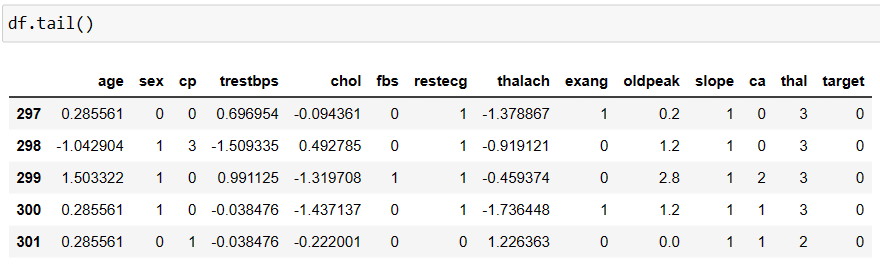
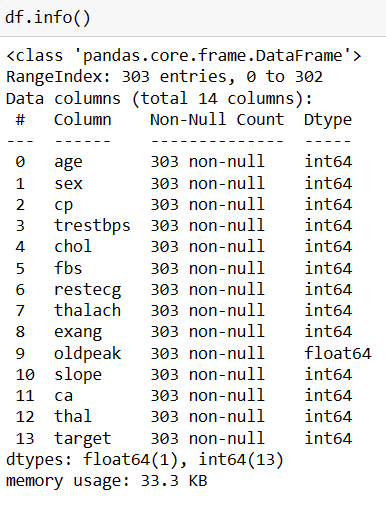
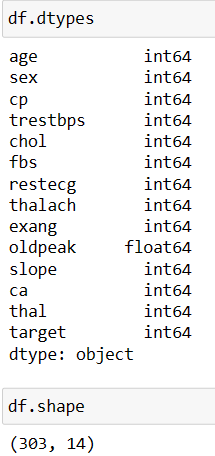
**Course End Project : HealthCare**

# 1. Preliminary Analysis:

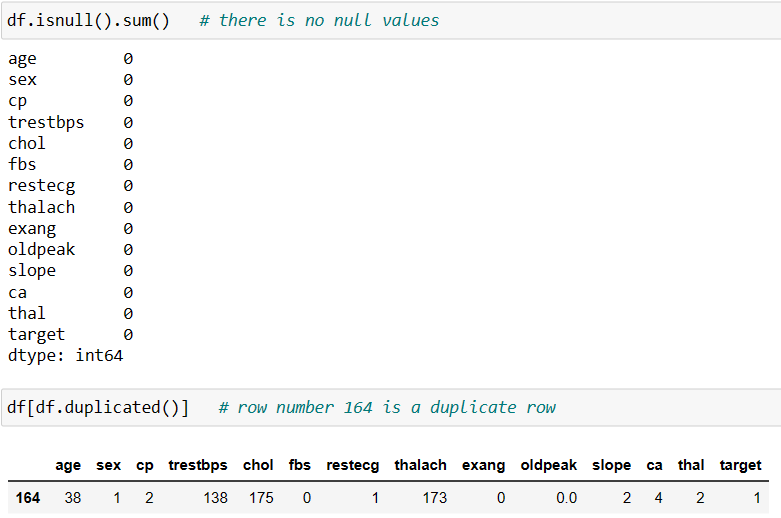
# # Reading the data and showing some analysis with it:

