Agriculture Data Analytics in Crop Yield Estimation using IBM Cognos

Prepared by,

Santhosh Kumar C,

Assistant Professor,
Department of Information Technology,
Sona College of Technology, Salem,
Tamil Nadu, India.
E-Mail: santhoshkumar.it@sontech.ac.in

1. **INTRODUCTION**
1.1 Overview

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

1.2 Purpose

Agriculture is the main pillar of the Indian economy and more than 60% of India's population are dependent on agriculture for their survival. Variations in weather, climate, and other such environmental conditions have become a major risk for the healthy existence of agriculture. The Cultivation process is usually planned by the farmers based on their previous experiences. Sometimes they end up in cultivating undesirable crops as their predictions are not precise. Crop yield prediction is an essential task for the decision-makers at all the levels for rapid decision-making. An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. There are different approaches to crop yield prediction. Innovations in Business Intelligence can bring a boom in the income by changing the scenario through cultivating the most optimal crops in the season. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. There are many factors influencing crop yield. Since analysing numerical data as it is would be complex, creating a dashboard that has the visualizations of the entire crop cultivation details would provide more insights on crop production.

2. **LITERATURE SURVEY**
2.1 Existing problem

Data visualization is the practice of representing information in a visual form and making it easy to understand the real-time trends and business insights present in the data. Python is a general-purpose programming language commonly used for data visualization in the data science community.

● Matplotlib, seaborn, plotly, bokeh, and many more best graphing packages are available in Python for data visualization. These help in creating interactive and highly customizable plots.

● Python has a large community and a vast number of in-built modules.

● Python provides all you need for accurate, appealing, and intelligible graphics when combined with add-ons.

As per the current requirement of data visualization there are many players in the Business Intelligence field such asTableau, Microsoft Power BI, Qlik Sense, etc.. Tableau software is a popular visualization tool that can also perform business intelligence. It simplifies the data and shows it in an easily understandable format.

The platform can generate results from any kind of data—structured, unstructured, or semi-structured. Besides, Tableau online is capable of creating complex visualizations and toggling between various representations. It uses natural language to answer the users' queries.

Cons

- An expensive solution in terms of scaling up.

- No option for scheduling or automatic report refreshing.

- Offers a limited number of columns by default.

- Requires SQL knowledge for creating complex datasets from multiple data resources.

- Tableau is compatible with Windows and Mac environments only. You can run the on-premise application on Windows 8 and later, macOS 10.14 and later.

- Whether Mac or Windows, the CPU should support POPCNT instruction sets and SSE4.2. Tableau is not compatible with RHEL OS.

There are different forecasting methodologies developed and evaluated by the researchers all over the world in the field of agriculture.

In general, basic programming knowledge is mandatory to apply data visualization using Python programming or basic Database knowledge is required to work with other integrated environments.

2.2 Proposed solution

Data Analytics is necessary approach for accomplishing practical and effective solutions for Crop yield Prediction from the large set of data. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions.

Given a framework that gathers the necessary data, the decision making to be performed requires knowledge extraction from these data. Towards this direction, we can use IBM Cognos, a web based integrated business intelligence suite by IBM. It is an analytical tool that integrates reporting, modelling, analysis, dashboards, stories and event management so that effective business decisions can be made.

