Which cardholders should receive a retention offer?

The answer to this question can be answered by identifying which credit card customers are most likely to cancel their accounts.

Introduction

Credit card companies generate most of their revenue through fees charged to their cardholders. These fees may include interest charged to balances carried over from the previous month, late fees if the minimum payment amount is not paid by the due date, an annual fee to keep the account open, cash advance fees charged when cards are used to obtain cash, balance transfer fees charged when the balance of one card is transferred to another, etc. In 2019, U.S. credit card companies made approximately $180 billion through customer fees alone (Egan, 2020).

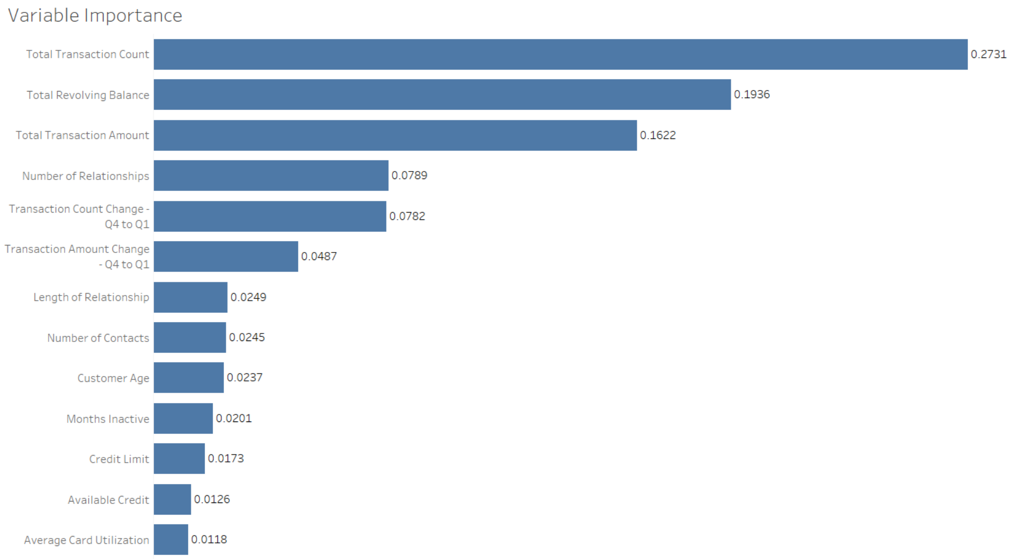
Business Problem

Customer attrition, or customer churn, can be defined as the rate at which customers stop doing business with a company. Customer churn is an important metric to consider since it is generally more expensive to acquire new customers than to retain existing customers. By determining which factors contribute to customer churn, the company can identify individual cardholders who are likely to cancel their credit card. In turn, the company can extend a retention offer to these customers to encourage them to retain their card.

Methods

I obtained the Credit Card Customers dataset from Kaggle, which contained over 10,000 credit cardholder records. It contained 20 variables, including the target variable *Attrition_Flag* – a categorical variable with binary classes, *Existing Customer* and *Attrited Customer*. First, I split the data into a training, validation, and test sets. The training set contained 70% of the data and the remaining 30% was split evenly between the validation and test sets. After converting the categorical variables (*Gender*, *Education_Level*, *Marital Status*, *Income_Category*, *Card_Category*), I used trained a decision tree on the training data to determine feature importance. The selected the variables that represented 97% importance, as seen in *Figure 1*. Using these variables, I trained and fit five different classification models on the validation data. To evaluate the models, I used the Area Under the Precision-Recall Curve and Negative Log-Loss. The Area Under the Precision-Recall Curve is stable under class-imbalance. Therefore, if the number of cardholders who cancel their account makes up a small percentage of the dataset, this metric will allow me to focus on the

Figure 1: Variables with top 97% importance



trade-off of precision and recall instead of giving credit for true negative predictions that often overwhelm the set of predictions. This allows for additional requirements to potentially quantify the impact of Type 1 (customers predicted to cancel their card but do not) and Type 2 (customers predicted to keep their card but cancel) errors to set a threshold of prediction. Average precision is a metric with scores between 0 and 1, where higher is better. Negative Log-Loss uses calibrated probabilities to assess how confident and correct the model is on average. It heavily penalizes confident and wrong predictions. Negative Log-Los is a metric of positive numbers, where lower is better. I also noted the accuracy of each model. The results can be seen in *Table 1*.

