

September 22, 2021

FACTORS THAT LEAD TO HEART DISEASES

EXPLORATORY DATA ANALYSIS PROJECT

BY TINOTENDA MHLANGA



Problem	2
Approach	2
UNIVARIATE ANALYSIS	3
BI-VARIATE ANALYSIS	9
MULTI-VARIATE ANALYSIS	19
Results and conclusion	26

TINOTENDA MHLANGA

EXPLORATORY DATA ANALYSIS PROJECT

FACTORS THAT LEAD TO HEART DISEASES

PROBLEM

We are presented with a dataset that has some attributes of a patient including one which determines whether a patient has heart disease or not. Using the attributes that are provided, I have to figure out the factors that are likely to cause heart diseases in patients.

APPROACH

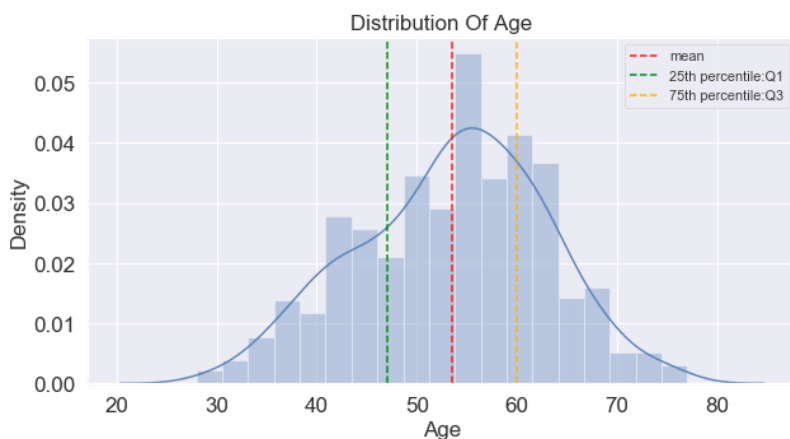
I used python, pandas and its packages and also numpy and its packages to perform the analyses and figure out the factors that cause heart diseases in patients. The Jupyter Notebook can be found on Github. I followed three main approaches which were basically:

1. **Perform Univariate Analysis:** In this approach, I first analyzed relevant columns(only one column at a time) and obtained some insights.
2. **Perform Bivariate Analysis:** Next, I performed bivariate analysis whereby I analyzed columns in pairs so as to get a deeper understanding of the data and gain meaningful insights.
3. **Perform Multivariate Analysis:** Finally, I performed some multivariate analysis, whereby I analyzed mutple columns at once so as to get even more meaningful insights.

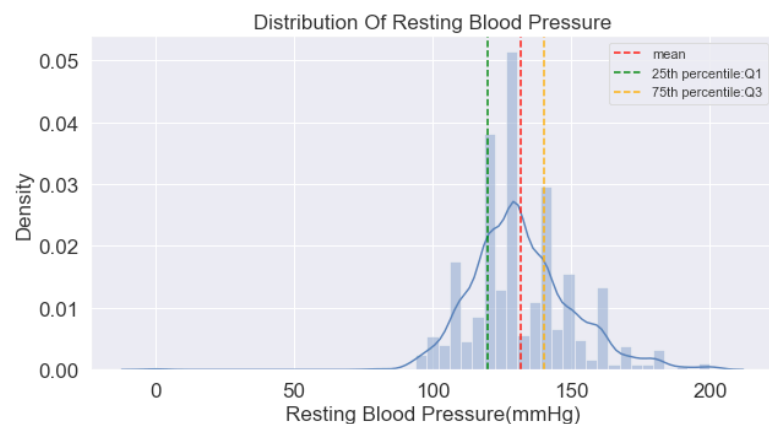
UNIVARIATE ANALYSIS

I obtained some insights when I performed univariate analysis, and these are outlined below.

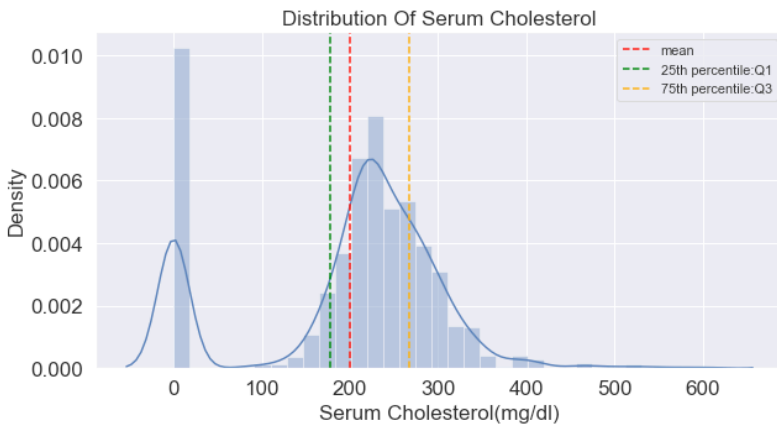
First, I looked at the distributions of different variables.



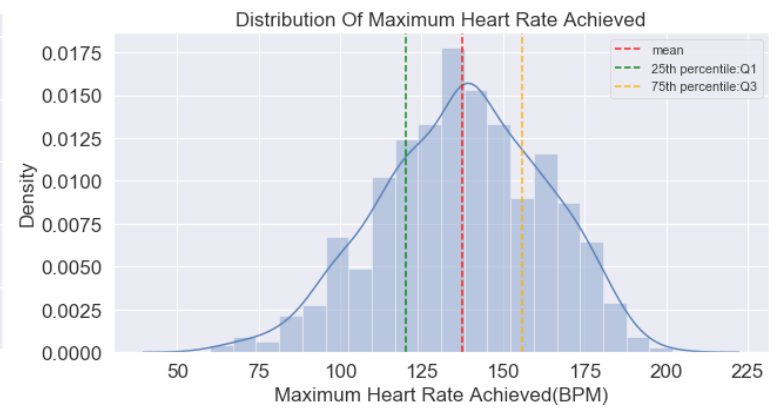
a. Distribution of Age



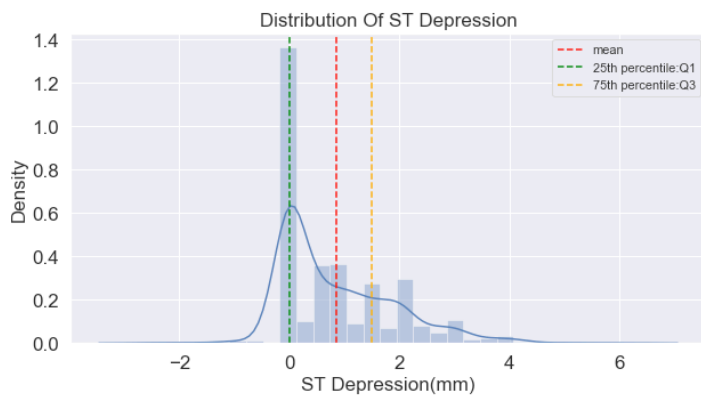
b. Distribution of Resting Blood Pressure



c. Distribution of Serum Cholesterol



d. Distribution of Maximum Heart Rate Achieved



e. Distribution of ST-Depression

A. DISTRIBUTION OF AGE

Here, we can see that the data is normally distributed, with the mean lying around 55 years.

B. DISTRIBUTION OF RESTING BLOOD PRESSURE

The mean of the resting blood pressure is 130mmHg(millimetres of Mercury). The data is skewed to the left as noone has a resting blood pressure bewteen 0-80mmHg.

C. DISTRIBUTION OF SERUM CHOLESTEROL

The data is skewed to the right, with the mean being 200mg/dl(milligrams per decilitre).

D. DISTRIBUTION OF MAXIMUM HEART RATE ACHIEVED

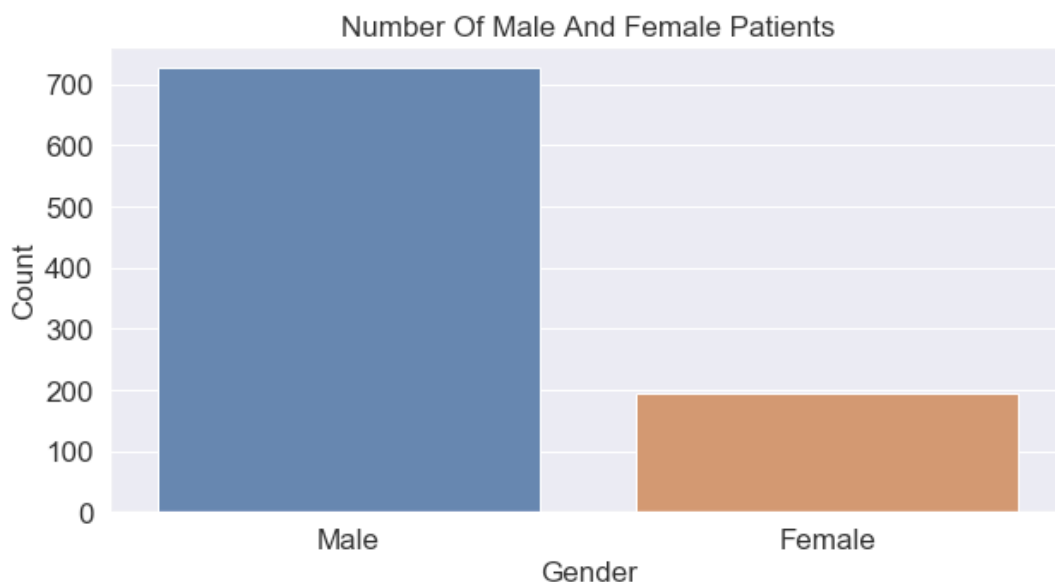
The mean is around 153 Beats per Minute, and the data is normally distributed.

E. DISTRIBUTION OF ST-DEPRESSION

Most patients have an ST-Depression of 0. The mean is 1mm and the data is skewed to the right.

NUMBER OF MALE PATIENTS COMPARED TO FEMALE PATIENTS

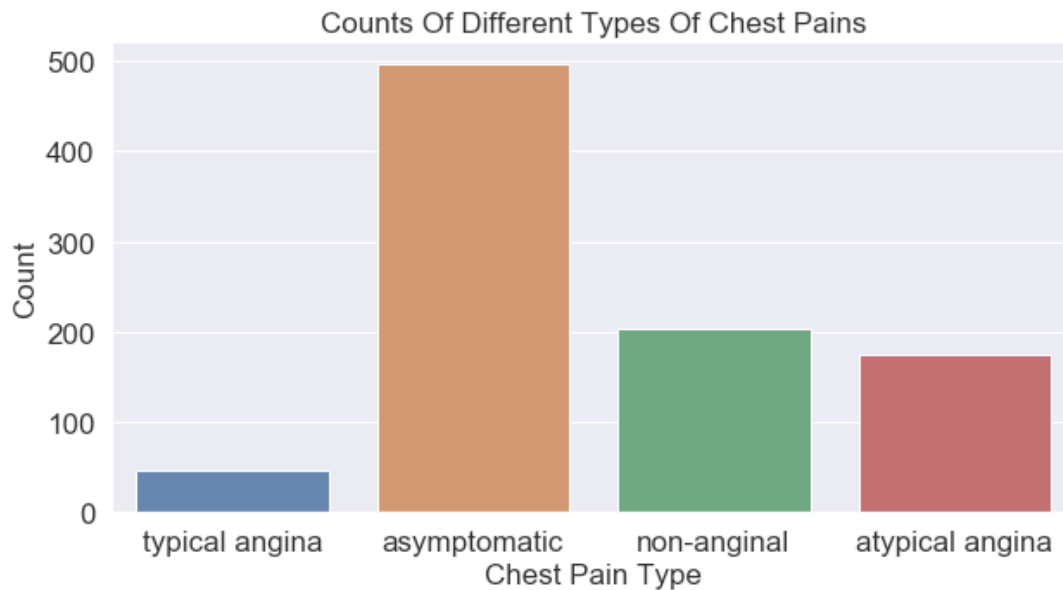
Next, I visualized the number of male and female patients.



