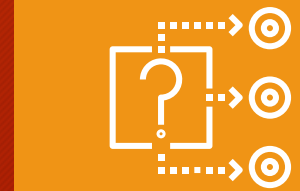
Heart Disease Diagnostic Analysis

Data Science

Upender konda
Upenderkonda2001@gmail.com

UNIFIED MENTOR INTERNSHIP



INTODUCTION

The goal of this project is to predict the presence of heart disease in patients based on various medical and demographic features. Heart disease is a leading cause of mortality worldwide, and early prediction can help in timely medical intervention, potentially saving lives. This project aims to build a predictive model that can classify whether or not a patient has heart disease based on a set of features.

Data Description

The dataset used in this project has 1025 rows and 14 columns. Each row represents an individual patient's data, with features describing various health metrics, such as age, cholesterol levels, and blood pressure. The dataset includes the following columns:

- **age**: Age of the patient
- **sex**: Gender (1 = male; 0 = female)
- **cp**: Chest pain type (categorized as 0, 1, 2, 3)
- **trestbps**: Resting blood pressure in mm Hg
- **chol**: Serum cholesterol in mg/dL
- **fbs**: Fasting blood sugar (1 if > 120 mg/dL, 0 otherwise)
- **restecg**: Resting electrocardiographic results (0, 1, or 2)
- **thalach**: Maximum heart rate achieved
- **exang**: Exercise-induced angina (1 = yes; 0 = no)

- **oldpeak**: ST depression induced by exercise relative to rest
- **slope**: Slope of the peak exercise ST segment
- **ca**: Number of major vessels (0–3) colored by fluoroscopy
- **thal**: Thalassemia (3 = normal; 6 = fixed defect; 7 = reversible defect)
- **target**: Heart disease diagnosis (1 = presence; 0 = absence)

No missing values are present in this dataset, making it ready for analysis without additional data imputation.

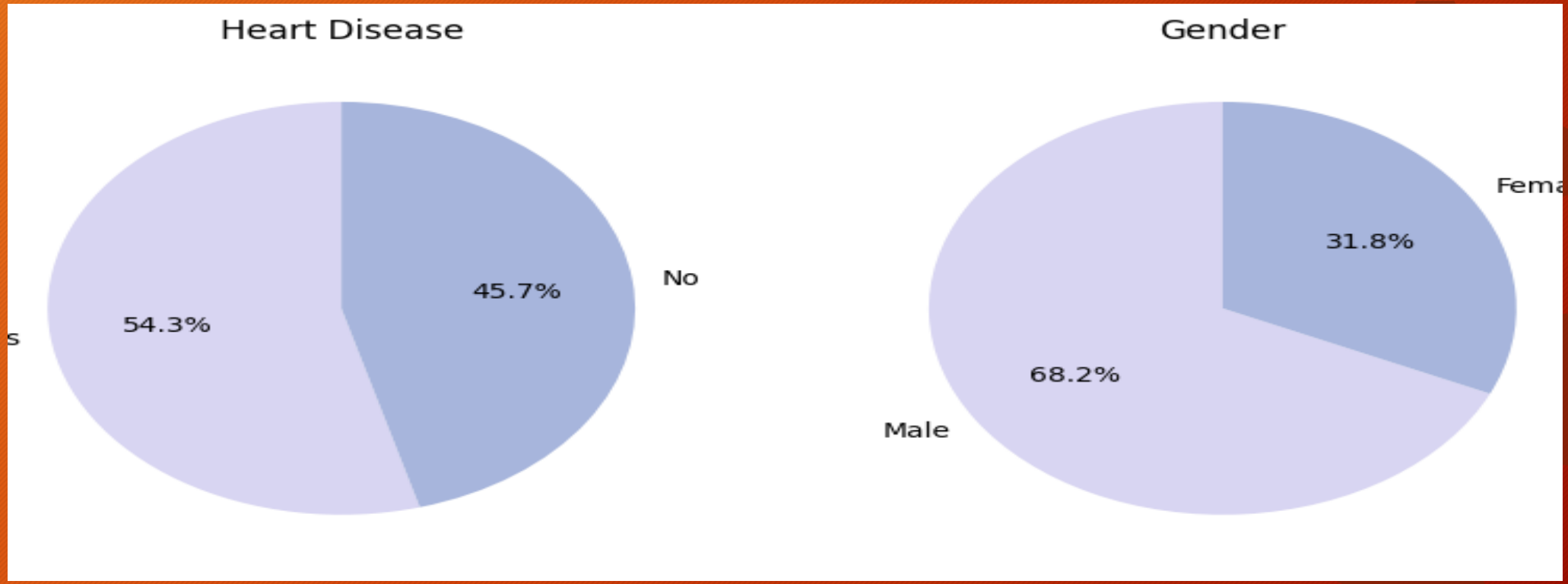
Methodology

- The methodology likely involves:
- **Exploratory Data Analysis (EDA)**: To understand relationships between features and identify patterns in the data.
- **Data Preprocessing**: Steps to prepare data for modeling, such as normalization or encoding categorical variables.
- **Model Selection**: Choosing algorithms like Logistic Regression, Decision Trees, or other classifiers suitable for binary classification.
- **Evaluation**: Using metrics such as accuracy, precision, recall, and the F1 score to evaluate model performance.

