Thera Bank

Contents

Business Problem Overview

The Thera Bank recently experienced a significant decline in the number of users for their credit cards. Since credit cards are a steady source of income, the bank wishes to figure out which of the customers are most likely to renounce their credit cards

Solution Approach :

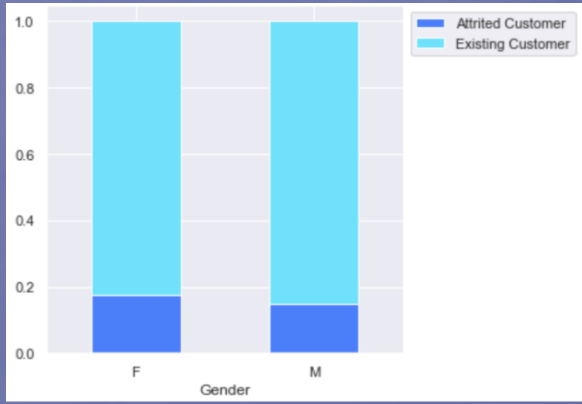
- To build and optimize a classification model to identify the potential customers
- To analyze the data and generate a set of insights and recommendations to improve its services

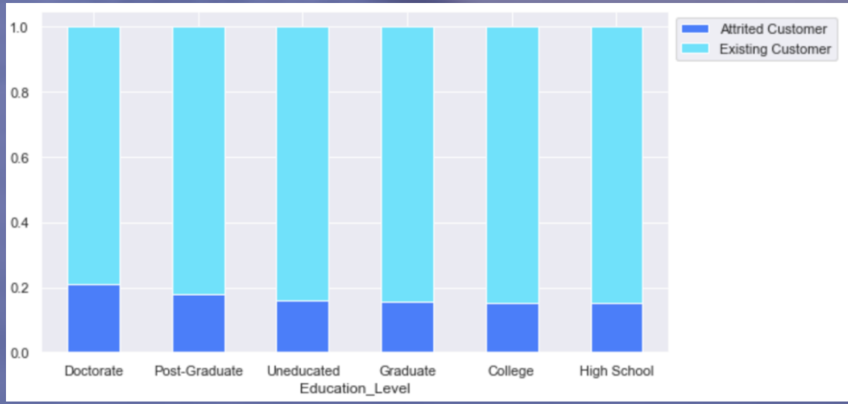
Data Overview

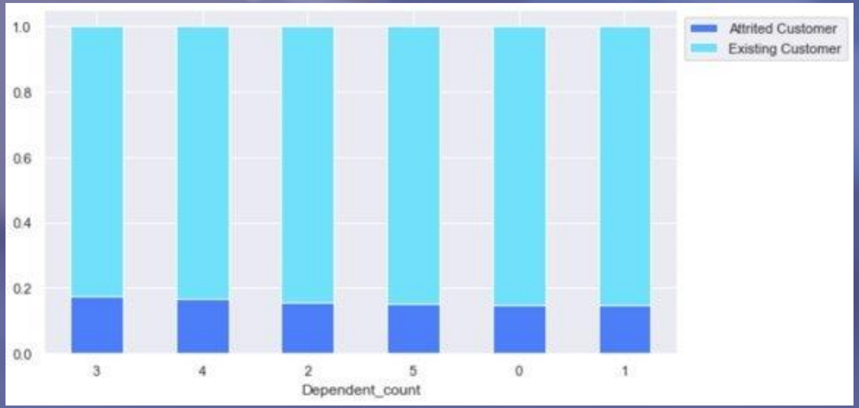
- The data contains information of 10127 customers
- Customer information includes:
  - Age, Gender, Education_Level, Marital Status, Income_Category and Dependent Count
  - Information on customer's relationship with the bank, type of credit card, credit limit, transaction amount, transaction count, revolving balance, the corresponding ratios in the 4th and 1st quarters and the ratio of available credit utilized by the customer
- Some customers may choose not to disclose their Education, Marital Status and Income
  - Hence imputed the missing values in Education_Level and Marital_Status with 'Unknown'
  - Replaced the 'abc' in the Income_Category as 'Unknown' and treated it as a new category
- Excluded the Avg_Open_To_Buy column while model building as this feature was in perfect correlation with Credit Limit

