HEART DISEASE PREDICTION

Shiva Sankar Modala
A20517528

Keerthana Reddy Mucherla
A20517254

Nehal Juvvisetty
A20511956

Illinois Institute of Technology

CSP 571 – Data Preparation and Analysis

Prof. Jawahar Panchal

Abstract

The healthcare industry is a vital sector, and heart disease is one of the most significant health concerns worldwide. Heart disease is a leading cause of death globally and a major cause of disability. Early detection can greatly improve patient outcomes. Heart disease is the top cause of death in the United States for men, women, and individuals of all races and ethnicities. Heart disease kills one person every 34 seconds in the United States. In 2020, over 697,000 people in the United States died from heart disease, accounting for one in every five fatalities. From 2017 to 2018, the United States spent roughly $229 billion on heart disease. This covers the expense of medical services, medications, and lost productivity as a result of mortality. In this project, we will use data analysis to predict the possibility of heart disease based on different factors. The data set contains information such as age, sex, chest pain type, blood pressure, cholesterol levels, and other health metrics. By analysing this data, we aim to gain insights into factors that may contribute to heart disease and develop a model to predict the likelihood of heart disease occurrence. The process of our analysis involves several stages including data cleaning, exploratory data analysis, feature selection, model building, and performance evaluation. We will start by cleaning the data, checking for missing values, and removing any irrelevant features. Next, we will perform exploratory data analysis to better understand the relationships between different variables and identify any patterns or trends that may exist in the data. Based on our findings, we will select the most relevant features and build several machine learning models such as logistic regression, decision trees, and random forest classifiers to predict heart disease. Our goal is to use machine learning algorithms to predict heart disease risk with high accuracy, providing valuable insights to healthcare professionals and helping to improve patient outcomes.

Overview

Heart disease prediction is an important problem in healthcare, as cardiovascular disease is a leading cause of death worldwide. The goal of heart disease prediction is to develop models that can accurately identify individuals who are at high risk of developing heart disease, which can enable healthcare providers to take preventative measures and provide early interventions to reduce the risk of complications. Machine learning techniques can be applied to datasets containing demographic, clinical, and lifestyle data to develop predictive models for heart disease. These models can help identify the key risk factors associated with heart disease, identify high-risk subgroups of patients, and provide insight into novel or unexpected risk factors. Overall, heart disease prediction can play an important role in improving patient outcomes and reducing the burden of cardiovascular disease.

Background work

Heart disease is a leading cause of death worldwide and it affects individuals of all ages and genders. According to the World Health Organization (WHO), an estimated 17.9 million people die each year from cardiovascular diseases, accounting for 31% of all global deaths. With the increasing prevalence of heart disease, predicting the likelihood of an individual developing heart disease is becoming more important. Heart diseases are a group of conditions that affect the heart and blood vessels. They include conditions such as coronary artery disease, heart failure, arrhythmias, and heart valve diseases. Heart diseases are the leading cause of death worldwide, accounting for an estimated 17.9 million deaths each year. In addition to the human cost, heart diseases have a significant impact on society, including increased healthcare costs, lost productivity, and reduced quality of life for patients and their families. Risk factors for heart diseases include age, family history, high blood pressure, high cholesterol, smoking, diabetes, obesity, lack of physical activity, and poor diet. However, many of these risk factors can be controlled or managed through lifestyle changes and medical interventions. Early detection and treatment of heart diseases are crucial in reducing the risk of complications and improving outcomes for patients. This highlights the importance of developing accurate and effective prediction models to identify individuals at higher risk of developing heart diseases, as well as developing effective prevention and treatment strategies.

Data Preprocessing

Data Overview

The data for heart disease prediction was collected from the UCI Machine Learning Repository. The dataset contains information on various risk factors for heart disease, such as age, sex, blood pressure, cholesterol levels, and more. The data includes a total of 303 patients, and each patient has 14 attributes. The goal of the project is to predict the presence of heart disease in patients based on these risk factors. The data was last updated in 1988, and is a widely used dataset for predictive modeling in healthcare.

The dataset contains 14 columns and 303 rows, with each row representing an individual. The columns in the dataset include:

- **Age**: Age of the individual in years.
- **Sex**: Gender of the individual (1 = male, 0 = female).
- **Chest pain type (cp)**: Type of chest pain experienced by the individual (1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic).
- **Resting blood pressure (trestbps)**: Resting blood pressure of the individual in mm Hg.
- **Serum cholesterol (chol)**: Serum cholesterol level of the individual in mg/dl.
- **Fasting blood sugar (fbs)**: Fasting blood sugar level of the individual (> 120 mg/dl = 1, <= 120 mg/dl = 0).
- **Resting electrocardiographic results (restecg)**: Results of the resting electrocardiogram (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy).
- **Maximum heart rate achieved (thalach)**: Maximum heart rate achieved during the exercise test.
- **Exercise-induced angina (exang)**: Whether or not the individual experienced angina during the exercise test (1 = yes, 0 = no).
- **Oldpeak**: ST depression induced by exercise relative to rest.
- **Slope**: The slope of the peak exercise ST segment (1 = upsloping, 2 = flat, 3 = downsloping).
- **Number of major vessels (ca)**: Number of major vessels colored by fluoroscopy (0-3).
- **Thal**: Type of thalassemia (3 = normal, 6 = fixed defect, 7 = reversible defect).
- **Target**: Presence of heart disease in the individual (1 = yes, 0 = no).

This dataset was collected from the UCI Machine Learning Repository and is commonly used for predicting the presence of heart disease in individuals.

Packages used:

1. **Rpart, rpart.plot and ROCR**: These packages are used for building classification and regression models using decision trees. Further, we can visualize the tree structure and evaluate the performance of the models.
2. **Tidyverse**: This package consists of 6 core packages out of which the below 3 are most important for this project: dplyr: Used for data manipulation tidyr: Used for data modifications ggplot2: Used for creating powerful visualizations.

Proposed Methodology:

Data Pre-processing

An information set was gathered from University of California, Irvine (UCI) machine learning repository [1]. Later, the data is processed in accordance with requirements, the types of data gathered, the amount of processing time available, and various other considerations.

Data Cleaning

In the dataset used, it contains 303 rows and 14 attributes. No missing values, incomplete data or missing values or null values are found. No duplicate values are found. So, we did not find any need to make any changes to the data. But if the dataset is huge in the future, then we need to make necessary data cleaning to make predictions using the data

Exploratory Data Analysis:

Exploratory Data (EDA) is an approach in analyzing data sets to summarize their main characteristics using statistical graphics and other data visualization methods. This helps in getting a better understanding of the data, discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

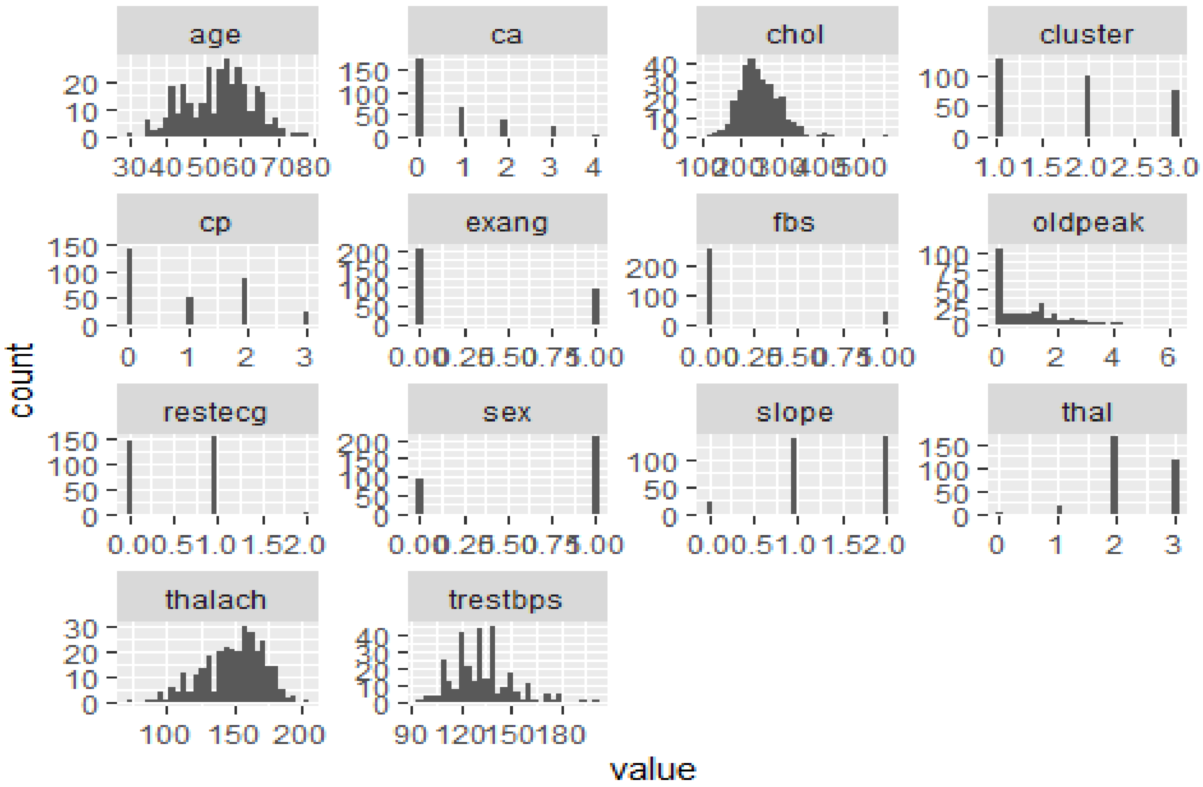
Using the EDA, we are outlining the main traits. To create the methods and connections between them, we established the relationships between the variables. By using multivariate visuals to showthe correlation between different data features. Python and R are utilized as the EDA tools.

We are going to analyze a dataset that contains medical data related to heart disease prediction. This dataset is obtained from the UCI Machine Learning Repository. The dataset includes various demographic, behavioral, and medical features of patients. It consists of a total of 14 attributes, including age, sex, chest pain type, blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and the diagnosis of heart disease status. The dataset aims to predict the presence or absence of heart disease in patients.

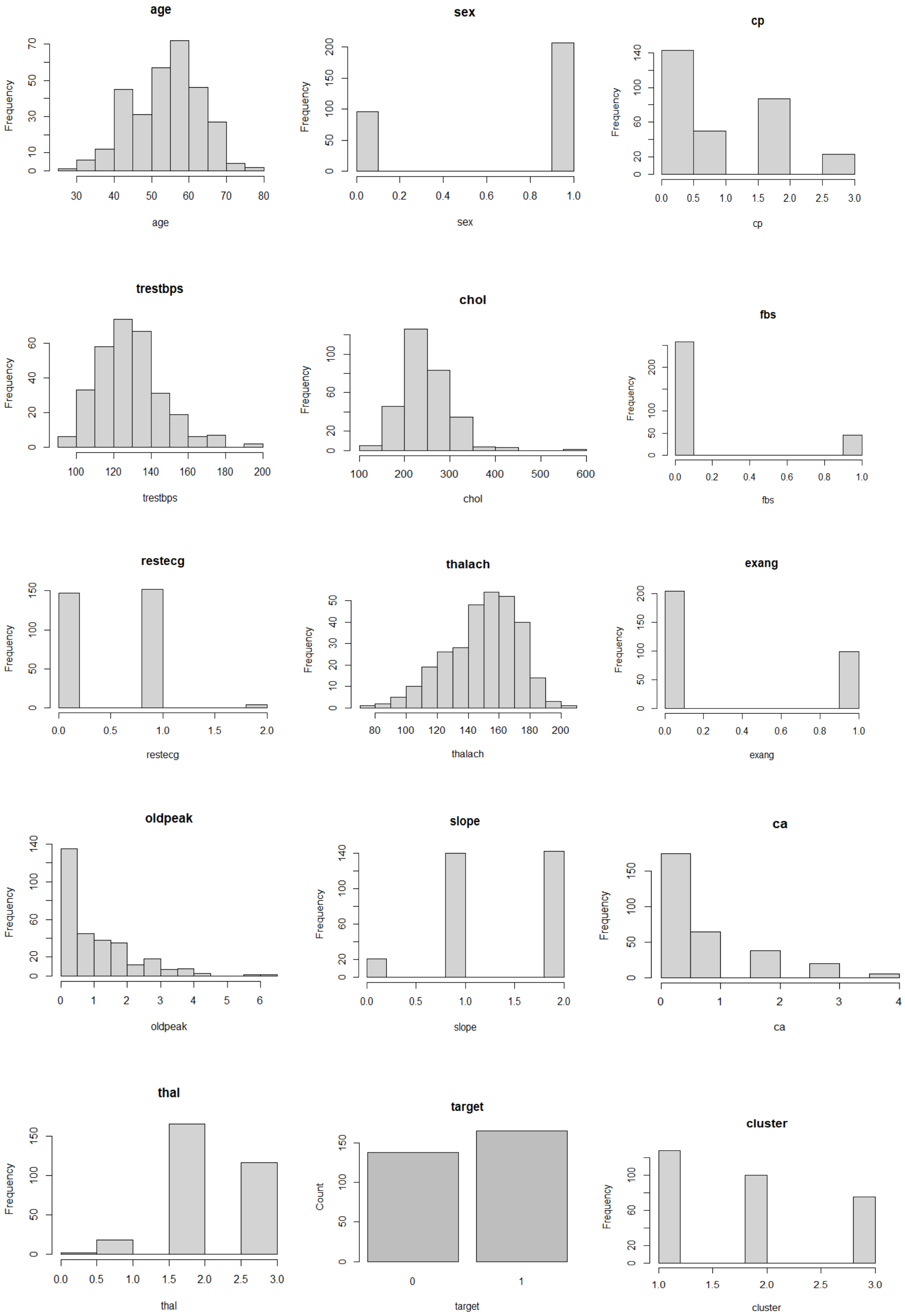
Data Preparation:

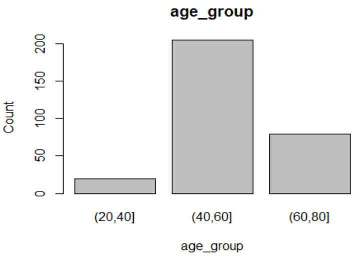
Data preparation is accomplished by converting data types, including date-time, nominal, and numerical values. By labelling the raw data into an algorithm friendly format. To create a thorough summary of the data for the data analysis process, data aggregation is done.

Histogram of all the numeric data:

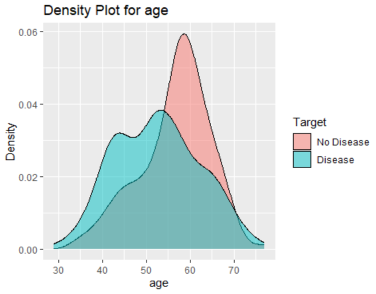


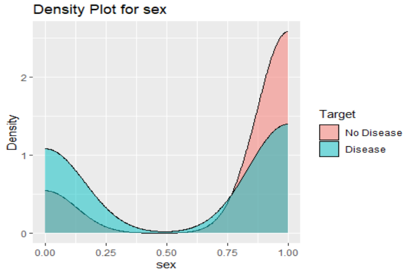
Bar plots:

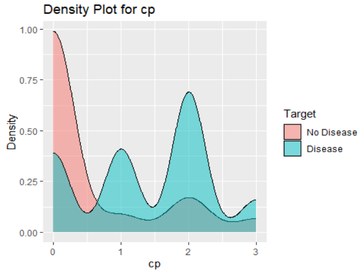

age
Frequency
0
10
30
50
70
30
40
50
60
70
80
age
sex
Frequency
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
sex
cp
Frequency
0
20
60
100
140
0.0
0.5
1.0
1.5
2.0
2.5
3.0
cp
trestbps
Frequency
0
20
40
60
100
120
140
160
180
200
trestbps
chol
Frequency
0
20
60
100
100
200
300
400
500
600
chol
fbs
Frequency
0
50
100
200
0.0
0.2
0.4
0.6
0.8
1.0
fbs
restecg
Frequency
0
50
100
150
0.0
0.5
1.0
1.5
2.0
restecg
thalach
Frequency
0
10
20
30
40
50
80
100
120
140
160
180
200
thalach
exang
Frequency
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
exang
oldpeak
Frequency
0
20
60
100
140
0
1
2
3
4
5
6
oldpeak
slope
Frequency
0
20
60
100
140
0.0
0.5
1.0
1.5
2.0
slope
ca
Frequency
0
50
100
150
0
1
2
3
4
ca
thal
Frequency
0
50
100
150
0.0
0.5
1.0
1.5
2.0
2.5
3.0
thal
target
Count
0
50
100
150
0
1
target
cluster
Frequency
0
20
60
100
1.0
1.5
2.0
2.5
3.0
cluster

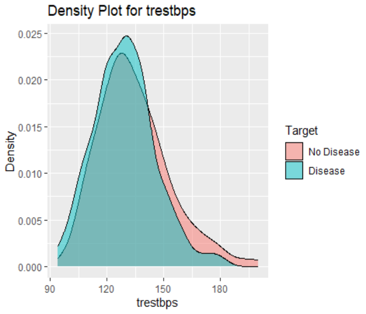

age_group

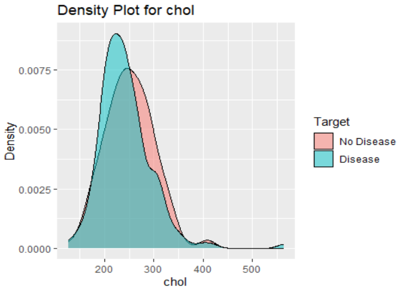
Density plots:

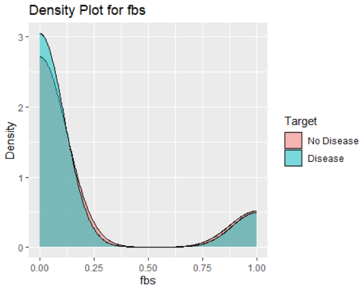

Density Plot for age

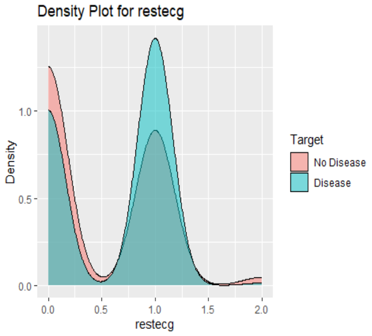

Density Plot for sex

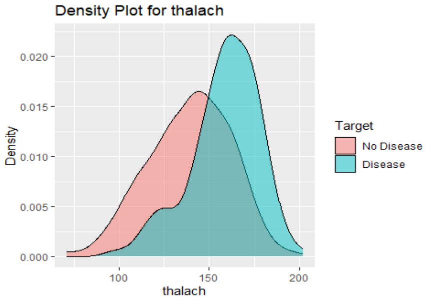

Density Plot for cp

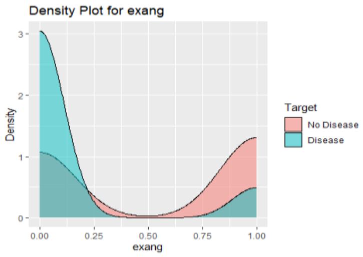

Density Plot for trestbps

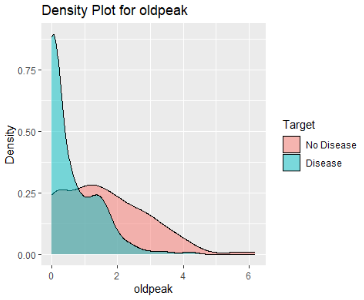

Density Plot for chol

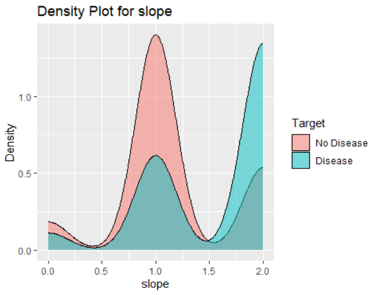

Density Plot for fbs

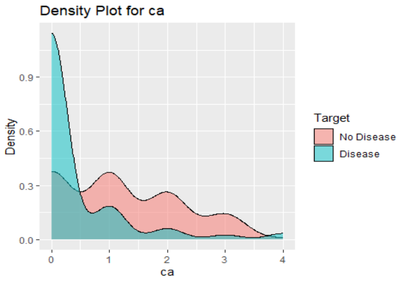

Density Plot for restecg

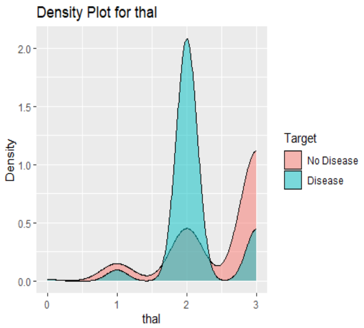

Density Plot for thalach

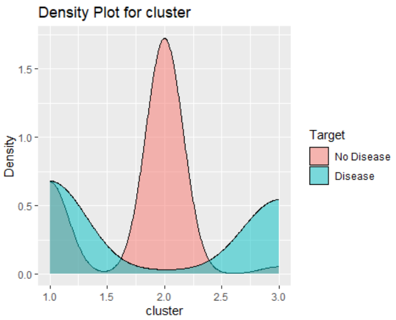

Density Plot for exang

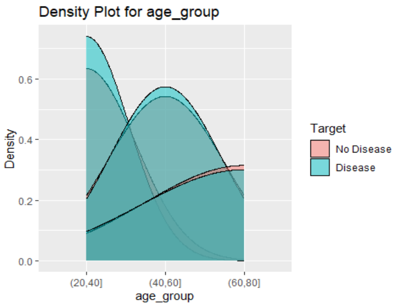

Density Plot for oldpeak


Density Plot for slope

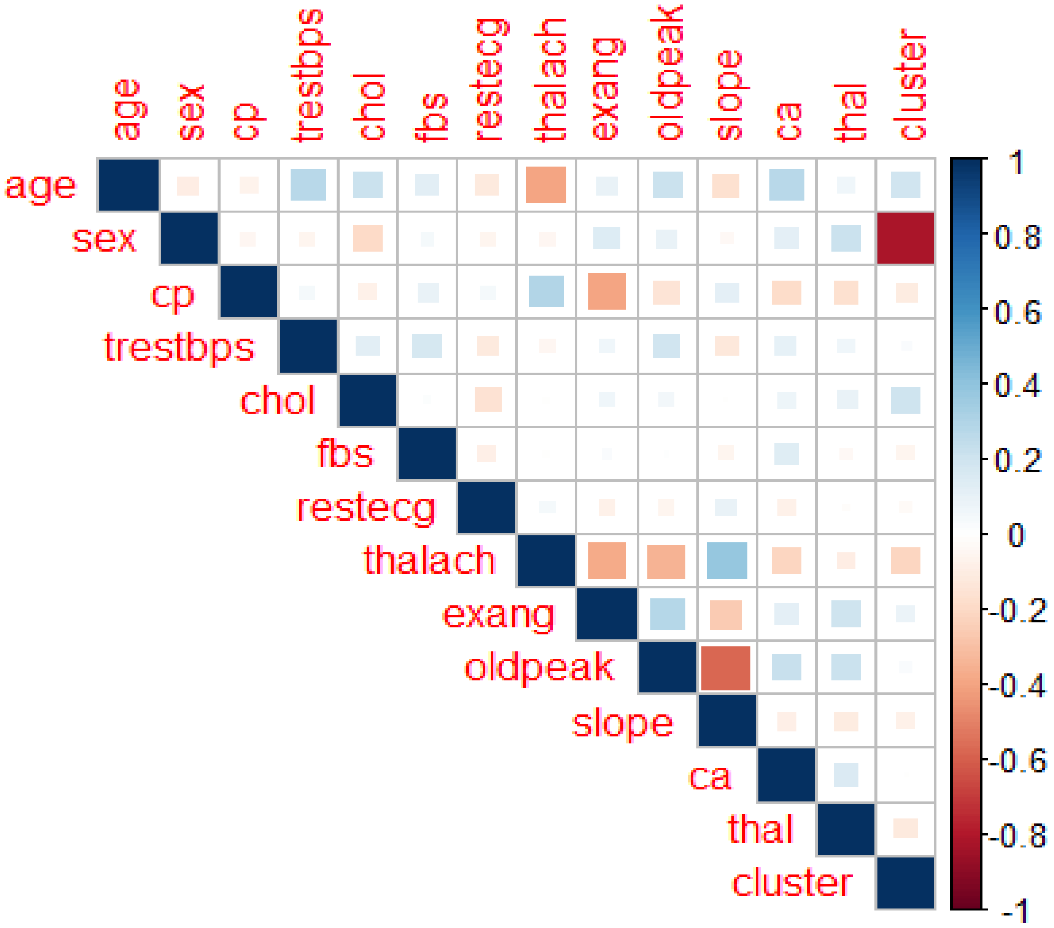

Density Plot for ca







Correlation matrix:



Exploratory Data Analysis:

We can implement Exploratory Data Analysis tools and techniques to investigate, analyse, and summarize the main characteristics of datasets, often utilizing data visualization methodologies. EDA techniques allow for effective manipulation of data, to find the answers by discovering data patterns, spotting anomalies, checking assumptions, or testing a hypothesis. It can help detect obvious errors, identify outliers in datasets, understand relationships, unearth important factors, find patterns within data, and provide new insights.

