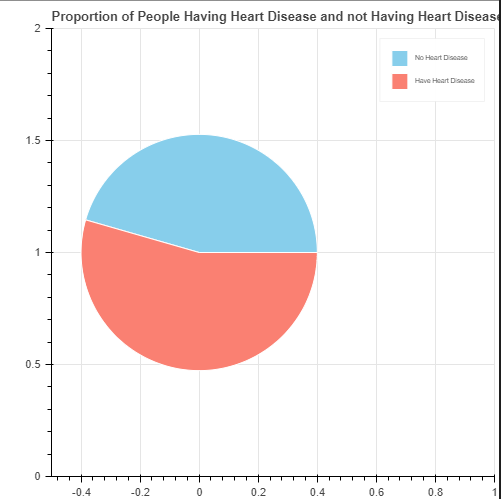
**CHAPTER 5 Conclusion**

**In this notebook we are performing Exploratory Data Analysis and use various Machine Learning Models to predict whether the patient has heart disease or not depending on the values of various features.** In this project by the help of various parameters about a patient, can we be predicting whether or not they have heart disease**. We have used Bokeh and a little bit of Seaborn to plot the graphs.** Let's have a look at what each of the terms used throughout this report means:

1. **Age**: Age of the person.
2. **Sex**: Sex of the person. (1 = male; 0 = female)
3. **Cp**: Chest Pain Type. It can take values of 0, 1, 2, 3.
4. **Trestbps**: Resting Blood Pressure (Measured in mm Hg on admission to the hospital). It can take continuous values from 94 to 200.
5. **Chol**: Serum Cholesterol in mg/dl. It also takes continuous values.
6. **Fbs**: Fasting Blood Sugar. It can take value of either 1 or 0.
7. **Restecg**: Resting Electrocardiographic Results. It can take value of 0, 1 or 2.
8. **Thalach**: Maximum Heart Rate achieved. It can take continuous value from 71 to 202.
9. **Exang**: Exercise Induced Angina. It can take value either of 0 or 1.
10. **Oldpeak**: ST depression induced by exercise relative to rest. It takes continuous decimal values.
11. **Slope**: the slope of the peak exercise ST segment. It can take value of either 0, 1 or 2.
12. **Ca**: Number of major vessels colored by fluoroscopy. It can take value of either 0, 1, 2, 3 or 4.
13. **Thal**: 3 = normal; 6 = fixed defect; 7 = reversable defect
14. **Target**: Indicates the presence or absence of heart disease. (= the predicted attribute)

**Output Graphs**

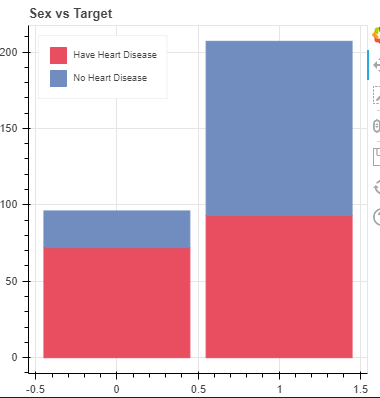


*Fig 5.1 Proportion of People Having Heart Disease and not Having Heart Disease*

Percentage of people having Heart Disease 54.0

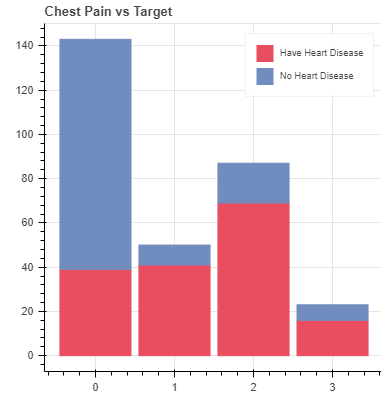
Percentage of people not having Heart Disease 46.0

We can see that the dataset is balanced as there is no major difference between the proportion of people having heart disease and those not having heart disease.



*Fig 5.2 Sex vs Target*

The above graph is for analyzing how much proportion of male or female have heart disease. We might think that more number of men have heart disease but if we observe closely, we can see that more proportion of female have heart disease as compared to men.



