

SPRINT – 1

WORKING WITH THE DATASET (UNDERSTANDING THE DATASET)

Date	25 October 2022
Team ID	PNT2022TMID35537
Project Name	Estimate The Crop Yield Using Data Analytics

“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 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.

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). The page includes a sidebar with navigation options like Home, Competitions, Datasets, Code, Discussions, Learn, and More. The main content area displays the dataset details, including a table with columns: State_Name, District_Name, Crop_Year, Season, Crop, and Area. The table shows data for various states and districts, including Uttar Pradesh, Madhya Pradesh, and Andaman and Nicobar Islands. A 'Data Explorer' panel on the right shows the dataset version (1) and the file name (crop_production.csv).

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

The screenshot shows the Excel spreadsheet for the 'crop_production' dataset. The spreadsheet has columns for State_Name, District_Name, Crop_Year, Season, Crop, Area, and Production. The data is organized into rows, with the first row being the header. The spreadsheet shows data for various states and districts, including Andaman and Nicobar Islands, Uttar Pradesh, and Madhya Pradesh. The 'Area' and 'Production' columns contain numerical values.

State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
Andaman	NICOBARS	2000	Kharif	Areca nut	1254	2000
Andaman	NICOBARS	2000	Kharif	Other Kha	2	1
Andaman	NICOBARS	2000	Kharif	Rice	102	321
Andaman	NICOBARS	2000	Whole Year	Banana	176	641
Andaman	NICOBARS	2000	Whole Year	Cashewnu	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 pot	5	15
Andaman	NICOBARS	2000	Whole Year	Tapioca	40	169
Andaman	NICOBARS	2001	Kharif	Areca nut	1254	2061
Andaman	NICOBARS	2001	Kharif	Other Kha	2	1
Andaman	NICOBARS	2001	Kharif	Rice	83	300
Andaman	NICOBARS	2001	Whole Year	Cashewnu	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 pot	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	62	12.5