to report with the global decline in cultivable area and cultivable water supplies. Agricultural-based big data analytics is one approach that is thought to play a significant role and have a positive impact on crop yield increase by providing the best conditions for plant growth and decreasing yield gaps, crop damage, and wastage.. With this goal in mind, the current research analyses the numerous breakthroughs, design models, software tools, and algorithms used in crop yield prediction, evaluation, and estimate. India is primarily an agricultural country, with agricultural crops accounting for over 70% of our country's economy. Paddy cultivation is the primary crop that occupies the biggest (60-70%) proportion of cultivable area in Indian soil, and it is the principal crop, particularly in central and southern India. Rice crop farming is critical to India's food security, accounting for more than 40% of total yield output. The increased yield of the rice crop is heavily dependent on water availability and climatic conditions. Low precipitation or temperature extremes, for example, can significantly reduce rice yield. Growing improved ways to predict yield efficiency in a variety of climatic situations can aid in understanding the impact of several primary components that influence rice crop production. Big data analytic methods for rice crop yield prediction and estimation will undoubtedly help farmers understand the optimum condition of the significant factors for rice crop yield, allowing them to achieve higher crop yield.

a. Crop Yield Prediction Using Machine Learning

A study group studied the use of several information mining approaches to forecast rice crop production for data obtained in the Indian state of Maharashtra. A total of 27 Maharashtra regions were chosen for the assessment, and data on the main rice crop yield influencing parameters such as different atmospheric conditions and various harvest parameters such as precipitation rate, minimum, average, maximum, and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield were collected from open sources for the years 1998 to 2002.WEKA, a Java-based dialect programming for less difficult help with information data sets, was used for dataset processing, and the overall methodology of the study involves (1) pre-processing of the dataset, (2) building the prediction model with WEKA, and (3) analysing the results. A cross validation research is conducted to see how a predictable information mining system would perform on an ambiguous dataset. To evaluate data subsets for screening and testing, the study used a 10-fold greater cross validation study design. The identified and gathered information was randomly allocated into ten sections, with one data section used for testing and the remaining data sections used for preparation information.Agriculture is one of India's key economic generators and a source of survival. Various seasonal, economic, and biological factors impact crop output, yet unpredictability in these aspects results in significant losses for farmers. These hazards may be quantified by employing appropriate mathematical and statistical model designs on soil, weather, and previous yield data. With the advent of data mining, crop production may be anticipated by drawing helpful insights from agricultural data, assisting farmers in deciding which crop to grow for the upcoming year in order to maximise profit. There are various systems that use various data mining technologies to manipulate data in order to derive insights and aid farmers in decision making.The current data mining techniques and algorithms are focused on a single crop and predict or forecast a single parameter such as yield or price. A study used to anticipate the production and price of Tamil Nadu's primary crops based on historical data is presented in a paper that gives a survey on the various algorithms used for crop yield prediction. Farmers may view the data and estimated production via a web application. This assists farmers in deciding which crop to sow for the future year. Furthermore, the web application provides a forum for farmers to sell their products without the use of middlemen, allowing them to obtain the best possible price for their products.

b. . Crop Yield Prediction Using Data Mining Techniques

India's primary economic resource is agriculture and agricultural-related businesses. It is also one of the countries that suffers from catastrophic natural disasters such as drought or flood, which destroy crops and inflict large financial losses for farmers as well as the country's economic stability. Predicting crop yields well in advance of harvest can assist farmers and government organisations in making appropriate planning decisions such as storing, selling, establishing minimum support prices, importing/exporting, and so on. Predicting a crop well in advance necessitates a systematic examination of vast amounts of data derived from various variables such as soil quality, pH, essential elements (N,P,K) quantity, and so on. Because crop prediction works with a huge number of databases, this prediction system is an ideal candidate for the use of data mining approaches, which greatly aids in obtaining information to attain improved crop production. The performance of any crop yield prediction system is strongly dependent on how correctly the characteristics are retrieved and how well classifiers are used. The study includes the findings of numerous algorithms employed by various authors for agricultural production prediction, along with their accuracy and recommendations.

'Weeds and pests were the principal crop damaging biotic agents, and farmers must be well informed in order to gain information on applications of efficient weed and pest control methods and crop management practises to decrease crop loss.' The major challenges in weed and pest control to protect crop damage are the collection of data related to various weeds and pests, modelling of the data to prepare for mining, selection of appropriate methodology, interpretation, and sharing of information.

A study was carried out to assess the primary problems, notable potential, and uses of Big Data in minimising weed and insect damage and thereby achieving improved agricultural output.According to the study, the form of the data collected, the type of assessment method, and the tools used are the major influencing factors in understanding the role of crop damaging agents such as weeds and pests, which provides knowledge on using improved crop management and crop yield prediction strategies.

Big Data cargo space and querying poses significant issues in terms of allocating data across several platforms, as well as continually developing data from various sources. When selecting data from various sources, semantic approaches play an important part in the evaluation, detecting elements with potential agricultural value and building linkages between data items in terms of meanings and units. The study offered a success story from the Netherlands in applying Big Data analytics and numerical algorithms to prevent crop damage and achieve better agricultural output. According to the study, the applicability and uses of Big data analytics for weed and pest management are extensive, particularly for invasive, parasitic, and herbicide-resistant weeds.Also introduced was the necessity for agricultural scientists to collaborate with data scientists in order to deploy approaches for the benefit of agricultural operations.

Data mining is critical for making decisions on many issues concerning agricultural operations. The goal of data mining methods is to extract information from a readily available data collection and transform it into a usable format for any meaningful application of the Agri process. Crop management in a certain agricultural region is dependent on the region's climatic circumstances since climate has a large influence on crop yield. Real-time weather data can assist in optimal crop management. The effective use of mined agricultural-based information and communications expertise allows for the automation of retrieving useful data in an effort to acquire knowledge, which allows for easier data acquisition from electronic sources directly, transfer to a secure electronic system of documentation, and dissemination.. In the field of agricultural bioengineering, scientists and engineers have worked together to develop and discuss the application of mathematical model designs such as fuzzy logic designs in crop yield optimization, artificial neural networks in validation studies, genetic algorithm designs in accessing the fitness of the model applied, decision trees, and support vector machines to assess soil, climate conditions, and water resource availability. The study outlines the use of data mining technologies such as neural networks, support vector machines, big data analysis, and soft computing in agricultural field evaluation depending on meteorological conditions.

c. Crop yield prediction using Big Data Analytics

Crop output in India is seasonally dependent and heavily impacted by each individual crop's biological and economic reasons. Reporting progressive agricultural production in all seasons is a large and beneficial undertaking for every nation in terms of assessing total crop yield forecast and estimation. Farmers are currently pressured in delivering larger agricultural yields due to the effect of unexpected climate fluctuations and considerable reductions in water resources globally. A research was conducted to collect data on global climate changes and available water resources, which may be utilised to support advanced innovative techniques prediction and estimation. such as big data analytics to recover information on prior crop production outcomes.According to the study, the selection and use of the most desirable crop based on the existing conditions helps to achieve a higher and improved crop yield.