Here, we can see that male patients are more compared to female patients. Male patients are over 700 while female patients are just 200.

CHEST PAIN THAT IS MOST COMMON

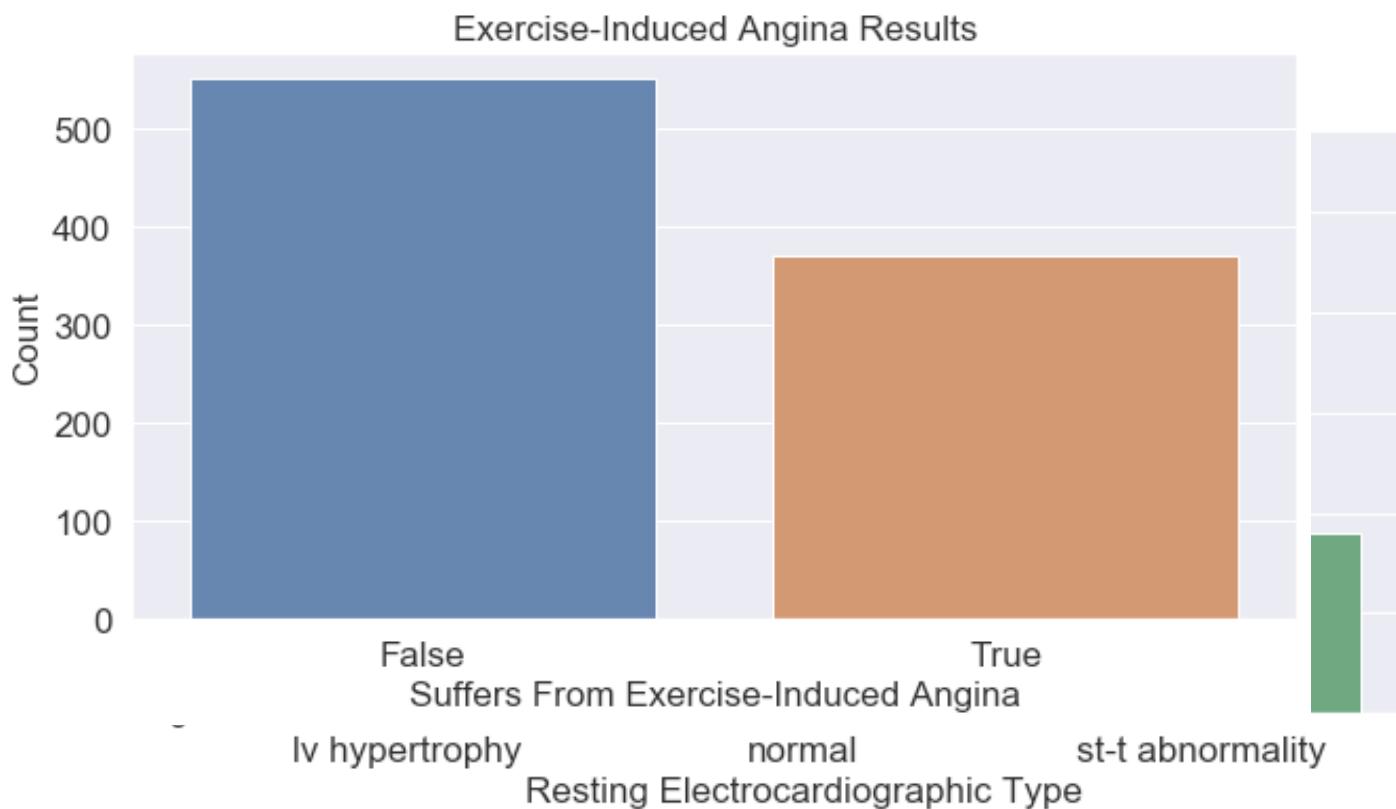
Next, I visualized the chest pain that is most common. Non-anginal means that its a regular chest pain which does not cause any heart disease.



From the results, we can see that asymptomatic angina is most common compared to the other types of angina.

VISUALIZE THE RESTING ELECTROCARDIOGRAPHIC RESULTS

Next, I visualized the resting electrocardiographic results. Electrocardiogram is basically a test that measures the electrical activity of the heart.



From the results, we can see that most patients have a normal resting electrocardiographic result. Iv hypertrophy and st-t abnormality are the same as around 190 patients have these kinds of results.

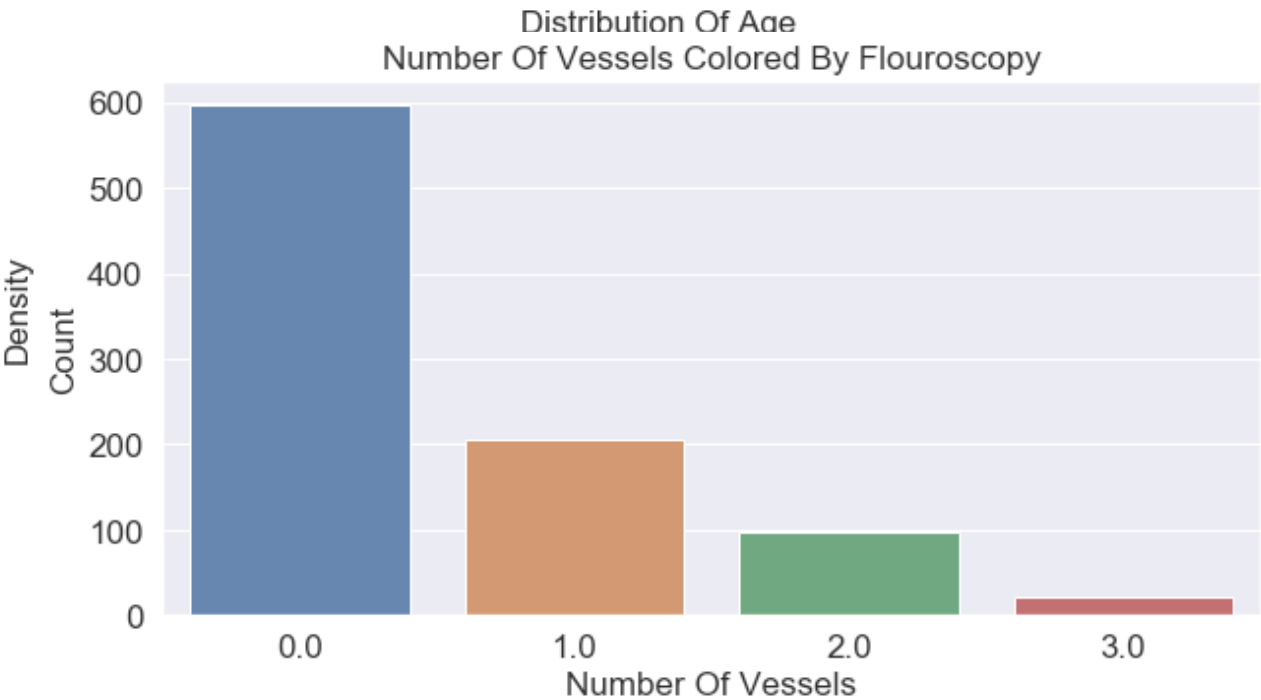
FIGURE OUT THE NUMBER OF PATIENTS WHO SUFFER FROM EXERCISE-INDUCED ANGINA VS THOSE WHO DO NOT

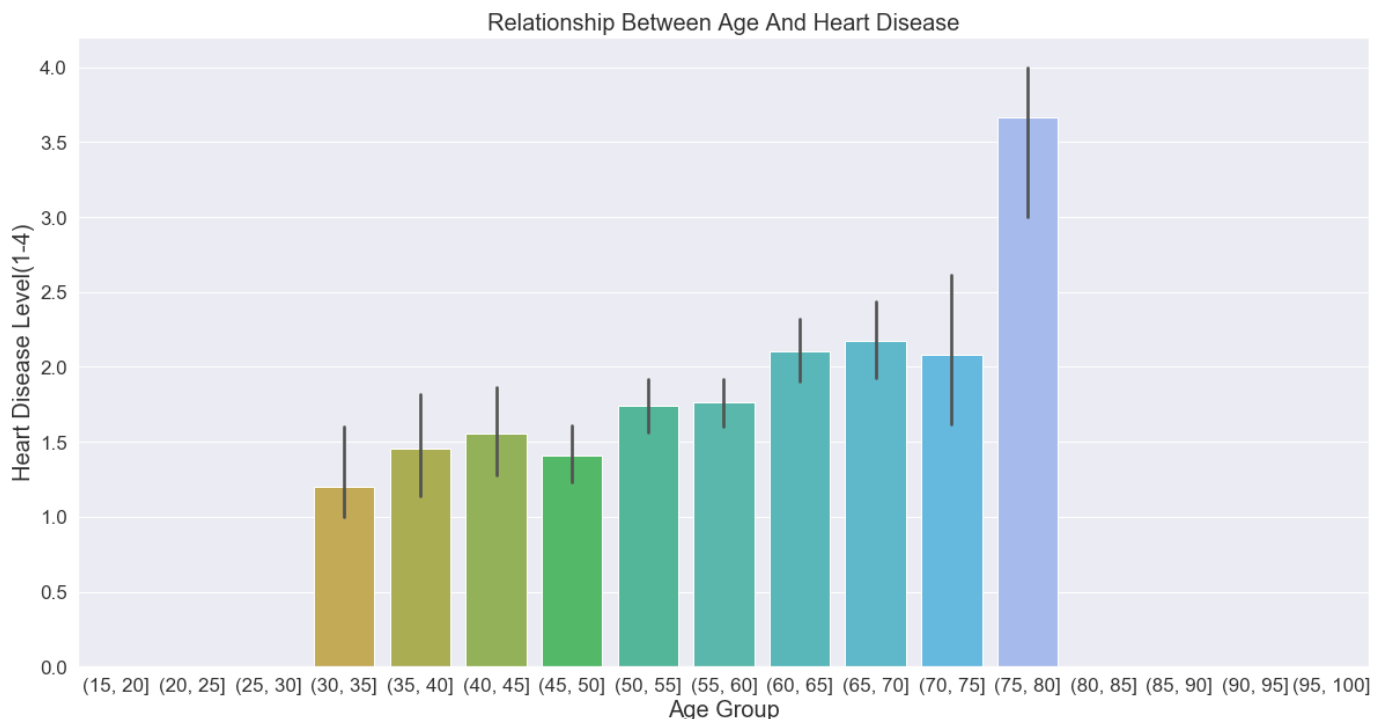
I then visualized the number of patients who suffer from exercise-induced angina and those who do not.

