Heart

2025-04-12

Table of contents

## 1 Introduction

The present notebook’s goal is to analyze the “Heart Failure Prediction” data set.

Data from Kaggle: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.

## 2 Context

Context

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this data set contains 11 features that can be used to predict a possible heart disease.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help. Attribute Information

Age: age of the patient [years]  
Sex: sex of the patient [M: Male, F: Female]  
ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]  
RestingBP: resting blood pressure [mm Hg]  
Cholesterol: serum cholesterol [mm/dl]  
FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]  
RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]  
MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]  
ExerciseAngina: exercise-induced angina [Y: Yes, N: No]  
Oldpeak: oldpeak = ST [Numeric value measured in depression]  
ST\_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]  
HeartDisease: output class [1: heart disease, 0: Normal]

## 3 Data load and first checks

Checking the individuals Age and Sex in the data set, we have ages in the range from 28 to 77 years with a fairly normal distribution as the similar mean and median show:

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 28.00 47.00 54.00 53.51 60.00 77.00

Checking the distribution of ages we have an uneven distribution of 193 female individuals and 725:

F M   
193 725

Checking all the medical and health related data we can see there are not missing data elements:

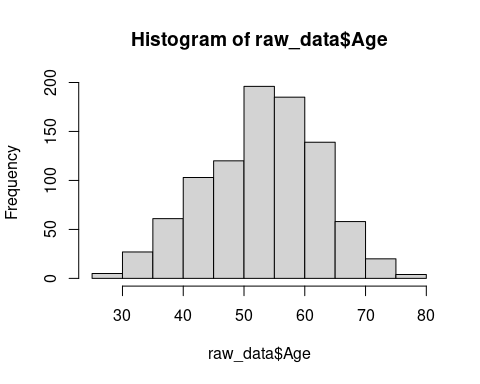
ChestPainType RestingBP Cholesterol FastingBS   
 Length:918 Min. : 0.0 Min. : 0.0 Min. :0.0000   
 Class :character 1st Qu.:120.0 1st Qu.:173.2 1st Qu.:0.0000   
 Mode :character Median :130.0 Median :223.0 Median :0.0000   
 Mean :132.4 Mean :198.8 Mean :0.2331   
 3rd Qu.:140.0 3rd Qu.:267.0 3rd Qu.:0.0000   
 Max. :200.0 Max. :603.0 Max. :1.0000   
 RestingECG MaxHR ExerciseAngina Oldpeak   
 Length:918 Min. : 60.0 Length:918 Min. :-2.6000   
 Class :character 1st Qu.:120.0 Class :character 1st Qu.: 0.0000   
 Mode :character Median :138.0 Mode :character Median : 0.6000   
 Mean :136.8 Mean : 0.8874   
 3rd Qu.:156.0 3rd Qu.: 1.5000   
 Max. :202.0 Max. : 6.2000   
 ST\_Slope   
 Length:918   
 Class :character   
 Mode :character

Finally the heart health condition is classified as 0 for healthy individuals and 1 for heart disease, distributed as below:

0 1   
410 508

## 4 Data visualization

As indicated age distribution is normal with a mean age of 54 years old.



Sex distribution of the sample is, on the other hand, a bit dissimilar.

