BIA Final Report: Credit Card Customer Analysis

Group3

Alexa Mavica, Saksham Bansal, Neel Patel, Gopika Sushrut Chaudhary, Sahil Yadav

BIA 500

**Introduction:**

Banks are spending a considerable amount of money trying to offer digital services and other amenities to new and existing customers; however, customer churning rates have continued to rise. With increased churning rates impacting revenue and profitability of banks, there have been major concerns about how to determine which customers will churn as well as how to lower the rate of churning. By evaluating if there is a relationship between gender and churning rates and determining which variables are most significant predictors of churning, bank managers are confident that customer churning rates will decrease, and overall profitability will increase.

**Dataset Explanation:**

The dataset used in this analysis was collected in 2021and was sourced from Kaggle. The dataset is a representation of the customers who use credit card services throughout the world and consists of 10,000 customers mentioning factors such as their age, salary, marital status, education level, gender, dependent amount, credit card limit, and credit card category. This dataset shows 16.07% of customers have churned.

Our dataset had a total of 19 variables which are as follows:

Customer age represents the age of the customer, which is ratio data.

Gender is a categorical variable with only 2 values, male and female.

Dependent count is a type of ratio data which has values from 0 to 5.

Education level represents the educational qualification of the account holder like graduate, high school, or other. It is a categorical variable.

Marital status is a categorical variable four categories- divorced, married, single, unknown.

Income category is also a categorical variable with values, less than $40k, $40k - $60k and others.

Card category has categories blue, gold, platinum, and silver.

Then there is months on book which represents period of relationship with bank.

Total number of products held by the customer, number of months inactive in the last 12 months, number of contacts in the last 12 months, credit limit on the credit card, total revolving balance on the credit card, open to buy credit line (average of last 12 months), change in transaction amount (Q4 over Q1), total transaction amount in the last 12 months, total transaction count in the last 12 months, change in transaction count (Q4 over Q1), and average card utilization ratio are ratio data.

Our aim was to find out whether gender plays an important role in the prediction of churning of a customer and find out the top 9 variables with the most significance and how they affect churning. To do this, we first used RapidMiner to find variables with the highest correlation with attrition and we found that total transaction count in the last 12 months, change in transaction count (Q4 over Q1), total revolving credit card balance, number of contacts (within the last 12 months), average card utilization ratio, total transaction amount in the last 12 months, number of months inactive (within past 12 months), total number of products held by a customer, and change in transaction amount (Q4 over Q1) are the top 9 variables. The results from RapidMiner can be seen in the graph below.

Insights on the key variables can be found in the table below. All the variables are of the type ratio data and most of them are desired to have a high value with the exception of number of contacts within the last 12 months and number of months inactive within past 12 months, which are desired to be low. As it can be clearly seen, mean of total transaction amount in the last 12 months is the highest, but it also has the highest standard deviation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Type of Data** | **Desired Value** | **Mean** | **Standard Deviation** |
| **Total Transaction Count in the last 12 months** | Ratio | High | 64.86 | 23.47 |
| **Change in Transaction Count (Q4 over Q1)** | Ratio | High | 0.712 | 0.238 |
| **Total Revolving Credit Card Balance** | Ratio | High | 1162.8 | 814.987 |
| **No. of contacts within the last 12 months** | Ratio | Low | 2.455 | 1.106 |
| **Average Card Utilization Ratio** | Ratio | High | 0.275 | 0.276 |
| **Total Transaction Amount in the last 12 months** | Ratio | High | 4404.08 | 3397.13 |
| **No. of months inactive (Within past last 12 months)** | Ratio | Low | 2.4 | 1.01 |
| **Total number of products held by the customer** | Ratio | High | 3.81 | 1.55 |
| **Change in Transaction Amount (Q4 over Q1)** | Ratio | Low | 0.76 | 0.22 |

**Analysis:**

To begin with, we needed to find out the role of gender in churning of a customer. So, we conducted Ch-square test tot find out whether gender and attrition are dependent on each other. Comparing all the values, we found that 930 customers who churned were females, while only 697 customers who churned were males, which can clearly be seen in the chart below, where 0 represents churned customer and 1 represents existing customers.

To conduct Chi-square test, our null hypothesis was gender and churning are independent of each other and alternative hypothesis was gender and churning are dependent on each other.