Looks like most patients do not suffer from exercise-induced angina as over 500 reported false and about 380 reported true.

VISUALIZE THE NUMBER OF VESSELS COLORED BY FLOUROSCOPY

Next, I visualized the number of vessels colored by flourosocopy.





Most patients had no vessels colored by fluoroscopy as 600 reported 0, and the least was 3 vessels as about only 20 reported this result.

This then concluded my Univariate analysis, and I then moved forward to the bivariate analysis as this was going to make me recognize the factors that cause heart diseases.

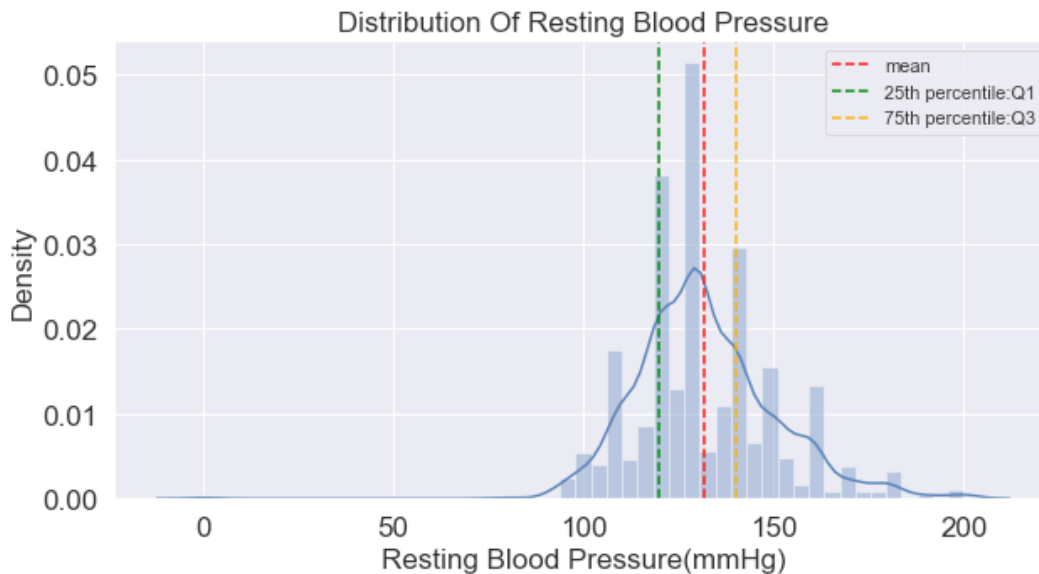
BI-VARIATE ANALYSIS

In Bi-Variate Analysis, I analyzed two attributes at a time, and the num attribute was used in all of my analyses. This is because it is the one that determines whether a person has a heart condition or not, and it has values ranging from 1-4. These values represent the presence of heart disease.

UNIVARIATE RESULTS COMPARED WITH BI-VARIATE RESULTS

FIGURE OUT THE RELATIONSHIP BETWEEN AGE AND HEART DISEASE

In the



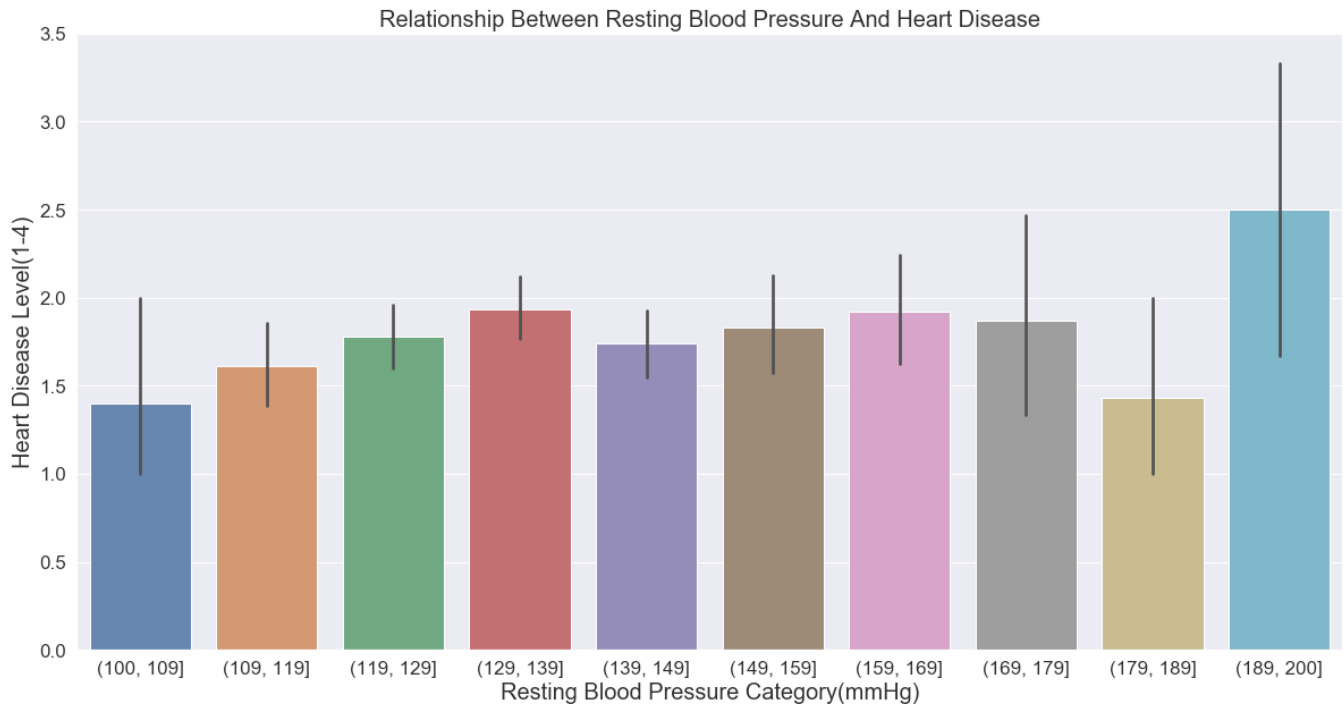
Univariate analysis, we saw that the mean of the age is 55. It is now time to figure out if this age is the one that has the most people that suffer from heart disease or not.

From the Univariate results, we saw that the mean of the age is 55 years. However, from the bivariate results, we can see that most people who suffer from heart diseases are old. They are in the range of 76-80 years old. People at this age group usually have a heart disease level that is around 3.6. Hence, it can be concluded that old age leads to a risk of getting a heart disease.

***FIRST FACTOR => OLD AGE**

FIGURE OUT THE RELATIONSHIP BETWEEN RESTING BLOOD PRESSURE AND HEART DISEASE

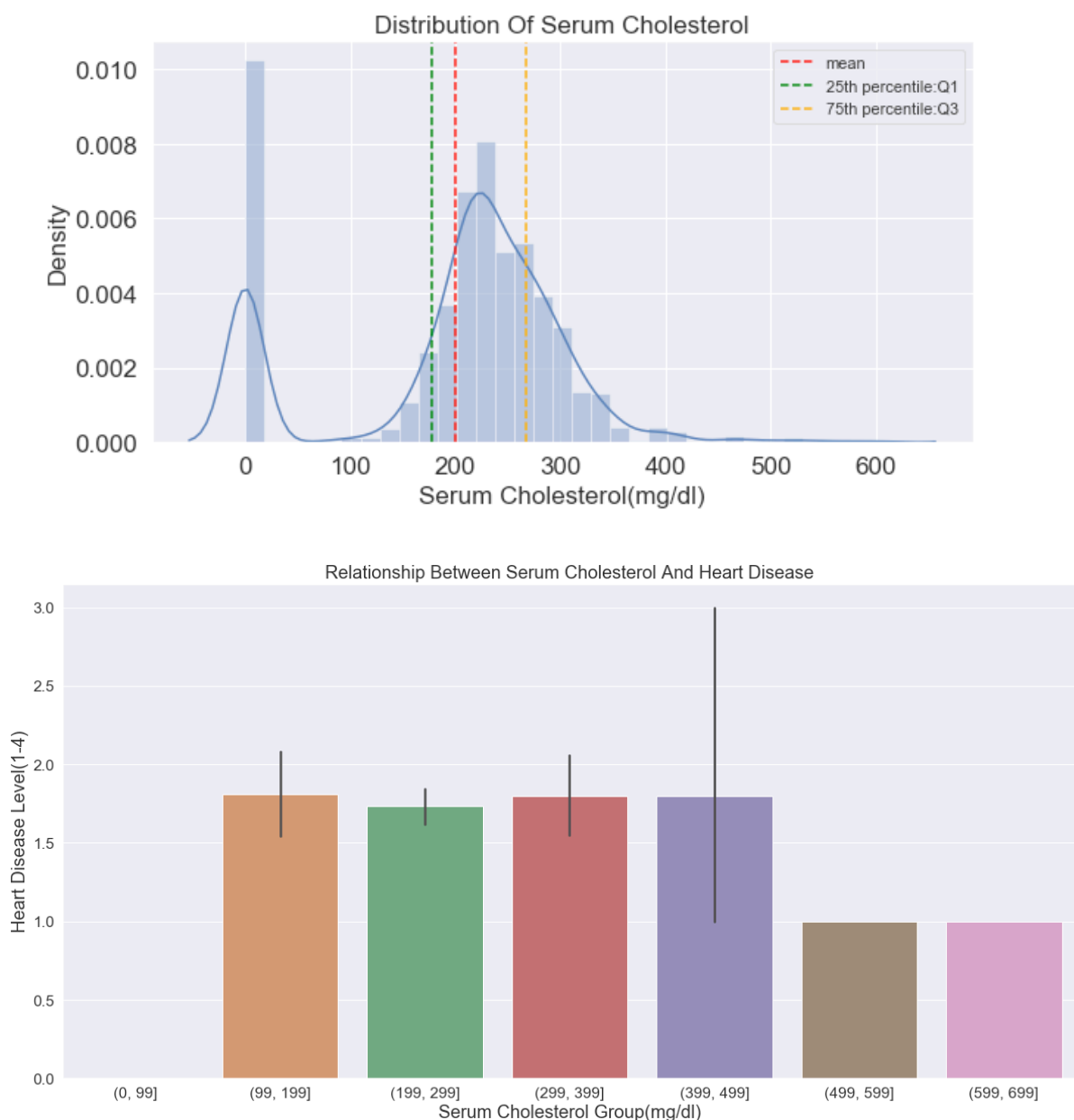
In the univariate analysis, we saw that the mean of the resting blood pressure is 130mmHg. It is now time to figure out if people with this resting blood pressure suffer from heart disease or not.



