Training Day 11 Report

Date: 7 July 2025

Topic: Pandas with Real Datasets (Titanic Example) & File Handling in Pandas

Overview

On Day 11, we explored **real-world data analysis using Pandas**.

- The dataset used was the **Titanic dataset** (built into Seaborn).

- Learned how to perform **basic data exploration, filtering, aggregation, and statistics**.

- Also practiced loading external data files (CSV), though one error (FileNotFoundError) occurred.

Key Concepts Covered

1. Importing Libraries and Dataset
import seaborn as sns
import pandas as pd

# Load Titanic dataset from seaborn
titanic = sns.load_dataset("titanic")

- **Seaborn** provides some datasets like Titanic for testing.

- Data was loaded successfully into a Pandas **DataFrame**.

2. Counting Total Passengers
total_passengers = len(titanic)
print("Total passengers:", total_passengers)

Output:

Total passengers: 891

3. Survival Analysis

```
survived_count = titanic['survived'].sum()
print("Survived:", survived_count)
```

- Column **survived** has values 0 (not survived) and 1 (survived).

- The sum gives total number of survivors.

4. Average Age of Passengers

```
average_age = titanic['age'].mean()
print("Average age:", average_age)
```

Output:

Average age: ~29.69 years

5. Filtering Passengers Below Age 18

```
below18 = titanic[titanic['age'] < 18].shape[0]
print("People below age 18:", below18)
```

Output:

People below age 18: 113

6. Checking the 'survived' Column

```
print(titanic['survived'])
```

- Displayed all values (0 or 1).

- Length = **891 rows** (same as passenger count).

7. Attempt to Load External CSV File

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

- Resulted in an **error**:

FileNotFoundError: [Errno 2] No such file or directory: 'pokemon.csv'

- Meaning the file was not in the working directory.

- Solution: Either place the file in the same folder as the notebook or give full path.

Summary of Learning

✔ Learned to use **Pandas with real datasets**.
✔ Performed **aggregation (sum, mean)**.
✔ Practiced **filtering rows with conditions**.
✔ Explored Titanic dataset for **basic statistics**.
✔ Understood **FileNotFoundError** in pd.read_csv() and how to resolve it.

Learning Outcomes

✅ Understood how to load built-in datasets using **Seaborn**.
✅ Gained practice in **data exploration (counts, averages, filtering)**.
✅ Learned about **real dataset handling** in Pandas.
✅ Identified importance of correct **file paths** when reading CSV files.