DATA MANIPULATION

What is Data Manipulation?

- Data manipulation is the process of organizing or arranging data in order to make it easier to interpret.

- Data manipulation in Python involves performing various operations on data to extract, transform, clean, and analyze it.

- The key feature of data manipulation is enabling faster business operations and also emphasize optimization in the process.

Most commonly used Python Libraries in Data Manipulation:

✓ **NumPy**

✓ **Pandas**

✓ **Matplotlib and Seaborn**

Numpy

- **NumPy is short for Numerical Python.**

- **NumPy is a fundamental library for numerical computations in Python.**

- **It provides support for multidimensional arrays and a wide range of mathematical functions, making it essential for data manipulation and scientific computing.**

Example of using NumPy:

import numpy as np

Create a NumPy array

data = np.array([1, 2, 3, 4, 5])

Perform operations on the array

mean = np.mean(data)

Pandas

- Pandas is a powerful library for data manipulation and analysis.
- It provides data structures like DataFrames and Series, which make it easy to work with tabular data, perform data cleaning, filtering, aggregation, and more.

Example of using Pandas:

```
import pandas as pd
# Create a DataFrame
data = pd.DataFrame({'A': [1, 2, 3], 'B': [4, 5, 6]})
# Perform operations on the DataFrame
mania = data['A'].mean()
```

Matplotlib And Seaborn

- These libraries are used for data visualization in Python, allowing you to create various types of plots and charts to explore and present your data visually.

Example of using Matplotlib:

import matplotlib.pyplot as plt

Create a simple line plot

x = [1, 2, 3, 4, 5]

y = [10, 12, 5, 8, 9]

plt.plot(x, y)

plt.xlabel('X-axis')

plt.ylabel('Y-axis')

plt.title('Simple Line Plot')

plt.show()

Advantages

- **Easy to Use**
- **Simple Data Merging**
- **Scalability**
- **Open-source**
- **Versatility**
- **Cross Platform Compatibility**
- **Data is Flexible**

Disadvantages

- Slow Execution Speed

- Large Memory Consumption

- Dependency Management

- Not suitable for Mobile and Game Development

- Less suitable for real time Application