
Wireframe Document

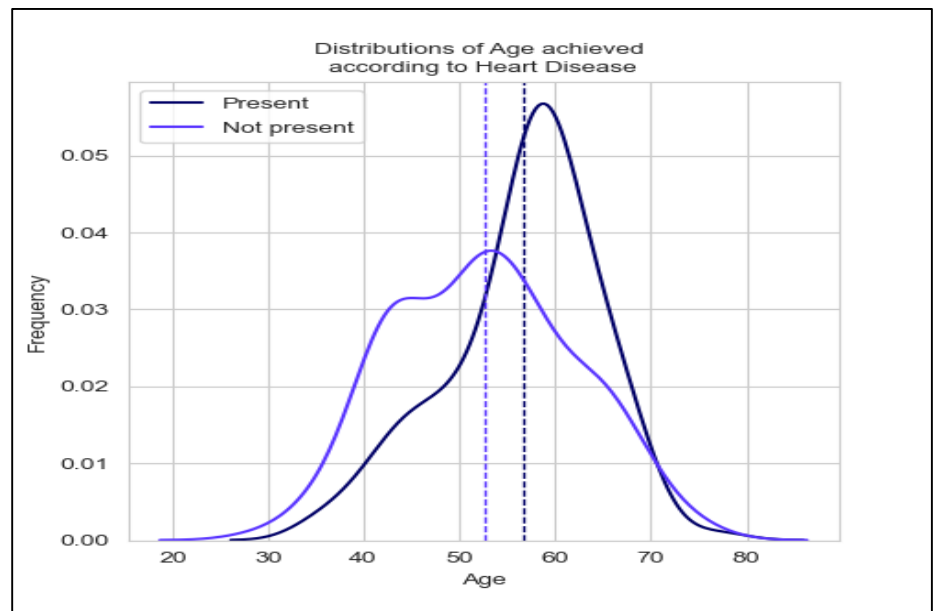
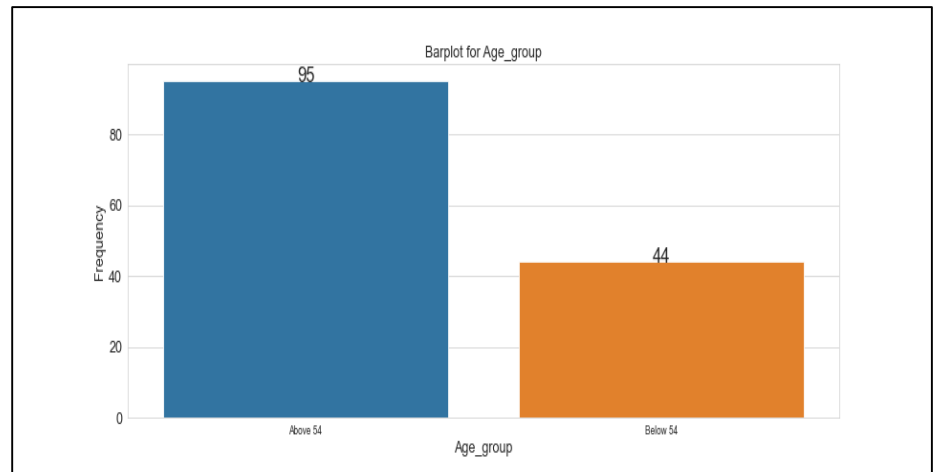
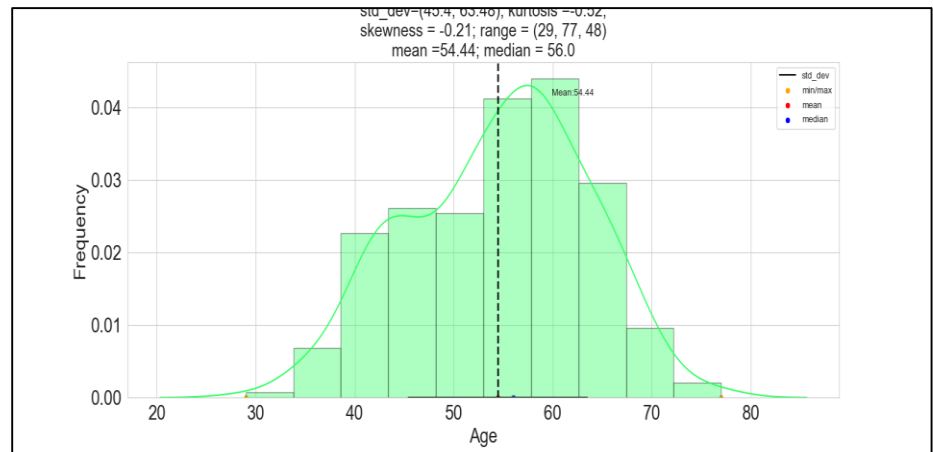
Heart Disease Data Analysis

