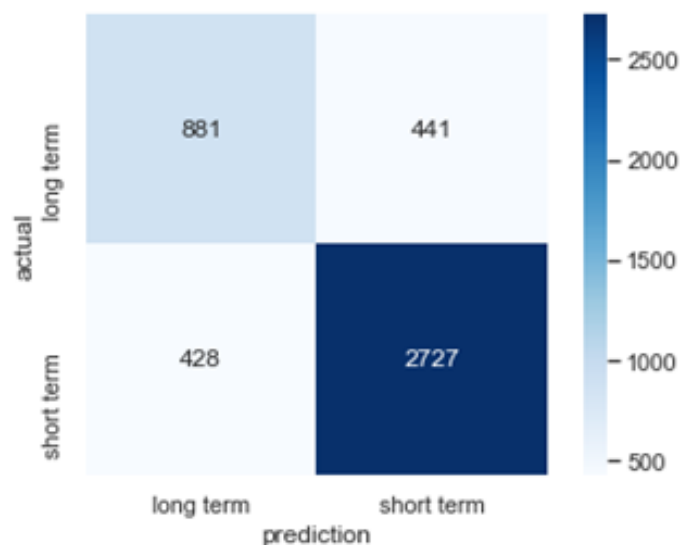


## Loan Prediction Analysis

The goal of this project is to predict the type of loan the bank should give to a customer whether a long-term loan or short-term loan by a classification model.

To start the classification, work a baseline model was needed to compare the affect of any modification in the data and the model. After cleaning the data by dropping duplicated rows and the outliers, the baseline model has been chosen to be Logistic Regression model and it gave 0.72 training accuracy and 0.73 for the validation.



The figure above shows the confusion matrix of the validation data for the baseline model, which compare the actual and the predicted value. It shows the four conditions:

True positive when the model predicts long term correctly (left top).

False positive when the model predicts long term while the actual value is short term (left bottom).

False negative when the model predicts short term while the actual value is long term (right top).

True negative when the model correctly predicts short term. (right bottom)

As the colors become darker that means more values, and it is clear that the baseline model has failed to predict some values.



The second figure is called ROC AUC curve. It represents how good the baseline predictive model is. When the threshold is large that means more values on the True positive (aggressive), and the lower threshold means more False positive value (more safe). The result that above figure shows is the balance between True positive and False positive where the value of the threshold on the baseline model is at the intersection point of the recall and precision.

To increase the accuracy of the baseline model, the complexity of the model should be increased by adding more features (feature engineering) and retrain the model until it reaches a point that cannot be improved any more. After that, multiple classification models will be evaluated based on their accuracy, recall, precision and f1 score, then the best model will be tested and chosen.