Classifing Bank Customers who are most likely to churn
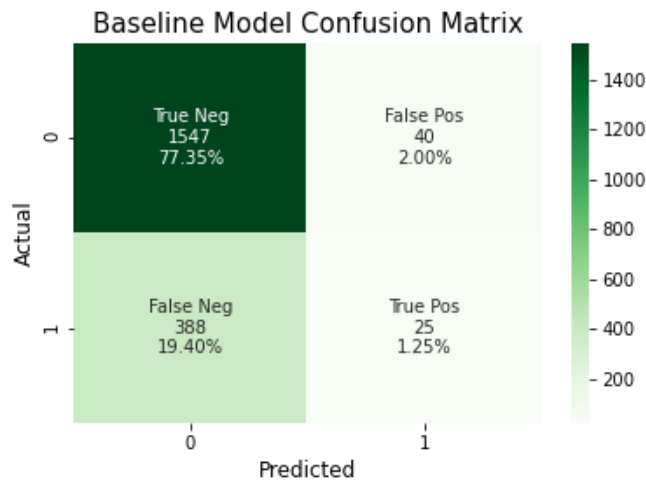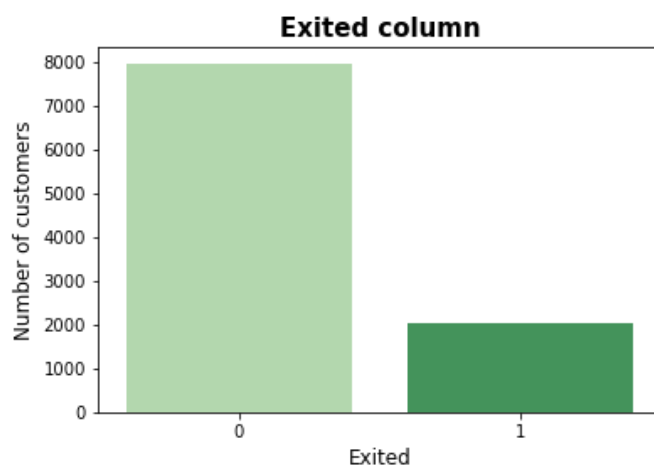

The goal of this project is to build an accurate model to classify the bank customers who are most likely churn (close their accounts). Above is the confusion matrix of the baseline model, which was created using Logistic Regression. The accuracy of the baseline model is: 0.786, which is not bad. But the F1 score is 0.104, which is very low.

The bar chart above shows that the target column (Exited) is imbalanced. Which justifies the low F1 score of the baseline model. This can be fixed by solving class imbalance, and by applying further feature engineering.