🕐 Hour 6 & Hour 7

🕐 Hour 6 – Pandas: Series & DataFrame Basics

1️⃣ Introduction to Pandas

🔹 What is Pandas?

Pandas is a Python library used for:

- Data analysis
- Data manipulation
- Handling **tabular data** (rows & columns)

🔹 Install & Import

```bash
pip install pandas
```

```python
import pandas as pd
```

2️⃣ Pandas Series

🔹 What is a Series?

A Series is a one-dimensional labeled array.

🔹 Create a Series

```python
s = pd.Series([10, 20, 30, 40])
print(s)
```

🔹 Series with Custom Index

```
s = pd.Series([85, 90, 78], index=["Math", "Sci", "Eng"])
```

🔹 Access Series Elements

```
print(s[0])
print(s["Math"])
```

3️⃣ Pandas DataFrame

🔹 What is a DataFrame?

A DataFrame is a 2-dimensional table (rows & columns).

🔹 Create DataFrame (Dictionary)

```
data = {
    "Name": ["Rahul", "Anita", "Suman"],
    "Marks": [85, 90, 88]
}
df = pd.DataFrame(data)
print(df)
```

🔹 Basic DataFrame Operations

Method	Use
`head()`	First 5 rows
`tail()`	Last 5 rows
`shape`	Rows & columns
`columns`	Column names
`info()`	Summary
`describe()`	Statistics

🔹 Access Columns

```
print(df["Name"])
```

🔹 Access Rows

```
df.loc[0]
df.iloc[1]
```

4️⃣ Reading Data from File

```
python

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

🔑 Hour 6 Exam Points

✓ Series = 1D
✓ DataFrame = 2D
✓ Pandas handles missing data
✓ Used in ML data preprocessing

🕐 Hour 7 – Pandas: Data Cleaning & Aggregations

1️⃣ Data Cleaning in Pandas

🔹 Missing Values (NaN)

Check Missing Values

```
python

df.isnull()
df.isnull().sum()
```

Remove Missing Values

```
python

df.dropna()
```

Fill Missing Values

```
df.fillna(0)
df.fillna(df.mean())
```

2 Removing Duplicates

```
df.duplicated()
df.drop_duplicates()
```

3 Renaming Columns

```
df.rename(columns={"Marks": "Score"}, inplace=True)
```

4 Sorting Data

```
df.sort_values("Marks")
df.sort_values("Marks", ascending=False)
```

5 Aggregation Functions

Function	Description
`sum()`	Total
`mean()`	Average
`max()`	Maximum
`min()`	Minimum
`count()`	Count

🔹 **Example**

```
df["Marks"].mean()
df["Marks"].max()
```

6️⃣ GroupBy (Very Important)

🔹 **Group Data**

```
df.groupby("Class")["Marks"].mean()
```

7️⃣ Simple Practice Programs

🔹 **Program 1: Handle Missing Data**

```
df.fillna(0, inplace=True)
```

🔹 **Program 2: Average Marks**

```
print(df["Marks"].mean())
```

🔑 Hour 7 Exam Points

✓ Data cleaning is mandatory in ML
✓ groupby() used for aggregation
✓ Pandas handles real-world data
✓ Aggregations help in data analysis

🧠 ML Connection

Pandas is used for:

Cleaning raw data

Feature selection

Data analysis before ML modeling