

Credit Card Users Churn Prediction

Business Presentation

Contents

- Business Problem Overview
- Data Overview
- Exploratory Data Analysis (EDA)
- Model Performance Summary
- Business Insights and Recommendations

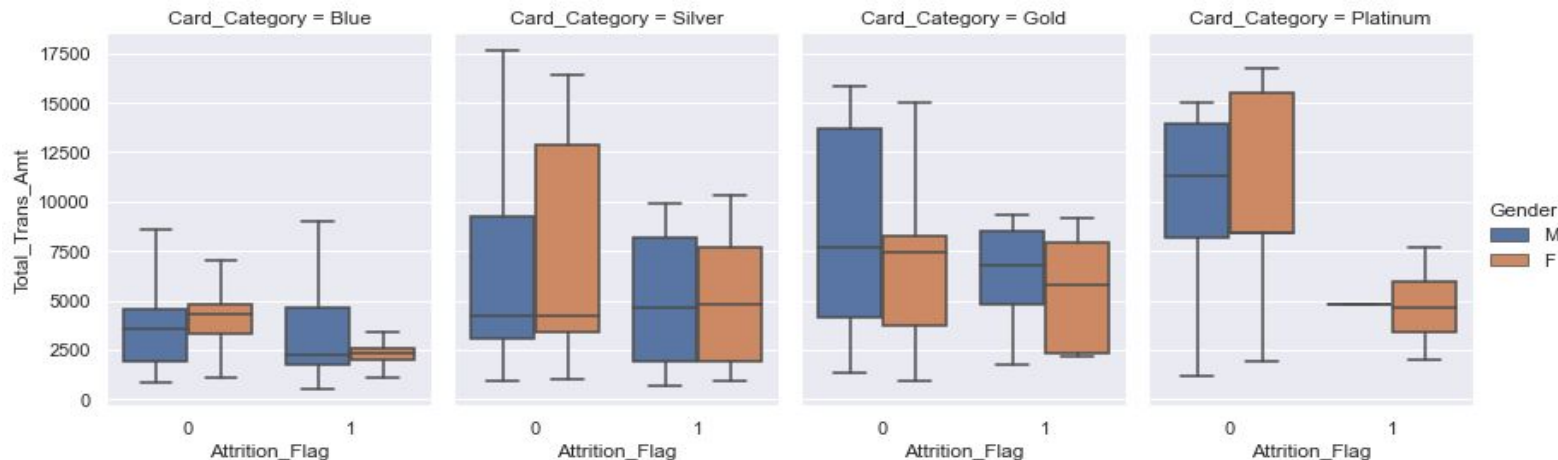
Business Problem Overview and Solution Approach

- The Thera bank recently saw a steep decline in the number of users of their credit card, credit cards are a good source of income for banks because of different kinds of fees charged by the banks like annual fees, balance transfer fees, and cash advance fees, late payment fees, foreign transaction fees, and others. Some fees are charged to every user irrespective of usage, while others are charged under specified circumstances.
- Customers' leaving credit cards services would lead bank to loss, so the bank wants to analyze the data of customers' and identify the customers who will leave their credit card services and reason for same – so that bank could improve upon those areas
- We'll build different models and use appropriate tuning techniques to get a final model that will help the bank predict which customers have a higher chance of leaving the bank services

Data Overview

- The data contains information about 10127 customers and their characteristics
- The characteristics include client number, age, gender, count of dependents, education level, marital status, income category, card category, period of relationship with the bank, number of products held, inactive months and number of contacts in past 12 months, credit limit, total revolving balance, average open to buy, change in transaction amount, total transaction amount (Last 12 months), Total Transaction Count (Last 12 months), change in transaction count(Q4 over Q1), avg utilization ratio
- 2 columns have missing values and 1 column have 'abc' values- we'll treat them as missing values

Exploratory Data Analysis



Blue Card

Female customers have higher total transaction amounts, for existing customers. Those female customers who spent less than 5000 using the blue cards attrited.

