Credit Card Users Churn Prediction

Business Presentation

Contents

Business Problem Overview and Solution Approach

Thera Bank saw a steep decline in the number of users of their credit card, which could be a good source of income for banks.

This leads Thera Bank to loss, hence Thera Bank wants to analyze the data to identify customers who will leave the credit card services so that Thera Bank could improve on those areas. Classification models are required for the case.

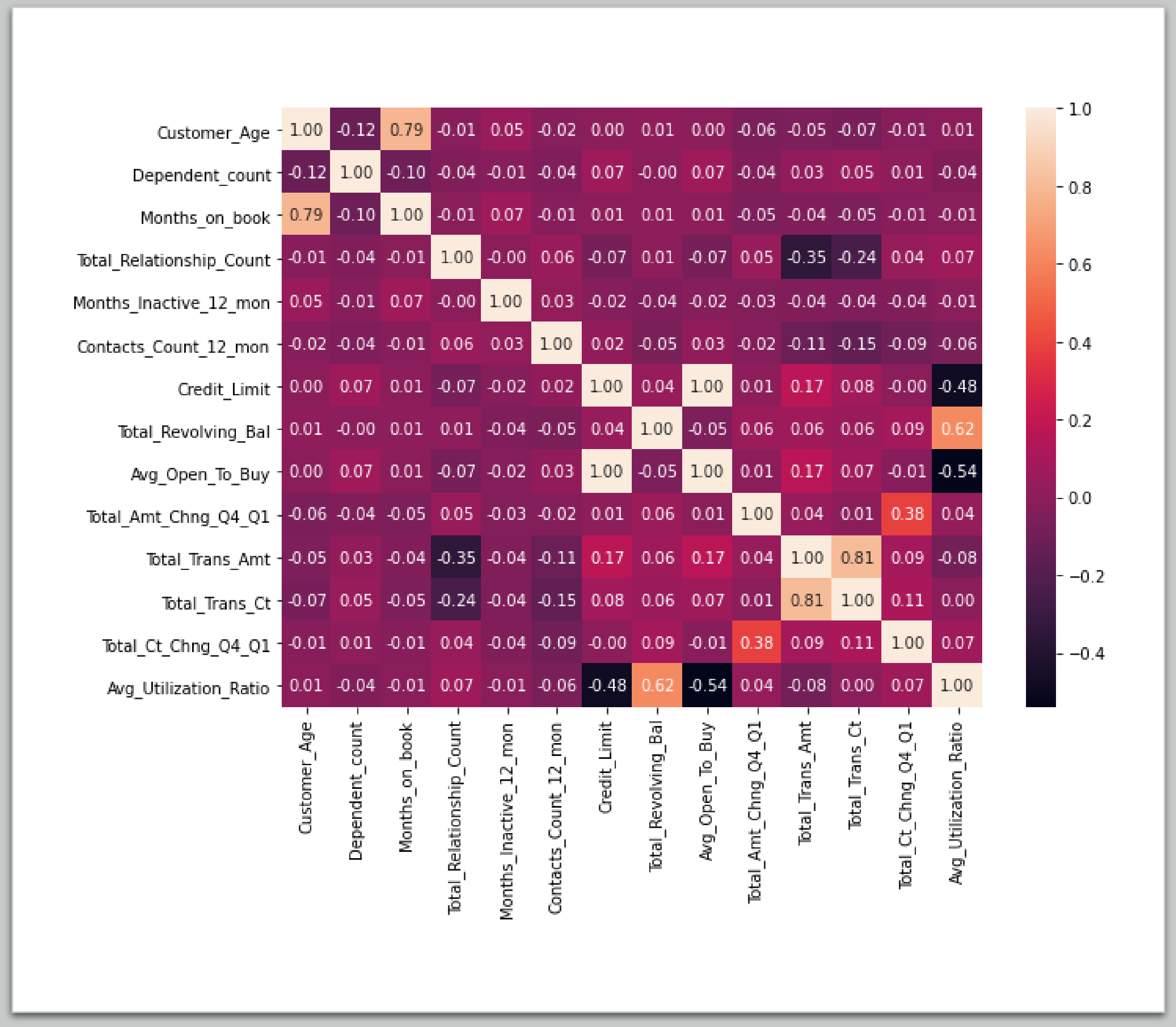
The metric of interest in model building would be recall. This is because we want to have as little false negatives as possible (false negatives are cases where the model fails to identify customers who want to renounce the credit card service), as more false negatives could lead the company to greater losses.

