Heart Disease Analysis

Presented by
Govarthini G S

Project overview

Reason for choosing the topic

- Heart disease is one of the leading causes of death globally.
- Analyzing data related to heart disease can help in understanding its patterns, risk factors, and potential preventive measures, thereby having a significant impact on public health.
- Wanted to analyze who got more affected by heart disease and what are the factors

Problem Statement : The primary objective is to analyse the likelihood of heart disease in individuals based on various health and demographic factors.

Hypotheses

Individuals(specially male) over the age of 50 who have high cholesterol levels, blood pressure are more likely to develop heart disease compared to those who have normal levels

About the dataset

- Age
- Sex
- Chest pain type(4 values)

  Value 0: typical angina,**Value 1:** atypical angina,,**Value 2:**non-anginal pain,,**Value 3:**: asymptomatic

- Trestbps: resting blood pressure(in mm Hg on admission to the hospital)
- Chol: serum cholesterol in mg/dl
- Fbs: (fasting blood sugar> 120 mg/dl) (1= True; 0 = false)
- Restecg: resting electrocardiographic results

  Value 0 : Normal,,**Value 1:** : having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) ,,**Value 2:** showing probable or definite left ventricular hypertrophy by Estes criteria

- Thalach: maximum heart rate achieved
- Exang: exercise induced angina (1=yes; 0=no)
- Oldpeak: ST depression induced by exercise relative to test
- Slope: the slope of the peak exercise ST segment

  Value 1:upsloping,,Value 2: flat, ,Value 3: downsloping

- Ca: number of major vessels (0-3) colored by fluoroscopy
- Thal: 3= normal; 6= fixed defect; 7= reversible defect
- Target : 0 = less chances of heart attack, 1=more chances of heart attack

Data Wrangling and analysis

 # check null values in the dataset

data.isnull()

data.isnull().sum()

**# check for duplicate data and drop them.**

data_dup = data.duplicated().any()

print(data_dup)

data = data.drop_duplicates()

Major Obstacles

- As I am little aware of medical concepts, it takes some research to understand some topics and its values
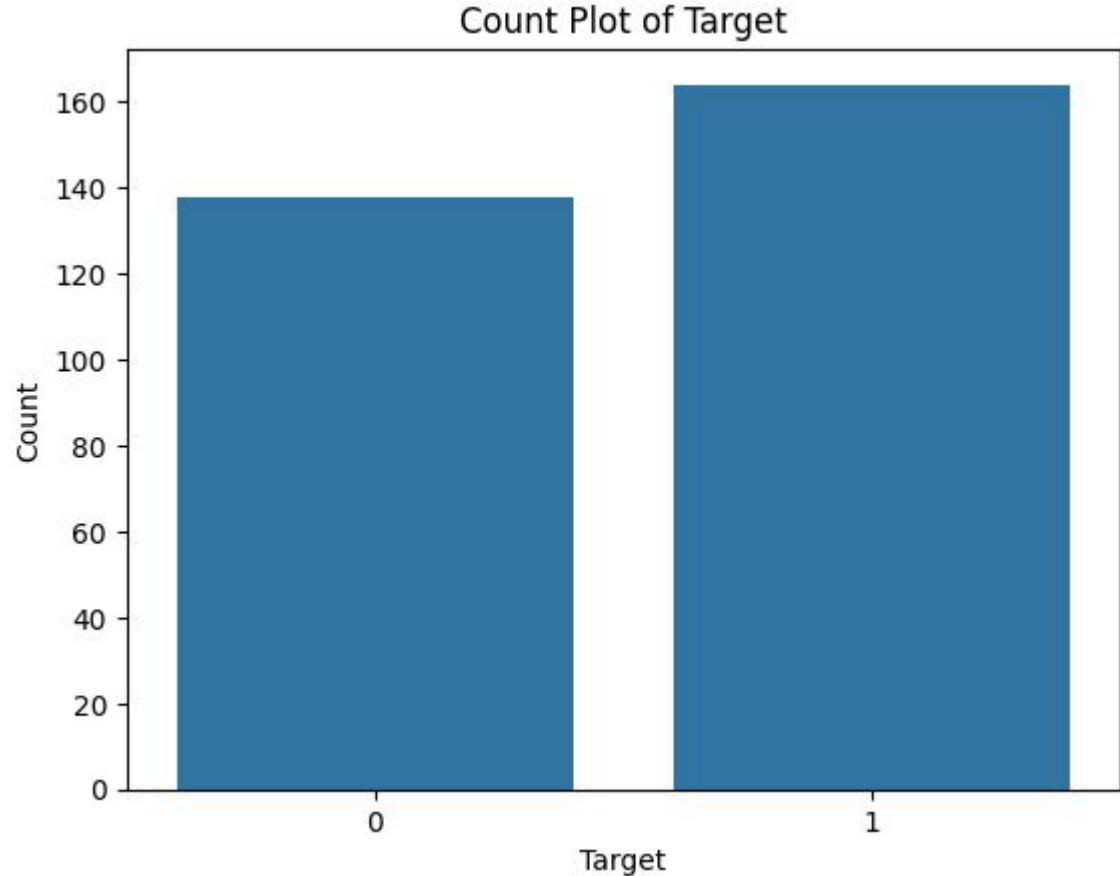- Anyhow with the interest and commitment everything is possible

