Lab 8: ROC Curve Analysis for Logistic Regression

The data file **Credit Default.xlxs** contains data from customers of a credit institution (adapted from the Kaggle.com “Give me Credit” data set).

Variable Name	**Description**
ID	Borrower’s ID
Default	Default (1= yes, 0 = no)
Age	Age
DebtRatio	Debt ratio
YearlyIncome	Yearly income (thousands of euro)
LatePayment	The borrower had a late payment in the last 2 years (1= yes, 0 = no)

Questions:

1. Fit a logistic regression model where Default is the dependent variable of interest by including Age, YearlyIncome and LatePayment as predictors. Interpret the results of the logistic regression model.

2. Save the predicted probabilities to the dataset.

3. Fit an ROC curve to the predicted probabilities and interpret the output.

4. What percentage of defaulters are correctly classified by the logistic regression algorithm if a cut-point a predicted probability of 0.5 is used to predict defaulters and non-defaulters.

5. Identify an optimal cut-point for classifying defaulters based on a sensitivity value of 75%. What percentage of the non-defaulters would be misclassified if this cut-point was be used?