FINAL PROJECT

# ANALYSIS OF THE VARIABLES OF CREDIT CARD CUSTOMERS TO PREDICT THE CUSTOMERS WHO ARE LIKELY TO DROP THEIR SERVICES FROM THEIR CREDIT CARD PROVIDERS

Rekik Ziku
Guided By - Dr. Colby Long
DATA 325: Applied Data Science
The College of Wooster, Wooster, OH

May 4, 2021

# Contents

**Introduction** **3**
Hypothesis . . . 3
Goal . . . 4

**Causes of Short Term Credit Card Usage** **4**
What is Unethical About Credit Cards? . . . 4

**Data Preprocessing** **4**
Relevant Variables . . . 5

**Methods** **5**
Assumptions . . . 5
Modeling . . . 5

**Analysis** **6**
Customer's Age . . . 6
Total Transaction Amount and Count . . . 7
Average Utilization Ratio . . . 8
Gender . . . 9

**Conclusion** **10**
Limitations . . . 10
Future Work . . . 10

**Appendix** **12**

**References** **12**

**Abstract**

In this paper, we will we will investigate which categories of customers tend to terminate their subscriptions from their credit card providers by using the Bank Churners data set [6]. We will discuss the causes of short term subscription and how we used the data to analyze the variables. We will then discuss our assumptions then create models to compare their accuracy to see which model is best fit. We will finally use that model to choose a few variables that will help us predict the reason behind customers dropping off from their credit card providers and to predict customers who are likely to drop off.

# Introduction

A credit card is a financial card that allows a user to borrow some money from the bank to purchase any products and return the amount to the bank when requested, usually monthly. By using the data set, "BankChurners.csv" that was originally from LEAPs and was found from Kaggle.com, we can see the different variables of a credit card customer. It also indicates that each year, higher percentage of customers tend to leave from their credit card services.

Dropping from a credit card provider might happen because of multiple reasons. Some customers are involved in credit card churning in which they are "frequently applying for new credit cards, not necessarily to use or even keep them, but rather to take advantage of lucrative sign-up or welcome bonuses in the form of cash or rewards miles or points" [4]. According to Experian, even if credit card churning is legal and somewhat beneficial, it might negatively affect the customers once their card issuers find out about their actions.

Another reason that causes customers to discontinue from their credit card services is the little understanding of credit cards and their uses before obtaining a credit card. Credit card companies tend to target and market their credit card service to undergraduate students without being fully transparent with them about the terms and conditions. Due to this, customers will keep their cards for a short period of time and cancel their subscription once they understand how it has affected them.

The data set overall contains 10,000 customers and has 18 different columns of the customers information like age, gender, income etc. We will be mainly analyzing which category of people are more likely to get a credit card and out of those people, which once are more likely to drop it. We will use the results to try to predict if some customers are how long a customer will maintain their credit cards.

We created a Variable Importance Plot for our random forest model and found that Customer Age, Total Transaction Amount, Total Transaction Count, Average Utilization Ratio, and Gender were the top five variables that predicted the Attrition Flag. We will mainly look into those variables to further understand how they will help us which types of customers tend to maintain their credit cards for a long term and which do not.

### Hypothesis

We expect the customers age, education level and Total Transaction Amount to have more effect on the number of months customers have their credit cards for and if the customers account is still active.

### Goal

Our ultimate goal is to analyze the data, Bank Churners, to find out the reason behind customers dropping off from their credit card providers and to predict customers who are likely to drop off.

## Causes of Short Term Credit Card Usage

There are multiple causes that affects the time length of the use of a credit card. To list a few: overspending and debt, credit card churning, credit card fraud, high interest rates, annual fees, and misunderstanding the use of credit cards. In this section, we will mainly focus on how some people might get into using credit cards without fully understanding the methods and how some customers only get their credit cards for a short period of time and only use the new card rewards.

### What is Unethical About Credit Cards?

In the book "Ethics in Finance," John R. Boatright discusses the issues with credit cards and how it might negatively impact the customers. First and foremost, he argued that these cards "should be made available to consumers with full, accurate disclosure of relevant information without deception, concealment or guile." He stated that despite the information being provided when a consumer starts using a credit card,it is usually unreadable as the terms are complicated and hard to understand for an average person. This will increase the debt and legal issues faced by a customer. In addition, credit card companies target college students to convince them to utilize credit cards. However, many college students get their first credit card and excessively use it until their debt starts to pile up. Hence, the younger customers tend to drop their credit card services sooner.