People with a resting blood pressure of 130-129mmHg seem to have a heart disease level of around 1.8. However, the majority of the people who have heart diseases have a resting blood pressure in the range 190-200mmHg. Their heart disease level is 2.5. Hence, it can be concluded that a resting blood pressure that is in the range 190-200mmHg can cause a heart disease.

***SECOND FACTOR => RESTING BLOOD PRESSURE THAT IS IN THE RANGE OF 190-200MMHG**

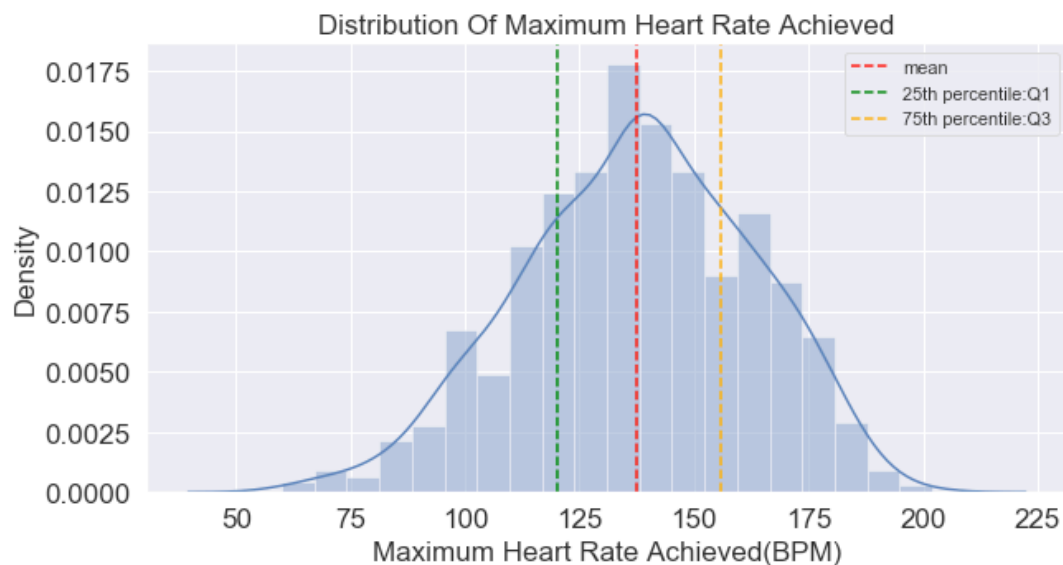
FIGURE OUT THE RELATIONSHIP BETWEEN SERUM CHOLESTEROL AND HEART DISEASE



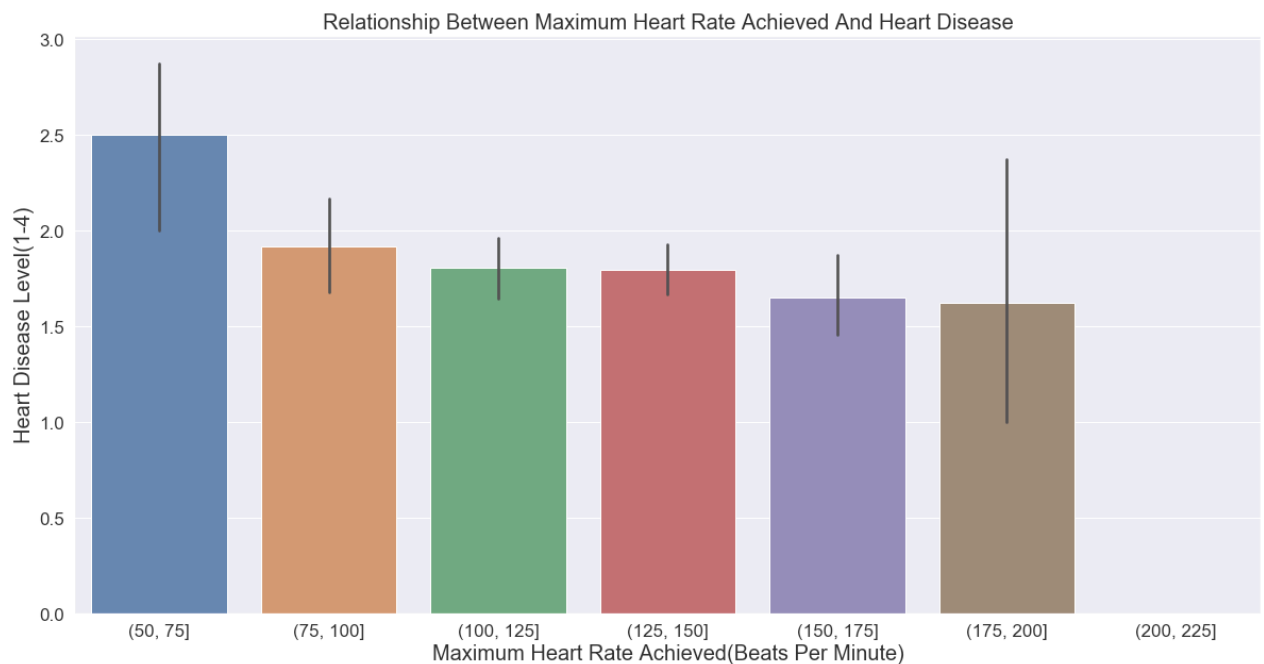
The univariate analysis showed that mean of serum cholesterol was 200mg/dl. However the bi-variate analysis shows that most people with a serum cholesterol of 100-199mg/dl are the ones who suffer from heart disease. Their heart disease level is about 1.7.

***THIRD FACTOR => SERUM CHOLESTEROL IN THE RANGE 100-199MG/DL**

FIGURE OUT THE RELATIONSHIP BETWEEN MAXIMUM HEART RATE ACHIEVED AND HEART DISEASE



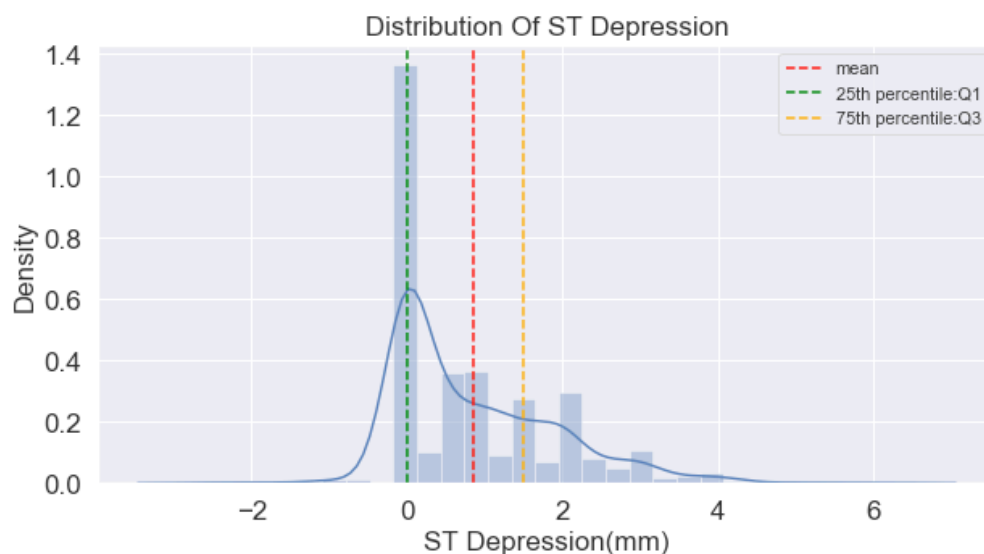
The

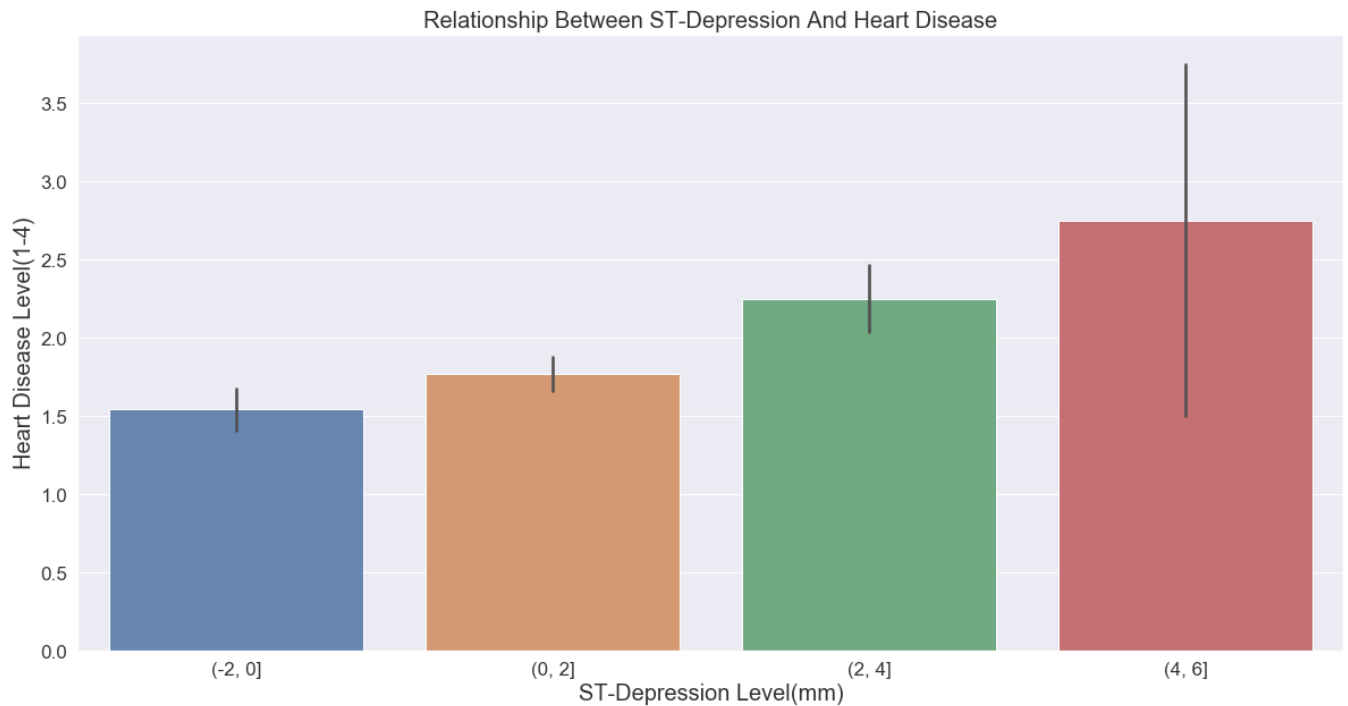


univariate analysis showed that the mean of the maximum heart rate achieved is around 153 Beats Per Minute. From the bi-variate analysis however, it can be concluded that people with a maximum heart beat rate that is in the range 51-75 usually suffer from heart disease, and their heart disease level is 2.5.