Table 1: Evaluation metric values for each classification model

Model	AUC Precision-Recall	Log Loss	Accuracy
Naïve Bayes	0.551	3.16	0.908
Decision Tree	0.672	2.23	0.935
Random Forest	0.789	1.36	0.961
Gradient Boosting	0.823	1.14	0.967
XGBoost	**0.840**	**1.02**	**0.970**

Results

The XGBoost classification model performed better than the other models. Therefore, I concatenated the training and validation data and performed a cross-validation randomized grid search to choose the best parameters. The fine-tuned XGBoost model was fit on the test data and produced the results seen in *Table 2*.

Table 2: Evaluation metric values of selected model after fine-tuning of parameters

Model	AUC Precision-Recall	Log Loss	Accuracy
XGBoost	0.847	0.98	0.972

The classification results can be observed in the confusion matrix (*Figure 2*). Of the 1520 records in the test data, 209 customers were correctly classified and 9 were incorrectly classified as closing their account. Furthermore, 34 customers were incorrectly classified as existing cardholders but have closed their account. However, approximately 86% of customers at risk of churn were identified. The credit card company can extend retention offers to these cardholders; however, the additional 14% represents missed opportunity. The Area Under the Precision-Recall Curve can be seen in *Figure 3*.

Figure 3: Confusion matrix of results

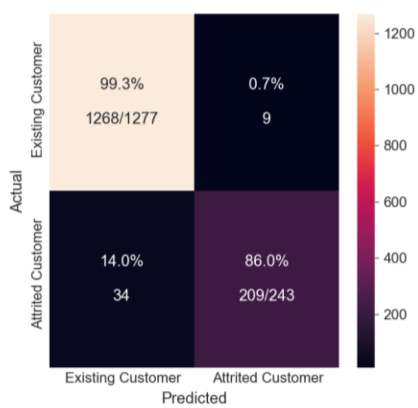
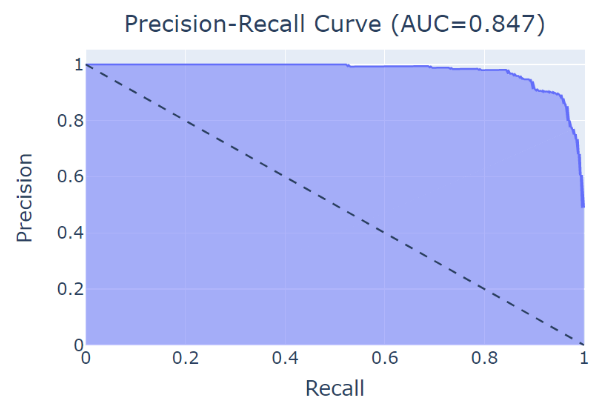


Figure 2: Area under the Precision-Recall Curve



Conclusion

Once cardholders that have been identified as *Attrited Customers*, the company must decide how they will proceed with extending retention offers. It may choose one or multiple approaches to compare the effectiveness of each. Based on the responses of each cardholder, further analysis through uplift modeling can be conducted to determine which customers are most likely to respond to the retention offer. During my research, I discovered two groups of cardholders to which company may not wish to extend retention offers: inactive cardholders and “churners”. First, credit card companies can cancel customer accounts due to inactivity. This is because they only have so much credit they can extend to their customers. They cannot simply give lines of credit to new customers without canceling others (Rathner, 2020). Therefore, the company may also be contributing to customer churn. Second, there are some customers who actively use credit card churn as a strategy to take advantage of card benefits. They open new credit cards accounts to take advantage of the initial signup benefits, cancel their card once the benefit is received, open a new account, and repeat the process. Prior to making a retention offer, the company may want to exclude cardholders who fall into one of these two groups.

References

Alpine Jennings, P. (2015, December 25). The 4 D's of Customer Attrition. Retrieved March 17, 2021, from https://thefinancialbrand.com/55772/banking-customer-attrition-analysis/

Egan, J. (2020, July 03). How Do Credit Card Companies Make Money? Retrieved March 18, 2021, from https://www.creditcards.com/credit-card-news/how-credit-card-companies-make-money/

Goyal, S. (2020, November 19). Credit Card Customers. Retrieved March 17, 2021, from https://www.kaggle.com/sakshigoyal7/credit-card-customers

Hurd, E. (2021, February 16). Should I Try Credit Card Churning? Retrieved March 17, 2021, from https://www.nerdwallet.com/article/credit-cards/credit-card-churning

