**MatPlot Lib**

**Matplotlib: An Overview**

**Matplotlib** is a powerful **data visualization library** in Python that allows users to create a wide range of static, animated, and interactive plots. It is widely used in data analysis, machine learning, and scientific computing.

**🔹 Why Use Matplotlib?**

* Helps visualize **data trends** and **patterns** easily.
* Supports a variety of plots, including **line plots, scatter plots, bar charts, histograms, and more**.
* Works seamlessly with **NumPy, pandas, and other scientific computing libraries**.
* Provides **customization options** for colors, labels, titles, legends, and styles.

**🔹 Installing Matplotlib**

If you don’t have it installed, use the following command:

pip install matplotlib

or in Jupyter Notebook:

!pip install matplotlib

**🔹 Types of Plots in Matplotlib**

| **Plot Type** | **Usage** |
| --- | --- |
| **Line Plot** | Trends over time or sequences |
| **Scatter Plot** | Relationship between two variables |
| **Bar Plot** | Comparing categories |
| **Histogram** | Distribution of a continuous variable |
| **Pie Chart** | Proportions of different categories |

**🔹 Matplotlib’s Key Components**

1. **pyplot (plt)**: The main module that provides functions to create and customize plots.
2. **Figure**: The overall container that holds one or more plots.
3. **Axes**: The individual plotting area inside a figure.
4. **Labels, Titles, and Legends**: Customization options to make plots more readable.

**Plots**

**1️. Scatter Plot**

* **Purpose**: Shows the relationship between two continuous variables, with each point representing a pair of values.
* **Analysis Type**: Bivariate
* **Types of Variables:**  
  ✅ **Independent Variable (X-axis)**
  + **Continuous Variable** (e.g., height, temperature, advertising spend)

✅ **Dependent Variable (Y-axis)**

* + **Continuous Variable** (e.g., weight, revenue, exam scores)
* **Examples of Suitable Variables:**
  + **Height (X) vs. Weight (Y)** → Understanding body mass trends
  + **Advertising Spend (X) vs. Sales Revenue (Y)** → How marketing affects sales
  + **Temperature (X) vs. Ice Cream Sales (Y)** → Warmer weather increases sales

**2️. Pie Chart**

* **Purpose**: Displays the proportion of different categories as segments of a circle, illustrating part-to-whole relationships.
* **Analysis Type**: Univariate
* **Types of Variables:**  
  ✅ **Categorical Variable (X-axis)**
  + **Nominal Categories** (e.g., gender, product categories, political parties)

✅ **Proportions or Percentages (Y-axis is not applicable)**

* + Each category is expressed as a **percentage** of the whole
* **Examples of Suitable Variables:**
  + **Market Share of Smartphone Brands** (Apple, Samsung, Xiaomi, etc.)
  + **Proportion of Students by Field of Study** (Engineering, Science, Arts, etc.)
  + **Distribution of Transport Modes Used** (Car, Bike, Bus, Train)

**3️. Line Plot**

* **Purpose**: Shows trends over time or a sequence, with data points connected by lines.
* **Analysis Type**: Bivariate
* **Types of Variables:**  
  ✅ **Independent Variable (X-axis)**
  + **Continuous Variable** (e.g., time, temperature, distance)
  + **Ordered Categorical Variable** (e.g., months of the year, years, age groups)

✅ **Dependent Variable (Y-axis)**

* + **Continuous Variable** (e.g., sales revenue, stock prices, sensor readings)
  + **Aggregated Discrete Data** (e.g., average exam scores per year, total rainfall per month)
* **Examples of Suitable Variables:**
  + **Stock Prices Over Time (Years vs. Price in $)**
  + **Website Traffic Over Days (Days vs. Number of Visitors)**
  + **Average Monthly Rainfall (Month vs. Rainfall in mm)**

**4. Histogram**

* **Purpose**: Displays the distribution of a single continuous variable by grouping data into bins and counting the number of observations in each bin.
* **Analysis Type**: Univariate
* **Types of Variables:**  
  ✅ **Independent Variable (X-axis)**
  + **Continuous Variable** (e.g., age, income, test scores)

✅ **Dependent Variable (Y-axis)**

* + **Frequency Count** (e.g., number of students in each age group)
* **Examples of Suitable Variables:**
  + **Distribution of Student Exam Scores (Score Ranges vs. Number of Students)**
  + **Income Distribution in a Country (Income Brackets vs. Population Count)**
  + **Number of Employees by Age Group (Age Ranges vs. Number of Employees)**

**5. Bar Plot**

* **Purpose**: A **bar plot** (or **bar chart**) is a type of data visualization that represents categorical data using rectangular bars. The length or height of each bar is proportional to the value it represents. Bar plots are commonly used to compare different categories of data.
* **Analysis Type**: Univariate or Bivariate
* **Types of Variables:**  
  ✅ **Independent Variable (X-axis)**
  + **Categorical Variable** (e.g., product categories, countries, departments)

✅ **Dependent Variable (Y-axis)**

* + **Continuous Variable** (e.g., revenue, population, sales count)
  + **Discrete Variable** (e.g., number of students, number of cars sold)
* **Examples of Suitable Variables:**
  + **Sales Revenue by Product Category (Categories vs. Sales in $)**
  + **Number of Students Enrolled in Different Majors (Majors vs. Count of Students)**
  + **Average Salary by Job Title (Job Titles vs. Average Salary in $)**