

PANDA'S AN ADVANCE LEVEL LIBRARY OF PYTHON .

1 What is Pandas?

Pandas is an open-source Python library used for **data manipulation and data analysis**. It is designed to work with real-world data such as Excel files, CSV files, and database tables in a fast and efficient way. We use Pandas because real data is often large, unorganized, and contains missing or duplicate values, which are difficult to handle using basic Python lists or loops. Pandas provides powerful data structures like **Series** and **DataFrame** that allow us to easily read data, clean it, filter rows and columns, perform calculations, and analyze information with very few lines of code. It is built on top of NumPy, making it fast and memory-efficient, and is widely used in fields like **Data Analysis, Data Science, Machine Learning, and AI** to turn raw data into meaningful insights.

2 Why do we use Pandas?

Real-world data is:

- Large in size
- Messy
- Contains missing values
- Has duplicate records
- Stored in multiple formats

✗ Using Python lists, loops, or dictionaries is slow and complex

✓ Pandas provides **fast, optimized, and easy tools** for handling such data

Pandas is built **on top of NumPy**, so it is:

- Fast
- Memory efficient
- Suitable for large datasets

3 Installing and Importing Pandas

Install:

```
pip install pandas
```

IMPORTING OF PANDA'S

Import:

```
import pandas as pd
```

pd is the standard alias used everywhere.

4 Core Data Structures in Pandas

Pandas mainly works with **two data structures**:

◆ A) Series (1-Dimensional Data)

What is a Series?

- A one-dimensional array
- Contains values with an index
- Similar to a **single column in Excel**

CODE EXAMPLE:-

```
python

import pandas as pd

salary = pd.Series([25000, 30000, 28000, 35000])
print(salary)
```

OUTPUT:

```
0    25000
1    30000
2    28000
3    35000
dtype: int64
```

◆ B) DataFrame (2-Dimensional Data)

What is a DataFrame?

- **Rows and columns**

- Tabular structure
- Similar to an Excel sheet or SQL table

Example:

```
data = {
    "Name": ["Aman", "Riya", "Neha"],
    "Age": [20, 21, 22],
    "Marks": [85, 90, 78]
}
```

```
df = pd.DataFrame(data)
print(df)
```

OUTPUT:-

	Name	Age	Marks
0	Aman	20	85
1	Riya	21	90
2	Neha	22	78

5 How Data Comes into Pandas (Input)

Read CSV file:

```
df = pd.read_csv("students.csv")
```

Read Excel file:

```
df = pd.read_excel("students.xlsx")
```

Read from Database (concept):

```
pd.read_sql(query, connection)
```

6 Understanding the Data (Exploration Step)

This is the first and most important step after loading data.

```
df.head()      # First 5 rows
df.tail()      # Last 5 rows
df.shape       # Number of rows and columns
df.columns     # Column names
df.info()       # Data types and null values
df.describe()   # Statistical summary
```