## **Problem Statement:**

One of the hospitals has a patient dataset that contains a wide range of heart-related features. This data allows hospital staff to conduct detailed analyses of heart-related conditions and treatments. We must build a logistic regression model to predict whether a patient has heart disease or not. Calculate the feature importance as well. The dataset contains data for around 303 patients.

## **Data Description:**

**age:** Age of the patient in years.

**gender:** Gender of the patient.

**cp:** Chest pain type.

**trestbps:** Resting blood pressure (in mm Hg on admission to the hospital).

**chol:** Serum cholesterol in mg/dl.

**fbs:** fasting blood sugar > 120 mg/dl (1 = true; 0 = false).

**restecg:** Resting electrocardiographic results.

**thalach:** Maximum heart rate achieved.

**exang:** Exercise induced angina (1 = yes; 0 = no).

**oldpeak:** ST depression induced by exercise relative to rest.

**slope:** The slope of the peak exercise ST segment.

**ca:** Number of major vessels (0-3) colored by fluoroscopy.

**thal:** 3 = normal; 6 = fixed defect; 7 = reversible defect.

**heart\_diagnosis:** Diagnosis of heart disease (angiographic disease status) (0 = No heart disease, >0 = heart disease).

## **Tasks/Activities List:**

Your code should contain the following activities/Analysis:

* Read the heart dataset.
* Exploratory Data Analysis (EDA) - Show the Data quality check, treat the missing values, etc if any.
* Transform the categorical data.
* Apply the Logistic Regression model.
* Print the model results.
* Get the feature importance