Data Overview

- The data contains information about 10127 customers and 21 variables.
- The data include customer information such as age, gender, no of dependents, education level, marital status, income category and card category.
- The data also has customer statistics such as period of relationship with bank, no of products held, no of months inactive, no of contacts between customer & bank, credit limit, balance carry over, open to buy, transaction amount & count, ratio of transaction amount & count in 4th quarter & 1st quarter, and average utilization ratio.
- The CLIENTNUM variable from the data will be dropped as it is unique for each customer and will not add value to our analysis.
- There are no missing values from the dataset.
- There are outliers in several variables in the dataset, which have been treated by removing those outside the interquartile range.
- Avg_Open_To_buy variable is removed from the data as it has perfect correlation with Credit_Limit. This is to prevent multicollinearity inside the model.

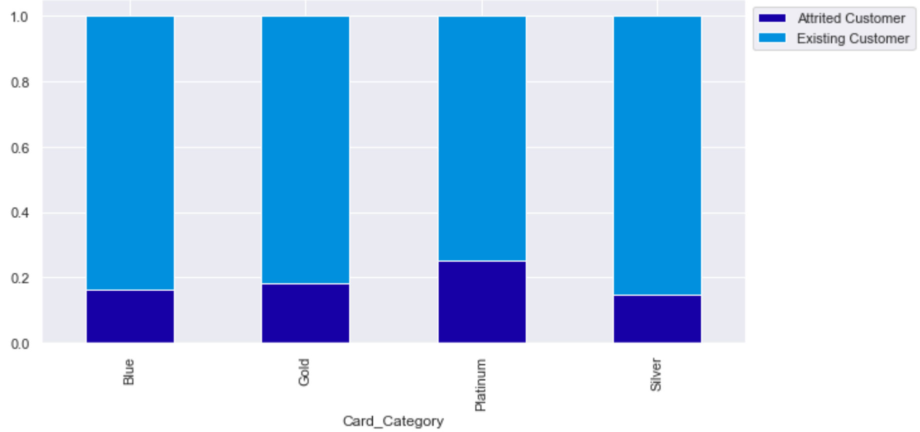
Exploratory Data Analysis

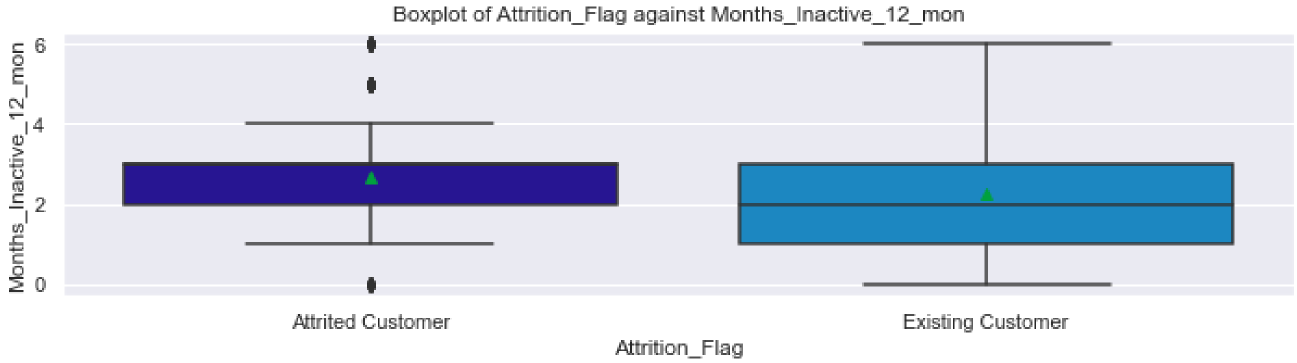
- The most important findings will be highlighted in this section.
- From the heatmap, it can be seen that Avg_Open_To_buy has perfect correlation with Credit_Limit (hence the removal of one of them).
- Total_Trans_Ct and Total_Trans_Amt are strongly correlated.
- Months_on_book and Customer_Age are strongly correlated.