Silver Card

There's not much difference between attrited and not-attrited customers or male and female customers.

Customers spending less than 10,000 attrited.

Gold Card

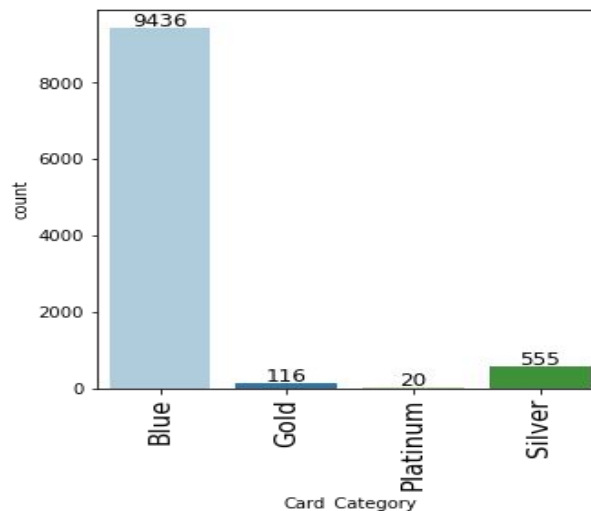
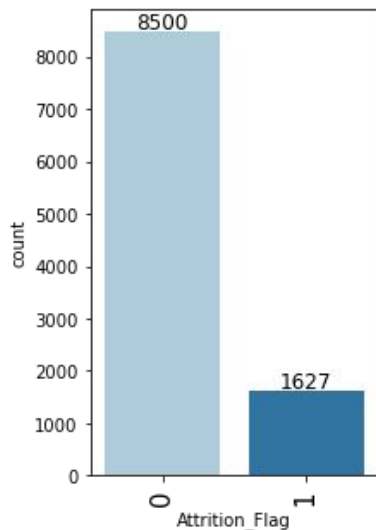
The spending for both male and female customers is the same. Customers spending less than 9000 attrited.

Platinum Card

Both males and females have equal spendings. Those customers who spent less than 7500 attrited.

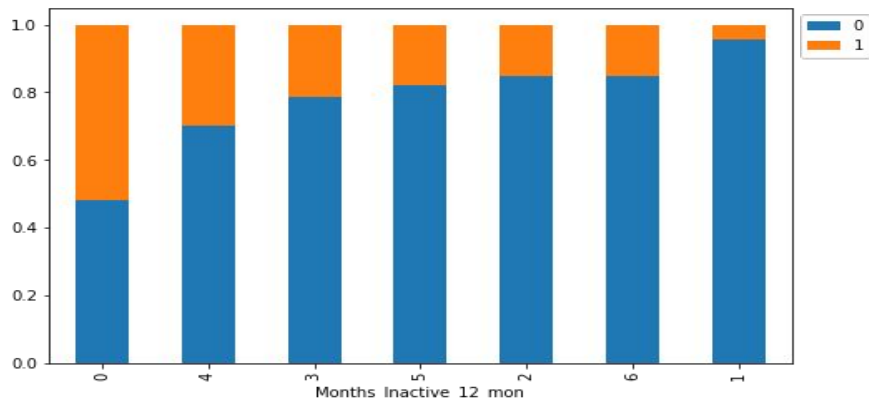
Exploratory Data Analysis

- 93.2% of the customers have a Blue card.
- The Blue card would be a standard card given by the bank to all its customers.



