**Fundamental knowledge of the “Agriculture data”**

The dataset is a collection of information about various agricultural, marine, and handicraft products from different states and districts in India. The data appears to be descriptive, providing a detailed overview of each product.

Python

import pandas as pd

# Load the CSV file into a DataFrame

df = pd.read\_csv('Agricultural data.csv')

# Display the first 5 rows to understand the data

print(df.head().to\_markdown(index=False, numalign="left", stralign="left"))

# Display the data types and non-null counts of each column

print(df.info())

Code output

| State | Product | District | LGD Code | Category | Sector | Description | GI Status | Photo | Ministry/ Department |

|:----------------------------|:---------------------------------|:-------------------------|:-----------|:-----------|:------------|:-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|:------------|:--------------------------------------------------------------------------------------|:-------------------------------------------|

| Andaman and Nicobar Islands | Coconut & Coconut based products | Nicobars | 603 | Primary | Agriculture | The Nicobar Islands are known for their abundant coconut plantations, that make available coconut oil, coconut milk, coconut water, and copra (dried coconut meat used for oil extraction). The islands are also known for their traditional handicrafts made from coconut shells and fibers, such as baskets and mats. | No | https://drive.google.com/file/d/1ZKef7mLS4iub1jTlqv9gbc8oBhJ8ljct/view?usp=drive\_link | Ministry of Agriculture & Farmers' Welfare |

| Andaman and Nicobar Islands | Fisheries/Marine products | North and Middle Andaman | 632 | Primary | Marine | The North and Middle Andaman islands, are known for their rich marine life and as a result, fishing and the production of marine products is a significant part of the local economy. These products include a wide variety of fish, shellfish, and other seafood, which are popular in the local market and also exported to other parts of the country. | No | https://drive.google.com/file/d/1caPuTWCA35HHLk0aEbqhiLlPAWAlB3zQ/view?usp=drive\_link | Department of Fisheries |

| Andaman and Nicobar Islands | Marine Products | South Andamans | 602 | Primary | Marine | A delight for seafood enthusiasts, South Andaman has a thriving sector of marine products. Jumbo prawns, plump, juicy shrimps, fish cooked in an olio of masalas, stir fried squid - a panoply of marine products are safely, following top standards of hygiene are processed in this district. | No | https://drive.google.com/file/d/15DIayCr05l9033ZxDtvkQHXa\_qCGHasz/view?usp=drive\_link | Department of Fisheries |

| Andhra Pradesh | Coffee (Araku) | Alluri Sitarama Raju | 745 | Primary | Agriculture | Andhra Pradesh is one of the major coffee producing states in India. The state is known for its high-quality Arabica and Robusta coffee beans. The coffee grown in Andhra Pradesh is known for its unique flavor profile, with notes of chocolate, nuts, and spices. The coffee is mostly grown in the hilly regions at an altitude of 900-1500m above sea level. | Yes | https://drive.google.com/file/d/18ptepDrMFsm5l6OVZl\_jcwOOz9-lp\_Hb/view?usp=drive\_link | Ministry of Agriculture & Farmers' Welfare |

| Andhra Pradesh | Bamboo Craft | Alluri Sitarama Raju | 745 | Secondary | Handicraft | Crafted with finesse in Andhra Pradesh's Alluri Sitarama Raju district, Bamboo Craft exemplifies the region's rich artisanal heritage. Skillfully woven by local artisans, these pieces showcase intricate designs and sturdy construction, reflecting the eco-conscious ethos of the community. Each creation embodies a blend of tradition and innovation, making Bamboo Craft not just a product, but a testament to the cultural legacy of the region. | No | https://drive.google.com/file/d/1HT-KGuEaGPfvf7jbfYc-\_II\_o7x-T6Y-/view?usp=drive\_link | DC Handicraft |

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1241 entries, 0 to 1240

Data columns (total 10 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 State 1241 non-null object

1 Product 1241 non-null object

2 District 1241 non-null object

3 LGD Code 1241 non-null int64

4 Category 1241 non-null object

5 Sector 1241 non-null object

6 Description 1241 non-null object

7 GI Status 1241 non-null object

8 Photo 1241 non-null object

9 Ministry/ Department 1241 non-null object

dtypes: int64(1), object(9)

memory usage: 97.1+ KB

None

**Data Overview**

* **Total Entries:** The dataset contains 1241 entries, each representing a specific product.
* **Completeness:** All columns are fully populated with no missing values.
* **Structure:** The data is organized into 10 columns, each representing a different attribute of the products.

**Breakdown of Columns**

The columns can be classified by their data type and the kind of information they contain:

* **Qualitative (Text-based) Columns:**
  + State: The name of the Indian state the product is from (e.g., 'Andaman and Nicobar Islands', 'Andhra Pradesh').
  + Product: The name of the product itself (e.g., 'Coconut & Coconut based products', 'Coffee (Araku)').
  + District: The specific district where the product is sourced.
  + Category: The main category of the product, such as 'Primary' (raw goods) or 'Secondary' (processed goods).
  + Sector: The specific industry sector, such as 'Agriculture', 'Marine', or 'Handicraft'.
  + Description: A detailed textual explanation of the product, its origin, and its characteristics.
  + GI Status: A categorical variable indicating whether the product has a Geographical Indication status ('Yes' or 'No').
  + Photo: A URL linking to an image of the product.
  + Ministry/ Department: The governmental body responsible for the product.
* **Quantitative (Numerical) Column:**
  + LGD Code: This is the only numerical column, which represents the Local Government Directory code for the corresponding district. This is a discrete numerical variable, as it is a unique code used for identification.