Exploratory Data Analysis

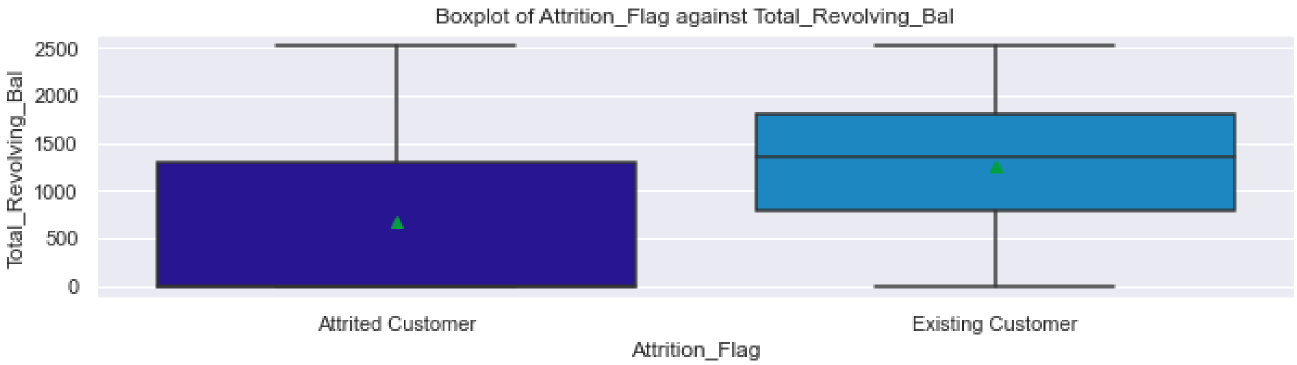
- There aren’t much difference between the percentage of attrited customers among categorical variables, except that percentage of attrited customers among platinum card holders are higher than others.
- Attrited customers have slightly higher number of Months_Inactive_12_mon.

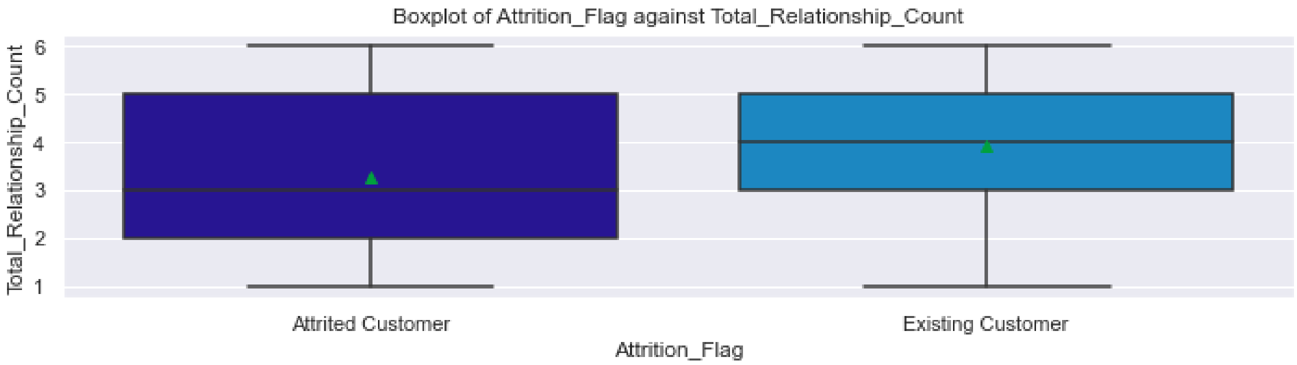




Exploratory Data Analysis

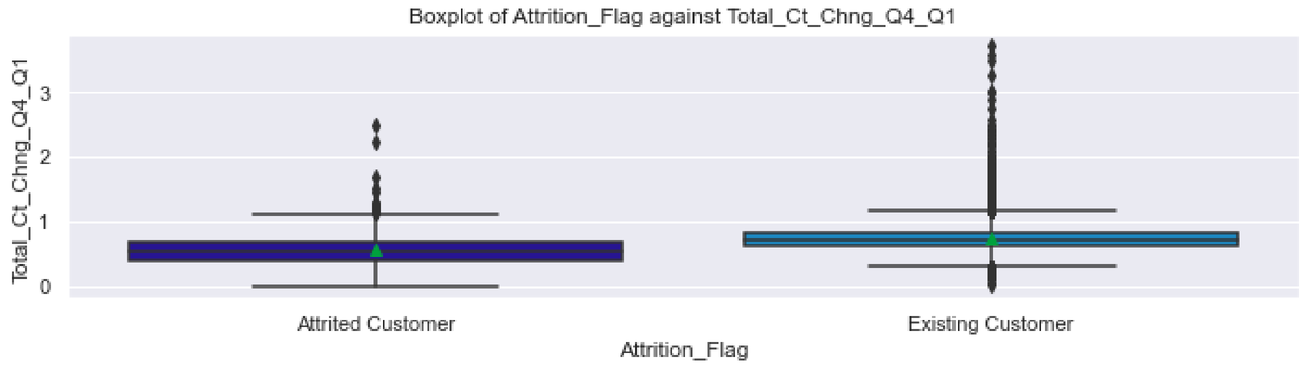
- Attrited customers have lower Total_Relationship_Count and Total_Revolving_Bal.

