

WORKING WITH DATASET

Date	29 October 2022
Team ID	PNT2022TMID23966
Project Name	Estimate The Crop Yield Using Data Analytics

To work on the given dataset, you need to first Understand the Dataset and the Load it to Cloud platform then Build the required Visualizations to provide various visual analytical solutions.

Understanding the Dataset:

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

State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
Andaman and Nicobar Islands	NIOBARIS	2000	Kharif	Areacnut	1254	2000
Andaman and Nicobar Islands	NIOBARIS	2000	Kharif	Other kharif pulses	2	1
Andaman and Nicobar Islands	NIOBARIS	2000	Kharif	Rice	102	321
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Banana	176	641
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Cashewnut	720	185
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Coconut	18168	65100000
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Dry ginger	36	100
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Sugarcane	1	2
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Sweet potato	5	15
Andaman and Nicobar Islands	NIOBARIS	2000	Whole Year	Tapioca	40	189
Andaman and Nicobar Islands	NIOBARIS	2001	Kharif	Areacnut	1254	2061
Andaman and Nicobar Islands	NIOBARIS	2001	Kharif	Other kharif pulses	2	1
Andaman and Nicobar Islands	NIOBARIS	2001	Kharif	Rice	83	300
Andaman and Nicobar Islands	NIOBARIS	2001	Whole Year	Cashewnut	719	192
Andaman and Nicobar Islands	NIOBARIS	2001	Whole Year	Coconut	18190	64430000
Andaman and Nicobar Islands	NIOBARIS	2001	Whole Year	Dry ginger	46	100
Andaman and Nicobar Islands	NIOBARIS	2001	Whole Year	Sugarcane	1	1
Andaman and Nicobar Islands	NIOBARIS	2001	Whole Year	Sweet potato	11	33
Andaman and Nicobar Islands	NIOBARIS	2002	Kharif	Rice	189.2	510.84
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Areacnut	1258	2083
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Banana	213	1278
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Black pepper	63	13.5
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Cashewnut	719	206
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Coconut	18240	67490000
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Dry chillies	413	28.8
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Dry ginger	47.3	133
Andaman and Nicobar Islands	NIOBARIS	2002	Whole Year	Sugarcane	5	40
Andaman and Nicobar Islands	NIOBARIS	2003	Kharif	Rice	52	90.17
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Areacnut	1261	1525
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Banana	266	1763
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Black pepper	75.5	15.86
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Cashewnut	717	208.5
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Coconut	18284.74	68580000
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Dry chillies	60	102
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	Dry ginger	102	326.4
Andaman and Nicobar Islands	NIOBARIS	2003	Whole Year	other oilseeds	4.2	0.3
Andaman and Nicobar Islands	NIOBARIS	2004	Kharif	Rice	52.94	72.57
Andaman and Nicobar Islands	NIOBARIS	2004	Whole Year	Areacnut	1264.7	805.85

Let's understand the data we're working with and give a brief overview of what each feature represents or should represent

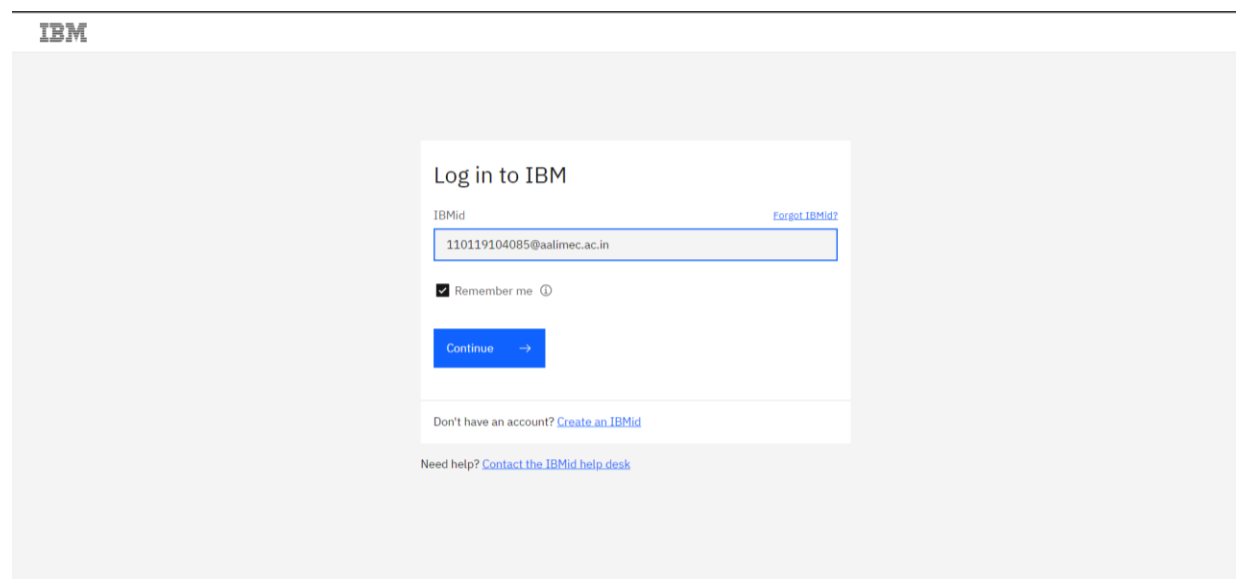
1. State Name - All the Indian State names.
2. District Name - Different District names.
3. Crop Year- contains the crop years.
4. Season – Different seasons for crop production.
5. Area- Total number of areas covered.
6. Production- production of crops.

