Student Name: SADIF
Student ID: 23083542
Git Hub:https://github.com/sadifshaik/Student-Name-SADIF.git

Introduction

Heart failure prediction using Python relies on machine learning and data analytics to predict the risk of heart conditions. Predictive models identify patterns and risk factors by analyzing clinical data such as age, blood pressure, and cholesterol. This approach improves early diagnosis, enhances patient outcomes, and supports proactive healthcare interventions.

Age Distribution

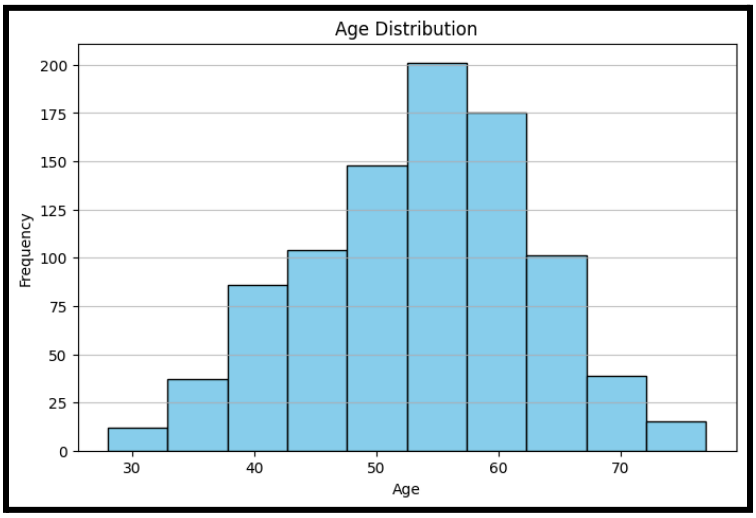

Figure 1: Histogram

The histogram shows the distribution of ages in the dataset. It groups the ages into 10 bins, which are shown using sky-blue bars. Frequency is displayed on the y-axis, and age ranges on the x-axis. To make the gridlines even more readable, insights will be obtained about the demographics and patterns of the dataset's age.

Cholesterol vs MaxHR

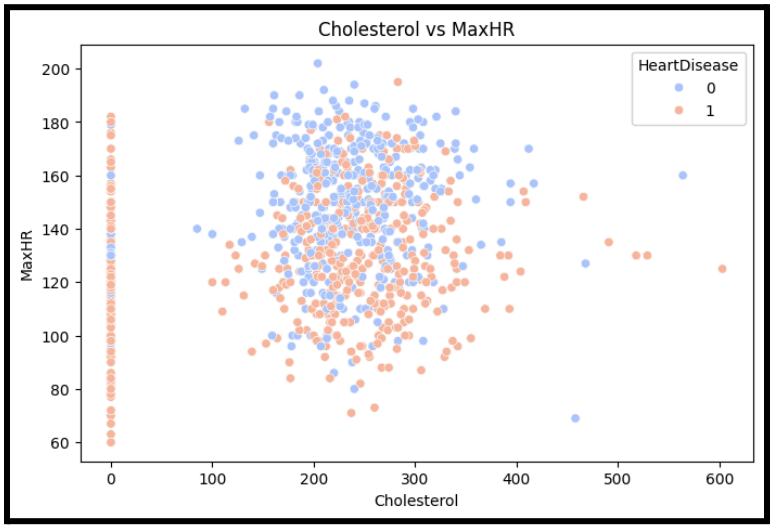

Figure 2: Scatter plot

The scatter plot shows a relation between cholesterol levels and maximum heart rate, MaxHR. The data points are colored with coolwarm color mapping as they indicate the presence of heart disease. The x-axis represents cholesterol levels, while the y-axis represents MaxHR, thus giving insights about any possible patterns or relations between these variables.

