CREDIT CARD USER CHURN ANALYSIS THERA BANK

Project 6 - Model Tuning

Authored by Chad Kamakani Kahunahana

BUSINESS OVERVIEW
PROBLEM AND SOLUTION APPROACH

PROBLEM

Customers' leaving credit cards services leads to revenue loss. Thera Bank wants to identify the customers most likely to churn and determine how to prevent these losses.

The Bank recently saw a steep decline in the number of users of their credit card, credit cards are a good source of income

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

SOLUTION APPROACH

Analyze customer data and build a classification model to identify customers likely to churn and reason that could lead to these losses

Using data provided, we will identify key variables from their customer database to identify which characteristics are best at predicting which customers are more likely to purchase travel packages.

Objectives:

1. To predict which customers are more likely to churn
2. Identify customer characteristics which will help improve customer retention

BASE DATASET OVERVIEW

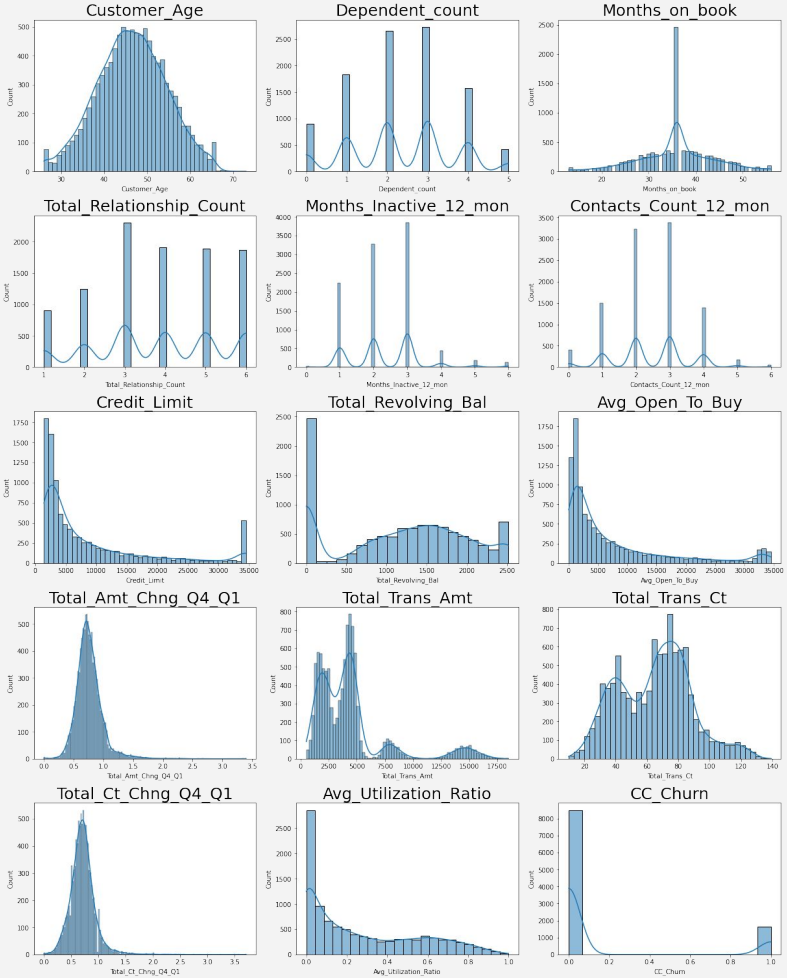
Details	Amount
Observations	10127
Variables	20
Missing Values	Education_Level (1519), Marital_Status (749)

This is a description of the base dataset prior to any pre-processing

Variable	Type	Description
CLIENTNUM	Integer	Unique Customer ID
Attrition_Flag	object	Target variable
Customer_Age	Integer	Age of customer
Gender	object	Gender of customer
Dependent_count	Integer	No. of dependents
Education_Level	object	Educational Qualification of account holder
Marital_Status	object	Marital status
Income_Category	object	Annual Income
Card_Category	object	Type of credit card
Months_on_book	Integer	Period of relationship with the bank
Total_Relationship_Count	Integer	Total no. of products held by the customer
Months_Inactive_12_mon	Integer	No. of months inactive in the last 12 months
Contacts_Count_12_mon	Integer	No. of Contacts with the customer in the last 12 months
Credit_Limit	Float	Credit Limit on the Credit Card
Total_Revolving_Bal	Integer	The balance that carries over from one month to the next
Avg_Open_To_Buy	Float	Amount left on the credit card to use (Average of last 12 months)
Total_Trans_Amt	Float	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Integer	Total Transaction Count (Last 12 months)
Total_Ct_Chng_Q4_Q1	Integer	Ratio of total transaction count 4Q to 1Q
Total_Amt_Chng_Q4_Q1	Float	Ratio of total transaction amount 4Q to 1Q
Avg_Utilization_Ratio	Float	Represents how much of the available credit the customer spent