*Fig 5.3 Chest Pain vs Target*

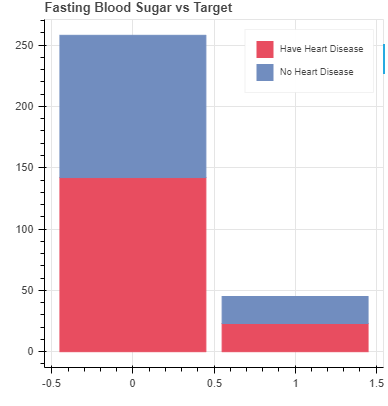
The above graph is for analyzing what proportion of different chest pain types patient have heart disease. It's really shocking to know that majority of the asymptomatic (Type 3) cases and Non-anginal pain patients (Type 2) ended up having heart disease. Different Chest Pain Types:

0: Typical angina: chest pain related decrease blood supply to the heart

1: Atypical angina: chest pain not related to heart

2: Non-anginal pain: typically esophageal spasms (not heart related)

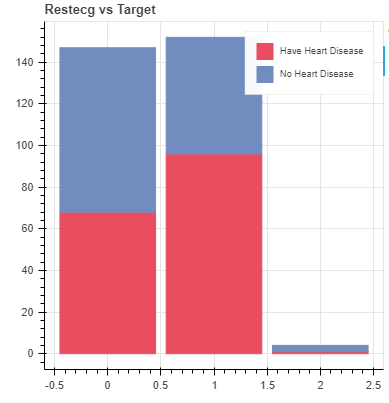
3: Asymptomatic: chest pain not showing signs of disease



*Fig 5.4 For analyzing how much proportion of diabetic and non-diabetic patients have heart*

*Disease*

The above graph is for analyzing how much proportion of diabetic and non-diabetic patients have heart disease. FBS > 120 mg/dl (1 = true; 0 = false). Those whose Fasting Blood Sugar is greater than 120 indicates that the patient is diabetic.

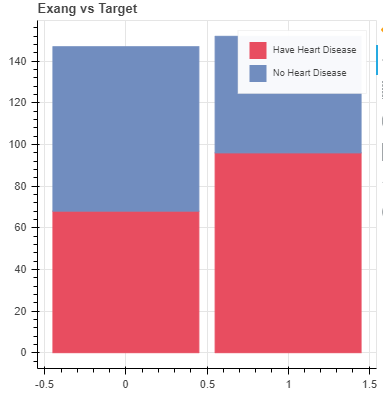


*Fig 5.5 Restecg vs Target*

A large proportion of people having restecg of type 1 actually have heart disease. We must take care of ST-T Wave abnormality as it can range from mild symptoms to severe problems.0: Nothing to note

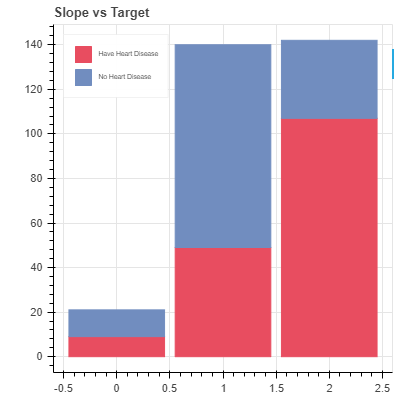
1: ST-T Wave abnormality can range from mild symptoms to severe problems signals non-normal heart beat

2: Possible or definite left ventricular hypertrophy. Enlarged heart's main pumping chamber



*Fig 5.6 Exercise induced Angina vs Target*

Exang means exercise induced angina (1 = yes; 0 = no). Angina is a type of chest pain caused by reduced blood flow to the heart. The above graph is for analyzing



*Fig 5.7 Slope vs target*

As type 2 means Downslopins which is a sign of unhealthy heart, most patients with type 2 slope had Heart Disease. slope - the slope of the peak exercise ST segment

0: Upsloping: better heart rate with exercise (uncommon)

1: Flatsloping: minimal change (typical healthy heart)

2: Downslopins: signs of unhealthy heart

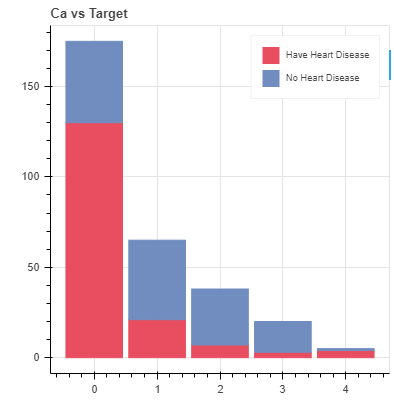


Fig 5.8 Ca vs target

ca - number of major vessels (0-3) colored by fluoroscopy

colored vessel means the doctor can see the blood passing through

the more blood movement the better (no clots)

We can see a large proportion of patients having 'ca' value of type 0 and type 4 had Heart

Disease.