To check correlation between different features available in our dataset
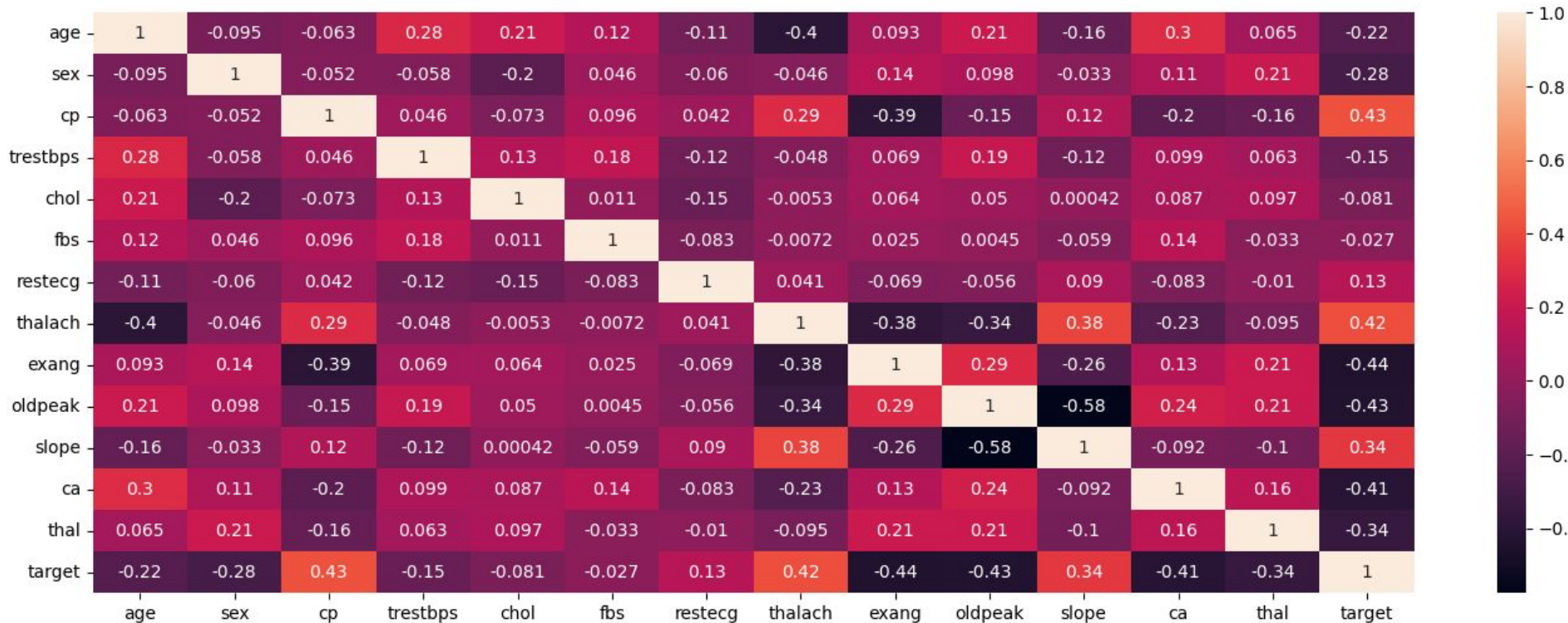
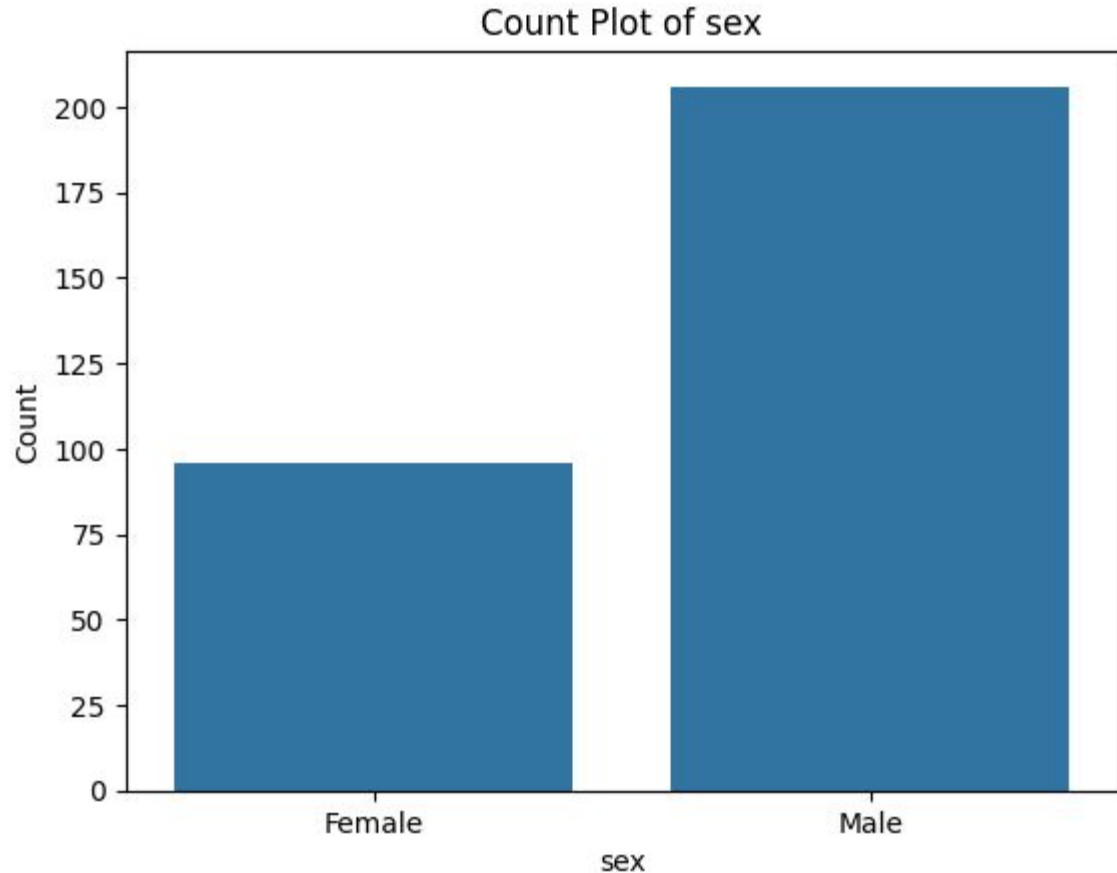# Analyzing how many people have heart disease and how many don't have

