Correlation Between Heart Activity, Certain Biomedical Traits and Heart Disease

Data Exploration

Heart Disease is a wide spread disease and is a leading cause of death, however it is preventable if detected in the very early stages before any damage has been done to the heart vessels. In the exploration of this data we will try to find any correlation between the biomedical traits of the individuals in relation to heart disease. For clarification the terms of the data is listed below:

1. age
2. sex
3. chest pain type (0,1,2,3)
  -- Value 0: asymptomatic
  -- Value 1: atypical angina
  -- Value 2: non-anginal pain
  -- Value 3: typical angina
4. resting blood pressure
5. serum cholesterol in mg/dl
6. fasting blood sugar > 120 mg/dl (fbs > 120 mg/dl = 1, fbs < 120 mg/dl = 0)
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9. exercise induced angina (1 = yes; 0 = no)
10. old-peak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment -
  -- 0: down-sloping;
  -- 1: flat;
  -- 2: up-sloping
12. number of major vessels (0-3) colored by fluoroscopy
13. thalassemia: 1 = normal; 2 = fixed defect; 3 = reversible defect
14. target: 0 = heart disease; 1 = no heart disease

There were a total of 10 charts made from the data comparing various medical traits. the first 4 charts are scatter plots where 3 variables could be determined the x variable, y variable and the color difference represented the presence or not of heart disease. While the last 6 charts were made form categorical data where the presence of heart disease was used to classify the other variable.

Pair Plot Table
Resting blood pressure VS Cholesterol
In the first chart (Scatter plot) there is a comparison between resting blood pressure and cholesterol, there is no correlation between resting blood pressure and cholesterol also there is no correlation between these two variables as it relates to heart disease increases.

Cholesterol VS Age
The next chart we are see a   slight positive correlation between the age and the cholesterol is also visible that as the older the individuals get the more likely they were to have heart disease. Hence there is a slight increase in the cholesterol levels as the individuals age.

Maximum heart rate achieved VS Age
There is a negative correlation between maximum heart rate and age, so as people age we can see that the heart rate slows down significantly, we can also see by that the number or yellow colored markers, heart disease is more prevalent in individuals as they age. And hence lower heart rate seems to be more prevalent with individuals with heart disease than not.

