Project Report on "Agriculture Data Analytics In Crop Yield Estimation Using IBM Cognos"

By,
Namratha M
Assistant Professor,
Department of CSE,
B.M.S. College of Engineering,
Bangalore-560019

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

Visualization of data in the form of grpahs is very useful for analyzing large amount of data. The crop production data is downloaded form KAggle which as the following parameters: State Name, District name, Crop Year, Season, Crop Area, crop Production.

Plotting of 6 different grpahs was done as part of the project.

The objectives of the project are:

- Know fundamental concepts and can work on IBM Cognos Analytics.
- Gain a broad understanding of plotting different graphs.
- Able to create meaningful dashboards

2 Literature Survey

2.1 Existing Problem

Excel charts allow spreadsheet administrators to create visualizations of data sets. By highlighting a set of data within an Excel spreadsheet and feeding it into the charting tool, users can create various types of charts in which the data is displayed in a graphical way. This can aid understanding of a data set, as well as communication of it, with Excel charts suitable for use within management or corporate presentations.

A chart can create a clearer picture of a set of data values than a table with rows of numbers in it, allowing managers to incorporate this understanding into analysis and future planning. Benefits of charts and graphs also come into play in presentations, where they can be used to quickly illustrate trends in data for others to see.

It's often easier to spot trends and patterns in graphs and charts than by simply looking through a long table of data, especially for big datasets, so one of the common uses of chart tools is in

simply coming to grips with what a new dataset actually means.

Automation of Chart Creation:

The Excel application automates the process of generating charts from existing data sets. If a spreadsheet already contains updated data, the chart function can transform this data into a chart with a minimal amount of user input. The Recommended Charts tool in Excel is a key part of this process. Using the chart function, spreadsheet administrators can generate a chart in a few clicks, choosing a chart type, as well as options such as labels, axes and titles.

This is often a big advantage over drawing a chart by hand with an illustration tool or even using a more complex visualization tool, many of which require programming or configuration.

Customization of Charts in Excel:

The chart function in Excel enables users to strike a balance between automation and customization. Although the program can generate a chart quickly and easily, the user is also able to assert control over many chart details if necessary. As well as customizing the data presented within a chart, users can finely tweak the chart options to any detail. Charts in Excel can also be altered after they have been initially generated, by selecting a chart and choosing from the available options to amend the chart type, data or other details.

Integration into Existing Sheets:

If a business or other organization is using spreadsheet data managed within Excel, using the chart function within Excel aids integration of the data. For example, when an Excel spreadsheet generates a chart from the data within a worksheet, the chart will automatically update whenever the data itself is edited. This allows business managers and administrators to manage their data and visualizations within a single application, with the results updating instantly. Excel spreadsheets can also be easily shared between users and computers.

This can be one of the advantages of Excel tables and charts over other data visualization systems where charts and datasets are effectively stored and managed separately.

Microsoft Excel is widely used spreadsheet software. Many organizations and colleges have adopted it due to its good reputation and useful features like time-saving formulas and the ability to produce professional charts and graphs. While Excel has many benefits for users, it also has a few drawbacks that we should be aware of.

The following are the 10 disadvantages of Microsoft Excel:
1. Lack of control and security
2. Excel is prone to human error
3. Excel is hard to consolidate
4. Excel is unsuitable for agile business practices
5. Excel can't help us make quick decisions
6. Excel is not designed for collaborative work
7. Excel is difficult to troubleshoot or test
8. Excel is vulnerable to fraud/corruption

9. Difficult to manage advanced pricing rules
10. Excel is Unfit for Agile business Practices

1. Lack of Control and Security

In Excel, if the data is too huge, it might slow down the Excel program, especially if all the data is included in one file. Attempting to divide the data down into smaller files may result in some of it being lost or missing. Excel is not user-friendly, and the application rounds off very large numbers using precise calculations, which compromises accuracy.

Excel is also a stand-alone tool that is not completely integrated with other business systems; it does not offer enough control because sales managers do not have clear and consistent visibility of the quotes that are sent by their representatives or the history of those quotes.