From this countplot, we can see that half of the people in this study had heart disease
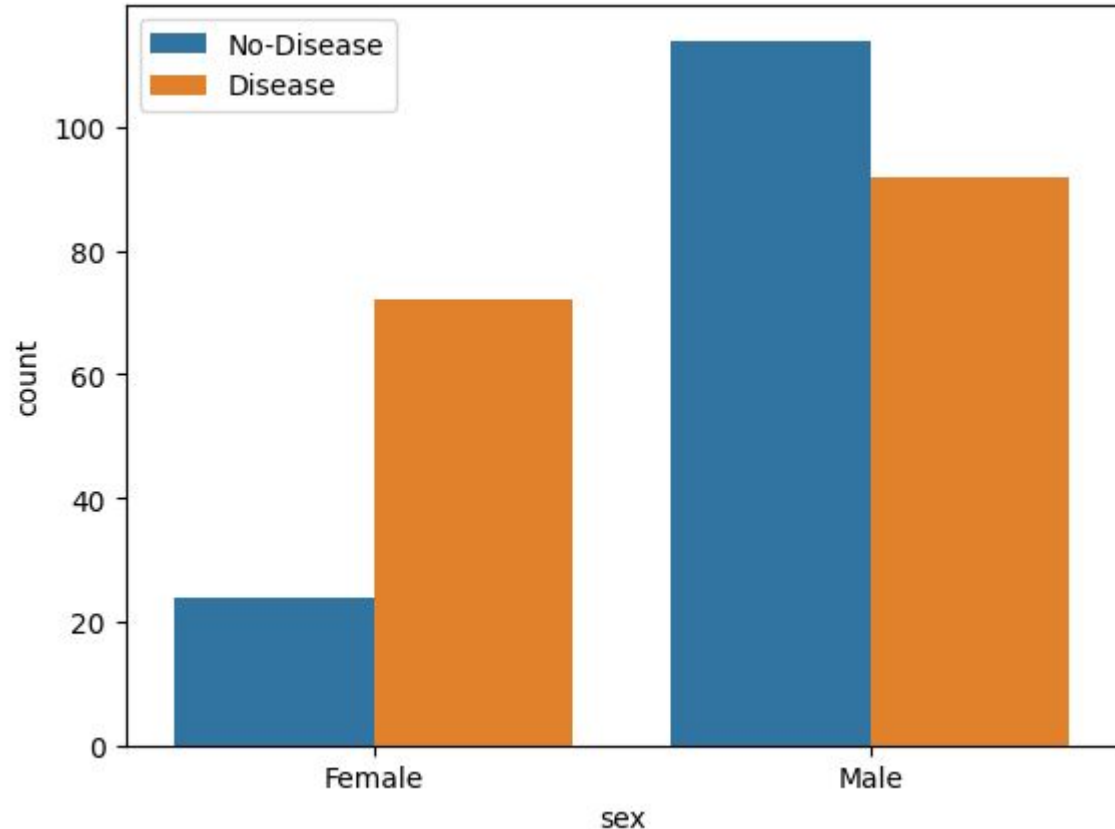