**Maximum heart rate achieved VS Resting blood pressure**
There is a no correlation between maximum heart rate and resting blood pressure, but as 0 signifies heart disease we see that at all levels of resting blood pressure there are cases of heart disease without any trend however as maximum heart rate increases we see a gentle decrease in heart disease cases. So we can again say that lower heart rate can be associated to heart disease.
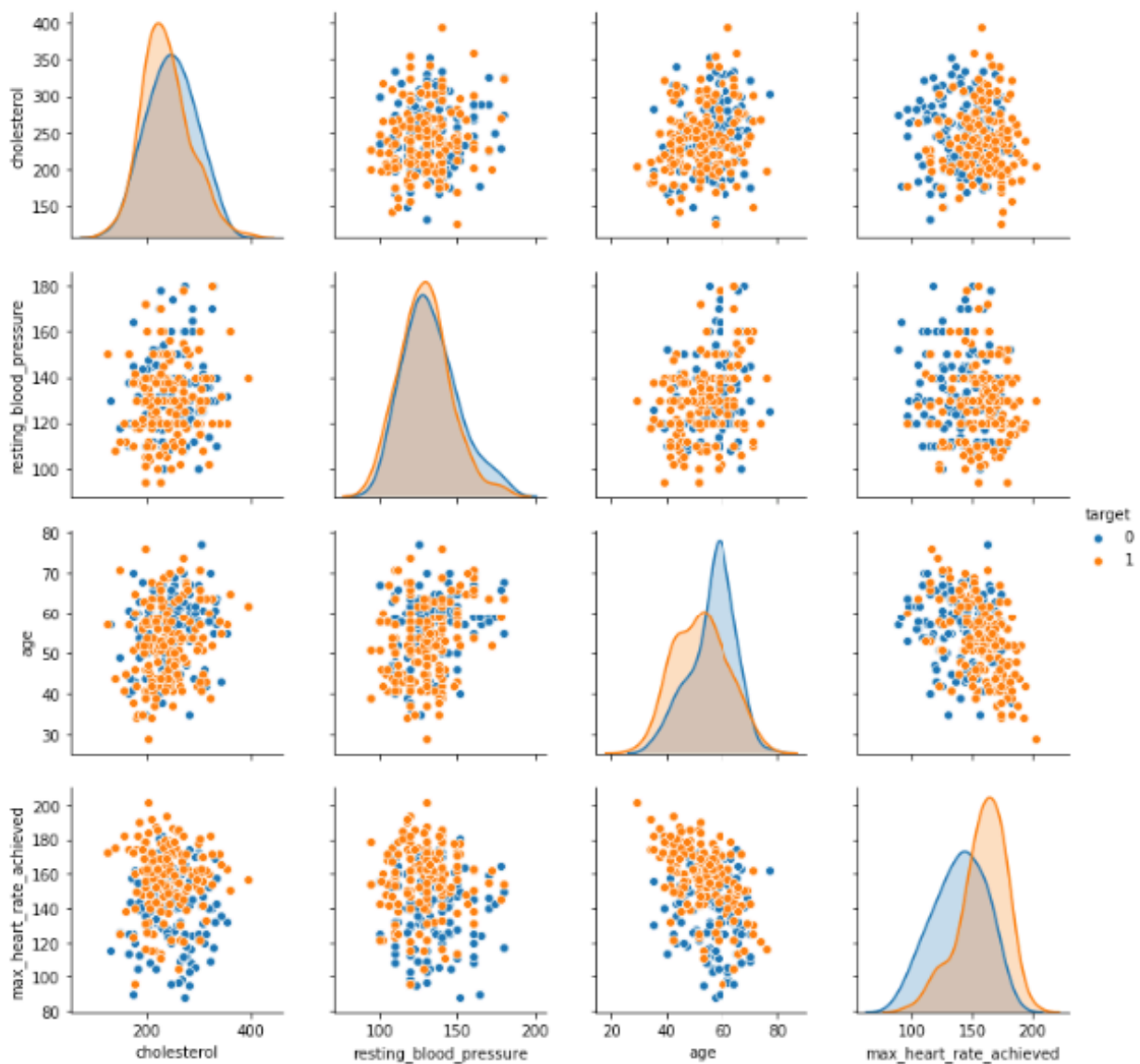
Pie chart Fasting blood sugar Percentage levels