### Credit Card Churning

Credit Card Churning is the act of "repeatedly opening and closing credit cards to earn cash, rewards points or miles" [4]. In order to operate churning, one will obtain a credit card and only use it until the intro bonus is complete, after that, they will terminate their use of this card and apply for another one. Although this is not illegal, it is considered controversial by many and is not appreciated by credit card providers. In response to this credit card companies have included multiple factors in their Terms and Condition in order to prevent credit card churning from happening. In addition to this, card issuers might terminate the client’s entire account or request that they repay the rewards if they’re caught in this act.

# Data Preprocessing

By using the Bank Churners data set that was initially made by LEAPs and that I found on Kaggle [6], we will investigate which categories of customers tend to terminate their subscriptions from their credit card companies. For the predictive variable, we mainly focus on: The Customers Age, The Total Transaction Amount and Count, Average Utilization Ratio, and Gender. For the response variable, we will mainly focus on: Months spent with the credit card company and if the customers account is still active. The table below shows the types of these categories.

From the given data set, we removed Client Number because we thought it irrelevant for our analysis. In addition, we removed the last two columns from the given data set (Native Base), because the provider of this data, Goyal mentioned that the last two columns are random and not relevant to the data set. While analyzing this data set, we looked at the rest of the categories. From the remaining categories, we will be looking at Months on Book and Attrition Flag as the response variable and the other 18 categories as the predictor variables.

## Relevant Variables

The Customers Age is represented by a whole number where the youngest was 26 years old and the oldest being 73. The Total Transaction Amount is the toal amount of money a customer has taken out or returned since they first obtained the card. Total Transaction Count is the number of times a customer has used their credit card since they first obtained the card. Average Utilization Ratio is a way of referring to a customer's average credit card debt by measuring how much debt they currently have on their credit cards compared to their credit card limit. Gender in this data set is represented as Male and Female. Months on Book is the number of months a customer has spent with the credit card provider. Attrition Flag shows if the customer still exists or if they have discontinued from their service.

Note: The name of the variables, their type and a brief description of their purpose is provided in the table at the end of this paper.

# Methods

## Assumptions

This data set was created in Kaggle on September, $19^{th}$ 2020. However, the dates this data set was collected is not provided. In order to get solutions for normal conditions, we assumed that this data was collected during years with This means that we did not consider situations like the 2008 recession and the 2020 pandemic. In short, there are no extreme factors that financially affected the customers. Thus, we assumed the number of new customers and customers terminating their services is normal and not affected by external factors.

## Modeling

From all the variables, the column, "Months on book" and "Attrition Flag" will be used as the response variables, so we will be comparing other variables to months on book and attrition flag to see the amount of time a person spends with a credit card and what factors contributes to terminating the use of credit cards.

To make the predictions more accurate, we created Supported-Vector Machines, Random Forest and Logistic Regression models for both Months on Book and Attrition Flag, then we compared their accuracy rate to see which method will help us predict better. We then found that Random Forest for Attrition Flag has the best accuracy rate with a 54.39% accuracy.
Note: The table below shows the the models and the accuracy rate they produced.

**Accuracy Rate**

| Predictive technique | Accuracy |
|---|---|
| Random Forest | 54.39% |
| Supported-Vector Machines | 44.45% |
| Logistic Regression | 15.73% |

**Figure 1:** Accuracy Rate

We then, created a variable importance plot for the random forest to help us predict which variables have the biggest influence on whether the customers keep their credit cards and for how long an individual they will do it. From our variable importance plot, the top five variables affecting the accuracy of the random forest were **Customer Age, Total Transaction Amount, Total Transaction Count, Average Utilization Ratio, and Gender**. In the next section, we will present visuals and explanations to show how we can use these variables to predict the months on book spent and the attrition flag.

# Analysis

In this section, we will dive into the analysis of Customer Age, Total Transaction Amount, Total Transaction Count, and Average Utilization Ratio, and Gender in comparison to Months on Book and Attrition Flag in order to find meaningful trends.

## Customers Age

In the data set, the customer's age was represented by a whole number where the youngest was 26 years old and the oldest being 73. The average age for customers to get a credit card is 46 according to the data set.

As shown on Figure 3, we can say that there is a linear relationship between Customers Age and Months on Book. As the age of the customer increases the amount of time the customer spends with that credit card provider also increases. From this, we can say that an increase in age will cause increase in months on book, however we cant say if the account is still existing because it's randomly situated.