Chest Pain Type vs Age by Heart Disease

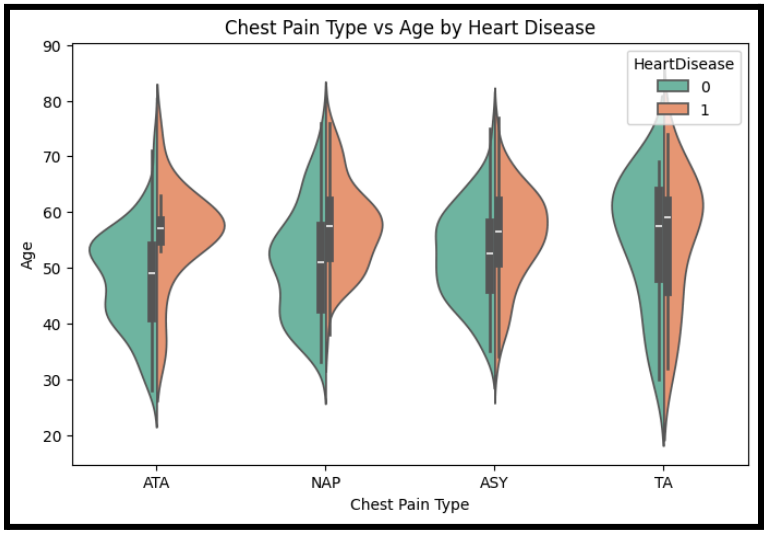

Figure 3: Violin plot

This violin plot shows the distribution of age for different types of chest pain, stratified by heart disease presence. The plot uses a split view for comparison, color-differentiated from the Set2 palette. The x-axis shows chest pain types, while the y-axis shows age and the patterns across these variables.

Elbow Method for Optimal K

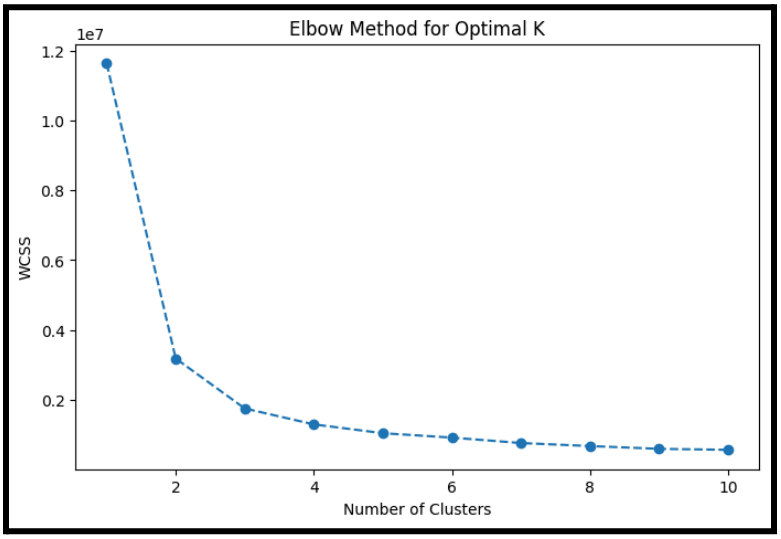

Figure 3: Elbow plot

The above figure provides the elbow plot that would determine the optimal number of clusters for K-Means clustering using the metric WCSS (Within-Cluster Sum of Squares), plotting

WCSS against the cluster counts, k from 1 to 10. The "elbow point" is where WCSS takes its greatest drop, which makes it possible to identify more meaningful clusters in the data set.

K-Means Clustering (3 Clusters)

Figure 4: K-Means clustering

The three clusters K-Means clustering are applied to this data set using age, cholesterol, and MaxHR. The results are visualized by a scatter plot where color of the clusters is assigned viridis palette, which indicates the patterns and relationships existing within formed clusters between cholesterol and MaxHR.

Line Fitting: Cholesterol vs Age

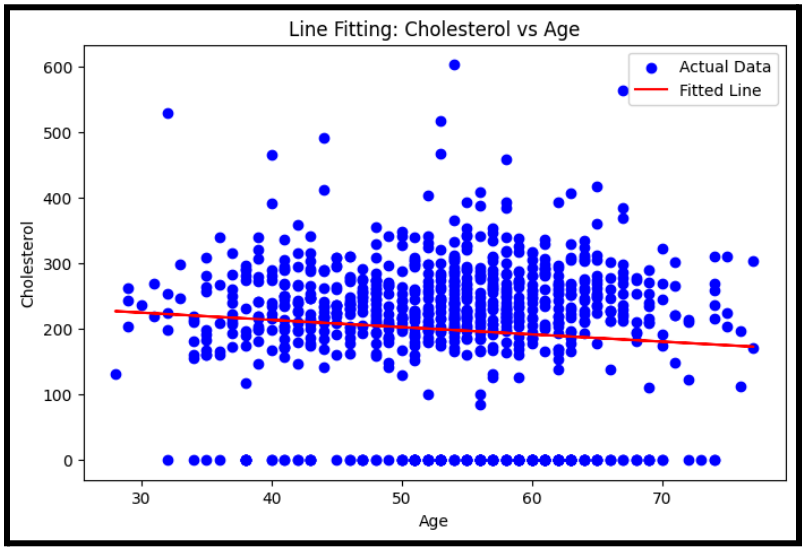

Figure 5: Linear regression

The above figure provides a linear regression model, fitting the relationship between X and y, which can be age and cholesterol levels. A fitted line will be derived and plotted along with the actual data points. The scatter plot will be in blue, and red will represent the model's predictions. The Mean Squared Error represents the accuracy of the prediction of the model.