Exploratory Data Analysis

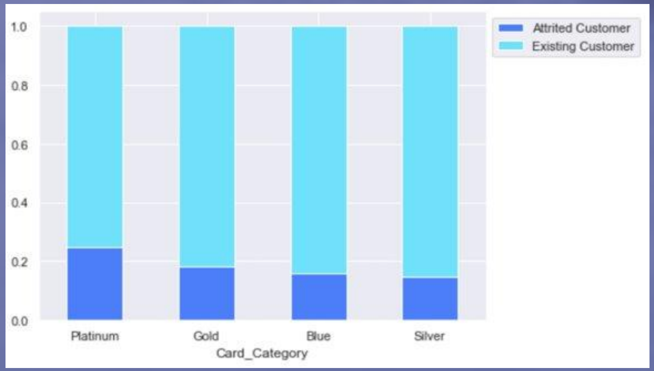
- 53% of the total customers are Female and 17.4% of them have attrited
- 31% of the customers are graduates and 15.5% of them have attrited
- Out of 4.5% of Doctorate customers, about 21% of them have attrited
- Out of 5.1% of Post-Graduate customers, 17.8% of them have attrited
- Uneducated customers and customers with high school education show 15% attrition rate
- Large number of customers have either 3 or 2 dependents and also show a higher attrition rate

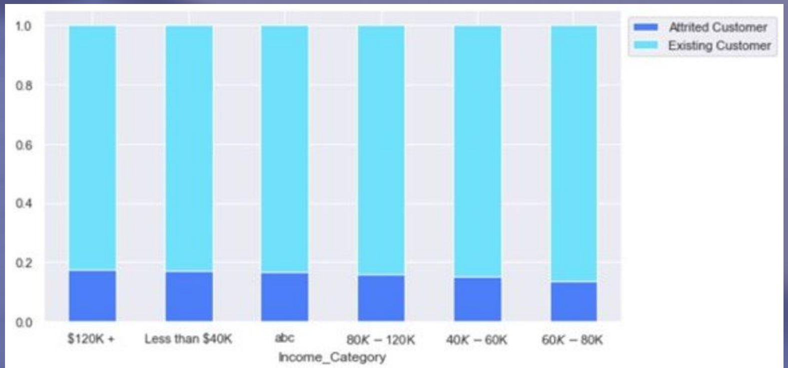


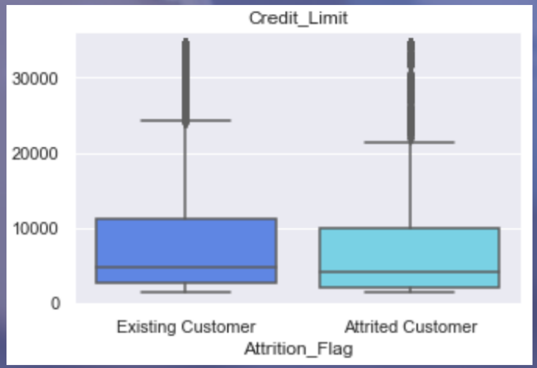




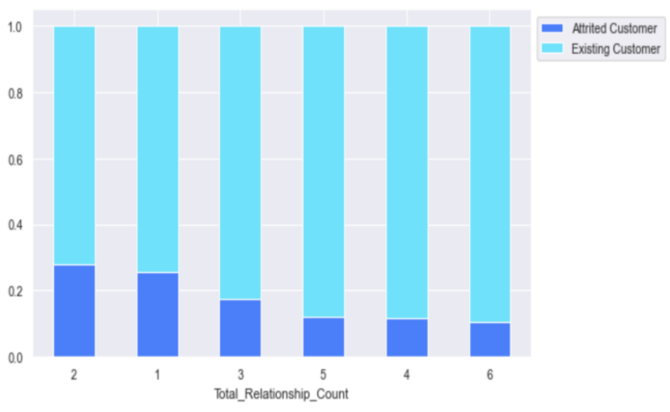
- 'Blue' card is the most common among, and is largely owned, by customers with an annual income less than $40K
- 18% of 'Gold', 16% of 'Blue' and 15% of 'Silver' card customers have attrited in the past 12 months
- Out of 20 'Platinum' card holders, five have renounced their cards
- 17% of customers with income above $120K, 16.8% of customers with income 'Unknown', and 15.8% of customers with income between $80-$120K have attrited in the past
- Customers with lower Credit Limit tend to have a higher attrition rate

