Modeling Credit Card Churning from Customer Data

Introduction:

In recent years, credit cards have been an essential tool for financial transactions and purchases. With the increase of a competitive credit card market, customers tend to churn their credit cards more often. Customer churn, also known as customer attrition, refers to the rate at which customers switch from one service provider to another. Therefore, understanding the factors that influence credit card churn is crucial for credit card companies to retain customers and maintain a competitive edge in the market. In this study, we aim to investigate how customer characteristics such as age, gender, marital status, credit card type, income category, utilization ratio, and transaction count impact the likelihood of credit card churn. By examining these factors, we hope to determine which combination of factors leads to the most accurate prediction of customer churning. Doing so will allow banks to have a stronger understanding of customers who may be a higher liability when opening a credit card.

Analysis:

This data, presented by Predicting Credit Card Customer Segmentation, was collected from a consumer credit card portfolio of an undisclosed business who was dealing with customer attrition problems. The dataset was published on December 15, 2020 and contains information pertaining to each customer such as their age, gender, marital status, total transaction count, income category, card category, and average utilization ratio of the credit card. Additionally, the data also contains information whether the customer has churned or not, which may be correlated with some of the customer data. This analysis focuses on these variables to determine which are significant in providing an accurate prediction of whether a customer will churn or not. The following table explains the selected variables of the dataset to focus our analysis on where Attrition_Flag is the response variable and the remaining are explanatory variables. Note, the Attirtion_Flag of 0 indicates that a customer has churned.

Table 1. Variables from the dataset to perform model analysis

Variable Name	**Description**
Attrition_Flag	Flag indicating whether or not the customer has churned out (0 if yes, 1 if no). (Boolean)
Customer_Age	Age of customer (Integer)
Gender	Gender of customer. (String)

Marital_Status	Marital status of customer. (String)
Income Category	Income category of customer (6 levels: <$40 000, between $40 000 - $60 000, $60 000 - $80 000, $80 000 - $120 000, >$120 000, Unknown)
Avg_Utilization_Ratio	Average utilization ratio of customers defined as $\frac{amount\ of\ credit\ in\ use\ (\$)}{total\ credit\ available\ (\$)}$. (Numeric)
Total_Trans_Ct	Total transaction count. (Integer)
Card_Category	Type of card held by customer. (String)

To analyze the data, we decided to do multiple data model selection methods like comparing graphically the attrition to the individual variables that we have considered, comparing AIC results, and using the BIC, and Residual Deviance model results.

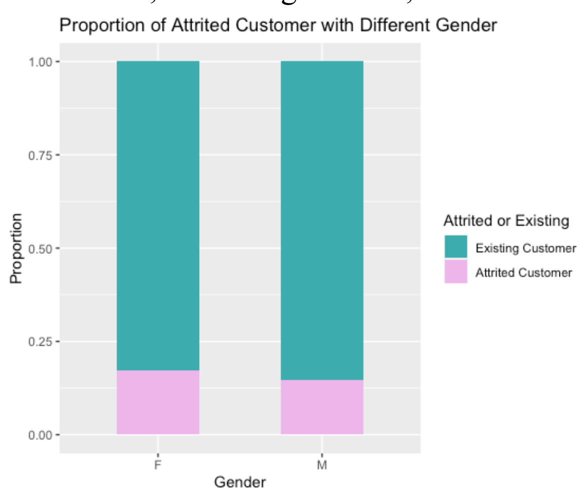

Figure 1: Proportion of Attrition by Gender

We begin by exploring the distribution of the data with the Attrition_Flag to see if there are any graphically noticeable patterns. In figure 1, we explore the differences between gender and attrition proportion. From the plot, it seems the proportion of males that churned out was ~11% and ~13% for females. This suggests there may not be a significant difference in the likelihood of attrition between genders.