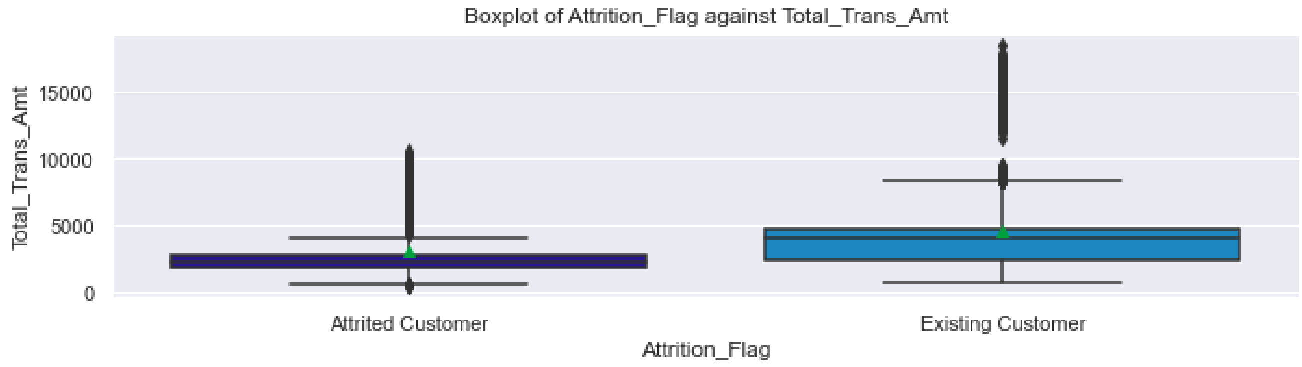




Exploratory Data Analysis

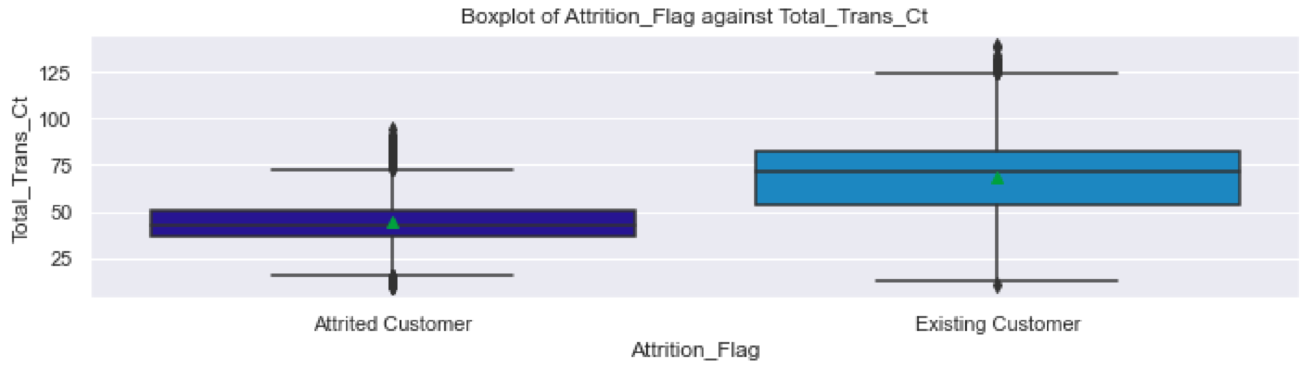
- Attrited customers have significantly lower Total_Trans_Amt and Total_Ct_Chng_Q4_Q1.