IBM Cognos reporting tool mainly deals with analytics and data monitoring. Powered by IBM, this robust Business Intelligence (BI) platform comes with various dashboards and reports to display data insights. It offers a smart and self-service analytical solution to the users that can query, track, and process analytical data. Due to the attractive dashboards, visualizations, ease of use, and forecasting capabilities, Cognos BI is highly popular among users.
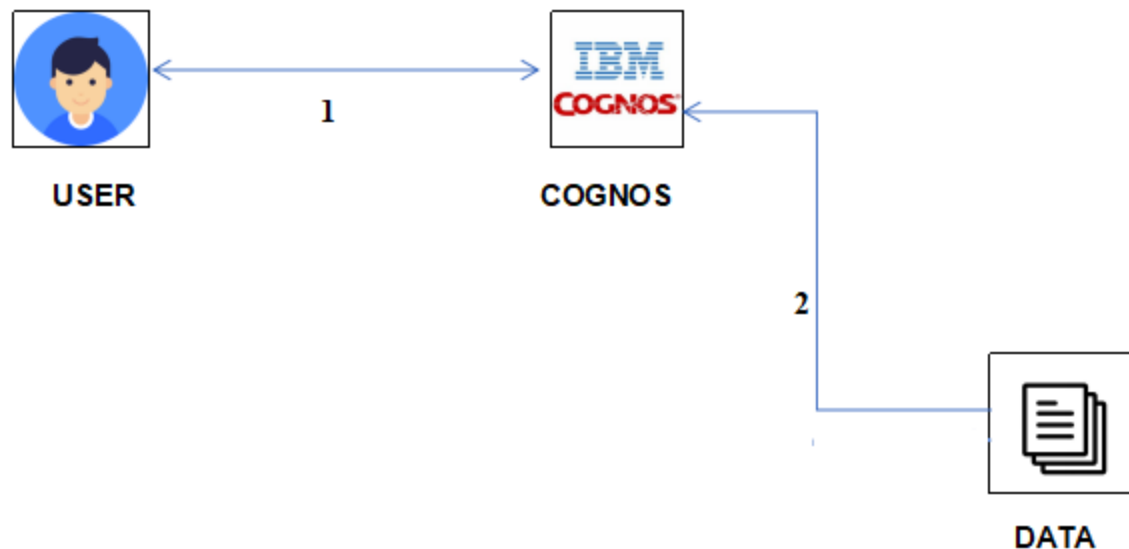
3. THEORITICAL ANALYSIS
3.1 Block diagram

Figure 1. Block Diagram

3.2 Hardware / Software designing

- IBM Cloud
- IBM Cognos

Cognos supports various operating systems such as:

- IBM Cognos Analytics for Mobile and Reports require Android 10 and later; iOS 14 and later; iPad OS 14 and later.

- The latest Cognos Analytics on-premise software requires Windows 8 or later; Windows Server 2012 and newer, Ubuntu 16.04 and later, Red Hat Enterprise Linux (RHEL) 8, Red Hat Enterprise Linux (RHEL) Server 7, and IBM AIX 7.1 and later versions.

4. EXPERIMENTAL INVESTIGATIONS

- The Crop Production dataset provides a huge amount of information on crop production in India ranging from several years.

- It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

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

5. FLOWCHART
Project Flow

- Users create multiple analysis graphs/charts.
- Using the analyzed chart creation of Dashboard is done.
- Saving and Visualizing the final dashboard in the IBM Cognos Analytics.

To accomplish this, we have to complete all the activities and tasks listed below

- IBM Cloud Account
- Login to Cognos Analytics
- Working with the Dataset
    - Understand the Dataset
    - Loading the Dataset
- 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)
- Dashboard Creation
- Export the Analytics