# 

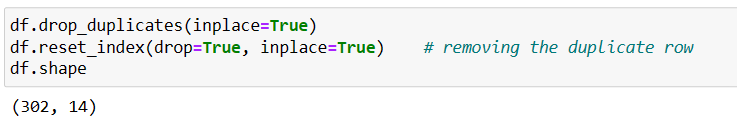


****

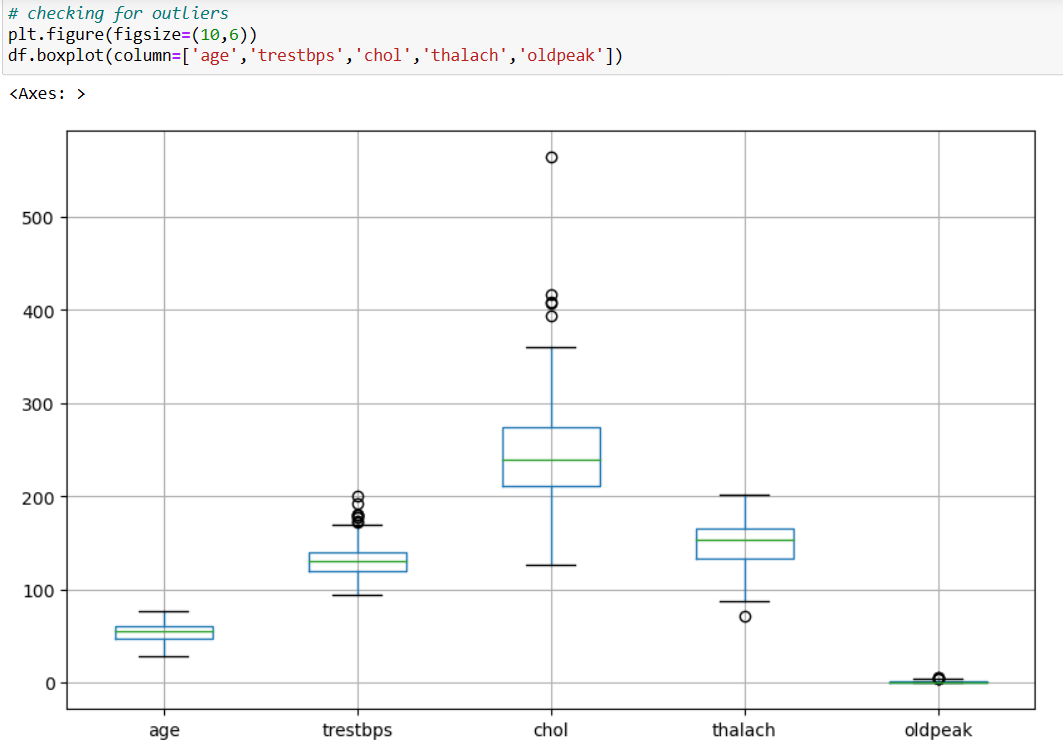
# # Checking for null and duplicate values:

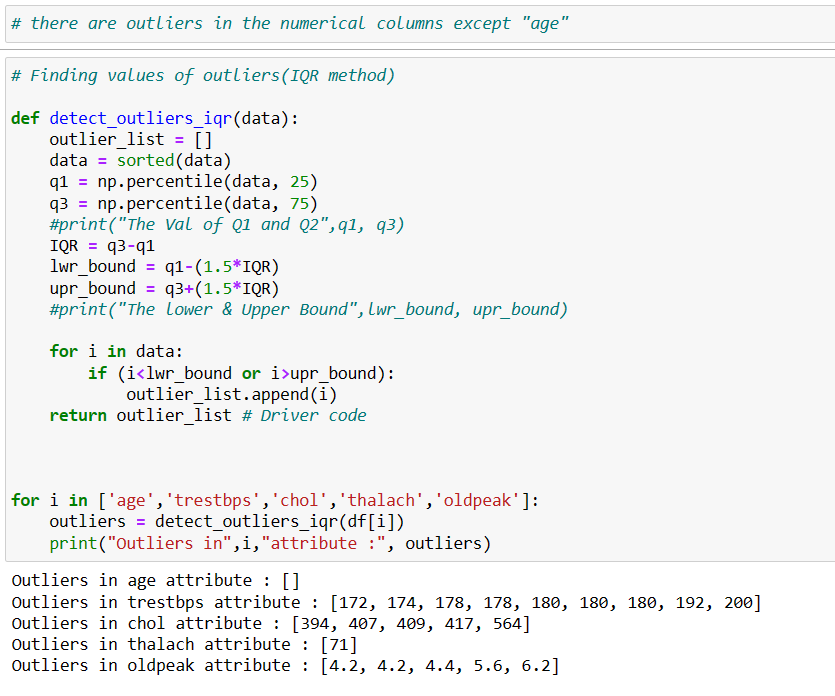


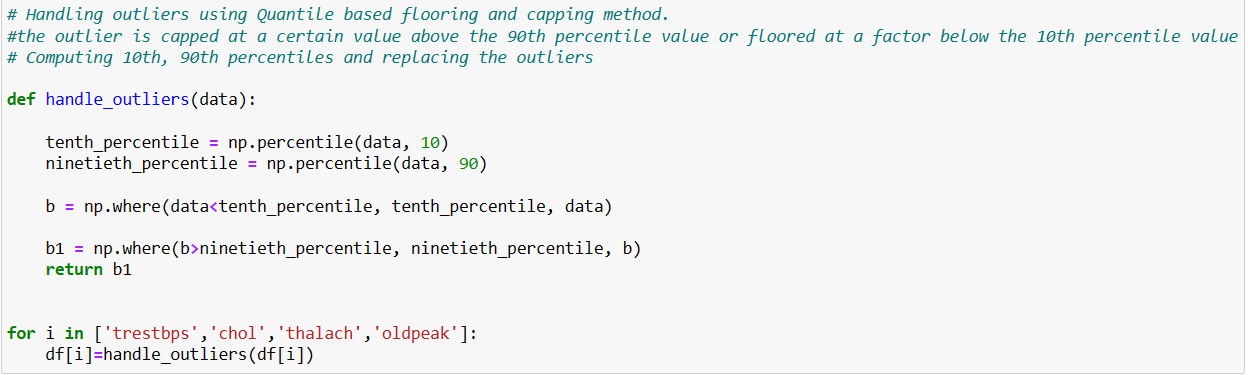
# # Removing the duplicate value:

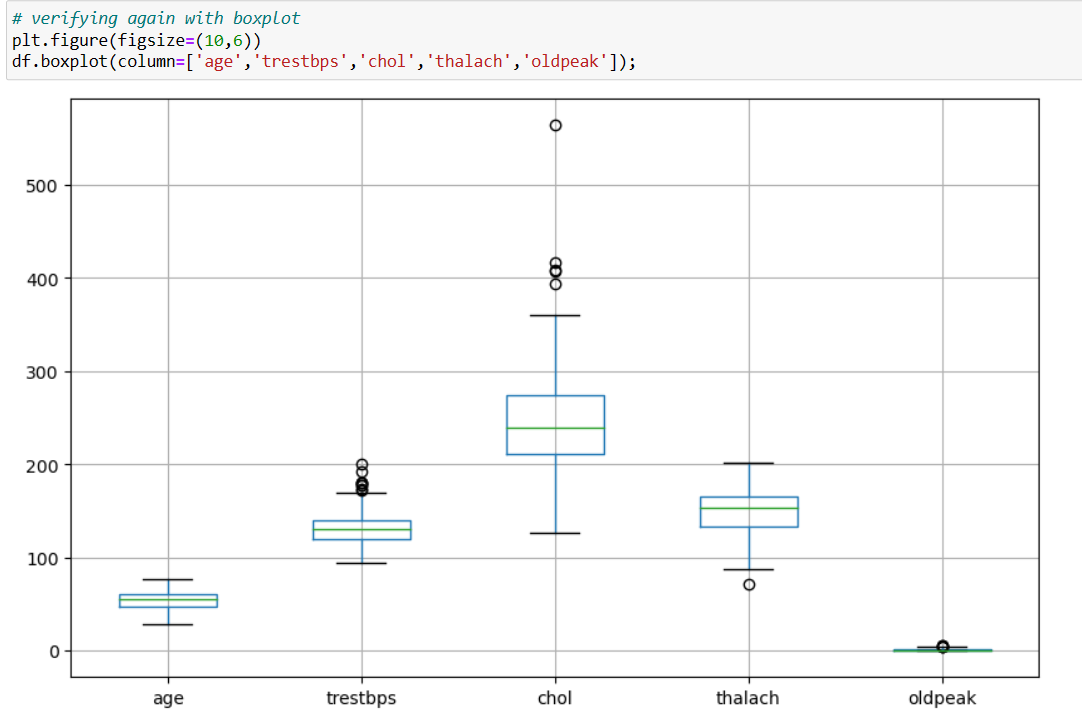
****

# # Outlier detection and removal of outliers:

****



****

****

# # Statistical summary of the data:

# 

# # Measuring various central tendencies:

# 

# 

# 

# # Exploring categorical features:

# 

# 

# 

# 

# 

# 

# # Occurance of CVD across Age Category:

# 

# 

# # Composition of All patients with respect to SEX category:

# 

# 

# # Composition of All patients with respect to “trestbps” category:

# 

# 

# 

# # Relationship between cholesterol levels and