Hurwood, J. (2016, November 30). Credit Card Colour Comparison: What's The Difference? Retrieved March 26, 2021, from https://www.canstar.com.au/credit-cards/colour-credit-card-mean/

Irby, L. (2020, October 7). What is the credit utilization ratio? Retrieved April 03, 2021, from https://www.thebalance.com/what-is-a-good-credit-utilization-ratio-960548

Johnson, H. (2020, February 27). What is Credit Card Churning? Retrieved March 17, 2021, from https://www.bankrate.com/finance/credit-cards/what-is-credit-card-churning-and-how-does-it-affect-your-credit-score/

Johnson, H. (2020, April 21). Before You Cancel Your Premium Credit Card, Ask About a Retention Offer of Bonus Points or a Reduced Annual Fee. Retrieved March 18, 2021, from https://www.businessinsider.com/personal-finance/credit-card-retention-offers

Lambarena, M. (2020, December 08). How Do Credit Card Companies Make Money? Retrieved March 18, 2021, from https://www.nerdwallet.com/article/credit-cards/credit-card-companies-money

Rathner, S. (2020, September 22). Credit Card Closed for Inactivity? What You Need to Know. Retrieved March 18, 2021, from https://www.nerdwallet.com/article/credit-cards/credit-card-cancelled-due-inactivity

Reducing Attrition Among Credit Card Clients. (n.d.). Retrieved March 17, 2021, from https://www.sas.com/en_nz/customers/tatra-banka.html

Tripathi, S. (2018, November 30). How Can Analytics Help Control Attrition in the Credit Card Industry? Retrieved March 17, 2021, from https://analyticstraining.com/how-can-analytics-help-control-attrition-in-the-credit-card-industry-2/

U.S. News Staff. (2021, March 15). 2019 Credit Card FEE Study: What's Normal and What's Not? Retrieved March 18, 2021, from https://creditcards.usnews.com/articles/fee-survey

Appendix

Variable Description

CLIENTNUM: Client number. Unique identifier for the customer holding the account

Attrition_Flag: (Target variable) – Existing Customer/Attrited Customer

Customer_Age: Customer's age in years

Gender: M=Male, F=Female

Dependent_count: Number of dependents

Education_Level: Educational Qualification of the account holder - Uneducated, High School, College, Graduate, Post-Graduate, Doctorate, Unknown

Marital_Status: Married, Single, Divorced, Unknown

Income_Category: Annual income caetegory of the account holder - Less than $40K, $40K - $60K, $60K - $80K, $80K - $120K, $120K + and Unknown

Card_Category: Type of card (Blue, Silver, Gold, Platinum)

Months_on_book: Period of relationship with bank

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 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: Available credit (Average of last 12 months)

Total_Amt_Chng_Q4_Q1: Change in transaction amount (Q4 over Q1)

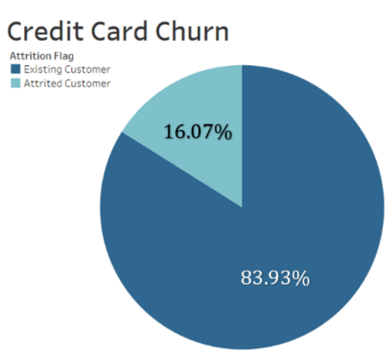
Total_Trans_Amt: Total transaction amount (Last 12 months)

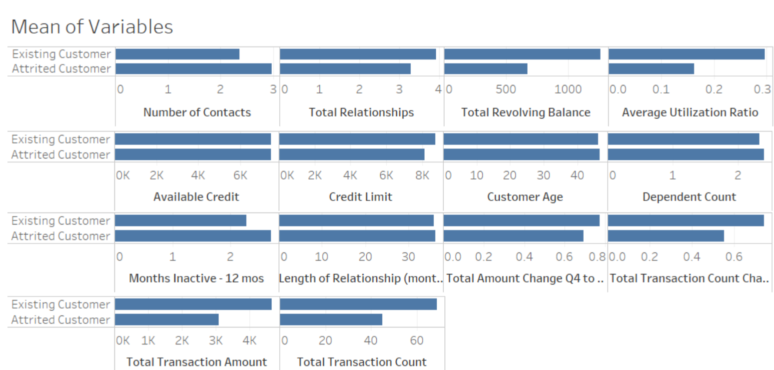
Total_Trans_Ct: Total transaction count (Last 12 months)

Total_Ct_Chng_Q4_Q1: Change in transaction count (Q4 over Q1)

Avg_Utilization_Ratio: Average card utilization ratio

Additional Figures





Q&A

Were there any missing data and how did you handle them?