***FOURTH FACTOR => MAXIMUM HEART RATE THAT IS IN THE RANGE 51-75 BEATS PER MINUTE**

FIGURE OUT THE RELATIONSHIP BETWEEN ST-DEPRESSION AND HEART DISEASE

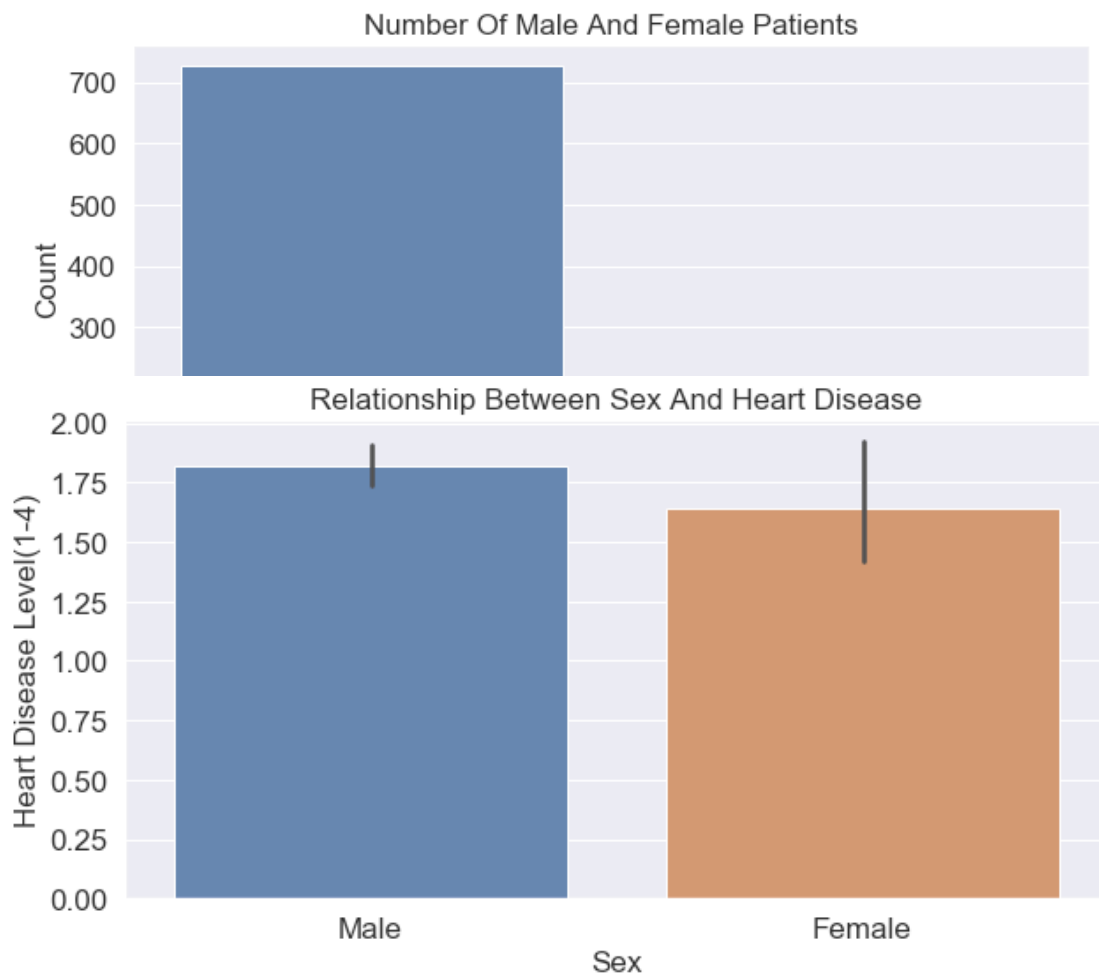




In

the univariate analysis, it can be seen that the mean of the ST-Depression is 1mm. From the bi-variate analysis, it can be concluded that an ST-Depression that is in the range 5-6mm leads to heart disease. People with this ST-Depression level have a heart disease level of about 2.6.

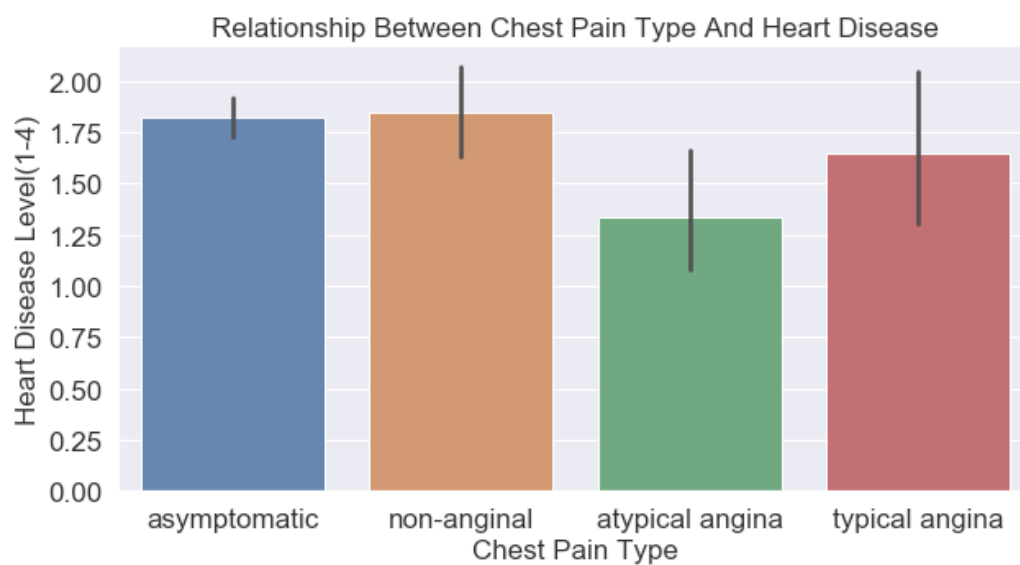
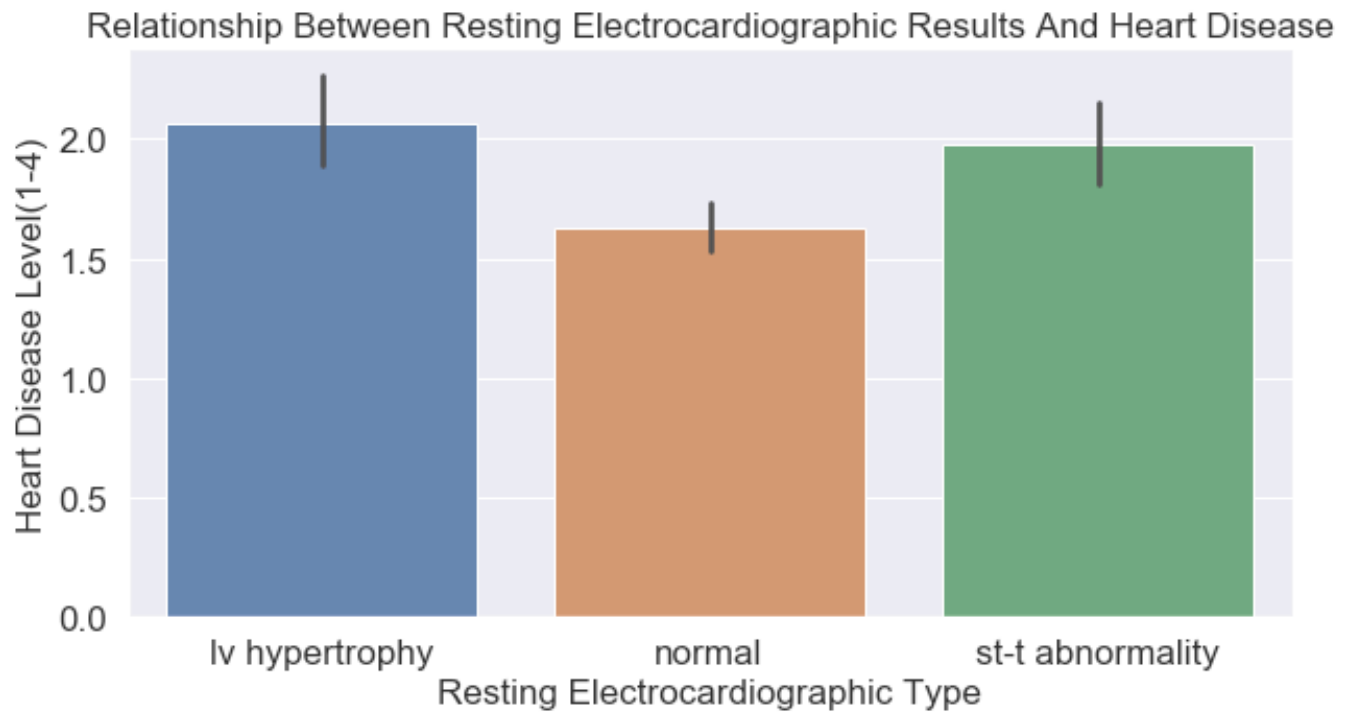
FIGURE OUT THE RELATIONSHIP BETWEEN SEX AND HEART DISEASE



The univariate analysis showed that there are more male patients compared to female patients. In the bi-variate analysis, it can be seen that males are more prone to heart disease as a large number if them have a heart disease level of about 1.78.

