**ASSIGNMENT : 11**

**BUILDING A LOGISTIC REGRESSION** THAT CAN PREDICT WHETHER A PERSON HAS **HEART DISEASE** BASED ON PHYSICAL FEATURES SUCH AS **AGE, SEX, CHOLESTEROL, ETC.**

1. Load the dataset (heart.csv) into a Pandas DataFrame and display the first 5 rows.
2. What method would you use to check for missing values in the dataset?
3. How can you generate a statistical summary (mean, median, standard deviation) of the numerical features in the dataset?
4. Create a bar plot to visualize the total counts of each target value (presence or absence of heart disease)?
5. How can you visualize the relationships between the following columns using a pairplot?
   1. ['age', 'trestbps', 'chol', 'thalach', 'target']
6. Why is it not recommended to create a pairplot with all features in large datasets?
7. Create a heatmap that shows the correlation between all features in the dataset?
8. How do you interpret the correlation between different features?
9. What does a high positive or negative correlation between two features indicate?
10. Separate the dataset into features (**X**) and labels (**y**)?
11. How can you split the dataset into a training and testing set using a **10% test size** and a random\_state of 101?
12. How can you apply **StandardScaler** to normalize the training and testing features while preventing data leakage?
13. How can you train a **Logistic Regression model** using Scikit-Learn?
14. Evaluate the model using accuracy