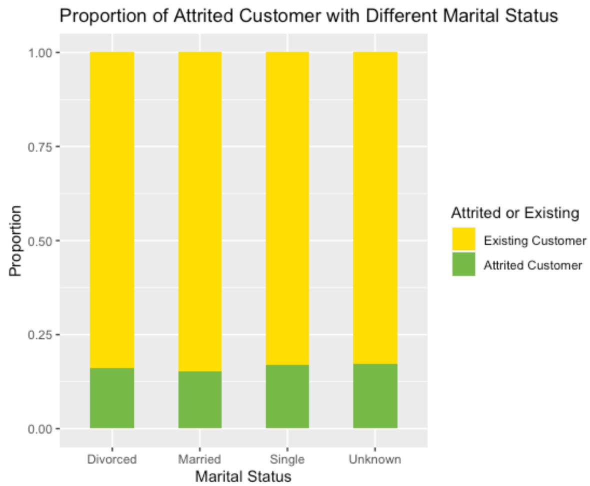

Figure 2: Proportion of Attrition by Marital Status

Now exploring the attrition proportion by marital status in figure 2, we see little difference between the different groups of marital status. All groups seem to have an attrition rate of around 13-14% with the lowest attrition rate being the married group and highest being the single and unknown groups.

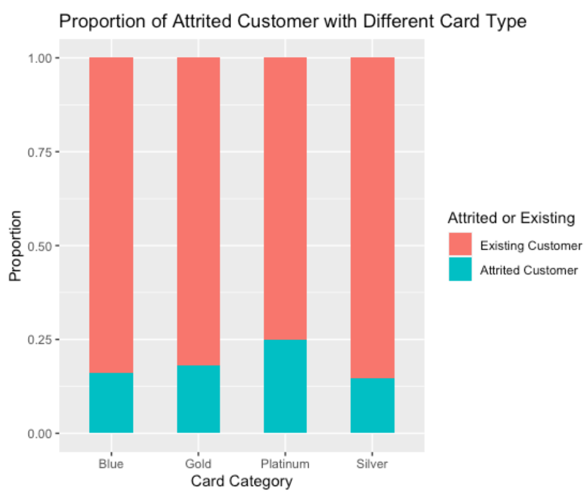

Figure 3: Proportion of Attrition by Credit Card Type

Taking a look at the relationship between the card type and attrition rate in figure 3, we see some significant difference between the groups. In particular, the proportion of customers who churned with platinum card types seems to be higher than the rest of the type of cards. The proportion of attrition in the other card types do not seem to be significantly different from each other. The highest tier card (Platinum) tends to be the target for churners as they provide the most benefits, which might explain why we see a higher proportion of attrition in this card type.

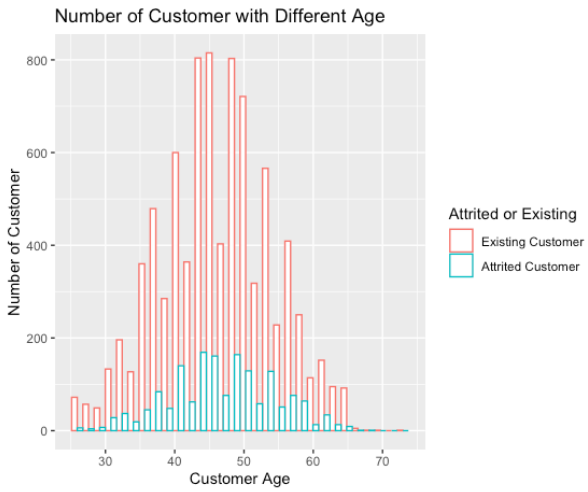

Figure 4: Proportion of Attrition by Age

From figure 4, it seems there is a normal distribution between the age and the number of customers attrited. The average age of those who have attrited is around 46 which is also the average age of the customers from this bank.

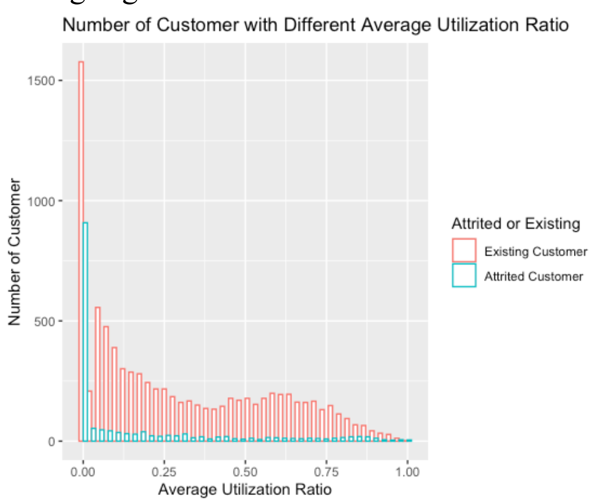

Figure 5: Proportion of Attrition by Utilization Ratio

From figure 5, we see the distribution of the average utilization ratio of existing customers and attrited customers. Most customers seem to have a low utilization ratio indicating most of available credit is not being used. Compared to the higher utilization ratios, it seems there is a higher proportion of attrited customers in the low utilization ratio bin.