Every time employees make changes to Excel manually; there is a risk of introducing errors and compromising accuracy. These errors are not only difficult to spot later in the process, but they can also have a detrimental impact on the bottom line.

2. Excel is Prone to Human Error

Excel's great sensitivity to minor human errors is more ordinary but equally destructive. Negative indications that are missed rows that are misaligned may appear innocuous, but they can impair investor trust or result in a significant loss of opportunity worth millions of pounds. On average, a spreadsheet will contain 1 error for every cell for every 20 cells that have data.

3. Excel is Hard to Consolidate

The usage of an Excel spreadsheet to generate reports makes the consolidation process lengthy. End-users must often collect data from various files, summarize it, then send their report to their department heads by email, portable storage media, or copying to a generally shared network folder. This procedure must be repeated until all of the data has reached the organization's senior decision- makers.

4. Excel is Unsuitable for Business Continuity

Spreadsheet data is hardly kept in a single location; instead, it is frequently in the hands of non-IT people who are unaware of data storage and backup best practices. Complete data recovery can be so tough, even impossible if a major disaster strike.

Even if a company has sufficient financial reserves, the lack of data to work with (e.g., accounts receivable records, customer records, stock and price lists) can stymie its ability to respond quickly.

5. Excel Can't Help Us Make Quick Decisions

In a spreadsheet-based environment, extracting data from many departments, aggregating that data, and summarizing the information in a way that supports the company's management in making the best decisions can be time-intensive.

In order to protect the integrity of the data, everyone participating in the information processing must be especially careful; double/triple checking will be required frequently. When the MD receives the final version, he or she may not have much time to deal with it.

6. Excel is Not Designed for Collaborative Work

Accurate pricing often necessitates information from a variety of people in many departments.

Multiple exchanges of data, ideas, and files will culminate in the final document.

When offices of the company are spread across the country, or specific team members are divided by significant distances, the only option to exchange spreadsheet data is via email. Duplication and even inaccurate data entry are a risk with this type of exchange. Different team members will struggle to keep track of comparable files sent back and forth and may submit an out-of-date version as a result.

7. Excel is Difficult to Troubleshoot or Test

Spreadsheets were never intended to be used for testing. It is not rare to have interconnected spreadsheet data scattered across multiple folders, offices, workstations, or even geographic locations. Even though the locations of all connected files can be pinpointed, it can take a long time to track the logic of formulas from one related cell to another. When troubleshooting any suspicious data, similar issues will arise.

8. Excel is Vulnerable to Fraud/Corruption

Several million pounds have already been lost as a result of fraudulent changes in firm Excel files. The main cause of this spreadsheet vulnerability is the lack of controls (as mention above) that makes it very easy to change formulas, values, or dependencies without being discovered.

9. Difficult to Manage Advanced Pricing Rules

Excel is commonly used in enterprises to manage a pricing plan based on cost-plus pricing, discount from list pricing approaches, or other straightforward pricing tactics.

10. Excel is Unfit for Agile Business Practices

Spreadsheets are frequently generated by people with little or no IT experience. Spreadsheets files eventually evolve into highly customized user-developed applications. Therefore, when it is time for a new person to take over as part of a business or personnel change, the newcomer may have to start from the beginning.

Below are some other disadvantages of Microsoft Excel:
1. Keeping up with the changing business world
2. Regulatory compliance challenges
3. Not prepared for disaster

1. Keeping Up with the Changing Business World

Major changes are shaping and redefining the corporate landscape in today's world. We have firsthand knowledge of the insurance industry's role in this transformation. This can be seen in large-scale business transformation programmes; M&A activity, and management buyouts are examples of this. Spreadsheets in reinsurance programmes are frequently extremely personalized for each user.

When a new person takes over as part of a large-scale corporate transition, spreadsheets may be

so individualized that the new person will have to start from scratch. Unlike a system, the spreadsheet user does not have a manual on how to utilize the function of the spreadsheet. This results in substantial productivity inefficiencies and exposes the data to errors once again if a new worker is left guessing what to do.

2. Regulatory Compliance Challenges

Using spreadsheets makes it difficult to ensure regulatory compliance for our reinsurance programs, as the data can be susceptible to fraud and errors.

