Note: Solution for the exercises will be on GitHub.

Day 9: Data Cleaning & Exploration using Python (Pandas)

1. Loading data with Pandas

2. Inspecting data (head, tail, info, describe)

3. Detecting missing values

4. Handling missing values

5. Filtering rows / columns

6. Basic aggregations (mean, sum, count)

7. Saving cleaned data

◆ **1. Load a CSV file**

```
import pandas as pd


df = pd.read_csv("your_file.csv")  # Replace with actual filename
```

◆ **2. View data overview**

```
print(df.head())        # First 5 rows

print(df.tail())        # Last 5 rows

print(df.info())         # Column types & non-null values

print(df.describe())     # Summary of numeric columns
```

◆ **3. Detect missing values**

```
print(df.isnull().sum())  # Count of missing values per column
```

◆ **4. Handle missing data**

```
df = df.dropna()          # Remove rows with any missing values

# or

df["Age"] = df["Age"].fillna(df["Age"].mean())  # Replace NaN in "Age" with average age
```

Note: Solution for the exercises will be on GitHub.

◆ **5. Filter data**

Rows where Age > 30

print(df[df["Age"] > 30])

Only keep Name and Age columns

print(df[["Name", "Age"]])

◆ **6. Basic aggregations**

print(df["Age"].mean())   # Average age

print(df["Salary"].sum())   # Total salary

print(df["Name"].count())   # Number of entries

◆ **7. Save cleaned data**

df.to_csv("cleaned_data.csv", index=False)

📝 **Mini Exercise**

Use a CSV file that contains columns like Name, Age, and Salary. Try:

- Checking how many missing values are in each column.

- Filling or dropping missing values.

- Filtering rows where Age > 25.

- Saving the cleaned result.