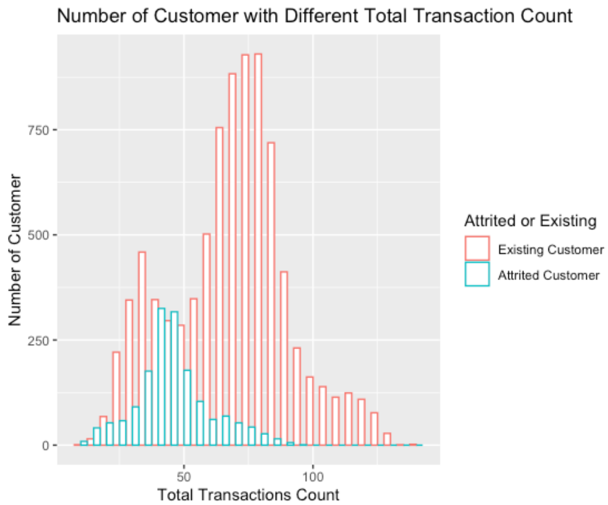

Figure 6: Proportion of Attrition by Total Transaction

From figure 6, we see that clients who have 30 to 60 total transactions seem more likely to churn their credit cards. In reference to figure 5, it seems like clients who have low utilization ratios and low transaction counts tend to churn their credit cards because the customers are reaping the benefits of the card and moving on to the next one.

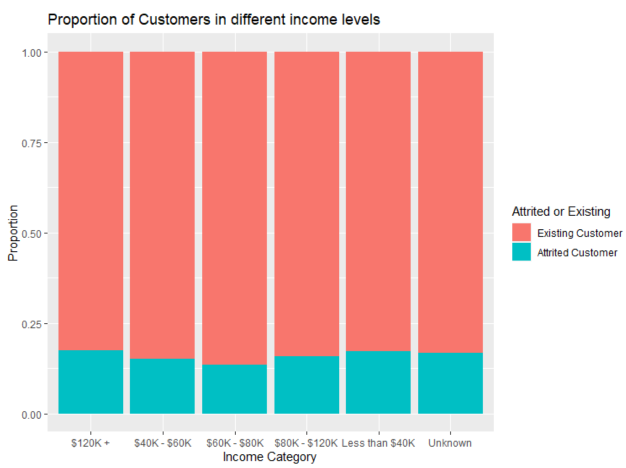

Figure 7: Proportion of Attrition by Income Level

Looking at the proportion of attrited customers for each income category, there does not seem a significant difference between the different income levels. The proportion attrited customers over all income levels seems to be between 13-18%. Although the difference may not be significant, the highest proportion of churning customers seem to be in the $120K+ income category and the lower proportion earning between $60k - $80k. From these graphical representations, credit card type, average utilization ratio, and total transaction seem to have some significance in modeling which customers are likely to churn credit cards.

To model the probability of credit card churning in customers in this dataset, we decided to use a logistic regression model. A logistic regression model is suitable here as the response variable (Attrition_Flag) is a binary variable and we are interested in which combination of variables will provide the most accurate prediction of customer attrition. After our graphical exploratory analysis with each of the explanatory variables, we decided to fit five different logistic regression models.

1. Full model with all explanatory variables (Age, Gender, Marital Status, Income Category, Card Category, Utilization Ratio, Transaction Count)
2. Model without Age
3. Model without Age and Income Category
4. Model without Income Category
5. Model with interaction between Income Category and Card Type

Additionally, we decided to explore the interaction between income category and card type in model 5 as we believe higher income customers would likely have a more premium card type which may influence the attrition likelihood. We also decided to limit our scope to these five models to keep the analysis tractable.

