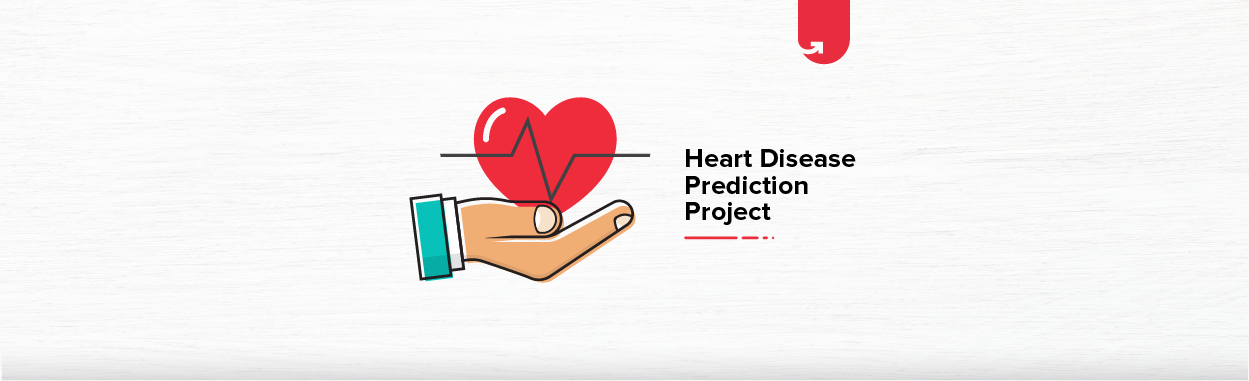
DATA SCIENCE PROJECT ON

HEART DISEASE PREDICTION



PROJECT ID : PRCP-1016-heart-disease-pred

BUISENESS CASE :

BASE ON GIVEN FEATURE AND MESUREMENT PREDICT WHETHER PATIENT WILL HAVE HEART DISEASE OR NOT.

Task:

BINARY CLASSIFICATION TASK

INTRODUCTION OF PROJECT:

Cardiovascular disease or heart disease is the leading cause of death amongst women and men and amongst most racial/ethnic groups in the United States. Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease. From the CDC, roughly every 1 in 4 deaths each year are due to heart disease. The WHO states that human life style is the main reason behind this heart problem. Apart from this there are many key factors which warns that the person may/may not getting chance of heart disease.

The term heart disease is often used interchangeably with the term cardiovascular disease. Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke.

DOMAIN ANALYSIS

TARGET COLUMN == HEART DISEASE PRESENT

In this project we are going to analyze that how other feature of dataset affecting heart disease

1.PATIENT ID:

Id of particular patient, Id is used to identify a patient, this is a unique column so that its not affect to heart disease

2.SLOPE OF PEAK EXERCISE ST SEGMENT:

While a high ST depression is considered normal & healthy. The “ slope ” hue, refers to the peak exercise ST segment, with values: 1: upsloping, 2: flat, 3: down-sloping). Both positive & negative heart disease patients exhibit equal distributions of the 3 slope categories.

3.THAL:

A blood disorder called thalassemia,[normal, reversible defect, fixed defect]

4.RESTING BLOOD PRESSURE:

blood pressure tells a lot about your general health. High blood pressure or hypertension can lead to several heart related issues and other medical conditions. Uncontrolled high blood pressure can lead to stroke.

5.CHEST PAIN TYPE:

Most of the chest pain causes are not dangerous to health, but some are serious, while the least cases are life-threatening.[TA: typical angina(1), ATA: Atypical angina(2), NAP: non-anginal pain(3), ASY: asymptomatic (4) ]

6.NUM OF MAJOR VESSELS:

Major Blood Vessels of the Heart: Blood exits the right ventricle through the pulmonary trunk artery. Approximately two inches superior to the base of the heart, this vessel branches into the left and right pulmonary arteries, which transport blood into the lungs.[number of major vessels: 0 to 3]

7.FASTING BLOOD SUGAR:

Your Fasting blood sugar level of 120 is a High Fasting blood sugar level. If your Fasting blood sugar is in between 74 mg/dL and 99 mg/dL, then you need not worry as 74-99 mg/dL is the normal range for Fasting blood sugar. But if your Fasting blood sugar is lesser or greater than the above values, then there may be some problem in your body.