Ashish Haldar
Ashishhaldar38@gmail.com

Observation from Heart Disease Exploratory data Analysis

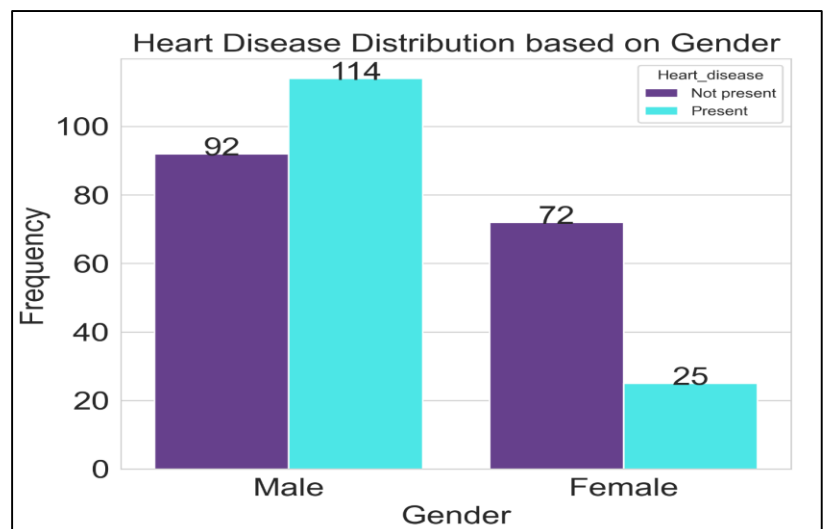
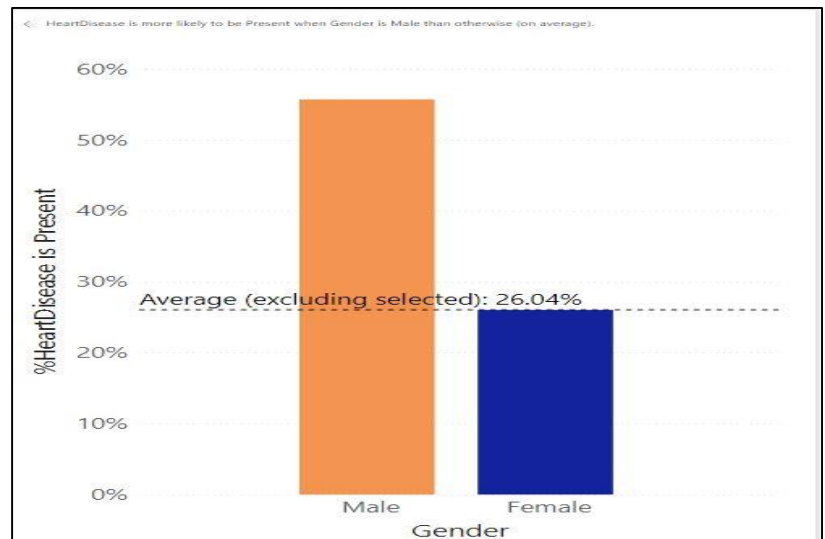
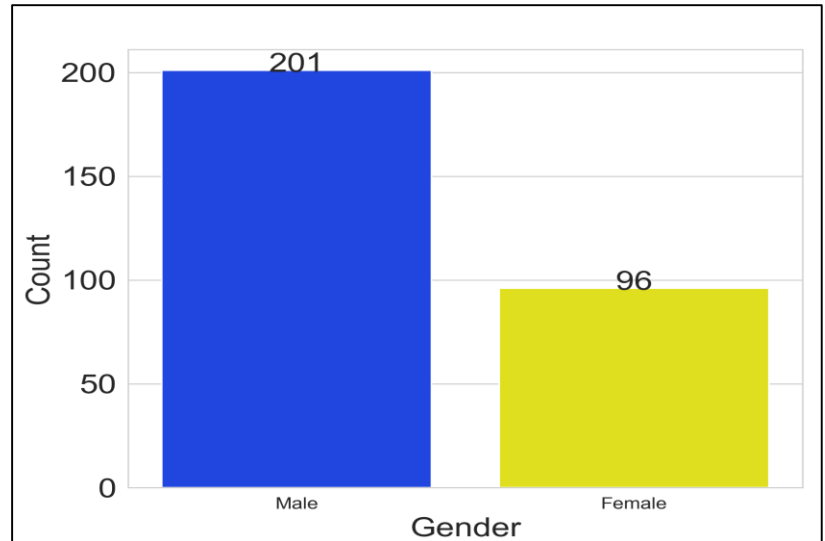
1. How population/patient for heart disease is distributed over Age?

- Mean Age is 54 years.
- Most of the people are in the age group of 45 to 63
- Skewness is -0.21 means age is left skewed i.e. people at young ages are most prevalent to heart disease.
- Kurtosis is -0.52; very less likely to have extreme/outlier values
- When age is more than 54 then heart disease is more likely to be present compared to all other age category
- **The distribution of age and the risk for the heart disease indicates that heart disease with the age ranging from 55 to 70 is having high risk of heart disease as the development of coronary fatty streaks starts in this age range.**
- We found Age group feature to be important which was indicated by Chi-square test of independence and t-test of independence.
- These observations conclude that as a person becomes older the risk of damaged & narrowing arteries also increases. It also weakens or thickens heart muscles that contributes to Ischemic heart disease & thus lead to heart attack.