Loading the Dataset:

Before you can build a view and analyze your data, you must first connect the data to IBM Cognos. Cognos supports connecting to a wide variety of data, stored in a variety of places.

The data might be stored on your computer in a spreadsheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in your enterprise.

Step 1: Login to the IBM Cognos Analytics with Watson.



The screenshot shows the IBM login interface. At the top left is the IBM logo. The main heading is "Log in to IBM". Below this, there is a label "IBMId" and a link "Forgot IBMId?". A text input field contains the email address "110119104085@aalimec.ac.in". Below the input field is a checkbox labeled "Remember me" with an information icon. A blue "Continue" button with a right-pointing arrow is positioned below the checkbox. At the bottom of the login box, there is a link "Don't have an account? Create an IBMId". Below the login box, there is a link "Need help? Contact the IBMId help desk".

Step 2: Launch the IBM Cognos Analytics.

The screenshot shows the IBM My IBM dashboard. At the top, there's a navigation bar with the IBM logo, a user profile icon, and links for My IBM, Profile, and Billing. The IBM ID is 110119104085@aalimec.ac.in. The main section is titled 'Products' and features a 'Trials' section with 2 offerings. The first offering is 'Cognos Analytics on Cloud Trial for Students', which is active and expires on Sep 5, 2023. It has 'Launch' and 'Manage' buttons. The second offering is 'IBM Cloud', also active and expires on Jan 2, 2023, with 'Launch' and 'Manage' buttons. To the right, there's a promotional banner for IBM solutions with a 'View catalog' button. A link for 'Need help? Get product support' is also present.

Step 3: Upload the Dataset

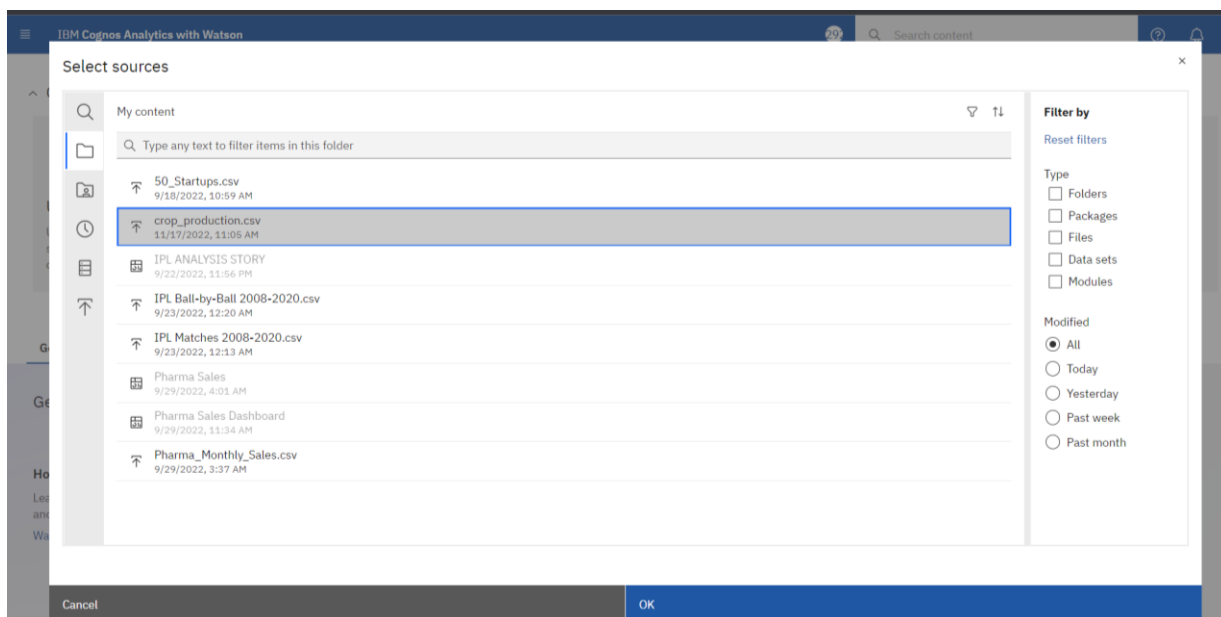
The screenshot shows the IBM Cognos Analytics interface. The top navigation bar includes the IBM Cognos Analytics with Watson logo, a search bar, and user profile icons. The main content area is divided into several sections. On the left, there's a 'Quick launch' section with an 'Upload data' button and a description: 'Upload or drag and drop spreadsheets, csv files, and other data sources.' Below this is a 'Get started' section with a 'Recent' tab. In the center, a Windows file explorer dialog is open, showing the 'This PC > Documents > IBM > Dataset' path. The file list contains 'crop_production' (Microsoft Excel file, 14,958 KB, modified 11/12/2019 7:20 AM). The 'File name' field is set to 'crop_production' and the 'Files of type' is set to 'Custom Files'. The 'Open' button is highlighted. On the right, there's a 'Present data' section with a description: 'Create sophisticated, multi-page, multi-query dashboards, reports, or stories.' At the bottom, there are three sections: 'How-to videos' (Learn how to use IBM Cognos Analytics and start gaining insights into your data. Watch videos), 'Accelerator Catalog' (Browse a growing catalog of industry use, best practice guides, custom visualizations, and more. Browse catalog), and 'Learning resources' (Expand your knowledge and get assistance by connecting with the community, reading documentation, and watching videos. Launch the learn pane).