There was a significant number of people had their fasting blood sugar below 120 mg/dl and out of those there was about 56.2% had fasting blood sugar below 120 mg/dl had heart disease, and about 50% of the people with blood sugar above 120 mg/dl had heart disease so we can see about the same percentage of people had heart disease in these two cases so from this chart there doesn't seem to be any correlation between fasting blood sugar and heart disease.
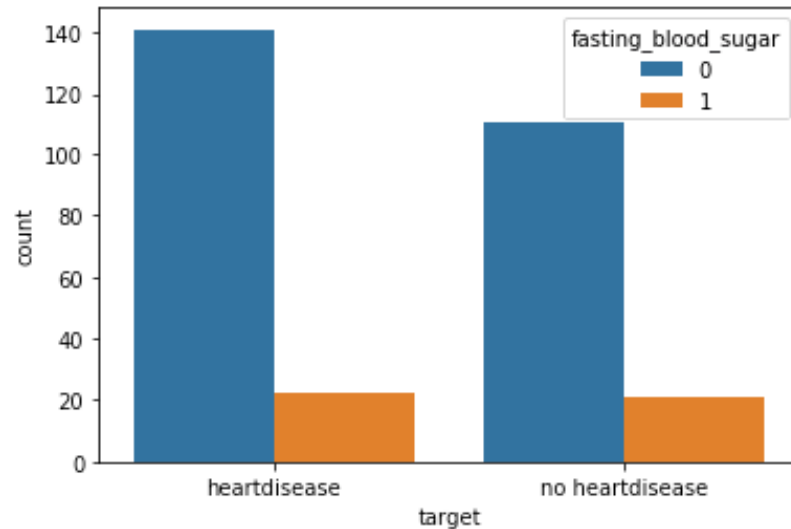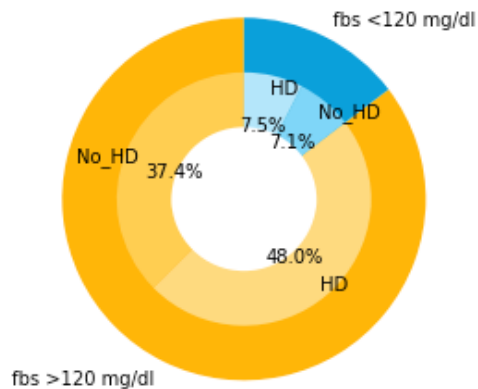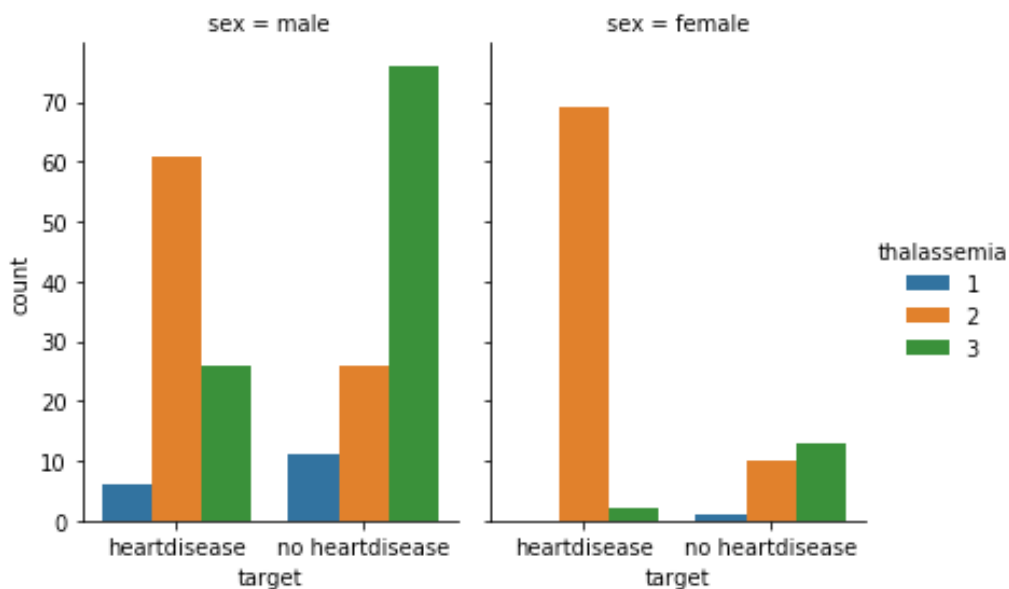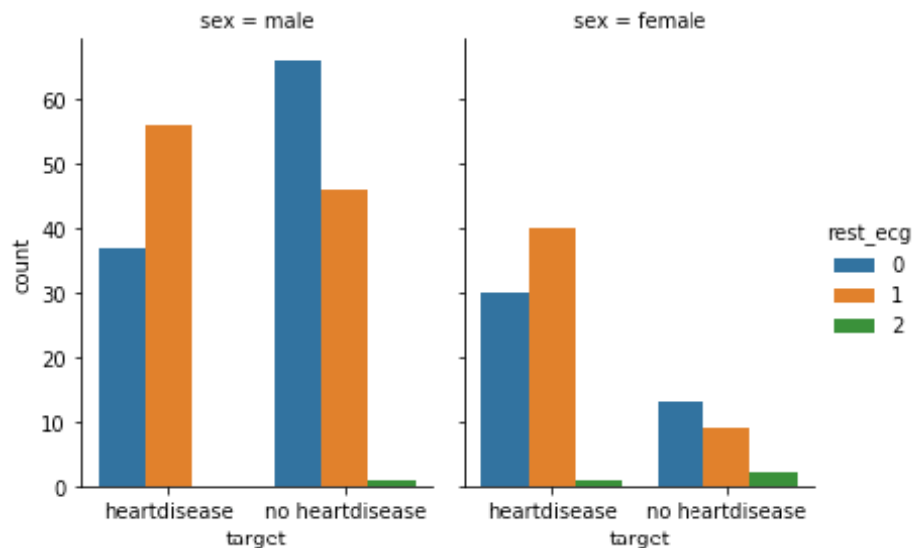