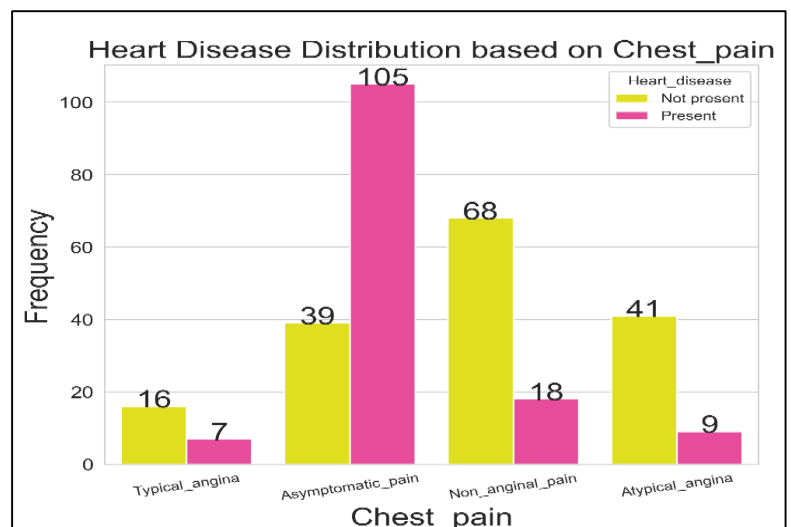
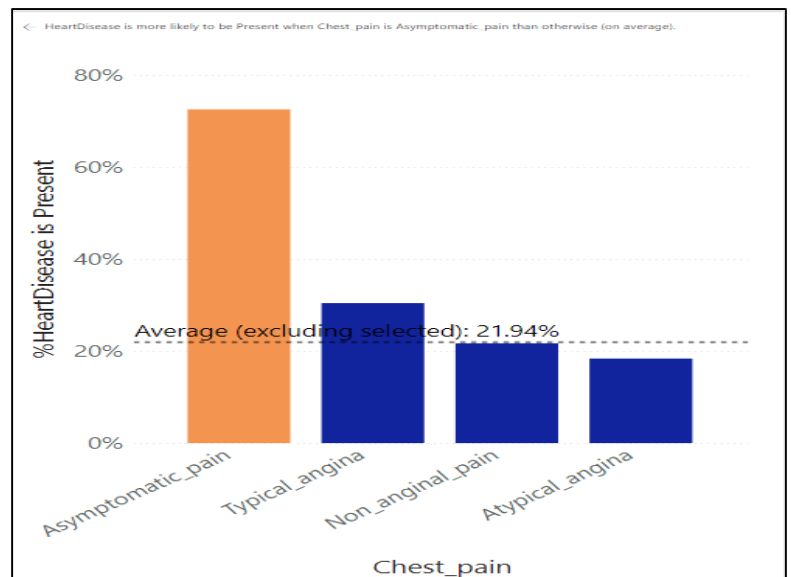
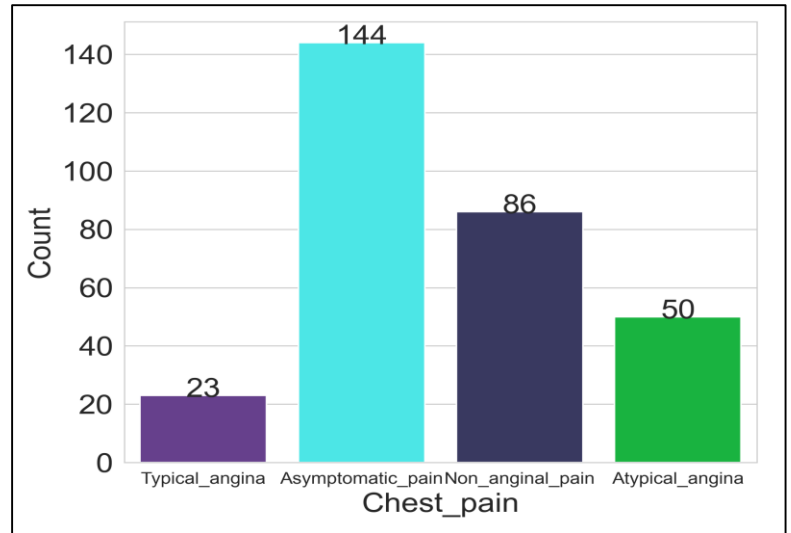
2. How population/patient for heart disease is distributed over Gender? Which gender is at greater risk?

- From all these figures we came to know that both men and women are affected by Heart disease but Males counterpart are most affected category.
- When gender is male then likelihood of heart disease present also increase than other.
- Gender is also an important attribute in contributing disease as chi-square's critical value is 23.22 which is very high.