Although the youngest customer in this data set is 26 year old, we also considered the credit card use by 18 - 24 years old, generally undergraduate students. This is because despite not being stated in the data, we can see from Figure 3 that some of the younger age group of people have had their credit cards for more that 20 months. This shows that they obtained their credit card as they were under 24 years old. A report produced by Nellie Mae called "Undergraduate Students and Credit Cards" shows that there is an increase in credit card usage by undergraduate students in 2001 by 24% since 1998. Similarly, the amount of debt they are in This report also discusses that increasing number of students have four or more credit cards and an increasing amount of loan. In addition to that the median credit card debt also increases as students progress through their four years of college. For example, a sophomore is in more debt than a freshman. Overall, there was an increase in the median loan by 43% from 1998 to 2001.

This happens for multiple reasons. One being credit cards are designed for people with fixed income since you have to pay back monthly. As discussed in the report, "the easy availability of credit cards, limited income while in school, and a greater comfort level with the accumulation of debt are contributors to increased credit card use among undergraduates." In addition to that, most customers in this age group are not educated well enough about credit card usage and the additional fees and interest rates. Although this study is two decades old, we can infer that this issue has increased ever since as credit cards are abundant and possessed by many more people, and the increase of college tuition.


**Figure 2:** Customer Age vs Months on Book and Attrition Flag

## Total Transaction Amount and Count

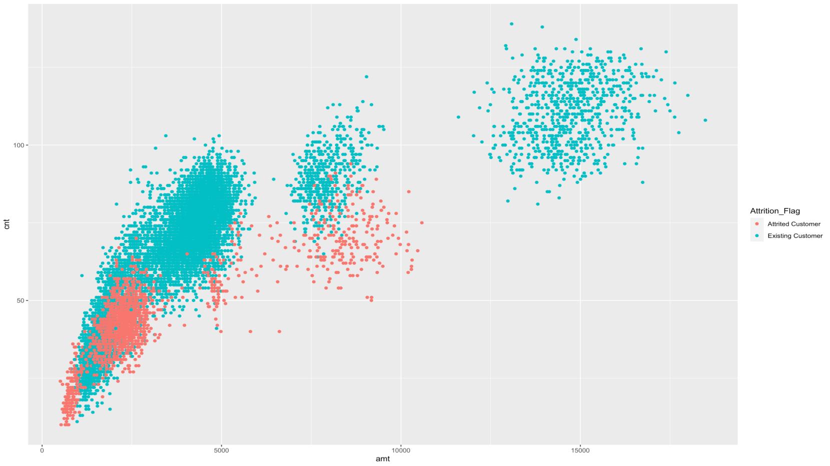


**Figure 3:** Total Transaction Count versus Total Average Count

In Figure 3, we can see that the increase in transaction amount indicates an increase in transaction count and vice versa. In addition the more transaction count and amount indicates that the customers is most likely to continue their subscription to their credit card company. This happens because of credit card limits and length of time. Since there are credit card limits assigned to each person, a customer will only use a limited amount of money at a time. Because of that, in order to take out more money, the customer needs to make additional transactions. This will increase both transaction amount and count. In addition to that, as time progresses, if that customer is an active user of their credit card, they will definitely have bigger number of transactions in comparison to a customer that only spent a few months with the bank. From this scatter plot, we can conclude that the increase of transaction count will indicate higher transaction count, and as both go higher, customers are less likely to terminate their credit cards.

## Average Utilization Ratio

Credit utilization is a way of referring to a customer's credit card debt by measuring how much debt they currently have on their credit cards compared to their credit card limit [8]. The average utilization ratio in this data set is calculated by finding the credit utilization ratio a customer had over the years they were a member of this credit card company.

$$\text{Credit Utilization Ratio} = \text{Credit Card Debt} \div \text{Credit Card Limit [8]}$$

As shown in the classification tree, a customer is considered to have terminated from their credit card company most likely have a total transaction count less than 57.5 and Average Utilization Ratio less than 0.0235 (2.35%). This indicates that if a customer has a low transaction count and a low utilization amount, that customer is likely to drop from their credit card services. As discussed above, customers who have lower transaction count are most likely to terminate there credit card.