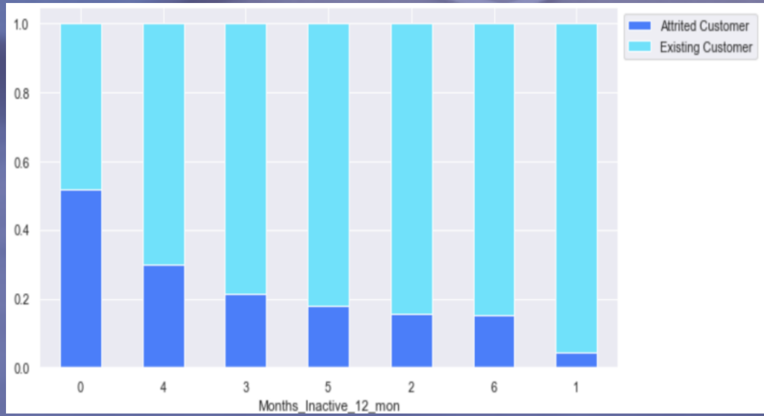


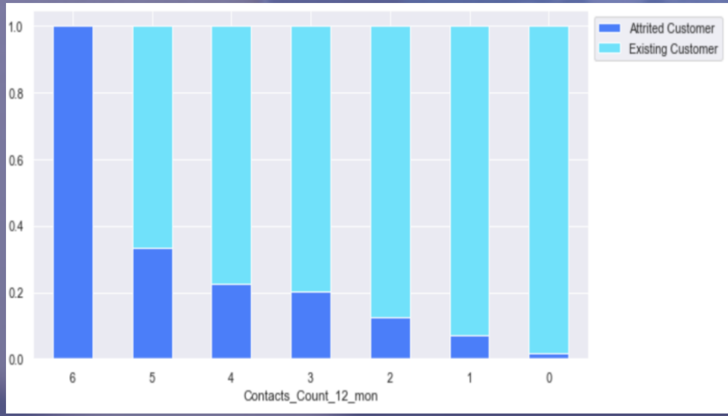




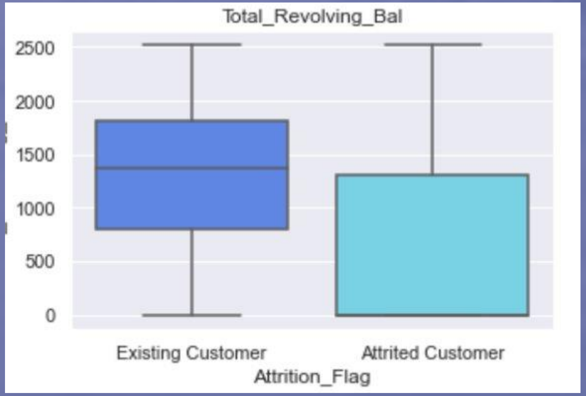
- Customers holding less number of products with the bank have a higher attrition rate
- Larger the number of months a customer is inactive, higher the chances to renounce the card
- Larger the number of contacts with the bank, higher the chances of the customer to attrite

