1. **Introduction**

Heart disease refers to several types of heart conditions. Some risk factors of heart disease are blood pressure, cholesterol, diabetes, smoking exposure, and physical inactivity etc.

The analysis was done on the Cleveland Heart Disease dataset. The chosen heart data set consists of 302 rows of individuals and 14 columns after removing the duplicates. The columns present in the data set are explained below. Analysis is performed to show how the factors like age, blood pressure, gender and cholesterol are significant in heart related issues.

|  |  |  |
| --- | --- | --- |
| **Column** | **Definition** | **Values** |
| age | Age in Years |  |
| sex | Sex | 0-Female 1 -Male |
| cp | Chest pain type | 0-Asymptomatic  1-Atypical Angina  2 - Non Anginal pain  3 - Typical Angina |
| trestbps | Resting blood pressure (mmHg) |  |
| chol | Serum cholesterol (mg/dL) |  |
| fbs | Fasting blood sugar > 120 (mg/dL) | 0 - False (< 120) 1 - True (> 120) |
| restsecg | Resting electrocardiographic results | 0 - Normal 1 - Abnormality 2 - Hypertrophy |
| thalach | Maximum exercise heart rate achieved (bpm) |  |
| exang | Exercise-induced angina | 0 - No 1 - Yes |
| oldpeak | Old peak = ST depression induced by exercise relative to rest | 0 - 6.2 |
| slope | Slope of the peak exercise ST segment | 0 - Upsloping 1 - Flat 2 - Down sloping |
| ca | Number of major vessels (0–3) colored by fluoroscopy |  |
| thal | Thallium Heart Rate | 1 - Fixed Defect 2 - Normal  3 - Reversible defect |
| target | Chance of Heart Disease Prediction | 0- Absence 1 - Presence |

1. **Source**

The data analysis is done purely based on the data set

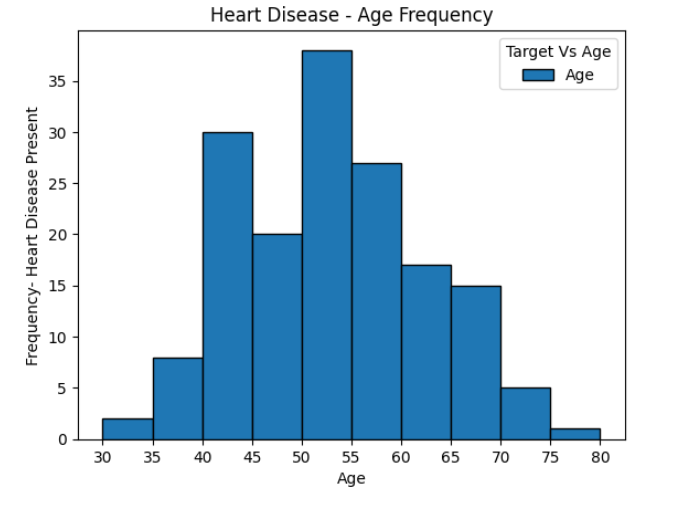
<https://raw.githubusercontent.com/hridyasg/Data_Analytics/main/Heart_Disease_Dataset.csv?classId=fa8be78a-1b27-4fb4-9128-5eddcd42f0aa&assignmentId=76d310f7-aa14-4426-84fc-ed565fa70432&submissionId=302c7c5f-7615-99fe-bb20-0c394a4bf2e0>

1. **Data Analysis**

The data analysis is performed on a few parameters like age, blood pressure, cholesterol, exercise heart rate and chest pain type. The parameters are selected according to the factors which affect the chances of heart disease.