Preparing the Data:

The uploaded data can be prepared using the option called “prepare data”.

Using this option we can create data modules and with the data module we can clean the data.

Step 1: Click the prepare data and choose our dataset.



After choosing our dataset click “OK” option in the bottom right corner.

Step 2: Clean the data by right clicking the specific category.

The screenshot displays the IBM Cognos Analytics interface. On the left, the 'Data module' pane shows a tree view with 'Crop_production Module' expanded, revealing 'Navigation paths' and 'crop_production.csv'. The 'Crop' folder is selected. The main area shows a data table with columns: Row Id, State_Name, District_Name, Crop_Year, Season, Crop, and Area. The table contains 10 rows of data. A context menu is open over the first row, with 'Filter...' selected. The menu options include: Filter..., Create data group..., Create navigation path..., Search for members..., Refresh members, Split..., Hide from users, Remove, Refresh properties..., Format data..., Clean..., Rename, Cut, Copy, and Properties.

Row Id	State_Name	District_Name	Crop_Year	Season	Crop	Area
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2
	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5
	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40
	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254
	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2
	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83

Step 3: We can do the required cleaning operations.

(eg: Changing State names to Uppercase)

Clean - State_Name

Whitespace

☐ Trim leading and trailing whitespace

Convert case to

☒ UPPERCASE ☐ lowercase ☐ Do not change

Return a substring of characters

Start Length

Preview

This is a preview

NULL values

☐ Replace this value with NULL

☐ Replace NULL values with

Properties

General **Navigation paths**

Crop

arecanut
other kh
rice
banana
cashew
coconut
dry gngy
sugarcar
sweet pc
tapioca
arecanut
other kh
rice

Label

Hide from users ☐

Expression [View or edit >](#)

