

# SPRINT – 1

## WORKING WITH THE DATASET(UNDERSTANDING THE DATASET)

Date	02 Nov 2022
Team ID	PNT2022TMID08893
Project Name	Estimate The Crop Yield Using Data Analytics

The “Estimate The Crop Yield Using Data Analytics” project is based on 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 detail.

**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.

### UNDERSTAND THE DATASET:

The screenshot shows the Kaggle dataset page for 'Crop Production in India'. The dataset is a CSV file named 'crop\_production.csv' (15.32 MB) with 76 versions. The 'Data Explorer' shows a preview of the data with 7 columns: State\_Name, District\_Name, Crop\_Year, Season, Crop, and Area. The preview shows data for Uttar Pradesh, Madhya Pradesh, and other states, with a total of 246,091 values for Crop\_Year. The 'About this file' section describes it as a state-wise Indian crop production dataset.

State_Name	District_Name	Crop_Year	Season	Crop	Area
Uttar Pradesh	14%	646 unique values	246091 total values	Kharif	39%
Madhya Pradesh	9%		Rabi	27%	6%
Other (189842)	77%		Other (183153)	34%	88%
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Areca nut	125
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.8
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102
Andaman and Nicobar	NICOBARS	2000	Whole Year	Banana	17%

- Download the dataset from the given dataset link.
- Data Given in the CSV file is understood.
- The data format is as shown in the below image:

The screenshot shows an Excel spreadsheet titled 'crop\_production.csv'. The data is organized in columns A through W. The first column (A) lists the state as 'Andaman : NICOBARS'. The second column (B) lists the year. The third column (C) lists the crop type. The fourth column (D) lists the crop name. The fifth column (E) lists the quantity. The sixth column (F) lists the value. The seventh column (G) lists the total value. The data is sorted by year and crop type.

State_Name	Year	Crop Type	Crop Name	Quantity	Value	Total Value
Andaman : NICOBARS	2000	Kharif	Areca nut	1254	2000	
Andaman : NICOBARS	2000	Kharif	Other Kharif	2	1	
Andaman : NICOBARS	2000	Kharif	Rice	102	321	
Andaman : NICOBARS	2000	Whole Year	Banana	176	641	
Andaman : NICOBARS	2000	Whole Year	Cashew nut	720	165	
Andaman : NICOBARS	2000	Whole Year	Coconut	18168	65100000	
Andaman : NICOBARS	2000	Whole Year	Dry ginger	36	100	
Andaman : NICOBARS	2000	Whole Year	Sugarcane	1	2	
Andaman : NICOBARS	2000	Whole Year	Sweet potato	5	15	
Andaman : NICOBARS	2000	Whole Year	Tapioca	40	169	
Andaman : NICOBARS	2001	Kharif	Areca nut	1254	2061	
Andaman : NICOBARS	2001	Kharif	Other Kharif	2	1	
Andaman : NICOBARS	2001	Kharif	Rice	83	300	
Andaman : NICOBARS	2001	Whole Year	Cashew nut	719	192	
Andaman : NICOBARS	2001	Whole Year	Coconut	18190	64430000	
Andaman : NICOBARS	2001	Whole Year	Dry ginger	46	100	
Andaman : NICOBARS	2001	Whole Year	Sugarcane	1	1	
Andaman : NICOBARS	2001	Whole Year	Sweet potato	11	33	
Andaman : NICOBARS	2002	Kharif	Rice	189.2	510.84	
Andaman : NICOBARS	2002	Whole Year	Areca nut	1258	2083	
Andaman : NICOBARS	2002	Whole Year	Banana	213	1278	
Andaman : NICOBARS	2002	Whole Year	Black pepper	63	13.5	
Andaman : NICOBARS	2002	Whole Year	Cashew nut	719	208	
Andaman : NICOBARS	2002	Whole Year	Coconut	18240	67490000	
Andaman : NICOBARS	2002	Whole Year	Dry chillies	413	28.8	
Andaman : NICOBARS	2002	Whole Year	Dry ginger	47.3	133	
Andaman : NICOBARS	2002	Whole Year	Sugarcane	5	40	
Andaman : NICOBARS	2003	Kharif	Rice	52	90.17	
Andaman : NICOBARS	2003	Whole Year	Areca nut	1261	1525	