* **Age**

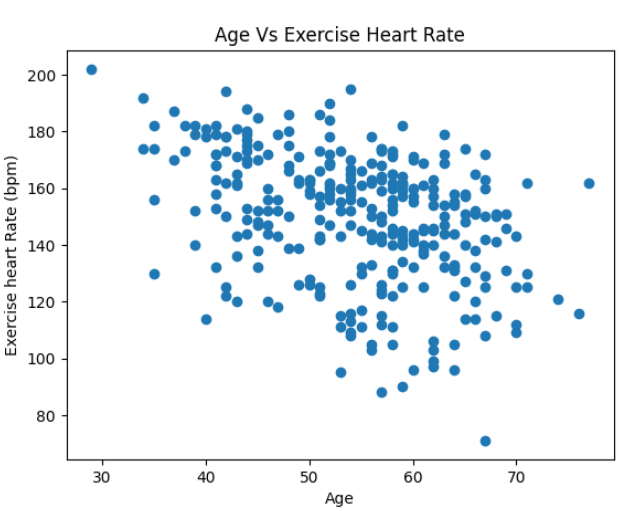
Aging can cause changes in the heart and blood vessels and that can be one of the risk factors of developing heart diseases.



The age limit between 50-55 has the chances of having heart disease compared to other age gaps. The age limit between 75- 80 are less chances to heart disease. Histogram plotted to show the age limit to have the chances of heart disease. X axis shows the age limit between 30 and 80.

* **Exercise heart rate**

Exercise stress testing is one of the most utilized tests in cardiology. The younger age group have high exercise heart compared to older age group.



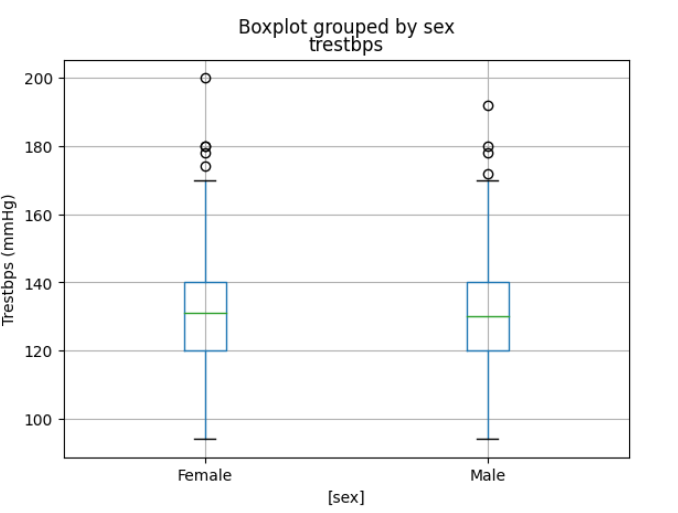
The scatter plot is plotted to show the exercise heart rate in different age groups. The plot clearly shows that the exercise hear rate decreases with age. Correlation between age and exercise heart rate is calculated as -0.35.

* **Resting blood pressure**

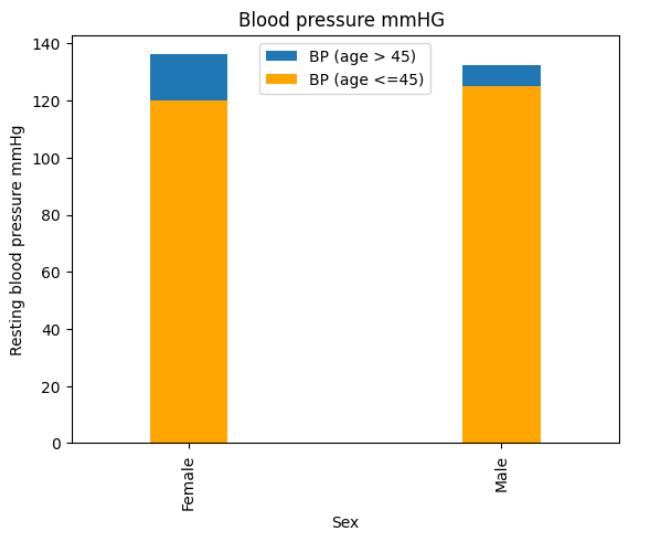
High blood pressure is the major risk for heart disease. Blood pressure is same in young men and women. The blood pressure increases in women after menopause.

A normal blood pressure level is between 120/80 mmHg.

The below box plot shows that the interquartile range of blood pressure is in between 120-140 mmHg. Blood pressure of few female and male falls above the maximum. The minimum to lower quartile shows the resting blood pressure between 80 -120 mmHg. Upper quartile to maximum shows the individuals have the resting blood pressure between 120-160 mmHg.