Fortunately, the dataset was not missing any data. However, the categorical variables *Education_Level*, *Marital_Status*, and *Income_Category* each had an "unknown" class. I thought about removing these observations from the dataset but decided it was best to keep them as future observations may also contain unknown data for these variables.

What type of retention offers could be made to identified customers?

Reasons for customer churn may vary. The company may choose to use different approaches based on the individual customer. Some options may be offering bonus rewards, statement credits, lower interest rates, increased credit limit, and annual fee waiver or reduction. The offer may vary based on the individual customer. For example, if a customer has a high revolving balance, the company may suggest other credit cards with lower interest rates and a waived transfer fee. If the goal is to get the customer to use their card to make more transactions, the company may offer bonus rewards points for using their card at certain locations or by spending a certain amount.

How can inactive cardholders be identified?

Inactive cardholders may be identified by looking at *Months_Inactive_12_mon*.

Why would a credit card company not want to extend a retention offer to an inactive cardholder?

Credit card companies do not benefit financially when their cardholders are inactive. Furthermore, they may not be able to recruit new customers if they do not have the line of credit available.

How can "churners" be identified?

"Churners" may be a bit more difficult to identify. Some variables that could help gain insight are *Card_Category* (blue, silver, gold, or platinum), *Income_Category* (less than $40K, $40K to $60K, $60K-$80K, $80-$120K, and $120K+), *Avg_Open_To_Buy* (the average of the difference of credit limit and monthly balance), and *Months_on_book* (how many months the customer held a relationship with the company).

Why would a credit card company not want to extend a retention off to a "churner"?

"Churners" may be more likely to take advantage of any offers provided by the company. Though the goal is to retain customers, the "churner" will most likely cancel their account once the benefit is received.

What challenges did you face during your analysis?

The variables *Total_Revolving_Bal* and *Credit_Limit* were used to create two of the other variables in the dataset. *Avg_Open_To_Buy* was calculated by subtracting *Total_Revolving_Bal* from *Credit_Limit*, and *Avg_Utilization_Ratio* was calculated by dividing *Total_Revolving_Bal* by *Credit_Limit*. My original thought was to remove *Credit_Limit* and *Avg_Open_To_Buy.* I also wanted to create two new variables, *Avg_Trans_Amt* and *Avg_Amt_Chng_Q4_A1,* by dividing *Total_Trans_Amt* by *Total_Trans_Ct* and *Total_Amt_Chng_Q4_Q1* by *Total_Ct_Chng_Q4_Q1*, and removing the original two variables. After making these changes to my data, the results were much lower than those of the data without the changes. Therefore, I decided to keep the data as it was.

What observations did you make that you found interesting?

Credit limit was generally greater based on the card category. The Blue, Silver, Gold and Platinum cards have a credit limit average of $7,382, $25,236, $28,420, and $30,626, respectively. It also increased as income increased. The average credit limit was considerably lower for women ($5,039) than men ($12,705). Initially, I believed this was a major concern that could cause bias. However, I discovered that all of the female cardholders had incomes of $60K and below. Therefore, I felt the difference of credit limit was based off income, which is a problem out of the scope of the credit card company. The average utilization ratio compares the revolving balance to the credit limit. It tells how much is currently being borrowed compared to how much could be borrowed (Irby, 2020). I noticed the average utilization was much higher for women (34.3%) than men (19.8%).

What is the importance of the average utilization ratio?

Average credit utilization directly affects one's credit score, accounting for approximately 30% of it. Credit utilization below 30% are generally considered "good". When someone applies for a credit card, the company takes their credit score into consideration. If they have poor credit, it may be difficult to gain approval. If they are approved, they will most likely be given a high interest rate. I would be interest to know the interest rates associated with each observation.

Is there any additional data that you feel would help with your analysis?

It would have been helpful to know if an account was voluntarily or involuntarily closed. Also, the card category tells whether the card was a blue, silver, gold, or platinum card. It would be interesting to know the details about each card type such as any associated benefits and fees. Based on my research, blue is usually a standard card. Silver and gold are similar, with gold having a few additional features. Platinum cards tend to have reward benefits and typically have high interest rates.