01

EDA &
CUSTOMER PROFILE



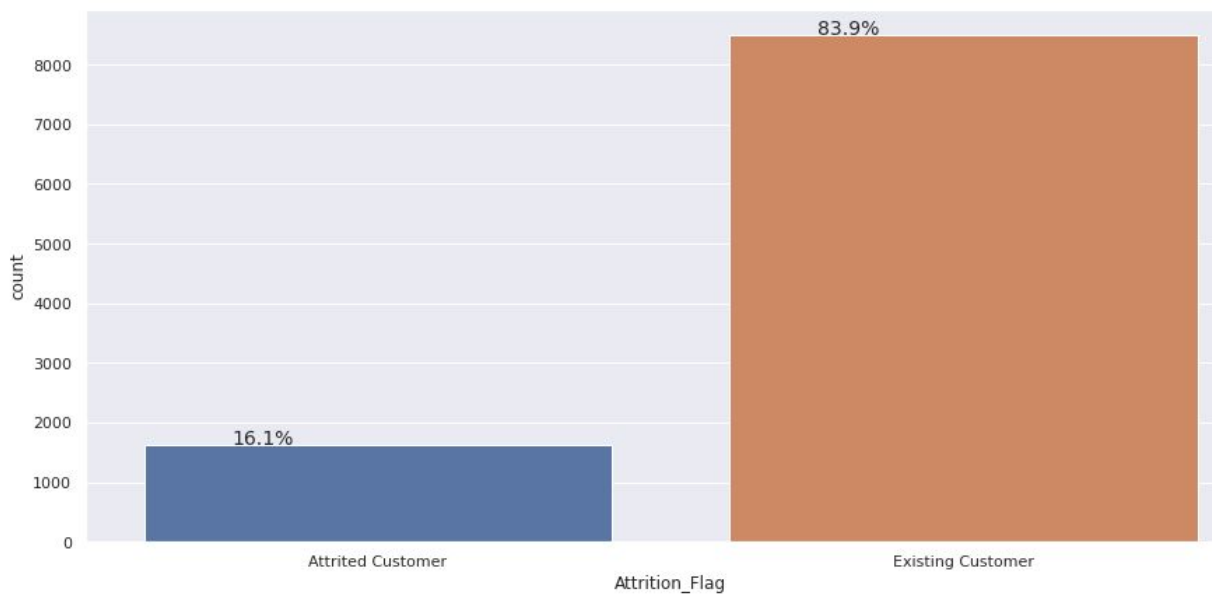
DISTRIBUTION ANALYSIS

- Customer_Age, Months_on_Book, Credit_Limit, Total_Revolving_Balance, Avg_Open_To_Buy, Total_Amt_Chng_Q4_Q1, Total_Trans_Amt, Total_Trans_Ct, Total_Ct_Chng_Q4_Q1, Avg-Utilizatin_Ratio are continuous variables
- Attrition_Flag a categorical variable has been removed and replaced in the final dataset with CC_Churn a numerical feature for better modeling
- There are heavy skewness in the dataset. We tried multiple transformations but they did not affect model results. Tree based boosting algorithms are well equipped to deal with skewness.

  Skewed values were sanity checked for abnormality and found to be within market trend and therefore were not clipped
- We will look at these variables in closer detail on the following pages

CUSTOMER ANALYSIS

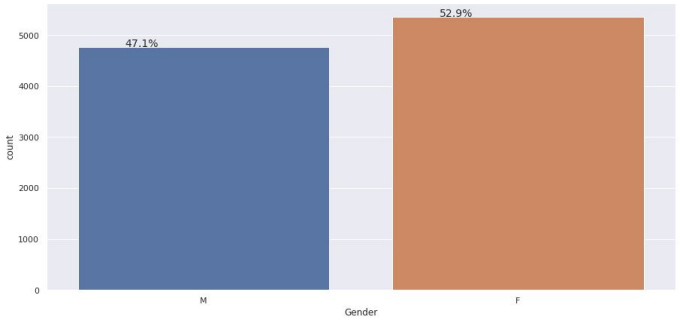
ATTRITION FLAG (CREDIT CARD CHURN)



