Task 1: Understanding Dataset & Data Types

Objective

The objective of this task is to understand what a dataset is, identify different types of data present in a dataset, and learn how to inspect them using Python libraries.

Tools Used

Python (Pandas, NumPy)

Jupyter Notebook / Google Colab

Alternatives: VS Code with Python, Kaggle Notebooks

What is a Dataset?

A dataset is a collection of related data organized in rows and columns.

Rows represent individual records or observations

Columns represent features or variables

Datasets are commonly stored in formats such as CSV, Excel, JSON, or databases.

Understanding Data Types

Data types define the kind of data stored in each column. In data analysis, data types are mainly classified as:

1. Numerical Data

Integer (int): Whole numbers (e.g., age, quantity)

Float (float): Decimal numbers (e.g., price, salary)

2. Categorical Data

String/Object: Text data (e.g., name, city, gender)

Category: Limited set of values (e.g., Yes/No, Male/Female)

3. Date & Time Data

Date, time, or timestamp values (e.g., order date, login time)

4. Boolean Data

True or False values

Understanding Dataset Using Python

Using Pandas, we can load and inspect a dataset easily.

Common steps:

Load dataset using read_csv()

View first few records using head()

Check column names and structure

Identify data types of each column

This helps in understanding:

The structure of the dataset

Type of data stored in each column

Any inconsistencies or missing values

Importance of Understanding Data Types

Helps in selecting correct analysis methods

Prevents errors during data processing

Improves data cleaning and transformation

Essential for visualization and machine learning models

Conclusion

Understanding the dataset and its data types is the first and most important step in data analysis. Using Python tools like Pandas and NumPy makes it easy to explore datasets, identify data types, and prepare data for further analysis or machine learning tasks.

If you want, I can also:

Add Python code

Customize it for college / university format

Shorten it for exam or viva submission