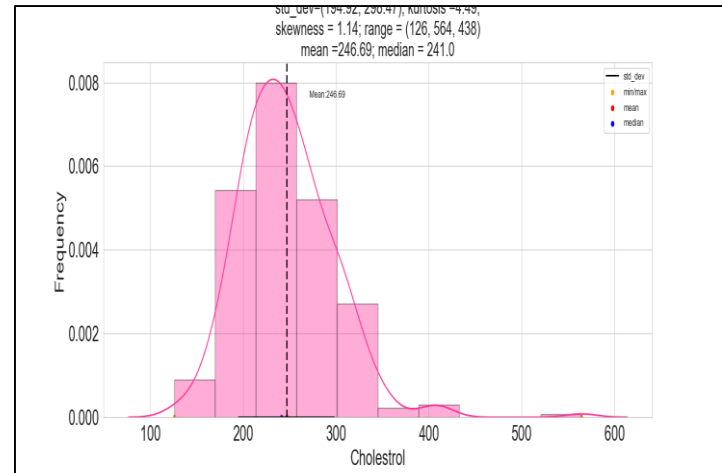
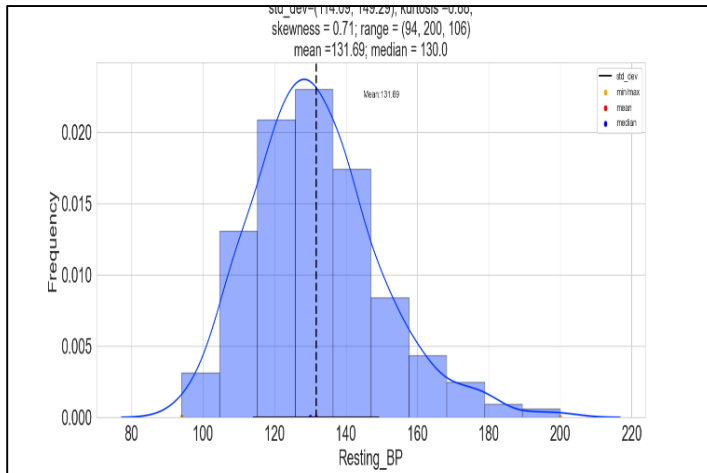


3. How Chest pain types is distributed? What chest pain type is causes heart disease more? Is chest pain an important indicator/factor in explaining the heart disease?

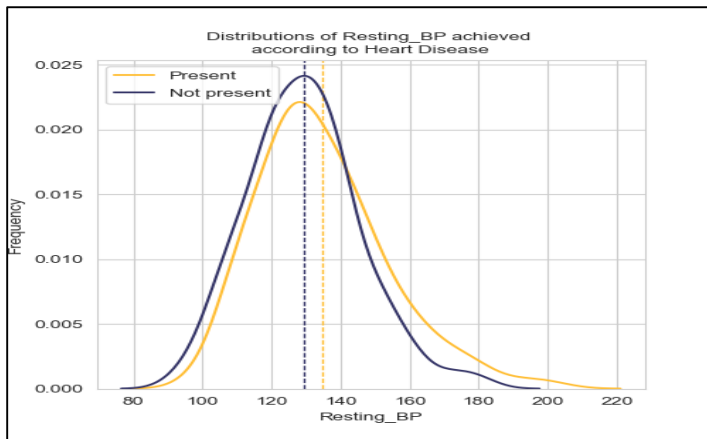
- Patients with Asymptomatic chest pain are more in number than patients with non- anginal and Atypical agina chest pain.
- When chest pain type is Asymptomatic then the likelihood of heart disease present increase than other pains.
- Yes, Chest pain is a very important variable which was indicated by high chi-square critical value.



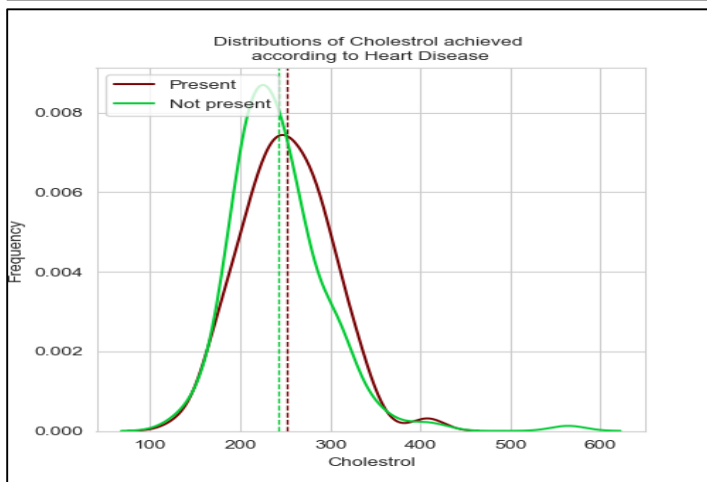
4. How are distributions of Resting blood pressure & cholesterol level distributed for Heart & Non-heart disease patients? Is there any statistical difference between average cholesterol level between heart & non-heart disease patients?



- Mean blood pressure and cholesterol is 131.69 and 246.69 respectively.
- Average BP indicates the **Stage 1 hypertension** and for cholesterol it is highly dangerous.