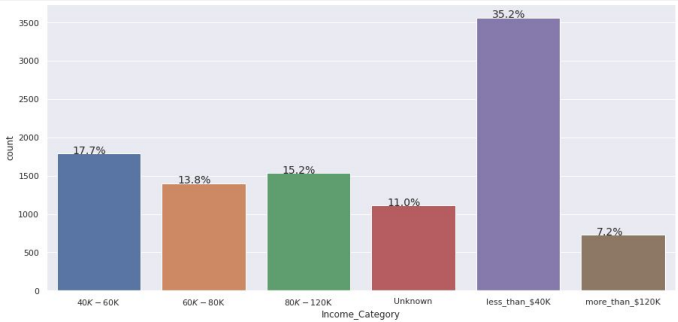
- 16.1% of customers left the bank

CUSTOMER ANALYSIS

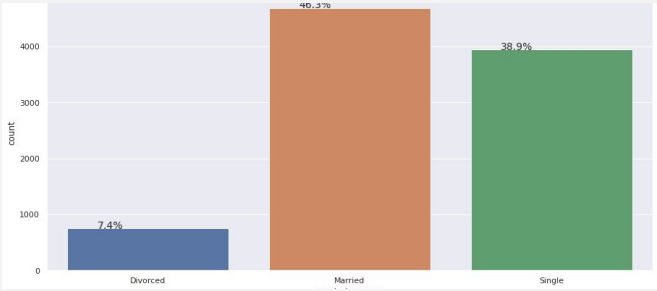
GENDER



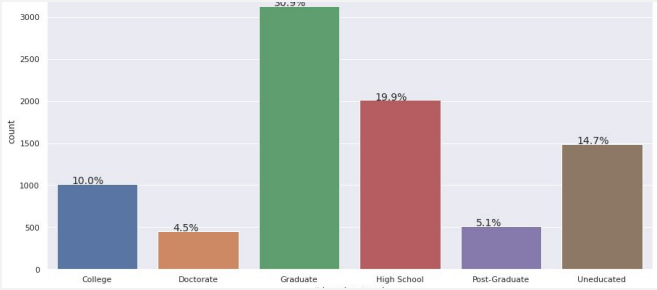
INCOME



MARITAL STATUS



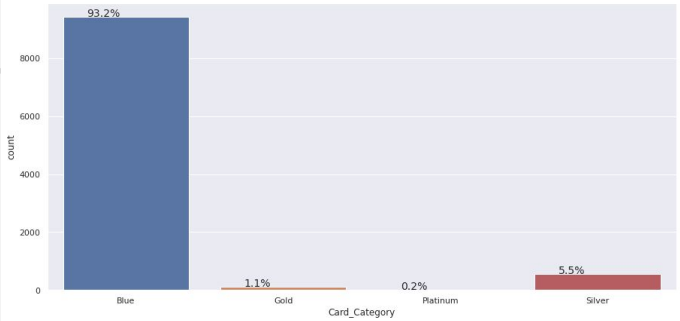
EDUCATION



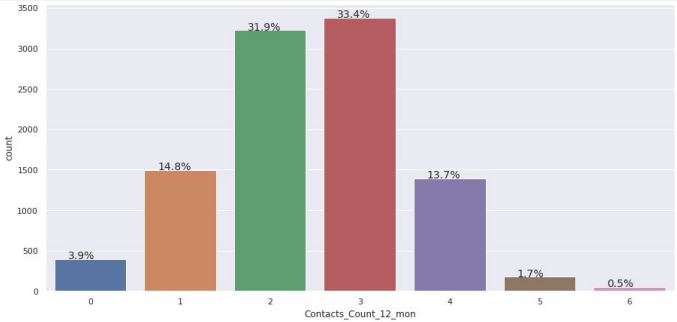
- There are slightly more females (52.9%) then males (47.1%)
- 61.1% of customers are married or divorced and just 38.9% are single
- 35.2% of customer make < $40K, 66.7% make less than $80K, only 22.4% make more than $80K
- Customers tend to be educated with just 14.7% having no education and 65.4% having at least some college education or more

CUSTOMER ANALYSIS

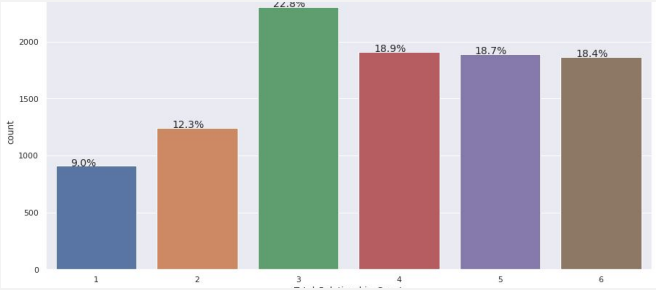
CREDIT CARD TYPE



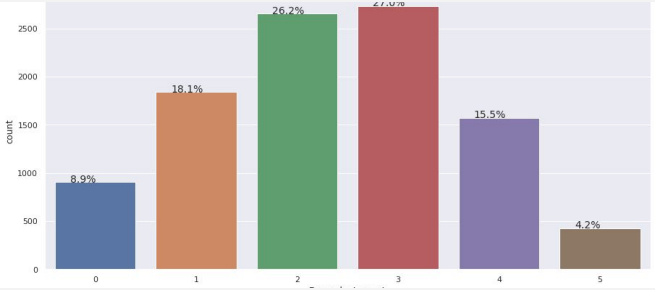
CONTACT COUNT



TOTAL RELATIONSHIP COUNT



NO. OF DEPENDENTS



- 93.2% of customers hold the Blue card, while only 0.2% hold platinum
- Bank customers tend to hold 3 or more bank products
- On average, customers tend to contact the bank 2 or 3 times per year
- Customer report having on average 2 to 3 dependents

EXPLORATORY DATA ANALYSIS
CUSTOMER PROFILE BY CARD & CHURN

BLUE

- Count 9436
- Median age 46
- Gender Female
- 2.3 dependents
- Graduate level education
- Married
- Earn less than $40K salary
- Median credit limit of $4105
- Balance of $1267
- Has 4 products
- Contact the bank 2.5x per year
- Transaction count ratio Q4 to Q1 is 0.713 on average