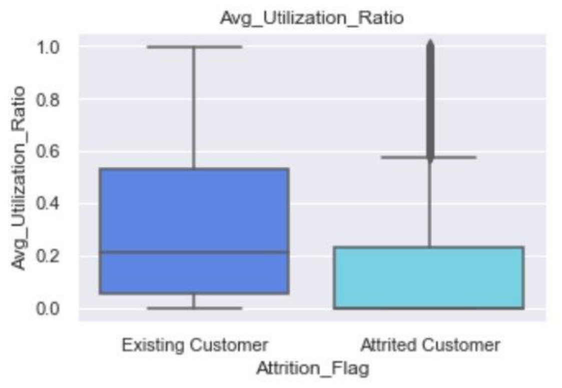


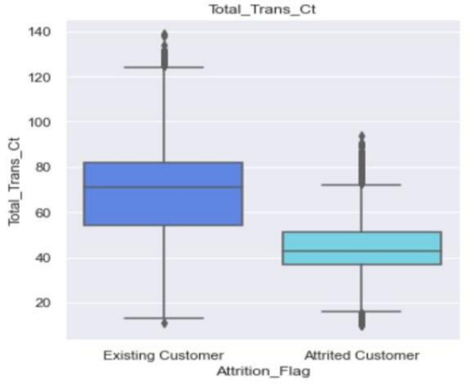


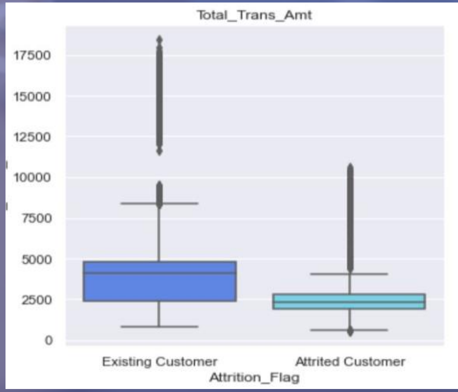


- Customers with lower revolving balance had shown high attrition in the past 12 months
- Customers who spend very less available credit seem to have higher attrition rate
- Customers with lower transaction amount have renounced their credit cards
- As Total_Tans_Ct is derived from Total_Tans_Amt, it follows the same pattern as the latter
- Lower the Total_Amt_Chng_Q4_Q1 and Total_Ct_Q4_Q1 ratio, higher the chances that the customer will attrite









Key Insights from EDA

- About 52% of customers are Female
- Large number of customers have either 3 or 2 dependents
- More number of graduate customers, followed by high school and uneducated customers
- Large number of customers are married
- Most customers have an average annual income less than 40K
- About 93% of customers hold the 'Blue' card. This being the basic card issued by the bank, it is affordable to the customers with income less than 40K
- 'Blue' card is the basic credit card and it is owned by maximum number of customers
- 'Platinum' card is owned by least number of customers

Key Insights continued...

Attrition rate is more common among :

- Female customers
- Customers with three or less number of dependents
- Postgraduate or Doctorate customers
- Married customers with an income less than 40K and who own a 'Blue' card
- Customers with a large number of contacts with the bank
- Customers holding less number of products with the bank
- Customers who do not have any revolving balance left
- Cutomers having 1K or less balance amount left on their credit card
- Customers who do not utilize the credit available on their card
- Customers with a lrger monthly inactivity number

Model Evaluation Criterion

Model can make wrong predictions such as :

- Predicting a customer will renounce their credit card but in reality the customer will not
- Predicting a customer will continue with their credit card services but in reality the customer will not

Prediction of concern :

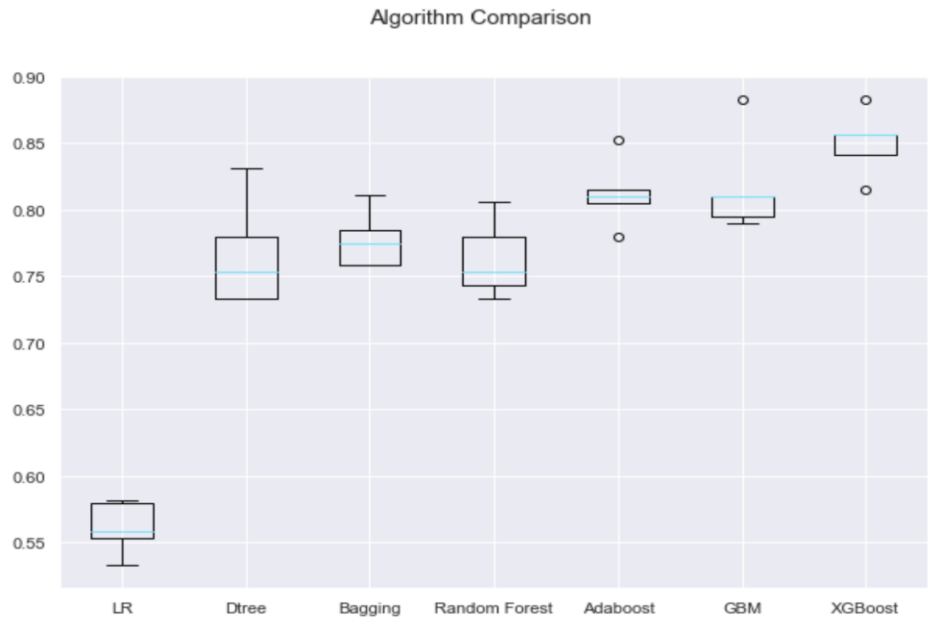
- The second prediction is our major concern as customers' leaving credit card services would lead the bank to loss and our aim is to build a model to help the bank improve its services so that the customers do not renounce their credit cards.