(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

8.RESTING EKG/ECG RESULT:

The electrocardiogram (ECG or EKG) is a test that measures the heart’s electrical activity, and a resting ECG is administered when the patient is at rest. It involves noninvasive recording with adhesive skin electrodes placed on specially prepared spots on the skin, and it plots out the heart's activity on a graph. It is used to determine the health of the heart and circulatory system and to help diagnose issues with associated body systems.[0: normal, 1:having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), 2:showing probable or definite left ventricular hypertrophy by Estes’ criteria]

9.SERUM CHOLESTEROL:

A person’s serum cholesterol level represents the amount of total cholesterol in their blood. A person’s serum cholesterol level comprises the amount of high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides in the blood. Triglycerides are a type of fat bundled with cholesterol.

10. oldpeak\_eq\_st\_depression:

oldpeak = ST depression induced by exercise relative to rest, a measure of abnormality in electrocardiograms

11.SEX:

sex (1 = male; 0 = female)

12.AGE:

age in years

13.MAX HEART RATE:

It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.[Average heart rate: 60 to 100 bpm]

14.EXERCISE INDUCED ANGINA:

Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood.[0: no, 1: yes]

15.HEART DISEASE PRESENT:

[0: no heart disease present, 1: heart disease present]

EXPLOTARY DATA ANALYSIS

UNIVARIATE ANALYSIS

OBSEVATION:

1.SOP:

1. The slope of peak exercise st segment with 3 unique observation, [1:upsloping, 2:flat, 3:downsloping]

2. 52% patient has upsloping peak of exercise st segment, 42% patient has flat slope of peak exercise st segment and remaning 7% patient has downsloping peak of exercise st segment.

2.THAL:

1. In thal 3 unique categories [normal,reversible defect,fixed defect]

2. In 98 patient normal blood disorder, 74 patient reversible blood disorder and remaing 8 patient fixed blood disorder

3.RESTING BLOOD PRESSURE:

1. The blood pressure range between 94 to 180 with 131.3 average

2. Most of the patient has 130 resting blood pressure

4.CHEST PAIN TYPE:

1. Four type of chest pain [1: typical angina(TA), 2: Atypical angina(ATA), 3: non-anginal pain(NAP), 4: asymtomatic(ASY) ]

2. Most number of (46%) patient has asymtomatic chest pain and (32%) patient has non-anginal pain and remaning (16% & 7%) patient has atypical angina nas typical angina chest pain

5.MAJOR VESSELS:

1. The range of major vessels between 0 to 3

2. Most (106)number of major vessels is 0, and remaining major vessels is [37,23,14]

6.FASTING BLOOD SUGAR:

1. Two categories in fasting blood sugar (1 = true, 0 = false)

2. 151 patient fasting blood sugar is less than 120 mg/dl

3. 29 patient fasting blood sugar is greter than 120 mg/dl

7.RESTING EKG/ECG RESULT:

1. In resting ekg/ecg result 3 uqiue value(0,1,2)

2. The most (52%)number of ekg result is 2, as well as 0 (normal) ekg result is 47%

8.SERUM CHOLESTEROL:

1. The range of serum cholesterol between 126 to 564

2. The Most frequent value of serum cholesterol is(239 & 204)

9.OLD PEAK ST DEPRESSION:

1. The range of old peak st depression between 0 to 6.20

10.SEX == [0:Female, 1:male]:

1. Most (124) number of patient is male and 56 patient is female

11.AGE:

1. The age range between 29 to 77 year

2. The average age of patient is 54.8 year

12. MAX HEART RATE:

1. The most obesrvation of heart rate between 140 to 180

2. maximum heart rate is 202 and minimum heart rate is 96

13.EXERCISE INDUCED ANGINA: [0: no, 1: yes]