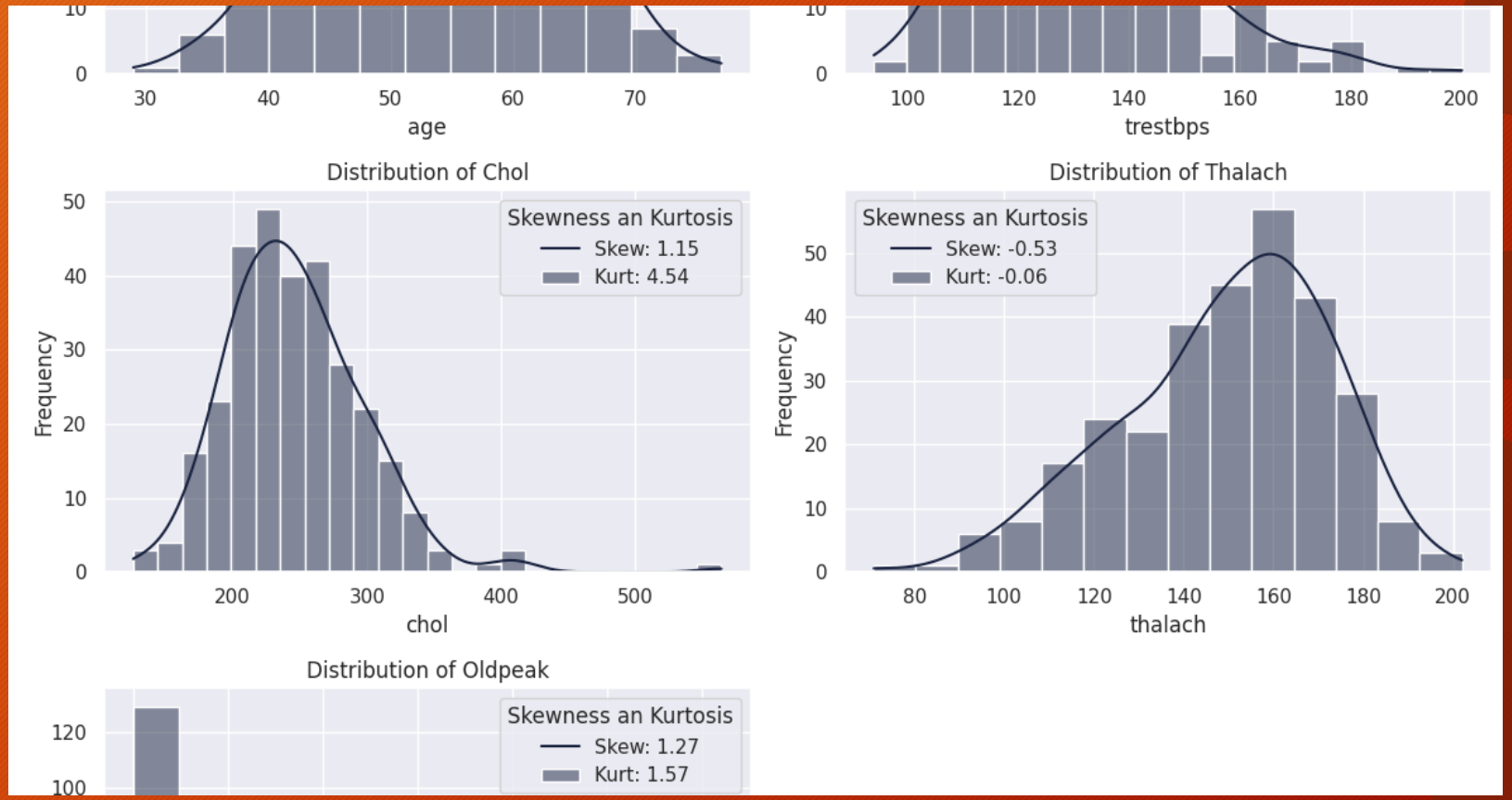
Data Preprocessing

- Data preprocessing steps might include:
- **Scaling**: Features like blood pressure and cholesterol might need normalization.
- **Encoding Categorical Data**: Transforming categorical data (e.g., chest pain type, thalassemia) into a numerical format suitable for machine learning algorithms.
- **Splitting Data**: Dividing the data into training and test sets to evaluate model performance on unseen data.

