Project Development Phase – Sprint 1

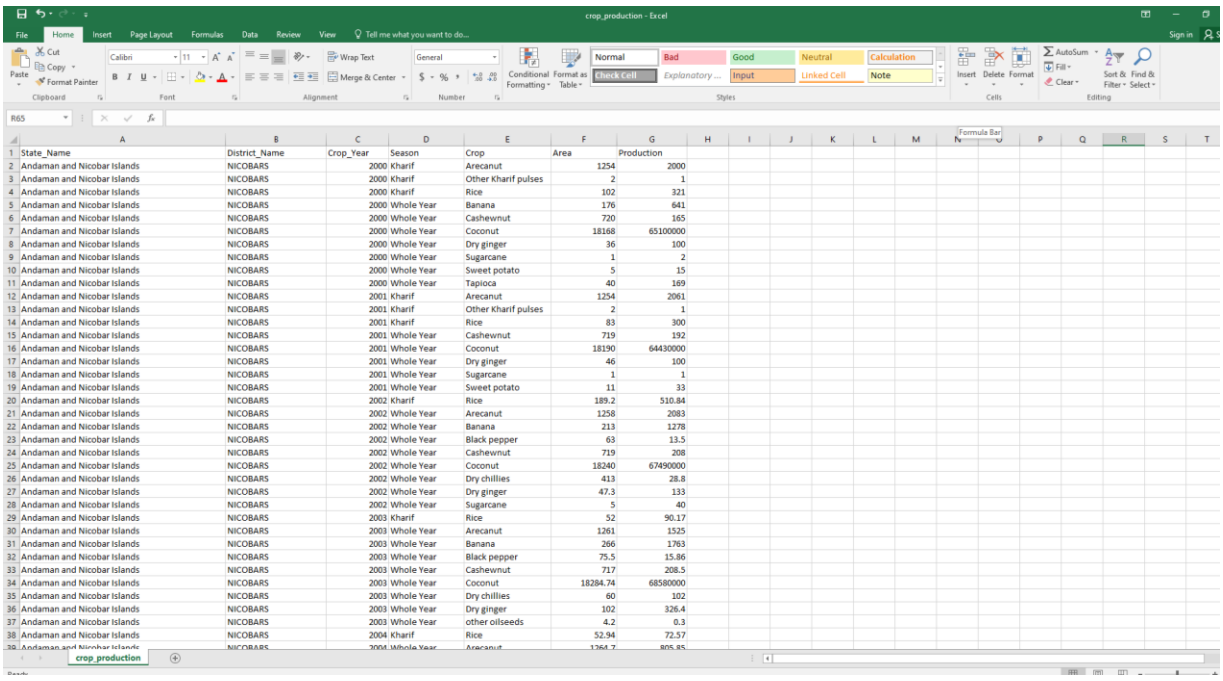
WORKING WITH DATASET

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 Visalizations 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	NICOBARS	2000	Kharif	Arecanut	1254	2000
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176	641
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720	165
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168	65100000
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36	100
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1	2
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5	15
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40	169
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254	2061
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83	300
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Cashewnut	719	192
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Coconut	18190	64430000
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Dry ginger	46	100
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sugarcane	1	1
Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sweet potato	11	33
Andaman and Nicobar Islands	NICOBARS	2002	Kharif	Rice	189.2	510.84
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Arecanut	1258	2083
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Banana	213	1278
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Black pepper	63	13.5
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Cashewnut	719	208
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Coconut	18240	67490000
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Dry chillies	413	28.8
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Dry ginger	47.3	133
Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Sugarcane	5	40
Andaman and Nicobar Islands	NICOBARS	2003	Kharif	Rice	52	90.17
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Arecanut	1261	1525
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Banana	266	1763
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Black pepper	75.5	15.86
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Cashewnut	717	208.5
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Coconut	18284.74	68580000
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Dry chillies	60	102
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Dry ginger	102	326.4
Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	other oilseeds	4.2	0.3
Andaman and Nicobar Islands	NICOBARS	2004	Kharif	Rice	52.94	72.57