Which regulations apply to reinsurance based on spreadsheets? We have seen a spike in worldwide regulation during the previous two decades, including:

- GDPR
- Solvency III
- Dodd-Frank, Basel III
- Serbanes-Oxley (SOX)
- FAS 157

3. Not Prepared for Disaster

When tragedy strikes, our reinsurance programme will be jeopardized if best practices for spreadsheet storage and backup are not in place. If something goes wrong, recovering all of our data will be difficult, if not impossible, and will significantly impact on our organization.

2.2 Proposed Solution

The proejct aims the usage of IBM Cognos Analytics for creating visualization charts.

There are various advantages to integrating with Cognos 10.2.2 for generating reports. Some of these are:

- Lower costs—reduces maintenance due to complete report coverage and a zero-footprint environment.
- Faster results—shortens reporting time due to seamless integration and adaptive authoring.
- Improved decision making—reports and dashboards present data in easily-understood formats.
- Adaptive authoring automatically adjusts report layout when objects are added, moved, or removed.
- Ability to work with data using familiar business terms.
- Ability to use a variety of charts—crosstabs, bar or 3D bar, pie or doughnut, line, gauge, funnel, scatter, dot density, waterfall, and so forth.
- Ability to create complex, multi-page layouts using different data sources.
- High performance data access across all sources.
- Complete connectivity regardless of environment.
- Open architecture that leverages XML, SOAP, and WSDL.
- Multiple export formats—Excel, Portable Document Format (PDF), Extensible Markup Language (XML), Hypertext Markup Language (HTML), and Comma Separated Value (CSV).

- Multilingual capabilities automatically deliver reports in the users' working language.
- Ability to integrate seamlessly with the Selling and Fulfillment Foundation, without the user having to log in to the application again.

3 Theoretical Analysis

3.1 Block Diagram
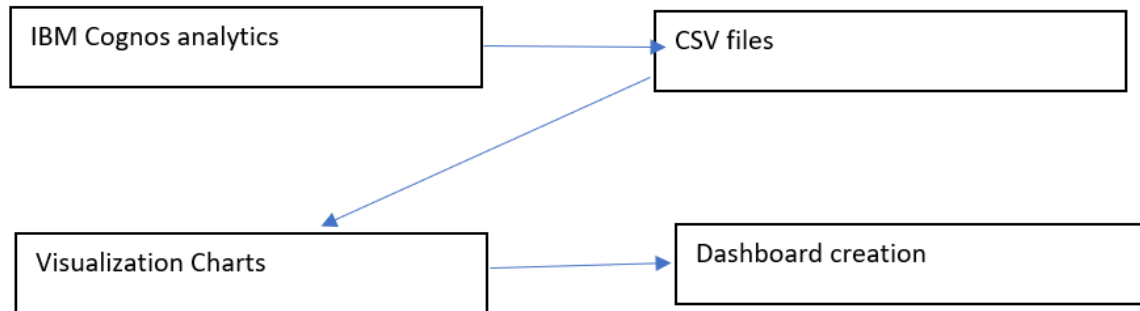
3.2 Hardware/Software Designing

Hardware Requirements-NIL

Software Requirements-IBM Cognos Analytics

4 Experimental Investigations

Six different visualization charts are plotted for the crop production data.

As production of crops depends on different seasons, we plot the graph to visualize the average production based on different seasons.

n our dataset we also have a year's columns by which we will plot a line and area graphs to see the change in these both data with respect to increase in years.

As we have an area data in our dataset, we will be plotting some graphs to visualize the top 10 Indian states with the most area.

There are so many different crops produced in Indian and most of us don't know which crop is belongs to which state so we will be plotting and highlight the states in map according to different crops.

Taking forward the previous plot we will be fetching the state name and showing it in a text table whenever different crops are chosen.