Bar chart
comparing the types of thalassemia with presence of heart disease
There are 2 types of thalassemia which are either reversible which can be treated and corrected and the fixed type that comes from gene mutation and is not reversible. From the charts we can see that most people that do not have heart disease also do not thalassemia while approximately 64% of the people with heart disease had the reversible form of thalassemia which is a large percentage of people and from this chart we can suggest a correlation between reversible thalassemia and heart disease.

Bar chart comparing Resting electrocardiographic results with presence of heart disease
There is no significant difference in results for the resting electrocardiograph so we can say there is no significant correlation directly between resting electrocardiograph results and heart disease.

Bar chart Comparing Exercise induced agina Percentage levels with presence of heart disease
So we can see that out of tested patients the number of people with heart disease that have exercise induced agina is approximately 77% and the number of patients that have heart disease without exercise induced agina is approximately 30% there fore we can say that from these charts we can assume that there is a higher chance that someone with exercise induced agina to develop or have heart disease.
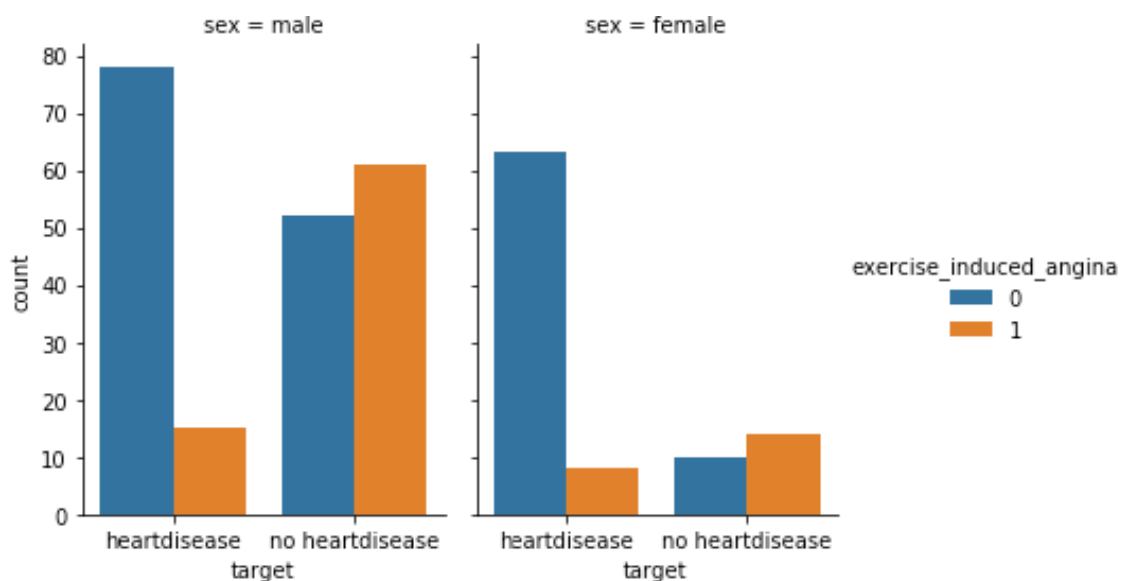
Bar chart comparing chest pain and heart disease
From the results of this graph the most prevalent type of chest pain in the individuals with heart disease was asymptomatic chest pain, this was found in high number, though there was also a sizable number of people without heart disease that also have that type of heart pain so having asymptomatic chest pain may be an indication of heart disease.

