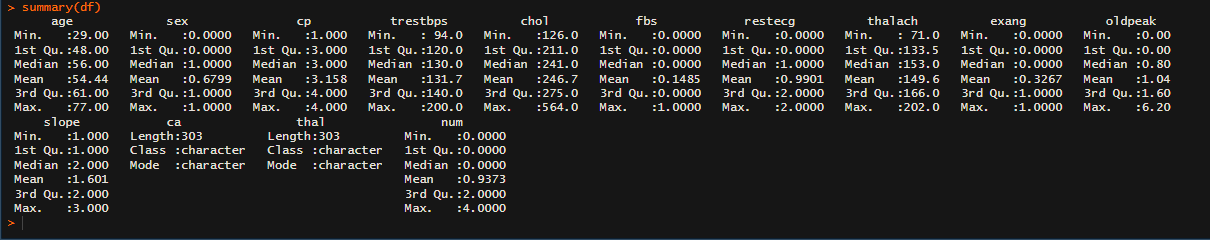
1. **Data Exploration**

Text

Description automatically generated

*Removed all ‘?’ values and NULL values if there are any in the given ‘Cleveland’ dataset.*

1. **Data Analysis**



Chart, histogram

Description automatically generatedChart, box and whisker chart

Description automatically generated

Chart, scatter chart

Description automatically generated

1. **Data Modeling**

Yes, heart disease can be predicted using logistic regression. A prominent statistical technique for modeling binary outcomes, such as the presence or absence of a disease, is logistic regression. In this technique, the numbers 1 - 4 indicate the existence of heart disease at various levels, while 0 indicates the lack of heart disease.

You would require a dataset including predictors (or independent variables) like age, sex, blood pressure, cholesterol levels, and other pertinent parameters that are known to relate to heart disease to perform logistic regression for heart disease prediction.

One of the various predictive modeling methods that can be used to forecast cardiac disease is logistic regression. Depending on the type and complexity of the data, other techniques including decision trees, random forests, and neural networks may also be useful.

Furthermore, since the training data used to train logistic regression models is only as good as the data itself, it is critical to make sure that the dataset is representative and free of biases to produce reliable results.

1. **Report & Summary**

In this assignment I did the part, when a user enters all the patient details (age, sex, cp, trestbps, chol, fbs, restecg, thalch, exang, oldpeak, slope, ca, thal), then it will predict whether the patient has a defective heart with defective levels (1/2/3/4) or a healthy heart.

First, I downloaded the “processed.cleveland.data” dataset from the given site and performed the data exploration part by importing the dataset into R Studio. Then I checked whether there are any NULL values or ‘?’ value fields in the dataset and replaced them with NA.

After that step I did the data analysis part by performing some analyses such as summary(), histogram, boxplot, and scatterplot. In the .ipynb file, I have also included a summary test, heatmap, histogram, boxplot, and a plot. After getting a proper understanding by the analysis of data started doing the prediction model.

1. **GitHub Repository Link**

<https://github.com/KavinduKasthurirathne/Heart-Disease-Prediction-Model.git>