5 Flowchart
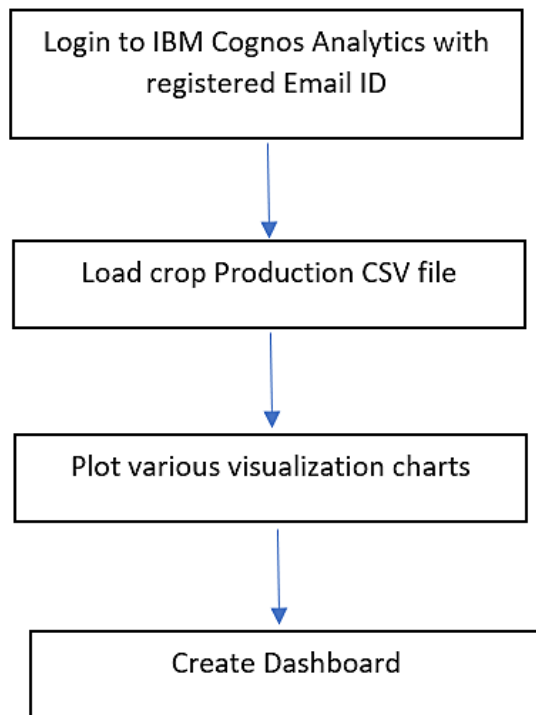
6 Results

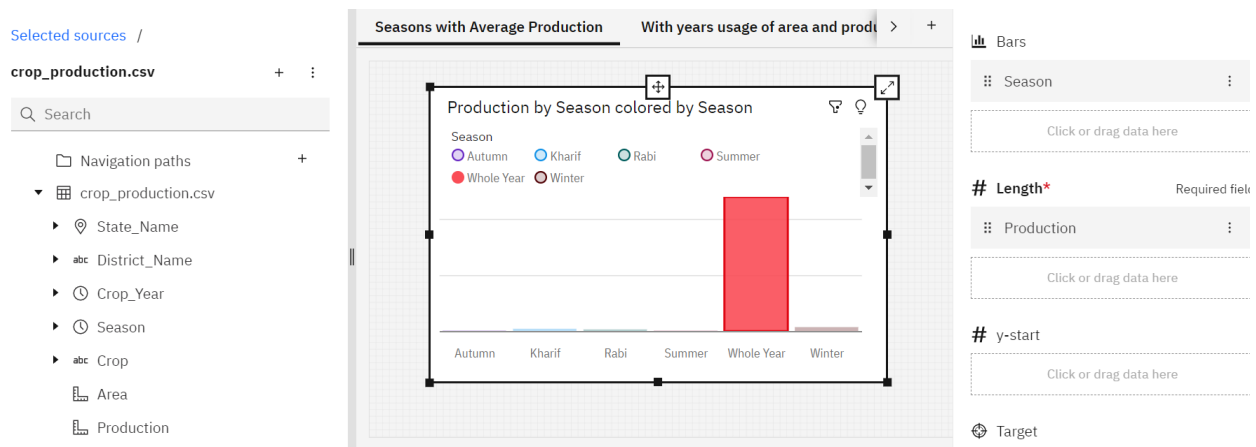Figure 1 shows the average production of various crops according to the season.

Figure 1: Average production of crops by season

Figure 2 shows two plots one for the average area accding to the crop and the aera grpah type is plotted. another line grpah is also plotted for the production by year and data labels are added.
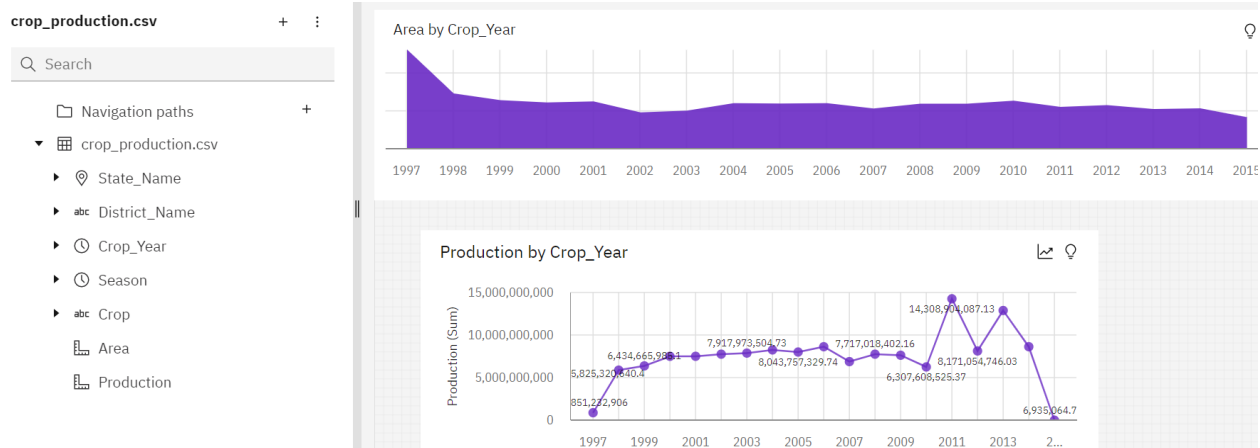
Figure 2: Average area and production by year