Count Plot of Target

To find count of male & female in the dataset

From this countplot we can see that around 30% are female and 70% are male

Count Plot of sex

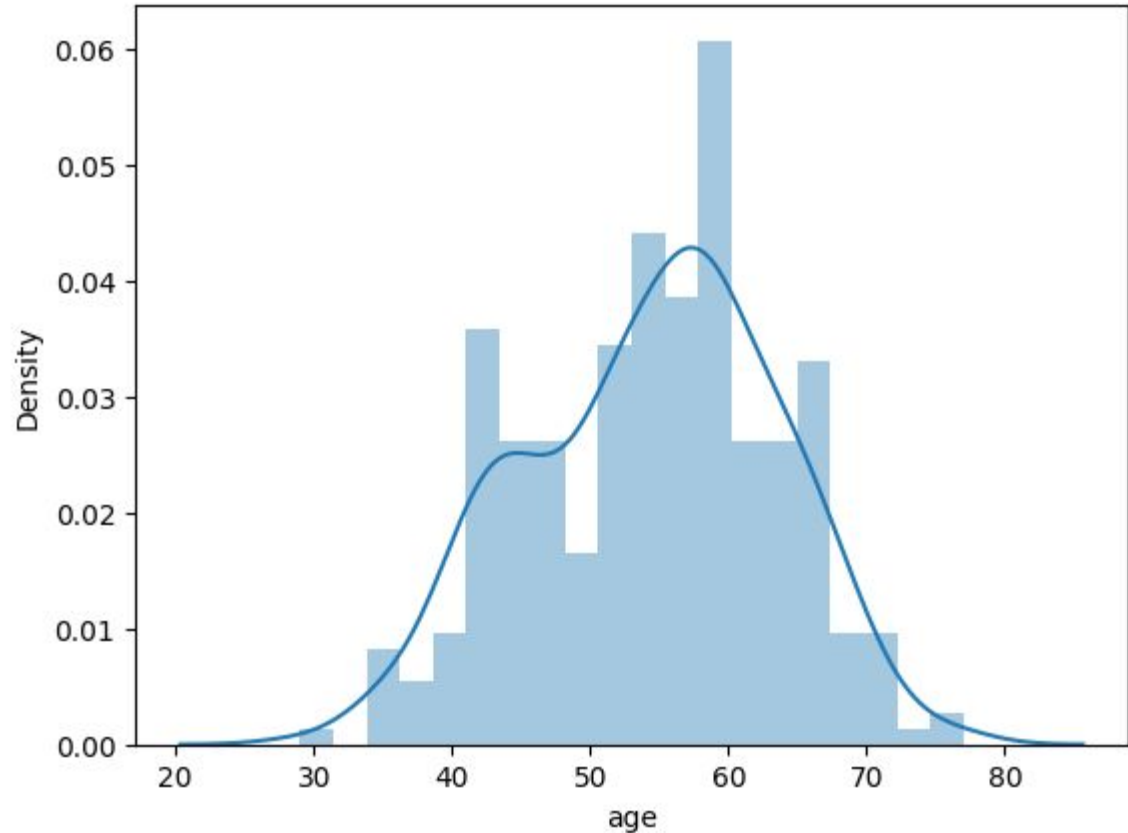# Finding gender distribution according to the target variable

From this countplot,we can see that there are more mens with disease and non-disease in the dataset