The exploratory data analysis steps that analysts have in mind when performing EDA include:

- Asking the right questions related to the purpose of data analysis like what are the key risk factors for heart disease, how age affects heart disease, etc.

- Obtaining in-depth knowledge about problem domains
- Setting clear objectives that are aligned with the desired outcomes.

There are four exploratory data analysis techniques that data experts use, which include:

Graphical

Univariate

Multivariate

Non-Graphical

Univariate Non-Graphical:

This is the simplest type of EDA, where data has a single variable like age, chol, fbs, etc. Since there is only one variable, we do not have to deal with relationships.

Univariate Graphical:

Non-graphical techniques do not present the complete picture of data. Graphical methods are therefore required. Common types of univariate graphics include Stem-and-leaf plots, which show all data values and the shape of the distribution. Histograms, a bar plot in which each bar represents the frequency (count) or proportion (count/total count) of cases for a range of values. Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum. Also, to detect outliers in the dataset.

Multivariate Non-Graphical:

Multivariate data consists of several variables. Non-graphic multivariate EDA methods illustrate relationships between two or more data variables using statistics or cross-tabulation.

Multivariate Graphical:

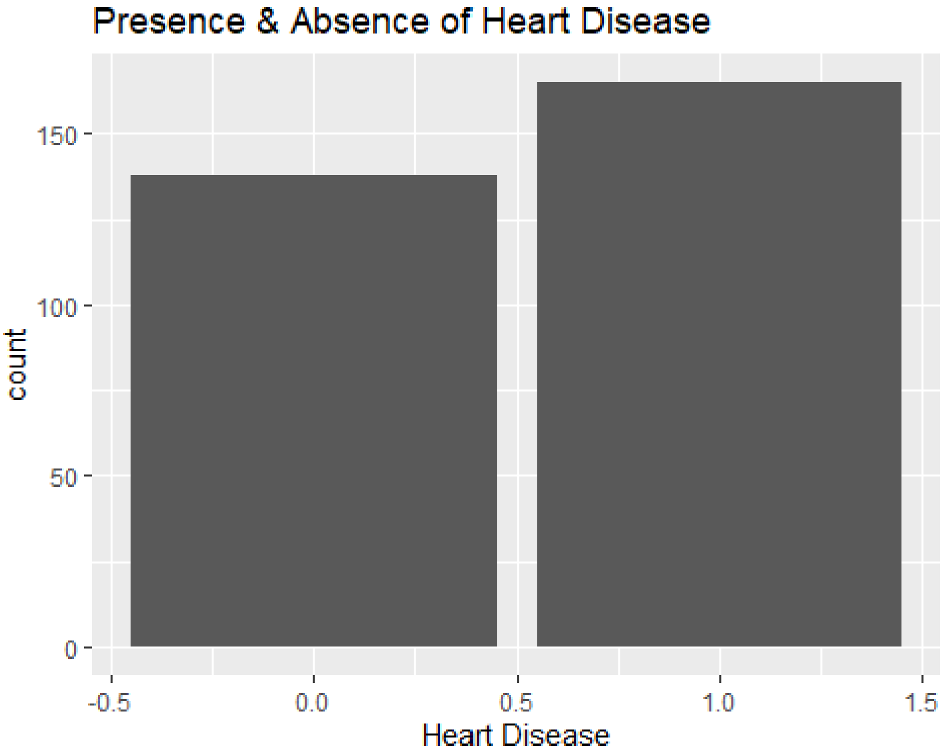
This EDA technique makes use of graphics to show relationships between 2 or more datasets like age vs trestbps, age vs chol, etc. Other common types of multivariate graphics include:

1. Scatter plot, which is used to plot data points on a horizontal and a vertical axis to show how much one variable is affected by another.
2. Multivariate chart, which is a graphical representation of the relationships between factors and a response.
3. Run chart, which is a line graph of data plotted over time.
4. Bubble chart, which is a data visualization that displays multiple circles (bubbles) in a two- dimensional plot.
5. Heat map, which is a graphical representation of data where values are depicted by colour.
6. Different ways to look at the data Different ways to look at the heart dataset include clustering and dimension reduction techniques, which help create graphical displays of high-dimensional data containing many variables.

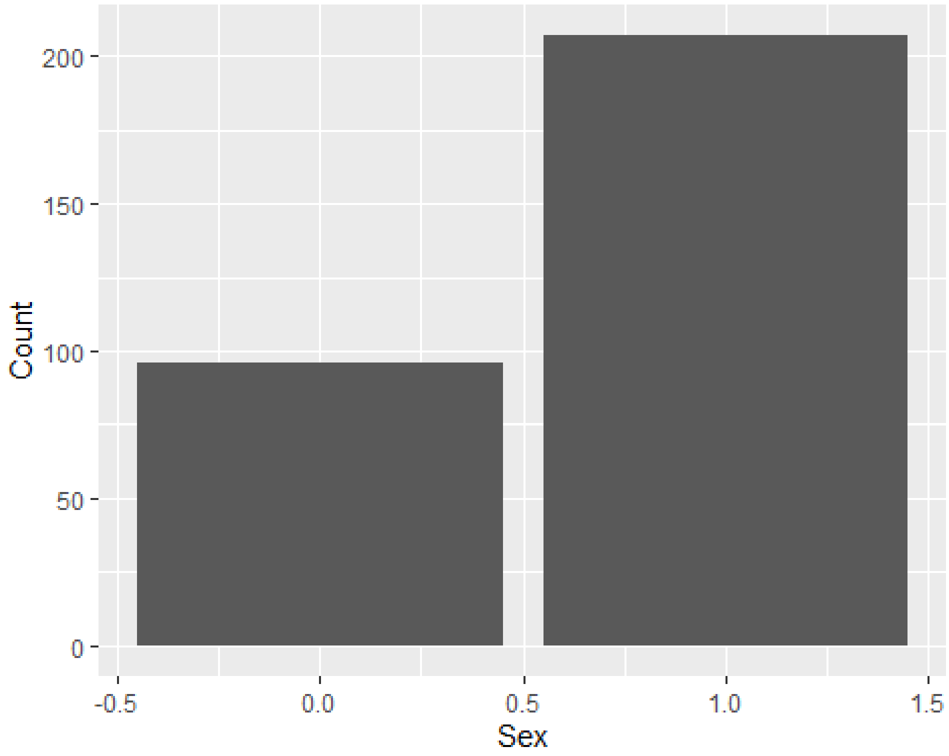
Benefit of Analysis (data analysis):

Data analysis is of utmost importance when it comes to heart disease prediction. By analyzing heart disease data, we can uncover hidden patterns, correlations, and relationships that may not be evident by simply looking at the raw data. This can lead to better understanding of the disease and its risk factors. Additionally, data analysis can help in the identification of new risk factors that may not have been previously identified. Furthermore, data analysis can be used to develop predictive models that can help in the early detection and prevention of heart disease. These models can identify individuals who are at higher risk of developing the disease and recommend preventive measures. Data analysis can also aid in the evaluation of treatment efficacy. By analysing patient data, we can determine which treatments are most effective for different subpopulations of patients. This information can be used to improve treatment protocols and ultimately improve patient outcomes. In summary, data analysis plays a crucial role in heart disease prediction and management. It can lead to the discovery of new risk factors, aid in the development of predictive models, and improve treatment efficacy.

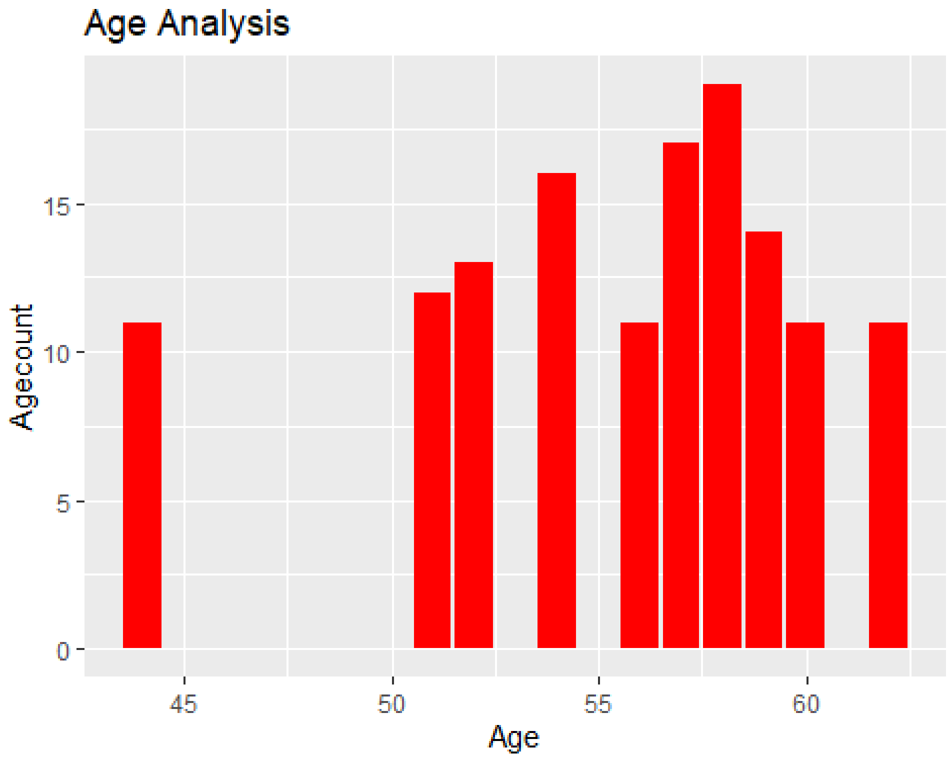
Presence and Absence of Heart Disease:



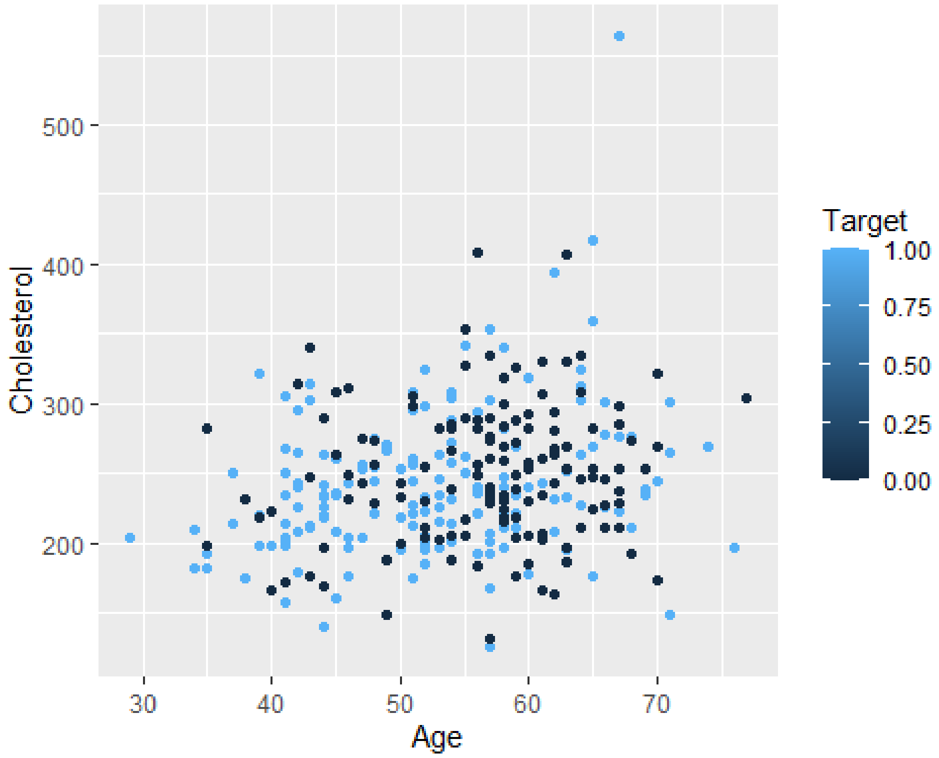
Sex Variable:



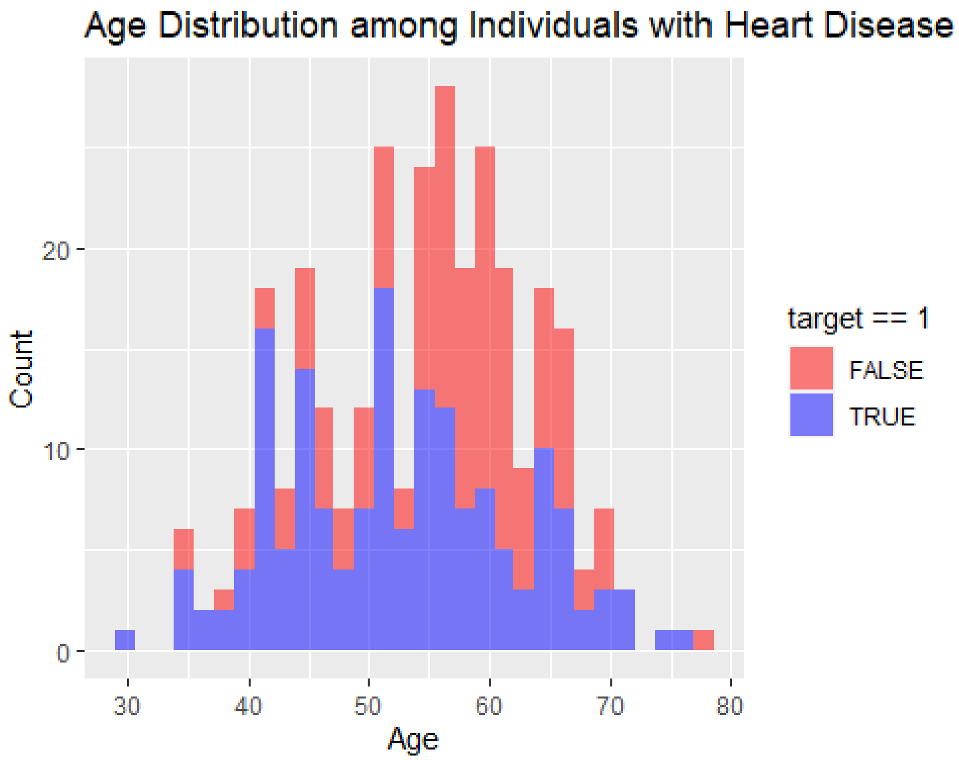
Age analysis:



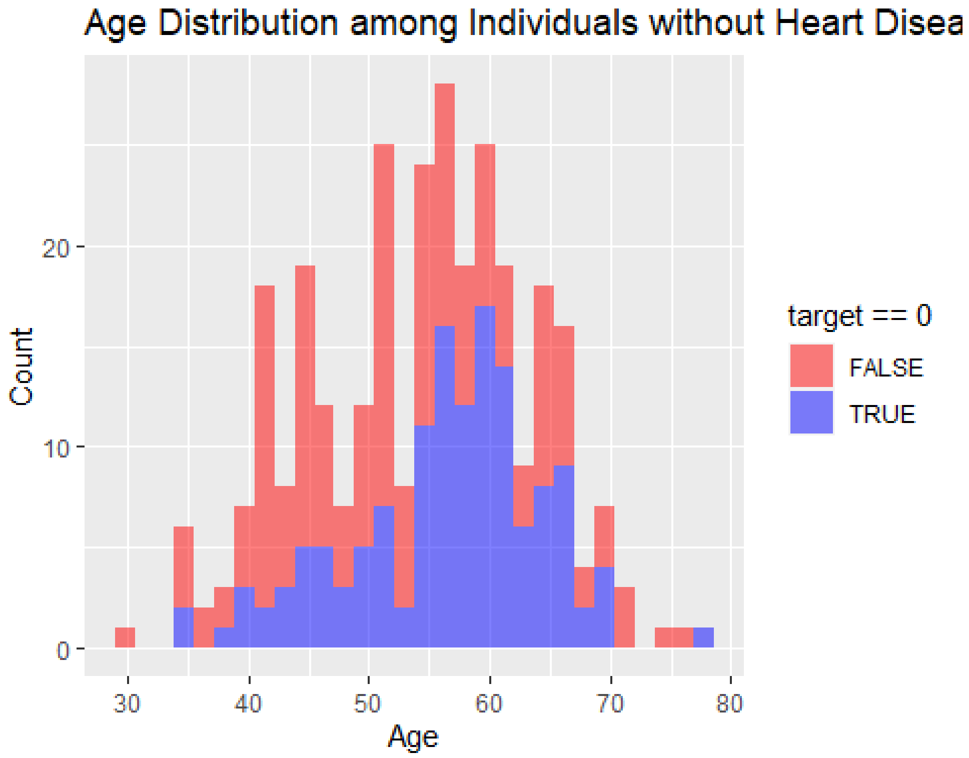
Scatter plot of Age and Cholesterol:



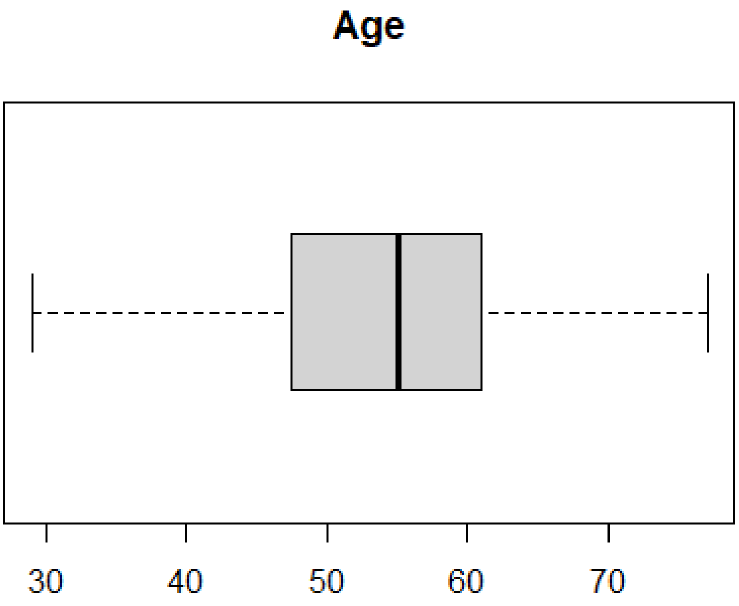
Age distribution among individuals with Heart Disease:



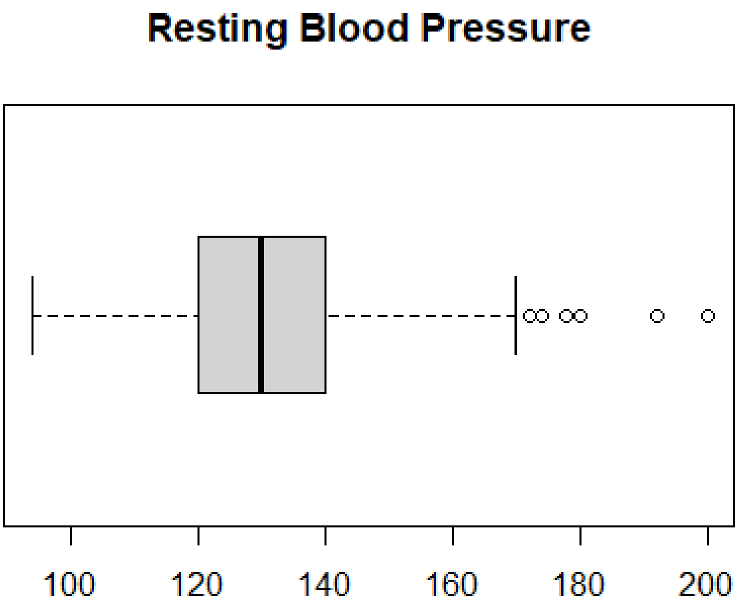
Age distribution among individuals with Heart Disease:



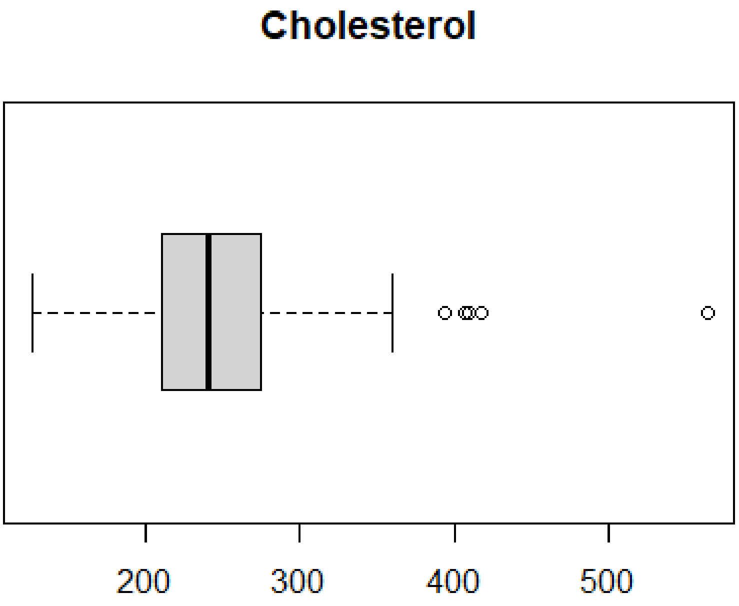
Boxplot of Age:



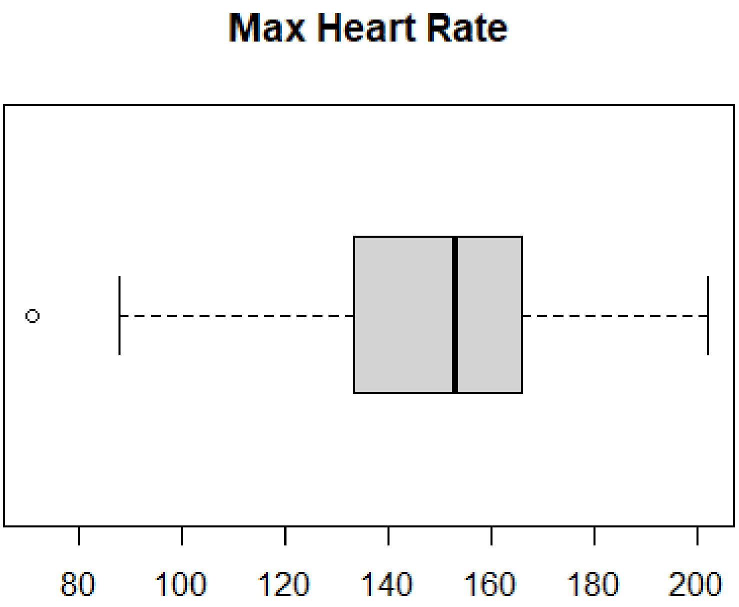
Boxplot of Resting blood pressure:



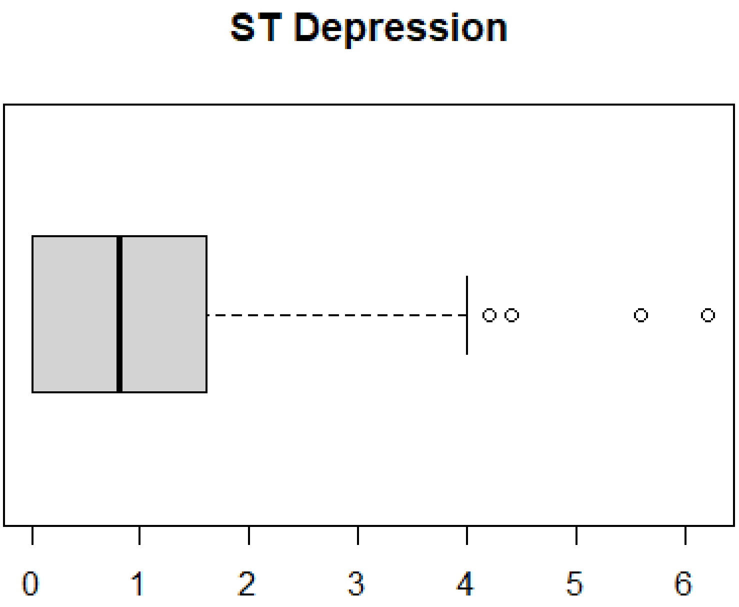
Boxplot of Cholesterol:



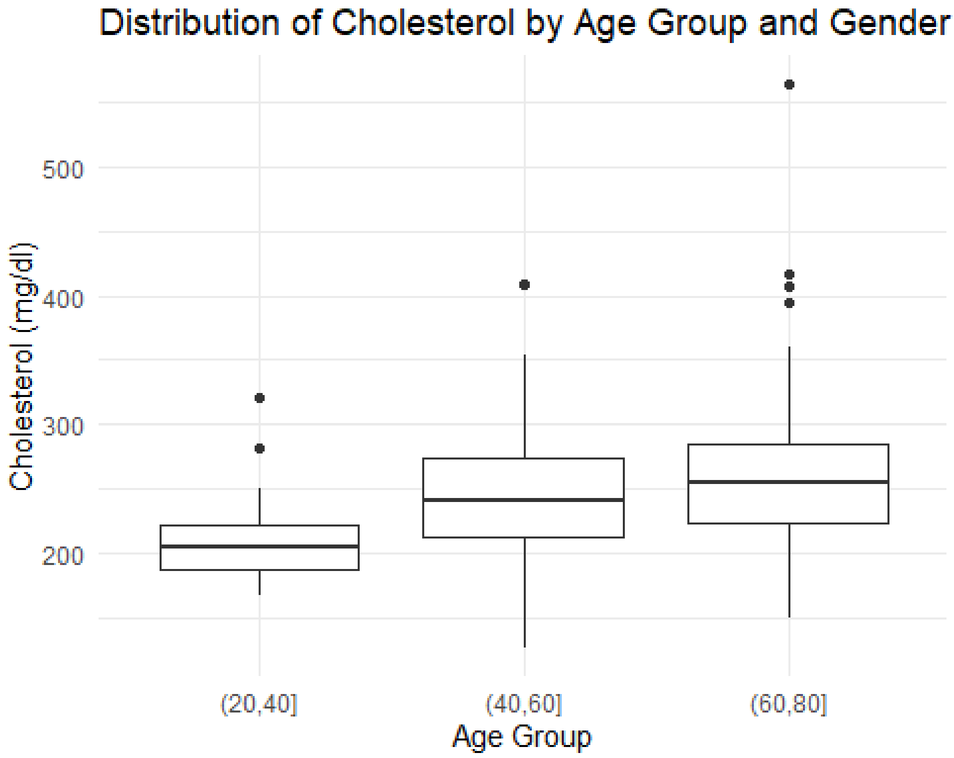
Boxplot of Max heart rate:



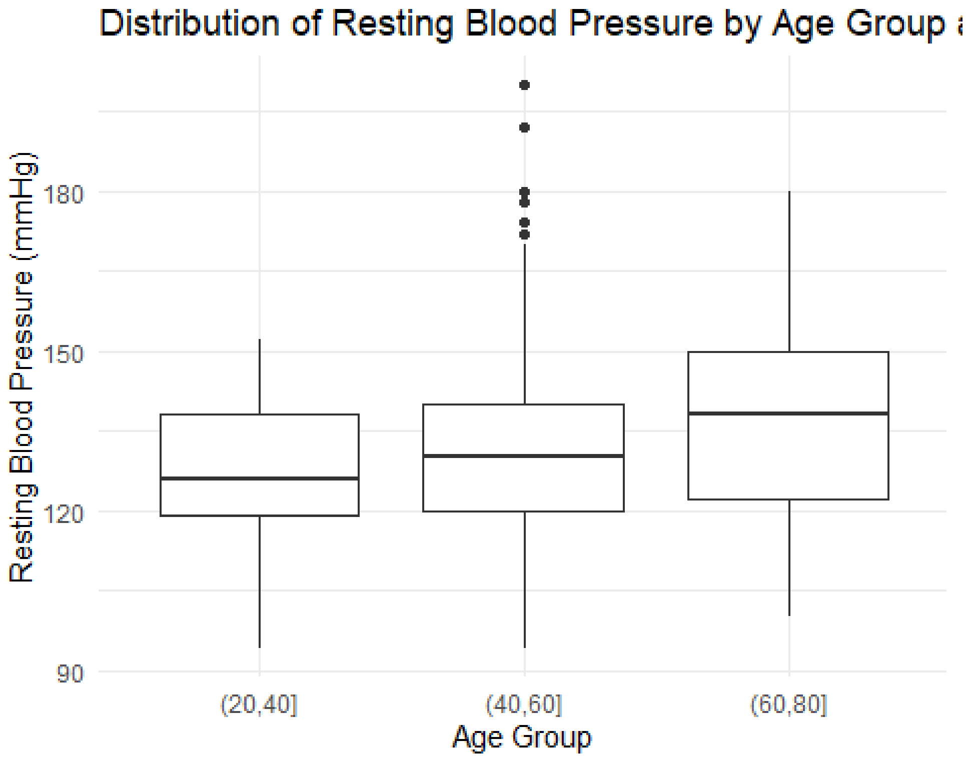
Boxplot of ST Depression:



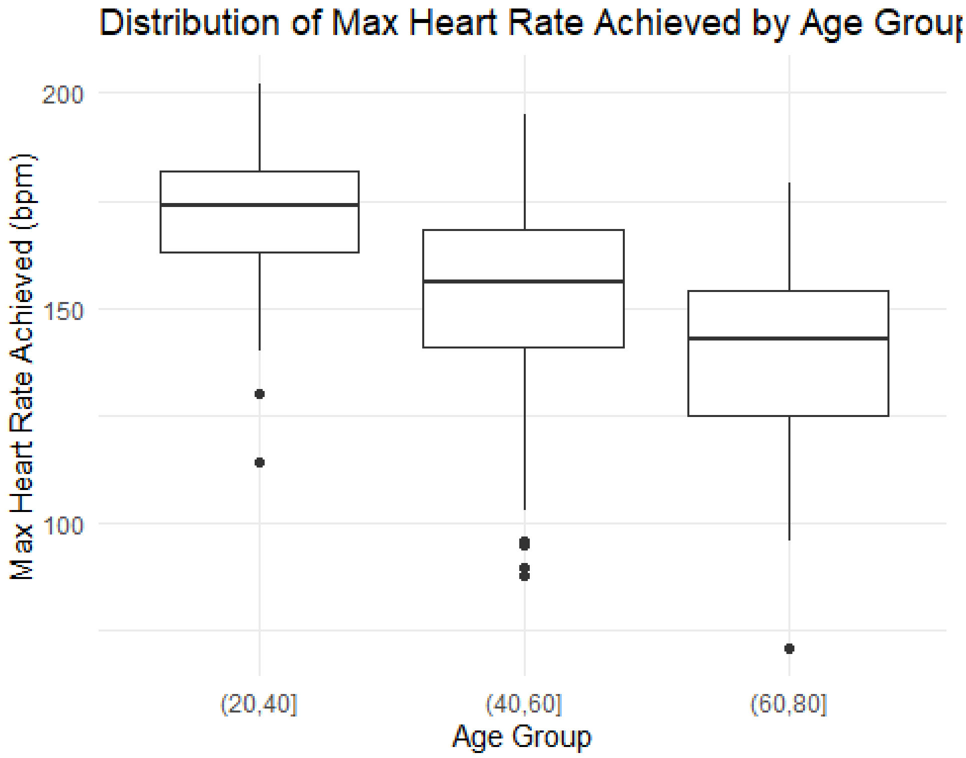
Distribution of Cholesterol by Age group and Gender:



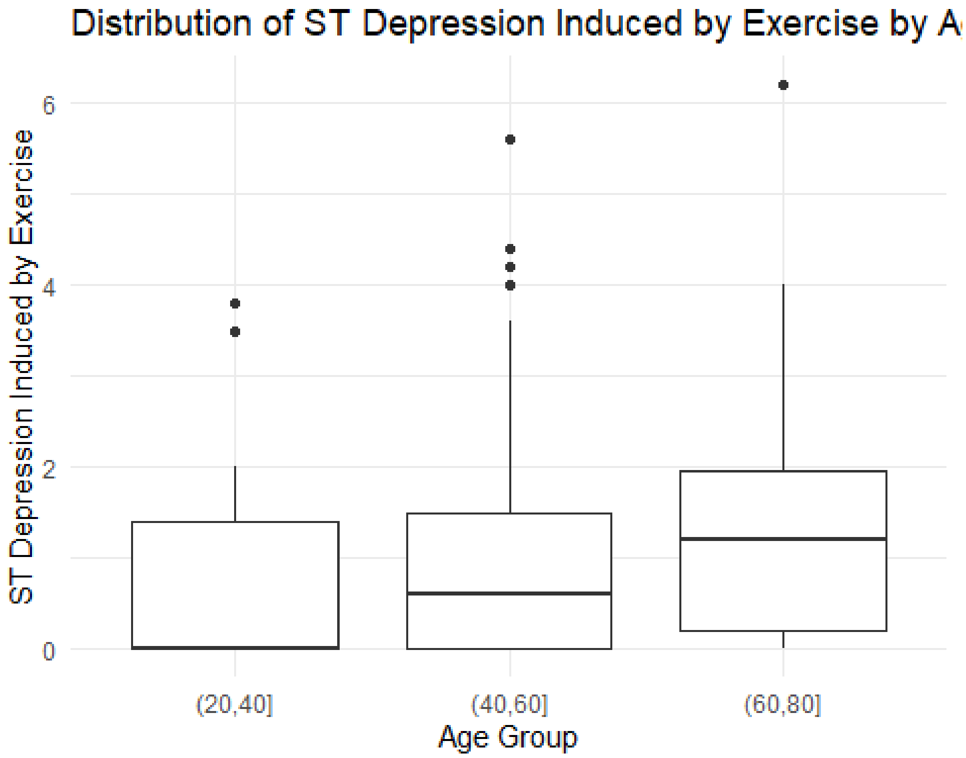
Distribution of Resting blood pressure and Age group:



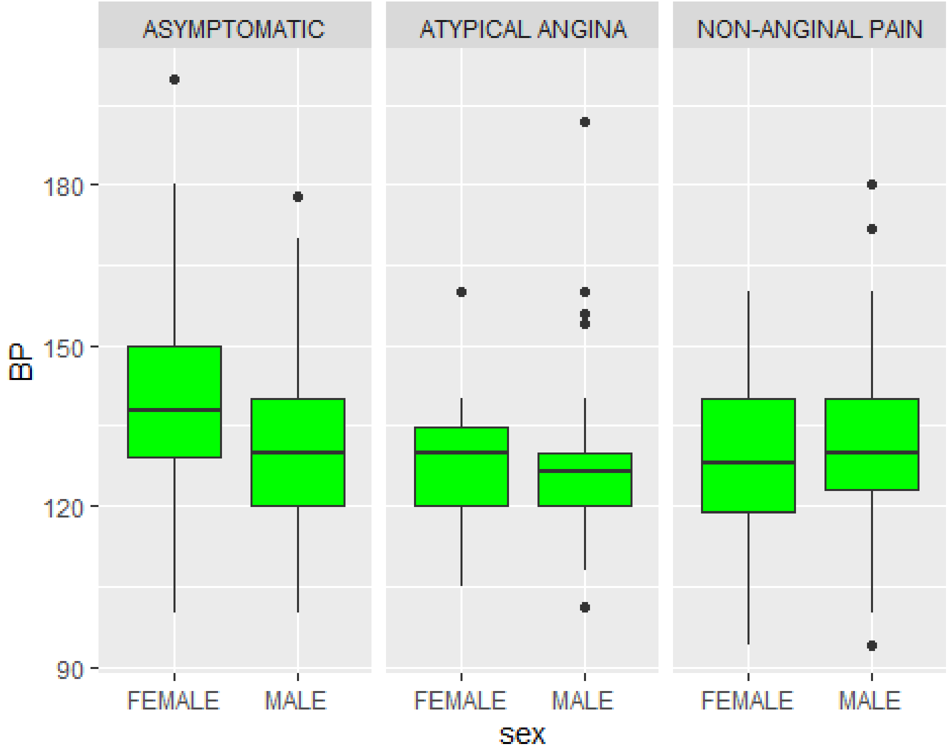
Distribution of Max heart rate achieved by Age group:



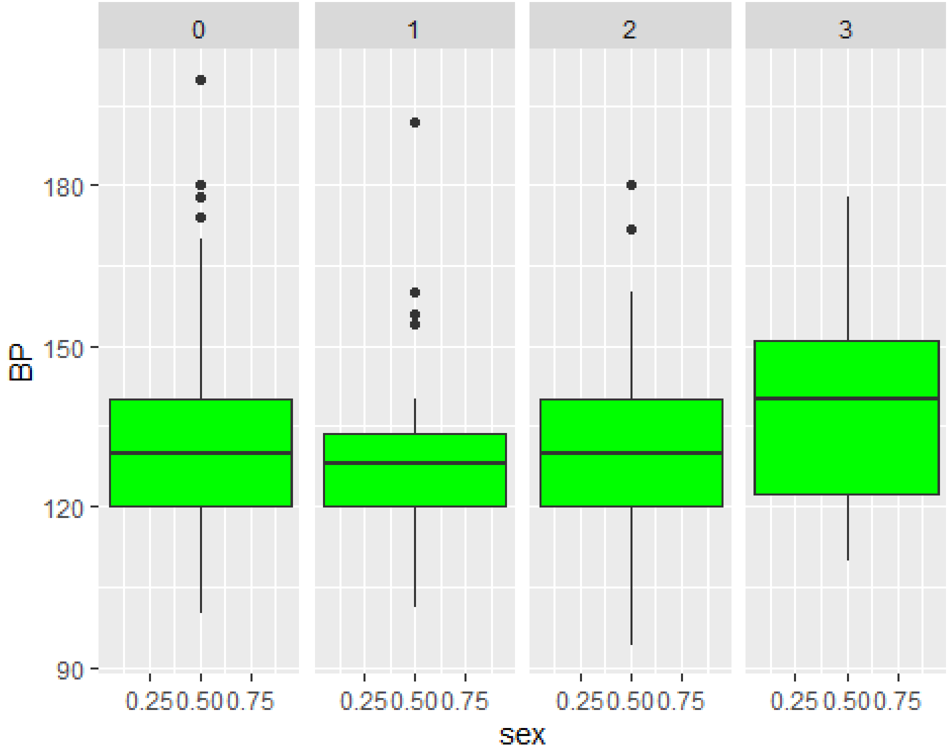
Distribution of ST depression induced by exercise by Age group:



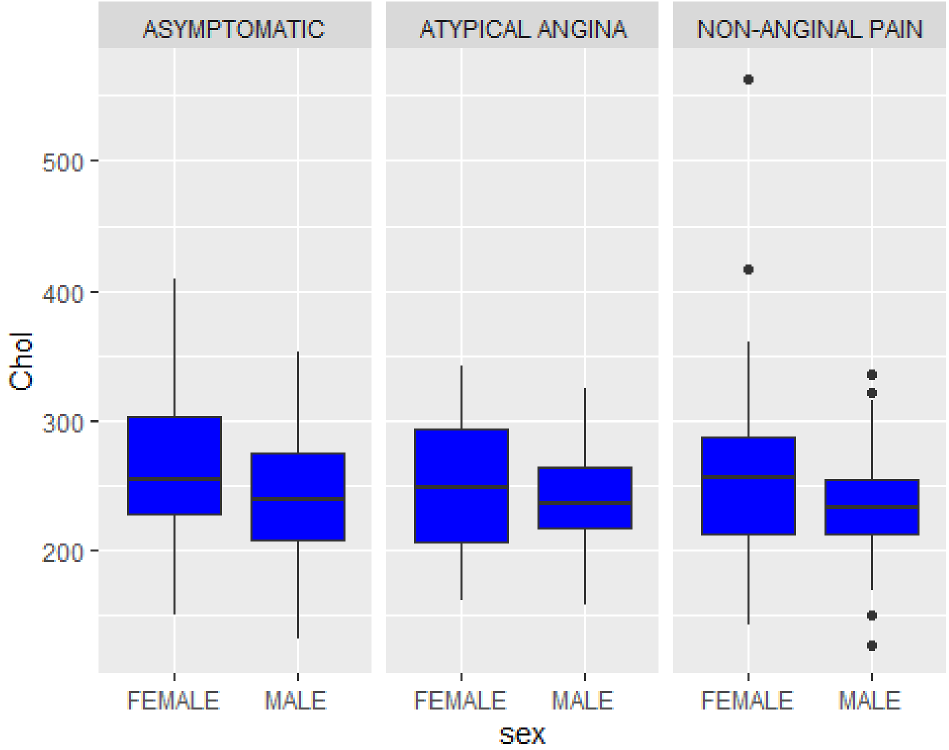
Comparing Blood pressure across Chest pain:



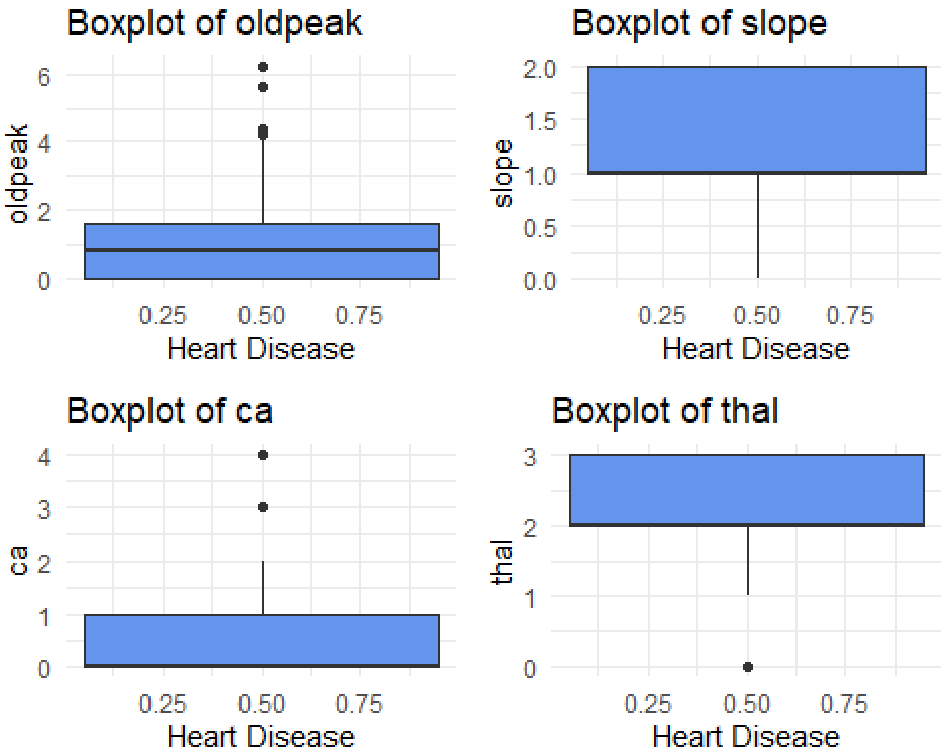
Comparing Blood pressure across Sex:



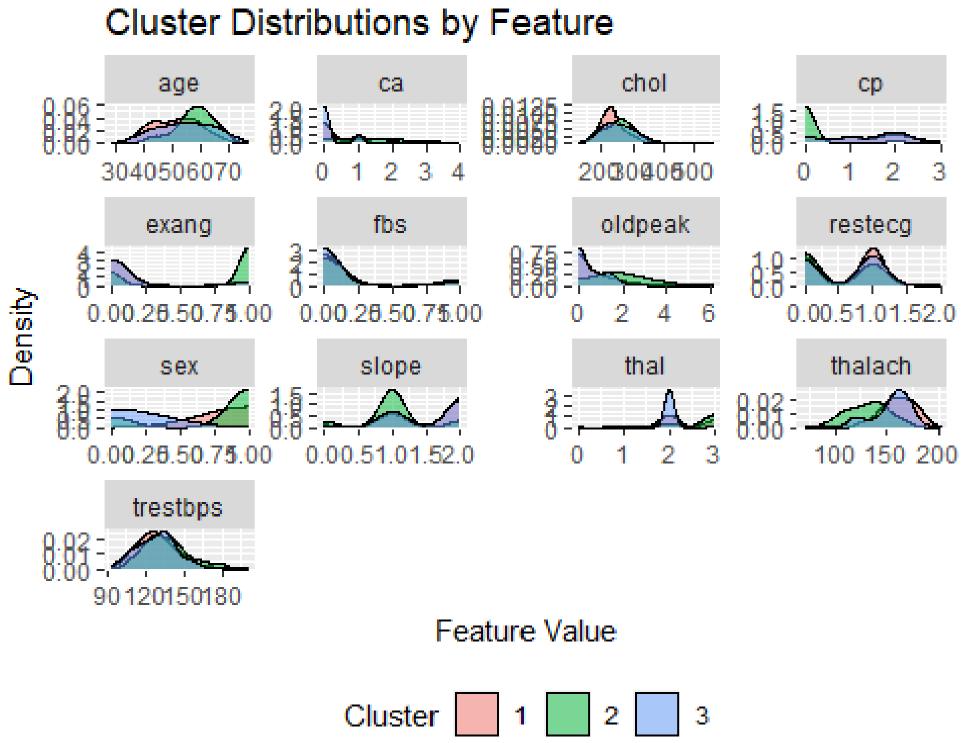
Comparing Cholesterol across Sex:



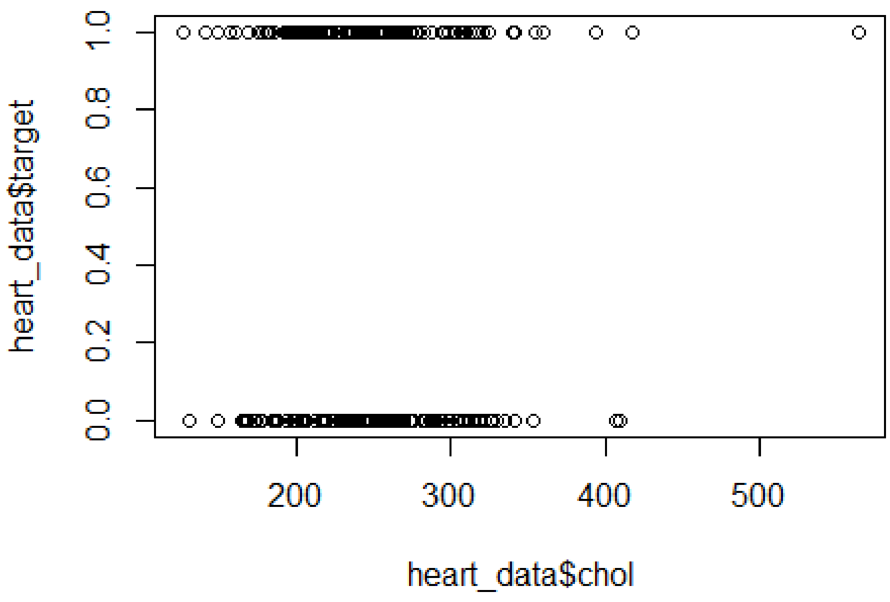
Boxplot to plot lifestyle factors:



Cluster distributions by features:



Relationship between Maximum heart rate achieved and Heart disease status:



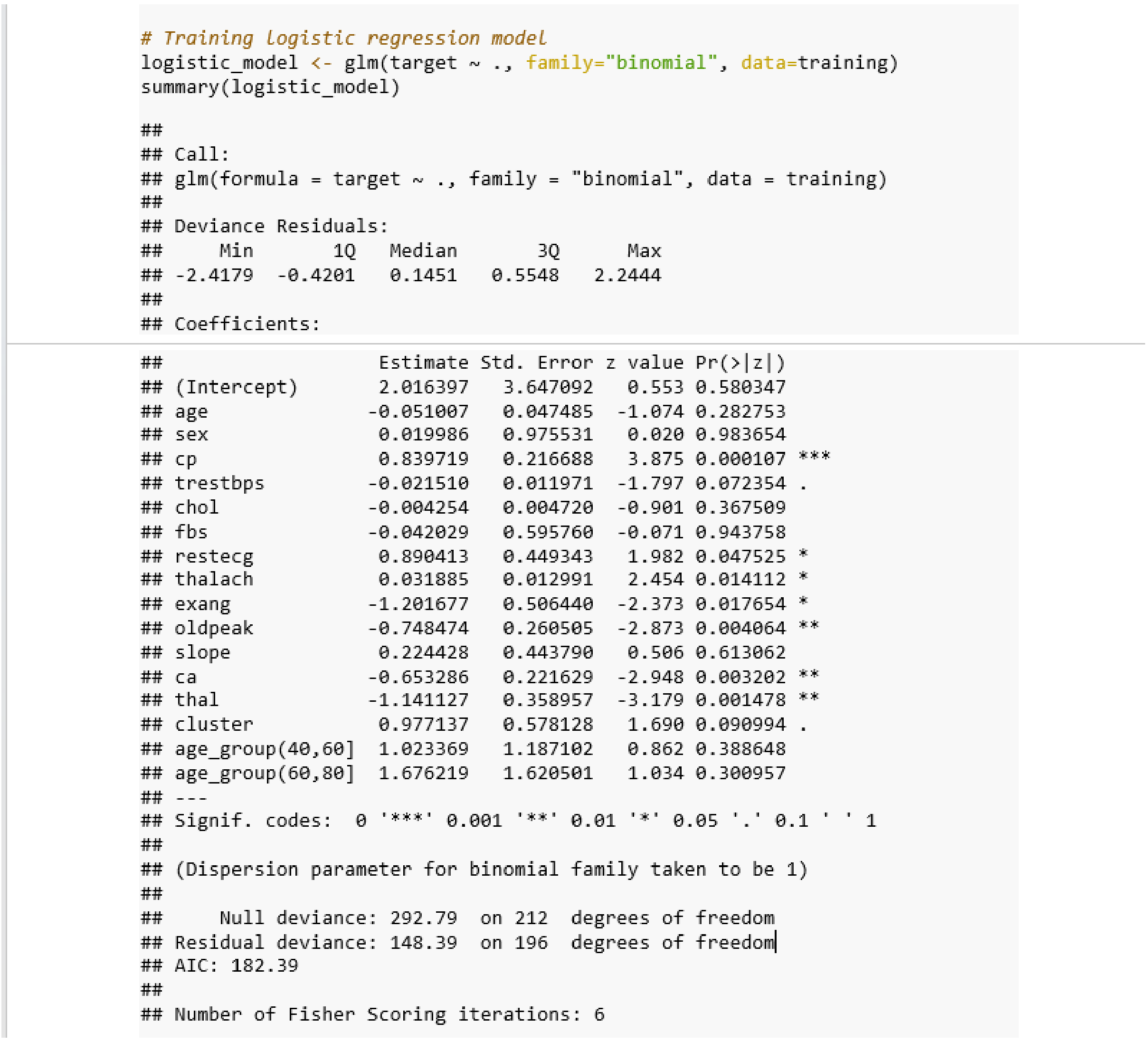
Model Building:

Model building in heart disease prediction involves using statistical and machine learning techniques to develop a model that can accurately predict the presence or absence of heart disease based on a set of input variables or features. For this part, we tried to predict whether a person is suffering from heart disease or not. We split the dataset into training and testing sets, where the training set is used to train the model and the testing set is used to evaluate its performance. There are various algorithms and techniques that can be used to build a heart disease prediction model, such as logistic regression, decision trees, random forests, support vector machines, and neural networks. The choice of algorithm depends on the characteristics of the data, the complexity of the model, and the accuracy required. Once a model is trained, we try to predict whether a person has heart disease or not and test the predictions using testing data.

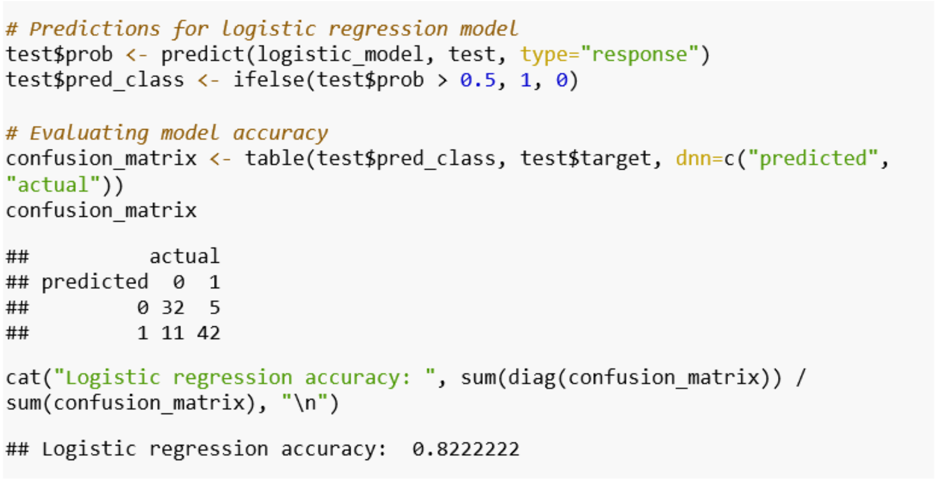
Logistic Regression:

We build the model from the dataset to predict the presence of heart disease. We split 70% of dataset into training data and 30% of dataset into testing data. We got an accuracy of 82% with testing data.

Prediction and confusion matrix:



```
# Training logistic regression model
logistic_model <- glm(target ~ ., family="binomial", data=training)
summary(logistic_model)

##
## Call:
## glm(formula = target ~ ., family = "binomial", data = training)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4179  -0.4201   0.1451   0.5548   2.2444
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        2.016397   3.647092   0.553 0.580347
## age               -0.051007   0.047485  -1.074 0.282753
## sex                0.019986   0.975531   0.020 0.983654
## cp                 0.839719   0.216688   3.875 0.000107 ***
## trestbps          -0.021510   0.011971  -1.797 0.072354 .
## chol              -0.004254   0.004720  -0.901 0.367509
## fbs               -0.042029   0.595760  -0.071 0.943758
## restecg            0.890413   0.449343   1.982 0.047525 *
## thalach            0.031885   0.012991   2.454 0.014112 *
## exang             -1.201677   0.506440  -2.373 0.017654 *
## oldpeak           -0.748474   0.260505  -2.873 0.004064 **
## slope              0.224428   0.443790   0.506 0.613062
## ca                -0.653286   0.221629  -2.948 0.003202 **
## thal              -1.141127   0.358957  -3.179 0.001478 **
## cluster            0.977137   0.578128   1.690 0.090994 .
## age_group(40,60]   1.023369   1.187102   0.862 0.388648
## age_group(60,80]   1.676219   1.620501   1.034 0.300957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 292.79  on 212  degrees of freedom
## Residual deviance: 148.39  on 196  degrees of freedom
## AIC: 182.39
##
## Number of Fisher Scoring iterations: 6
```

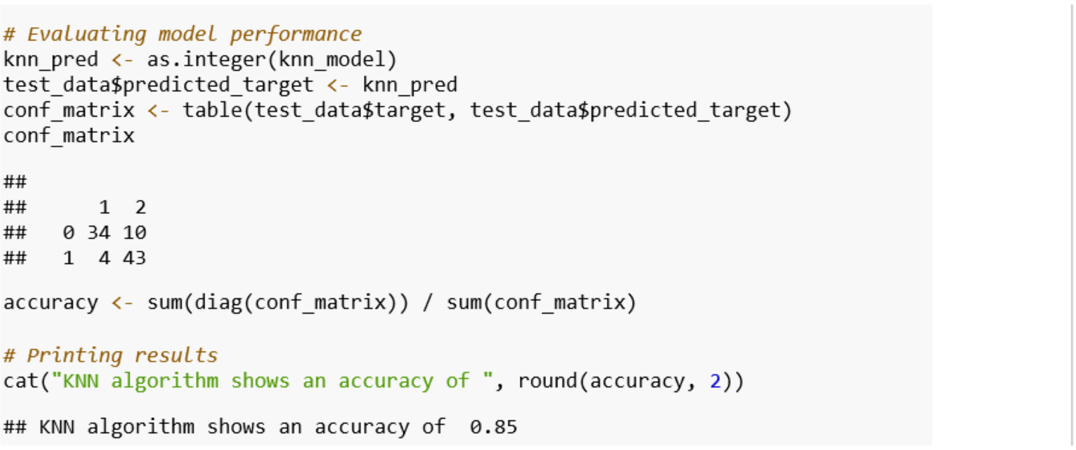


```
# Predictions for logistic regression model
test$prob <- predict(logistic_model, test, type="response")
test$pred_class <- ifelse(test$prob > 0.5, 1, 0)

# Evaluating model accuracy
confusion_matrix <- table(test$pred_class, test$target, dnn=c("predicted",
"actual"))
confusion_matrix

##            actual
## predicted  0  1
##         0 32  5
##         1 11 42

cat("Logistic regression accuracy: ", sum(diag(confusion_matrix)) /
sum(confusion_matrix), "\n")

## Logistic regression accuracy:  0.8222222
```

K – Nearest Neighbors(KNN):

We build the model from the dataset to predict the presence of heart disease. We split 70% of dataset into training data and 30% of dataset into testing data. We got an accuracy of 85% with testing data.

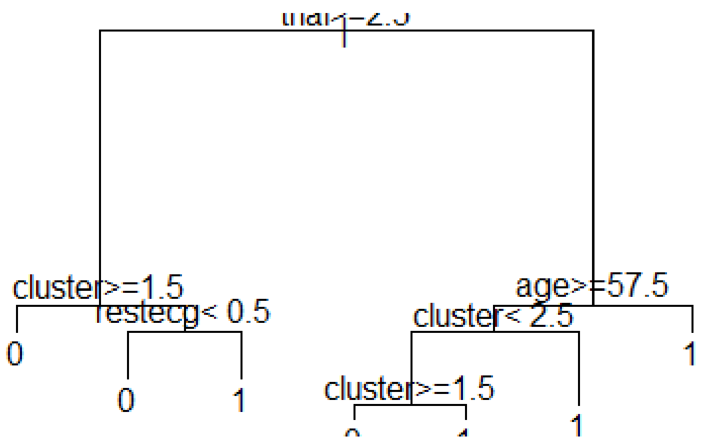
Prediction and confusion matrix:



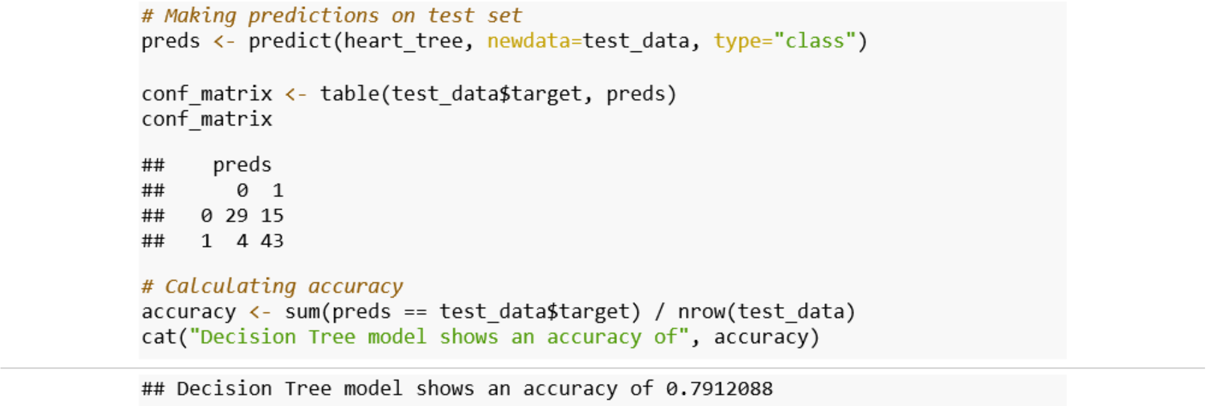
```
# Evaluating model performance
knn_pred <- as.integer(knn_model)
test_data$predicted_target <- knn_pred
conf_matrix <- table(test_data$target, test_data$predicted_target)
conf_matrix

##
##      1  2
##   0 34 10
##   1  4 43

accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)

# Printing results
cat("KNN algorithm shows an accuracy of ", round(accuracy, 2))

## KNN algorithm shows an accuracy of  0.85
```

Decision Tree:

We build the model from the dataset to predict the presence of heart disease. We split 70% of dataset into training data and 30% of dataset into testing data. We got an accuracy of 79% with testing data.