1. 123 patient has no chest pain and reamining 57 patient chest pain

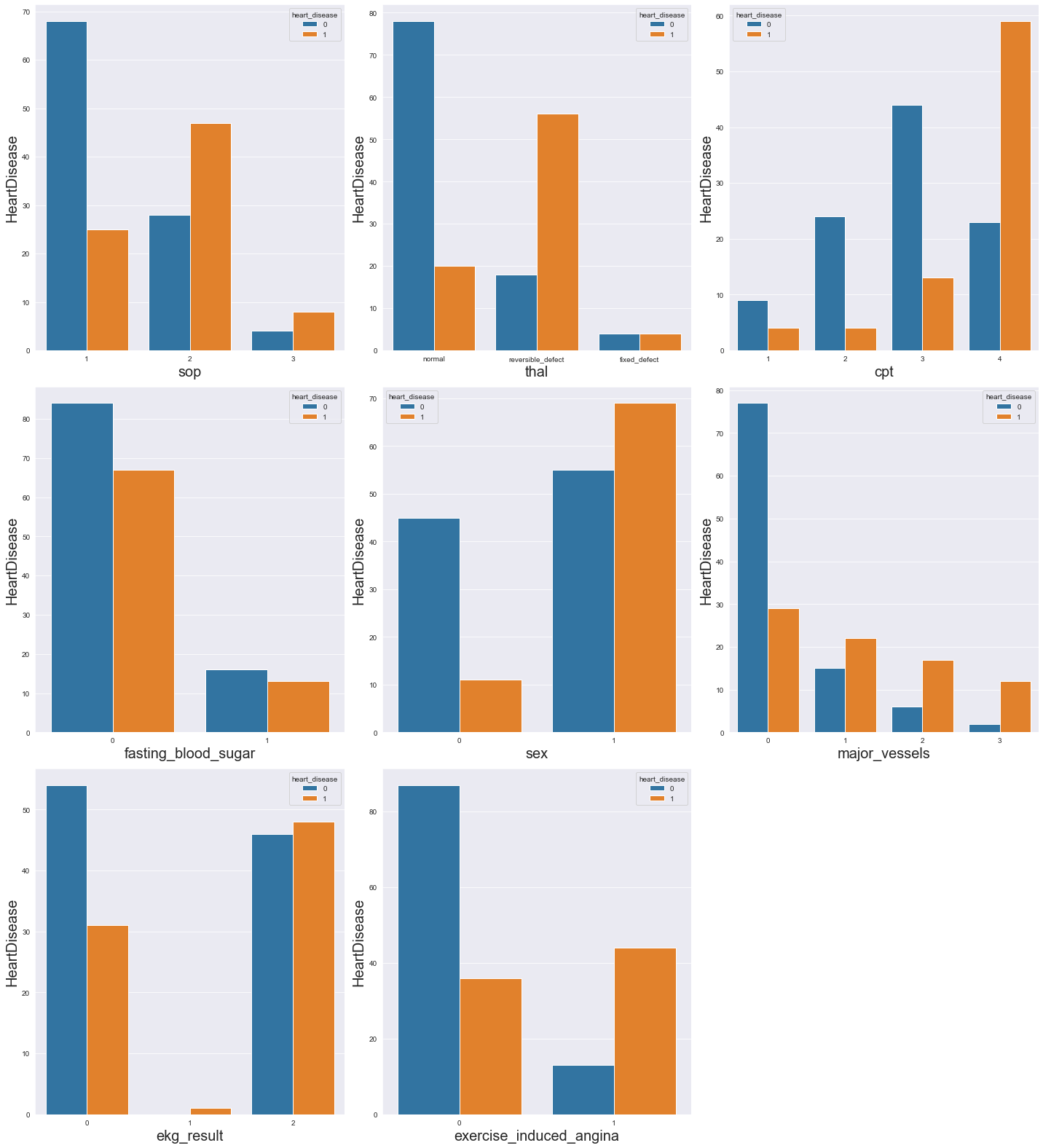
14. HEART DISEASE == TARGET VERIABLE

[0:No heart disease, 1:heart disease]

1. 100 patient has no heart disease and 80 patient has heart disease

BIVARIATE ANALYSIS:

ANALYSIS ON CATEGORICAL VERIABLE WITH RESPECT TO TARGET VERIABLE(HEART DISEASE)