The flow chart given below describes the activities and tasks to be done:
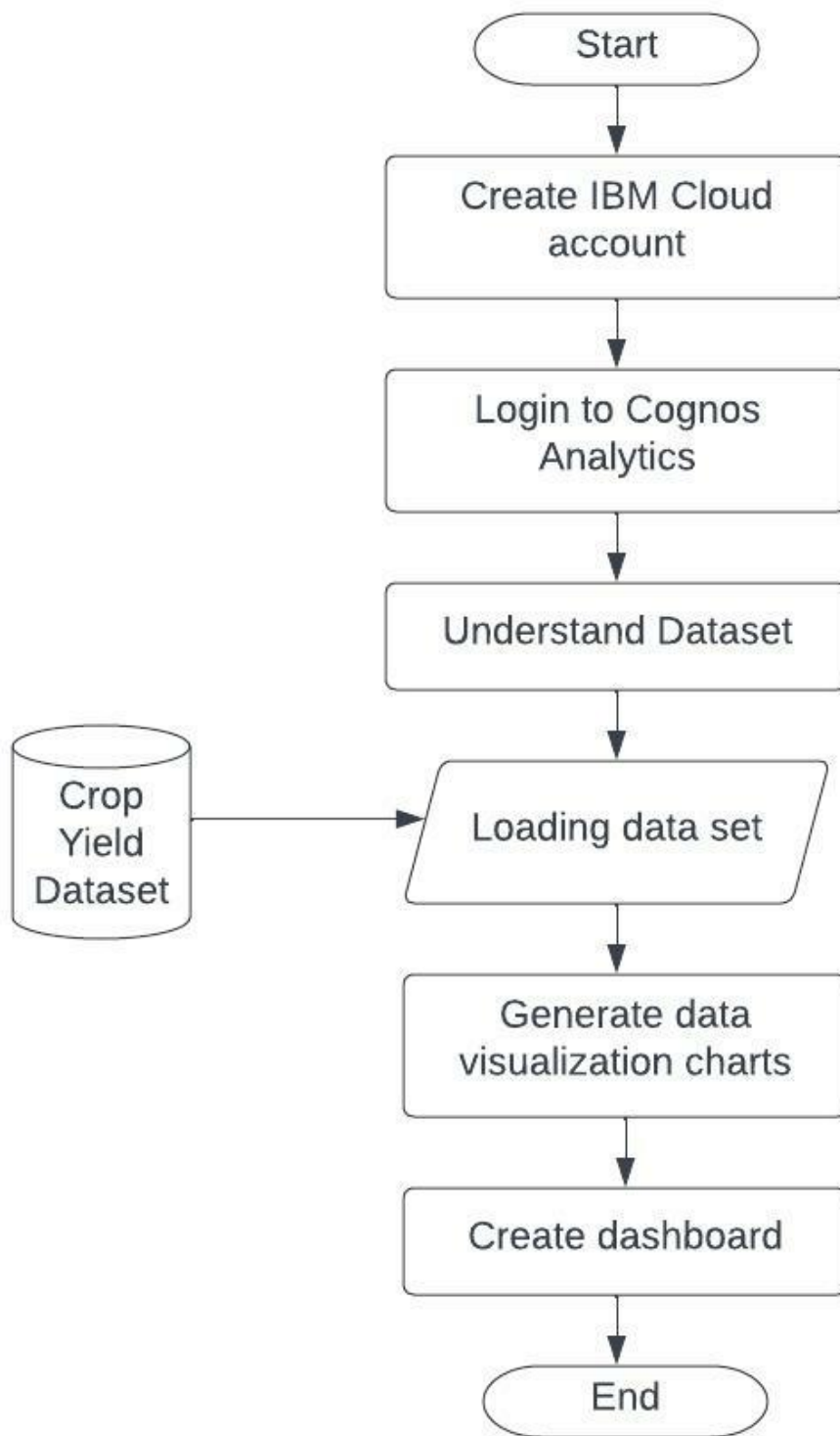


```
          ┌───────────┐
          │   Start   │
          └─────┬─────┘
                ▼
       ┌──────────────────┐
       │ Create IBM Cloud │
       │     account      │
       └────────┬─────────┘
                ▼
       ┌──────────────────┐
       │ Login to Cognos  │
       │    Analytics     │
       └────────┬─────────┘
                ▼
       ┌──────────────────┐
       │ Understand Dataset│
       └────────┬─────────┘
                ▼
  ┌──────┐   ┌──────────────────┐
  │ Crop │──▶│ Loading data set │
  │ Yield│   └────────┬─────────┘
  │Dataset│          ▼
  └──────┘   ┌──────────────────┐
             │  Generate data   │
             │visualization charts│
             └────────┬─────────┘
                      ▼
             ┌──────────────────┐
             │ Create dashboard │
             └────────┬─────────┘
                      ▼
                ┌───────────┐
                │    End    │
                └───────────┘
```