SILVER

- Count 555
- Median age 45
- Gender Male
- 2.4 dependents
- Graduate level education
- Single
- Earn less than $40K salary
- Median credit limit of $29808
- Balance of $1378
- Has 3 bank products
- Contact the bank 2.5x per year
- Transaction count ratio Q4 to Q1 is 0.709

GOLD

- Count 116
- Median age 46
- Gender Male
- 2.7 dependents
- Graduate level education
- Single
- $60K-$80K salary
- Median credit limit of $34516
- Balance of $1459
- Has 3 bank products
- Contact the bank 2.4x per year
- Transaction count ratio Q4 to Q1 is 0.706

PLATINUM

- Count 20
- Median age 48
- Gender Male
- 2.5 dependents
- Graduate level education
- Single
- Income Unknown
- Median credit limit of $34516
- Balance of $1512
- Has 2 bank products
- Contact the bank 2.5x per year
- Transaction count ratio Q4 to Q1 is 0.666

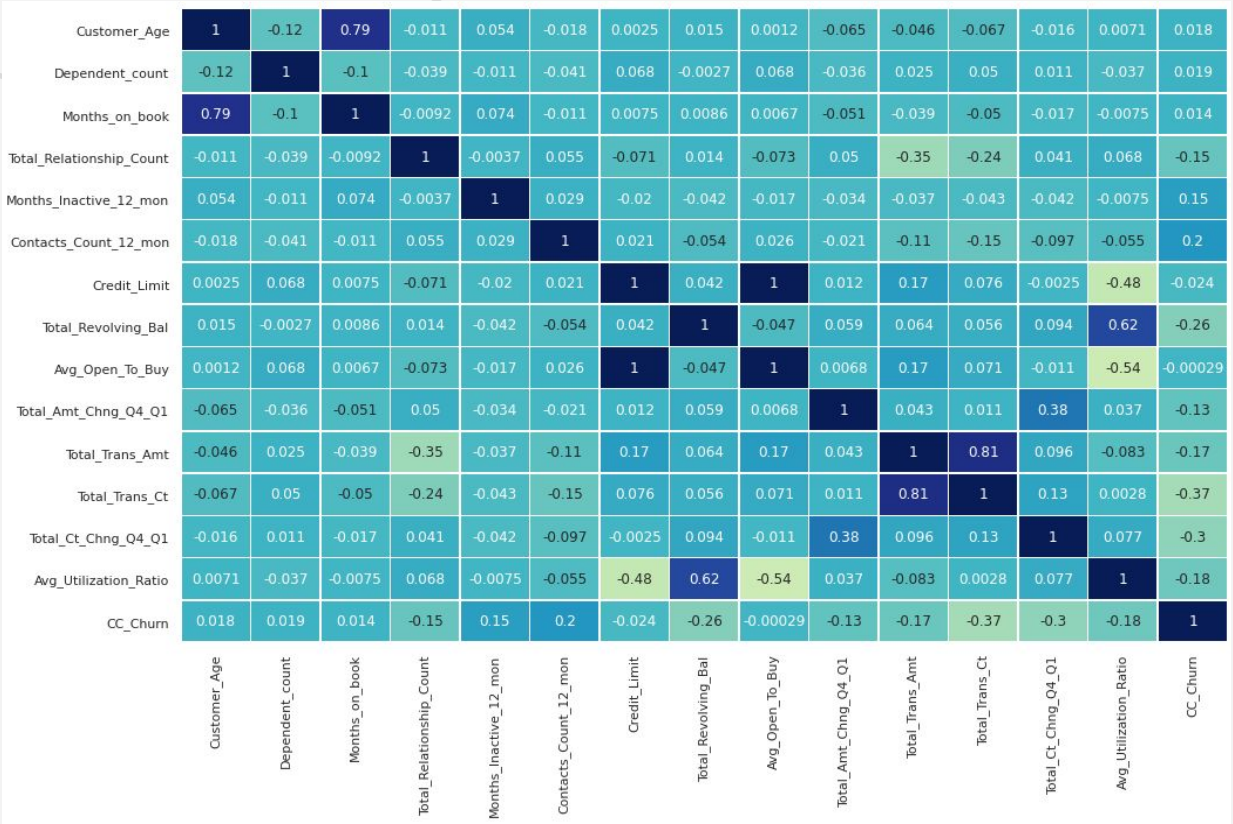
CHURNED CUSTOMER

- Count 1627
- Median age 47
- Gender Male
- 2.4 dependents
- Graduate level education
- Married
- Earn less than $40K salary
- Median credit limit of $4178
- Balance of $0.00
- Has 3 bank products
- Contact the bank 3.0x per year
- Transaction count ratio Q4 to Q1 is 0.554

