

## Sprint – 1

### Understanding The Dataset

Date	08 November 2022
Team ID	PNT2022TMID53863
Project Name	Estimate the crop yield using Data Analytics

#### Understanding The Dataset:

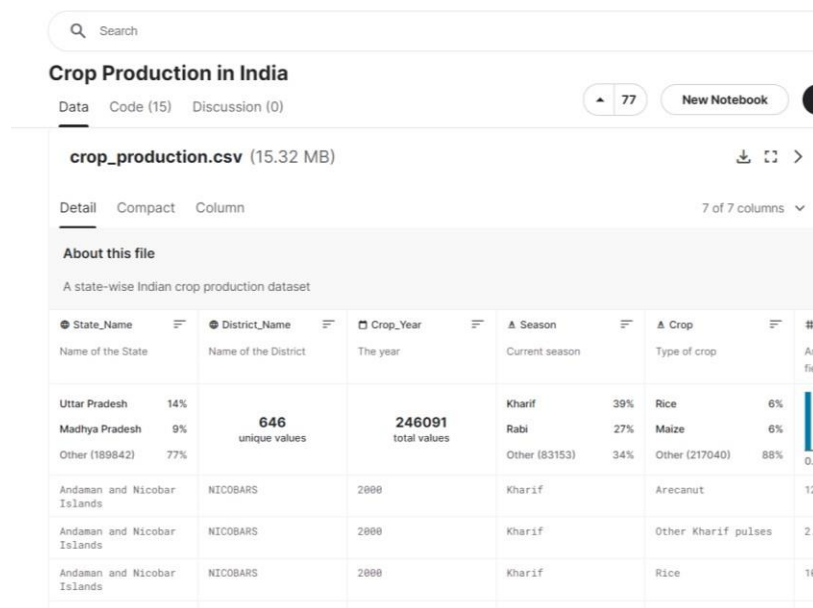
This project is based on a understanding the crop production of India. It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.

**Dataset Link:** [Dataset](#)

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.

The data format is as shown in the below image:



The screenshot shows a dataset viewer interface for 'Crop Production in India'. It includes a search bar, tabs for Data, Code (15), and Discussion (0), and a 'New Notebook' button. The dataset is named 'crop\_production.csv' (15.32 MB). Below the file name, there are tabs for Detail, Compact, and Column, and a dropdown for '7 of 7 columns'. The 'About this file' section describes it as a state-wise Indian crop production dataset. The main table has columns: State\_Name, District\_Name, Crop\_Year, Season, Crop, and Area. The table shows data for Uttar Pradesh, Madhya Pradesh, and Other (189842) for the year 2000, with a total of 246091 values. It also shows data for Andaman and Nicobar Islands for the year 2000, with a total of 162 values.

State_Name	District_Name	Crop_Year	Season	Crop	Area
Uttar Pradesh		2000	Kharif	Rice	14%
Madhya Pradesh		2000	Rabi	Maize	9%
Other (189842)		2000	Other (83153)	Other (217040)	77%
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	125
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.8
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	162