Credit Card Customer Attrition

Business Analytics with R

- Prof. Zhe Zhang

Group – 5 Members

- *Suma Harshini Samanthula*
- *Catherine Christina Ganduri*
- *Srija Mannam*
- *Ramakrishna Chepuri*
- *Dharmik Mahendra Bhanushali*

Business Problem:

A manager at the bank is disturbed by an alarming number of customers leaving their credit card services. There will be numerous reasons which are not consistent with other customers who have left the bank but what if there is a pattern to the reasons customers decided to leave the bank?

The manager would appreciate it if someone could predict who is going to leave their company so that they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction. Furthermore, it will be helpful for the company to prevent further customer attritions in the future.

Objective:

The main goal of this case is to understand what attributes are making clients leave the bank, which increases credit card attrition. We have a few tasks to perform to come up with a solution. We first need to explore our data which will help us draw insights to have a general idea of what could be the main reasons which are causing credit card customers to leave the bank. We will be able to understand what are the top attributes which affect attrition in our model. We then perform classification models for predictive analysis.

DATA SOURCE

https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers

DATA DESCRIPTION

The study uses a data set that consists of 10127 observations and 21 columns with generic credit card details, one for each customer in the bank. In this data set, there is detailed information that the bank obtains from different sources to analyze the use of each customer's credit card.

This dataset comprises a variety of elements and their values, which we can use to create a model and make a strategic decision that will benefit the bank. The data model demonstrates how we may use R to solve a variety of business challenges by applying classification models.

- Attrition_Flag: This is our target attribute, which tells us whether the customer is an existing customer or an attired customer.
- Customer_Age: Customer's age in years
- Gender: Customer's gender (male or female)
- Dependent_count: Number of people who depend upon the customer for their support and welfare.
- Education_Level: Customer's educational qualification (high school, graduate, etc.)
- Marital_Status: Customer's marital status (married, single, etc.)
- Income_Category: Customer's income bracket in dollars (less than 40K, 40K-60K, etc.)
- Card_Category: Credit card category (Blue, Silver, etc.)

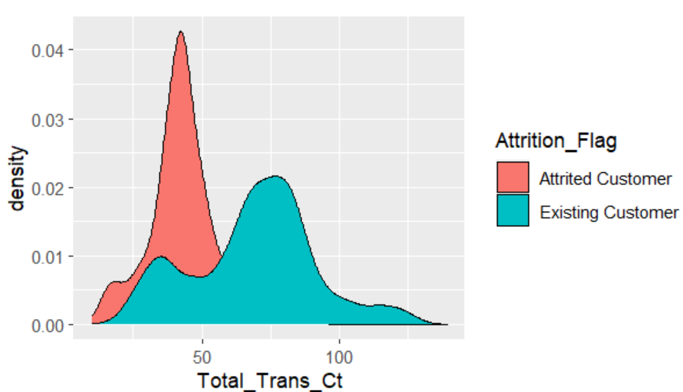
- ➢ Months_on_book: Period of relationship with the bank in months
- ➢ Total_Relationship_Count: Total number of products held by the customer
- ➢ Months_Inactive_12_mon: Number of months inactive in the last 12 months
- ➢ Contacts_Count_12_mon: Number of contacts (phone calls) in the last 12 months
- ➢ Credit_Limit: Credit limit on the credit card
- ➢ Total_Revolving_Bal: Total revolving balance on the credit card
- ➢ Avg_Open_To_Buy: Open to buy credit line (average of last 12 months). This also turns out to be the difference between the credit limit (Credit_Limit) assigned to a cardholder account and the present balance on the account (Total_Revolving_Bal).
- ➢ Total_Amt_Chng_Q4_Q1: Change in total transactions amount (Q4 over Q1)
- ➢ Total_Trans_Amt: Total transactions amount (last 12 months)
- ➢ Total_Trans_Ct: Total number of transactions (last 12 months)
- ➢ Total_Ct_Chng_Q4_Q1: Change in the total number of transactions (Q4 over Q1)
- ➢ Avg_Utilization_Ratio: Average card utilization ratio