02

BIVARIATE & MULTIVARIATE ANALYSIS

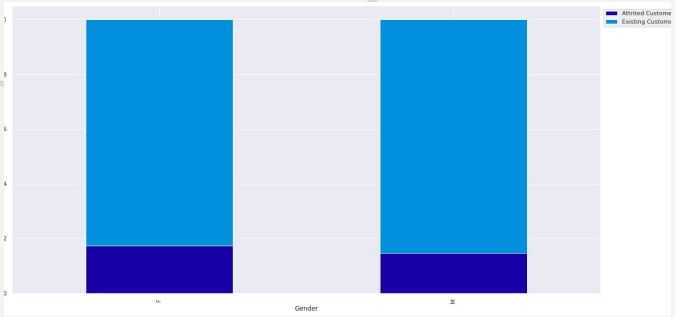
VARIABLE ANALYSIS
CORRELATION MATRIX



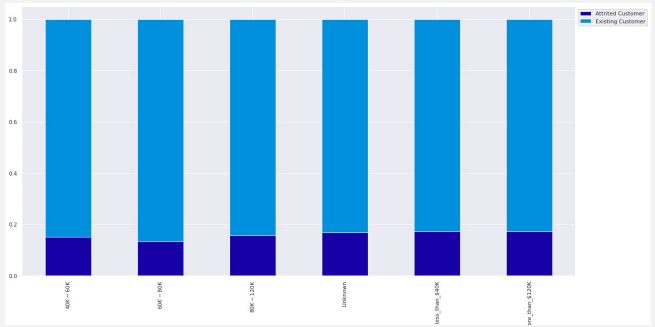
- Months_on_Book and Age are correlated which is not surprising (Months on book was dropped)
- Credit_Limit and Avg_Open_to_Buy has perfect correlation which makes sense (Avg_Open_to_Buy was dropped)
- Total_Revolving_Balance and Avg_Utilization Rate are related which also seems normal (both are kept)
- Total_Trans_Amt and Total_Trans_Ct are also highly correlated (Total_Trans_Ct was dropped)

