Heart Disease Dataset Description and Feature Engineering Proposal

The dataset we are working with includes 1,025 patient records and 14 features related to heart health. Each record represents an individual, with details such as age, sex, blood pressure, cholesterol, heart rate, and ST depression. The goal is to predict whether a person has heart disease (target = 1) or not (target = 0). While all columns are numeric, some represent categories that will need to be properly encoded before modeling.

The main objective of this project is to get the dataset ready for machine learning by applying several preprocessing steps. These include one-hot encoding for categorical features, filling any missing or null values if they appear, and scaling and normalizing numerical data to make everything consistent. We will also apply label encoding for binary columns like sex, fasting blood sugar, and exercise-induced angina.

By the end of preprocessing, the data will be clean, balanced, and ready for algorithms such as logistic regression, decision trees, or neural networks. This process should improve both the accuracy and interpretability of the final model while setting a strong foundation for future analysis of heart disease prediction.