Figure 2. Flow Chart

6. RESULTS

The Data visualizations results obtained from Crop Production dataset using IBM Cognos is presented below:

a. **Seasons with Average Productions**
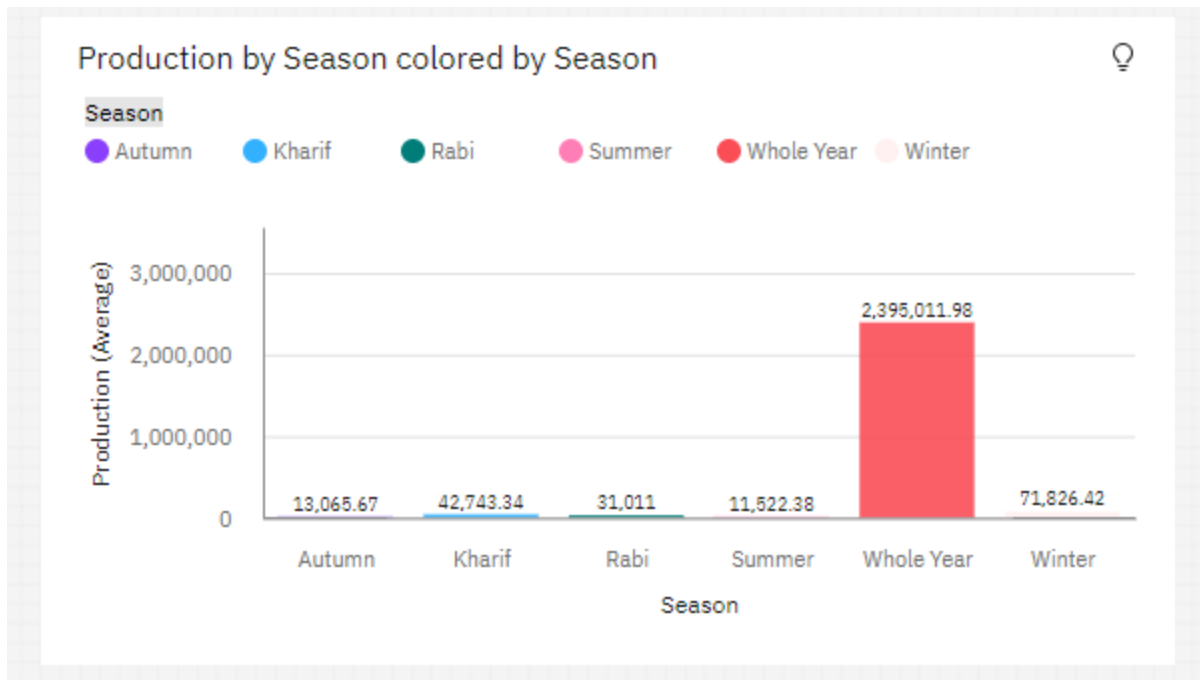
Figure 3. Seasons with Average Productions

b. Years usage of Area and Production
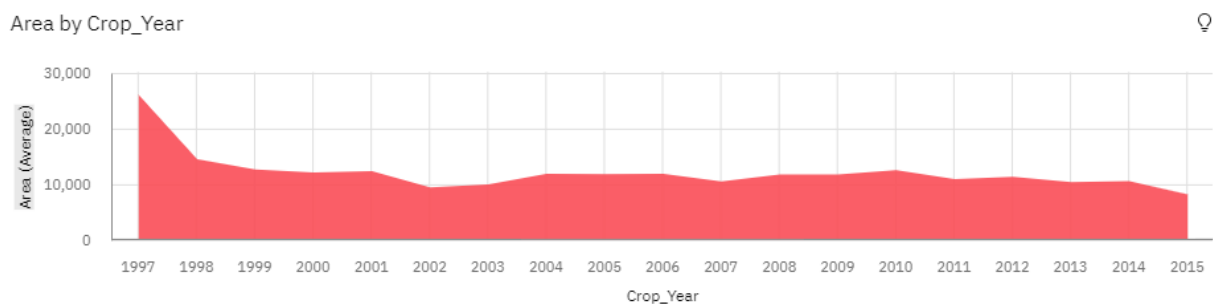
Figure 4 . Years usage of Area

