Assignment 1:

Question1: (Data Analysis of Employee's performance dataset)

a) Constructing a Dataframe

Below is a dataset containing details about employees, including their demographics, work experience, and job performance. Use the following data to **create a pandas DataFrame** in Python.

ID	Age	Gender	Department	Salary ($)	Experience (Years)	Work Hours/Week	Projects Completed	Job Satisfaction (1-10)	Training Hours	Promotion Status
1	28	Male	IT	70000	3	40	5	7	20	No
2	35	Female	HR	60000	7	38	4	8	15	Yes
3	42	Male	Finance	90000	12	45	6	6	18	Yes
4	30	Female	Marketing	65000	5	42	7	9	25	No
5	50	Male	Sales	85000	15	50	8	5	12	Yes
6	41	Female	HR	62000	9	37	5	7	16	No
7	29	Male	IT	75000	4	39	6	8	22	No
8	33	Female	Finance	88000	11	44	7	6	20	Yes
9	45	Male	Sales	87000	14	48	9	5	14	Yes
10	39	Female	Marketing	67000	6	41	6	8	19	No

b) Exploratory Data Analysis (EDA)

Overview of the Dataset:

i) Determine the total **number of rows and columns** in the dataset.

ii) Generate **summary statistics** for all numerical attributes/variables.

iii) Identify and count **distinct values** in the **Department** and **Promotion Status** columns/varibales.

Handling Missing Values:

i) Check whether the dataset contains any missing values.

ii) If there are missing values, describe an appropriate strategy to handle them (**Hint: Mean, Median, or Mode**).

Visual Analysis and Interpretation:

i) Construct a **histogram** to illustrate the distribution of employee **Salaries**.

ii) Develop a **bar graph** to visualize the number of employees working in each **Department**.

iii) Create a **box plot** for **Experience (Years)** and analyze the spread of values and outliers if exist.

Question2: (Data Analysis of Heart Disease Dataset)

a) Load the attached csv file of the dataset on PHYTON using pandas

b) Basic Insights

i) How many rows and columns does the dataset have?

ii) Display the summary statistics of the numerical columns.

iii) Find the count of unique values in the **Gender** and **Smoking** columns.

c) Handling Missing Data

i) Check if there are any missing values in the dataset.

ii) If missing values exist, explain how you would handle them (Hint: Application of mean, median, and mode).

d) Data Visualization

i) Plot a **histogram** of the **Age** column.

ii) Create a **bar chart** showing the count of people with and without **Heart Disease**.

iii) Draw a **box plot** of **Cholesterol Level** and explain its distribution

Submission Instructions:

- Submit a **Jupyter/Colab Notebook (.ipynb)** containing your **Python code and observations**.
- Implement **pandas, seaborn, and matplotlib** for data handling and visualization.
- Ensure that you add explanations and observations in markdown cells.