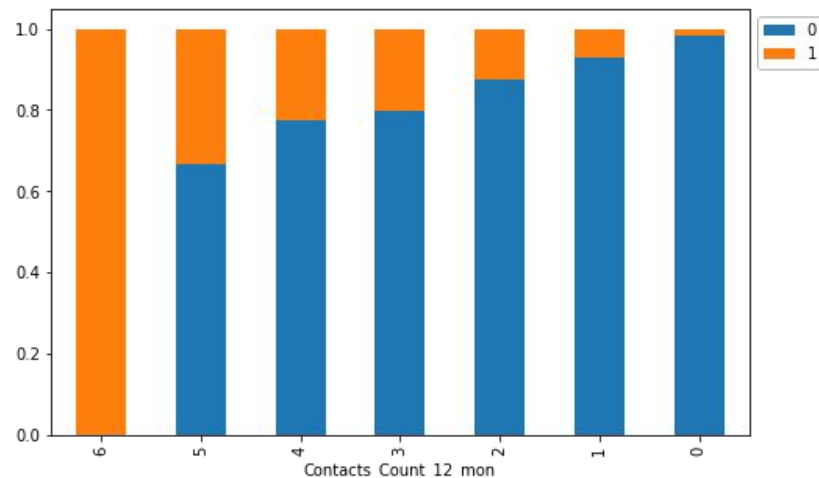
- 16.1% of the customers attrited.
- This indicates an imbalance in the data.

Exploratory Data Analysis



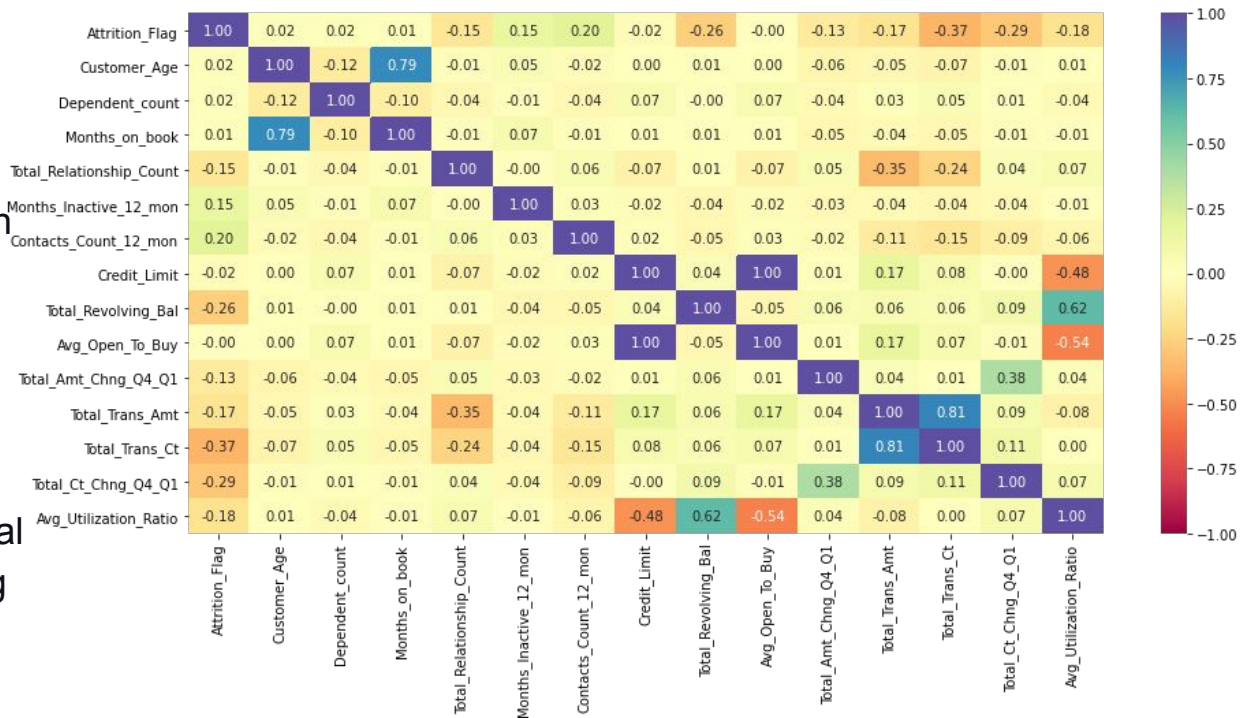
- The highest attrition is among the customers who interacted the most with the bank.
- This signifies that the bank is not able to resolve the problems faced by customers leading to attrition
- A preliminary step to identify attriting customers would be to look out for customers who have reached out to them repeatedly.

