OBSERVATIONS

1. There is total 303 datapoints in each column
2. Each columns have 0 null values so need of dropping filling the columns
3. The maximum age of the patient is 77 whereas the minimum age is 29
4. 75 % of paitents are below 61 years of age
5. Target variable has only 0 and 1 entry so it is a clearly Classification task
6. Categorical variables:

   Sex, chest pain type(cp), fasting blood sugar > 120mg/dl(fbs), resting electrocardiographic results(restecg), exercise induced angina(exang), slope, ca, thal

7. Numerical variables:

   Age, trestbps, chol, thalach, oldpe

8. 165 patients have cardiovascular disease whereas 138 patients are not affected
9. At slope 0 and 1 the patient is getting less heart attacks
10. And at slope 2 heart attacks of the patients have been increased
11. As the age is getting increased the maximum heart rate is getting decreased
12. Whereas the cholesterol is getting increased
13. And the resting blood pressure is not having any linear relation with the age
14. It is clear from the heatmap that the serum cholesterol in mg/dl and fasting blood sugar > 120 mg/dl are less correlated with the target
15. Male age less than 40 are only affected by heart disease whereas female age ranges from 55 - 65 are not affected by heart disease
16. Following are the features which are having p-values less than 0.05 and thus are the useful features for logistic model:
    'cp', 'ca', 'sex', 'thal', 'oldpeak', 'exang', 'thalach'.

We have used the **Logistic Regression algorithm** to predict the values and used the **cross validation** for finding the hyperparameters. Following are the results:

Hyperparameters:

'C': 0.3333333333333333,

'l1_ratio': 0.0,

'multi_class': 'auto',

'penalty': 'l1',

'solver': 'liblinear'

Confusion Matrix:

array ([[25, 4],

[4, 28]]

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.86	0.86	29
1	0.88	0.88	0.88	32
accuracy			0.87	61
macro avg	0.87	0.87	0.87	61
weighted avg	0.87	0.87	0.87	61

After applying **Random Forest Algorithm** following are the metrics:

Confusion Matrix:

```
array ([[24, 5], [ 6, 26]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.80	0.83	0.81	29
1	0.84	0.81	0.83	32
accuracy			0.82	61
macro avg	0.82	0.82	0.82	61
weighted avg	0.82	0.82	0.82	61

The metrics, after apply random forest, has been decreased.