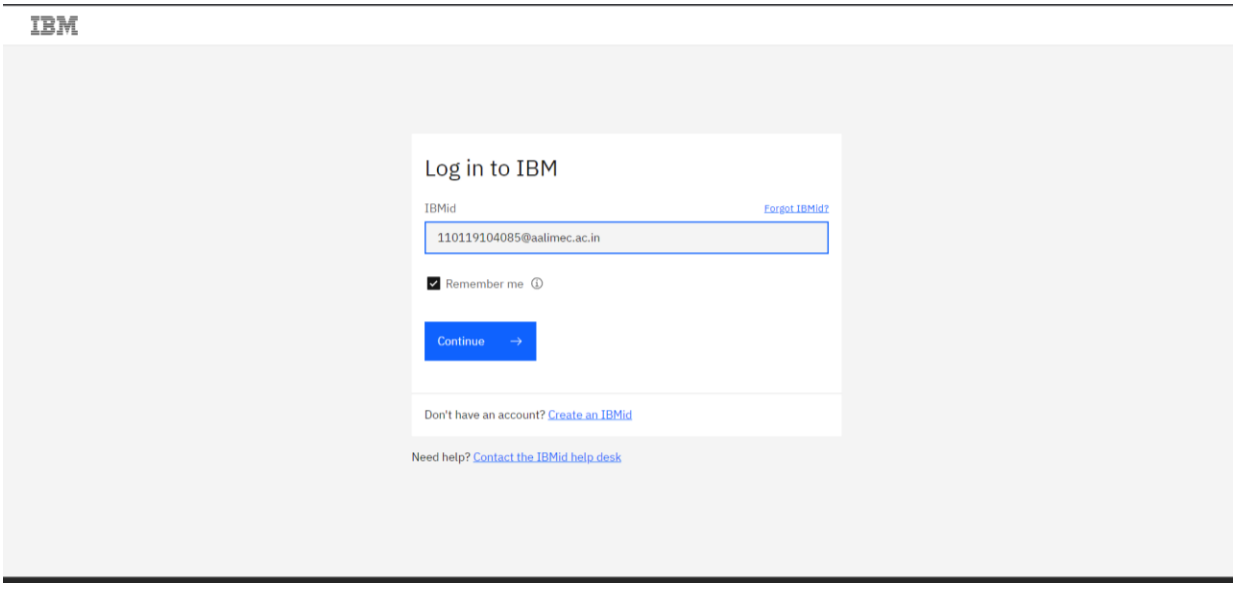
Let’s understand the data we’re working with and give a brief overview of what each feature represents or should represent

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.

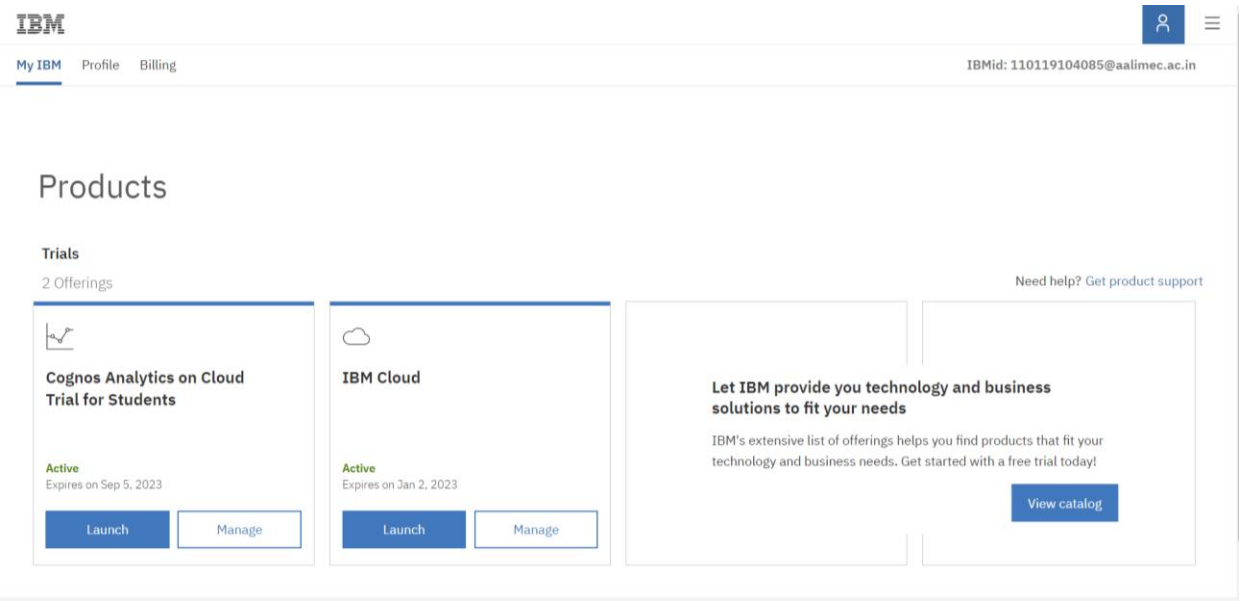
Loading the Dataset:

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.

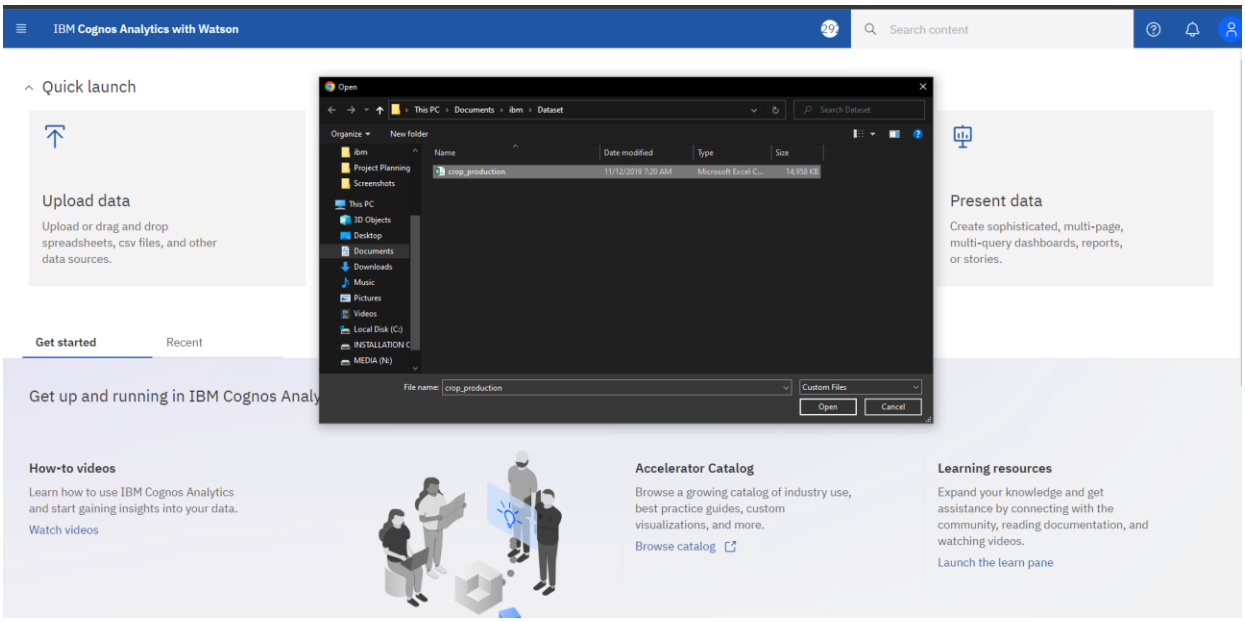
Step 1: Login to the IBM Cognos Analaytics with Watson.



Step 2: Launch the IBM Cognos Analaytics.