This data is best used for descriptive analysis, such as categorizing products by state or sector, and for gaining a deeper understanding of the diverse range of agricultural and handcrafted goods across India. The textual descriptions and qualitative information are the primary focus of this dataset.

**About state column**

The **State** column in the file is a **qualitative** variable that serves as a primary geographical identifier. Its purpose is to specify the Indian state where each product originates.

### What's included in this column?

The column contains the official names of various Indian states and Union Territories, such as 'Andhra Pradesh', 'Tamil Nadu', and 'Andaman and Nicobar Islands'. Each entry in this column acts as a label, linking a specific product to its geographical source.

### Significance of the State Column

This column is fundamental to the dataset for several reasons:

* **Geographical Context:** It provides a crucial geographical context for the data, allowing you to understand the distribution of different products across India.
* **Data Grouping:** It enables you to group and analyze data by state. For example, you can use this column to find all agricultural products from a specific state or compare the number of unique products from different regions.
* **Regional Analysis:** This column is essential for any form of regional analysis, helping to answer questions like: "Which states are the top producers of a particular type of product?" or "Are there any regional specializations in certain sectors like marine or handicraft products?"
* **Primary Key:** In combination with other columns like District and Product, the State column helps uniquely identify each entry in the dataset.

**About product column**

The **Product** column contains **qualitative data**. Specifically, it holds **nominal data**.

### What does this mean?

* **Qualitative Data:** This type of data describes characteristics or qualities and is not numerical. It can't be measured or counted in a meaningful way.
* **Nominal Data:** This is a sub-type of qualitative data where categories are used for labeling or naming without any inherent order or ranking. In the Product column, each entry—like 'Coconut & Coconut based products' or 'Coffee (Araku)'—is a distinct label for a type of good. There is no numerical value associated with them, nor is there a logical order among them. For example, 'Coffee (Araku)' is not "better" or "worse" than 'Bamboo Craft' based on its name alone; they are simply different categories.

**About District column**

The **District** column contains **qualitative** data, specifically **nominal** data.

### What does this mean?

* **Qualitative Data:** This type of data describes a quality or characteristic. It's non-numerical and can't be measured.
* **Nominal Data:** This is a subcategory of qualitative data where values are used as labels or names. The districts, such as 'Nicobars' or 'North and Middle Andaman', are distinct categories without any specific order or ranking.

**About Category column**

The **Category** column contains **qualitative data**, specifically **ordinal** data.

### What does this mean?

* **Qualitative Data:** This type of data describes a characteristic or quality rather than a numerical value.
* **Ordinal Data:** This is a sub-type of qualitative data where the categories have a meaningful order or rank. In this case, the categories 'Primary' and 'Secondary' indicate a clear progression or hierarchy in the production process. A **Primary** product is a raw good, while a **Secondary** product is processed from primary goods. This establishes a natural order: Primary → Secondary.

**About Sector column**

The **Sector** column contains **qualitative data**, specifically **nominal** data.

**What does this mean?**

* **Qualitative Data:** This type of data describes a characteristic or quality. It is non-numerical and cannot be measured.
* **Nominal Data:** This is a sub-type of qualitative data where the values are distinct names or labels without any inherent order or ranking. The different sectors, such as 'Agriculture', 'Marine', and 'Handicraft', are simply categories for grouping products. There's no logical hierarchy or order between them.

**About Description column**

The **Description** column contains **qualitative data**, specifically **textual data** or **unstructured data**.

### What does this mean?

* **Qualitative Data:** This type of data is non-numerical and describes a characteristic or quality.
* **Textual/Unstructured Data:** The information is in the form of free-flowing text, consisting of sentences and paragraphs. Unlike other columns with predefined categories (like 'State' or 'Sector'), the content of the Description column does not follow a specific format. It's meant to be read and interpreted rather than being easily categorized or sorted.

This type of data is rich in information but difficult to analyze with traditional statistical methods. It's often used for tasks like:

* **Content Analysis:** Extracting themes, keywords, or sentiment from the descriptions.
* **Text Mining:** Using algorithms to find patterns and relationships within the text.
* **Providing Context:** Giving a comprehensive overview of each product that simple categories can't capture.

**About GI Status column**

The **GI Status** column contains **qualitative data**, specifically **nominal** data.

### What does this mean?

* **Qualitative Data:** This type of data describes a characteristic or quality, rather than a numerical value.
* **Nominal Data:** This is a sub-type of qualitative data where the values are used as labels or names for categories that do not have any intrinsic order or ranking. In this case, 'Yes' and 'No' are simply distinct labels indicating the presence or absence of a Geographical Indication status. While they are a binary choice, there is no hierarchy implied (one is not "better" or "more" than the other).

**About Photo column**

The **Photo** column contains **qualitative data**, specifically **unstructured data** in the form of Uniform Resource Locators (URLs).

### What does this mean?

* **Qualitative Data:** This type of data is non-numerical and describes a characteristic or quality.
* **Unstructured Data:** This refers to information that doesn't have a pre-defined format. While URLs have a specific structure, the column itself is a collection of unique links, not a set of categories. The actual value of the data lies in the image at the other end of the URL, not in the text of the URL itself.

The URLs are essentially pointers to external image files. They provide a visual representation of each product, which is a key qualitative attribute of the data.

**About Minstry/Department column**

The **Ministry/Department** column contains **qualitative data**, specifically **nominal** data.

### What does this mean?

* **Qualitative Data:** This type of data describes a characteristic or quality. It is non-numerical and cannot be measured.
* **Nominal Data:** This is a sub-type of qualitative data where values are used as labels or names for categories that do not have any inherent order or ranking. The different ministries or departments, such as 'Ministry of Agriculture & Farmers' Welfare' or 'Department of Fisheries', are distinct labels used for classification. There is no logical hierarchy or order among them.