How to reduce False Negatives :

- **Recall** score should be maximized. Greater the Recall score, higher the chances of predicting the potential customers who may renounce their credit cards.

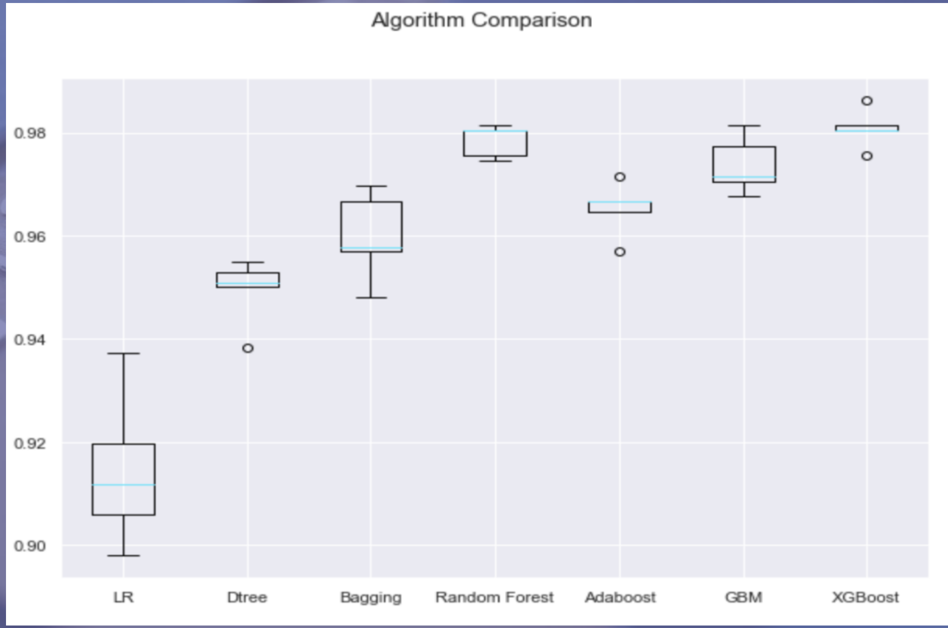
Model Performance – Original Data

- XGBoost is giving the highest K-Folds CV Recall follower by GBM and AdaBoost Classifiers
- However, all these models are slightly overfit
- Although, Logistic Regression has the lowest CV score, it is generalizing well on train and validation sets
- Highest Validation Recall achieved is 89