Step 3: Upload the Dataset

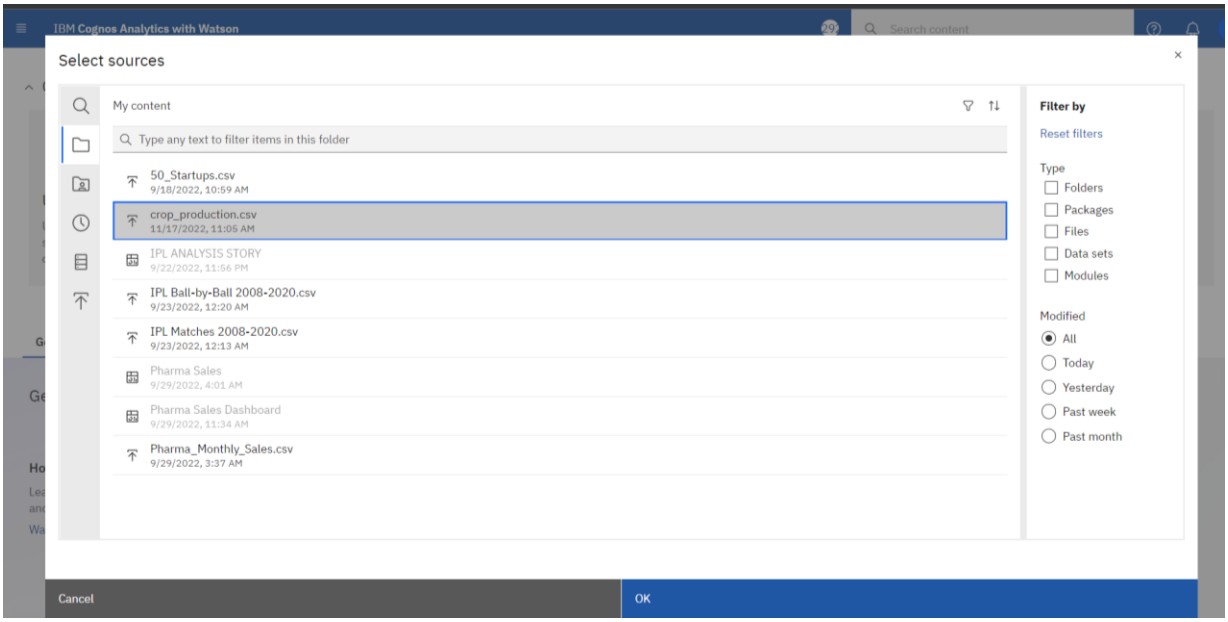


Preparing the Data:

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

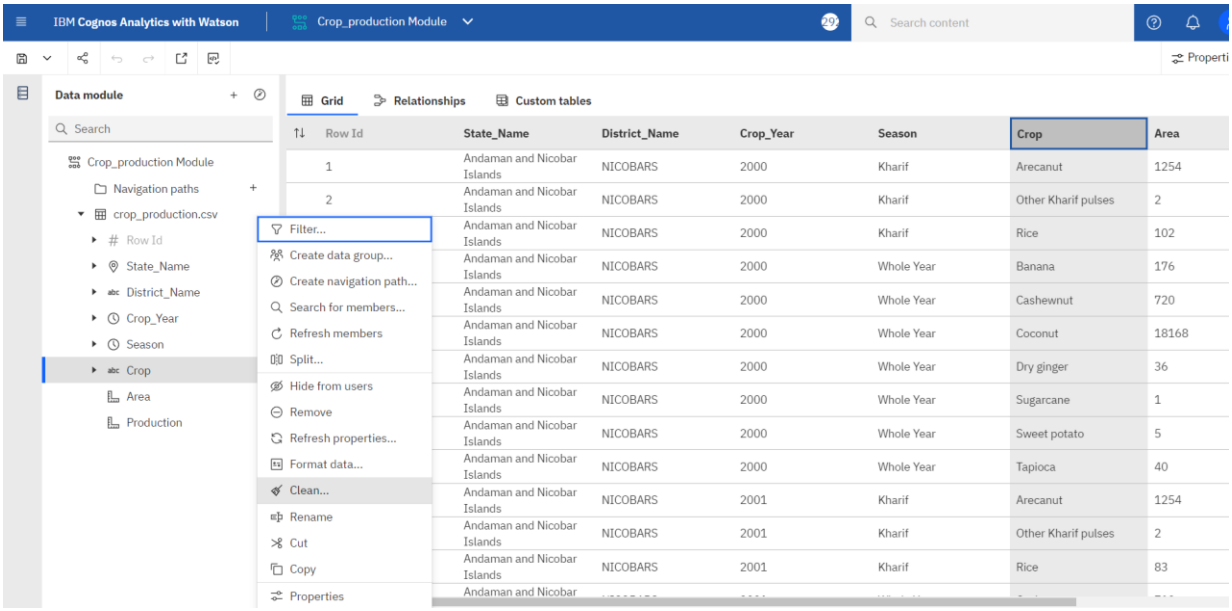
Using this option we can create data modules and with the data module we can clean the data.