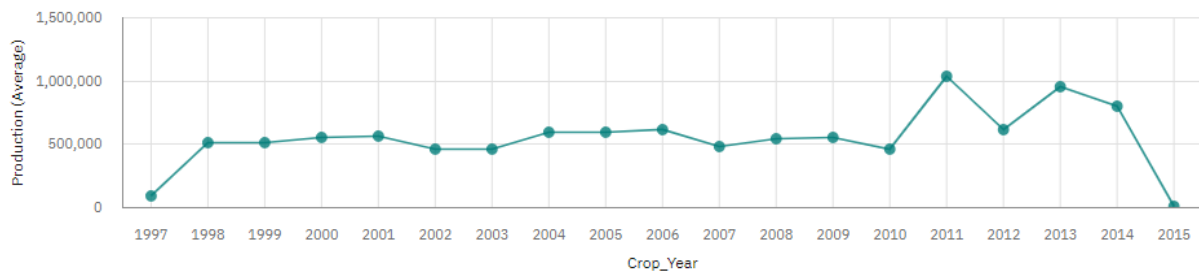Figure 5 . Years usage of Production

c. Top 10 states with Most area
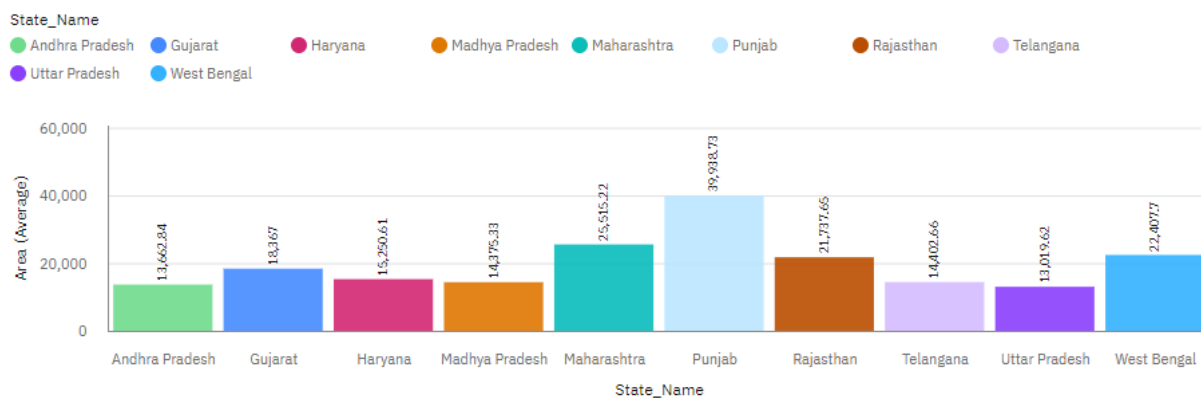
Figure 6 . Top 10 states with most Area