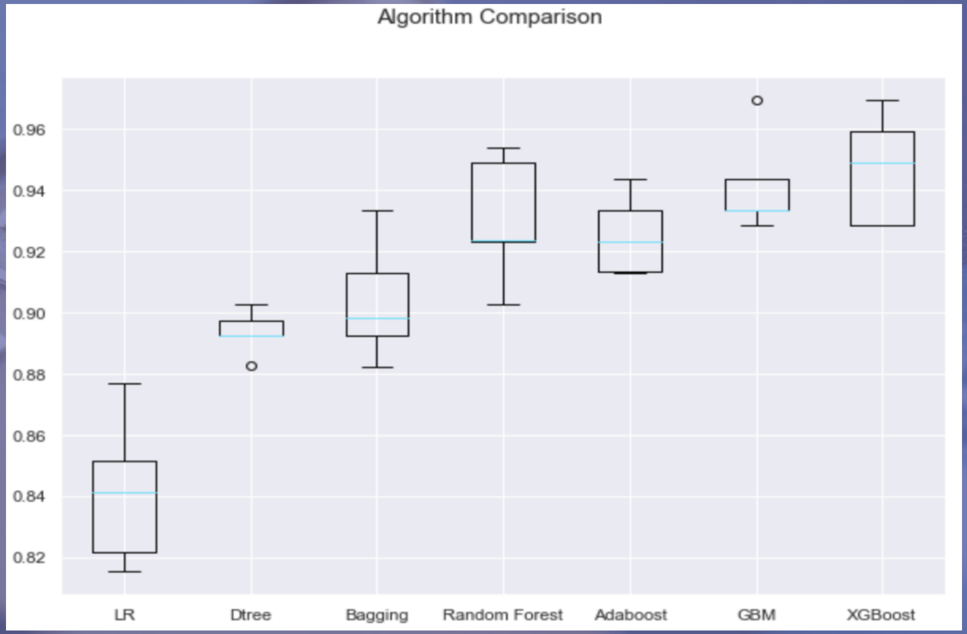
Model Performance – Oversampled Data

- XGBoost is giving the highest K-Folds CV Recall followed by Random Forest and GBM Classifiers
- All these models are slightly overfit
- Logistic Regression model has the lowest CV score
- The performance of Bagging, Random Forest, GBM and LR models are all consistent without any outliers
- Highest Validation Recall achieved is 91



Model Performance – Down-sampled Data

- XGBoost is giving the highest K-Folds CV Recall followed by GBM and Random Forest Classifiers
- LR, AdaBoost, GBM and XGBoost models are generalizing well on train and validation sets
- Highest Validation Recall achieved is 96
- Overall, all the models are giving really good scores on the validation set



Performance Summary - Hyper-tuned models

Estimators	*Train Accuracy*	*Validation Accuracy*	*Train Recall*	*Validation Recall*	*Train Precision*	*Validation Precision*	*Train F1*	*Validation F1*
LR on Down sampled data	0.845	0.846	0.859	0.877	0.836	0.513	0.847	0.647
AdaBoost on Down sampled data	0.986	0.936	0.992	0.969	0.980	0.725	0.986	0.829
GBC on Down sampled data	0.983	0.947	0.989	0.966	0.978	0.766	0.983	0.855
XGB on Over sampled data	0.861	0.764	1.000	0.991	0.783	0.404	0.878	0.574

Best Model

- Logistic Regression is giving a generalized performance on the validation set.
- XGBoost gives a near perfect recall on the validation set. However, the precision score is too low.
- AdaBoost and Gradient Boost models are both giving the best scores fulfilling the problem requirement.
- **Hypertuned Gradient Boosting Classifier** has accuracy, precision and F1 scores are a little over the edge in comparison to the hypertuned AdaBoost model, while the recalls are equal

Best Model – Test Performance

Metrics	*Accuracy*	*Recall*	*Precision*	*F1*
Validation set	0.947	0.966	0.766	0.855
Test set	0.940	0.975	0.737	0.840

- The model is giving a generalized performance on validation and test sets while fulfilling our requirements

Business Insights

Potential customers are likely :

- Customers who have been inactive for longer periods
- Ones who do not have any revolving balance left
- Having 1K or less balance amount left on their credit card
- Who do not utilize the credit available on their card
- Who have had high number of contacts with the bank in the last 12 months
- Customers with lower transaction amount / count in the past 12 months
- Customers with lower total transaction amount / count change ratio of 4th and 1st quarters

Recommendations

The bank can :

- Provide deals like cash back or 0% interest over a period of time, so that the customers are likely to become more active
- Promote their products with attractive offers as customers holding more number of products with the bank show lower attrition rate
- Resolve any customer issues so as to reduce the number of contacts with the bank
- Provide less interest rates on cash transfers with credit cards
- Provide rewards for online shopping over a fixed amount
- Provide introductory offers to attract new customers
- Constantly update the marketing strategies based on the customers' interest and competitive bank offers

Thank You