Checking age distribution in the dataset

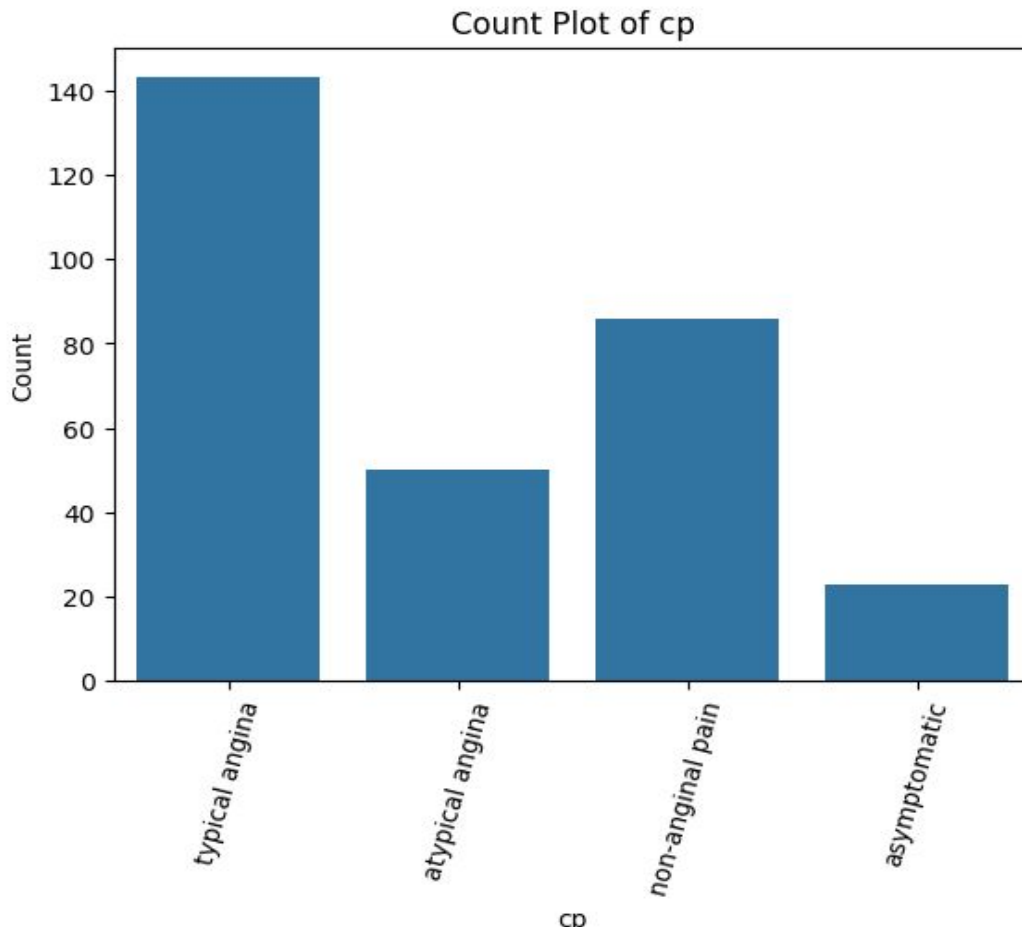From this plot,we can see that most of the people in this dataset is aged from 50 to 60.

Checking pain types

Chest pain types(4 values)
- **Value 0:** typical angina
- **Value 1:** atypical angina
- **Value 2:** non-anginal pain
- **Value 3:** asymptomatic

From this plot,we can see that chest pain type 0 which is typical angina is more common between people


