# **Bank Customer Churn**

# MSBA 5324: Marketing Analytics

# RamaKrishna Velamala

# Devaasheesh Pavuluru

# Pravallika Vasantham

# Smit Patel

**Table of contents**

1. Executive Summary…………………………………………………….3
2. Introduction……………………………………………………………..4
3. Literature review………………………………………………………..6
4. Research design…………………………………………………………7
5. Data dictionary………………………………………………………….8
6. Data analysis…………………………………………………………….9
7. Research findings……………………………………………………….10
8. Interpretation……………………………………………………………11
9. Market implementation ………………………………………………...15
10. Conclusion………………………………………………………………16
11. Reference………………………………………………………………..17
12. Appendix………………………………………………………………...18

**Executive summary**

The goal of this research is to investigate customer attrition in a financial setting by applying survival analysis, especially the Cox proportional hazard model. The study looks at a range of customer characteristics, such as age, balance, credit score etc., in an effort to pinpoint important variables affecting client attrition. The alternative hypothesis asserts that at least one attribute influences churn hazard, whereas the null hypothesis predicts no substantial impact. The study uses survival analysis to calculate hazard ratios and evaluate statistical significance using a Kaggle dataset. Results show that, in line with assumptions, turnover is highly influenced by credit score, age, active membership, country and gender. Some presumptions, however, are incorrect about the balance, product number, credit card ownership & expected wage. Notably, the risk of churn is greatly decreased by being an active member. According to the report, individual consumer categories should be targeted with different marketing techniques. Some examples of these strategies include concentrating on creditworthy customers, developing age-specific loyalty programs, and adjusting retention campaigns according to gender and nationality. Marketing implementations include personalized communication, targeted retention campaigns, and continuous analysis of customer behavior. Recommendations emphasize the importance of adapting retention tactics to evolving market conditions and customer trends. In conclusion, the Cox model provides valuable insights for enhancing customer retention strategies in the financial sector, emphasizing the significance of credit-related factors and active customer engagement.

**Introduction**

**Importance of the Topic :**

Survival analysis is a powerful tool in many industries, including banking, healthcare and customer retention, because it is essential to comprehending the factors that impact the amount of time until an event occurs. Our goal in this project is to study customer churn in a financial setting using the Cox proportional hazards model.

**Research Goal :**

In our project the main objective is to find important indicators of customer attrition and evaluate how they affect customers' chances of surviving. By doing this, we hope to provide practical advice for enhancing client retention tactics.

**Theoretical background :**

A popular survival analysis method is the Cox proportional hazards model, which lets us analyze the connection between variables and the hazard rate and reveal the variables affecting an event's timing. In this instance, customer churn is the event of interest, and we will investigate how different customer attributes affect the likelihood of churn over time.

**Story line :**

Credit\_Score: A customer's credit score is represented numerically by their credit score, which also includes information about their risk tolerance and level of financial responsibility. A borrower with a better credit score is thought to be more dependable and is less likely to churn. Our hypothesis is that because they are generally more responsible and stable with their finances, clients with better credit scores are less likely to leave.

Country : The country of the customer adds a geographical and cultural element that could have a big impact on their financial choices and actions. The varying economic situations, regulatory frameworks, and cultural values throughout nations can have an impact on client loyalty and attrition trends. Our goal is to investigate the ways in which these nation-specific characteristics influence the financial institution's churn risk.

Gender : Gender influences how people behave financially and react to marketing tactics. Gender differences in spending patterns, risk tolerance, and financial decision-making are frequently influenced by societal and cultural conventions. We hypothesize that variations in customer churn are influenced by gender disparities, which may shed light on the efficacy of retention measures designed for particular gender segments.

Age : Different life stages, financial demands and risk tolerances are reflected in the age of the customer. While older clients can be looking for stability and long-term connections, younger customers might be more impulsive and experiment with different financial products. We hypothesize that age influences the risk of turnover, identifying distinct traits and actions connected to various age groups.

Tenure : The duration of a customer's relationship with a financial institution is measured by their customer tenure. A longer tenure could be interpreted as an indication of customer satisfaction, loyalty and faith in the organization's services. Newer customers, however, can be more likely to churn as they explore the many goods and services that are offered. Our hypothesis is that longer tenure leads to a decreased risk of turnover, indicating a solid and established relationship with the organization.

Balance : The balance or total amount of money in an account shows a customer's financial situation. A greater balance might be interpreted as a sign of good financial standing and a deep dedication to the financial organization. Since clients with larger balances show a larger commitment to the bank, we hypothesize that they are less likely to churn.

Products number : A customer's number of products indicates how involved they are with the financial organization. More products could be a sign of loyalty and a more varied partnership. Because they are more deeply integrated with the institution's offerings, we predict that customers who have more goods will be less likely to churn.