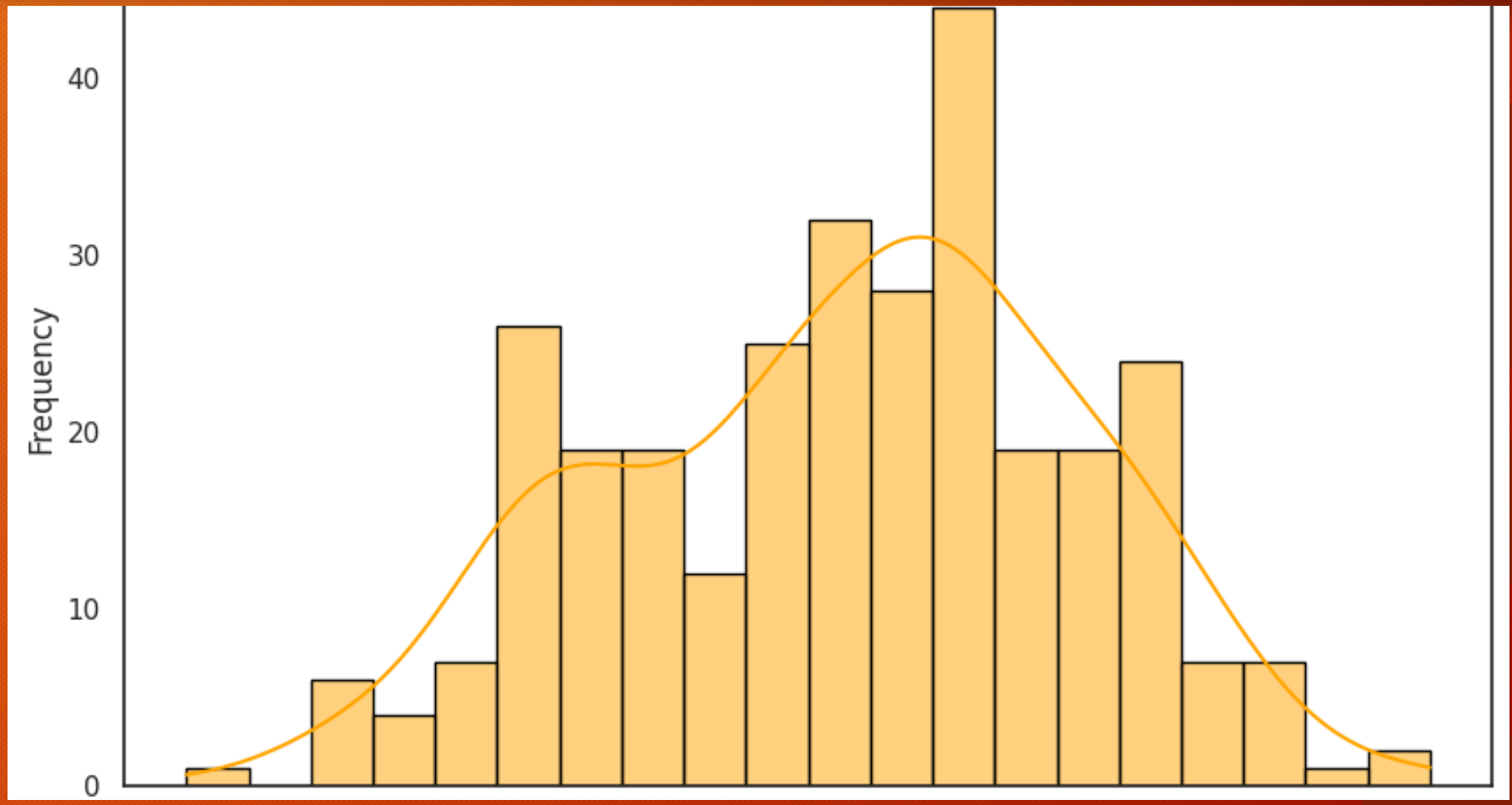
Pie Charts for HPie Charts for Heart Disease Distribution and Gender Distribution