CHALLENGES ENCOUNTERED AND APPROACHES TAKEN

- The main challenge with this data set is the unbalanced target variable, that is, Attrition_Flag only counts for about 16.07% of the desired output.
- Columns like Clientnum are of no use and naïve Bayes columns were removed.
- Outliers in various variables, such as Credit limit and Customer Age, were addressed by imputing values below the IQR to 5% of the data distribution and values above the IQR to 95%.
- There were multiple null values in the dataset so we Omitted null values.

ANALYSIS AND DISCUSSION

Total_Trans_Ct



From the density graph below, it can be interpreted that the customers whose total transaction count is customers less than 75 has highly attrited the credit card and the customers with total transaction count greater than 75 are existing customers.

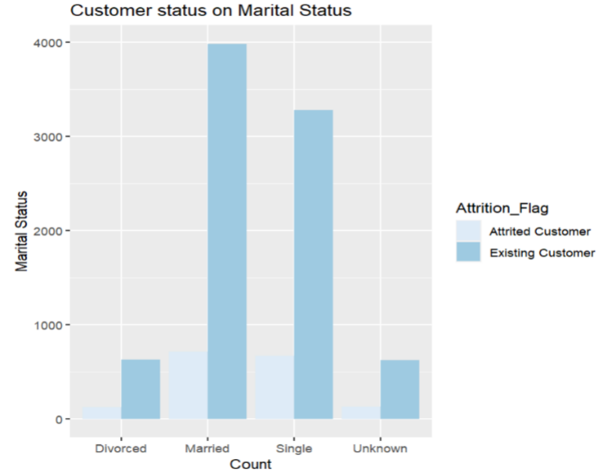
Customer_Age

Attritioned customers have higher median age which shows that it is important to focus on these to reduce attrition.

Customer Status on Gender

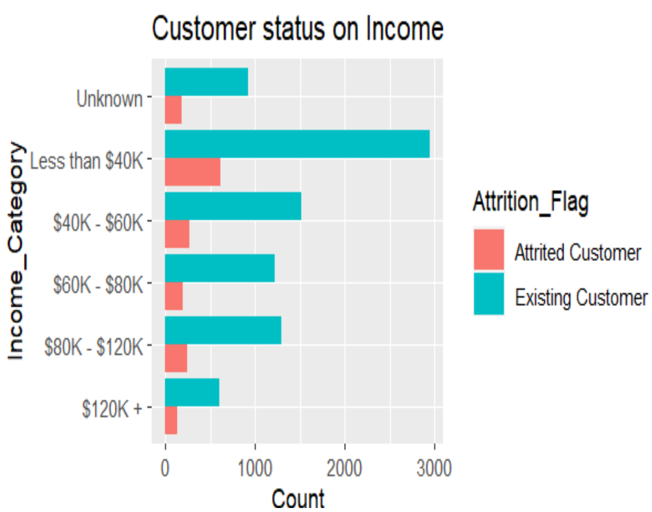
For more precise information, the below histogram graph shows that 9.2% customers who are attrited were female and 6.9% of attrited customers are male. And 43.7% of existing customers are female and 40.2% of existing customers are male.

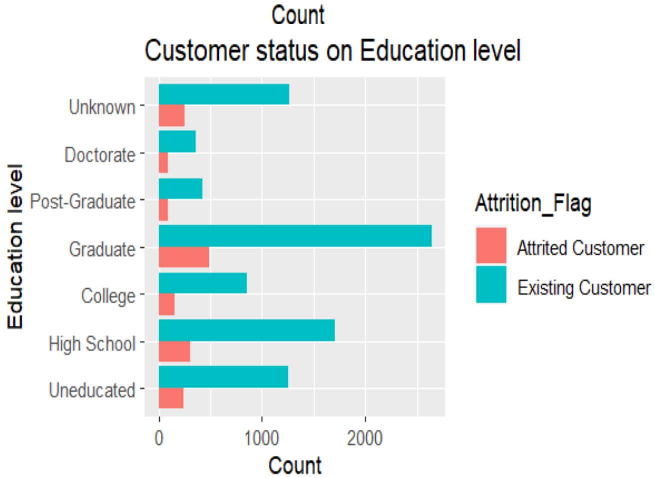
Attrition analysis based on Marital Status