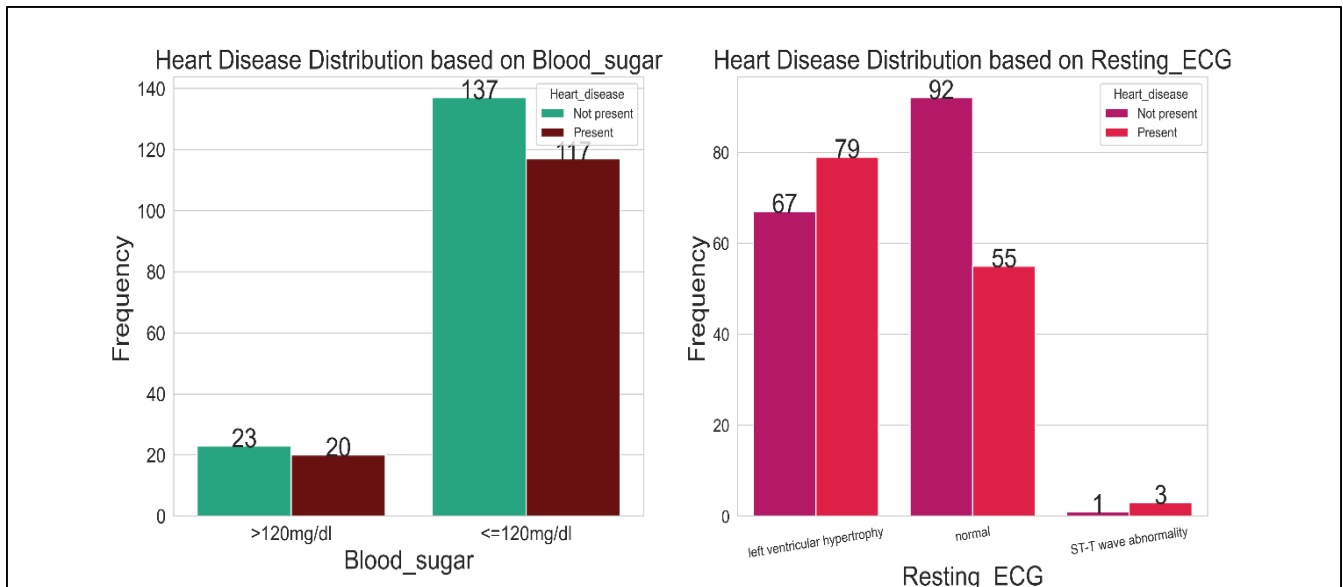


• But this variable is not significant indicated by chi-square test when we created the blood pressure categories. There is also statistical difference between Average cholesterol level between heart disease & non-heart disease patients. The ideal total cholesterol level should be less than 200 mg/dl but we observe more number of heart disease patients beyond 250 mg/dl level.



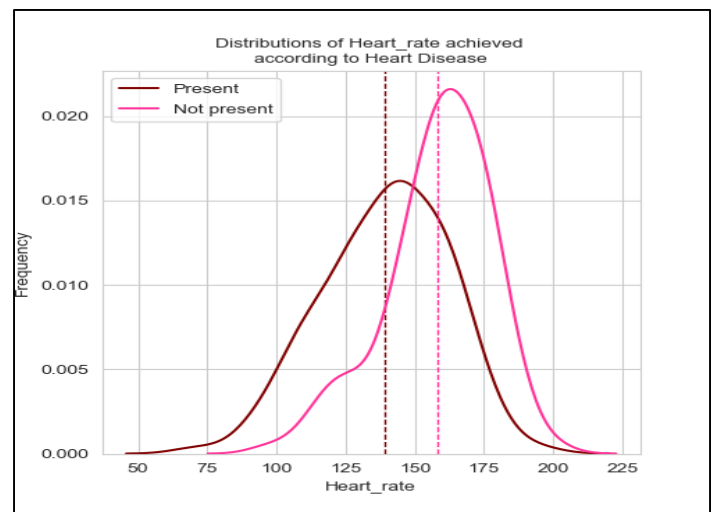
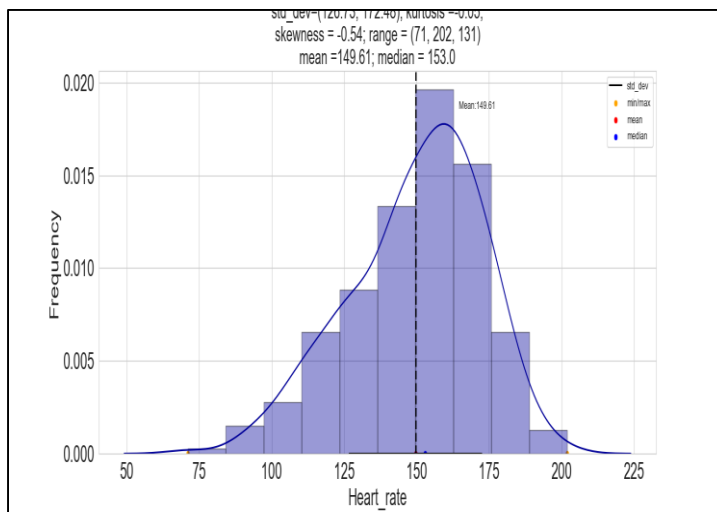
• However we do t-test on a sample of 50 we find that Resting_BP has different distribution as $p = 0.007$ is less than $\alpha = 0.05$ whereas Cholesterol has same distribution as $p = 0.159$ is greater than $\alpha 0.05$.