Prediction and confusion matrix:



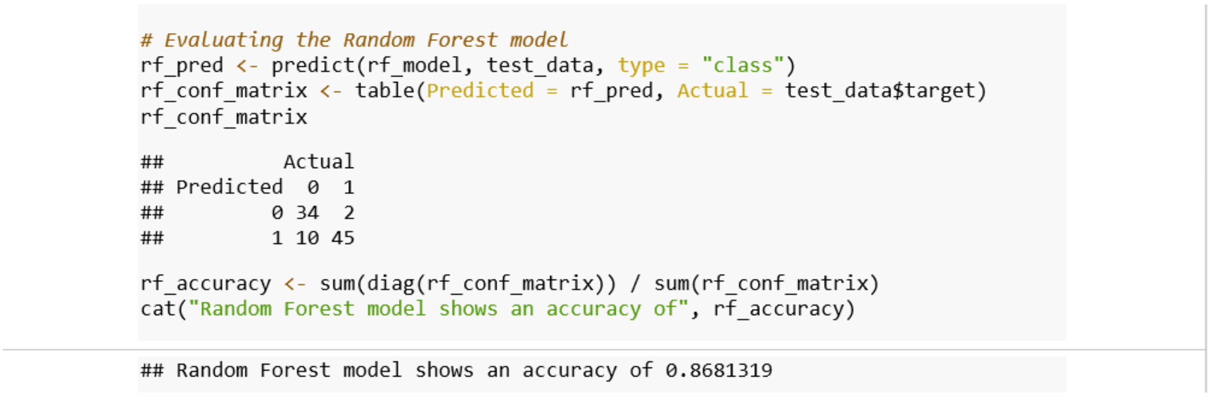
```
# Making predictions on test set
preds <- predict(heart_tree, newdata=test_data, type="class")

conf_matrix <- table(test_data$target, preds)
conf_matrix

##    preds
##      0  1
##   0 29 15
##   1  4 43

# Calculating accuracy
accuracy <- sum(preds == test_data$target) / nrow(test_data)
cat("Decision Tree model shows an accuracy of", accuracy)
```

```
## Decision Tree model shows an accuracy of 0.7912088
```

Random Forest:

We build the model from the dataset to predict the presence of heart disease. We split 70% of dataset into training data and 30% of dataset into testing data. We got an accuracy of 86% with testing data.

Prediction and confusion matrix:



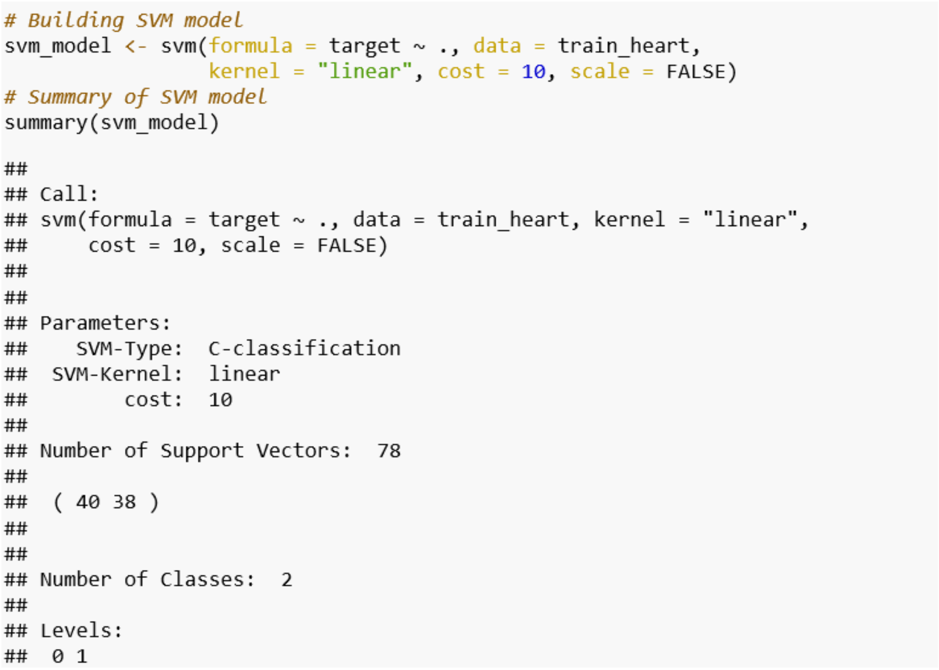
```
# Evaluating the Random Forest model
rf_pred <- predict(rf_model, test_data, type = "class")
rf_conf_matrix <- table(Predicted = rf_pred, Actual = test_data$target)
rf_conf_matrix

##          Actual
## Predicted  0  1
##         0 34  2
##         1 10 45

rf_accuracy <- sum(diag(rf_conf_matrix)) / sum(rf_conf_matrix)
cat("Random Forest model shows an accuracy of", rf_accuracy)
```

```
## Random Forest model shows an accuracy of 0.8681319
```

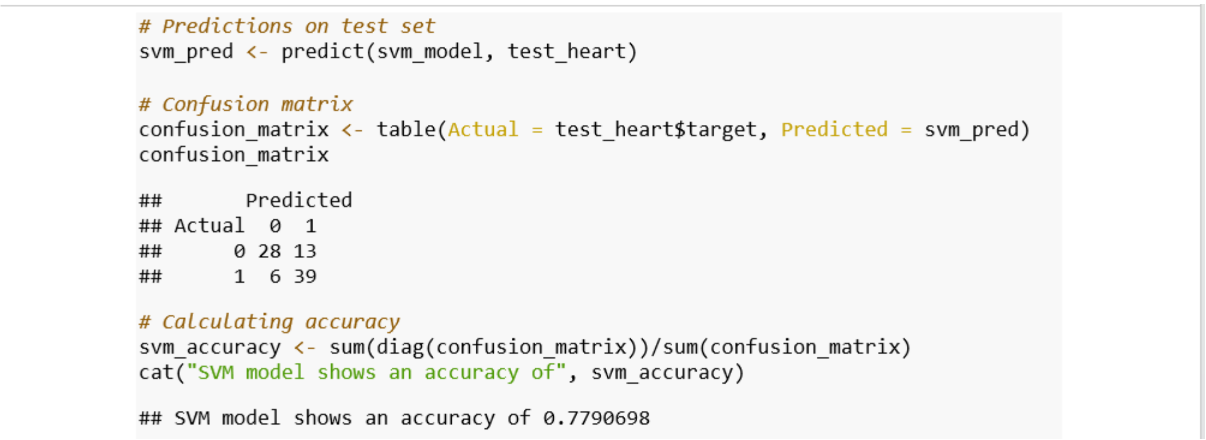
Support Vector Machine:

We build the model from the dataset to predict the presence of heart disease. We split 70% of dataset into training data and 30% of dataset into testing data. We got an accuracy of 77% with testing data.

Prediction and confusion matrix:



```
# Building SVM model
svm_model <- svm(formula = target ~ ., data = train_heart,
                 kernel = "linear", cost = 10, scale = FALSE)
# Summary of SVM model
summary(svm_model)

##
## Call:
## svm(formula = target ~ ., data = train_heart, kernel = "linear",
##     cost = 10, scale = FALSE)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  10
##
## Number of Support Vectors:  78
##
##  ( 40 38 )
##
##
## Number of Classes:  2
##
## Levels:
##  0 1
```



```
# Predictions on test set
svm_pred <- predict(svm_model, test_heart)

# Confusion matrix
confusion_matrix <- table(Actual = test_heart$target, Predicted = svm_pred)
confusion_matrix

##       Predicted
## Actual  0  1
##      0 28 13
##      1  6 39

# Calculating accuracy
svm_accuracy <- sum(diag(confusion_matrix))/sum(confusion_matrix)
cat("SVM model shows an accuracy of", svm_accuracy)

## SVM model shows an accuracy of 0.7790698
```

Results:

We divided the dataset such that 70% of data is used for training the models and 30% of data is used for testing the model predictions. We used five algorithms in our project to make predictions. They are Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest and Support Vector Machine.

- In Logistic Regression model, we achieved an accuracy of 82%.
- In K-Nearest Neighbors (KNN) model, we achieved an accuracy of 85%.
- In Decision Tree model, we achieved an accuracy of 79%.
- In Random Forest model, we achieved an accuracy of 86%.
- In support Vector Machine model, we achieved an accuracy of 77%

Out of all the models, Random Forest model achieved highest accuracy and Support Vector Machine model achieved least accuracy,

Conclusions

- Heart disease Predictions helps us in identifying the heart disease in the early stages so that required steps can be taken to prevent the death of patients.
- We were able to successfully explore lifestyle factors that are strongly associated with heart disease.
- The prediction model are able to predict the heart disease in the patients with high accuracy which helps the patients in taking measure to prevent its occurrence.

Limitations

- The dataset is relatively small which may not be representative of the entire population.
- The dataset does not contain information on other important factors such as genetic predisposition to heart disease, lifestyle factors such as diet, and other medical conditions.
- Since the dataset was relatively small, we can incorporate more advanced techniques if huge datasets are used for making predictions.

References:

1. Cheng, J., Liu, W., Sun, X., Wang, Y., & Ye, J. (2018). Risk prediction of cardiovascular disease based on machine learning algorithms. Journal of Healthcare Engineering, 2018, 1-7
2. Ahmed, A. S., Mohamed, M. A., & Ali, M. H. (2019). Predicting heart disease using machine learning algorithms. International Journal of Advanced Computer Science and Applications, 10(7), 344-349.
3. Mahmood, T., Shahzad, A., & Sultan, S. (2018). Predictive modeling for heart disease using machine learning techniques. Journal of Intelligent Learning Systems and Applications, 10(04), 1-11.
4. Deshpande, S., & Patil, A. (2020). Analysis of heart disease prediction using machine learning algorithms. Journal of Big Data, 7(1), 1-17.
5. Vijayarani, S., & Latha, P. (2021). Analysis and prediction of heart disease using machine learning algorithms. International Journal of Electrical and Computer Engineering, 11(3), 2393-2402.
6. Sabaté-Llobera, A., Prieto-Blanco, A., & Terrades-García, N. (2020). Prediction of heart disease using machine learning: A systematic review. International Journal of Environmental Research and Public Health, 17(18), 1-19.
7. Chen, Y., Wang, H., & Deng, Y. (2019). A machine learning approach for heart disease diagnosis. Journal of Medical Systems, 43(12), 1-10.
8. Ramesh, S., Palaniappan, S., & Rajendran, P. (2018). Heart disease prediction using machine learning algorithms. International Journal of Engineering and Technology(UAE), 7(2.34), 51-56.

9. N. T. Uzun, S. Ozturk, and H. A. Uzun, "Prediction of heart disease risk using machine learning algorithms," in Computer Methods and Programs in Biomedicine, vol. 171, pp. 35-45, Nov. 2019, doi: 10.1016/j.cmpb.2019.01.018.
10. Rattani, Ambarish, et al. "Predicting heart disease risk using machine learning and feature selection techniques: A review." Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.