***FIFTH FACTOR=>BEING MALE**

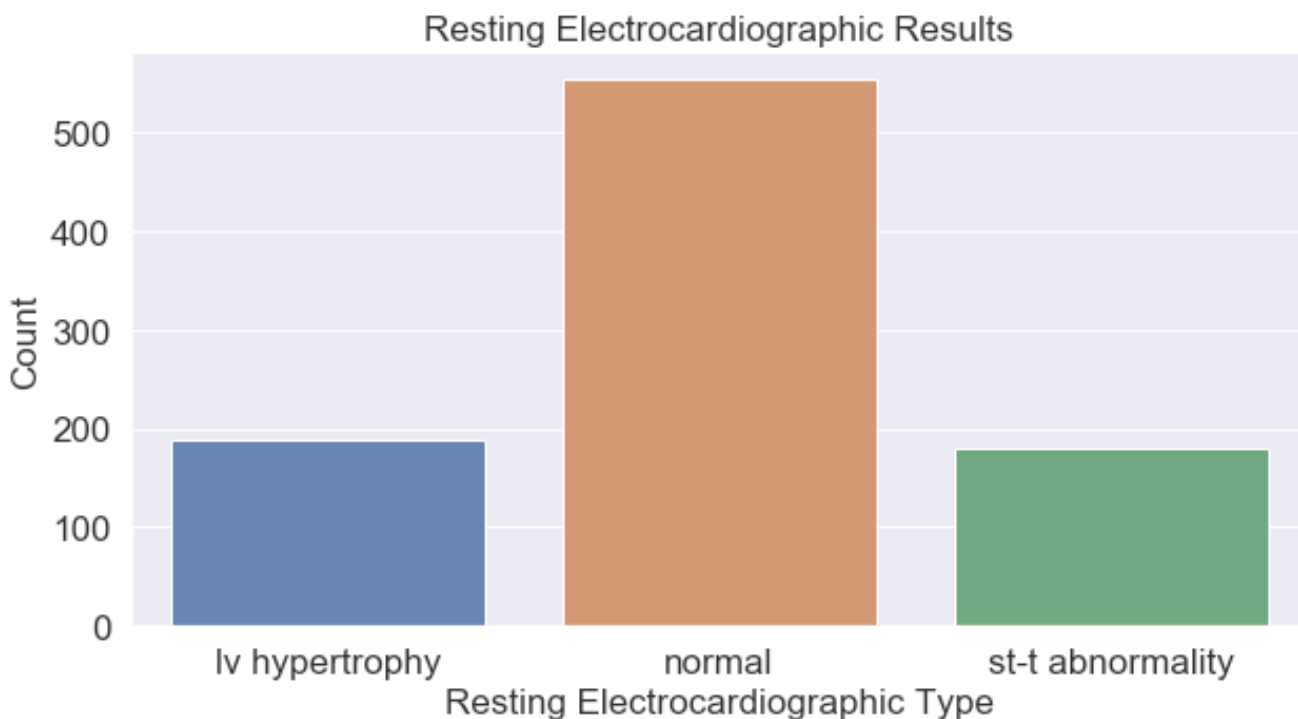
FIGURE OUT THE RELATIONSHIP BETWEEN CHEST PAIN AND HEART DISEASE

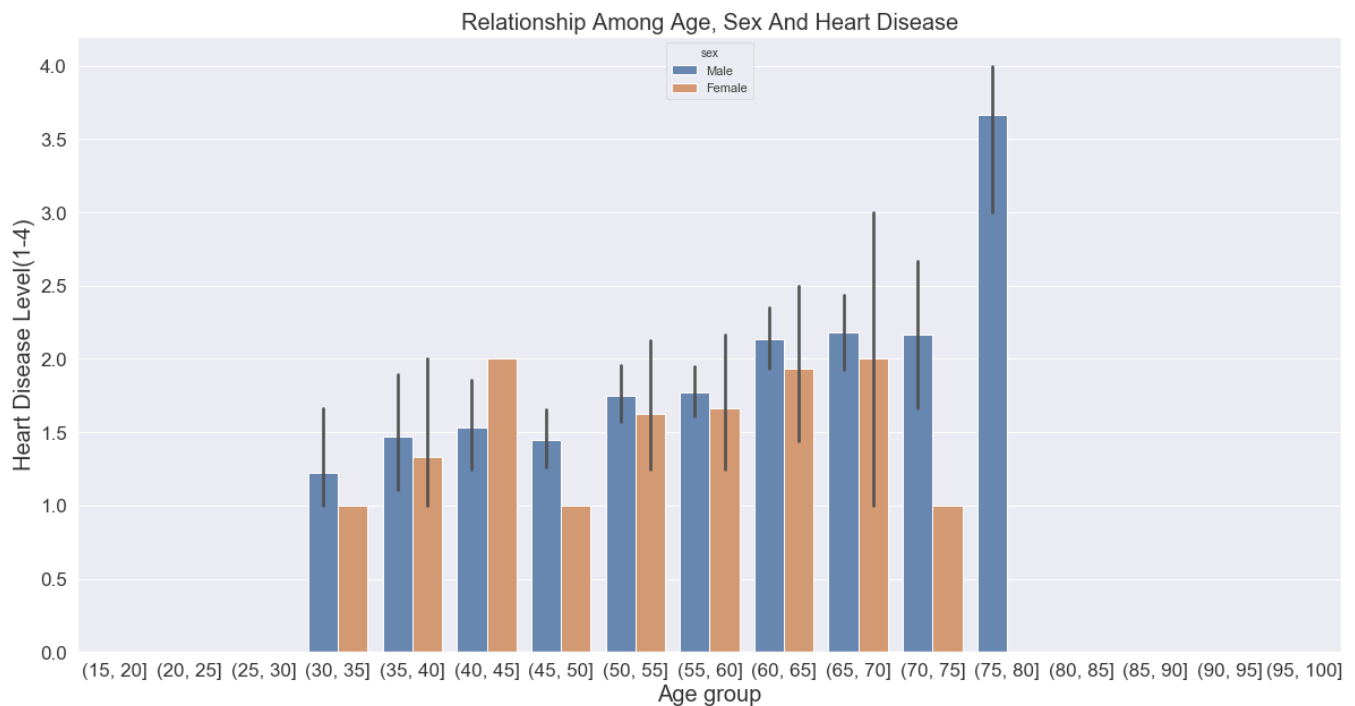


In univariate analysis, we saw that asymptomatic angina is most common. The bi-variate analysis results show that most people suffer from asymptomatic angina, as the heart disease level is above 1.75. Hence, it can be concluded that people with asymptomatic angina can have a heart disease. Non-anginal can be excluded as it is just normal pain that does not lead to heart disease. However, as it can be seen that a lot of people suffer from it, it can be concluded that people with non-angina frequently end up having heart diseases.

***SIXTH FACTOR=>ASYMPTOMATIC ANGINA AND ALSO FREQUENT NON-ANGINA PAINS CAN END UP RESULTING IN A HEART DISEASE.**

VISUALIZE THE RELATIONSHIP BETWEEN ELECTROCARDIOGRAPHIC RESULTS AND CHEST PAIN





Univariate Analysis showed that many people have normal electrocardiographic results. The bi-variate analysis results show that people with Iv hypertrophy electrocardiographic results usually suffer from heart diseases. Their heart disease level is about 2.1.

***SEVENTH FACTOR=>IV HYPERTROPHY ELECTROCARDIOGRAPHIC RESULTS**

MULTI-VARIATE ANALYSIS

After obtaining some of the factors from the bi-variate analysis, it was now time to get even deeper insights using multi-variate analysis.

The steps I followed are listed below.

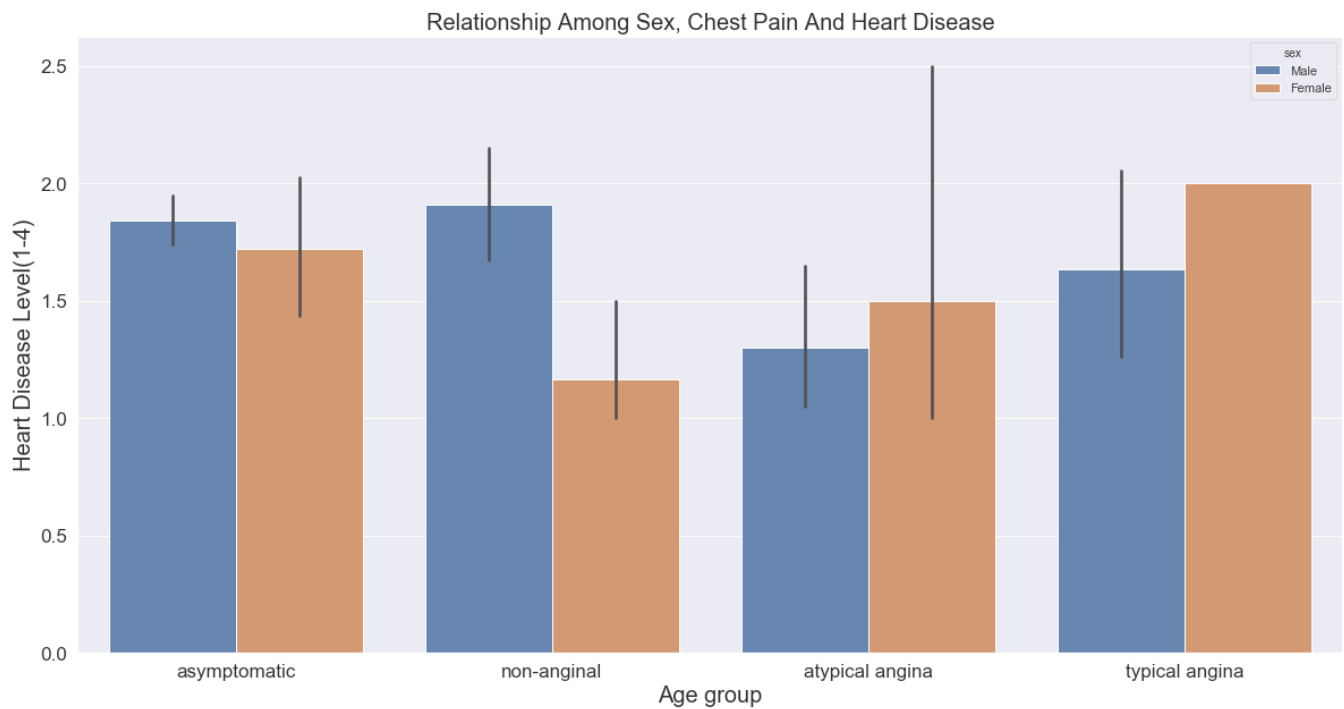
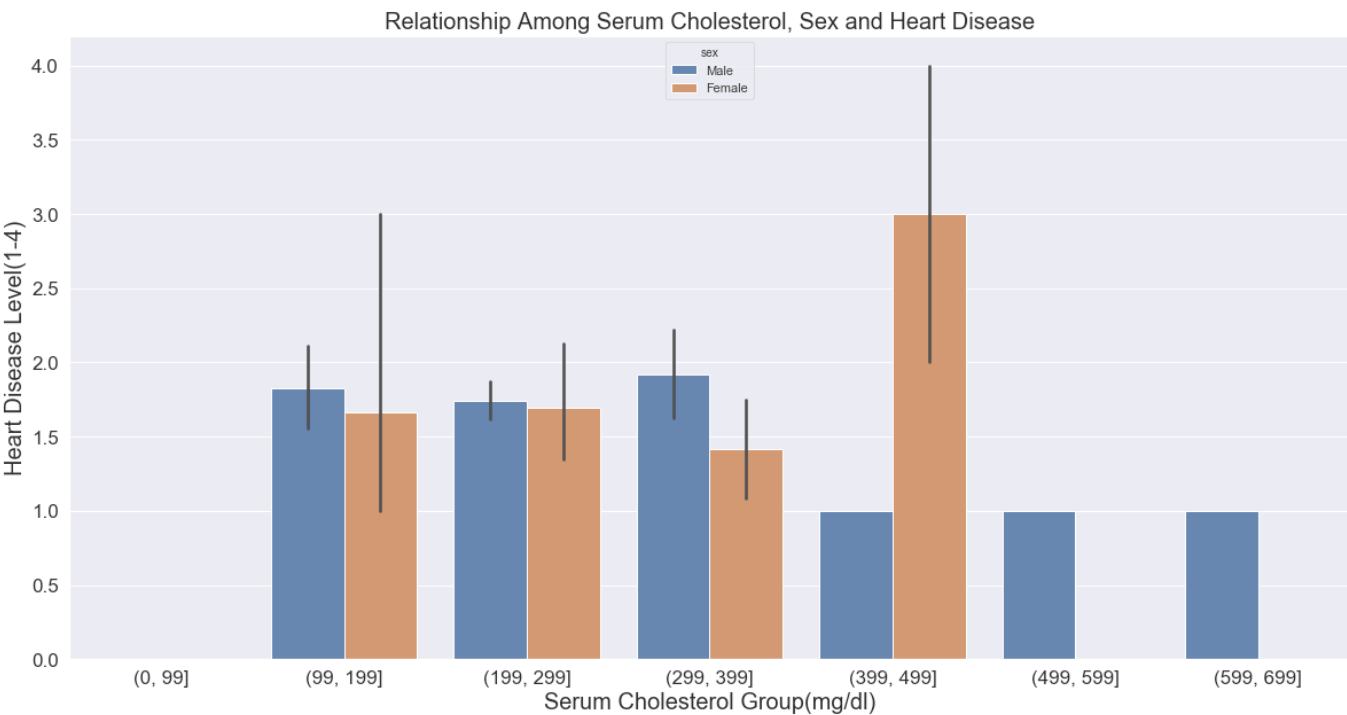
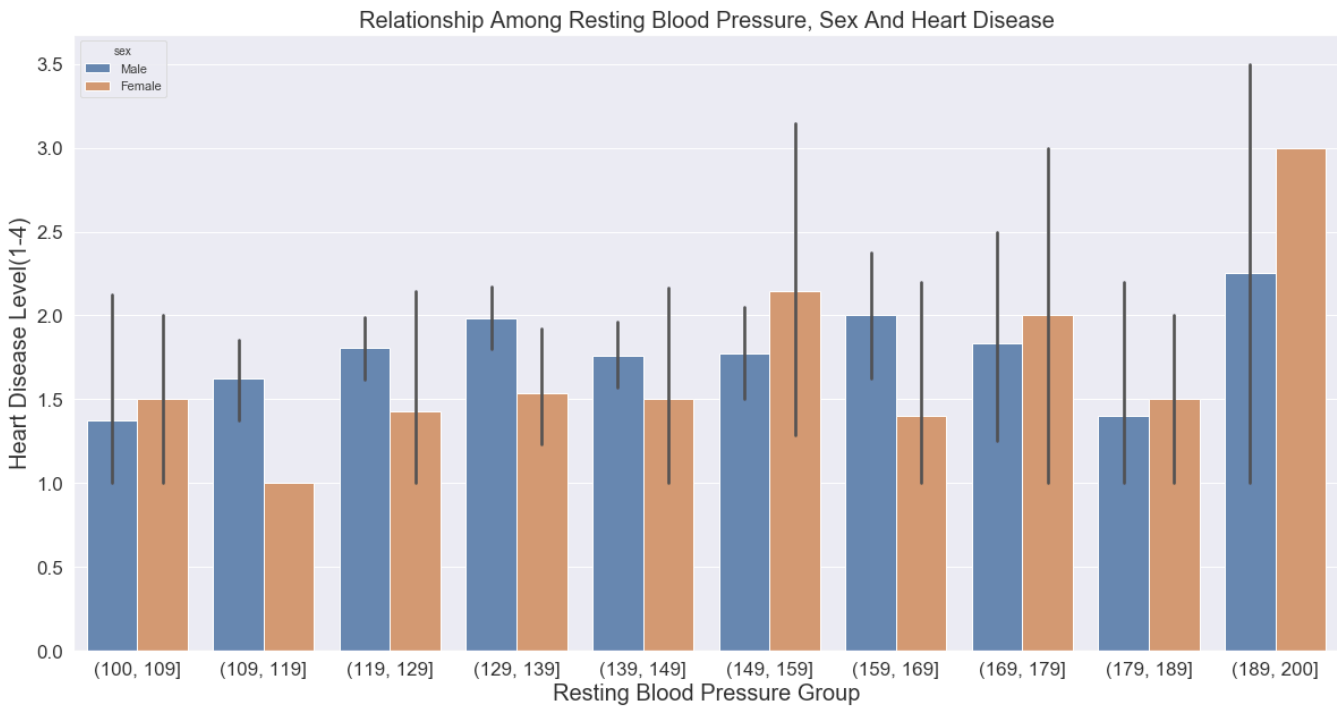