CHURN ANALYSIS

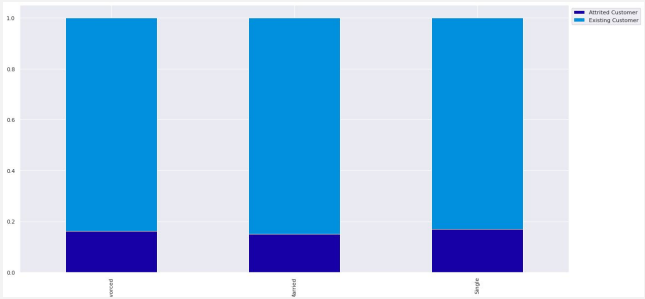
GENDER VS. CHURN



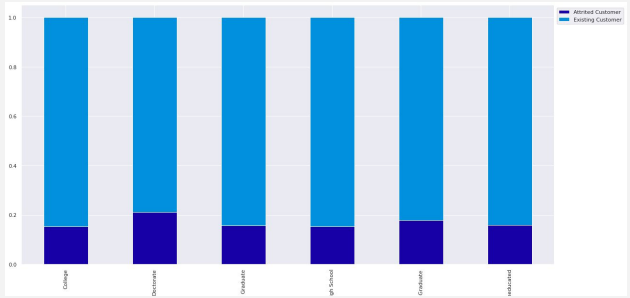
INCOME VS. CHURN



MARITAL STATUS VS. CHURN



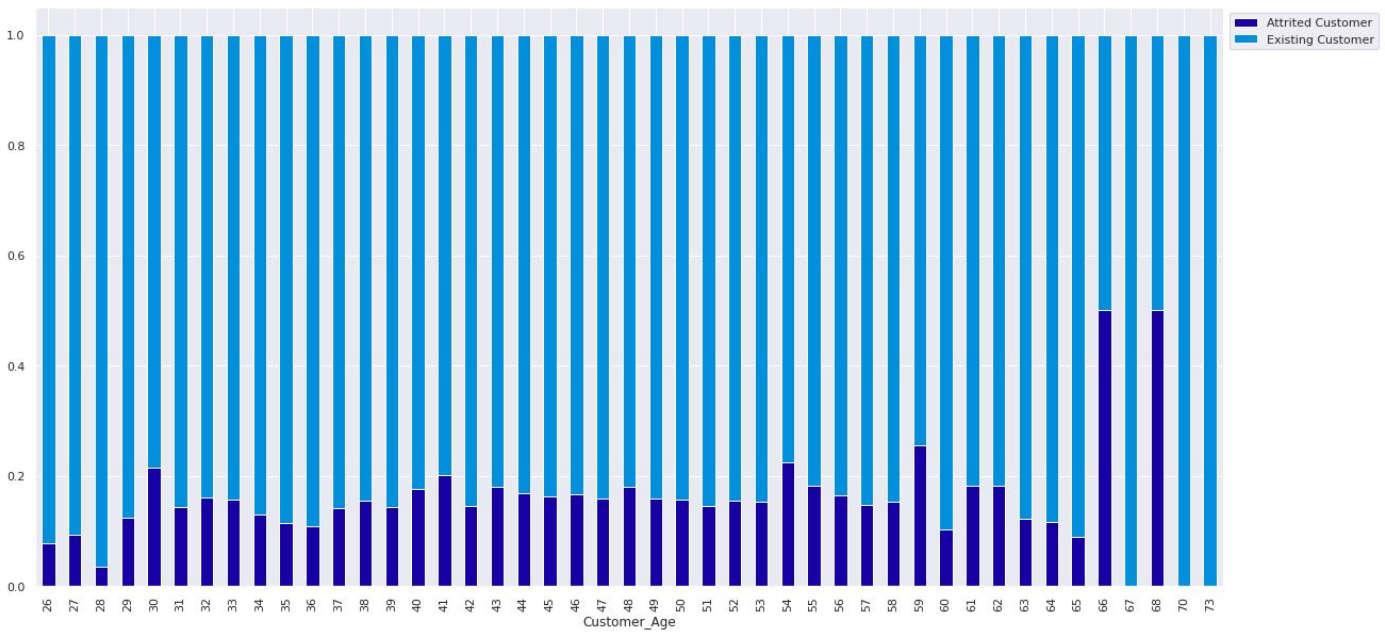
EDUCATION VS. CHURN



- Gender is not a major factor in churn
- Married customers are slightly less likely to churn
- Low income customers and highest income customers are slightly more likely to churn. This could be for different reasons 1. Financial distress (lower income) or possibly 2. More options to go to another bank (upper income)
- Those with Doctorate and Post-Graduate education are relatively more likely to churn

CHURN ANALYSIS
AGE

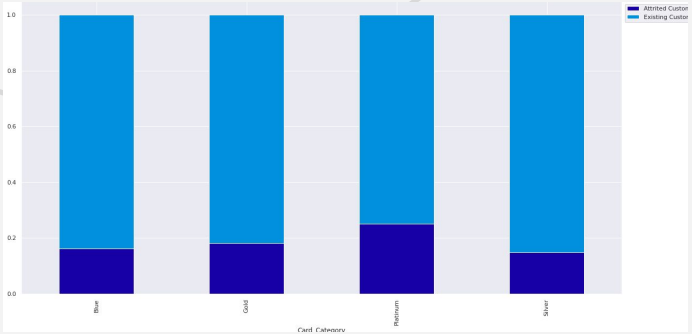
AGE VS. CHURN



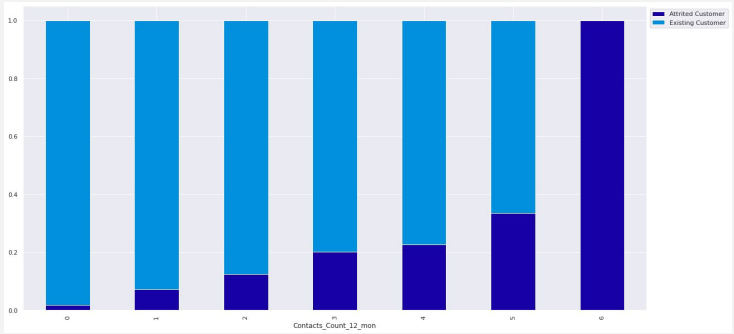
- Age does not seem to be a significant factor in churn except for those customers who are less than 30
- Spikes in churn for customers aged 66 and 68 are not reliable due to very small amount of observations