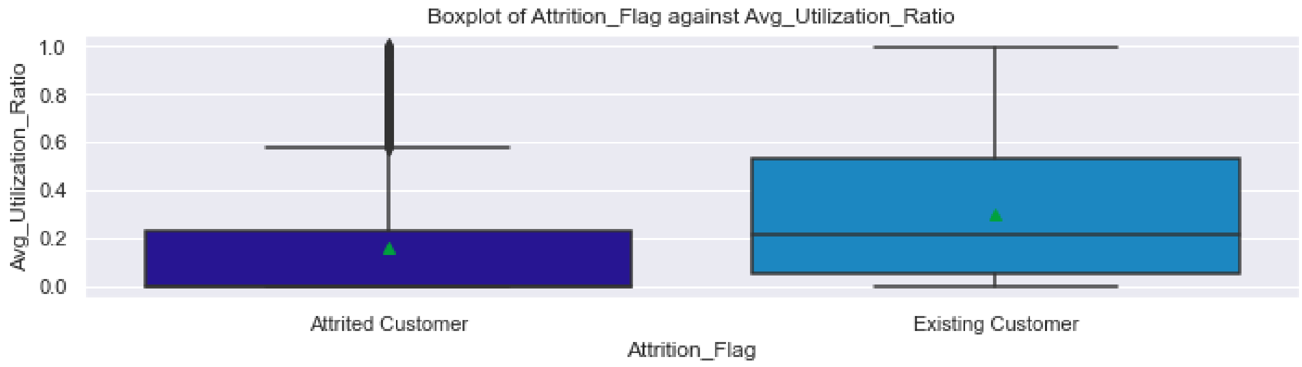




Exploratory Data Analysis

- Attrited customers have significantly lower Total_Trans_Ct and Avg_Utilization_Ratio.





Model Performance Summary

- Recall will be used as the metric for model evaluation as false negatives must be minimized to prevent losses to the bank.
- Different models will be built, such as Logistic Regression, Logistic Regression with Upsampling and Downsampling, Decision Tree, Bagging Classifier, Random Forest Classifier, Adaboost Classifier, Gradient Boosting, and XGBoost.
- Tuning will be done to the top three most performing models by using GridSearchCV and RandomizedSearchCV.

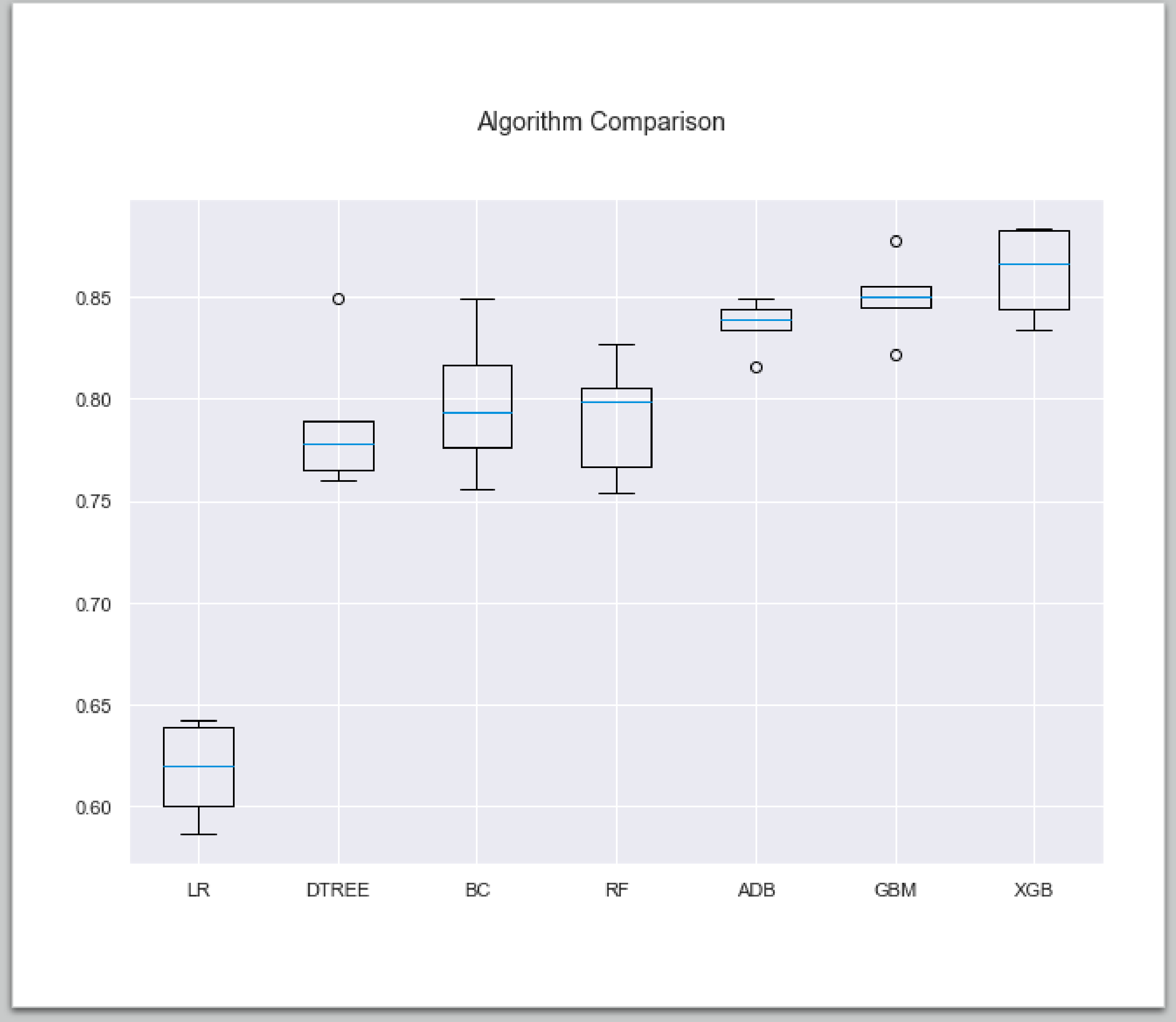
Model Performance Summary (Logistic Regression)

