**Pandas (Python Data Analysis Library)**

Pandas is a data manipulation and analysis library. It is primarily based on DataFrame and Series structures.

**Pandas** Its main functions are:

| **Category** | **Topic** |
| --- | --- |
| 1. **Data Structure** | Series, DataFrame |
| 1. **Data Loading** | read\_csv(), read\_excel(), read\_json() |
| 1. **Data Inspection** | head(), tail(), info(), describe() |
| 1. **Indexing & Selecting** | df['col'], df.loc[], df.iloc[] |
| 1. **Filtering** | Logical conditions (e.g. df[df['age'] > 30]) |
| 1. **Missing Data Handling** | isnull(), fillna(), dropna() |
| 1. **Grouping & Aggregation** | groupby(), agg(), mean(), sum() |
| 1. **Merging & Joining** | merge(), concat(), join() |
| 1. **Sorting & Ranking** | sort\_values(), rank() |
| 1. **Apply Functions** | apply(), map(), lambda |
| 1. **Exporting Data** | to\_csv(), to\_excel(), to\_json() |

**Example:**

python

CopyEdit

import pandas as pd

data = {'Name': ['Moni', 'Ali'], 'Age': [24, 30]}

df = pd.DataFrame(data)

print(df[df['Age'] > 25]) # Output: Only Ali's row

**NumPy vs Pandas Summary:**

| **Feature** | **NumPy** | **Pandas** |
| --- | --- | --- |
| Structure | ndarray (array) | Series, DataFrame |
| Performance | Faster (for numeric data) | Slower but flexible |
| Use Case | Mathematical operations | Data analysis and manipulation |
| File Support | Limited | CSV, Excel, SQL, JSON |

**Next step of learning:**

**Matplotlib / Seaborn** – Data visualization  
 **Scikit-learn** – Machine Learning  
 **Statsmodels** – Statistical modeling  
 **SQL + Pandas** – Data wrangling  
 **Project Ideas** – EDA, sales analysis, health data dashboard