Fitting the logistic regression models in R and comparing AIC scores between models, we found the AIC values listed in table 2.

Table 2. AIC values for all 5 models

Model	**AIC**
Full model	6766.91
Model without Age	6766.97
Model without Age and Income Category	6772.90
Model without Income Category	6773.88
Model with interaction between Income Category and Card Type	6775.90

We notice that the AIC is largest in the model with interaction between Income Category and Card Type and the smallest in the full model. Since a lower AIC suggests a better model, the full model including all the explanatory variables may be the best in terms of modeling customer attrition. These AIC results also suggest the model with interaction between Income Category and Card

Type may not be as significant when it comes to predicting customer attrition and it may be overfitting the data.

After looking at the AIC values for each model, we decided to also look at the BIC values to see if both analyses agree on which model is the best. Table 3 lists the BIC values across the different models.

Table 3. BIC values for all 5 models

Model	**BIC**
Full model	6882.48
Model without Age	6875.32
Model without Age and Income Category	6845.12
Model without Income Category	6852.41
Model with interaction between Income Category and Card Type	6999.79

Looking at the results in table 3, we see the lowest BIC value being the model without Age and Income Category and the highest being the model with interaction. These results differ from the best model with AIC as BIC suggests the best model is without Age and Income Category. BIC tends to penalize additional parameters much more harshly so the model without Age and Income Category (which has the fewest parameters) results in the lowest BIC value.

Finally, comparing the deviance of the five models to see how well each model is predicting customer attrition we find the following values in table 4.

Table 4. Residual Deviance for all 5 models

Model	**Residual Deviance**
Full model	6734.91
Model without Age	6736.97
Model without Age and Income Category	6752.90
Model without Income Category	6773.01

Model with interaction between Income Category and Card Type	6713.88

From the results in table 4, we find the deviance is lowest in the model with interaction and highest in the model without Income Category. Since the model with interaction has the most terms, it is expected to have the lowest deviance. The model without Income Category has the highest deviance which suggests Income Category may be a good predictor for customer attrition. This also agrees with our earlier graphical analysis. The full model has the second lowest deviance which may suggest it does well in predicting customer attrition.

From the analyses above, we decided the full model would be the best for predicting customer attrition. The full model has the lowest AIC value and second lowest deviance which suggests it does well modeling customer attrition without overfitting like the interaction model. The following equation represents the full model for the logistic regression.

$$\text{logit}(p) = -3.12 + 0.0054Age + 0.76Gender_M + 0.36S_M - 0.17S_s - 0.11S_u - 0.068I_{<40} + 0.097I_{40-60} + 0.19I_{60-80} + 0.10I_{80-120} + 0.32I_u - 0.60C_g - 1.30C_p + 0.095C_s + 2.56Avg\_Utilization\_Ratio + 0.062Transaction\_Count$$

Where p is the probability of a customer not attriting, S_M is a married, S_s is single, S_u is unknown marital status, I_x represents income in bracket x, C_g is gold card type, C_p is platinum card type, and C_s is a silver card type.

Conclusion:

In conclusion, our study identified several key factors that influence credit card churn, including age, gender, marital status, credit card type, income category, utilization ratio, and transaction count. In order to study this relationship, we fit a logistic regression model to the data where we set whether a customer is attrited or not as the response variable, and those key factors as the explanatory variables. By comparing multiple models, we found that the model with all the variables we identified (excluding interactions) provided the best representation of the data. Our findings suggest that males who are married, have low utilization ratios, and an average of 30 transactions are more likely to churn their credit cards.

While our study sheds light on the factors that influence credit card churn, it has some limitations. For example, our selection of variables was based on logical thinking and could be improved by incorporating additional variables or exploring interactions between variables. Additionally, our study was limited to the data available in the dataset we used, and the results may not generalize to other populations or contexts. Future research could build on our findings by using a larger and more diverse sample, as well as exploring other factors that may influence credit card churn.

Overall, our study provides valuable insights for credit card companies seeking to retain customers and maintain a competitive edge in the market. By understanding the factors that influence credit card churn, companies can better tailor their services to meet customers' needs and reduce attrition.