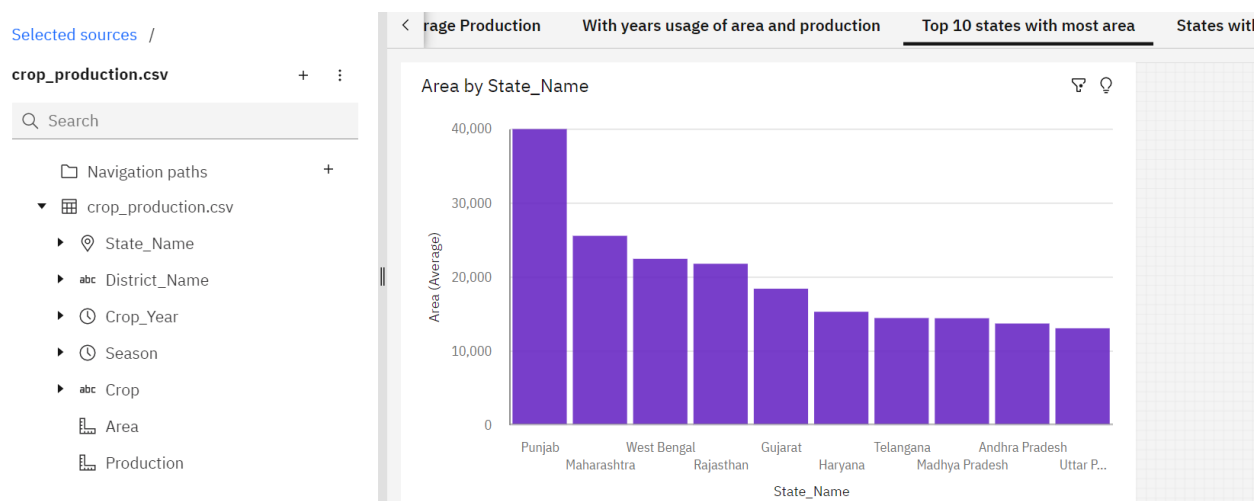Figure 3 shows the top 10 states producing the crops.

Figure 3: Top 10 states for crop

Figure 4 uses the map plot to show and apply filters and see which crops are being produced in which states of India.