5. Either blood sugar or ECG affecting heart disease more likely? Are Blood sugar & Resting Electrocardiographic results important factors?



Neither Blood sugar nor Resting Electrocardiographic results variables are important since by chi-square test of independence these variables have no connection with Heart Disease.

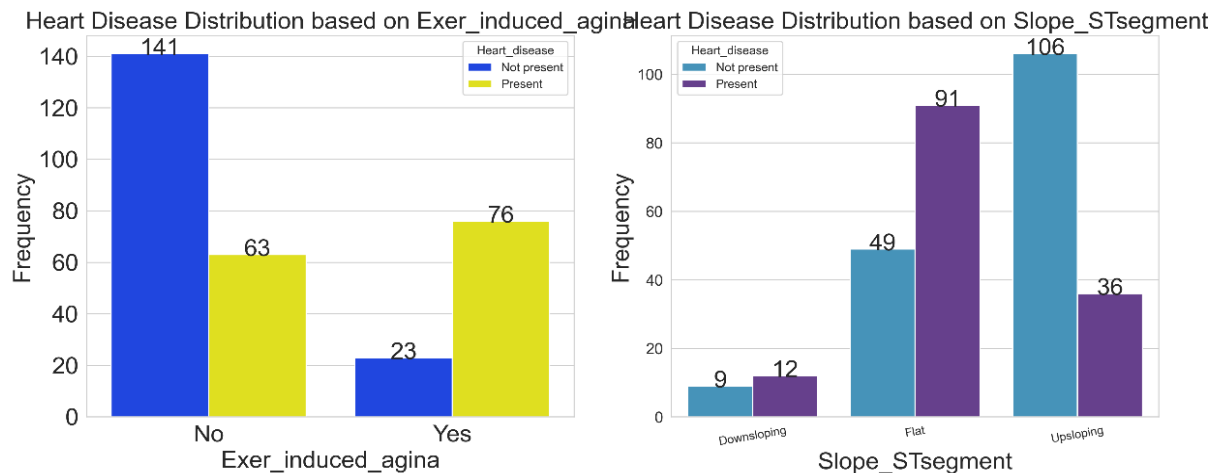
6. What insights we get for maximum heart rate when we do Univariate and Bivariate analysis? Is sample mean for maximum heart when disease is present different or same when disease is not present?



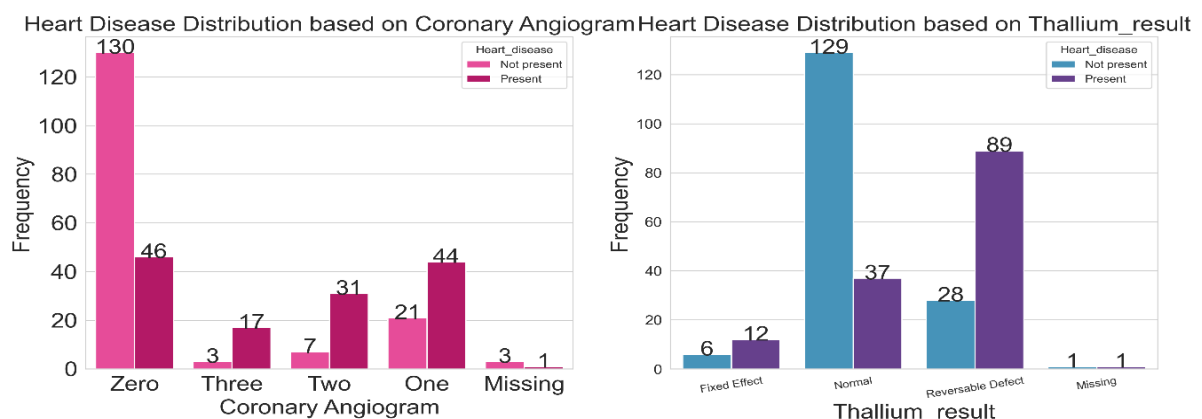
- Average heart rate is 149bpm
- minimum and maximum heart rate is 71 bpm and 202 bpm
- When we compared two sample mean then we find that both of the distribution i.e. heart rate when disease is present and not present is different.

7. What do Treadmill test result (indicated by Exercise induced angina) & Slope of ST segment say about Heart disease for the sample? Are they important factors?