Accurate crop yield forecast surely helps farmers choose the best approach to prevent crop damage and earn the best price for their crops. A research group conducted a study with the goal of accurately predicting crop yield using big data analytics. The study assessed various crop yield influencing factors such as Area under Cultivation (AUC) interims of hectors, Annual Rainfall (AR) rates, and Food Price Index (FPI) and developed relationships between these parameters. The Regression Analysis (RA) approach was used to investigate the selected parameters' influence on crop forecast and ultimate yield. RA methodology is a multivariable investigation practise that classifies factors into groups such as explanatory and response variables and aids in assessing their interaction to obtain a resolution. All of the selected components of the current research design, known as AR, AUC, and FPI, were assessed throughout a 10-year period between 1990 and 2000. Linear Regression (LR) is a revolutionary approach for analysing the association between explanatory factors (AR, AUC, FPI) and crop production as the response variable. According to the study, the R2 value for the analysed parameters clearly suggest that crop output is mostly determined by AR. The study also indicated that the other two screening parameters (AUC and FPI) had a significant influence following the AR. The study will be expanded to examine the influence of other significant elements such as the Minimum Support Price (MSP), Cost Price Index (CPI), Wholesale Price Index (WPI), and so on and their relationship on the yields of different crops .

Crop yield gaps are defined as the discrepancy between predicted yields based on potency and actual farm output. Farmers must address influencing variables such as the impact of changing climate conditions on crop yield prospects and changes in agricultural land management to analyse and eventually eliminate crop yield gaps in order to attain better crop yields. Several studies reported using bio modelling models to predict agricultural yield gaps during the previous decade.. The quality and resolution of climate and soil data, as well as unscientifically held expectations about crop yield prediction systems and crop yield assessment modelling designs calibration method, all had a negative impact on the impact of crop yield gap assessment studies conducted using bio simulation-based methodologies. An explicit logic model that can be effectively utilised at various degrees of quality information availability for finding data sources to assess agricultural production and evaluating yield gaps at specific geographical regions and is based on the rise in titer method. The model is extremely beneficial in recovering relevant data from poor quality, less rigorous data sources, or if the data is not accessible.A case study on the deployment of a selected model design to measure maize crop yield gaps in the state of Nebraska (USA), as well as at several geographical areas representing Argentina and Kenya at the national scale level, was explored. Various geographical sites, including Nebraska (USA), Argentina, and Kenya, were chosen to represent the various situations of Agri-based data availability and quality for the selected variables analysed to forecast and estimate agricultural output gaps. The intended method's ultimate goal is to provide transparent, simply available, reproducible, and technically solid and powerful guidance for anticipating yield gaps.The recommended principles were also useful for understanding and simulating the impact of changes in climatic conditions on the use of cultivable land at national and global scales. As previously said, a better knowledge of data relevance and usefulness for assessing agricultural yield and predicting yield gaps, as depicted, may aid in detecting data gaps in crop yield and allow focusing on various worldwide initiatives to address the most significant issue.

Crop yield analysis is required to adjust policy to guarantee food security. A research group performed a study with the goal of proposing an unique data mining approach to estimate crop yields based on agricultural big data analytics methods, which were gradually contrasted with standard data mining methodologies in the process of data handling and modelling designs. According to the study, the technology utilised should be user pleasant, function on a progressive big-data responsive processing structure, utilise current agricultural-based substantial datasets, and be used with higher quantities of data rising at gigantic rates.Nearest neighbours modelling is one such unique data mining approach that works on the results acquired from farmers based on data processing structures and suggests a well unbiased outcome based on accuracy and forecast time in advance. The paper went on to give a case study on the evaluation of actual agricultural datasets (numerical examples on) in China from 1995 to 2014. According to the study, the innovative model used has published a better performance and was discovered to be progressive in reporting prediction accuracy % of the comparable techniques with traditional designs [7].

Simulation models based on field experiments are essential tools for researching and analysing agricultural production gaps, but one key problem with these approaches is scaling up to examine data collected across many time periods from a larger geographical region. Satellite data has repeatedly been shown to offer data sets that, by themselves or in combination with other information and model designs, may exactly estimate crop yields on agricultural fields. The yield maps created will give a one-of-a-kind chance to overcome both geographical and temporal scaling up obstacles, hence improving agricultural yield gap forecast ideology.A evaluation of the applications of remote sensing technologies to identify the impact and causes of yield gaps was undertaken. Although the study group's example indicates the use of remote sensing in the prediction of production gaps, many areas of potential application in agricultural yield evaluation, prediction, and improvement remain unexplored. The study offered two less difficult, simply assessable approaches for determining and quantifying production differences between distinct agricultural areas.

