PROJECT WHITE PAPER

Customer Churning Evaluation

Abstract

In today's business keeping customers for long term has become a challenge. They must find ways to prevent customers from leaving.

EDRIS SAFARI

Contents

Figures

No table of figures entries found.

Business Problem

Customer churn for any company is expensive but it is especially expensive in the customer service area. The financial and insurance sector has a large amount of competition and with new digital only institutions coming into the area the amount of competition has only grown. The biggest reason sighted for leaving a bank was "poor service" and high fees. The high fees are a profit and loss question, and each institution tries to show is value. The Poor service was the driver for 56% of the individuals that changed banks (Reducing customer churn for banks and financial Institution). Financial institutions and insurance companies struggle with customer churn as the institutions do not usually see the customer leaving before, they have closed all their accounts and left. Normally never get a chance to try and attempt to retain the customer.

It is important for service companies to understand how to predict customer churn and identify those customers before they leave. Usually, high value customers are easy to find and resolve but the middle to lower value customers need to be identified. The longer a person stays with a company the more likelihood of them being a profitable customer this increase customer lifetime value (CLTV). In our analysis we ran some linear regression models and visualized our input data to help determine what key inputs would help in determine correct traits and demographic information to use to better understand what customer good candidates for preventive attrition are.
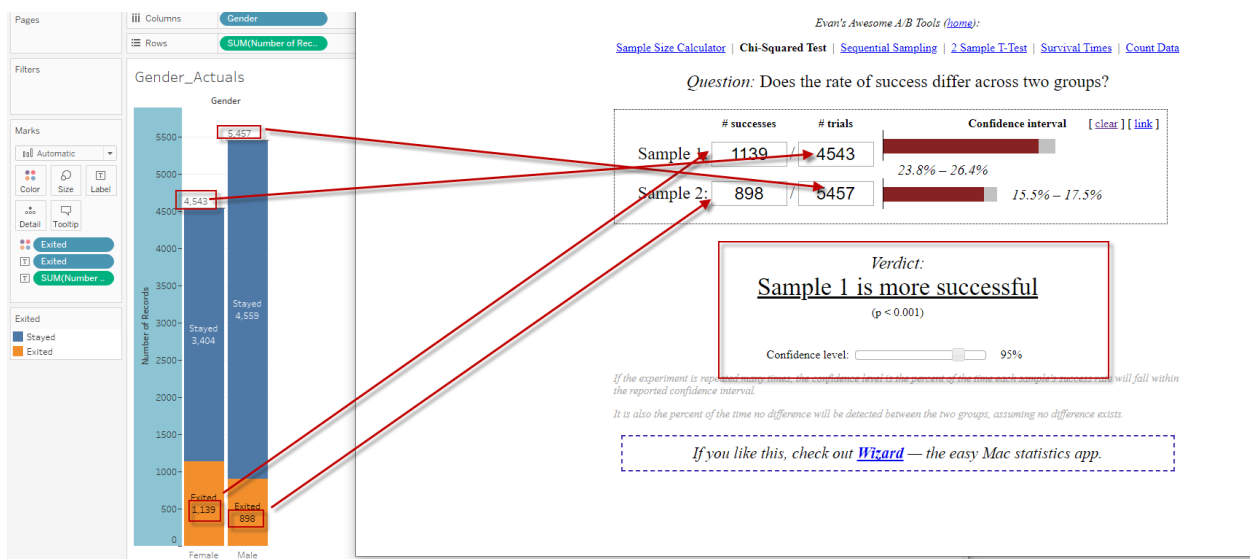
Methodology

Short version for poster

Customer churn in the financial and insurance sector is high. Companies struggle to identify customers who are likely to leave before they have left. Surveys are not frequent enough and a bad service might not show up on a survey. To increase customer lifetime value (CLTV) banks need to understand the correct behavioral attributes, build correct predictive models using new and traditional data science techniques like k-means or spatio-temporal (Behavioral attributes and financial churn prediction). This helps in selecting the correct behavioral traits based on transactions and other demographic behaviors to identify customer churn and determine if a customer is a good candidate to be retained.

Analysis

To accomplish this task, we decided to apply the regression logistics algorithm. Logistic regression algorithm is commonly used in predicting binary outcomes. In the case of churning, we want to predict if a new customer or existing customers are likely to stay or exit. We will examine a dataset composed of 10,000 records. 'exited' column in this dataset is regarded as the dependent variable which is the subject of this analysis, and the rest of the variables are the regressors or independent variables.

Tableau showed correlation with 'Has Credit Card' and ' 'IsActive' at 20% stayed/exited reference line.

After performing data mining in Tableau, we built a logistic regression model in Gretl and performed 5 backward eliminations. We made dummy variables 'Spain', Germany' and 'France' from the 'Geography' variable and 'Male' and 'Female' variables from 'Gender' variable. We included 'Female', 'Spain' and 'Germany' in the model along with the other independent variables in the dataset. In each run of the model, Gretl recommended removal of a variable. Table below shows the summary of each elimination. The main criteria for keeping a variable in the model was that the p-value to be below out threshold of 0.5 for the variables and the Adjusted R-Squared increasing for each model.