Usage

Aggregate

Data type

Represents

Geographic location

State/Province

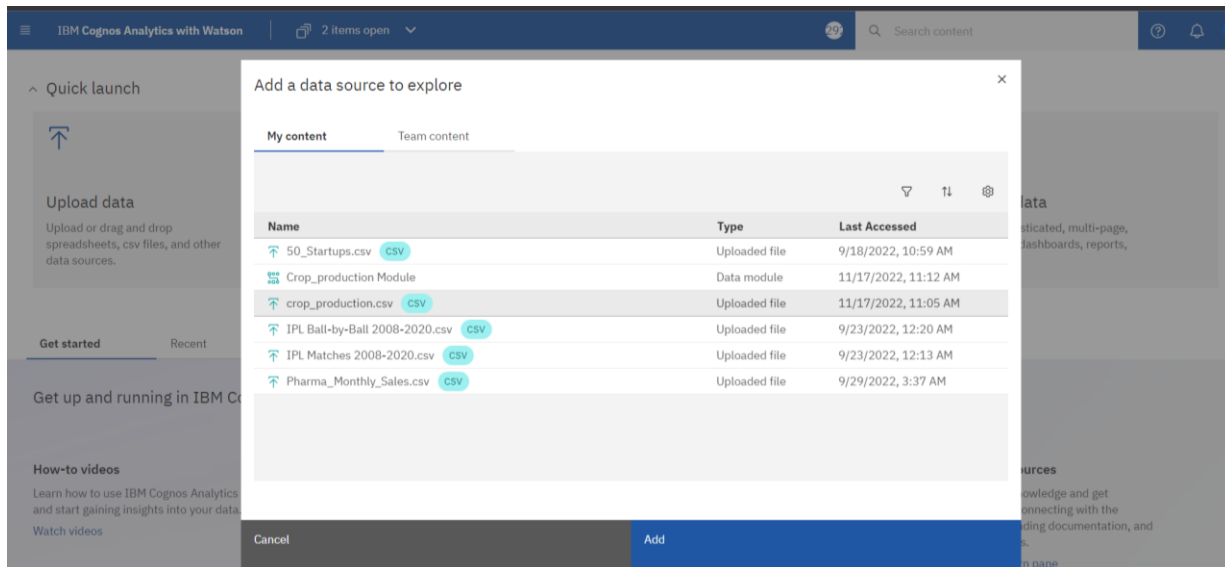
Description

Comments

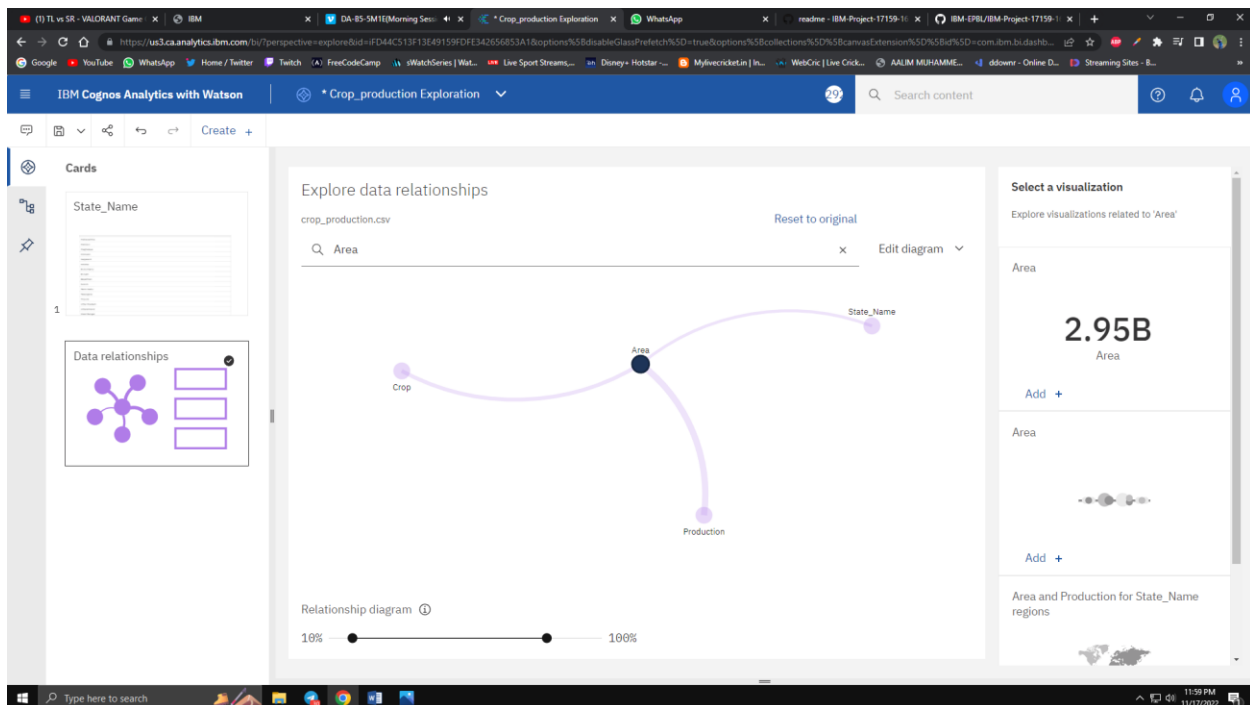
Data Exploration:

Quickly find unbiased answers by identifying trends in your data with data exploration.

Step 1: Upload the dataset using Exploration option in IBM cognos Analytics.



Step 2: Select the category you want to explore from the data you uploaded.



Step 3: You can select the required data subsets to explore more about them