More number of heart disease patients experience angina while doing treadmill test, whereas less number of heart disease patients are experiencing angina at rest. Since the chi-square critical value is coming very high, therefore, we conclude Exercise induced angina to be an important feature. Slowly upsloping ST segment usually indicates heart attack. Horizontal ST Segment depression is considerable abnormal response. Down sloping ST Segment depression represents severe heart attack. We observe more number of patients having abnormal Slope of the ST segment (indicated by Flat slope of the ST segment). Number of patients with down-sloping Slope of ST segment is considerably very less than Flat slope segment patients. This feature is also very important indicated by high chi-square critical value.



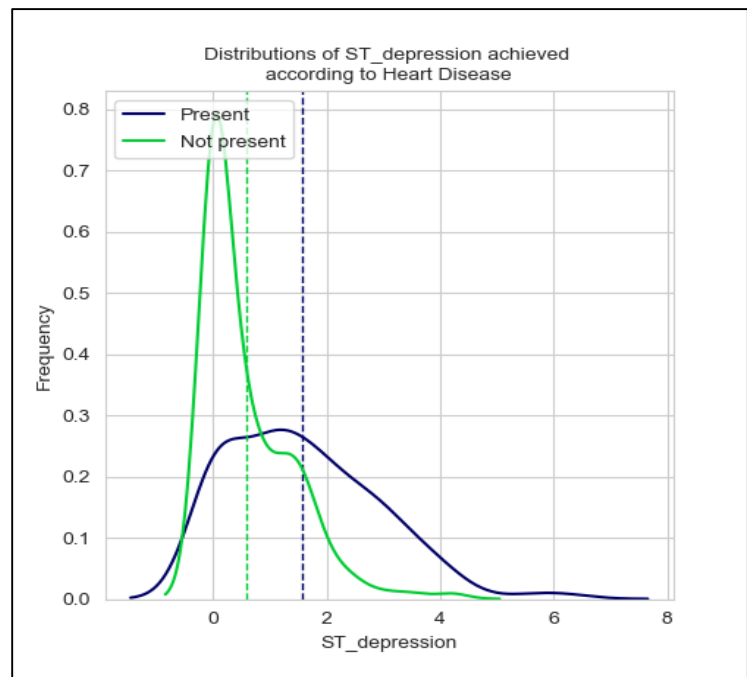
8. What do Coronary Angiogram results and Thallium test say about heart disease?



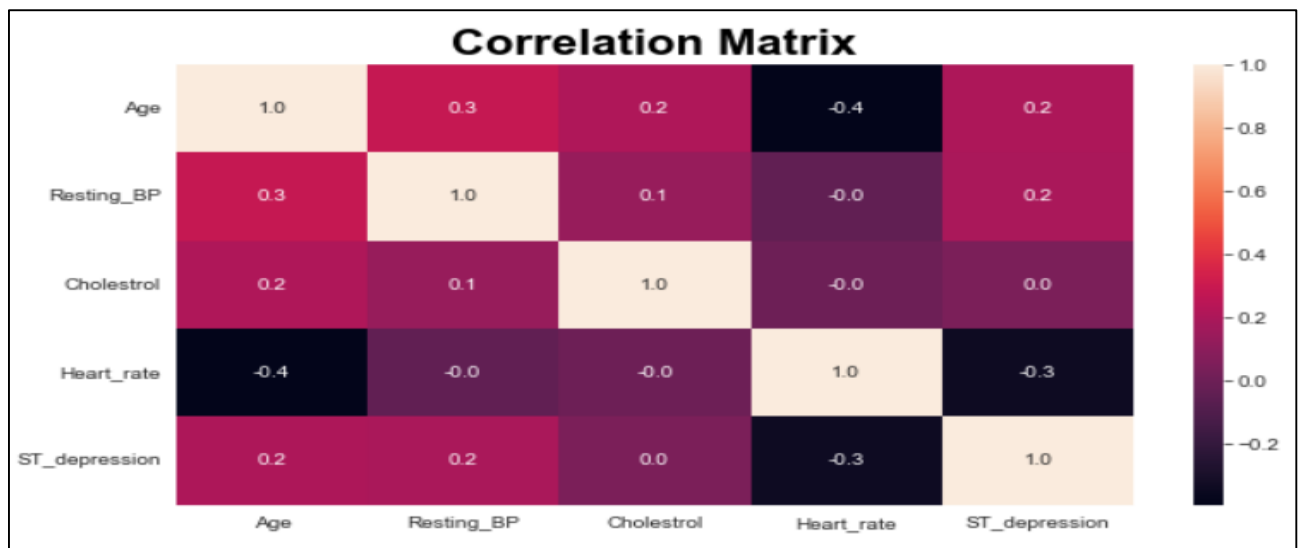
We notice that more number of non-heart disease patients with zero/no defect in their major vessels that supply blood, oxygen & nutrients to the heart than heart disease patients. But as the number of major vessels blocked increases we observe more number of heart diseases patients than with non-heart disease patients. These patients need an immediate attention (indicated by angiogram results). We observe a greater number of heart disease patients with reversible defect (indicated by Thallium test). And both features are important indicated by a very high chi-square critical value.

9. How ST_depression relative to rest look like for heart disease & non-heart disease patients?

- Mean value is 1.06 which suggest that ST_depression is normal.
- Skewness values suggest that it is highly right skewed.
- Kurtosis value is less than 3 which suggest that it has fewer outliers
- Both the averages are drawn from different distribution as t-test reject null hypothesis.



