Objectives:

- **Use Python libraries** to handle and analyze data.
- **Learn to perform basic data manipulation** with pandas.
- Understand **data analysis techniques** such as filtering, sorting, and aggregating data.
- Gain insight into **data visualization** using matplotlib.

Introduction to Pandas: Creating, Reading, and Manipulating DataFrames

What is pandas?

pandas is an open-source Python library that provides high-performance, easy-to-use data structures and data analysis tools. It is primarily used for working with tabular data in the form of DataFrames. It allows you to efficiently manipulate, clean, and analyze structured data.

- **Series**: A one-dimensional labeled array that can hold any data type (integers, strings, etc.).
- **DataFrame**: A two-dimensional, size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns).

Key Operations in pandas:

- **Creating DataFrames**: You can create a DataFrame by loading data from various formats like CSV, Excel, or SQL databases. For example:

```
import pandas as pd

# Creating DataFrame from a dictionary
data = {'Name': ['Alice', 'Bob', 'Charlie'],
        'Age': [25, 30, 35],
        'City': ['New York', 'Los Angeles', 'Chicago']}

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

Reading Data: pandas allows you to load data from multiple sources, such as:

```
# Read data from a CSV file
df = pd.read_csv('data.csv')
```

Manipulating DataFrames:

- **Selecting Columns**: You can select a specific column like this:

```
df['Age']   # Selects the Age column
```

- **Filtering Rows**: You can filter rows based on a condition:

```
df[df['Age'] > 30]   # Returns rows where Age > 30
```

- **Adding/Removing Columns**:
- ```
  df['Country'] = ['USA', 'USA', 'USA']   # Adds a new column
  df.drop('Country', axis=1, inplace=True)   # Drops the 'Country'
  column
  ```