Bar chart comparing st_slope and heart disease
The predictions from the chart suggest that people without heart disease have up-sloping ST segment while those with heart disease have a high number of flat ST segment, suggesting a correlation between flat ST segment and heart disease.
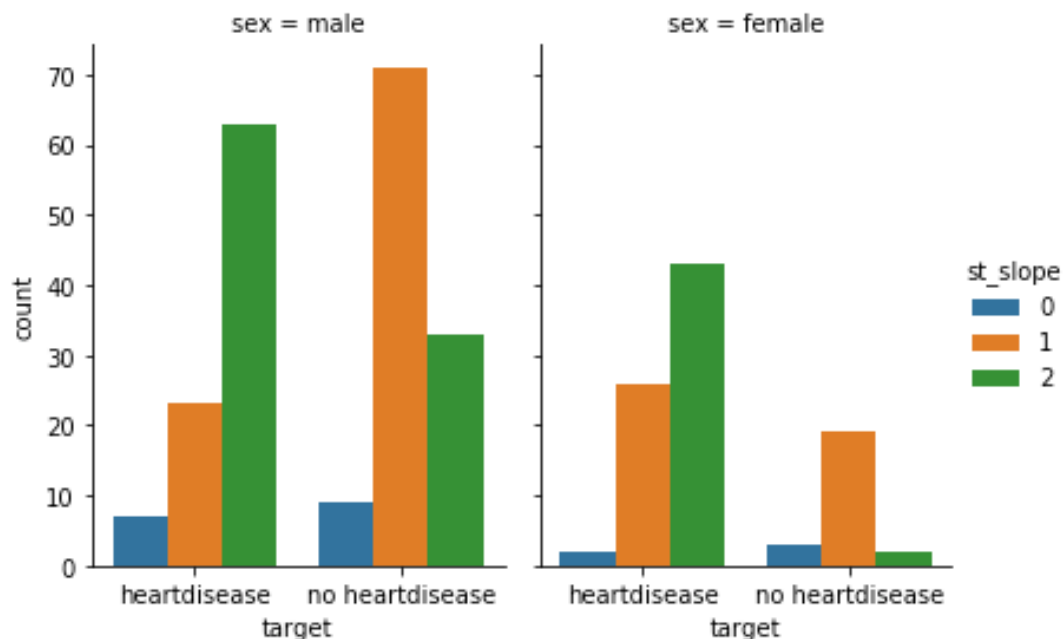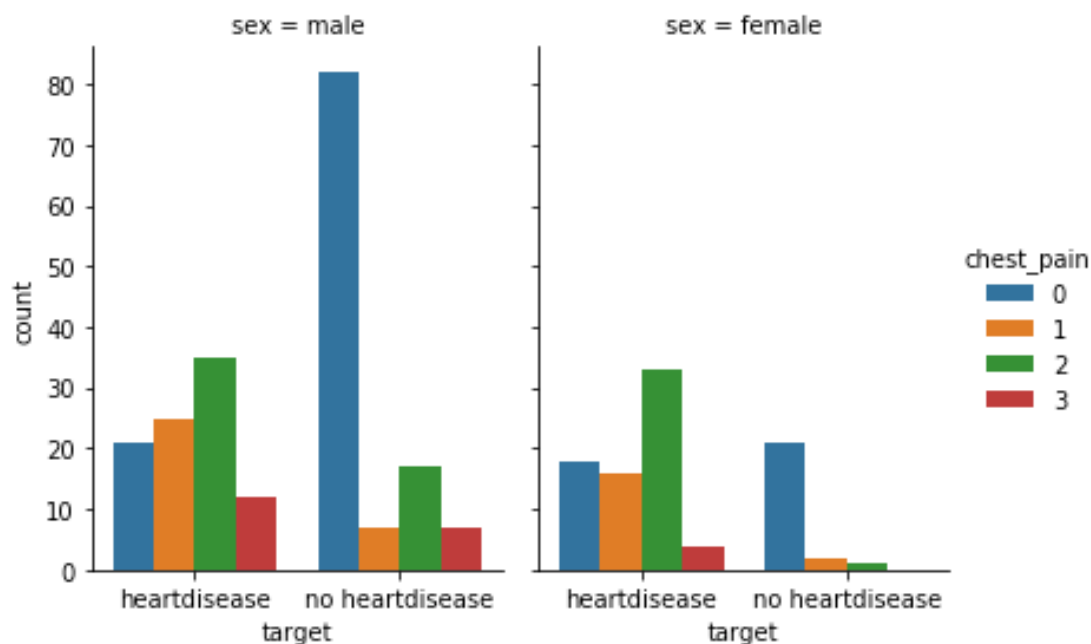
Summary

Out of the 14 attributes selected we are trying to determine which of these attribute exhibit any behaviors that can help us predict the presence of heart disease. By using simple charts we will try to find any surface level correlation to heart disease and from our first four analysis. Out of these four characteristics: resting blood pressure, cholesterol, maximum heart rate and age, In plot 4 Comparing maximum heart rate and resting blood pressure, heart disease patients are found spread out through all ranges of resting blood pressure but on maximum heart rate axis they concentrated on the lower region where the maximum heart rates are lower, similarly looking at plot 3 where maximum heart rate is compared to age, as a person gets older we see that their maximum heart rate decreases also we see that heart disease patients increase, showing again correlation between heart disease and maximum heart rate. In plot 1 resting blood pressure to cholesterol there is no correlation in this scatter plot because the patients with the heart disease is evenly mixed in with those without heart disease on both axes. Hence for the individual data from our surface level graph observation a low maximum heart rate is a strong variable in heart disease detection.

As we look into the categorical data from plot 6 to plot 9, there are six attributes that we make observations of. From the graphs we see that there seems to be high correlation between heart disease and the following variables: exercise induced agina (plot 6), reversible form of thalassemia in (plot 7), asymptomatic chest pain (plot 9) and a flat st_slope (plot 10). While fasting blood sugar (plot 5) and resting electrocardiograph (plot 8) does not show strong correlation to heart disease. So in conclusion from the results of this surface level overview of the data there are five attributes that if are found in a patient that predicts a high probability of heart disease which are low maximum heart rate, exercise induced agina, reversible form of thalassemia, asymptomatic chest pain and a flat st_slope.