- As inactivity increases attrition also increases (2-4 months)
- The interpretation from here for 0 months and 6 months is difficult as customers who recently used the card attrited the most while those who were inactive for 6 months attrited less.



Exploratory Data Analysis

- The attrition flag shows a bit of a negative correlation with total transactions count and total transaction amount
- There's a strong positive correlation between months on book and customer age, total revolving_Bal and Avg utilization ratio, total trans amt and Total trans count
- There's a negative correlation of Total relationship count with the total trans amt and total trans count, avg utilization ratio with the credit limit and avg open to buy.



Model Building - steps

- Split the dataset into train, test and validation sets
- Impute missing values using mode
- Choose Model performance evaluation metric
- Build different models - Logistic Regression, DecisionTreeClassifier, RandomForestClassifier, BaggingClassifier, AdaBoostClassifier, GradientBoostingClassifier, XGBClassifier
- Tune hyperparameters using RandomSearchCV
- Train the tuned models using oversampled and undersampled data
- Compare all models' performance on train and validation data
- Choose a final model and find the performance on test data
- At last create a pipeline to productionize the final model

Comparing Model Performance

Training Performance

	Tuned XGBoost trained with original data	Tuned XGBoost trained with oversampled data	Tuned XGBoost trained with undersampled data	Tuned Gradient boosting trained with original data	Tuned Gradient boosting trained with oversampled data	Tuned Gradient boosting trained with undersampled data	Tuned AdaBoost trained with original data	Tuned AdaBoost trained with oversampled data	Tuned AdaBoost trained with undersampled data
Accuracy	0.946	0.909	0.912	0.988	0.989	0.996	0.995	0.993	1.000
Recall	0.999	1.000	1.000	0.946	0.990	0.998	0.980	0.994	1.000
Precision	0.749	0.846	0.851	0.980	0.987	0.994	0.990	0.992	1.000
F1	0.856	0.917	0.919	0.962	0.989	0.996	0.985	0.993	1.000

