**Summary of Heart Disease Dataset:**

The dataset consists of 303 instances (patients) and contains 14 attributes. These attributes include information about age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, the number of major vessels (0-3) colored by fluoroscopy, a thalassemia score, and the target variable indicating the presence or absence of heart disease. The attribute Characteristics was Categorical, Integer, Real. The dataset contains missing values. The heart disease dataset has four databases: Cleveland, Hungary, Switzerland, and the VA Long Beach. The dataset of Cleveland is been used for Data Pre-processing.

**Challenges Faced During Data Pre- processing:**

* One of the challenges encountered while importing the heart disease dataset from the UCI Machine Learning Repository was that the dataset format was not in a CSV file. This required additional steps to handle the alternative format.
* As not learn seaborn early on, I faced challenges while using it in the dataset for data exploration.
* Dealing with missing data is a significant challenge. You may need to decide whether to remove rows or columns with missing values, impute missing values using statistical techniques or machine learning algorithms.
* While data cleaning, duplicate records or observations in a dataset can introduce bias and affect analysis outcomes. Identifying and removing duplicate data is essential to ensure the integrity and accuracy of the analysis.
* Other challenges face was on understanding on which attributes feature scaling needs to be applied.