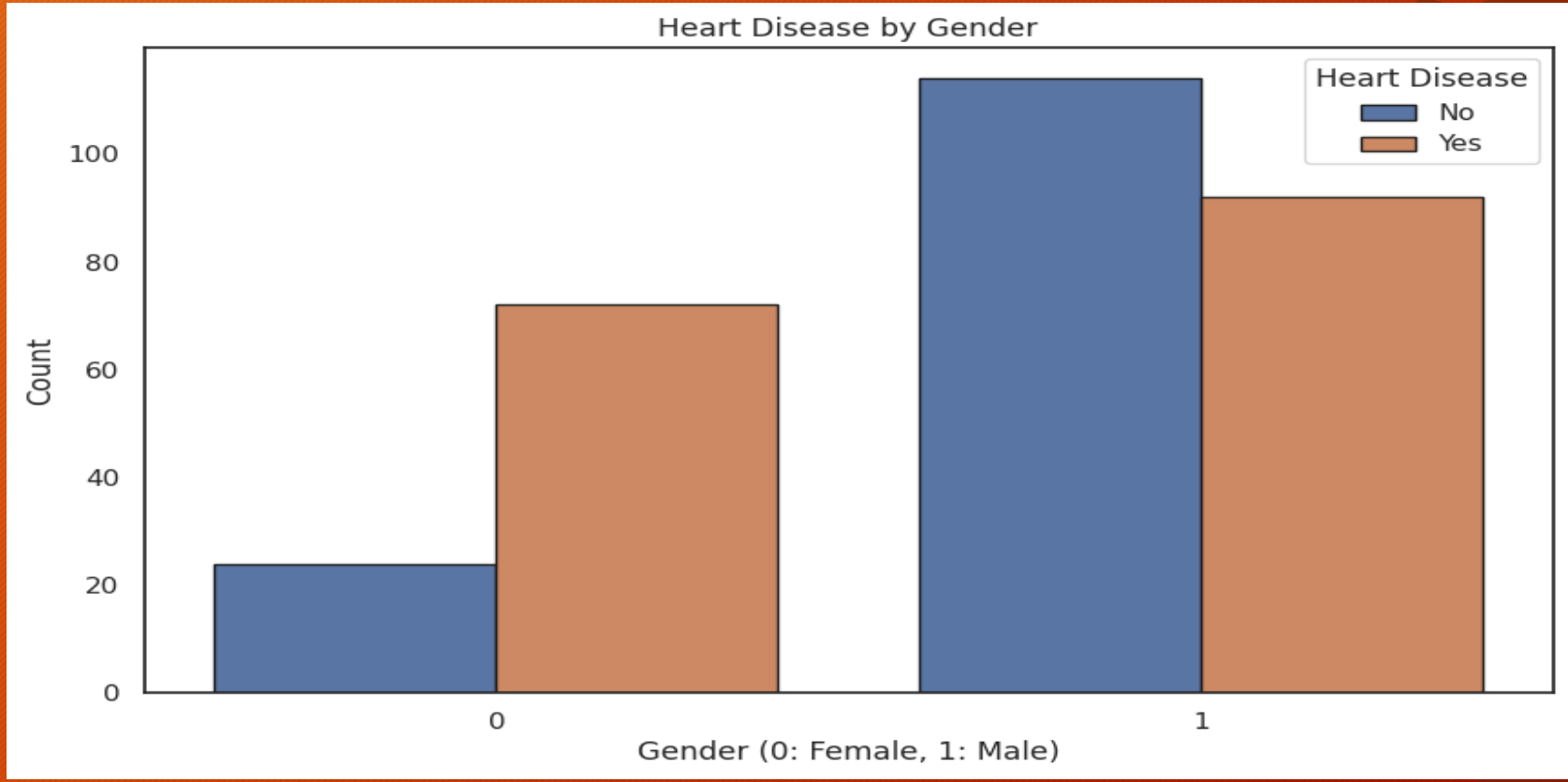
Distribution Analysis of Numerical Data



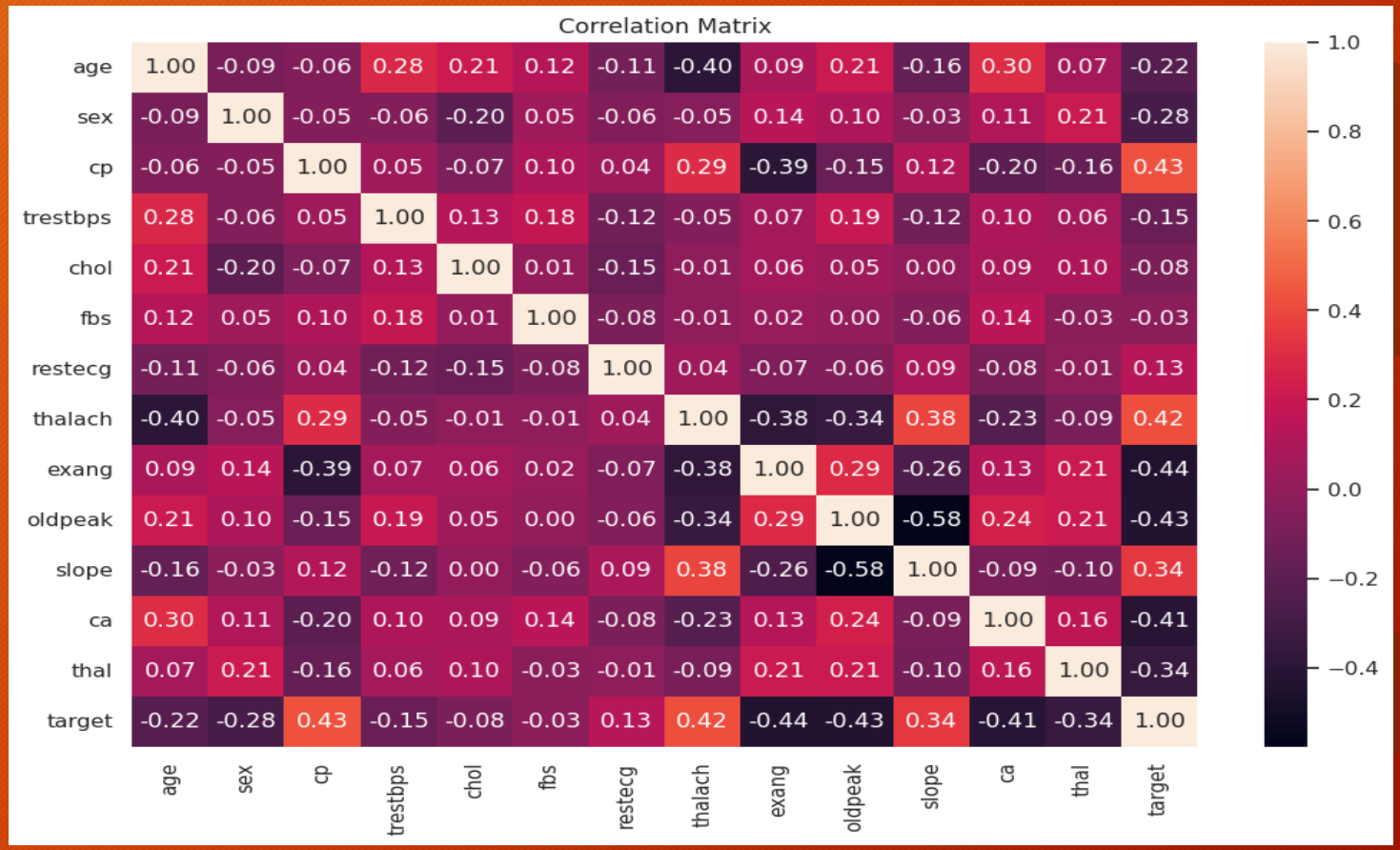
Age distribution



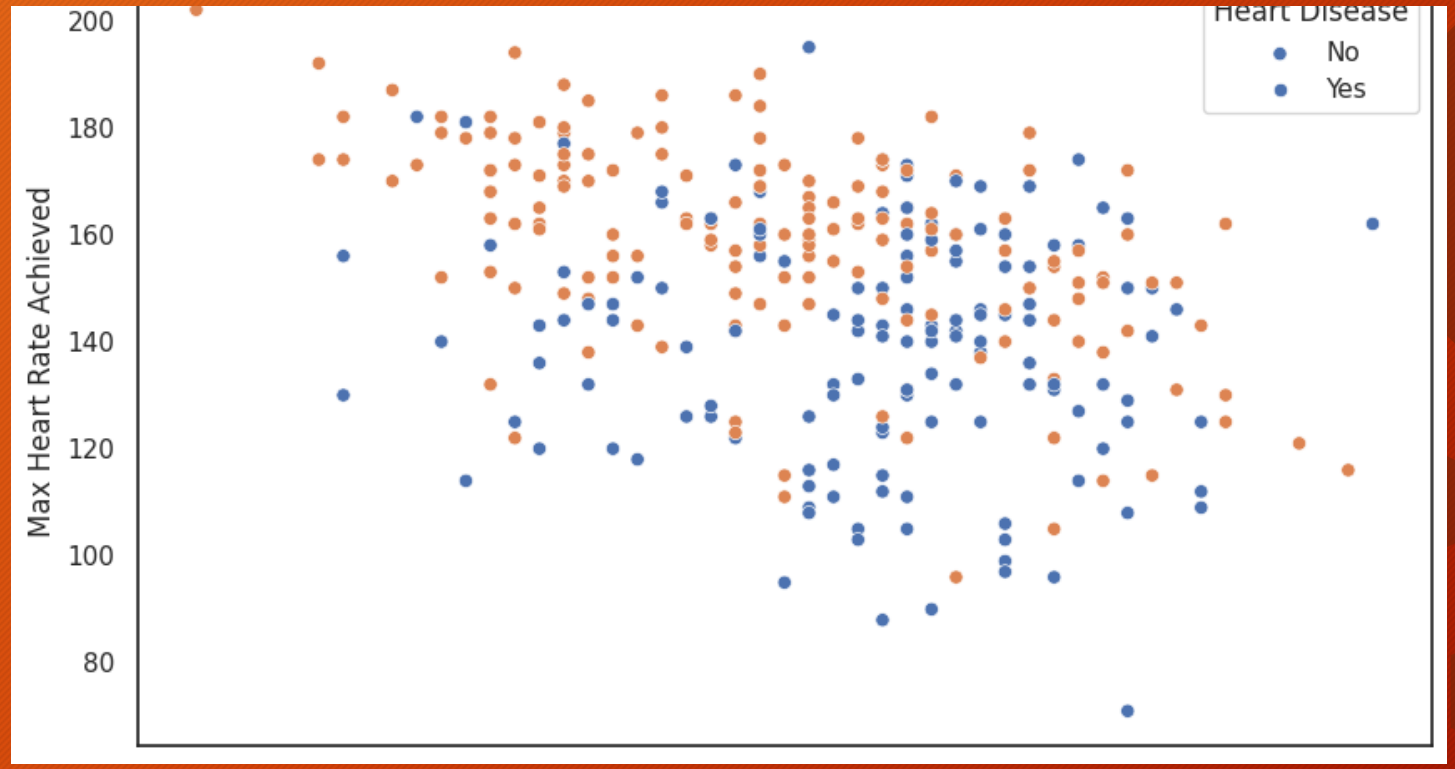
Heart disease by Gender



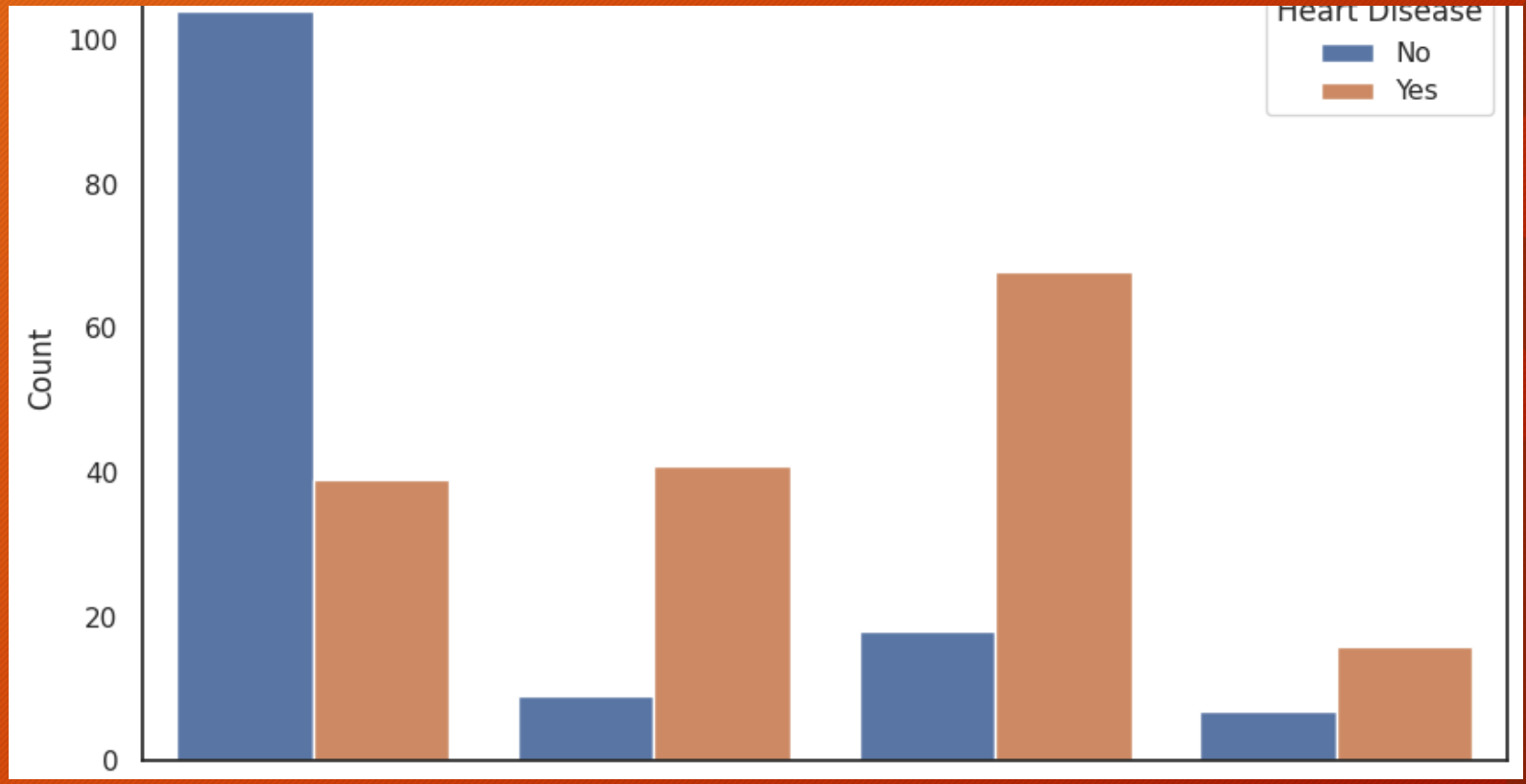