d. State with Crop Production
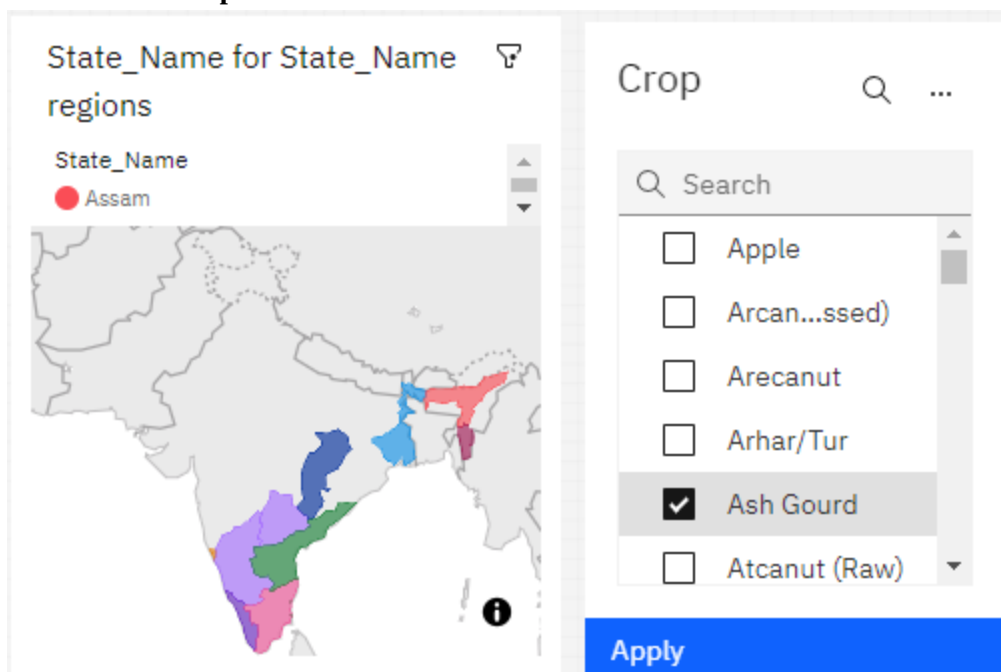
Figure 7 . States with Crop Production

e. States with the Crop Production along with season(Text table)

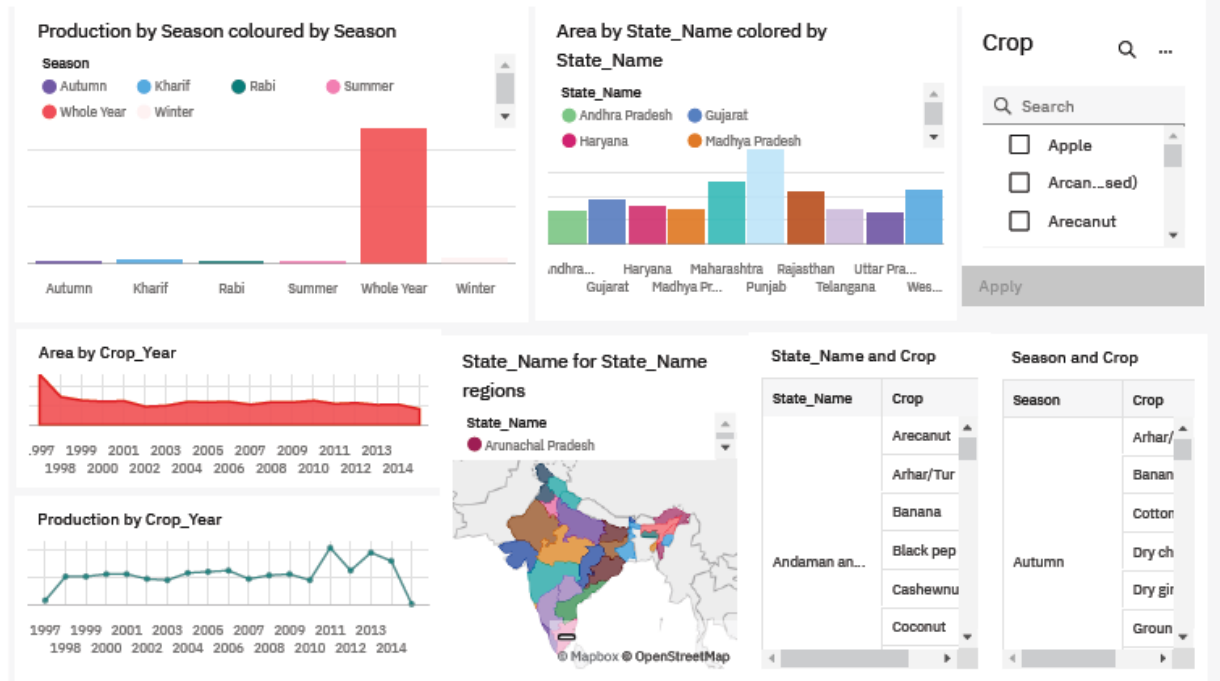Figure 8 . States with Crop Production along with season(Text Table)