. The first method works closely with constructive maps that represent average crop yields and can be used directly to access specific crop yield influencing factors for further studies, whereas the second method uses remote sensing technology to retrieve data for providing useful information about crop yield prediction and estimation. The two most major and crucial elements identified to impact crop yield in the next decades are a growth in world population and economy, which considerably needs greater and sustainable agricultural-based crop yields.The global food production capacity will be severely constrained due to a lack of cultivable land and water resources, difficulty in sustaining sustainable crop production levels, the consequences of changes in global meteorological conditions, and many biophysical elements that influence crop output. Farmers must be educated on the use of scientifically proven methods to quantify crop yield capacities, and this information must be shared with higher authorities in order to maintain transparency in sharing actual information. Interns assist in making policy-based, research-oriented, development, and investment-related decisions that aim to influence future crop yield.Crop production abilities and yield gaps may be analysed and estimated by comparing possible yields under normal conditions to crop production under irrigated and rain fed situations, respectively, using crop yield levels restricted by water availability as benchmarks. Yield gaps are defined as the difference between predicted crop yields and actual crop yields, and precise, geographically unambiguous knowledge and information about output gaps is required to ensure sustainable agricultural yield amplification.A study group conducted a survey on the many ways used to estimate yield gaps with the goal of examining the influence of the various approaches used in assessing yield gaps with a focus on the local-to-global relevance of outcomes. The study reported on a few standard operating procedures used in evaluating agricultural production potential based on data acquired from farmers in western Kenya, Nebraska (USA), and Victoria (Australia). The study advocated for the use of reliable and recent yield data assessed by calibrated crop model designs and further up scaling validated approaches in agricultural yield gap prediction. Bottom-up implementation of this global protocol enables for verification of predicted production gaps using on-farm data and tests.

d. **M. A. Jayaram and Netra Marad, "Fuzzy interference Systems for Crop Prediction"**, Journal of Intelligent Systems, 2012, 21(4), pp.363-372[1]. Prediction of crop yield is significant in order to accurately meet market requirements and proper administration of agricultural activities directed towards enhancement in yield. Several parameters such as weather, pests, biophysical and morphological features merit their consideration while determining the yield. However, these parameters are uncertain in their nature, thus making the determined amount of yield to be approximate. It is exactly here that the fuzzy logic comes into play. This paper elaborates an attempt to develop fuzzy inference systems for crop yield prediction. Physio morphological features of Sorghum were considered. A huge database (around 1000 records) of physio morphological features such as days of 50 percent? powering, dead heart percentage, plant height, panicle length, panicle weight and number of primaries and the corresponding yield were considered for the development of the model. In order to? nd out the sensitivity of parameters, one-to-one, two-to-one and three-to-one combinations of input and output were considered. The results have clearly shown that panicle length contributes forth yield as the lone parameter with almost one-to-one matching between predicted yield and actual value while panicle length and panicle weight in combination seemed to play a decisive role in contributing for the yield with the prediction accuracy rejected by very low RMS value.

P. Vindya "Agricultural Analysis for Next Generation High Tech Farming in Data Mining" ,

Anna University, Trichy, Tamil Nādu, India, 5 May 2015. Recent developments in Information Technology for agriculture field have become an interesting research area to predict the crop yield.

4. IDEATION AND PROPOSED SOLUTION

The application of systems theory to the production of a project is known as system design. The system design defines the architecture, data flow, use case, class, sequence, and activity diagrams of a project's development.

IBM Cognos Analytics

IBM Cognos Analytics is a set of business intelligence tools that is available both on-premises and in the cloud. The primary focus is on descriptive analytics, which use dashboards, expert reporting, and self-service data exploration to assist users in understanding the information in your data. We used IBM cognos data analytics to analyse agricultural yield data in this study.

Following are important features of IBM Cognos:

1. Get Connected - Connect your data effortlessly Import data from CSV files and spreadsheets. Connect to cloud or on-premises data sources, including SQL databases, Google BigQuery, Amazon, Redshift, and more.
2. Prepare your data – Prepare and connect data automatically Save time cleaning your data with AI-assisted data preparation. Clean and prep data from multiple sources, add calculated fields, join data, and create new tables.

3. Build visualizations - Create dynamic dashboards easily Quickly create compelling, interactive dashboards. Drag and drop data to create auto- generated visualizations, drill down for more

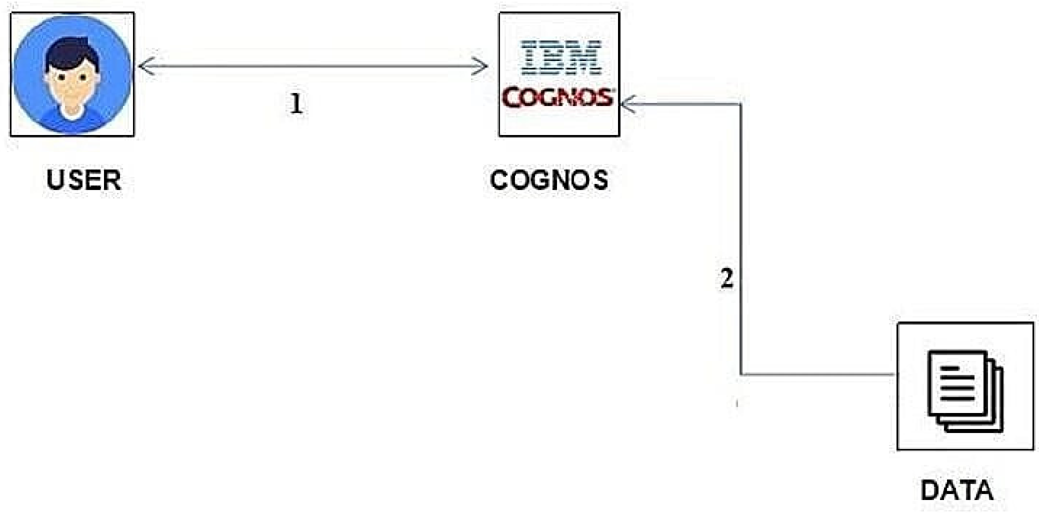
detail, and share using email or Slack.

4. Identify Patterns – Uncover hidden patterns Ask the AI assistant a question in plain language, and see the answer in visualization. Use time series modelling to predict seasonal trends.
5. Generate Personalised Reports – Create and deliver personalized reports Keep your stakeholders up-to-date, automatically. Create and share dynamic personalized, multipage reports in the formats your stakeholders want.
6. Gain Insights - Make confident data decisions Get deeper insights without a data science background. Validate what you know, identify what you don't with statistically accurate time-series forecasting and pinpoint patterns to consider.
7. Stay Connected – Go Mobile Stay connected on the go with the new mobile app.