CHURN ANALYSIS

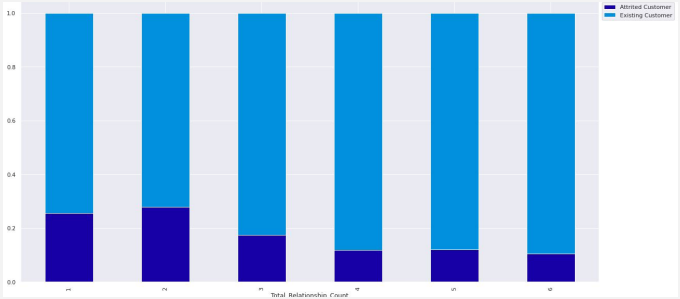
CREDIT CARD TYPE VS. CHURN



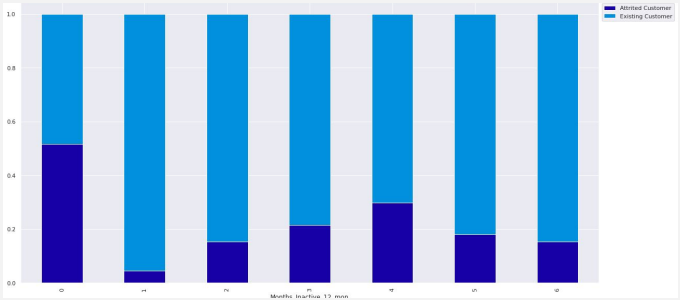
CONTACT COUNT. VS CHURN



TOTAL RELATIONSHIP COUNT VS. CHURN

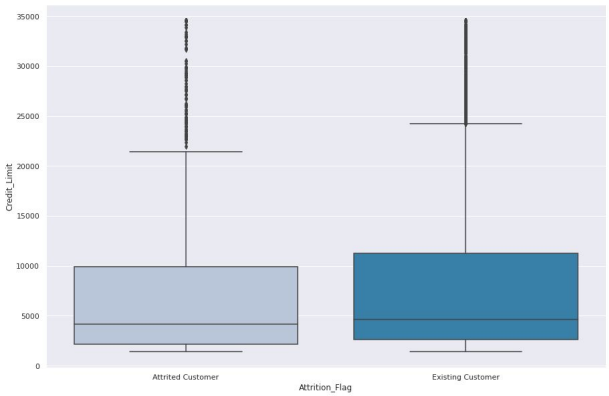


MONTHS INACTIVE VS. CHURN

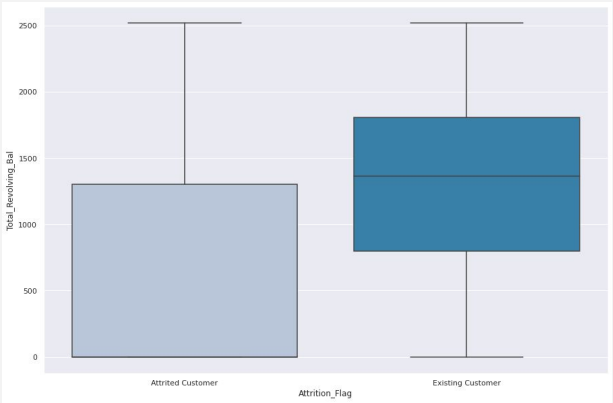


- Platinum card holders tend churn the most relatively speaking although they account for only 0.2% of customers
- Customers with more bank products tend to churn less which makes sense
- Customers who contact the bank the most are more likely to churn which is shown by month Contact Count and Month Inactive

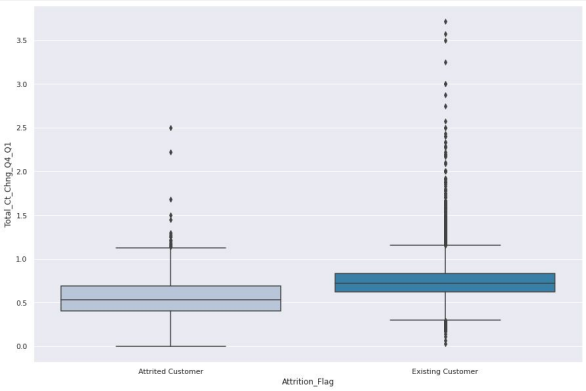
CREDIT LIMIT VS. CHURN



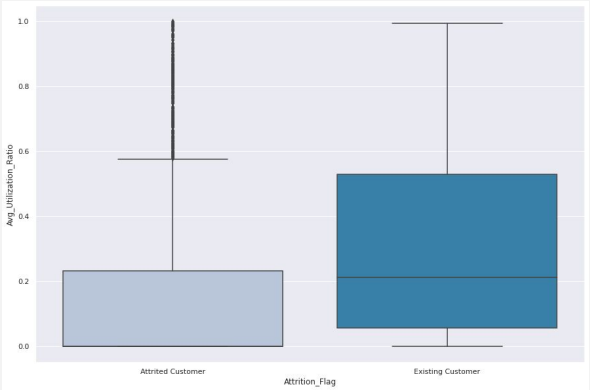
TOTAL REVOLVING BALANCE VS. CHURN



TOTAL TRANS COUNT CHANGE Q4/Q1 VS. CHURN



AVERAGE UTILIZATION VS. CHURN



CHURN ANALYSIS

- Customers with higher credit limits tend to churn slightly less
- Customers with more transaction in the Q1 vs Q4 tend to trend less
- Customers with higher revolving balances and higher average utilization tend to churn less

04

MODEL BUILDING & BUSINESS RECOMMENDATIONS

MODEL SUMMARY

Objectives

- To predict which customers are more likely to churn
- Identify customer characteristics which will help improve customer retention

Model can make wrong predictions as:

- Predicting a customer will churn and the customer does not churn (increase costs; False Positive)
- Predicting a customer will not churn and the customer does churn (lose customers, loss opportunity and revenue; False Negative)

Which case is more important?

Predicting that customer will not churn but the customer does churn. In this case, we will have a business loss and have missed an opportunity to prevent this loss. We want to reduce False Negatives and increase True Positives.

How to reduce this loss i.e need to reduce False Negatives?

Company wants Recall to be maximized, greater the Recall higher the chances of minimizing false negatives. Hence, the focus should be on increasing Recall or minimizing the false negatives or in other words identifying the true positives(i.e. Class 1) so that the company can expand its customer base and increase profits.

Accuracy is also a good score as it can help the company better refine its marketing to be more capital efficient. However, Recall is our primary metric.

