Training Day 10 Report

Date: 4 July 2025

Topic: Pandas – DataFrames and Data Manipulation

Overview

After learning about **Series** on Day 9, Day 10 focused on **Pandas DataFrames**, which are the most important structure in Pandas.

- A **DataFrame** is a **2D labeled data structure** (like a spreadsheet or SQL table).
- It consists of **rows and columns**, where each column is a Pandas **Series**.
- Provides powerful tools for **data cleaning, selection, and analysis**.

Key Concepts Covered

1. Creating a DataFrame

```
import pandas as pd

data = {
   'Name': ['Alice', 'Bob', 'Charlie', 'David'],
   'Age': [24, 27, 22, 32],
   'Marks': [85, 90, 75, 88]
}

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

Output:

```
     Name  Age  Marks
0   Alice   24     85
1     Bob   27     90
2 Charlie   22     75
3   David   32     88
```

2. Accessing Data

- **Columns**

```
print(df['Name'])   # Access a column
print(df[['Name', 'Marks']])   # Multiple columns
```

- **Rows**

```
print(df.loc[1])   # Access by index label
print(df.iloc[2])  # Access by row position
```

3. Data Selection & Filtering

```
# Select rows where Age > 25
print(df[df['Age'] > 25])

# Select rows where Marks >= 85
print(df[df['Marks'] >= 85])
```

4. Adding and Modifying Columns

```
df['Grade'] = ['B', 'A', 'C', 'A']   # Add new column
df['Marks'] = df['Marks'] + 5        # Modify existing column
print(df)
```

5. Basic DataFrame Operations

```
print(df.head(2))   # First 2 rows
print(df.tail(2))   # Last 2 rows
print(df.shape)     # (rows, columns)
print(df.info())    # Data types and summary
print(df.describe())# Statistical summary
```

6. Handling Missing Data

```
import numpy as np

df.loc[2, 'Marks'] = np.nan   # Introduce missing value
print(df)

# Fill missing values
```

```
df['Marks'].fillna(0, inplace=True)

# Drop rows with missing values
df.dropna(inplace=True)
```

7. Sorting and Grouping

```
# Sorting
print(df.sort_values(by='Marks', ascending=False))

# Grouping
print(df.groupby('Grade')['Marks'].mean())
```

Summary

On 4 July (Day 10), we explored:

✔ Creation of **DataFrames** from dictionaries.
✔ Data **accessing techniques** (loc, iloc, column selection).
✔ **Filtering rows** with conditions.
✔ **Adding, modifying, and handling missing data**.
✔ Performing **sorting** and **grouping** operations.
✔ Using info(), describe() for quick insights into data.

Learning Outcomes

✅ Understood the **structure of Pandas DataFrames** and their usage.
✅ Learned how to **select, filter, and manipulate** rows and columns.
✅ Gained knowledge of **data cleaning techniques** (handling missing values).
✅ Practiced **sorting and grouping**, useful for real-world analysis.
✅ Built a strong foundation for **data analysis and visualization** tasks.