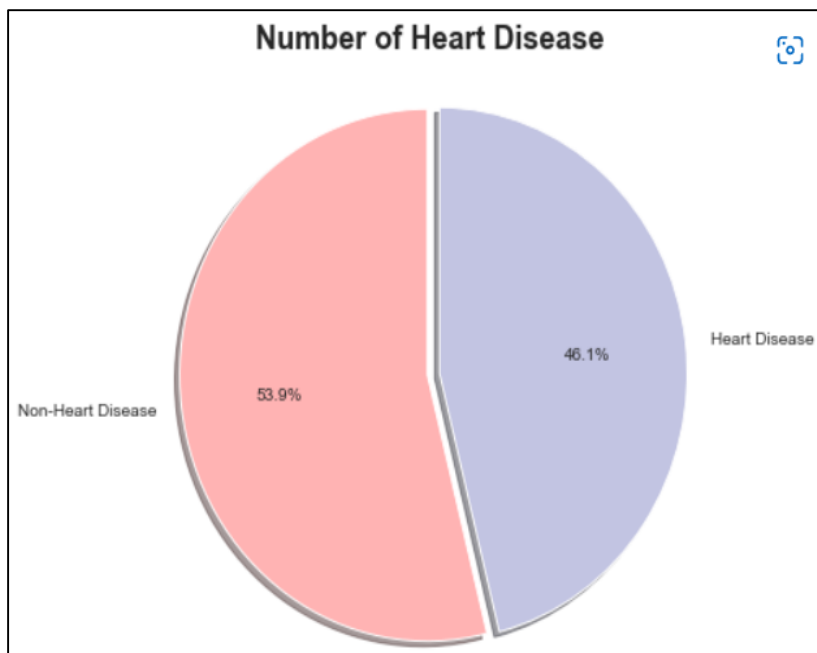
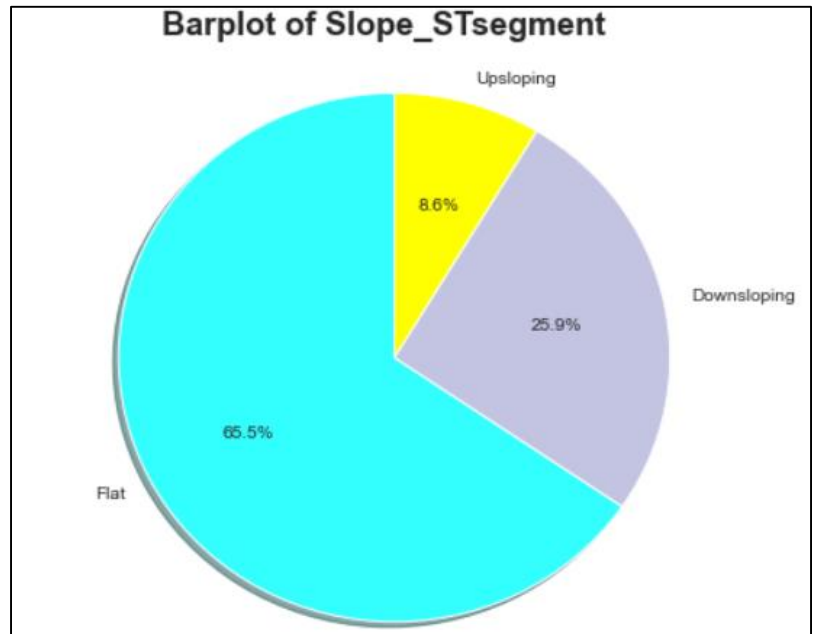
10. What is correlation between all the attributes?



As we can observe from the correlation matrix that there is very minor or no correlation between any of the attributes.

11. How Slope ST_segment is look like when heart disease is actually present? How total disease is divided into Present and Not Present?

This shows that when Slope ST_segment is Flat then there is high chance that the person has Heart disease compared to other.



Approximately 54% of the patients are Non- Heart disease patients and 46% are Heart disease patients.

Finding & Conclusion

From the above analysis it is referred that Age, maximum heart rate, Blood Pressure, cholesterol plays a vital role in influencing the heart disease more along with some important symptoms like **if a person has Asymptomatic pain, gender is male, thallium has reversible defect, Slope_Stsegmentis Flat and Exercise_inducedaginais yes then chances of a person being affected by disease is more compared to other symptoms/test.**

Heart Stroke and vascular disease are the major cause of disability and premature death. Chest Pain is the key to recognize the heart disease. In this work, the heart disease are predicted by considering major factors with four type of chest pain.

As people becomes older they need to maintain their blood pressure, cholesterol level, Heart rates and they should visit to a doctor as well to get check their health check. To avoid any heart disease they can do some of the following things like avoid smoking, do exercise, avoid high fat consumption diet and adopt low fat diet, eat raw green vegetables, maintain their stress level. In short, people should change their lifestyle & adopt healthy habits.