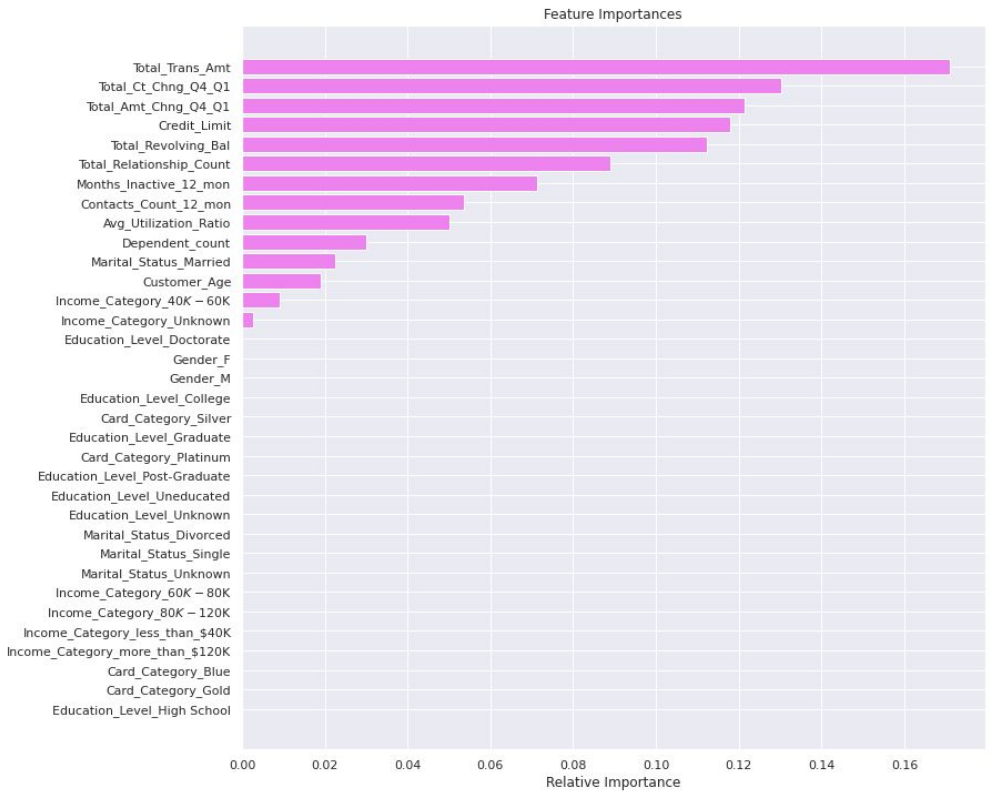
MODEL PERFORMANCE
SUMMARY

COMPARING MODEL PERFORMANCE

Model	Train Accuracy	Validate Accuracy	Train Recall	Validate Recall	Train Precision	Validate Precision
ADABoost Tuned	0.947	0.943	0.732	0.721	0.922	0.907
ADABoost Oversampling	0.969	0.945	0.956	0.785	0.981	0.859
ADABoost Undersampling	0.945	0.919	0.942	0.939	0.947	0.680
GBM Tuned	0.993	0.963	0.972	0.859	0.985	0.906
GBM Oversampling	0.993	0.957	0.992	0.862	0.994	0.870
GBM Undersampling	0.999	0.927	1.000	0.954	0.998	0.702
XGBoost Tuned	0.874	0.870	0.981	0.954	0.561	0.556
XGBoost Oversampling	0.857	0.748	0.997	0.982	0.779	0.388
XGBoost Undersampling	0.808	0.658	0.999	0.994	0.723	0.319

- ADABoost Tuned has the highest Precision but Recall below our >0.95 target
- GBM Oversampling has the highest Accuracy but Recall below our >0.95 target
- XGBoost has the highest Recall at 0.982 and seems to be a good model
- XGBoost Tuned has a Recall of 0.954 which exceeds our 0.95 target, it also has good Accuracy of 0.87
- **XGBoost Tuned is our best over model**

FEATURE IMPORTANCE
MODEL PERFORMANCE SUMMARY



- High total transaction amounts and changes in transaction counts and amounts from Q4 to Q1 are strong indicators for churn
- Credit Limit and Total Revolving Balance are also features which contribute to churn

BUSINESS RECOMMENDATIONS

- Analysis shows customers with a credit card balance of $0.00 are more likely to churn.
  - Bank should proactively reach out to these customers and inquire about service
  - Bank should offer these customers incentives such as lower interest rates or card upgrades to encourage spending

- Analysis shows customers with incomes <$40K and higher incomes >$120K are more likely to churn. Bank can create two different programs to address both groups. It also shows that customers who have more bank products are less likely to churn.
  - FIrst if should train agents to contact lower income customers and offer them other products such as savings accounts
  - Second, it should contact higher income customers and offer them perks such as rewards cards, investment accounts, and loans

- Customers that contact the bank often are more likely to churn. This could be an indication of poor customer service standards.
  - The bank should take contact from existing customers seriously and train agents to investigate the source of this contact and proactively solve customer complaints. The bank should track customer inquiries in a database for further analysis.

- Customer with lower credit limits, lower balances, and lower transaction volume are also more likely to churn. It is also true that “Blue” credit cards are by far the most common.
  - The bank should analyse customer traits such as education and income and offer good customers Silver and Gold cards which have increased credit limits, lower interest rates and more perks.
  - The bank should also partner with retailers and travel companies to offer discounts and other perks to increase transaction volume