As we can see below, Chi-square test gave a value of 0.000176298, which is significantly lower than the value of alpha which is 0.05. We reject the null hypothesis and accept the alternative hypothesis. Hence, we conclude that gender plays a significant role in predicting whether a customer would churn or not.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Attrition Flag: Existing: 1, Churned: 0** | **Column Labels** |  |  |
| **Row Labels** | **0** | **1** | **Grand Total** |
| F | 930 | 4428 | 5358 |
| M | 697 | 4072 | 4769 |
| **Grand Total** | **1627** | **8500** | **10127** |
|  |  |  |  |
| Gender | 0 | 1 | Grand Total |
| Female | 860.8142589 | 4497 | 5358 |
| Male | 766.1857411 | 4003 | 4769 |
| Grand Total | 1627 | 8500 | 10127 |
|  |  |  |  |
| Chi Square Test | **0.000176298** | <.05 |  |

Next, we wanted to know the affect of the highest correlated variables on churning. So, we performed regression analysis, in which our null hypothesis was that total transaction count in the last 12 months, change in transaction count (Q4 over Q1), total revolving credit card balance, number of contacts (within the last 12 months), average card utilization ratio, total transaction amount in the last 12 months, number of months inactive (within past 12 months), total number of products held by a customer, and change in transaction amount (Q4 over Q1) are not significant predictors of whether a customer would churn or not and alternate hypothesis was total transaction count in the last 12 months, change in transaction count (Q4 over Q1), total revolving credit card balance, number of contacts (within the last 12 months), average card utilization ratio, total transaction amount in the last 12 months, number of months inactive (within past 12 months), total number of products held by a customer, and change in transaction amount (Q4 over Q1) are all significant predictors of whether a customer will churn or not.

We compared various models in regression to find that all the above-mentioned variables are significant predictors. Although our model had R Square and Adjusted R Square value of only 0.36, it was the best model for the analysis.

Total transaction count in the last 12 months had a p-value of 0.

Change in transaction count (Q4 over Q1) had a p-value of 9.08e-104

Total revolving credit card balance had a p-value of 4.67e-112

Number of contacts (within the last 12 months) had a p-value of 2.8e-52

Average card utilization ratio had a p-value of .0019

Total transaction amount in the last 12 months had a p-value of 3.51e-105

Number of months inactive (within past 12 months) had a p-value of 6.6e-47

Total number of products held by a customer had a p-value of 5.12e-101

Change in transaction amount (Q4 over Q1) had a p-value of 9.6e-06

All this data can be clearly seen in the regression model below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Regression Statistics* | |  |  |  |
| Multiple R | 0.601063 |  |  |  |
| R Square | 0.361277 |  |  |  |
| Adjusted R Square | 0.360708 |  |  |  |
| Standard Error | 0.293625 |  |  |  |
| Observations | 10127 |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 0.031123 | 0.020294936 | 1.533514034 | 0.12518052 |
| Total transaction count in the last 12 months | 0.009574 | 0.000215579 | 44.41205086 | 0 |
| Change in transaction count (Q4 over Q1) | 0.294679 | 0.01346513 | 21.88457114 | 9.0834E-104 |
| Total revolving credit card balance | 0.000106 | 4.66384E-06 | 22.77952599 | 4.6748E-112 |
| Number of contacts (within the last 12 months) | -0.04106 | 0.002682948 | -15.30261756 | 2.8214E-52 |
| Average card utilization ratio | -0.04308 | 0.013884848 | -3.102567102 | 0.001923785 |
| Total transaction amount in the last 12 months | -3.4E-05 | 1.54683E-06 | -22.03936191 | 3.5167E-105 |
| Number of months inactive (within past 12 months) | -0.04186 | 0.002895759 | -14.45723406 | 6.61309E-47 |
| Total number of products held by a customer | 0.04351 | 0.002016205 | 21.58012386 | 5.1206E-101 |
| Change in transaction amount (Q4 over Q1) | 0.064141 | 0.014487492 | 4.42731211 | 9.64152E-06 |

**Interpretations:**

The analysis performed showed that there was statistically significant evidence that gender affects the customer churn rate for banks.

The regression model developed from our analysis showed that there was statistically significant evidence that all nine variables with the highest correlation with attrition are important predictors of churning customers.

**Recommendations:**

Based on the analyses performed we have found statistically significant evidence that nine key variables are all significant in predicting churning customers and that gender has an impact on churning customers. From these conclusions, we recommend that banks focus less on the duration of time with the bank, customer age, credit card limit, and number of dependents and instead look at factors such as total transaction count in the last 12 months, change in transaction count, total revolving credit card balance, and other key variables noted to predict churning customers. Overall, these variables that relate to the activity of a customer within a year period are going to be most accurate for bank manager, and strategies for engagements with the client should be centred around these factors to prevent them from churning.