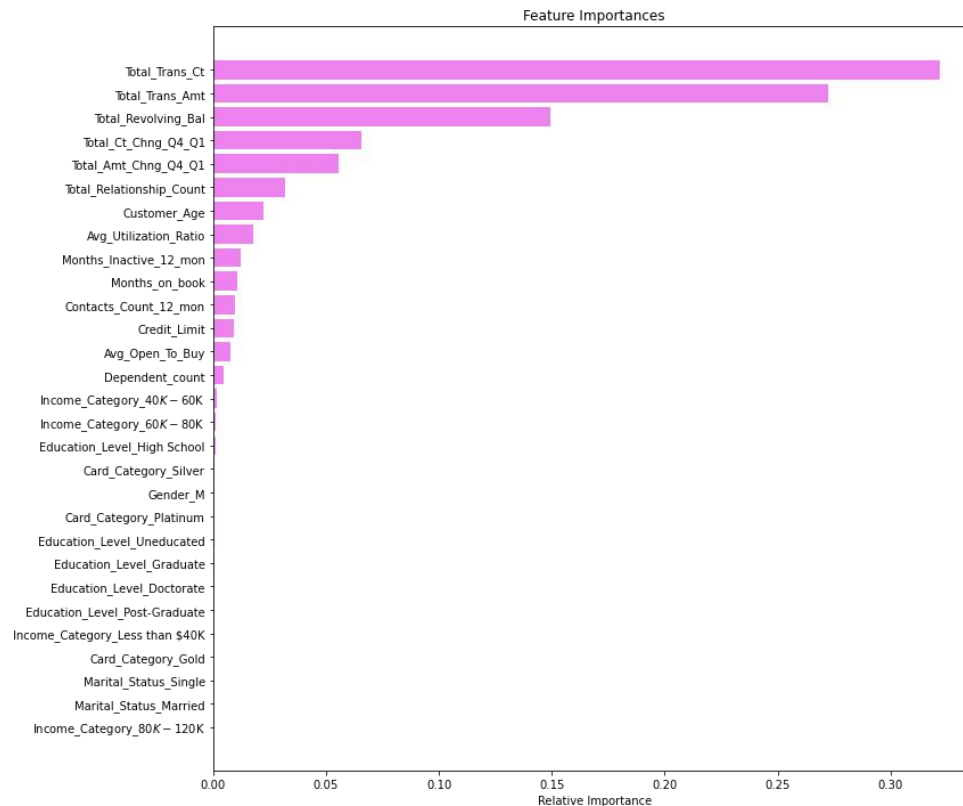
	Tuned XGBoost trained with original data	Tuned XGBoost trained with oversampled data	Tuned XGBoost trained with undersampled data	Tuned Gradient boosting trained with original data	Tuned Gradient boosting trained with oversampled data	Tuned Gradient boosting trained with undersampled data	Tuned AdaBoost trained with original data	Tuned AdaBoost trained with oversampled data	Tuned AdaBoost trained with undersampled data
Accuracy	0.930	0.840	0.818	0.969	0.966	0.945	0.967	0.967	0.941
Recall	0.960	0.972	0.988	0.880	0.914	0.957	0.871	0.911	0.963
Precision	0.710	0.502	0.469	0.926	0.882	0.763	0.922	0.889	0.746
F1	0.816	0.662	0.636	0.903	0.898	0.849	0.896	0.900	0.841

Gradient boosting model trained with undersampled data has best performance.

Validation Performance

Feature Importance

- Total_Trans_Ct is the most important variable in predicting credit card churn followed by Total_Revolving_Bal, Total_Trans_Amt, Total_Relationship_Count and Total_Ct_Chng_Q4_Q1 .



Business Insights and Recommendations

- Factors that drive the attrition - Total_Trans_Ct, Total_Revolving_Bal, Total_Trans_Amt, Total_Relationship_Count
- Total_Trans_Ct: Less number of transactions in a year leads to attrition of a customer -
 - a) To increase the usage of cards the bank can provide offers like cashback, special discounts on the purchase of something, etc so that customers feel motivated to use their cards.
 - b) The cards can be made flexible for transactions so that customers can use them at more places (like shops, malls, etc).
- Total_Revolving_Bal: Customers with less total revolving balance are the ones who attrited, such customers must have cleared their dues and opted out of the credit card service. The bank should investigate it further to find the cause
- Total_Trans_Amt: Less number of transactions can lead to less transaction amount and eventually leads to customer attrition - Bank can provide offers on the purchase of costlier items which inturn will benefit the customers and bank both.

Business Insights and Recommendations

- **Total_Relationship_Count:** Attrition is highest among the customers who are using 1 or 2 products offered by the bank - together they constitute ~55% of the attrition - Bank should investigate here to find the problems customers are facing with these products, customer support, or more transparency can help in retaining customers.
- Female customers should be the target customers for any kind of marketing campaign as they are the ones who utilize their credits, make more and higher amount transactions. But their credit limit is less so increasing the credit limit for such customers can profit the bank.
- **Months_Inactive:** As inactivity increases the attrition also increases, 2-4 months of inactivity are the biggest contributors of attrition -Bank can send automated messages to engage customers, these messages can be about their monthly activity, new offers or services, etc.
- Highest attrition is among the customers who interacted the most with the bank, This indicates that the bank is not able to resolve the problems faced by customers leading to attrition - a feedback collection system can be set-up to check if the customers are satisfied with the resolution provided, if not, the bank should act upon it accordingly.

greatlearning
Power Ahead

Happy Learning !