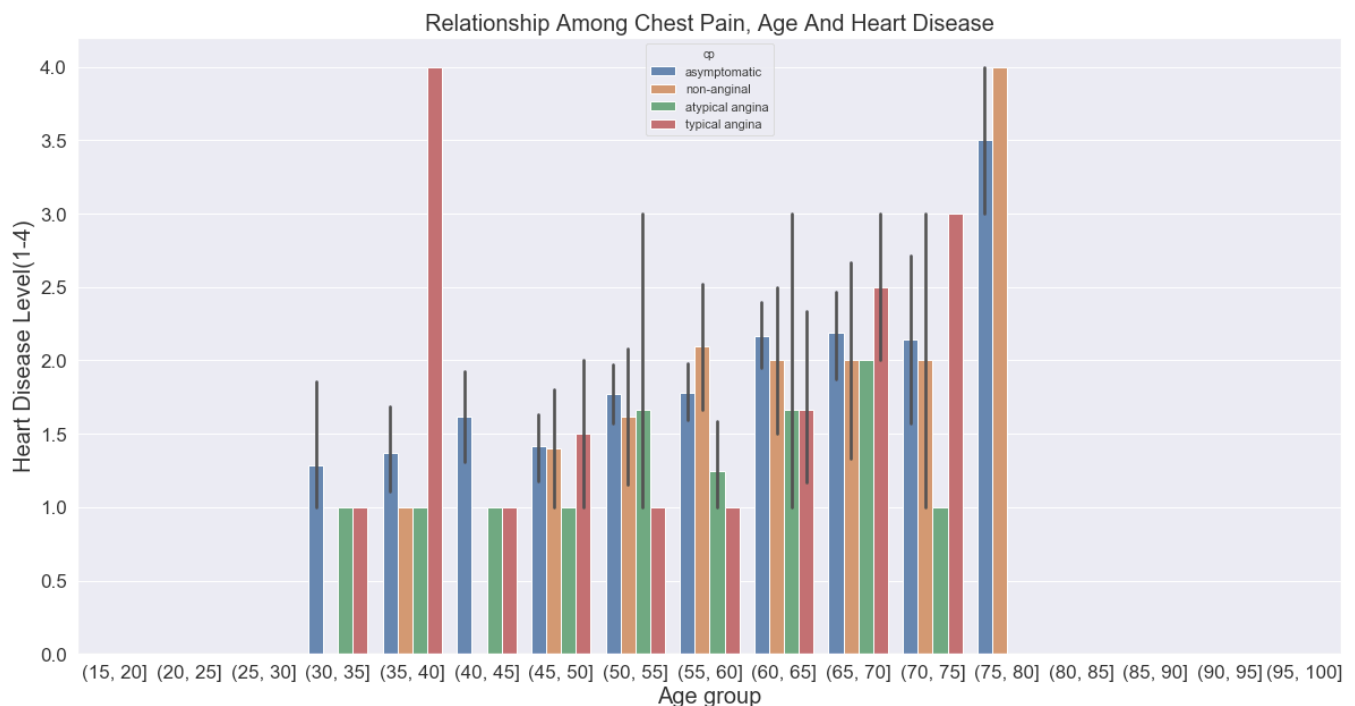
FIGURE OUT THE RELATIONSHIP AMONG AGE, SEX AND HEART DISEASE

From the results, it can be seen that most males suffer from heart disease, and they are in the 75-80 age group.

***FIRST AND FIFTH FACTOR COMBINED**

FIGURE OUT THE RELATIONSHIP AMONG SEX, CHEST PAIN AND HEART DISEASE



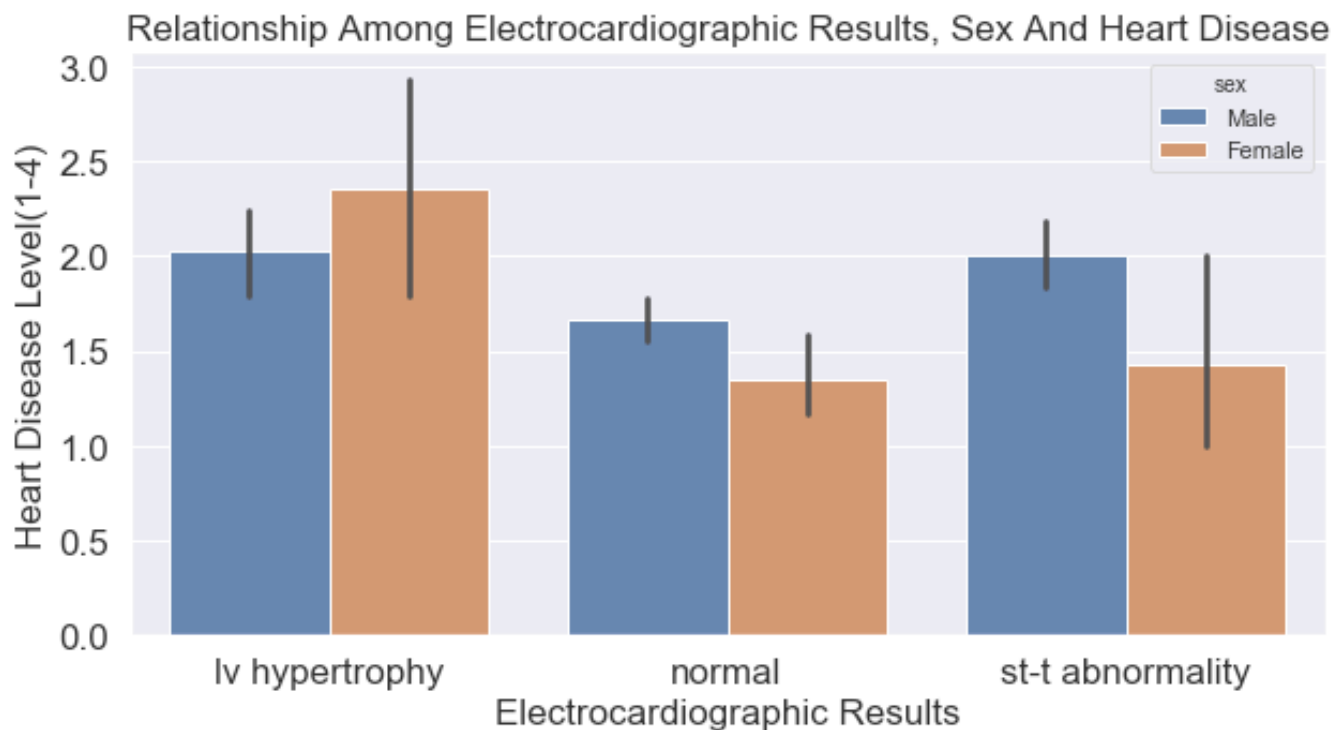


It can be seen that most males have asymptomatic angina and also non-anginal symptoms. However, as discussed in the bi-variate analysis, frequent non-anginal symptoms can actually lead to heart disease. A new insight can also be seen which suggests that most females suffer from typical angina.