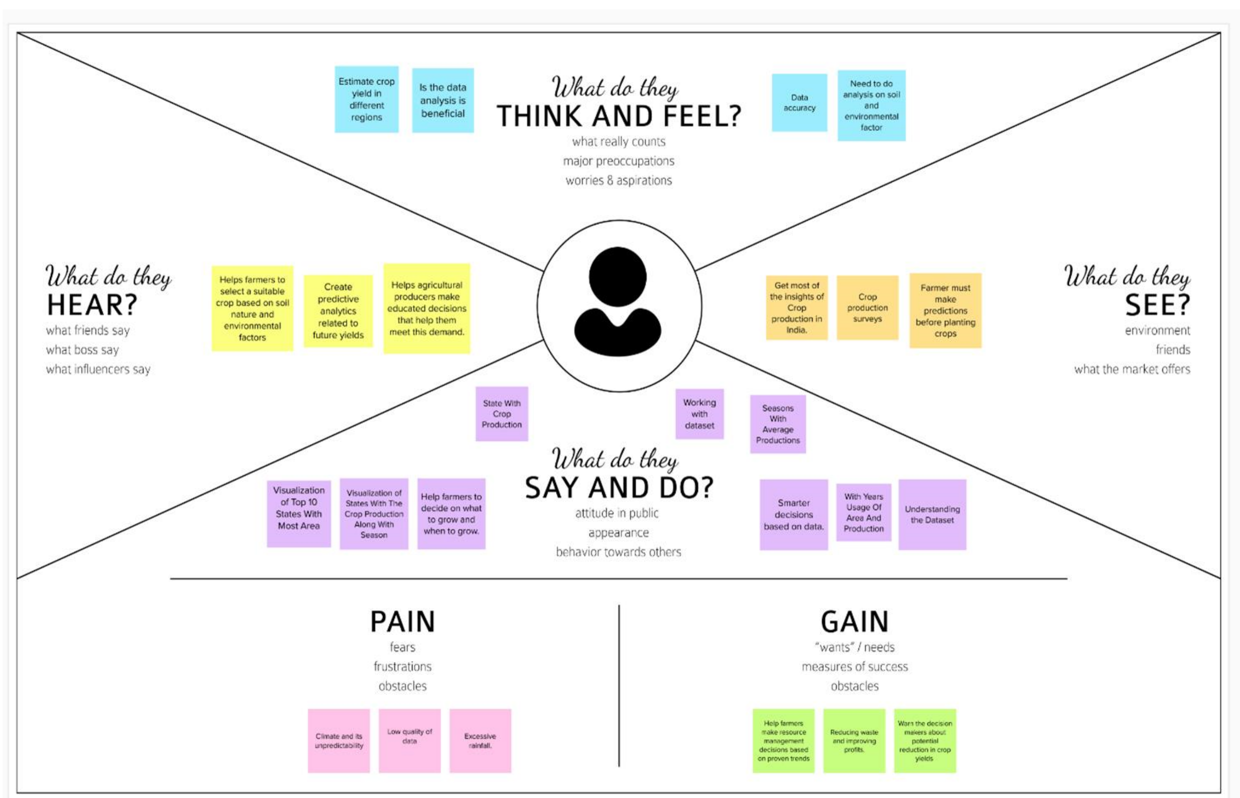
System Architecture

India is one of the top countries for agricultural production, making it one of the most significant sources of income. As part of this project, we will analyse some significant visualisations, build a dashboard, and then use this information to gain the majority of our understanding of crop output in India.

Technical Architecture:



EMPATHY MAP



5.REQUIREMENTS ANALYSIS

IBM Cognos Analytics

A collection of business intelligence tools called IBM Cognos Analytics is offered both on premises and in the cloud. The main emphasis is on descriptive analytics, which uses dashboards, expert reporting, and self-service data exploration to help users understand the information in your data. In this study, we analysed the crop yield data using IBM cognos data analytics.

Following are important features of IBM Cognos:

1. Get Connected - Connect your data effortlessly Import data from CSV files and spreadsheets. Connect to cloud or on-premises data sources, including SQL databases, Google BigQuery, Amazon, Redshift, and more.
2. Prepare your data – Prepare and connect data automatically Save time cleaning your data with AI-assisted data preparation. Clean and prep data from multiple sources, add calculated fields, join data, and create new tables.
3. Build visualizations - Create dynamic dashboards easily Quickly create compelling, interactive dashboards. Drag and drop data to create auto- generated visualizations, drill down for more detail, and share using email or Slack.
4. Identify Patterns – Uncover hidden patterns Ask the AI assistant a question in plain language, and see the answer in visualization. Use time series modelling to predict seasonal trends.
5. Generate Personalised Reports – Create and deliver personalized reports Keep your stakeholders up-to-date, automatically. Create and share dynamic personalized, multipage

reports in the formats your stakeholders want.

6. Gain Insights - Make confident data decisions Get deeper insights without a data science background. Validate what you know, identify what you don't with statistically accurate time-series forecasting and pinpoint patterns to consider.
7. Stay Connected – Go Mobile Stay connected on the go with the new mobile app.

B. System Architecture

India is one of the top countries for agricultural production, making it one of the most significant sources of income. As part of this project, we will analyse some significant visualisations, build a dashboard, and then use this information to gain the majority of our understanding of crop output in India.

IBM® Cognos® Analytics integrates reporting, modeling, analysis, dashboards, stories, and event management so that you can understand your organization data, and make effective business decisions.

After the software is installed and configured, administrators set up security and manage data sources. You can get started yourself by uploading local files and applying visualizations in dashboards or stories. For enterprise-level data, modelers are next in the workflow. After data modules and packages are available, report authors can then create reports for business users and analysts. Administrators maintain the system on an ongoing basis.

Whether you're an analyst, report author, data modeler, or an administrator, you start by signing in to the Welcome portal from your desktop or mobile device.