Count Plot of cp

Chest pain distribution as per the target variable

- From this plot, we can see that even healthy people have chest pain
- Chest pain can be subjective due to stress, physical activities and many more
- It varies between gender
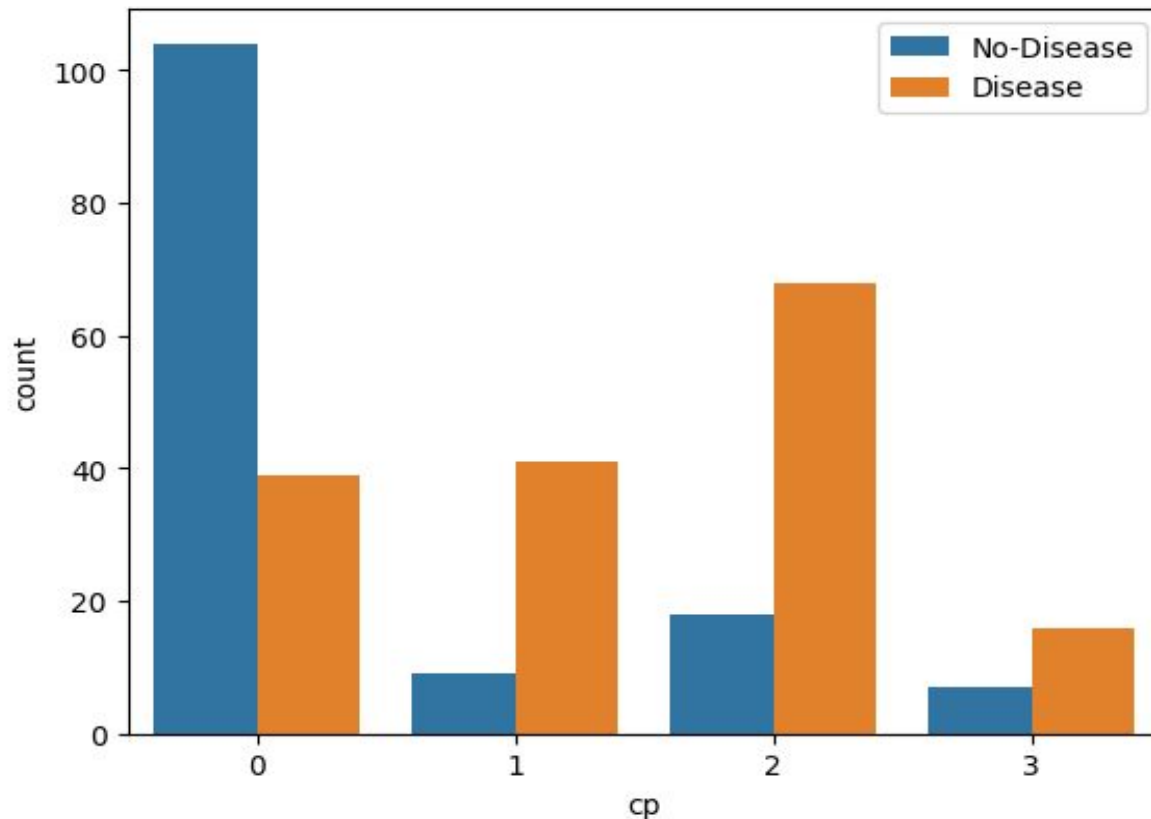
# Fasting blood sugar distribution according to the target variable

- Fbs is a diabetes indicator
- With fbs > 120 is considered as diabetic
- From this countplot,we can see that there are higher number of diabetes patients without heart disease
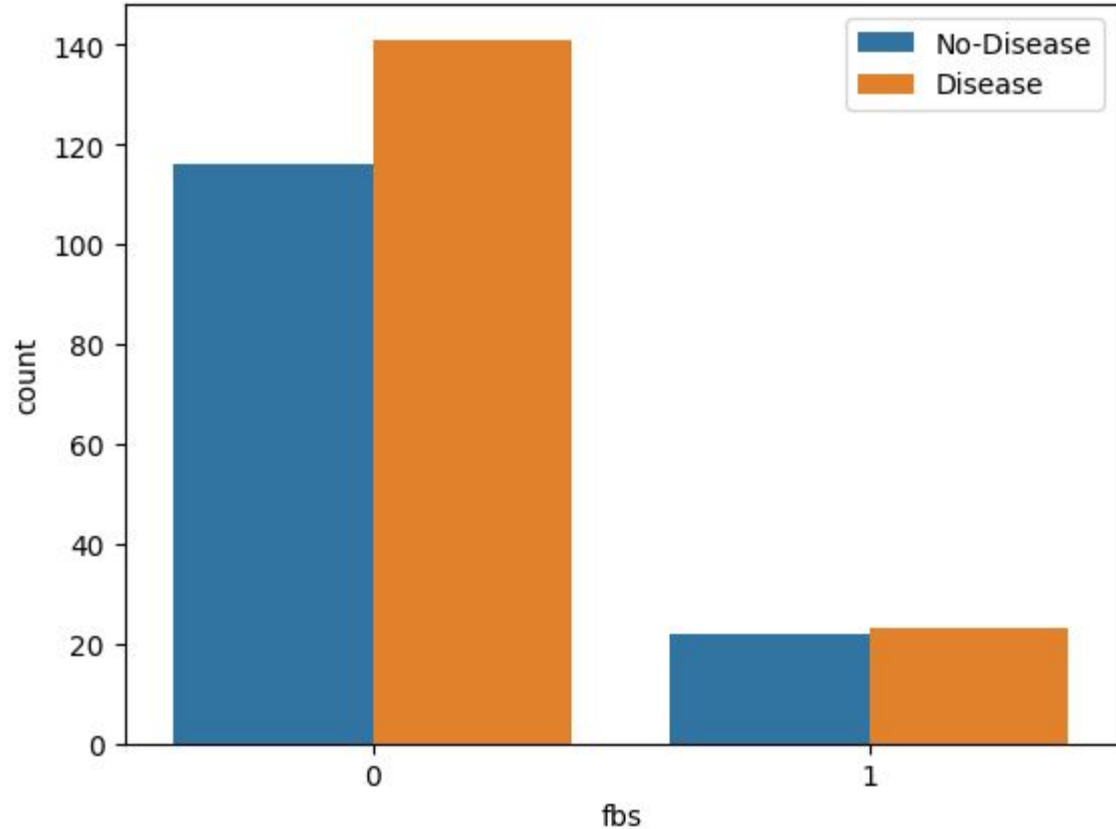
Check resting blood pressure distribution

