

Pandas task

◆ Required Tasks:

1 Reading Data

✓ Tasks:

- Load the data from a CSV file using **pandas**.
- Display the first 5 rows of the dataset to ensure it is loaded correctly.
- Verify that the column names are clear and understandable.
- Check the data types of each column.

2 Cleaning Data

✓ Tasks:

- Check for **missing values (NaN)** in the dataset and decide how to handle them (replace or remove).
- Ensure numeric columns contain valid values (e.g., convert prices to numbers).
- Remove any illogical data (e.g., negative prices or unexpected values).
- Ensure there are no unnecessary duplicates in the dataset.

◆ Exploring Data (Exploratory Data Analysis - EDA)

✓ Tasks:

- Display a **summary of the dataset**, including the number of columns, rows, and data types.
- Calculate **basic statistics** such as mean, standard deviation, minimum, and maximum for numeric columns.
- Determine the number of **unique values** in categorical columns, such as product names or categories.
- Check the **distribution of values** in each column to detect outliers.
- Display the **most frequently occurring values** in each column.

◆ Feature Extraction

✓ Tasks:

- Extract year and month from the **Date** column.
- Calculate total revenue per sale ($\text{Total Revenue} = \text{Quantity} * \text{Unit Price}$).
- Calculate total cost per sale ($\text{Total cost} = \text{Quantity} * \text{Unit cost}$).
- Calculate profit ($\text{profit} = \text{total revenue} - \text{total cost}$)
- create a new column than classify profit to positive or negative or break even.

◆ Data Analysis Using the Following Questions

- 1 How many records (rows) are in the dataset?
- 2 What is the total revenue (**Total Revenue**)?
- 3 What is the average selling price (**Unit Price**)?
- 4 Which sub-category (**Sub Category**) has the highest sales?
- 5 What is the highest profit achieved?
- 6 Which states generate the most revenue?
- 7 How many sales were made each year?
- 8 How many orders were placed for each product category (**Product Category**)?
- 9 Which category generates the most profit during different months?
- 10 Which months have the highest sales? Are there seasonal trends in sales?