It is recommended to have a low utilization ratio, preferably less than 30% [1], in order to be in a good standing and be able to apply easily to other credit card companies. In this data set, the customers who tend to terminate their credit cards have an Average Utilization Ratio of 2.35% or less. This suggests that these customers mostly pay their credit card balances in full, thus decreasing the amount of debt they're in. From here we can infer that these customers are not getting as much benefit from this loan program as they would expect. Hence, this is the cause of people with low average utilization ratio terminating from their service.

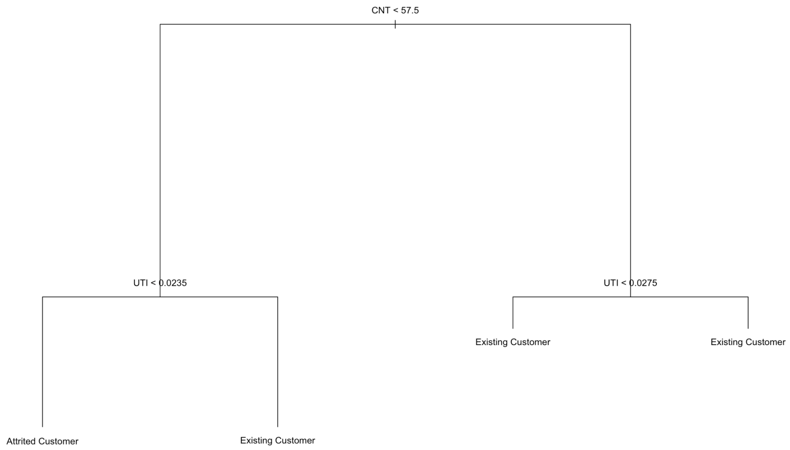


**Figure 4:** Average Utilization Ratio and Total Transaction Count versus Attrition Count

## Gender

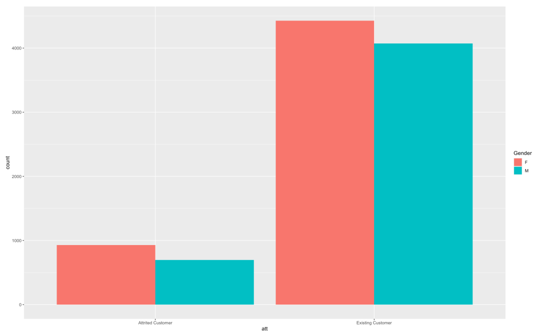

**Figure 5:** Gender vs Attrition Flag

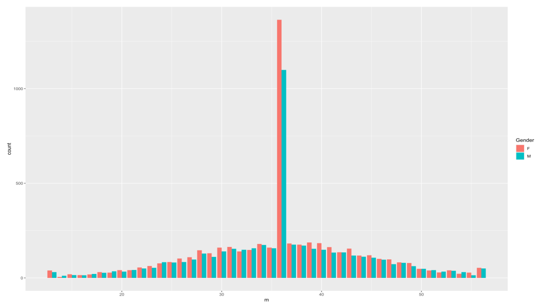

**Figure 6:** Gender vs Months on Book

Gender in this data set is categorized as Male and Female and generally there are slightly more females in the data set as opposed to male. In addition, as we can see in Figure 5 that the number of females is slightly bigger than males for both categories of Attrition Flag, which is equivalent to the total population. Even though there are a couple instances in Figure 6 where the number of males in a month is slightly bigger than the number of females, females have more or less spent more time overall. According to Kaya, E., Dong, X., Suhara, Y. et al in Behavioral attributes and financial churn prediction male and female customer groups have no significant superiority over each other in terms of predictability [9]. Therefore, we can conclude that there is no visible trend that shows that Gender has a relation with who tends to drop from their credit card services.

# Conclusion

By creating a Variable Importance Plot for the random forest model, we obtained the top five variables that were relevant to predicting Attrition Flag, namely, Customer Age, Total Transaction Amount, Total Transaction Count, Average Utilization Ratio, and Gender. We then compared each variables to Attrition Flag and Months on Book to find any relevant trends between each variables.

From the comparison, we found that the relationship between customer's age and Months on Book is direct, when the age of the customer increases, the number of months spent with the credit card provider also increases. However, there was no visible relationship between customer's age and attrition flag. We also found that the increase in transaction amount indicates an increase in transaction count and vice versa. Besides, the higher transaction count and amount indicates that the customers are most likely to continue their subscription to their credit card company for a long term.