After testing various logistic regression models, it is found that logistic regression with IMBLearn Random Oversampling gives the best recall score.

Model	Train_Recall	Test_Recall
Logistic Regression	0.55	0.53
Logistic Regression with SMOTE	0.84	0.83
Logistic Regression with IMBLearn Random Undersampling	0.84	0.84
Logistic Regression with IMBLearn Random Oversampling	0.86	0.86
Logistic Regression with TomekLinks	0.56	0.55

Model Performance Summary

- Pipelines are applied to the models.
- Cross Validation is applied with number of splits equal to 5.
- After comparison, it is found that the model that yields the highest recall is **XGBoost**, followed by Gradient Boosting and Adaboost.



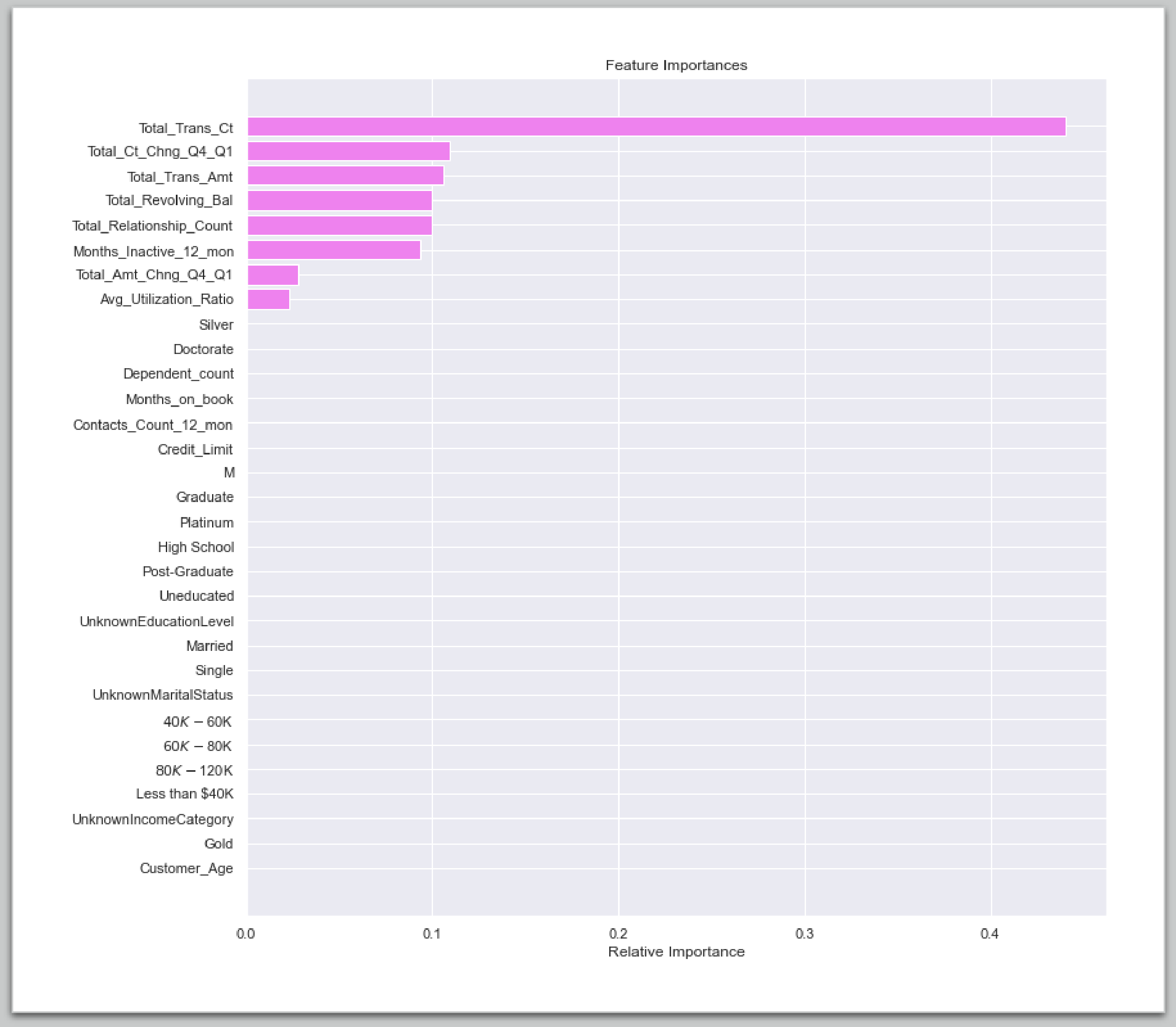
Model Performance Summary

- Bagging Classifier is added to the list of models tuned (due to long waiting time of XGBoost tuning with GridSearchCV).
- After comparison, it is found that the model with highest recall is **XGBoost with RandomizedSearchCV**, followed by XGBoost with GridSearchCV and Gradient Boosting with RandomizedSearchCV.
- In terms of overall performance, it seems that Gradient Boosting with RandomizedSearchCV is the best.

	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision	Train_F1	Test_F1
6	XGBoost with RandomizedSearchCV	0.76	0.75	0.98	0.97	0.41	0.40	0.58	0.57
2	XGBoost with GridSearchCV	0.92	0.91	0.99	0.95	0.69	0.66	0.81	0.78
5	Gradient Boosting with RandomizedSearchCV	0.99	0.97	0.97	0.91	0.99	0.92	0.98	0.91
1	Gradient Boosting with GridSearchCV	1.00	0.97	0.99	0.90	0.99	0.91	0.99	0.90
4	Adaboost with RandomizedSearchCV	0.99	0.97	0.97	0.89	0.98	0.92	0.97	0.91
