
📅 Day 3: Python – NumPy & Pandas (with Fun Data-Based Activities)

🎯 Objective

Learn the foundational data manipulation tools in Python:

- `NumPy` for numerical operations
- `Pandas` for working with structured data (CSV, Excel, etc.)

You'll also **play with data** through interactive mini projects.

🧰 1. Install and Import Required Libraries

pip install numpy pandas

import numpy as np
import pandas as pd

🔢 2. NumPy Basics

🔹 Arrays

```
import numpy as np

# Create an array
arr = np.array([1, 2, 3, 4])
print("Array:", arr)

# Reshape
arr2d = np.array([[1, 2], [3, 4]])
```

```
print("2D Array:\n", arr2d)
```

🔹 Array Operations

```
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
print("Sum:", a + b)
print("Square:", a ** 2)
print("Mean:", np.mean(a))
```

📊 3. Pandas Basics

🔹 Series

```
import pandas as pd

data = pd.Series([10, 20, 30, 40])
print(data)
print("Mean:", data.mean())
```

🔹 DataFrame

```
df = pd.DataFrame({
    "Name": ["Alice", "Bob", "Charlie"],
    "Age": [25, 30, 35],
    "Score": [85, 90, 95]
})
print(df)
```

🔹 CSV Loading

Use sample data or your own CSV:

```
df = pd.read_csv("https://people.sc.fsu.edu/~jburkardt/data/csv/airtravel.csv")
print(df.head())
```

🎮 4. Data Games and Mini Projects

📌 Game 1: Random Dice Roll Analysis with NumPy

Concepts: NumPy arrays, random generation, value counts

```
import numpy as np

rolls = np.random.randint(1, 7, size=1000)  # simulate 1000 rolls
unique, counts = np.unique(rolls, return_counts=True)

for val, count in zip(unique, counts):
    print(f"Face {val}: {count} times")

print("Most common face:", unique[np.argmax(counts)])
```

📌 Game 2: Student Score Analyzer with Pandas

Concepts: DataFrame, filtering, slicing, aggregation

```
import pandas as pd

data = {
    "Name": ["Alice", "Bob", "Charlie", "David"],
    "Math": [85, 78, 92, 60],
    "Science": [89, 76, 95, 70],
    "English": [91, 80, 85, 72]
}

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

# Average score per student
df["Average"] = df[["Math", "Science", "English"]].mean(axis=1)
print(df[["Name", "Average"]])

# Students scoring above 85 in Math
print(df[df["Math"] > 85])
```

📌 Game 3: CSV Data Quiz

Use a CSV with country data or scores. Ask questions like:

- “Which country has the highest population?”
- “What is the average literacy rate?”

```
url = "https://raw.githubusercontent.com/datasets/population/master/data/population.csv"
df = pd.read_csv(url)
df_latest = df[df['Year'] == df['Year'].max()]

print("Top 5 Populous Countries:")
print(df_latest.sort_values("Value", ascending=False).head(5))
```
