**Name: Aditee Mohanty**

**Class: CSE(AI) B div**

**Roll no.: 282023**

**PRN No.: 22310631**

**Subject: Machine Learning**

**Assignment 3 – Data Visualization**

**Objective**

The objective of this assignment is to visually explore and understand the dataset used in Assignment 1 and 2 using various graph types. Data visualization techniques such as Scatter Plot, Bar Chart, Box Plot, Pie Chart, and Line Chart are applied to analyze patterns, distributions, and relationships within the data.

**Dataset Overview**

The dataset used is a heart disease dataset that contains various features such as age, chest pain type, resting blood pressure, thalassemia test results, and more. It also includes the target variable indicating the presence of heart disease.

Before visualizing:

* Null values in the "Ca" column were handled by filling them with the mean.
* Basic exploration and data cleaning were done to ensure consistency in visualizations.

**1. Scatter Plot**

sns.scatterplot(x="Age", y="MaxHR", hue="ChestPain", data=df)

**Purpose**:  
The scatter plot shows the relationship between **Age** and **Maximum Heart Rate (MaxHR)**, with points colored by different chest pain types. This helps identify patterns or trends between age and heart performance under different types of chest pain.

**2. Bar Chart**

df["ChestPain"].value\_counts().plot(kind="bar")

**Purpose**:  
The bar chart visualizes the **frequency distribution of Chest Pain types**. It shows how many patients experienced each type of chest pain, helping to identify the most and least common types among the patients.

**3. Pie Chart**

plt.pie(x=df["Thal"].value\_counts(), labels=lables, autopct="%1.1f%%")

**Purpose**:  
The pie chart represents the **proportion of patients for each type of Thallium heart scan result** (normal, fixed defect, reversible defect). It is useful for understanding the relative distribution of different thalassemia outcomes.

**4. Box Plot**

sns.boxplot(y="Age", data=df)

**Purpose**:  
The box plot displays the **age distribution** of the patients, including the median, quartiles, and potential outliers. This provides insights into the spread and central tendency of age within the dataset.

**5. Line Chart**

sns.lineplot(x="Age", y="RestBP", data=df)

**Purpose**:  
The line chart shows how **Resting Blood Pressure (RestBP)** varies with **Age**. It helps detect any visible trends or fluctuations in blood pressure across different age groups.

**Conclusion**

This assignment demonstrates the power of data visualization in understanding and analyzing datasets. Each plot type offers unique insights:

* Scatter plots reveal relationships,
* Bar and pie charts depict categorical distributions,
* Box plots highlight spread and outliers,
* Line charts show trends over a continuous variable.