Credit card : Having a credit card is a crucial financial practice that shows a customer's faith in and dependence on the company for credit services. We speculate that since credit card ownership denotes a higher level of loyalty to the company, credit card holders are less likely to leave.

Active Member : The degree to which a consumer is actively involved with the financial institution is indicated by their active membership status. A regular user of the institution's services and frequent interactions with it are indicative of an active member. We predict that compared to less engaged customers, active members will be less likely to churn.

Estimated Salary : The customer's financial capability and level of participation are influenced by their estimated salary, which offers insights into their income. Greater financial stability and loyalty might be linked to higher estimated salaries. Based on our hypothesis, clients that have higher expected salaries are more likely to stay with the company longer since they can afford to do so.

The following hypothesis is prepared using this theoretical background.

Null Hypothesis (H0): The combination of customer characteristics including credit score, country, gender, age, tenure, balance, Product number, credit card, active member and estimated salary does not significantly influence the hazard of churn.

Alternative Hypothesis (H1): At least one of the customer characteristics significantly affects the hazard of churn.

**Literature Review :**

Over the past decade, customer rate analysis using machine learning has gained importance in several industries, including telecommunications, e-commerce, banking, insurance, retail, energy, gaming, and medicine. This study aims to analyze the growth rate of customers in the banking sector based on general results from the literature. Key factors affecting customer satisfaction and loyalty in banking include marketing factors such as trust, communication and relationships. Important factors that determine customer satisfaction include accuracy of information, responsiveness of staff, availability of services, reliability, security, personalization and consistency. Both relationship marketing dimensions and objective factors have a significant impact on customer loyalty in the banking sector. The study highlights the importance of predictive models that use demographic, psychographic and transactional data to predict customer spending through machine learning techniques. Previous research has investigated factors such as age, income, gender, credit card status and bank payment discounts offered using various models such as logistic regression, decision trees and neural networks to explore factors like age, income, gender, credit card status, and bank-offered discounts on customer churn in the banking sector. These are some of the methods used by other people to analyze the dataset.

**Research Design :**

The dataset for this research is obtained from Kaggle. The dataset includes information related to bank customer churn, and it can be accessed using the following Kaggle link:<https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset>.

Our dataset contains customer information from a certain point in time, including credit score, nationality, gender, age, tenure, balance, product number, credit card ownership, active membership status, and predicted salary unlike previous research which investigated factors such as age, income, gender, credit card status and bank payment discounts. The major goal of the research is to investigate customer churn in a financial setting using the Cox proportional hazards model.

Credit score, nationality, gender, age, tenure, balance, product number, credit card ownership, active member status, and expected salary are all independent variables. Churn among customers is the dependent variable.

For survival analysis, the Cox proportional hazards model is used as the primary methodology. This model allows for the investigation of the link between independent variables and the hazard rate, revealing information about the factors influencing the timing of customer churn.

**Data Dictionary :**

| **Variable** | **Description** | **Measurement Level** |
| --- | --- | --- |
| customer\_id | Unused | Unused |
| credit\_score | The customer's credit score. | Numerical |
| country | The country where the customer is located. | Categorical |
| gender | The gender of the customer. | Categorical |
| age | The age of the customer. | Numerical |
| tenure | The number of years the customer has been with the bank. | Numerical |
| balance | The customer's account balance. | Numerical |
| products\_number | The number of products the customer has with the bank. | Categorical |
| credit\_card | Whether the customer has a credit card or not. | Binary |
| active\_member | Whether the customer is an active member or not. | Categorical |
| estimated\_salary | The estimated salary of the customer. | Numerical |
| Churn | whether the customer has left the bank (1) or not (0). | Binary |

**Data Analysis**

**Summary statistics :**

A screenshot of a table

Description automatically generated

**Bar chart :**

A graph with numbers and a rectangular column

Description automatically generated with medium confidence

**Research Findings:**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A table with numbers and letters

Description automatically generated

**Interpretation :**

If the hazard ratio is less than 1 it indicates a decrease in hazard of churn.

If the hazard ratio is greater than 1 it indicates an increase in hazard of churn.

If hazard ratio is equal to 1 it indicates no difference in hazard of churn.

Credit Score : From the above output we can see that a one unit increase in credit score is associated with a 0.1% decrease in the hazard of churn. In other words, as credit score increases the likelihood of churn decreases.

We can also see that Credit Score is statistically significant since its chisq value is 0.0255 which is less than 0.05.

From the analysis we can see that the output aligns with our storyline i.e., as clients with better credit scores are less likely to leave.

Age: From the above output we can see that a one year increase in age is associated with a 4.9% increase in the hazard of churn. In other words, as age increases the likelihood of churn increases.

We can also see that Age is statistically significant since its chisq value is <0.0001 which is less than 0.05.

