SCREENSHOT – HEALTH CARE PROJECT
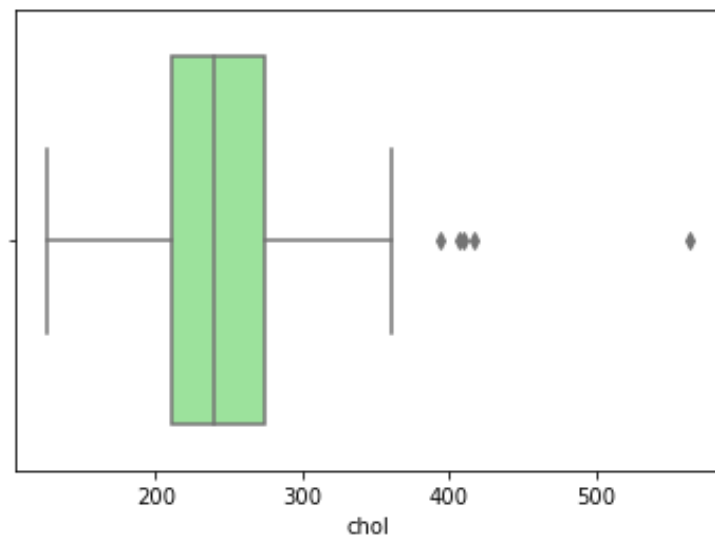
GIVEN DATA:

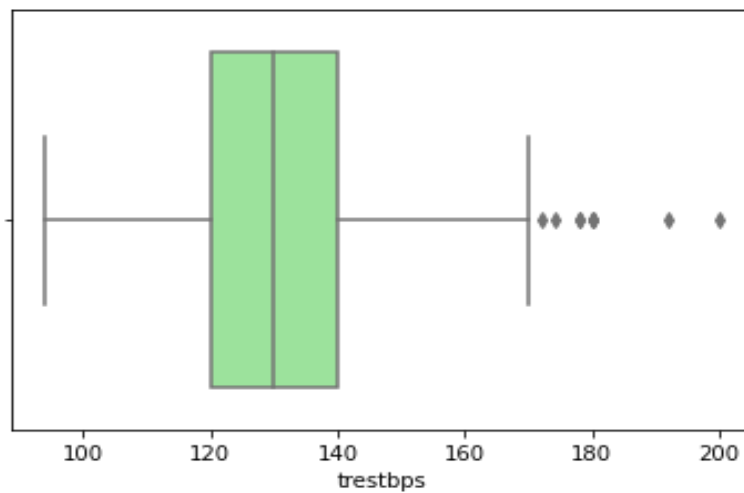Out[2]:

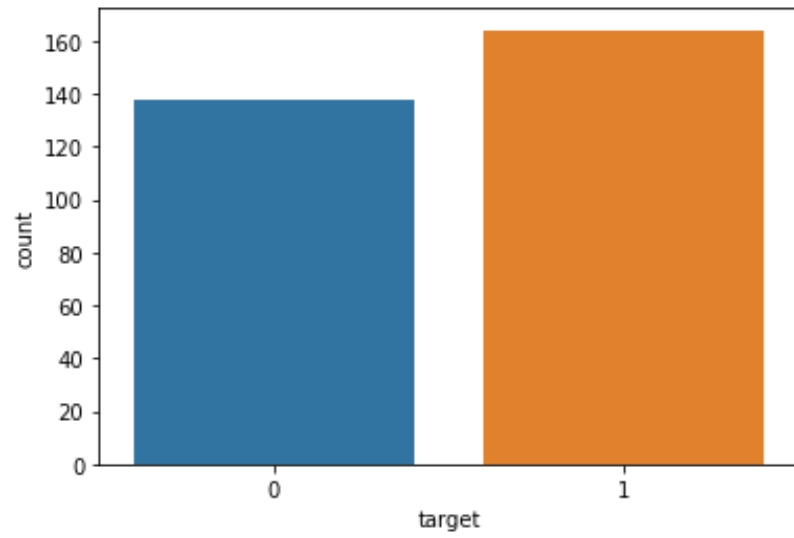|	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