For the comparison of the Average Utilization Rate, we found that a customer is most likely to discontinue their credit cards, if they have a total transaction count that is less than 57.5 and Average Utilization Ratio less than 2.35%.However, in terms of trends for gender, we were not able to find a relevant trend that helps us predict the dedication to the bank of the customer.

In the beginning of this project, I hypothesised that customers age, education level and Total Transaction Amount will have more effect on the data set to help predict if the customer is more likely to maintain their membership with the bank. Although customer's age and total transaction amount, among other variables, were beneficial to predict the chances of a customer dropping from their credit card providers, the customer's education level seemed to be quite irrelevant in predicting if the customer will stay with the bank for a long term.

## Limitations

There are many external factors that affect the chances of a customer remaining with the credit card provider. How a customer obtain the credit card and if they were initially fully informed about the usage of this credit card will mostly affect how long they will use the credit cards for. In addition, knowing the exact time this data set was collected is quite relevant to understand if there were any additional external factors that affected the usage of credit cards by these customers.

## Future Work

In order to make this prediction more accurate, we can look into additional variables that affect the chances of a customer remaining with the credit card provider for the long run. This might require an in depth interview with the customer and/or additional information on how they first obtained the card, if they had any prior knowledge about credit cards before that, and In addition, it will also be vital to know the year the data was collected in order to see if there are any economical changes that might have affected the long term uses of credit cards.

# Appendix

| | Variable | Type | Description |
|---|---|---|---|
| 1 | Customer Age | Numerical | Customer's Age in Years |
| 2 | Total Trans Amount | Numerical | Total Transaction Amount (Last 12 months) |
| 3 | Total Trans Count | Numerical | Total Transaction Count (Last 12 months) |
| 4 | Average Utilization Ratio | Numerical | Average Loan Amount |
| 5 | Months on Book | Numerical | Period of relationship with bank |
| 6 | Dependent_count | Numerical | Number of Dependents |
| 7 | Total Relationship Count | Numerical | Total no. of products held by the customer |
| 8 | Months Inactive | Numerical | No. of months inactive in the last 12 months |
| 9 | Contacts Count | Numerical | No. of Contacts in the last 12 months |
| 10 | Credit Limit | Numerical | Credit Limit on the Credit Card |
| 11 | Total Revolving Balance | Numerical | Total Revolving Balance on the Credit Card |
| 12 | Average open to buy | Numerical | Open to Buy Credit Line (Average of last 12 months) |
| 13 | Total Amount of Change | Numerical | Change in Transaction Amount (Q4 over Q1) |
| 14 | Total Count Change | Numerical | Change in Transaction Count |
| 15 | Attrition Flag | Categorical | If the account is open or closed |
| 16 | Education Level | Categorical | Educational Qualification of the account holder |
| 17 | Maritial Status | Categorical | Married, Single, Divorced, Unknown |
| 18 | Income Category | Categorical | Annual Income Category of the account holder |
| 19 | Card Category | Categorical | Type of Card (Blue, Silver, Gold, Platinum) |
| 20 | Gender | Categorical | If the customer is Male or Female |

**Figure 7:** Categories Table

# References

[1] What is a credit utilization ratio?

[2] Computer assisted customer churn management: State-of-the-art and future trends. Master's thesis, October October 2007.

[3] John R. Boatright. *Ethics in Finance.* John Wiley and Sons Inc, 350 Main Street, Malden, MA, 3rd edition, 2014.

[4] Louis DeNicola. What is credit card churning? *Experian*, February 2021.

[5] Robin Saks Frankel. What is a credit card? *Nerdwallet*, October 2020.

[6] Sakshi Goyal. Bankchurners.csv.

[7] Peng Zhang Yingjie Tian Yong Shi Guangli Nie, Guoxun Wang. Finding the hidden pattern of credit card holder's churn: A case of china. Master's thesis, Research Center on Fictitious Economy and Data Science, Springer, Berlin, Heidelberg, 2009.

[8] Latoya Irby. What is the credit utilization ratio? *The Balance*, October 2020.

[9] Dong X. Suhara Y. et al Kaya, E. Behavioral attributes and financial churn prediction. *EPJ Data Sci.*, 2018.

[10] Fareed Shaukat Memon Malick M. Abu Rehan Alvi Syed Faheem Hasan Bukhari, Ayesha Ismail. Global journal of management and business research: E marketing. *Global Journals Inc*, 15, 2015.