Filter can be applied and a specific crop can be specified and the various states that produce the crops are highlighted.
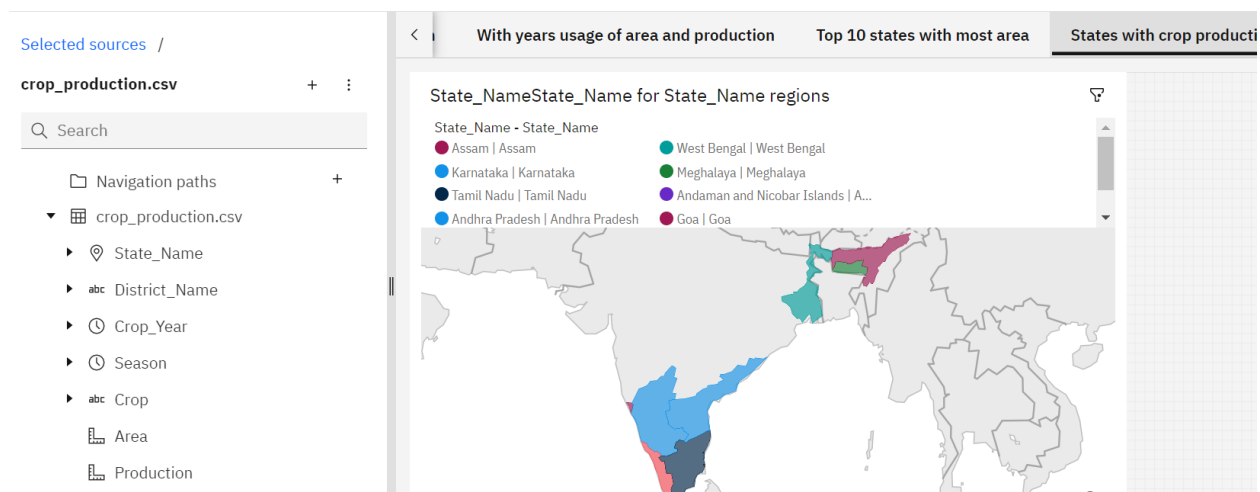
Figure 4: States with crop production

Figure 5 shows how to plot the etxt tbale visualization tool and shows the crop being grown in various states. again filter is applied here for a specific crop.

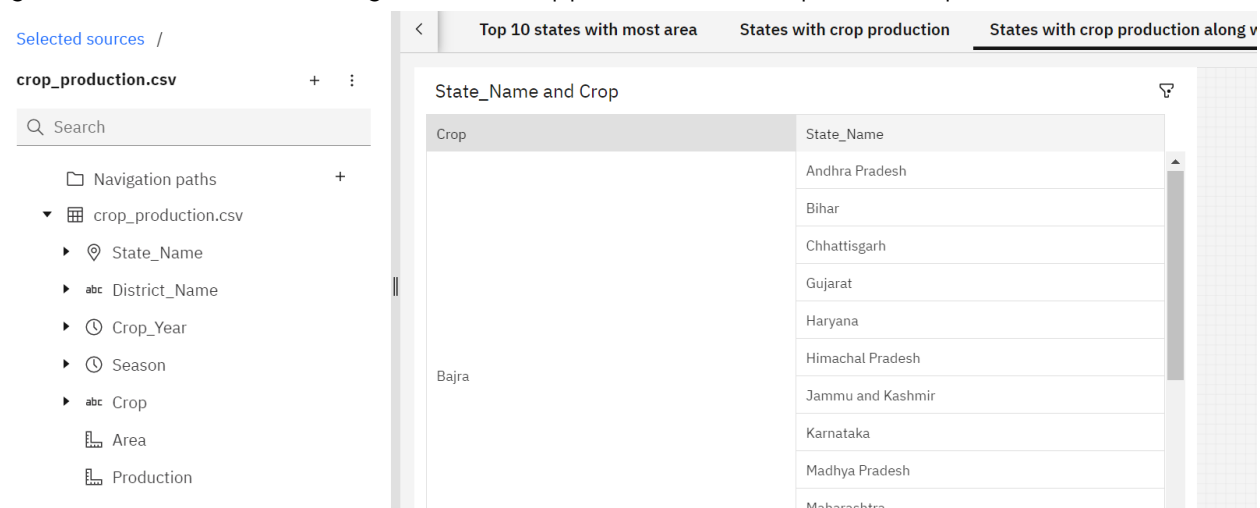
Figure 5: Text table for crop by states

Finally a dashboard is created and all the graphs plotted are copied there and pasted.

7 Advantages and Disadvantages
Advantages:
Able to produce grpahs quickly
The graphs can be generated in an interactive manner

Disadvantages

It is difficult to get access to the coount free whole time

8 Applications
Data Visulaization charts help us to depict data in a clear and precise fashion and easy for the user to understand.

9 Conclusion
As part of this Data Anlytics project, IBM Cognos Analytics Login was cretaed and the crop production dataset was loaded. Then the various visualization charts were plotted for the various parameters in the dataset. Different types of graphs and different options for each graph was explored.