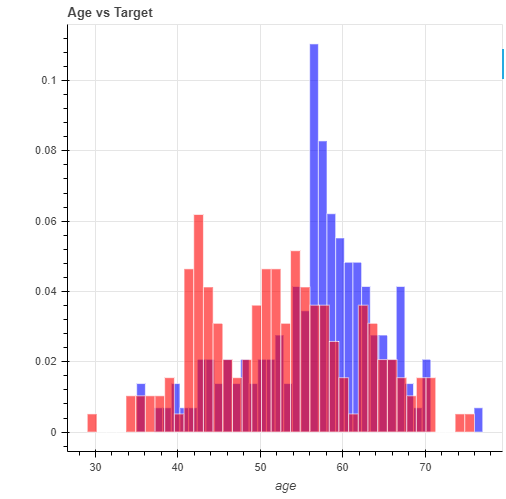


Fig 5.9 Age vs target

According to above graph there is no particular age at which the person is more prone to having heart disease, which proves that age is just a number.

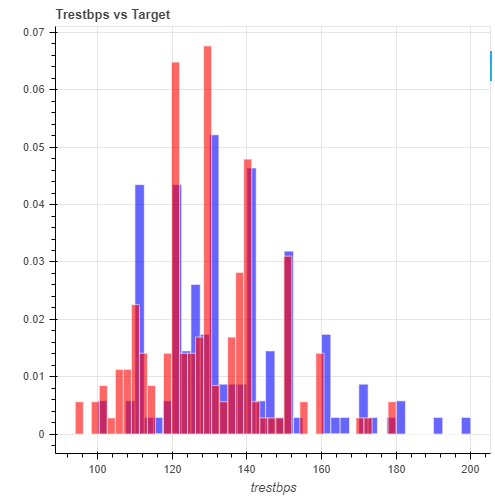


Fig 5.10 Tresbps vs target

As above graph shows testing Blood Pressure (in mm Hg on admission to the hospital) anything above 130-140 is typically cause for concern. Those patients having Blood Pressure in the range of 120 to 160 have the highest chance of having heart disease

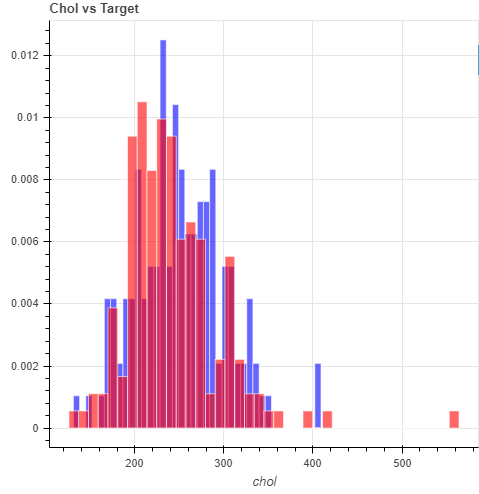


Fig 5.11 Chol vs target

As we can see in above graph patient having Cholesterol level greater than 200 had heart disease.

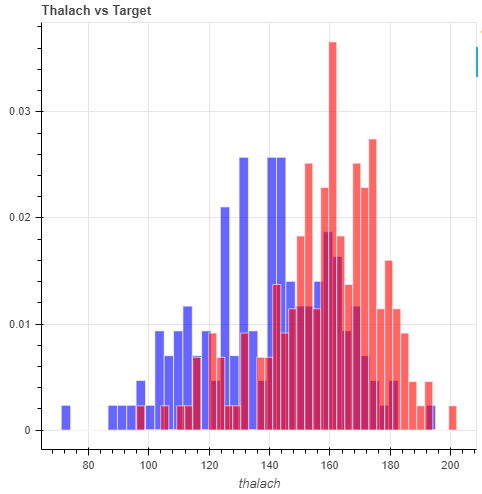


Fig 5.12 Thalach vs target

As the graph shows the patients having maximum heart rate greater than 150 are at a greater risk of having heart disease.