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



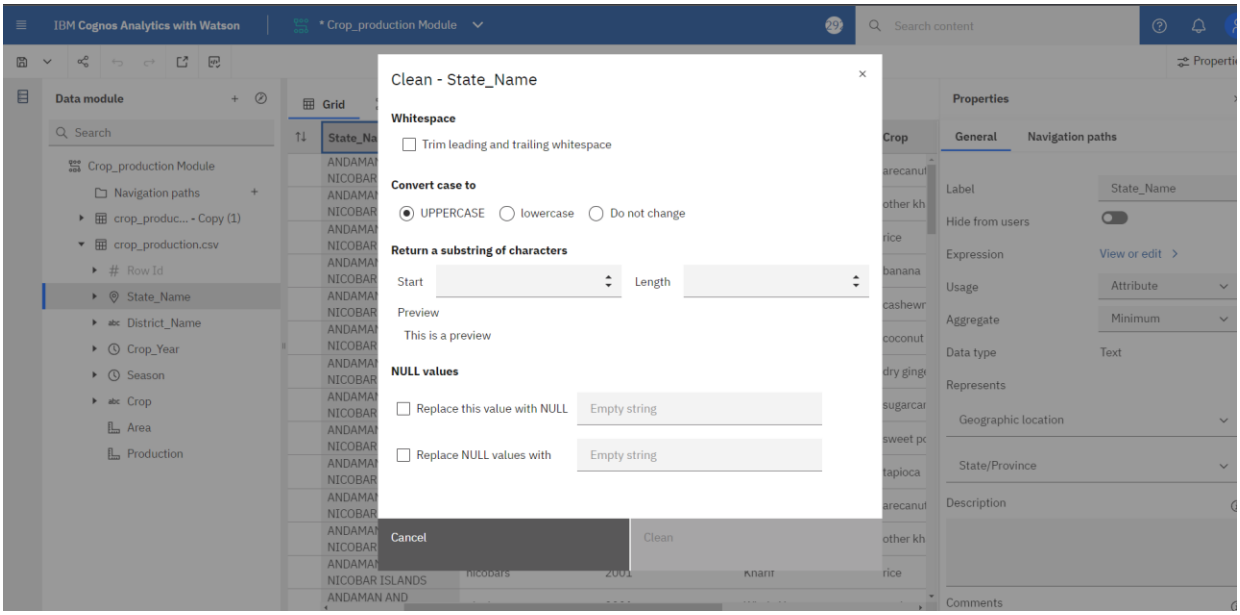
After choosing our dataset click “OK” option in the bottom right corner.

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



Step 3: We can do the required cleaning operations.

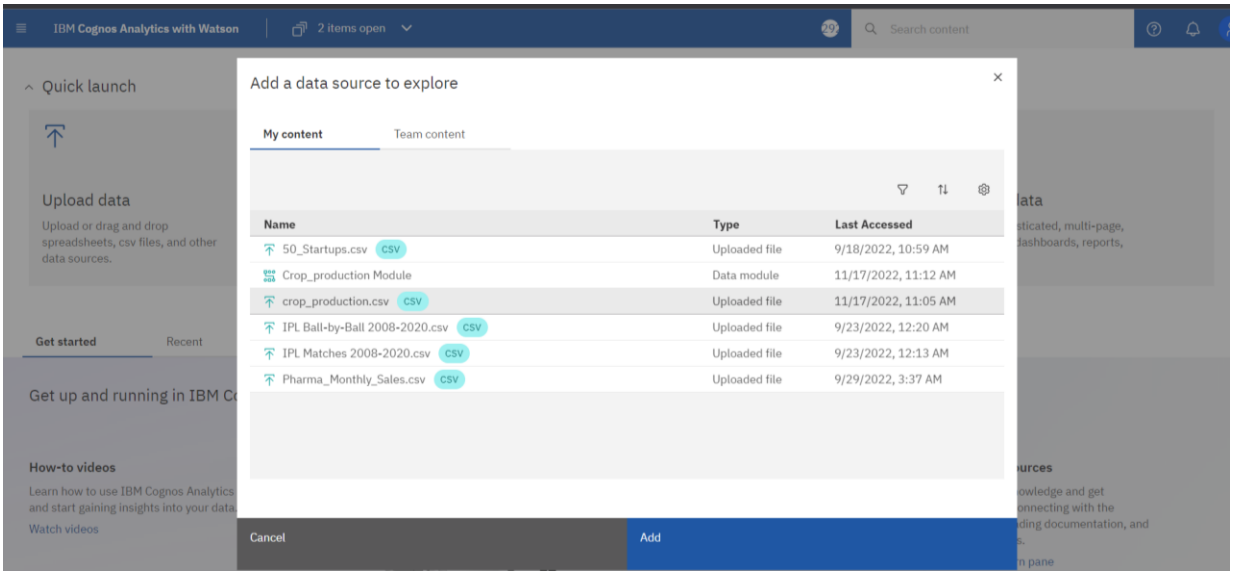
(eg: Changing State names to Uppercase)