BWElimination Number	Variable eliminated	Variable P-Value	Model's Adjusted R-squared before/after removal	Adjusted R-Squared Difference
1	Spain	0.6181	0.150787/ 0.150961	0.000174

2	HasCrCard	0.4489	0.150961/ 0.151102	0.000141
3	EstimatedSalary	0. 3091	0.151102/ 0.151197	0.000095
4	Tenure	0.0873	0.151197/ 0.151106	-0.000091

As shown in elimination 4 'Tenure' was removed, but not by recommendation from Gretl, but because we wanted to see the impact of removal to test the p-value threshold. It shows that the Adjusted R-Squared was not impacted by much, so we reincluded 'Tenure' in the model. After transforming the 'Balance' variable to Log10(Balance +1) for better uniformity, we got the result shown below.
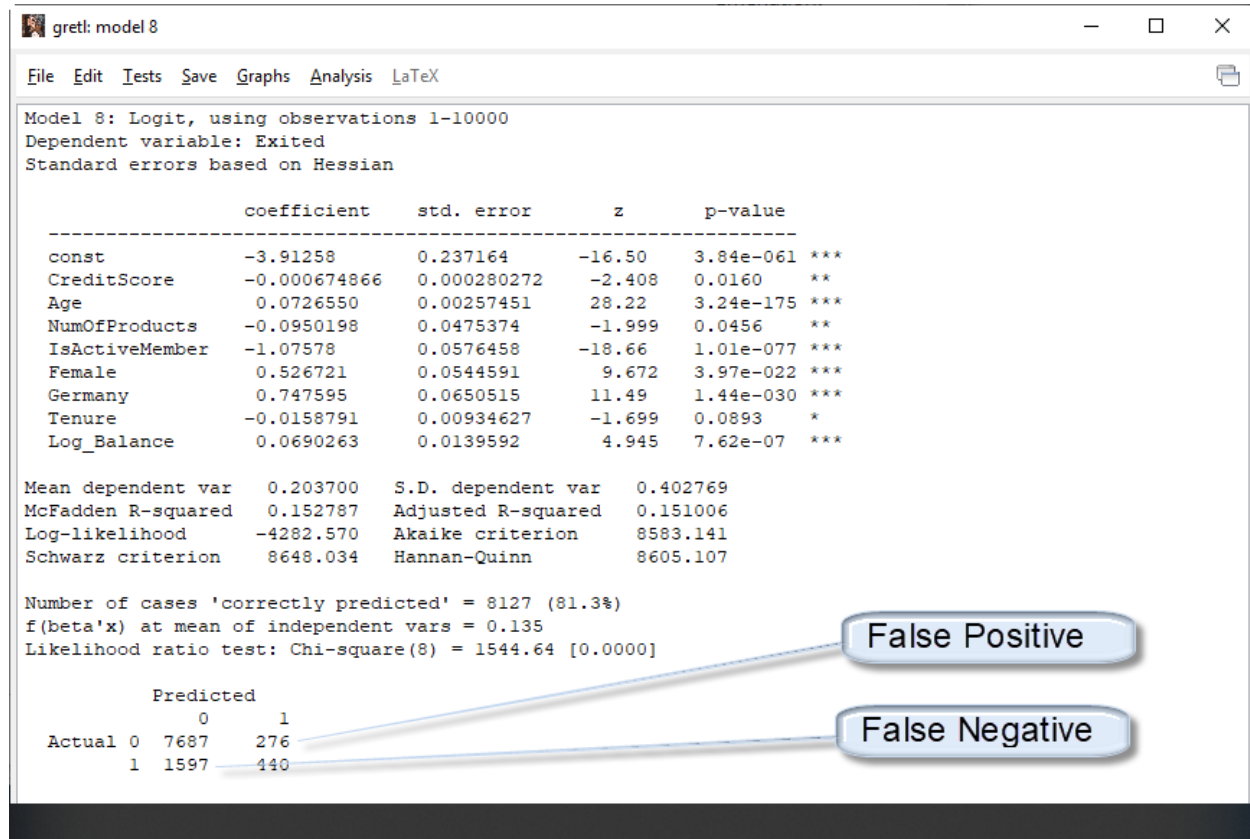
From the confusion Matrix, the accuracy and error rates we calculated as shown below.

Accuracy Rate = Correct/Total = (7687+440)/10000 = 81.27 %

Error Rate = Wrong/Total = (276+1597)/10000 = 18.73 %

Based on the current analysis, we deem the accuracy rate to be not satisfactory. We look to maximize this rate by introducing new variables from our data mine.

References

1. Diana Kaemingk (2018) "Reducing customer churn for banks and financial Institution https://www.qualtrics.com/blog/customer-churn-banking/
2. Plug Those Leaks: Stop Attrition From Stalling Your Growth Strategy https://thefinancialbrand.com/68371/banking-customer-acquisition-attrition-growth-strategy/
3. Kaya E, Dong X, Sutiara Y, Balcisoy S, Bozkaya B & Pentland A, Behavioral attributes and financial churn prediction, EPJ Data Science, Vol 7, Iss 1, Pp 1-18 (2018)
4. Jennings A, (2015), The 4 D's of Customer Attrition https://thefinancialbrand.com/55772/banking-customer-attrition-analysis/
5. Farquhar, JillianDawes (2005), Retaining customers in UK financial services: The retailers' tale., Service Industries Journal. Dec2005, Vol. 25 Issue 8, p1029-1044. 16p. 3 Charts