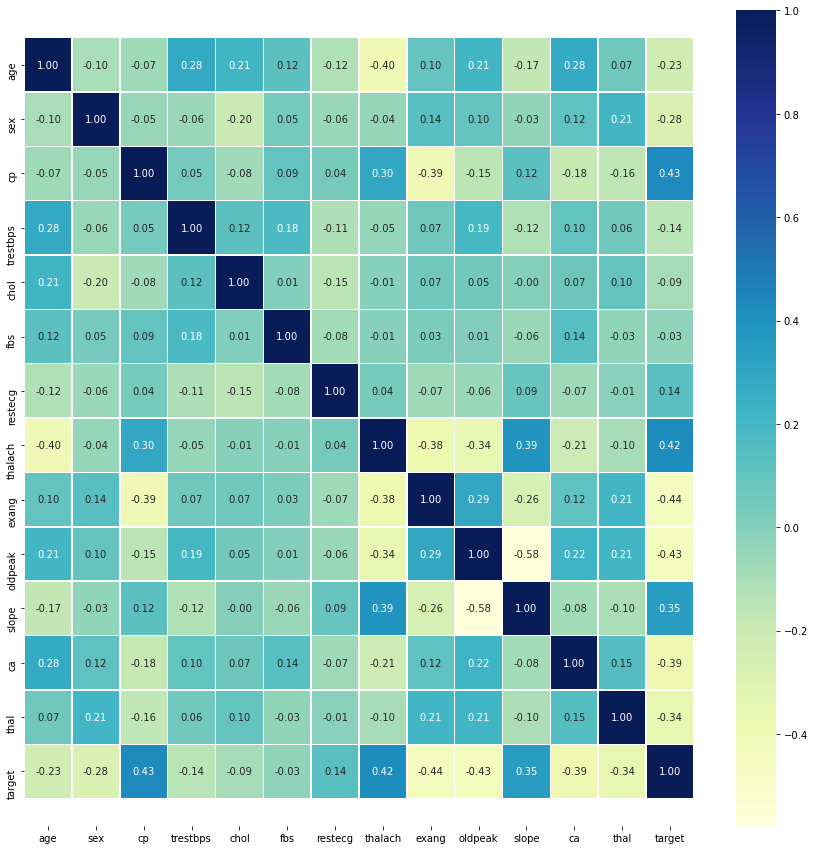
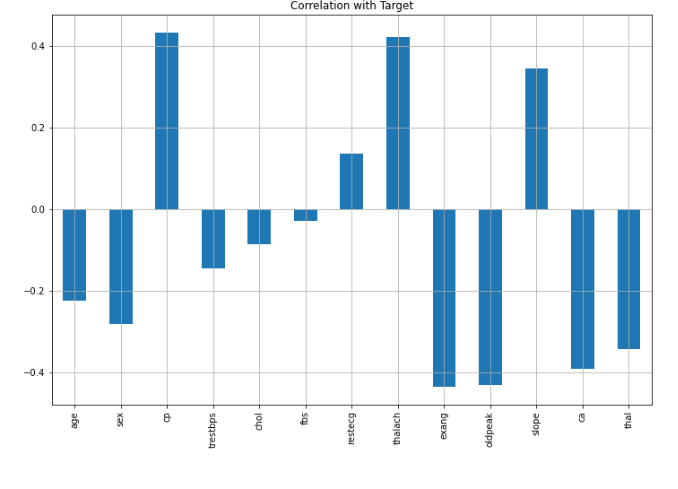


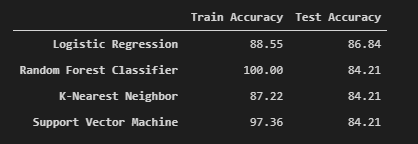
Fig 5.13 Correlation Matrix

We have used seaborn for plotting correlation matrix as its much faster and much easier than bokeh. The correlation between the features with target is not that clear in the correlation matrix as there are a large number of features, lets visualize it in another way.



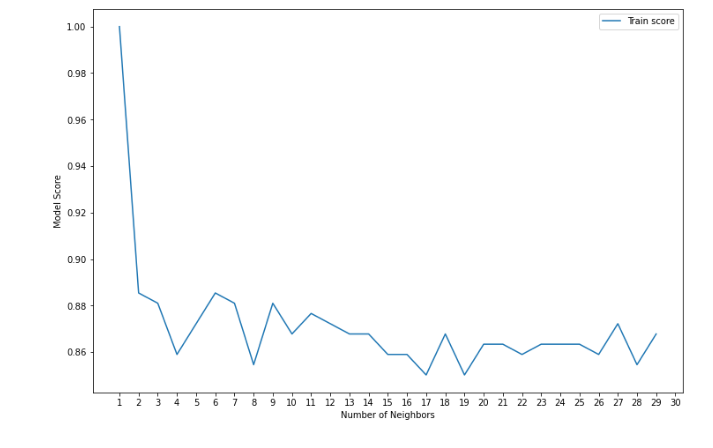
*Fig 5.14 Correlation with Target*

We can see that 'fbs' and 'chol' are least related with 'target' whereas other features are highly correlated with the 'target' variable.



*Fig 5.15 Summary of train and test accuracies of different classifiers*

The above figure shows the train accuracies and test accuracies that we are planning to use in our project.



*Fig 5.16 Result of K-nn*

*Maximum KNN score on the train data: 100.00%*