OBSERVATION:

1.IMPACT OF SOP TO HEART DISEASE:

1. In this table clearly seen the slope of peak st segment is upsloping the chance of heart disease is less than other.

2. If slop pf peak is flat the chance of heart disease in more than upsloping

3. downslope st segment patient is also chance to get heart diseaase

4. with the follwing observation we can say that slope of peak st segment is fullt impact to heart disease

2.IMPACT OF THAL TO HEART DISEASE:

1. Normal blood disorder patient has less chance of heart disease than other thal

2. reversible defect blood disorder has more chance of heart disease and fixed defect blood disorder has 50-50 chance of heart disease

3.IMPACT OF CPT TO HEART DISEASE:

1. if the patient have asymtomatic(4) chest paint the chance of heart disease is more high

2. non-anginal pain(3),typical angina(1), Atypical angina(2) chest pain has less chances of heart disease.

3. but all chest pain types are impacted to heart disease.

4.IMPACT OF FASTING BLOOD SUGAR TO HEART DISEASE:

1. If fasting blood suagar is less than 120mg/dl the chance of heart disease is high.

2. and fasting blood sugar is greter than 120mg/dl the chace of heart disease is slightly less.

5.IMPACT OF SEX TO HEART DISEASE:

Male patient has more chance of heart disease than female

6.IMPACT OF MAJOR VESSELS TO HEART DISEASE:

1. If the major vessels is zero the chance of heart disease is less but zero major vessels are also chance of heart disease

2. 1,2, and 3 major vessels are more(high) chance of heart disease

7.IMPACT OF EKG RESULT TO HEART DISEASE:

1. If the ekg/ecg result is normal(0) the chance of heart disease is less.

2. If ekg/ecg result is 1 the 100% patient has heart disease

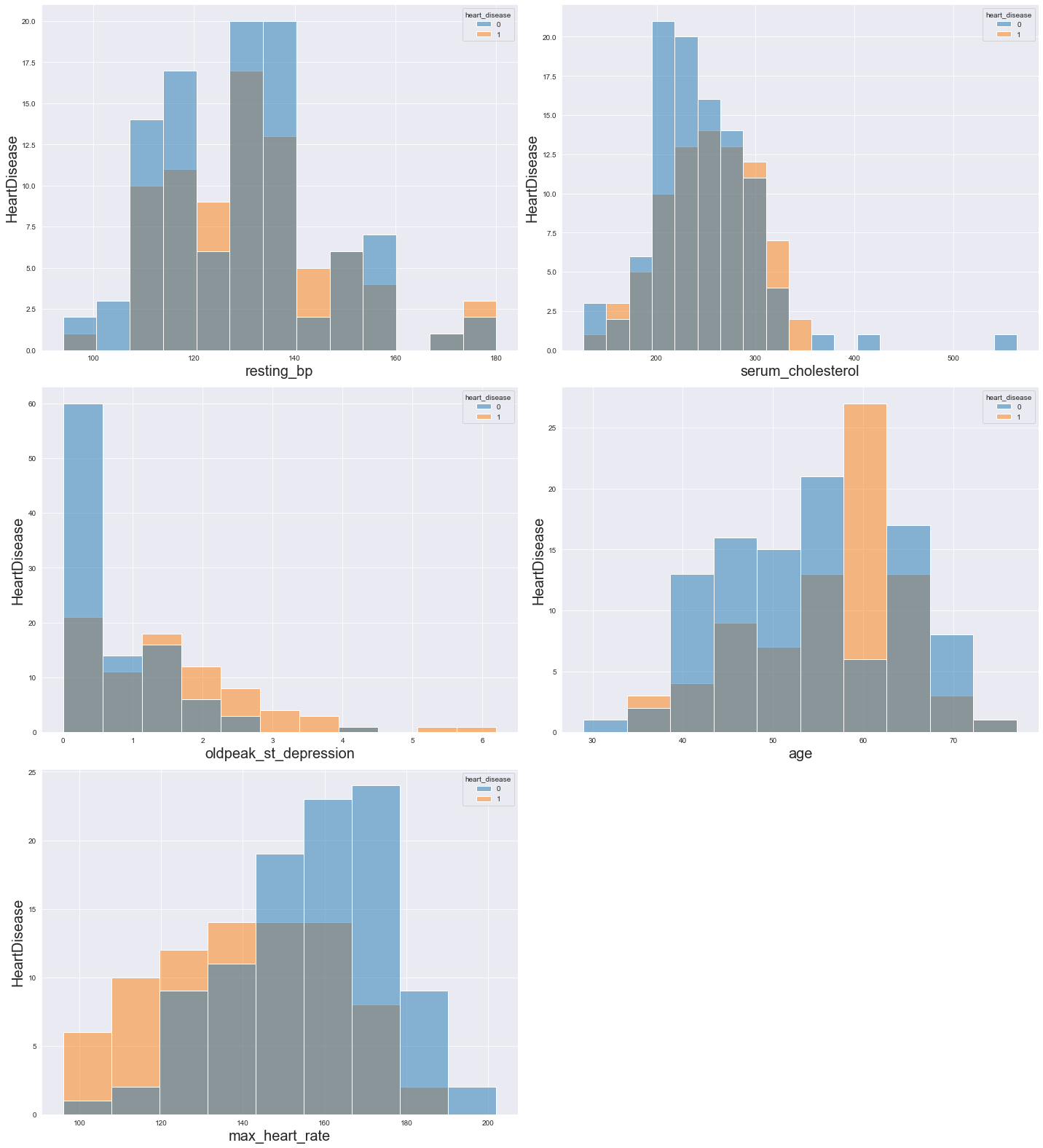
3. 2 ekg/ecg result is 50-50% chance of heart disease