From the above graph the attrition is comparatively less for the customers who got divorced and unknown and the attrition is higher for customers who are married and single. Moreover the existing customers are mostly married and single than the divorced and unknown.

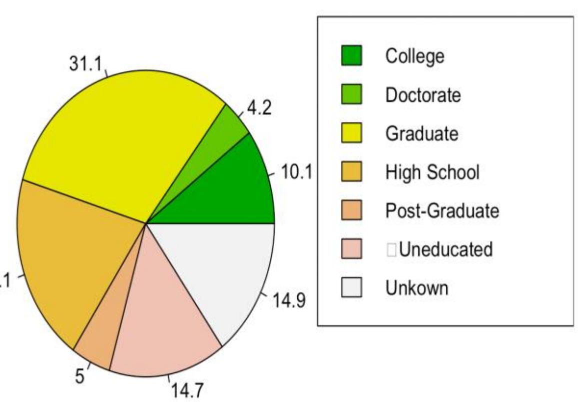
Customer status on Income and Education Level





From the below graph, the Graduate customers contribute more to the existing customers than the other categories and the customers with Doctorate and PG contribute less to the existing customers.
The attrition rate is also higher for customers with a graduate degree compared to others.

Existing People Education Level



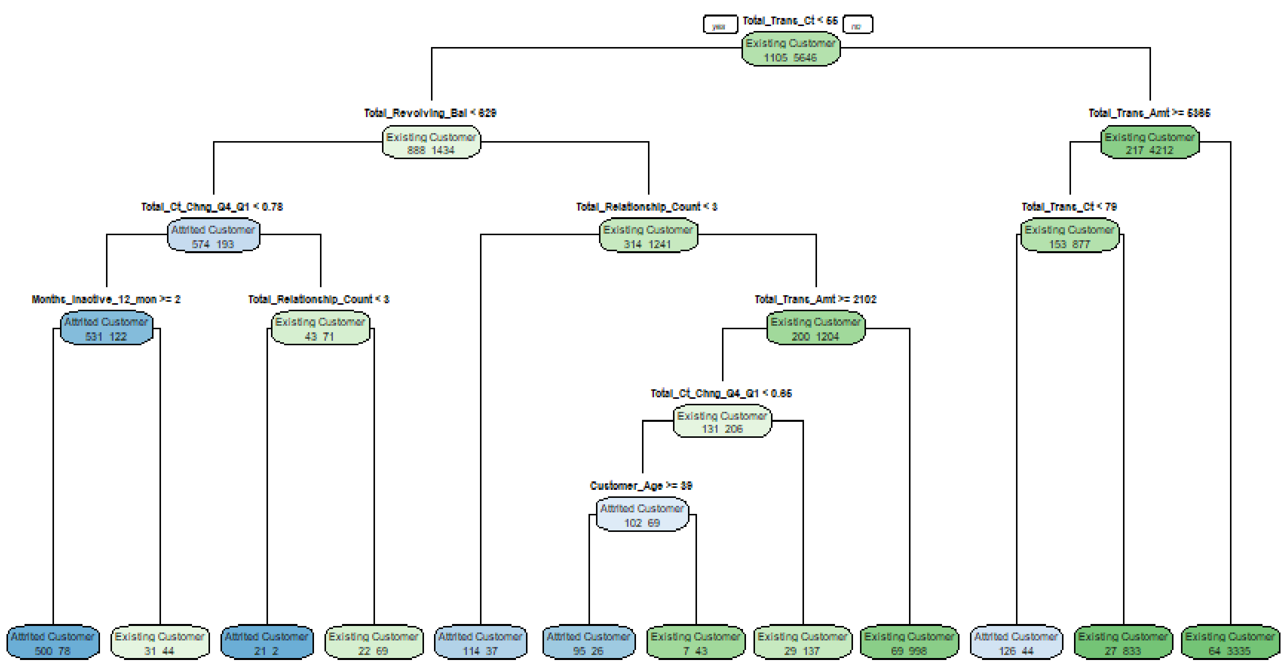
Main education level of the customers is graduated people, followed by high school. We can also notice here that there is no great difference between the ratios of the education level between the existing and attired customers.