From this histogram,we can see that blood pressure of the people is between 120 to 130

Distribution of Resting Blood Pressure
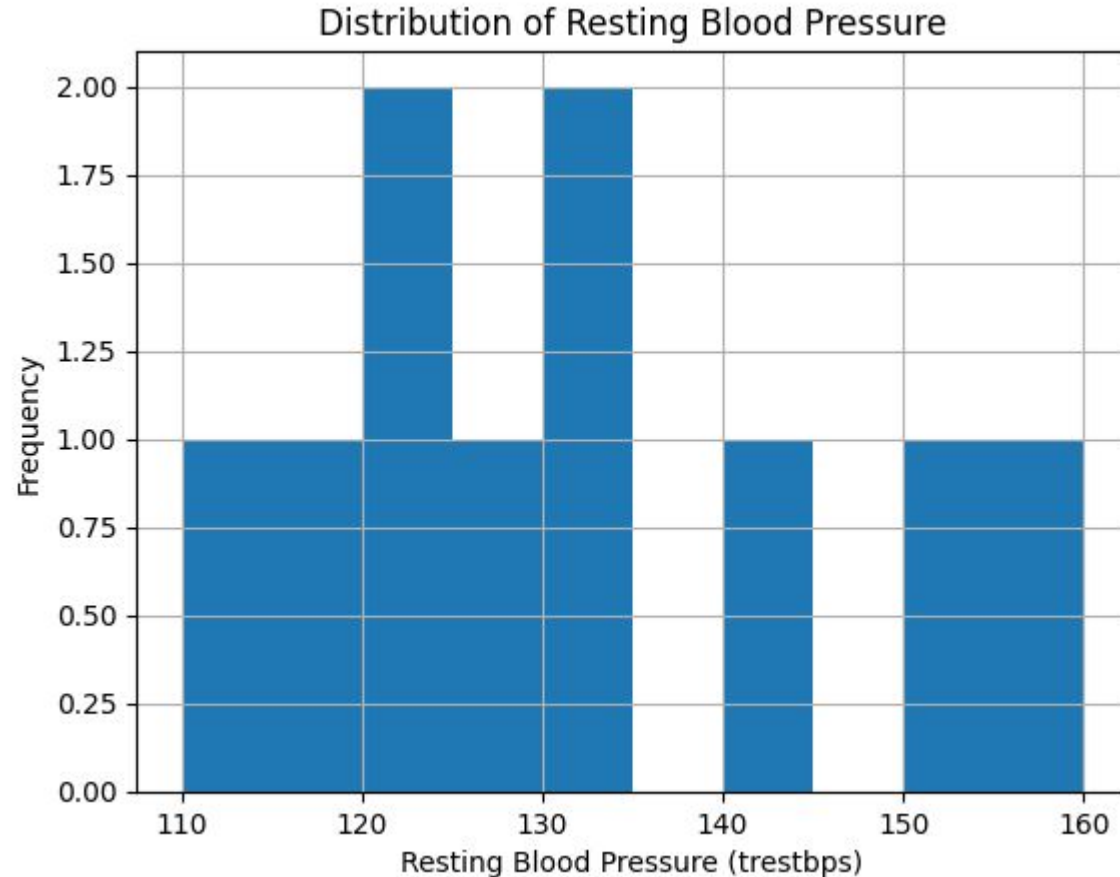
Compare resting blood pressure as per sex column

KDE Plot of Resting Blood Pressure by sex
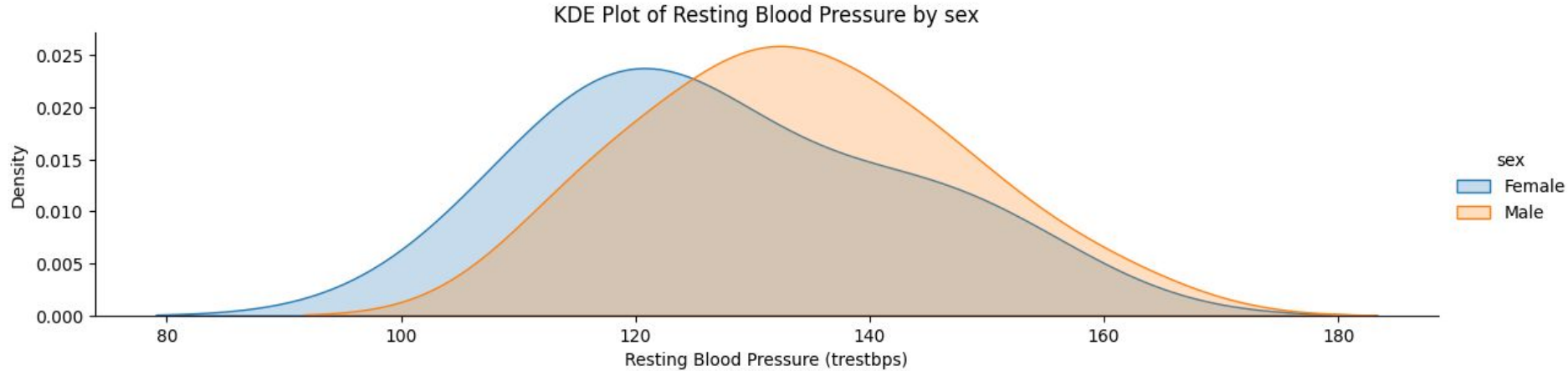
From this plot,we can see that women have lower resting blood pressure compared to men. For women mostly it is around 120 while for men is a little less than 150.

