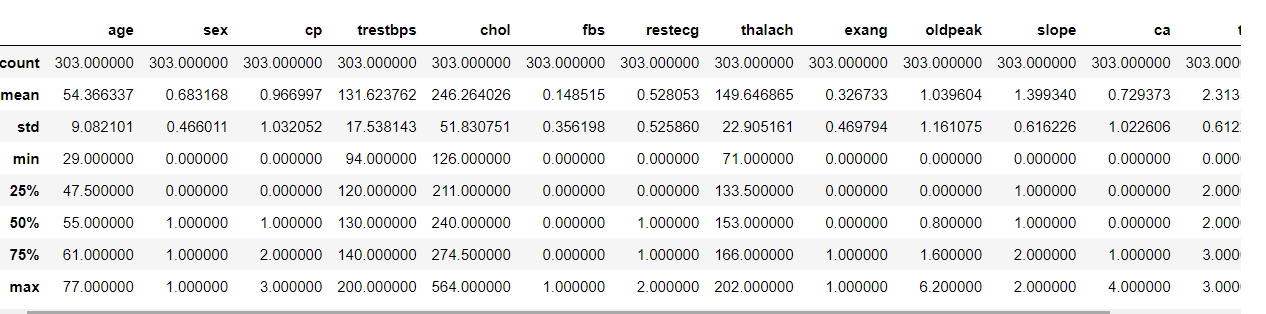
**Capstone Project**

**Domain-**Healthcare

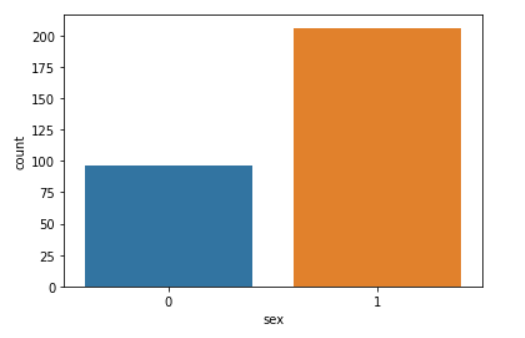
**Problem Statement-**Cardiovascular diseases are the leading cause of death globally. To identify the causes and to develop a system to predict heart attack in an effective manner is necessary. The presented data has all the information about all the relevant factors that might have an impact on heart health. The data needs to be explained in detail for any further analysis.

**Preliminary analysis-**

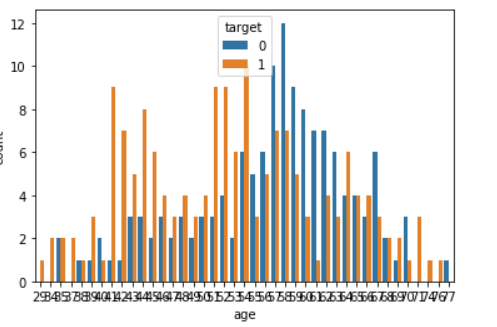
* Given dataset is in the excel format and using excel find out duplicates ,there is one duplicate found and removed.
* Load the dataset in the python for further analysis.
* After checking the null values ;no null values in the data.
* Statistical summery of the data