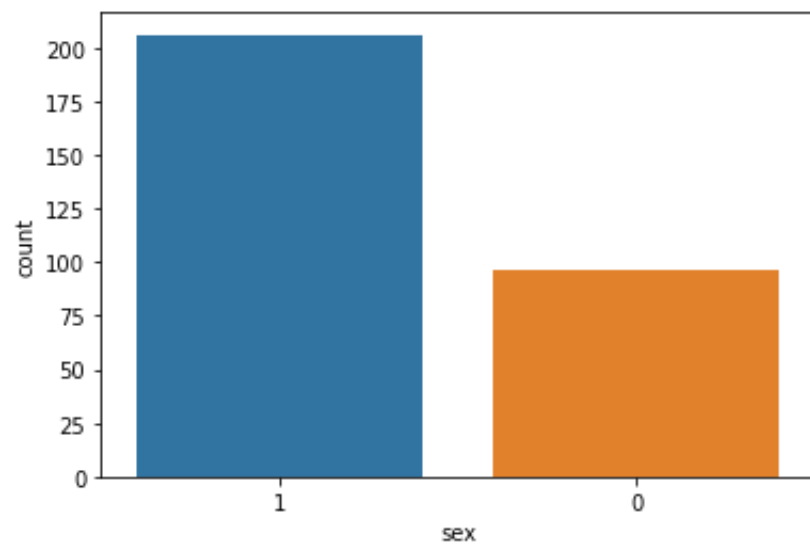## DATA PREPROCESSING

Outliers in Cholesterol:

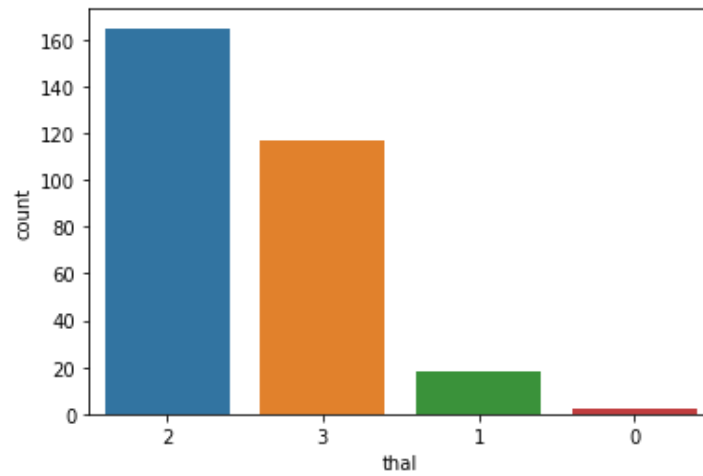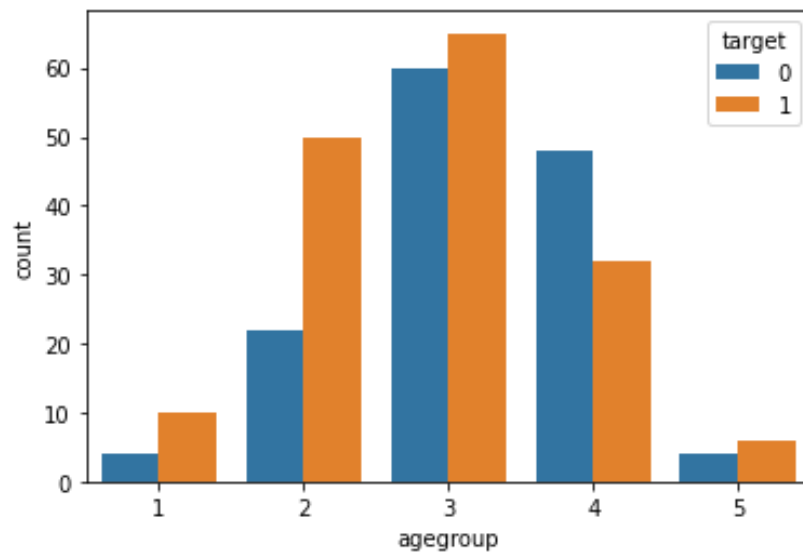**Outliers in Blood Pressure:**

