

## Day 3 Training Report

25 June 2025

### Python Libraries: NumPy, Pandas, Matplotlib — Overview and Basic Operations

On Day 3, the focus was on exploring three of the **most essential Python libraries** used in Artificial Intelligence and Data Science — **NumPy**, **Pandas**, and **Matplotlib**. These libraries provide powerful tools for **data manipulation**, **analysis**, and **visualization**, forming the foundation for working with datasets efficiently.

#### 1. NumPy (Numerical Python)

**NumPy** is a fundamental library for **numerical computations** in Python. It provides support for **multidimensional arrays** (ndarrays), mathematical functions, and linear algebra operations. Unlike Python lists, NumPy arrays are **faster**, **more memory-efficient**, and allow vectorized operations.

#### Basic Operations with NumPy:

```
import numpy as np

# Creating arrays
arr = np.array([1, 2, 3, 4, 5])
print(arr)

# Array operations
print(arr + 5)      # Add 5 to each element
print(arr * 2)      # Multiply each element by 2

# 2D arrays
matrix = np.array([[1, 2], [3, 4]])
print(matrix.shape) # Output: (2, 2)
```

#### Key Features:

- Fast mathematical computations on arrays
- Useful functions like `mean()`, `sum()`, `sqrt()`, etc.
- Linear algebra, random number generation, reshaping arrays

#### 2. Pandas

**Pandas** is a powerful library for **data manipulation and analysis**. It introduces two main data structures:

- **Series** — One-dimensional labeled array
- **DataFrame** — Two-dimensional labeled data structure (like a table or spreadsheet)

#### Basic Operations with Pandas:

```
import pandas as pd

# Creating a DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie'], 'Age': [25, 30, 35]}
df = pd.DataFrame(data)

print(df)          # Display the DataFrame
print(df['Name']) # Access a column
print(df.describe()) # Summary statistics
```

### Key Features:

- Data cleaning and transformation (handling missing values, filtering, grouping)
- Easy loading of data from CSV, Excel, JSON, and SQL
- Integration with NumPy and Matplotlib for analysis and visualization

## 3. Matplotlib

**Matplotlib** is the most commonly used library for **data visualization** in Python. It allows users to create a variety of plots to better understand and present data.

### Basic Plotting with Matplotlib:

```
import matplotlib.pyplot as plt

# Simple line plot
x = [1, 2, 3, 4, 5]
y = [10, 20, 25, 30, 40]

plt.plot(x, y)
plt.title("Simple Line Plot")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

### Other Plot Types:

- **Bar charts** for categorical comparisons
- **Scatter plots** for relationship visualization
- **Histograms** for distribution analysis

## Conclusion

By the end of Day 3, we **gained a clear understanding of NumPy, Pandas, and Matplotlib** and practiced their **basic operations**. NumPy enhanced numerical capabilities, Pandas simplified data handling, and Matplotlib enabled clear visual representation of data. These three libraries will play a critical role in upcoming modules involving data preprocessing, model building, and analysis.