6. PROJECT DESIGN

PROJECT FLOW

1. Users create multiple analysis graphs/charts.
2. Using the analyzed chart creation of the Dashboard is done.
3. Saving and Visualizing the final dashboard in the IBM Cognos Analytics.
4. To accomplish this, we have to complete all the activities and tasks listed below.
5. IBM Cloud Account
6. Login to Cognos Analytics
7. Working with the Dataset
8. Understand the Dataset
9. Loading the Dataset
10. Data visualization charts
11. Seasons with average productions
12. With years usage of Area and Production

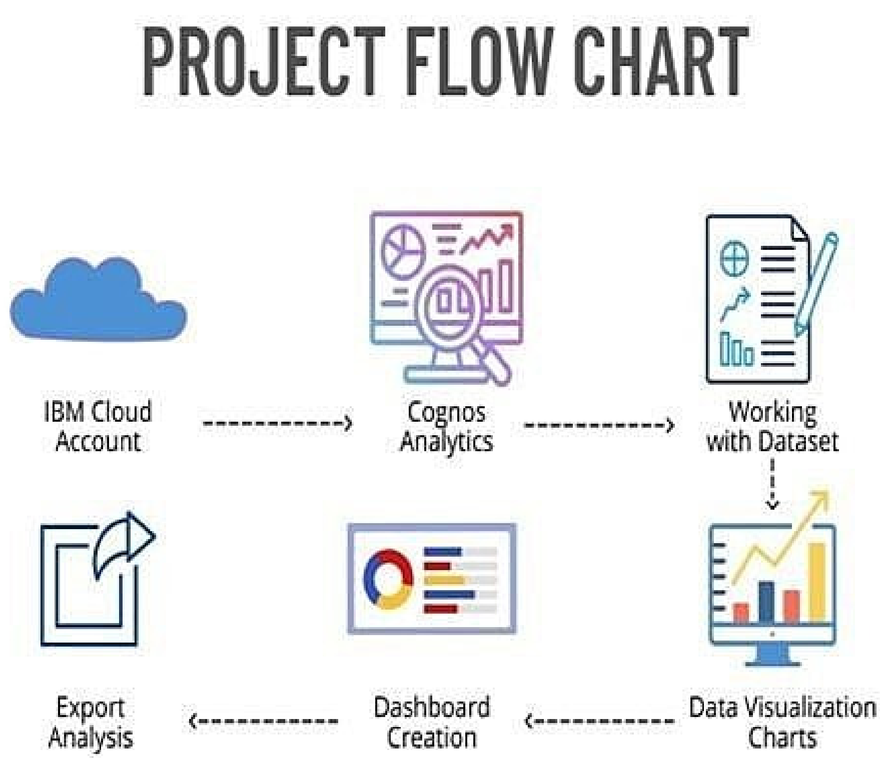
13.Top 10 States with most area

14.State with crop production

15.States with the crop production along with season (Text Table)

16.Dashboard Creation

17.Export the Analytics



SOLUTION REQUIREMENTS

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Sub-Task)
1	**User Registration**	Registration through Whatsapp Registration through Gmail Registration through Mail Registration through Agri-Consultancy modes.
2	**User Confirmation**	Confirmation via Email Confirmation via OTP through SMS. Confirmation via physical Letter.
3	**User Profile**	User Details Farm Details
4	**Required Data**	The past crop yield data and data of the Farmer to a their yield.
5	**Analysis**	Clean , Analyze the data by means of set of past data multiple users which is farmers.

NonFunctional Requirements:

FR No.	Non-Functional Requirement	Description
1	**Usability**	The data report is created according to the past data By considering these recommendation the sowing o will be decided.
2	**Security**	IBM Cognos have a high-secure user information.
3	**Reliability**	The interactive data visuals of the dashboard can easy to understand by the farmers.