COUNTPLOT ANALYSIS

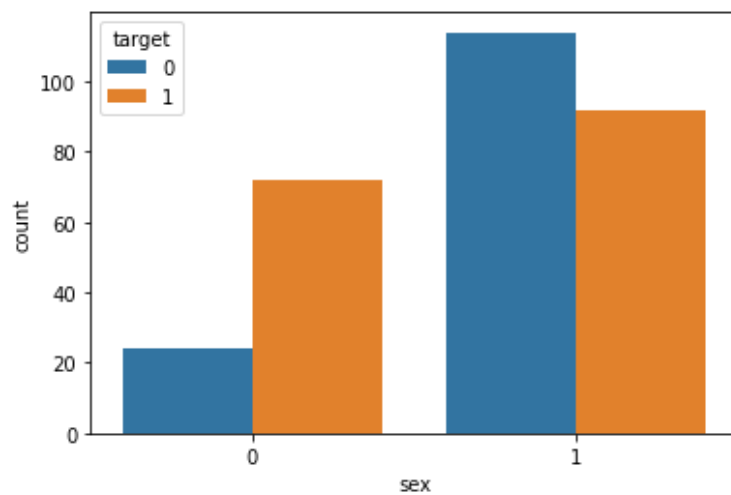**a) Target Variable:**

b) Sex

c) Ca
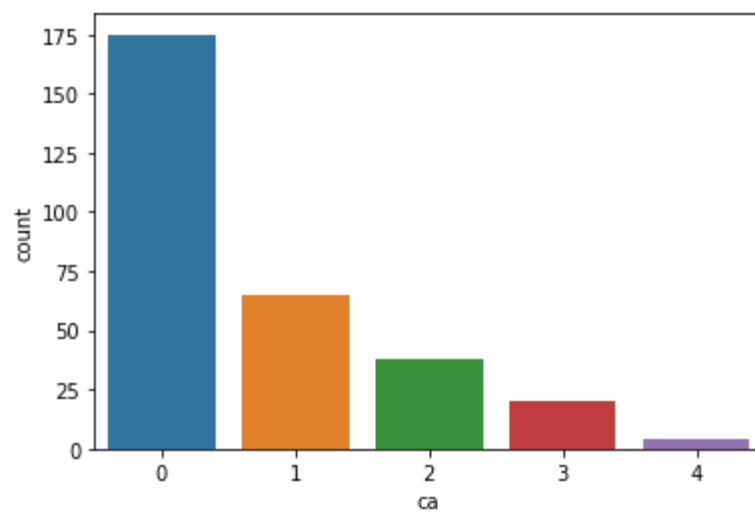
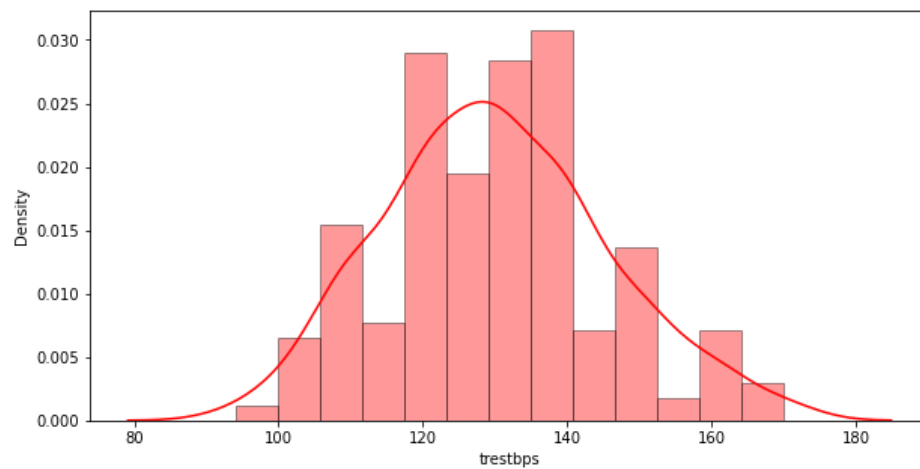**d) Thal**

Analysis of Age group:

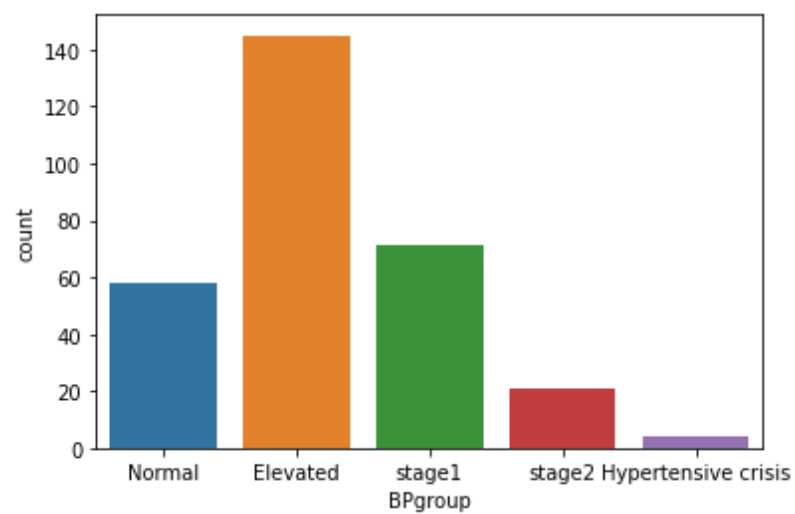## Gender Vs Target variable

Analysis of Blood Pressure:

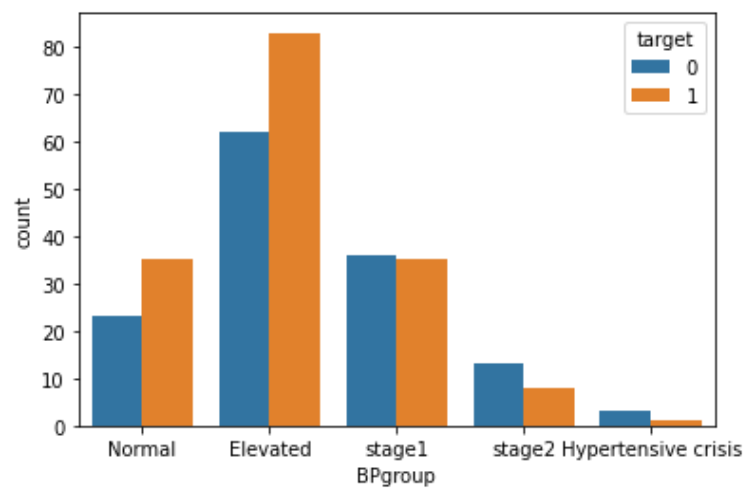a) **Distribution of BP:**

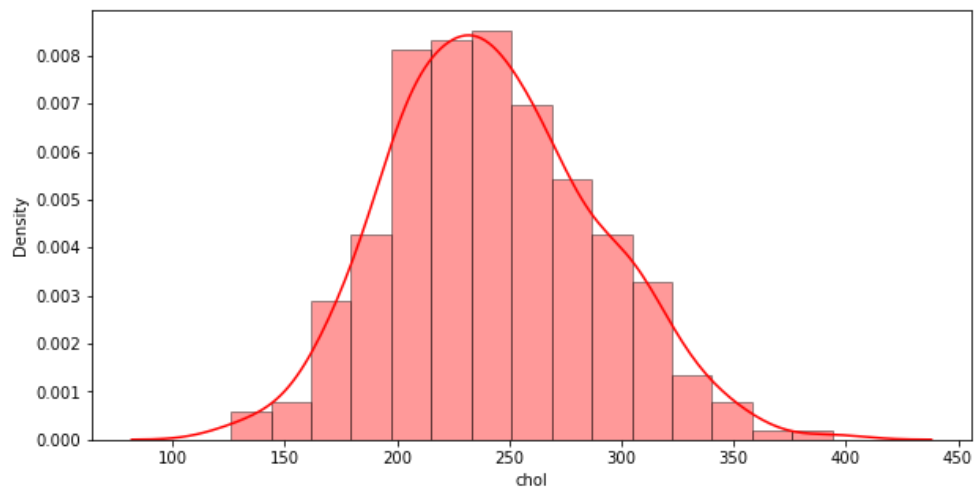b) **Grouping of BP:**

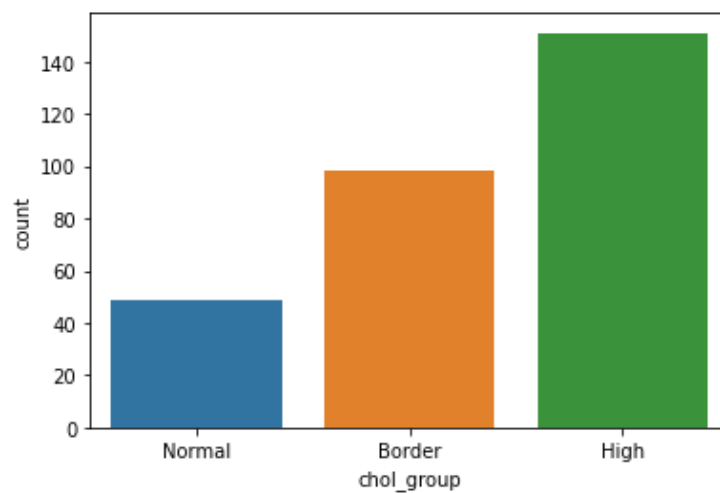c) **BP vs target variable:**

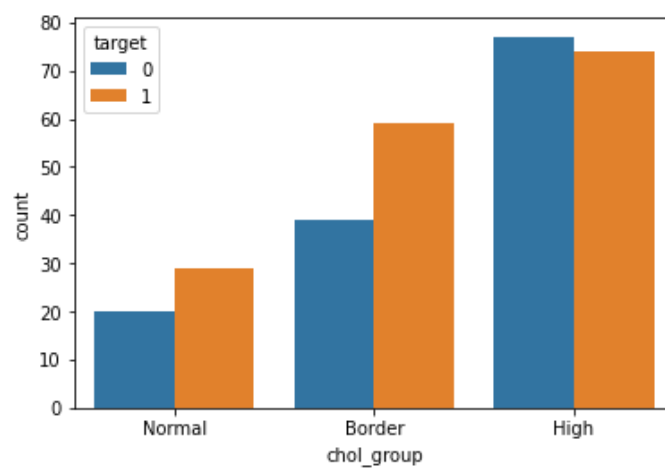**Analysis of Cholesterol:**

a) Distribution of data

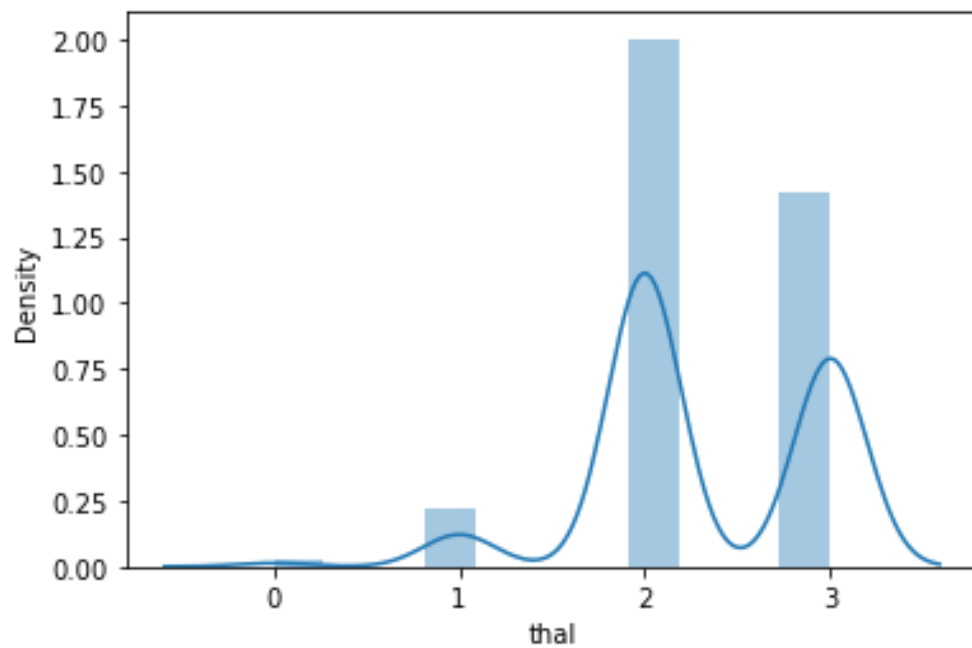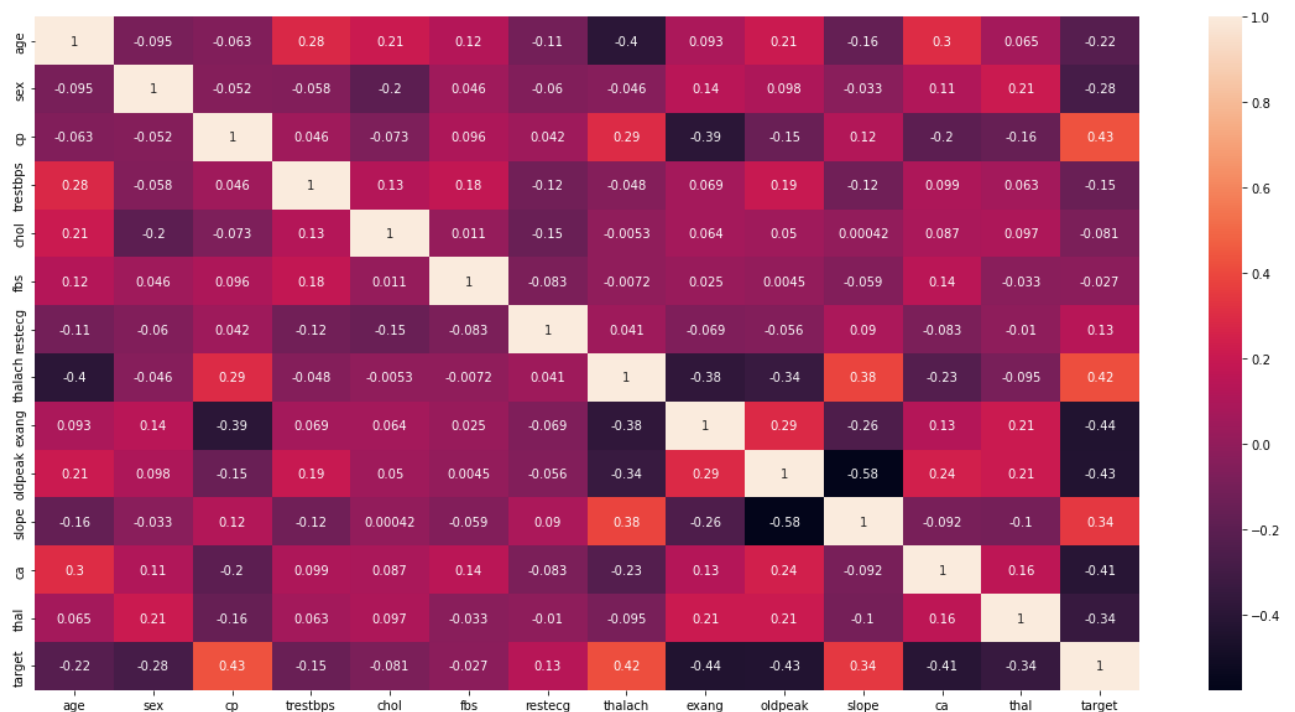**b) Grouping of cholesterol**

c) Cholesterol vs target variable

Study on Thalassemia

Correlation analysis between variables using Heat Map:

LOGISTIC REGRESSION:

Logistic Regression

```
In [101]:  ▶  classifier = LogisticRegression(solver='lbfgs',max_iter=10000)

In [102]:  ▶  classifier.fit(X_train, y_train)
  Out[102]:  LogisticRegression(max_iter=10000)

In [103]:  ▶  y_pred=lr.predict(X_test)

In [104]:  ▶  accuracy_score(y_test,y_pred)
  Out[104]:  0.8131868131868132
```

RANDOM FOREST:

```
  Out[86]:  RandomForestClassifier(max_depth=50, n_estimators=900, random_state=0)

In [87]:  ▶  y_pred=rfc.predict(X_test)

In [91]:  ▶  accuracy_score(y_test,y_pred)
  Out[91]:  0.8571428571428571

In [92]:  ▶  #Confusion Matrix
             confusion_matrix(y_test,y_pred)
  Out[92]:  array([[36,  4],
                   [ 9, 42]], dtype=int64)

In [95]:  ▶  print(classification_report(y_test,y_pred))

                       precision    recall  f1-score   support

                 0        0.80       0.90      0.85        40
                 1        0.91       0.82      0.87        51

          accuracy                            0.86        91
         macro avg        0.86       0.86      0.86        91
      weighted avg        0.86       0.86      0.86        91
```