## **Heart Disease Data Analysis & Insight Generation**

### **Overview:**

Heart disease remains one of the leading causes of death globally, and understanding its patterns and associated risk factors is critical for prevention and diagnosis. This project focuses on analyzing the **Heart Disease UCI Dataset** to uncover meaningful insights into patient health indicators, disease likelihood, and demographic trends. It combines data cleaning, visualization, and statistical analysis to assist healthcare professionals and researchers in data-driven decision-making.

### **Brief Description:**

The dataset consists of various features such as age, gender, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, maximum heart rate achieved, and others that may influence the presence of heart disease. The target variable indicates whether or not a patient has heart disease.

The analysis was conducted using Python and SQL. Python libraries like **Pandas**, **Matplotlib**, **Seaborn**, and **Plotly** were used to perform Exploratory Data Analysis (EDA) through visualizations such as histograms, box plots, KDE plots, and count plots. Column names were renamed to be more meaningful, missing values were handled, and categorical variables were converted into interpretable labels.

In addition to Python, **SQL** was used to perform subgroup analysis, such as comparing average cholesterol across different age groups and analyzing heart disease prevalence between male and female patients. Conditional aggregation, case-based binning, and grouping operations helped generate targeted insights.

The project helps identify factors that may increase the likelihood of heart disease, such as chest pain type, cholesterol levels, and age. It also highlights differences across demographics, enabling a better understanding of at-risk populations.

Overall, this analysis bridges medical data with actionable insights, supporting early diagnosis, risk assessment, and personalized healthcare strategies.

### **Summary:**

Visual analysis of the heart disease dataset reveals several important patterns. Middle-aged individuals, particularly those aged 40 to 60, are most commonly affected, suggesting higher vulnerability during these years. Males appear more frequently in the dataset and also show a higher incidence of heart disease, indicating a potential gender disparity in risk. Chest pain type is strongly associated with heart disease—especially the asymptomatic type, which is frequently linked to positive diagnoses despite the absence of symptoms. Cholesterol levels mostly fall within normal ranges and do not clearly distinguish between diseased and non-diseased patients, implying limited predictive power on their own. Patients with heart disease tend to have lower maximum heart rates, suggesting reduced cardiovascular performance. Exercise-induced angina, ST depression during stress (oldpeak), and a downsloping ST segment are all strongly associated with heart disease, reinforcing their diagnostic importance. In contrast, fasting blood sugar levels do not show a significant difference between those with and without heart disease, indicating it may not be a key factor in this dataset.