Distribution of serum cholesterol

- Serum cholesterol is measured in mg/dl
- Healthy cholesterol is less than 200 mg/dl
- In this dataset,there are more persons with the highest level of cholesterol

Distribution of Serum Cholesterol
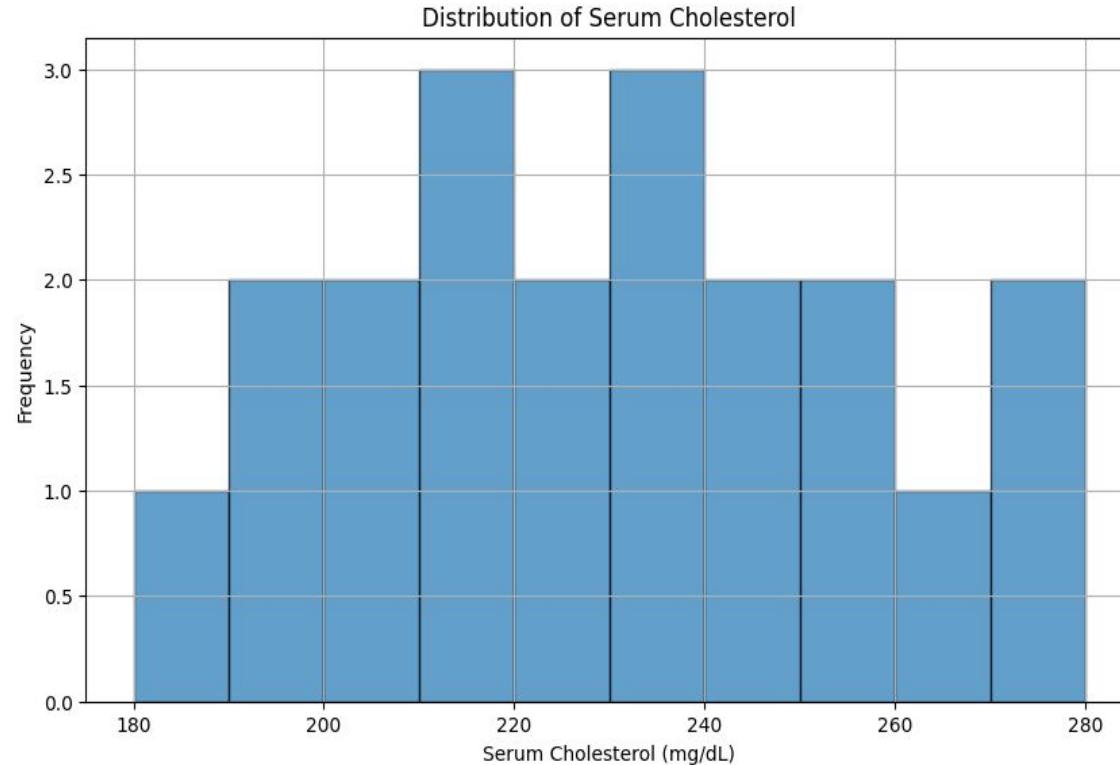
Potential business case & Conclusion

- From this dataset, we have identified that male above 50 has affected more with heart disease
- *Business case:* Male patients before the age of 50 who have either a cholesterol , blood pressure or diabetes could be tested for heart problems(For eg: from age 48). With this we can prevent the heart disease and the hospitals can also treat the patients effectively.
- Death rate can be prevented because of the early preventive measures
- Doctors can collaborate with each other (Blood pressure doctor, Diabetes doctor and heart specialist doctor(Cardiologist)
- Collaboration with physiotherapist, dieticians and Yoga centers

Thank you so much for your time guys!

Have a nice day!!

-Govarthini GS