Correlation Matrix



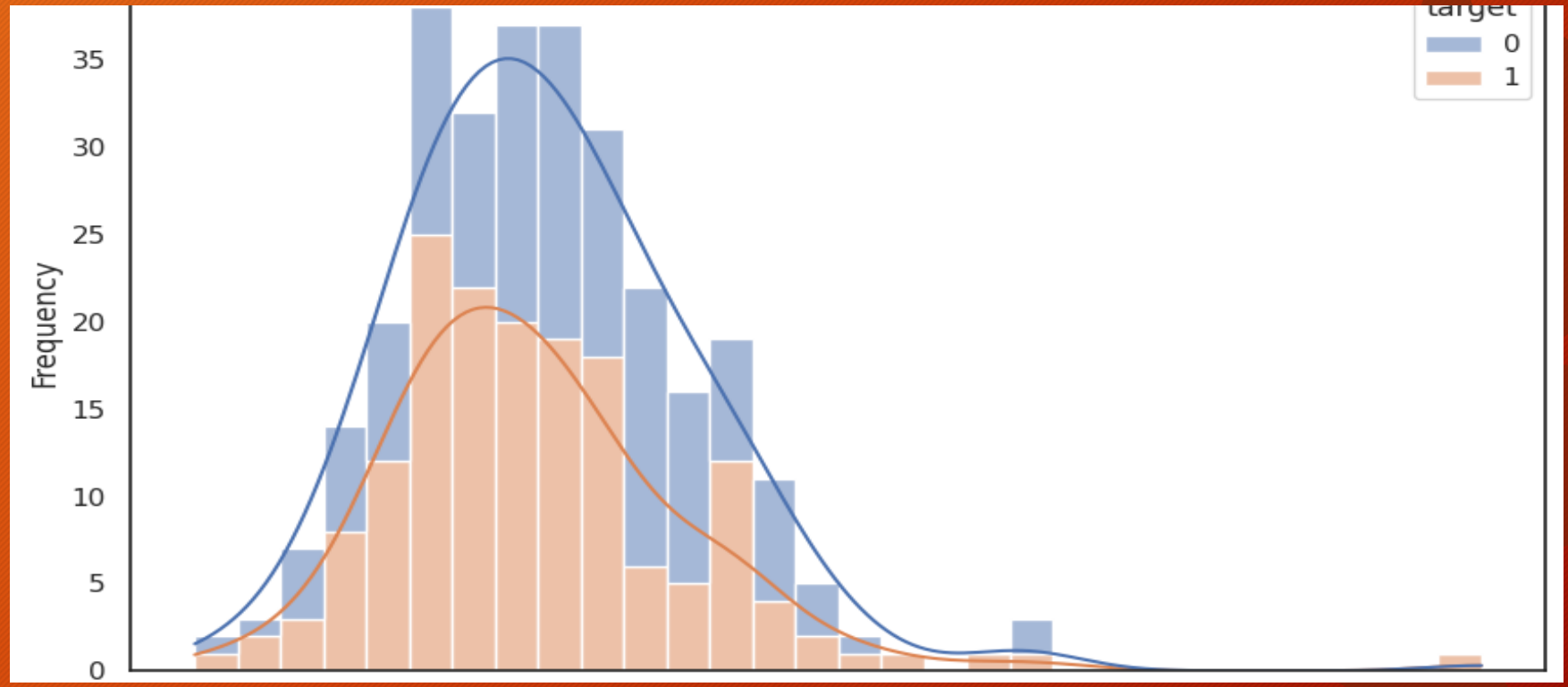
Age vs. Max Heart Rate



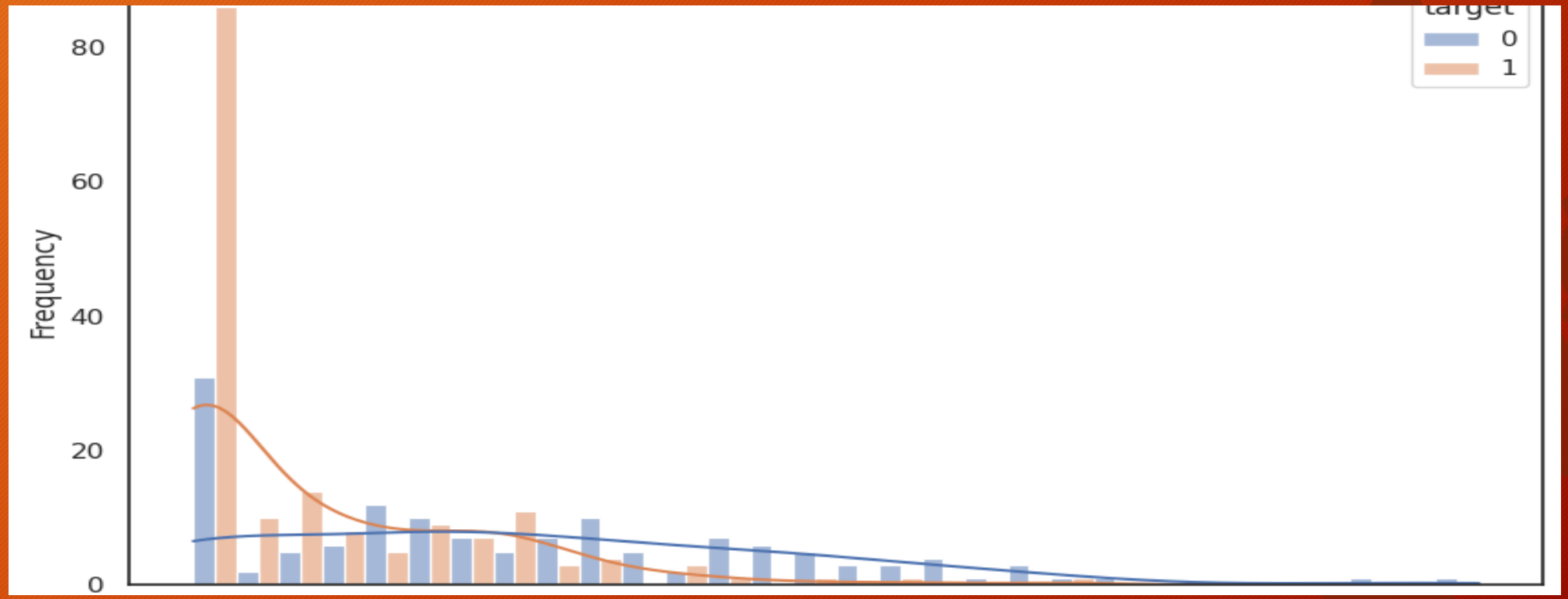
Chest pain type analysis



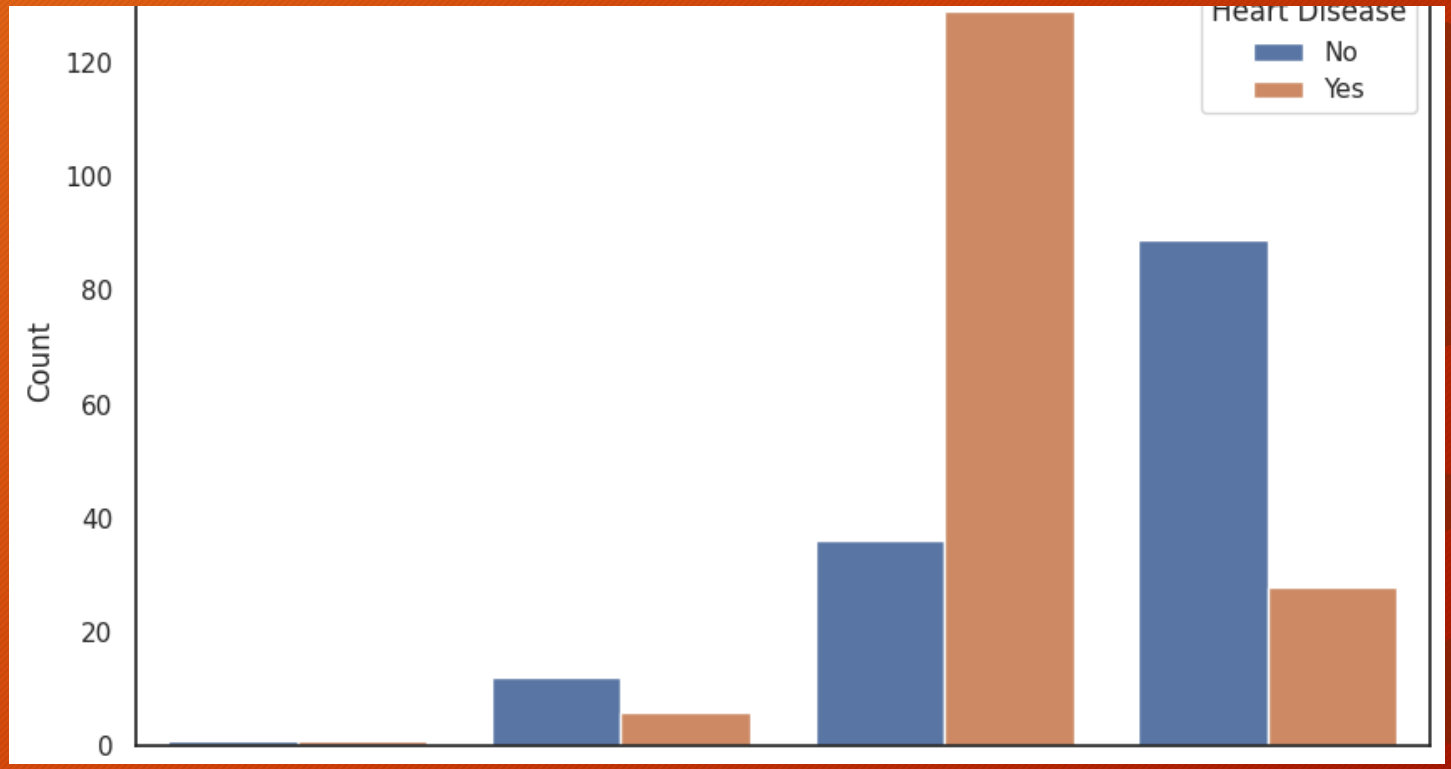
Cholesterol levels distribution



ST depression analysis



Thalassemia analysis



Results and Analysis

The analysis would involve comparing model performance and selecting the model that best balances all evaluation metrics. Graphical analysis, like ROC curves, confusion matrices, or feature importance plots, can offer insights into model effectiveness and feature contributions.

Conclusion

The selected model provides a predictive tool for assessing the likelihood of heart disease based on patient data. This model could be integrated into healthcare settings to support decision-making and early diagnosis. Future work may involve testing the model on more diverse datasets, refining feature engineering, or employing more complex models such as ensemble methods for improved accuracy.

THANK YOU