Modelling and Performance Metrics:

Metrics	Decision Tree	Logistic Regression	Naïve Bayes method
True positive	2755	2746	2670
True Negative	394	321	332
False Positive	128	201	190
False Negative	99	108	184
Accuracy	0.9327607	0.9084716	0.889218
Specificity	0.7547893	0.6149425	0.6360153
Sensitivity (Recall)	0.9653118	0.9621584	0.9355291
FPR (alpha)	0.2452107	0.3850575	0.3639847
FNR (beta)	0.03468816	0.03784163	0.06447092
Precision	0.9556018	0.931795	0.9335664
Prevalence	0.8453791	0.8453791	0.8453791

Decision Tree:
On the test data, the model predicted the attrited customers with an accuracy of 93.3% and sensitivity of 96.5% due to which we have selected, Decision tree as the best model for our prediction.

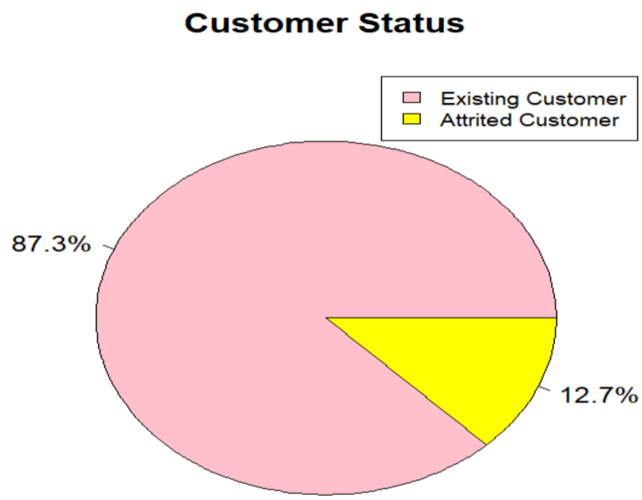
Classification Tree for Attrition Prediction



To predict attrition, we increased mini split to 500 and observed the following:

- When total transactions are less than 55 the and revolving balance is less than 629 customers are more attrited.
- When total transactions are greater than 55 the and total transaction amount is less than or equal to 5365, customers are less attrited which is a true positive.
- Similar way by changing our mini split, we can interpret churn ratios.

Logistic Regression:
This model predicted the attrited customers with an accuracy of 90.8% and sensitivity of 96.2%.



Our prediction through test data from logistics regression model tells us that 12.7% are predicted to be attrited.

Conclusion

Three prediction models have been built to predict whether a customer is attrited or not. Those models were Decision Tree, logistic regression model and Naïve Bayes method model. To guarantee a high accuracy Decision Tree model yielded the best results.

Presentation link:

https://cometmail-my.sharepoint.com/:v:/g/personal/sxm210319_utdallas_edu/EQUHLFhKYMVMhI3Fc9jlq_EBYbWssa_w9tx5IDmuFmNXqA

CSV file: https://cometmail-my.sharepoint.com/:x:/g/personal/sxm210319_utdallas_edu/Ec44iTI9N3BJufXkxYevW-UBq82t4RTuswqbxlfKqX1_Cg