target variables:

# 

# 

# 

# # Relationship between peak exercising and occurance of a heart attack:

# 

# 

# # Relationship between thalassemia type and occurance of a heart attack:

# 

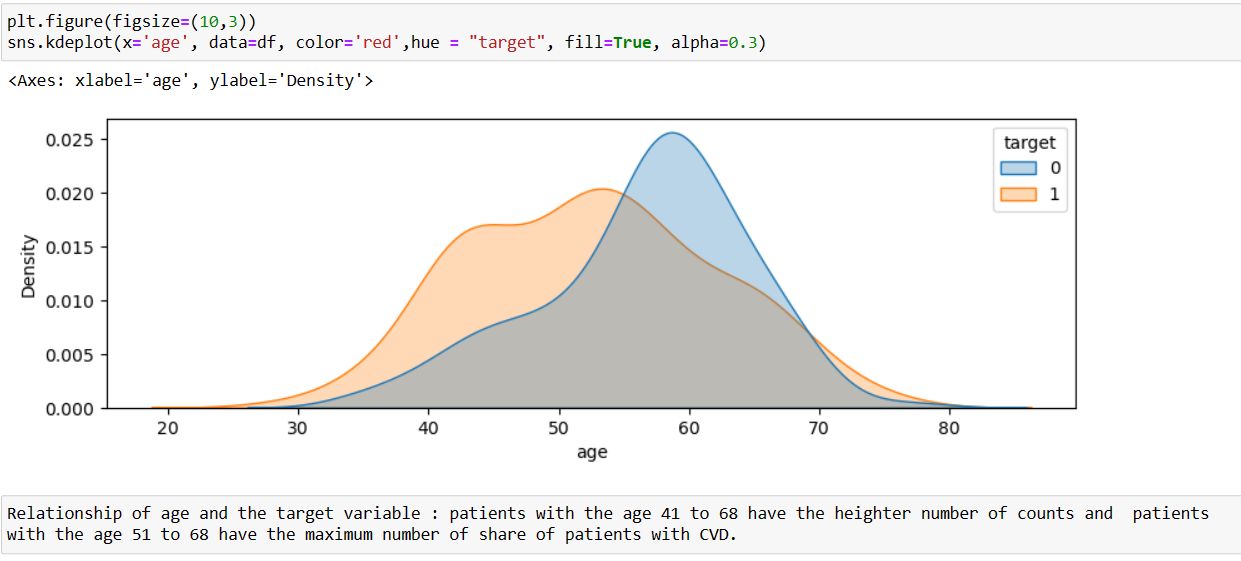
# 

# 

# # Other factors determining the occurance of a hear attack:

# “Age”:

# 



# “Chest pain type”:

# 

# 

# “Fasting blood Sugar”:

# 

# “Resting electrocardiographic results”:

# 

# 

# “Exercise induced angina”:

# 

# 

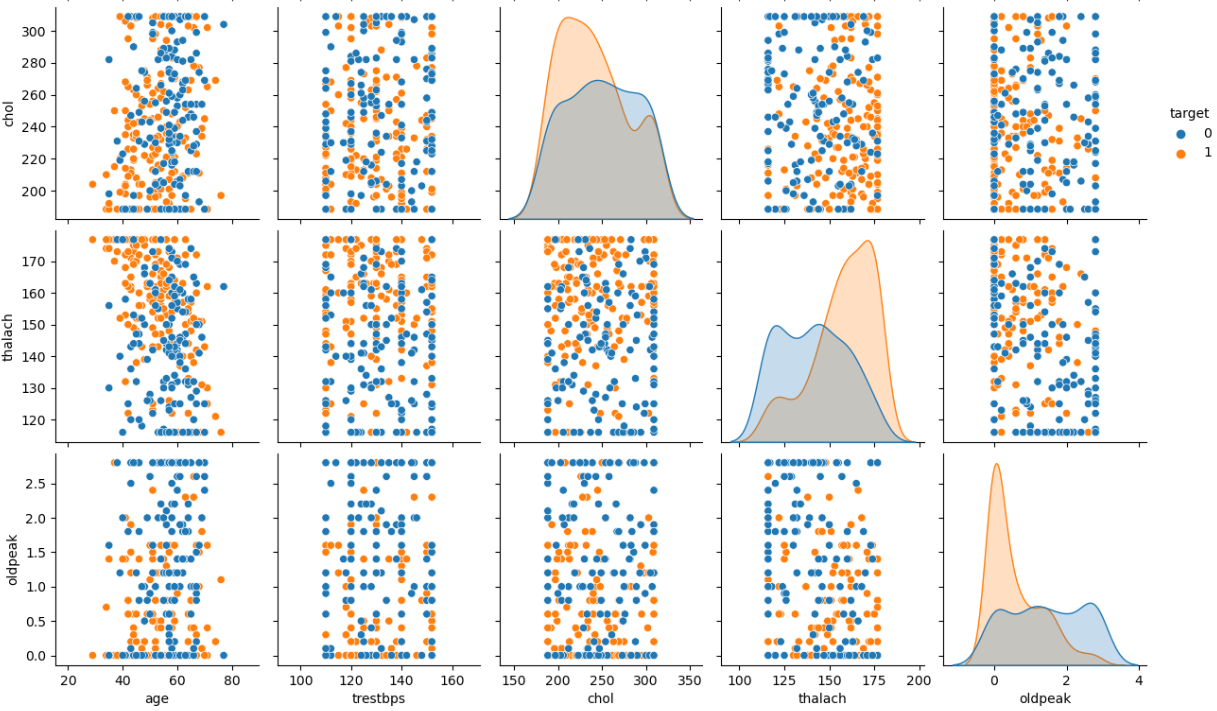
# “Oldpeak”:

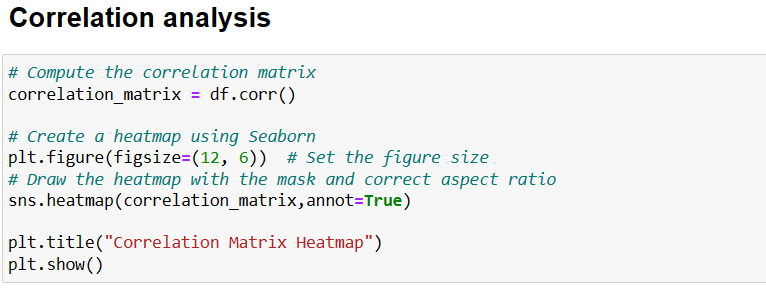
# 

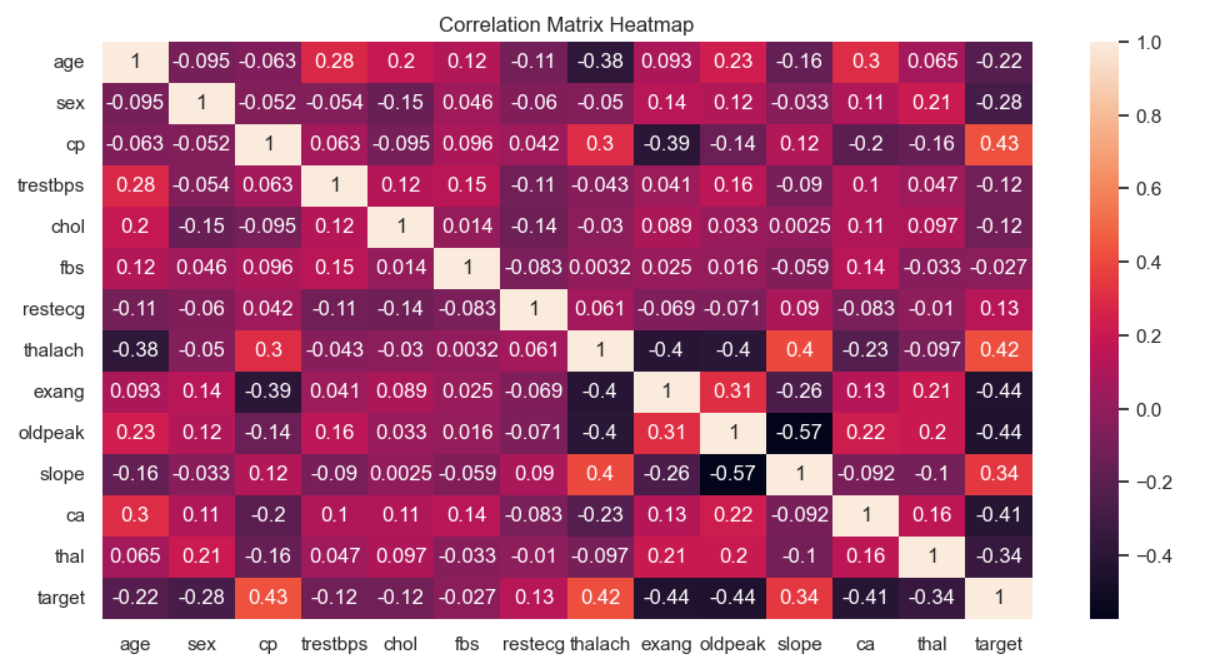
# 

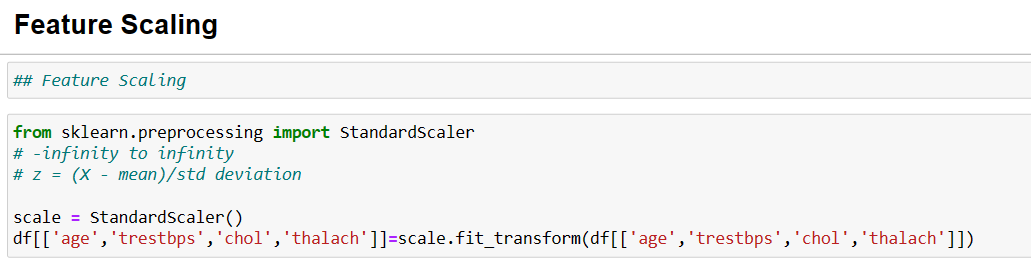
# ## Pairplot:

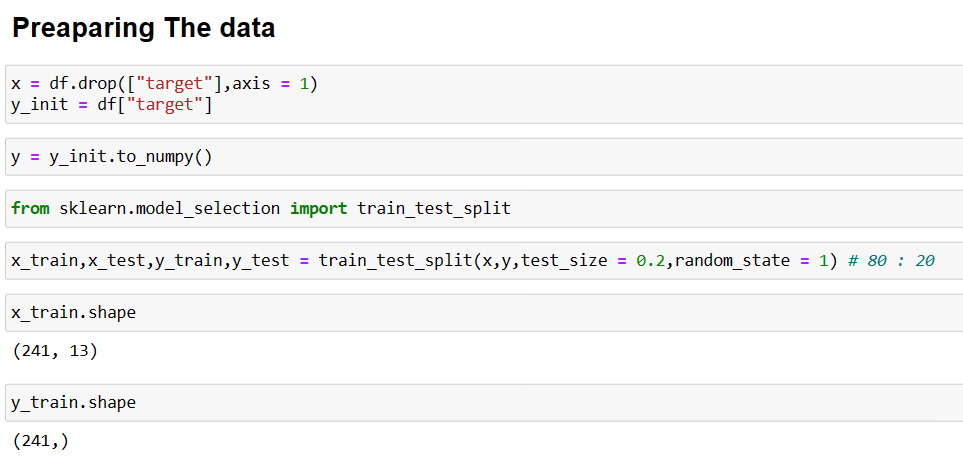
# 



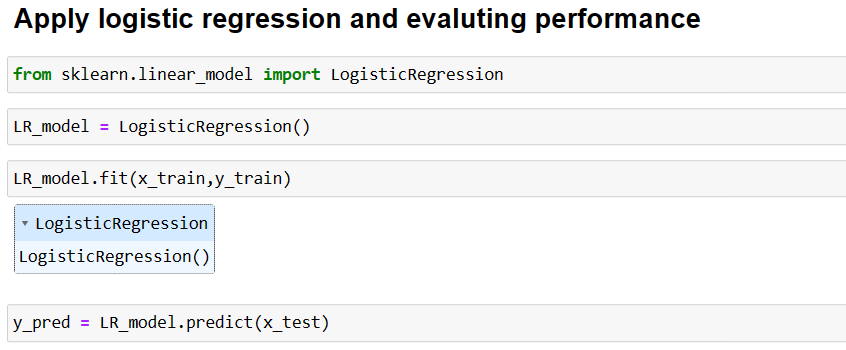


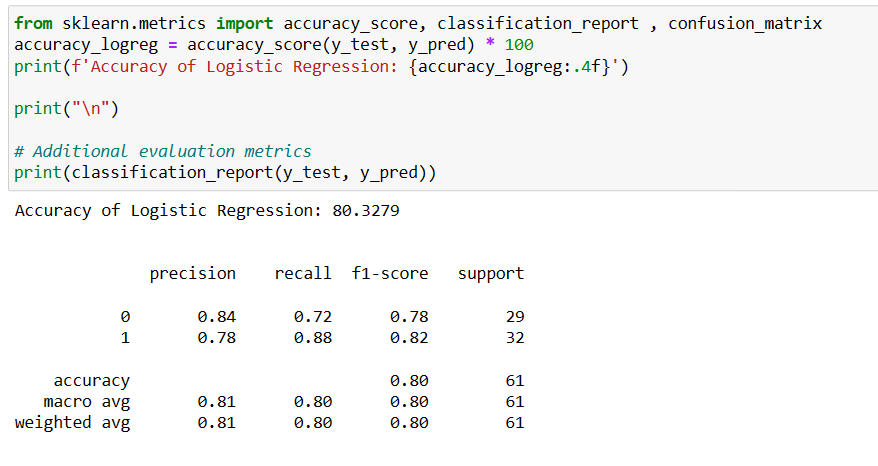


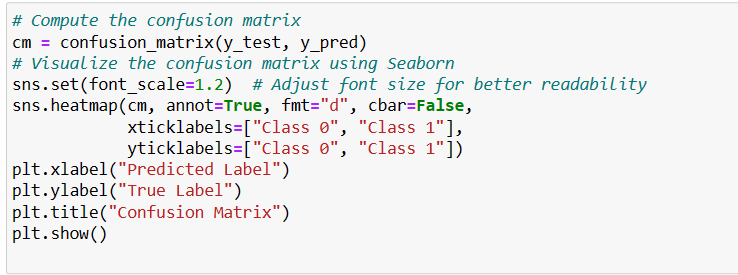


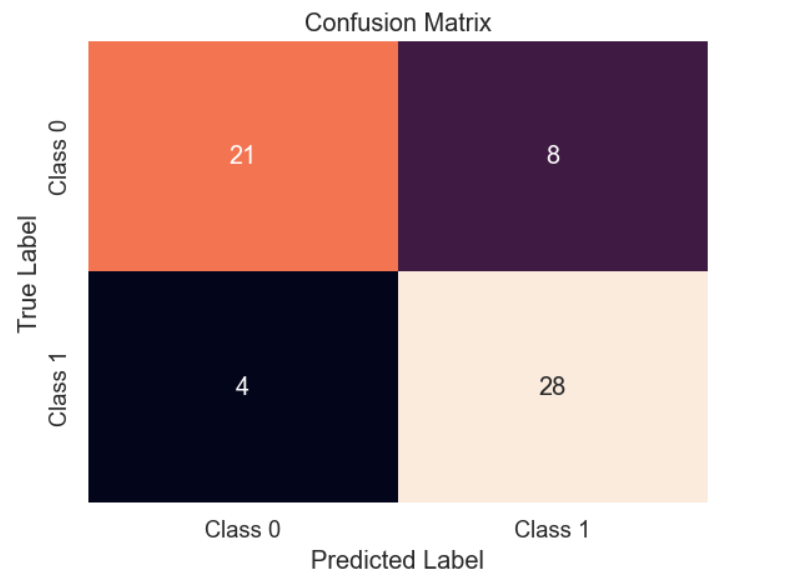


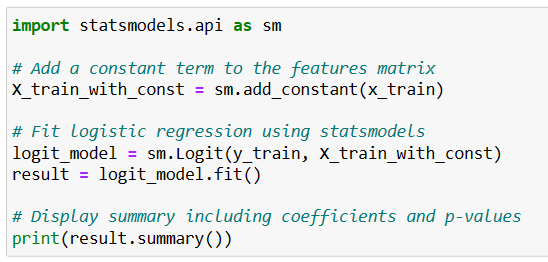
# 2. Model building:

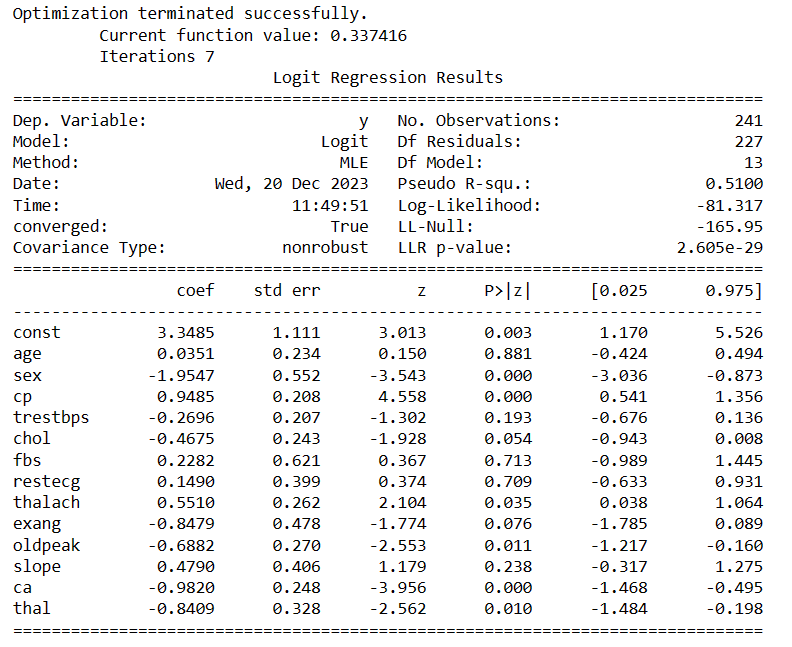


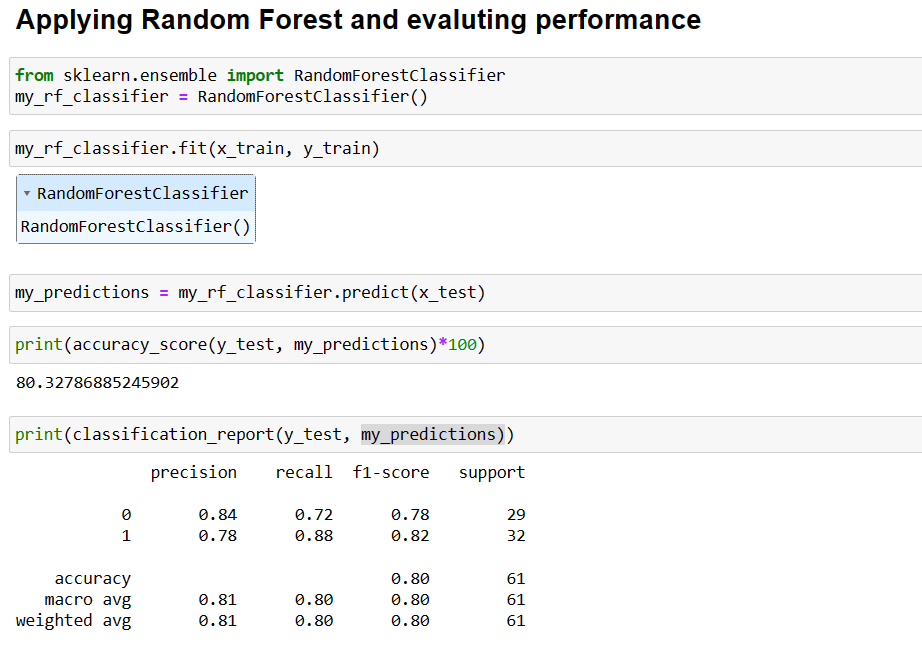


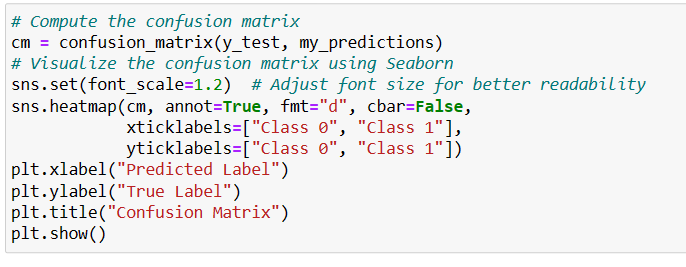


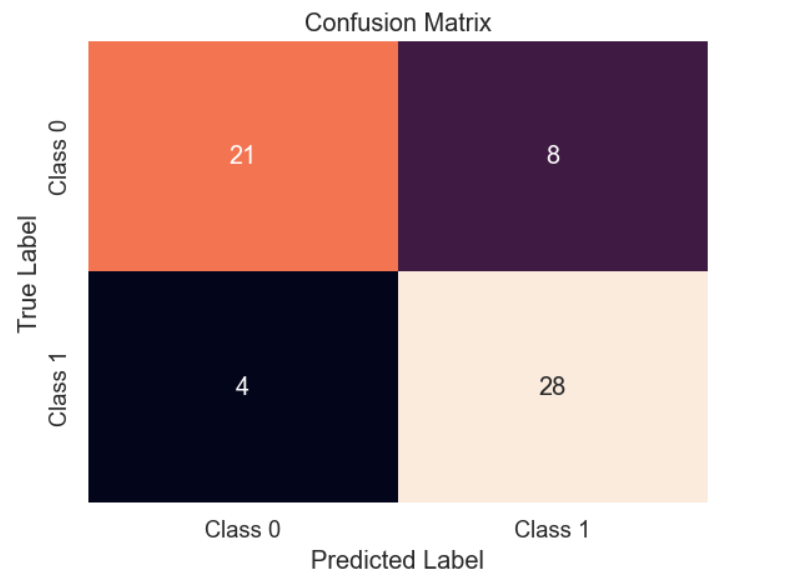






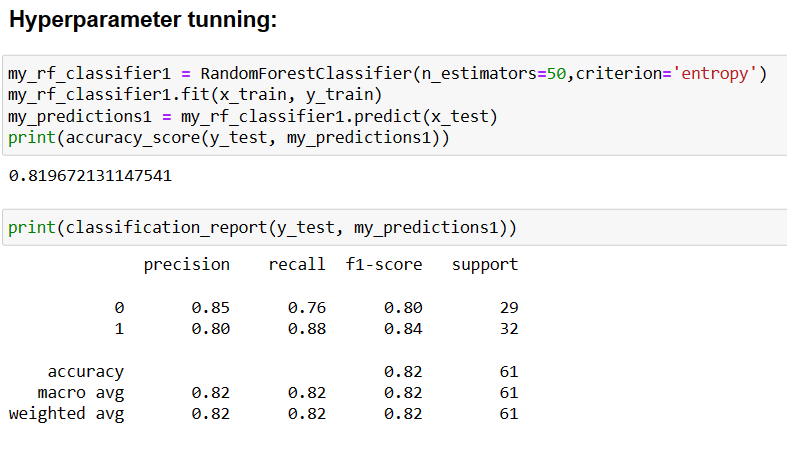






**Some important results :**

* **Performance using Logistic Regression : 80.3279 % .**
* **Performance using Random Forest: 80.32786 % .**
* **Both the models are performing equally.**



**Some important results :**

* **After hyperparameter tunning , performance using Random Forest is : 81.9672 %.**