f. Final Dashboard

Figure 9 . Final CPA Dashboard

7. ADVANTAGES & DISADVANTAGES
7.1 Advantages:

- Offers multiple BI capabilities like reporting and query, analysis, and score carding with a single architecture.

- Automatic real-time notifications and workflows for better productivity.

- Support for popular RDBMS tools for content management and reporting.

- Share insights with colleagues for collaborative decision-making and collective intelligence.

- Cognos Active Reports are transportable. They can be e-mailed

- Cognos Active Reports are great for offline consumption. They can be executed while away from the office.

- Cognos Active Reports can be flashy.

7.2 Disadvantages:

- Features like departmental deployments and multi-dimensional analysis are unavailable.

- Comes with a steep learning curve and is not ideal for non-technical users.

- Does not support MacOS, and no free trials are available.

8. APPLICATION

This can be used by farmers and business decision makers to analyze and investigate data sets and summarize their main characteristics. Also can be used to see what data can reveal beyond the formal modelling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate.

9. CONCLUSION

The proposed work deals with data visualization using IBM Cognos Analytics platform. Crop yield estimation dataset was used to perform the Data Visualization task. The visualization charts for the different categories were generated and a meaningful dashboard was generated for the most insights of Crop production in India. The dashboard created using this platform acts as a smart and self-service analytical solution to the users that can query, track, and process analytical data. Due to the attractive dashboards, visualizations, ease of use, and forecasting capabilities, IBM Cognos Business Intelligence will be highly popular among users.

10. FUTURE SCOPE

The future work may involve in data visualization using real time data rather than the static dataset. The various Machine Learning models may be applied to produce the analytical report to predict the crop yield based on the various environmental conditions.

11. BIBILOGRAPHY

1. https://www.ibm.com/academic/badges/improving-healthcare-the-role-of-cloud-computing

2. https://www.ibm.com/academic/badges/improving-healthcare-the-role-of-data-analytics-and-data-science

3. https://cognitiveclass.ai/courses/python-for-data-science

4. https://cognitiveclass.ai/courses/data-analysis-python

5. https://cognitiveclass.ai/courses/data-visualization-python

6. https://smartinternz.com/Student/guided_project_workspace/101562

7. https://smartbridge.teachable.com/courses/data-analytics-ibm-project-build-a-thon/lectures/41626857

8. https://smartbridge.teachable.com/courses/data-analytics-ibm-project-build-a-thon/lectures/41683271

9. https://smartbridge.teachable.com/courses/data-analytics-ibm-project-build-a-thon/lectures/41683270

10. https://smartbridge.teachable.com/courses/data-analytics-ibm-project-build-a-thon/lectures/41685557

11. https://smartbridge.teachable.com/courses/data-analytics-ibm-project-build-a-thon/lectures/41710977

12. https://smartbridge.teachable.com/courses/data-analytics-ibm-project-build-a-thon/lectures/41736401

APPENDIX

Source Code

Only Data Visualizations are generated using IBM Cognos, hence source code is not applicable for this proposed work.