8.IMPACT OF EXERCISE INDUCED ANGINA:

1. If the patient has no chest pain the chance of heart disease is less

2. If patient has chest pain the chance of heart disease is more

ANALYSIS ON NUMERICAL VERIABLE WITH RESPECT TO TARGET VERIBLE (HEART DISEASE)



OBSERVATION:

1.IMPACT OF RESTING BP TO HEART DISEASE:

1. If the blood pressure range between 110 to 150 the chance of heart disease is more

2. If resting blood pressure is low the chance of heart disease is slightly less

2.IMPACT OF SERUM CHOLESTREOL TO HEART DISEASE:

1. If serum cholestreol is less than 350 the heart disease chance is 50-50 percent.

2. serum cholestreol is more than 350 thier is no chance of heart disease

3.IMPACT OF OLD PEAK DEPRESSION TO HEART DISEASE:

1. If old peak depression is less the chance of heart disease is less

2. old peak depression is more than 1 the chance of heart disease is more

4.IMAPCT OF AGE TO HEART DISEASE:

1. At the age of 60 the more chance of heart disease and age range between 40 to 70 heart disease chance is 50-50 percent

2. If age is less than 30 their is no chance of heart disease

5.IMAPCT OF MAX HEART RATE TO HEART DISEASE:

1. If the heart rate is less than 140 the chance of heart disease is more

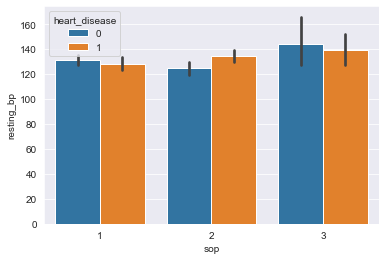
2. Above 140 heart rate chance of heart disease is 50-50 percent.

3. If the heart rate is more than 180 their is no chance of heart disease

MULTIVARIATE ANALYSIS

CHECK RELATION OF TWO VERIABLE WITH RESPECT TO TARGET VERIBLE (HEART DISEASE)

1. RELATION BETWEEN SOP AND RESTING BLOOD PRESSURE WITH RESPECT TO TARGET VERIABLE(HEART DISEASE)

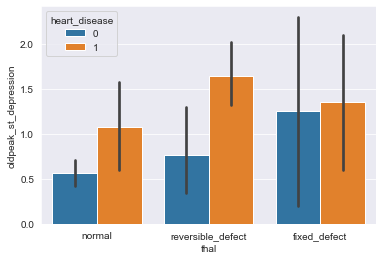


OBSERVATION:

In this plot clearly seen that if the resting blood presure is increases the chance of heart disease is equal

**OBSERVATION:**

1. **RELATION BETWEEN THAL AND OLD PEAK ST SEGMENT WITH RESPECT TO TARGET VERIABLE(HEART DISEASE)**

****

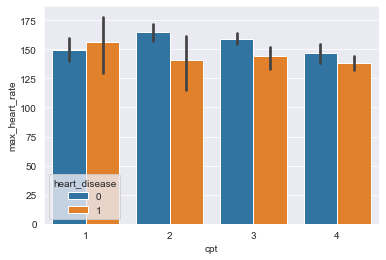
**OBSERVATION:**

**1. reversible defect blood disorder and high oldpeak st depression the chance of heart disease is more**

**2. fixed defect blood disorder and above 1 oldpeak st depression then the chance of heart disease is equal**

**3. If thal is normal and old peak st depression is more than 1 the chance of heart disease is more**

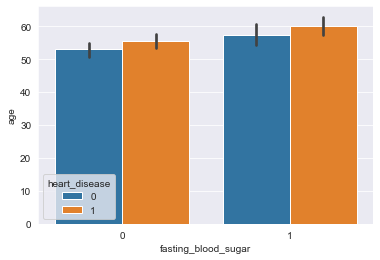
1. RELATION BETWEEN CPT AND MAX HEART RATE WITH RESPECT TO TARGET VERIABLE(HEART DISEASE)



OBSERVATION:

In this plot cleary seen that the all types of chest pain and below 150 heart rate are chnace of heart disease is equal

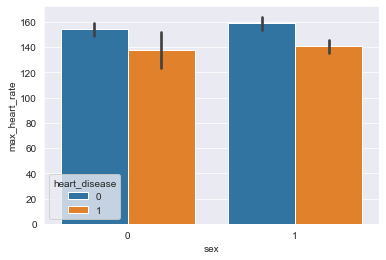
1. RELATION BETWEEN FASTING BLOOD SUGAR AND AGE WITH RESPECT TO TARGET VERIABLE HEART DISEASE



OBSERVATION:

If the fasting blood sugar is less than 120mg/dl & greter than 120mg/dl with age 0 to 60 the chance of heart disease is 50-50 percent

1. RELATION BETWEEN SEX AND MAX HEART RATE WITH RESPECT TO TARGET VERIBALE(HEART DISEASE)

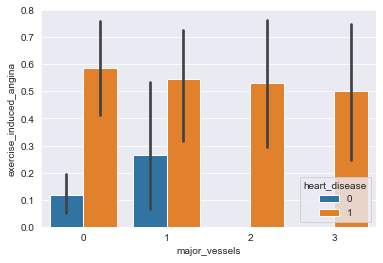


OBSERVATION:

1. Maximum heart rate is less chance of heart disease in male and female.

2. Minimum heart rate is more chance of heart disease in male and female.

1. RELATION BETWEEN MAJOR VESSELS AND EXCERCISE INDUCED ANGINA WITH RESPECT TO TARGET VERIBALE(HEART DISEASE)

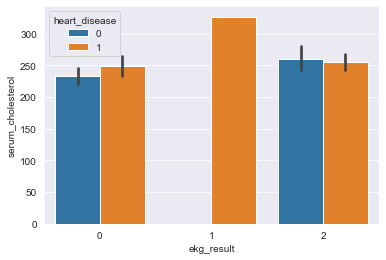


OBSERVATION:

1. If patient exercise induced angina is minimumn the chance of heart disease in less in 0 and 1 major vessels.

2. Maximum exercise induced angina is more chances of heart disease with all major vassels(0,1,2,3)

1. RELATION BETWEEN EKG RESULT AND SERUM CHOLESTEROL WITH RESPECT TO TARGET VERIBALE(HEART DISEASE)



OBSERVATION:

1. 1 ekg/ecg result with all serum cholesterol the 100% chance of heart disease

2. If ekg/ecg result is (0,2) with all serum cholesterol the chance of heart disease is equal.