7 .RESULTS

7.1 Seasons With Average Productions

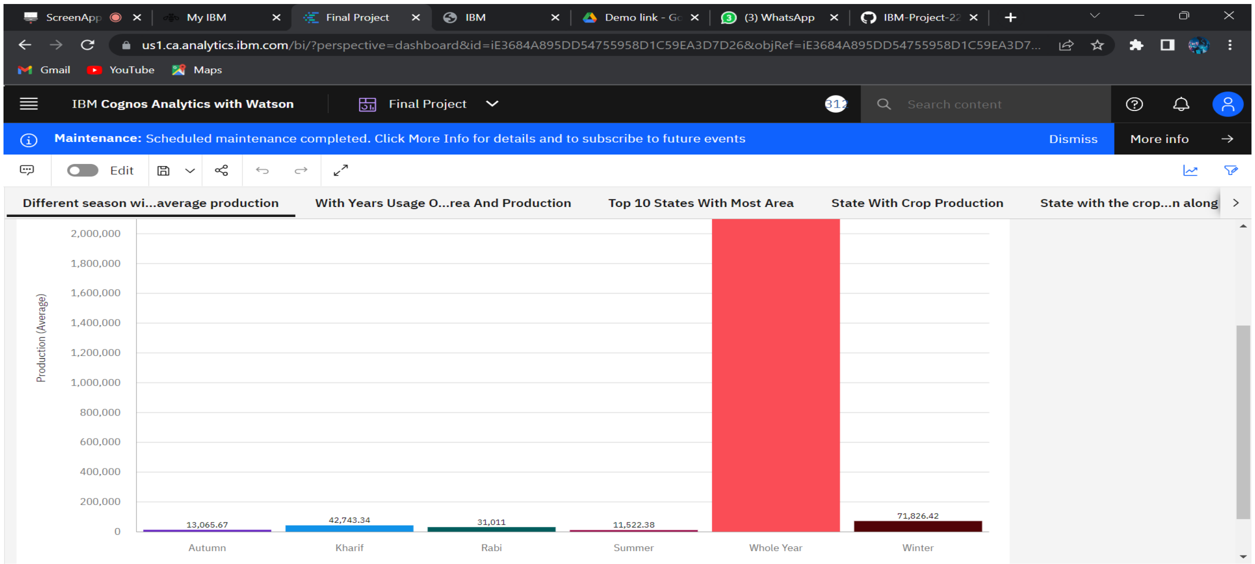


Fig 7.1

7.2 With Years Usage Of Area And Production

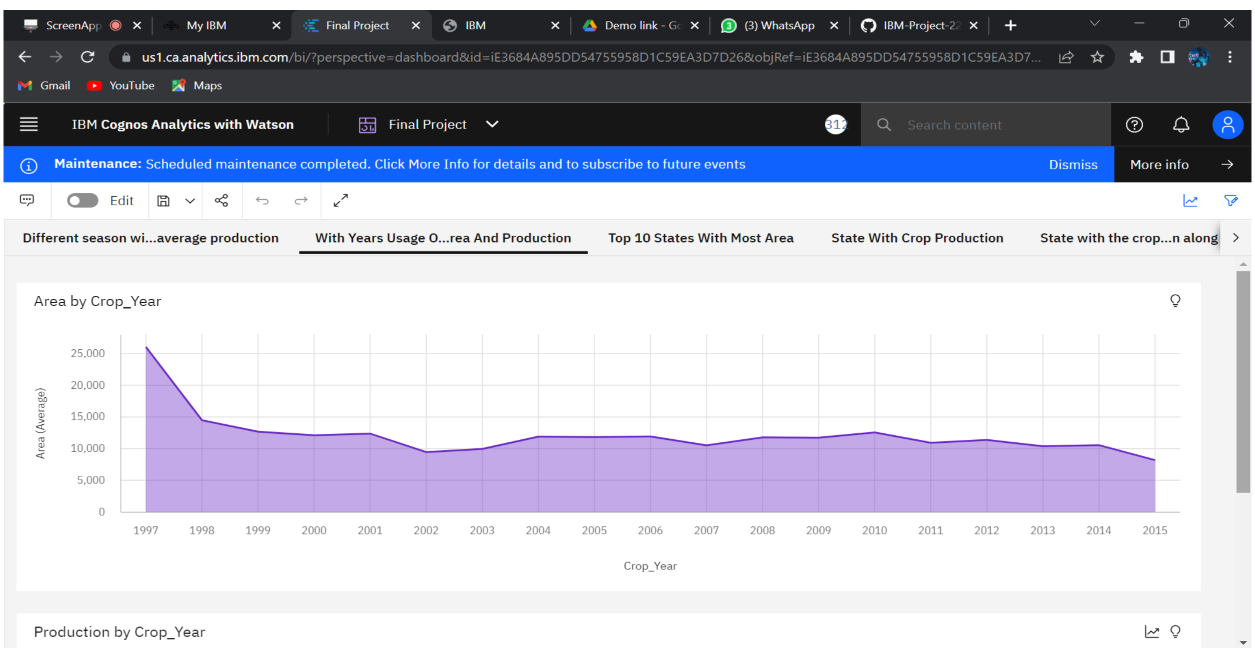


Fig 7.2

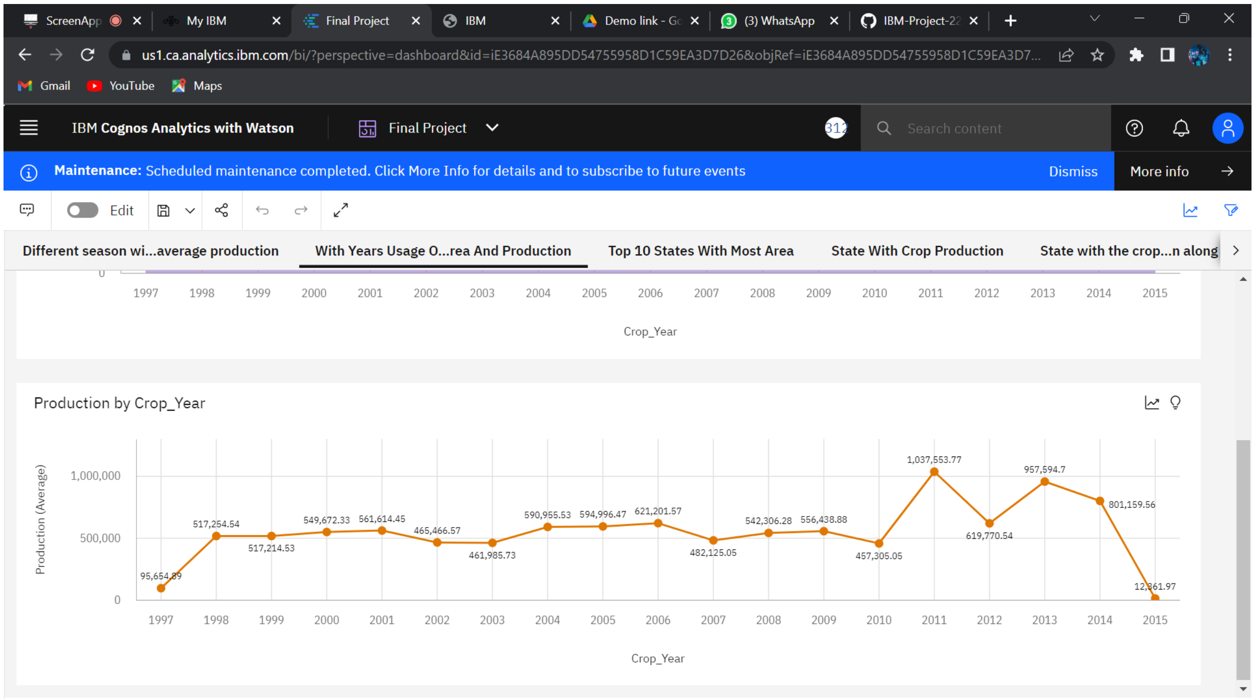


Fig 7.3

7.3 Top 10 States With Most Area

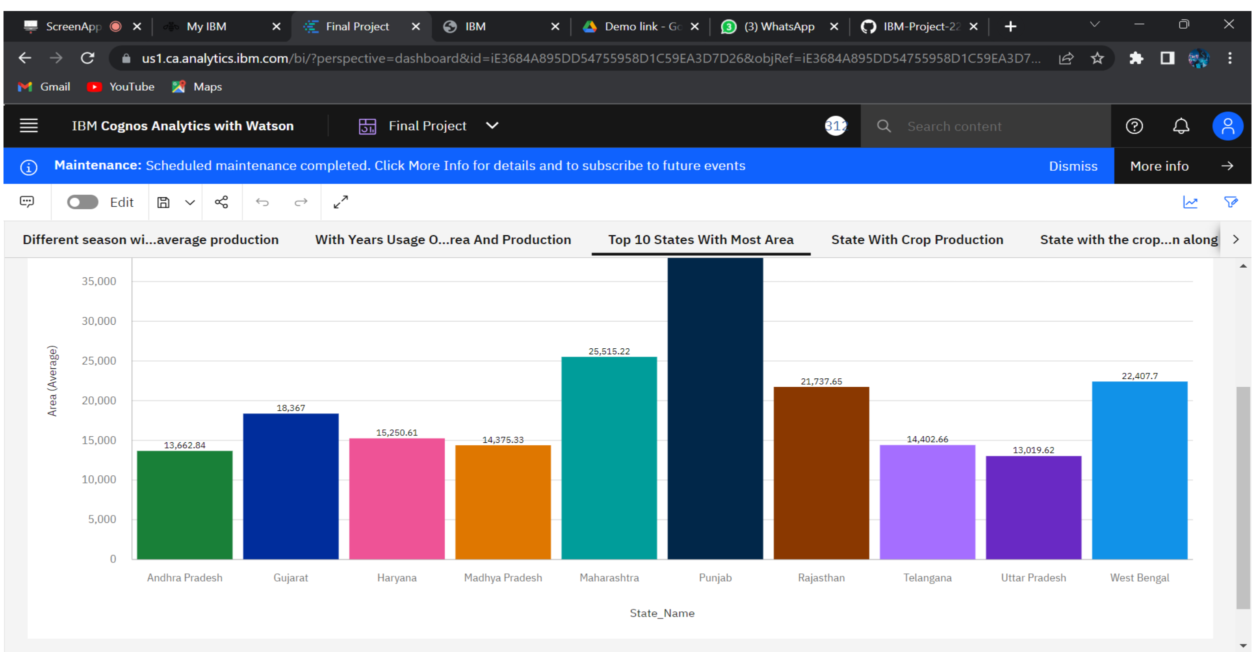


Fig 7.4

7.4 State With Crop Production