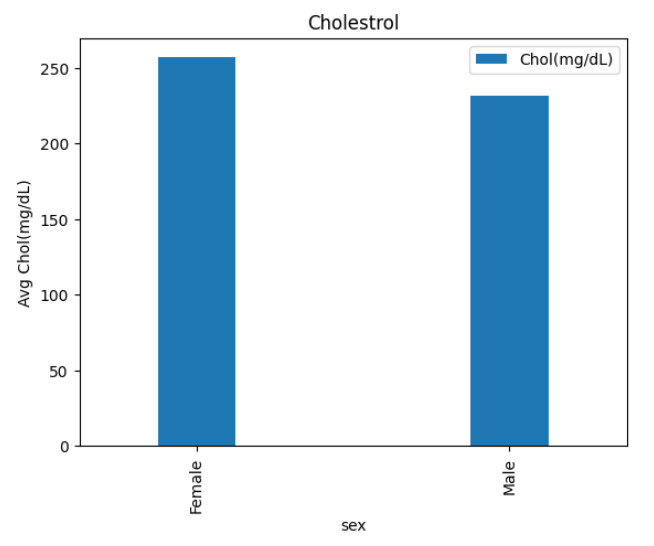
The box plot shows the distribution of blood pressure in men and women.



The bar plot shows the blood pressure in female and male after the age 45. Selected data set shows that blood pressure is high in women after the age of 45 compared to men.

* **Cholesterol**

Cholesterol is essential for human body to work although bad cholesterol can increase the chance of heart diseases. High cholesterol develops fatty deposits in blood vessels which in turn affect the blood flow through arteries.



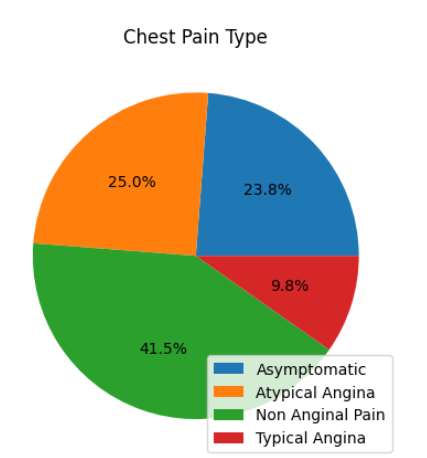
Bar plot is drawn below shows the average cholesterol in Male and Female who has the chance of heart disease. The average cholesterol in Female falls above 250 mg/dL and the average in male is in between 200-250 mg/dL.

* **Chest pain type**

Chest pain appears in many forms and most often it is associated with heart disease.

The heart disease shows 4 types of chest pains that are asymptomatic, atypical angina, non-anginal pain and typical angina.

Asymptomatic chest pain may not have any symptoms and intensity of classic heart attack. Atypical angina pain usually feels like a stabbing or burning pain in the chest and may sometimes have the characteristics like indigestion. Non anginal chest pain may also called as non- cardiac chest pain refers to the pain, they feel behind the breastbone. Typical angina pain is often described as squeezing, pressure, heaviness, tightness, or pain in the chest.



Pie chart is plotted to show the percentage of chest pain type in the heart disease dataset. 41.5 % have non anginal type pain. 25% of individuals have atypical angina and 23 % have asymptomatic chest pain.

1. **Summary**

Theory states that the parameters cholesterol, diabetes, and blood pressure are leading factors of heart disease. Highest values of these parameters are checked against the target value to verify the accuracy of Kaggle heart disease data set.

1. **References**

* <https://towardsdatascience.com/heart-disease-prediction-73468d630cfc>
* <https://libres.uncg.edu/ir/ecsu/f/Brandon_Simmons_Thesis-Final.pdf>

1. **Conclusion**

The Kaggle heart data is not accurate to perform heart disease analysis by considering all the parameters. The causes of heart attack can be smoking, high cholesterol, hypertension, a high fat diet, being overweight or obese etc. The analysis on Kaggle data set using cholesterol, diabetes and blood pressure is not showing accurate association with target value on Kaggle data set. So, the data set may be updated with more parameters and the recordings of more individuals to get a precise heart disease analysis.