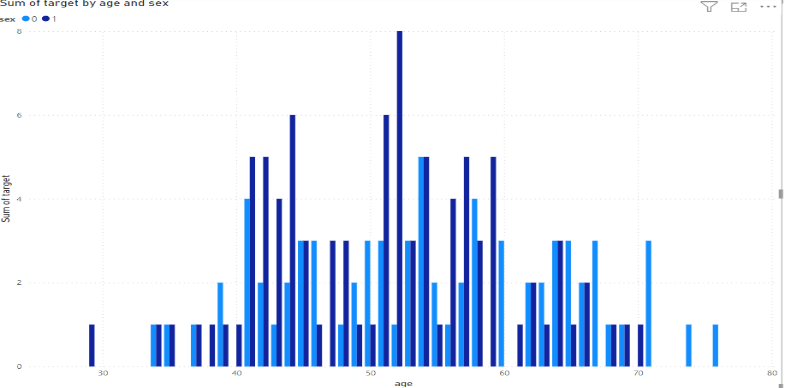


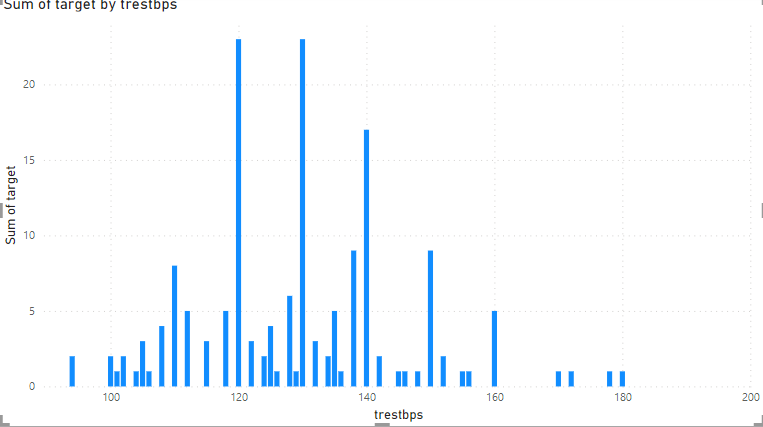
* All columns have integer datatype.
* Count plot for sex



* Occurrence of cvd across different ages

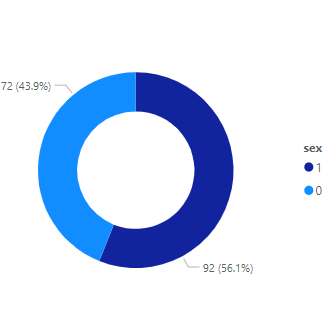


* Occurrence of cvd across different ages and sex
* 
* The above chart shows that at the age of 52(Male) have more chances of cvd i.e. 4.88% of overall survey
* Study the heart attack based on anomalies in the resting blood pressure

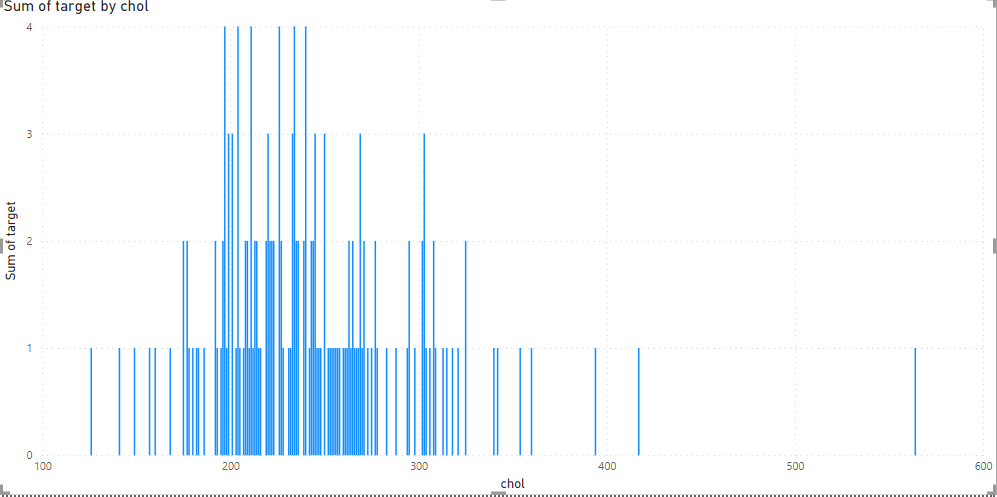
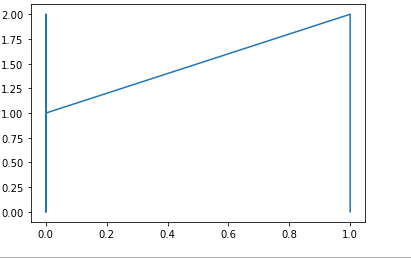
. 

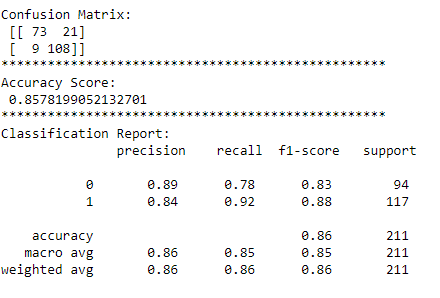
Above chart shows the rest blood pressure at the time of admission is between 120 to 130 have more number of heart attacks

* Study the age category using donut chart



Above chart shows more male affected from heart attack than female

* Study the cholesterol level 
* The cholesterol level 200 to 240 shows more number of heart attacks
* Relationship of peak exercise and heart attack
* Peak exercise is directly proportional to the number of heart attacks
* Relationship between all variables shown using pairplot.
* Prepare dummy columns for building model
* Accuracy of the Training dataset is 85.78%



* Accuracy of the Test dataset is 80.21%