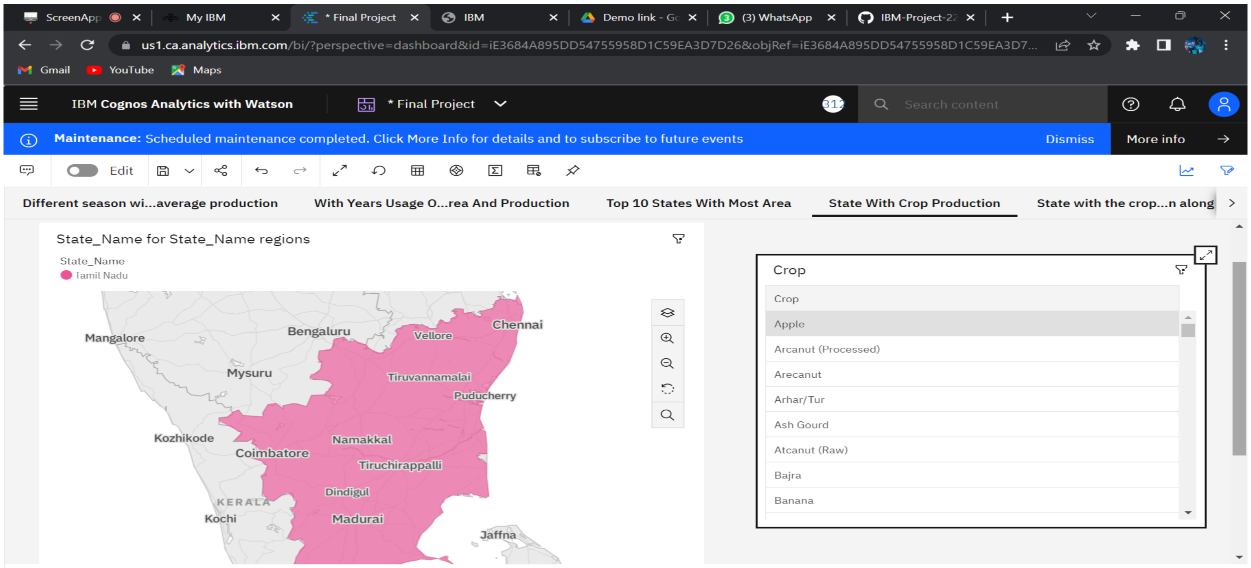


Fig 7.5

7.5 States With The Crop Production Along With Season (Text Table)

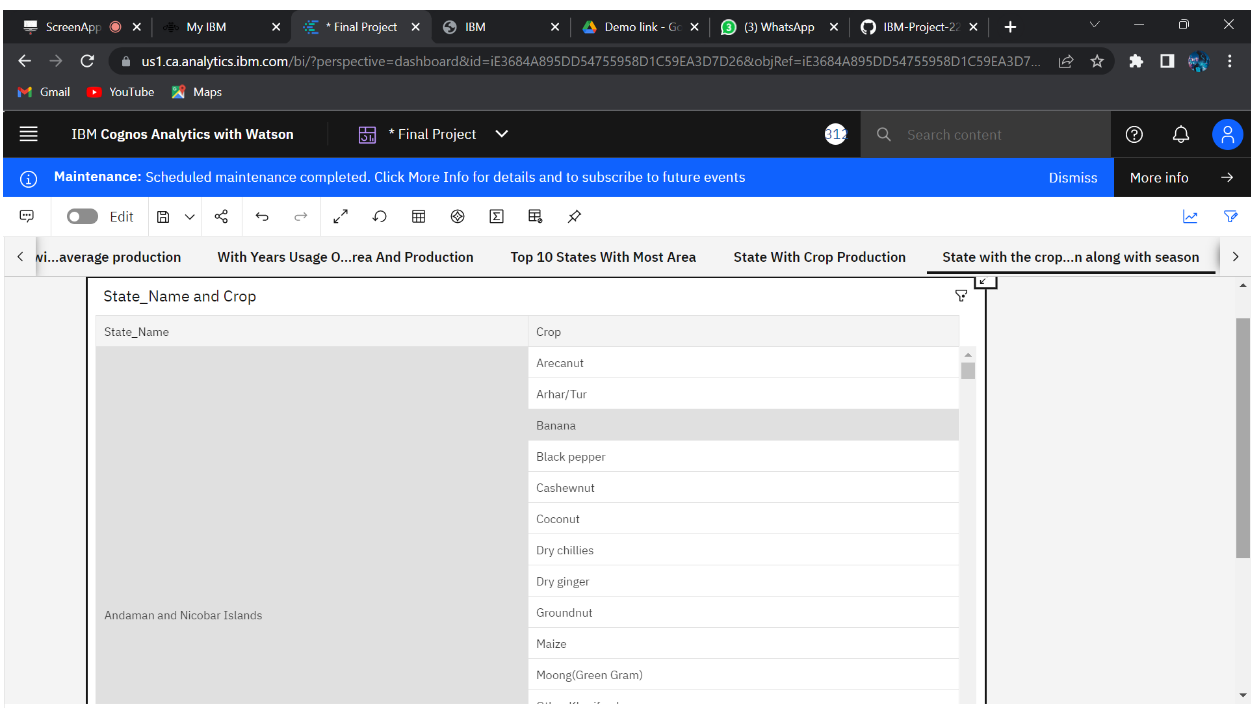


Fig 7.6

7.6 Final Output Overview

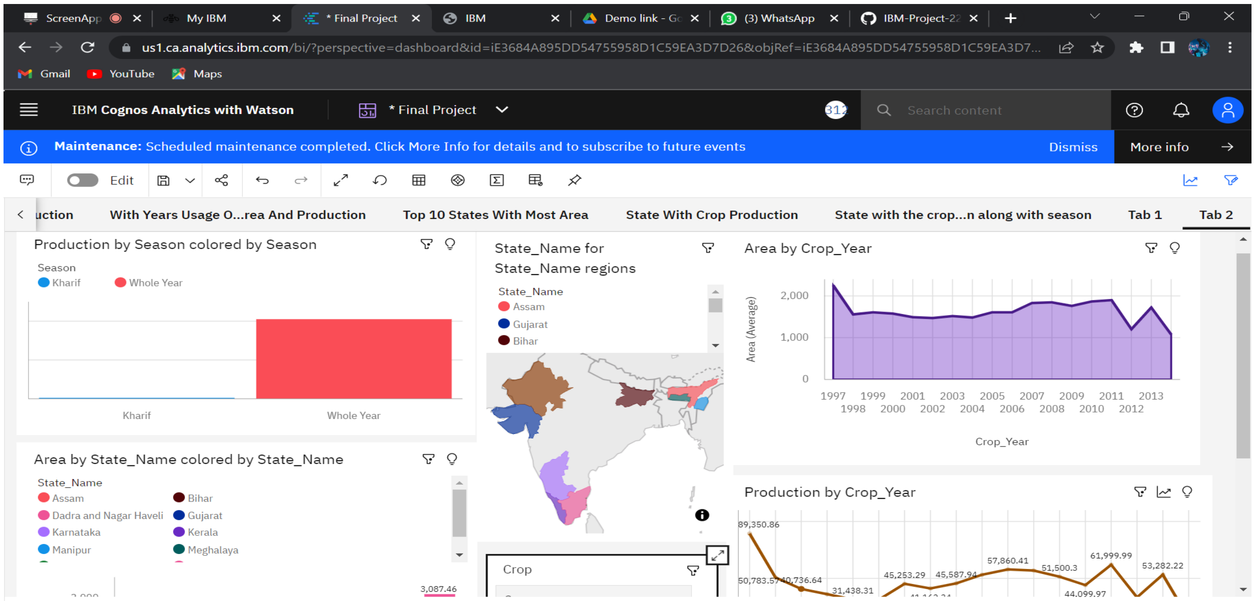


Fig 7.7

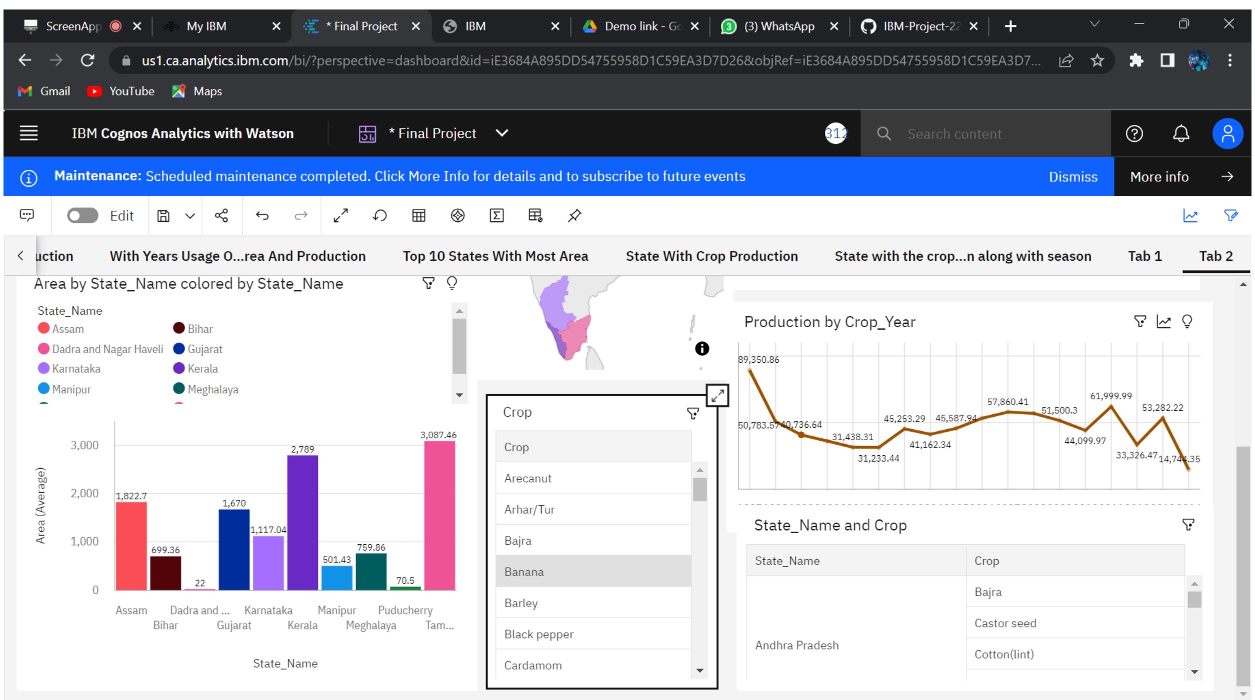


Fig 7.8

8. 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. The analysis of agricultural productivity and the uncovering of hidden patterns utilising data sets related to seasons and crop yields have been noted in the literature. Using IBM Cognos, we have observed and conducted analysis regarding various crops grown, areas, and productions in various states and districts.

Demo Link

https://drive.google.com/file/d/1n44rkzSmIIFfcC81C8wrERrjCn9te8iB/view?usp=share_link