***FIFTH AND SIXTH FACTOR COMBINED + MOST FEMALES WHO HAVE TYPICAL ANGINA HAVE HEART DISEASES**

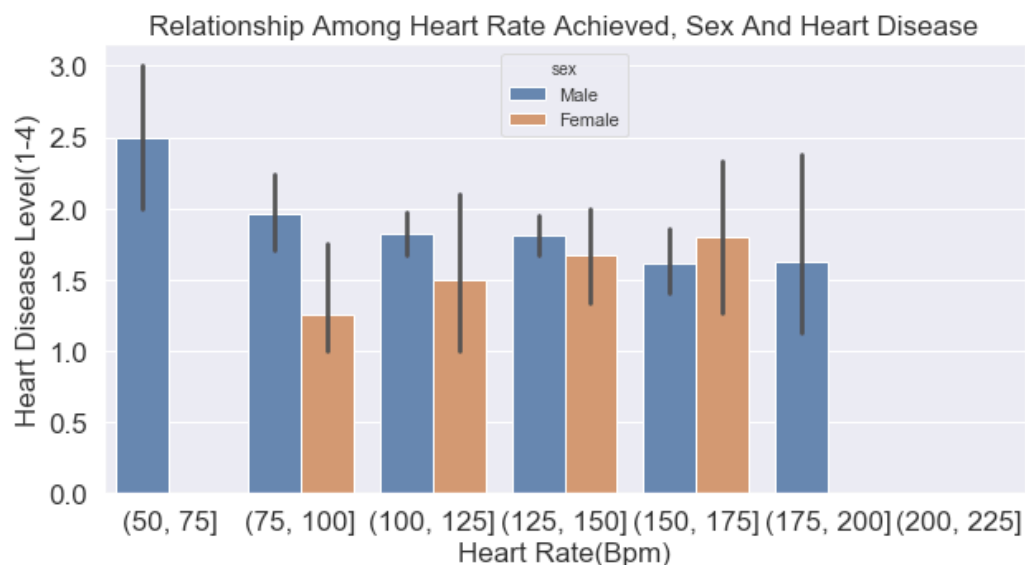
FIGURE OUT THE RELATIONSHIP AMONG CHEST PAIN, AGE AND HEART DISEASE

People in the 36-40 age group seem to have typical angina which causes heart disease. Same goes for people in the 46-50 range, 66-70 range and



71-75 range. 31-35, 41-45, 51-55, and 61-65 have asymptomatic angina which leads to heart diseases. The 76-80 group has both high asymptomatic angina and non-anginal effects. The 56-60 age group also has non-anginal effects. However, these frequent non-anginal effects can lead to heart diseases.

***SIXTH FACTOR=>TYPICAL ANGINA(36-40, 46-50, 66-70 AND 71-75YEARS), ASYMPTOMATIC ANGINA(31-35, 41-45, 51-55, 61-65 AND 76-80YEARS), NON-ANGINAL(56-60 AND 76-80YEARS)**



RELATIONSHIP AMONG RESTING BLOOD PRESSURE, SEX AND HEART DISEASE

Seems like most females with a resting blood pressure of 190-200mmHg have a high level of heart disease. The same goes for males too.

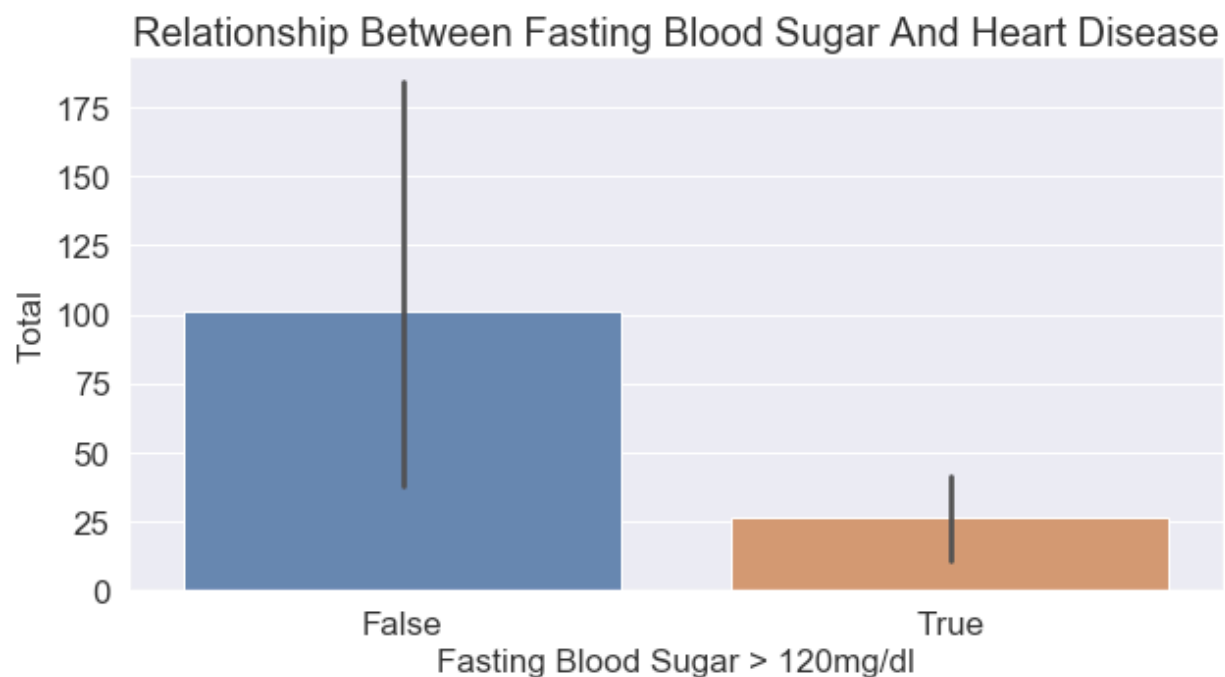
***SEVENTH FACTOR=> MALES AND FEMALES WITH A RESTING BLOOD PRESSURE THAT IS IN THE RANGE 190-200MMHG**

RELATIONSHIP AMONG SERUM CHOLESTEROL, SEX AND HEART DISEASE

Seems like females with a 400-499mg/dl serum cholesterol have a high heart disease level while males with a serum cholesterol of 300-399 have a high heart disease level.

***THIRD FACTOR ELIMINATED**

FIGURE OUT THE RELATIONSHIP BETWEEN FASTING BLOOD SUGAR AND HEART DISEASE



Seems like a blood-sugar level which is greater than 120mg/dl does not cause heart diseases.

FIGURE OUT THE RELATIONSHIP AMONG RESTING CARDIOGRAPHIC RESULTS, SEX AND HEART DISEASE LEVEL

Males and females with lv hypertrophy suffer from heart disease more often. The number of females exceeds that of males.

***ELIMINATE FIFTH FACTOR FROM BI-VARIATE ANALYSIS**

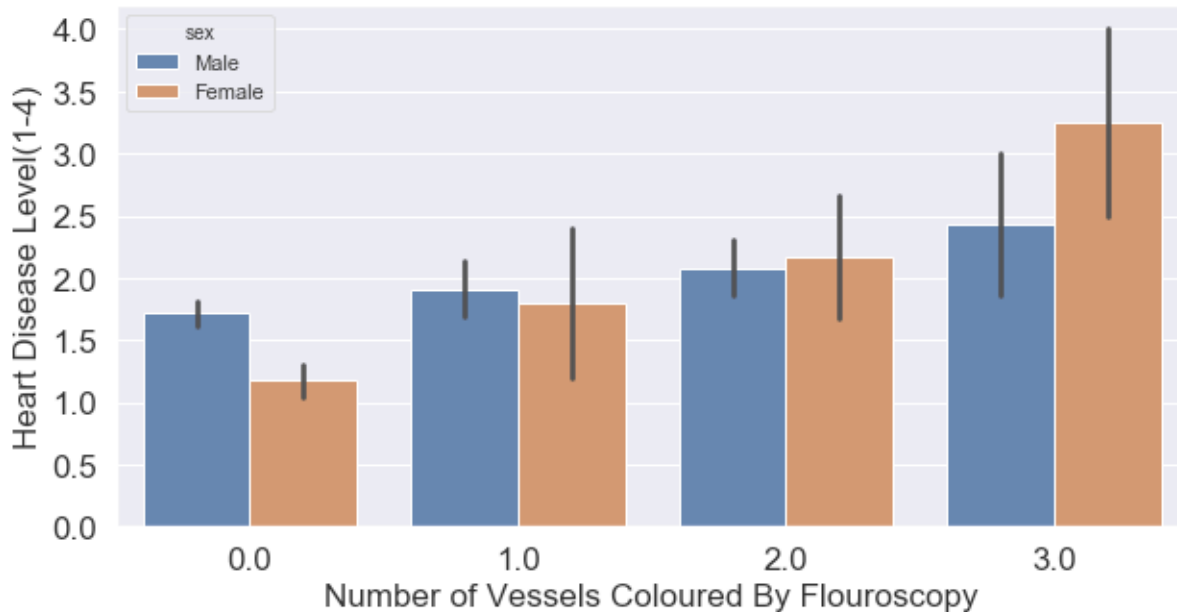
FIGURE OUT THE RELATIONSHIP AMONG MAXIMUM HEART RATE ACHIEVED, SEX AND HEART DISEASE

Most males with a maximum heart rate in the range 51-75 Beats Per Minute suffer from heart disease. Females with a maximum heart rate that is in the range 151-175 are the ones who suffer from heart disease.

***ELIMINATE FOURTH FACTOR FROM BI-VARIATE ANALYSIS**

FIGURE OUT THE RELATIONSHIP AMONG THE NUMBER OF VESSELS COLOURED BY FLOUROSCOPY, SEX AND HEART DISEASE

Relationship Among Number of Vessels Coloured By Flouroscopy, Sex And Heart Disease



Most males and females with 3 vessels coloured by flouroscopy suffer from heart disease.

RESULTS AND CONCLUSION

FACTORS THAT LEAD TO HEART DISEASES

- A. Being male and having an old age(75-80 year range).
- B. Asymptomatic angina in males and typical angina in females. An assumption is also made that frequent non-anginal effects can end up resulting in heart diseases for both males and females.

- C. The following age groups usually have typical angina which leads to heart diseases: (36-40, 46-50, 66-70, 71-75). The following age groups usually have asymptomatic angina which leads to heart disease: (31-35, 41-45, 51-55, 61-65, 76-80). The following age groups have non-anginal effects, but an assumption is made that frequent non-anginal effects will lead to heart diseases: (56-60, 76-80).
- D. Males and Females with a resting blood pressure that is in the range 190-200mmHg(millimetres of Mercury).
- E. Females with a serum cholesterol level that is in the range 400-499mg/dl(milligrams per decilitre) and males with a serum cholesterol level that is in the range 300-399mg/dl.
- F. Males and females with lv hypertrophy electrocardiographic results.
- G. Males with a maximum heart rate that is in the range of 51-75 Beats Per Minute and females with a maximum heart rate that is in the range of 151-175 Beats Per Minute.
- H. Males and females with 3 vessels coloured by flouroscopy.

As can be seen from the results, quite a number of factors lead to heart conditions for both males and females in different age groups.