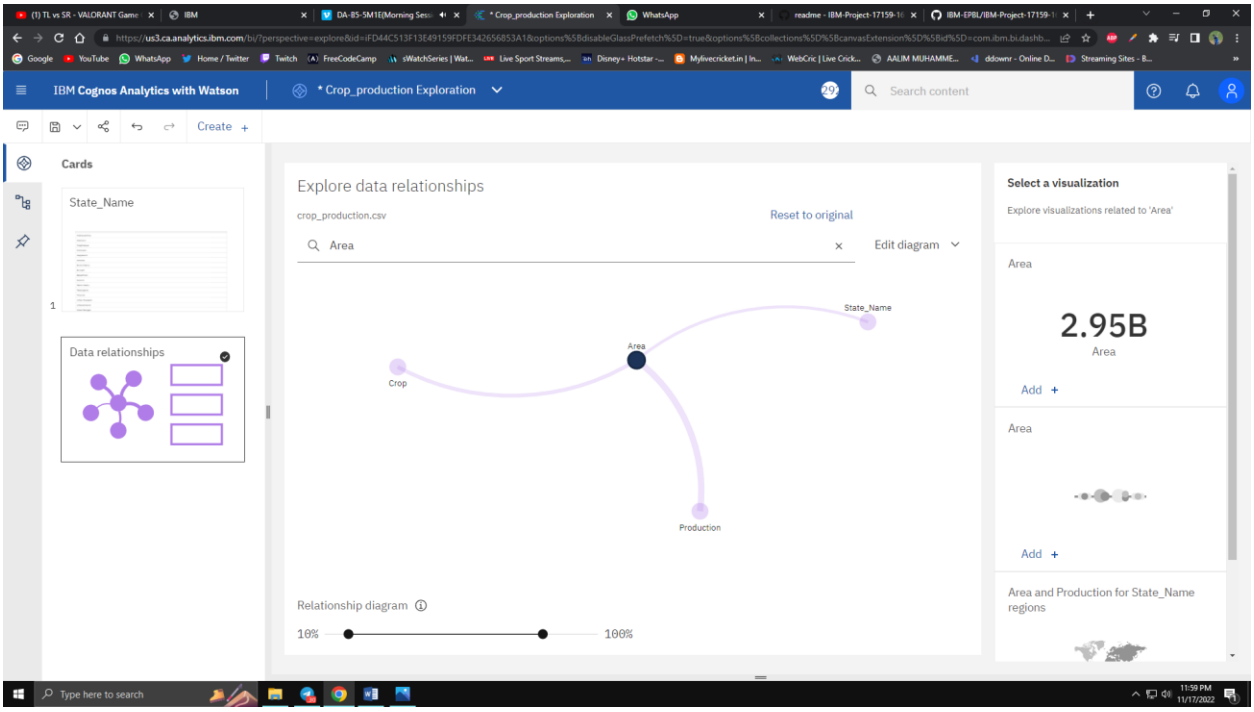
Data Exploration:

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