0	Adaboost with GridSearchCV	1.00	0.97	0.99	0.88	0.99	0.92	0.99	0.90
3	Bagging Classifier with GridSearchCV	1.00	0.96	1.00	0.86	1.00	0.91	1.00	0.89

Feature Importances

- The first most important feature is Total_Trans_Ct.
- The second most important feature is Total_Ct_Chng_Q4_Q1.
- The third most important feature is Total_Trans_Amt.
- Categorical variables hardly contribute to the model.



Business Insights & Recommendations

- Our analysis indicates that those who renounces credit card have:
  - Low amount of total transaction count & amount over the last 12 months.
  - Low amount of ratio of total transaction count & amount in 4th quarter and total transaction count & amount in 1st quarter.
  - High levels of inactivity in the last 12 months.
  - Low levels of available credit the customer spent.
- An early detection system for the above measures of the customer can be implemented. A certain threshold can be set, and if the numbers are below/above that threshold, the bank can consider providing special promotions to make sure customers retain their credit cards.
- An introduction of step-up bonus, which will be given to customers when the balance on a month is a certain number above the balance on the month before, can be implemented. This is because Total_Revolving_Bal is an important contributor to customers renouncing their credit card.
- The bank can also market more products to customers who have a low number of products held. This will make customers more attached to the bank and more willing to use the products & services the bank used, which includes credit cards.