From the analysis we can see that the output aligns with our storyline i.e., as age is an important factor influencing the customers leaving the bank.

Balance: From the output we can see that Balance has a negligible impact on the hazard of churn as its hazard ratio is 1 which tell us that there is no difference in hazard. In other words, as balance increases the likelihood of churn is very negligible i.e., the value is very small.

We can also see that balance is statistically significant since its chisq value is <0.0001 which is less than 0.05.

From the analysis we can see that the output does not align with our storyline i.e., although we thought customers with high balance are less likely to churn but the value is very negligible.

Products Number: From the output we can see that one unit increase in the number of products is associated with a 6.4% decrease in the hazard of churn. In other words, as product number increases the likelihood of churn decreases.

But we can see that the product number is not statistically significant since its chisq value is 0.0852 which is greater than 0.05.

From the analysis we can see that the output does not align with our storyline i.e., although we thought customers with high product number are less likely to churn but product number is not statistically significant.

Credit Card :From the output we can see that credit card is associated with a 6.1% decrease in the hazard of churn. In other words, as credit card ownership increases the likelihood of churn decreases.

But we can see that credit card is not statistically significant since its chisq value is 0.19 which is greater than 0.05.

From the analysis we can see that the output does not align with our storyline i.e., although we thought customers with high credit card ownership are less likely to churn data, credit card is not statistically significant.

Active Membership : From the above output we can see that being an active member is associated with a significant 52.2% decrease in the hazard of churn. In other words, as active membership increases the likelihood of churn decreases.

We can also see that Active membership is statistically significant since its chisq value is <0.0001 which is less than 0.05.

From the analysis we can see that the output aligns with our storyline i.e., as we predicted that compared to less engaged customers, active members will be less likely to churn.

Estimated Salary: From the above output we can see that Estimated salary has a negligible impact on the hazard of churn. as its hazard ratio is 1 which tell us that there is no difference in hazard. In other words, as estimated salary increases the likelihood of churn is very negligible i.e., the value is very small.

We can also see that estimated salary is not statistically significant since its chisq value is 0.9920 which is greater than 0.05.

From the analysis we can see that the output does not align with our storyline i.e., although we thought customers with high salary are less likely to churn, the estimated salary is not statistically significant.

Country : From the above output we can see that France has a 4.8% decrease, and Germany has a significant 55.1% increase in the hazard of churn compared to Spain. In other words, France has less churn rate compared to Spain and Germany has more churn rate compared to Spain.

We can also see that the country France is not statistically significant since its chisq value is 0.4123 which is greater than 0.05.

We can also see that Germany is statistically significant since its chisq value is <0.0001 which is less than 0.05.

From the analysis we can see that the output aligns with our storyline i.e., although we thought that these nations' specific characteristics influence the financial institution's churn risk from the analysis we can see that France is not statistically significant and Germany has significantly more churn rate than spain.

Gender : From the above output we can see that Female customers have a 47.2% higher hazard of churn compared to males. In other words, female customers are more likely to leave the bank compared to male.

We can also see that gender is statistically significant since its chisq value is <0.0001 which is less than 0.05.

From the analysis we can see that the output aligns with our storyline i.eAlthough we thought gender has an influence on churn, from the analysis we can clearly see that female customers are more likely to leave the bank compared to male.

**Marketing Implementation:**

Credit Score: Start focusing advertisements for clients with excellent credit ratings. To encourage them to stay, provide them with benefits like loyalty programs, individualized services, and special privileges.

Age: Design rewards programs or loyalty schemes that are age specific. Engage senior citizens with individualized product offers, retirement perks, and financial planning services.

Balance : Although balance by itself might not have a significant impact on churn, we can use it in combination with other elements. To increase loyalty, provide premium services or incentives to clients with large balances.

Active Membership: Create plans to maintain active and involved consumers. To keep people interacting, implement gamification, loyalty reward schemes, or customized materials.

Country : Customize retention tactics for every nation. Address certain issues or preferences in Germany.

Gender : Recognize and target to the particular requirements and tastes of female clients. Make sure that product offerings, incentives, and advertisements appeal to the unique needs and interests of  female target market.

**Recommendations to client :**

Personal communication : Using individual client profiles, interests, and behaviors as a basis, implementing specific communication methods.

Retention Campaigns: Create focused retention strategies for the particular client categories that the investigation revealed.

Feedback Systems: Set up feedback systems to continuously learn about the wants and needs of your customers. Utilize client input to enhance retention tactics.

Analysis in Progress: Keep an eye on and evaluate consumer behavior at all times. Retention tactics should be reviewed and modified on a regular basis in light of changing market conditions and consumer trends.

You can develop a more complex and successful client retention strategy by putting these improved tactics into practice and targeting particular elements found in the Cox proportional hazards model.

**Conclusion :**

A closer look at the Cox proportional hazards model uncovers interesting trends that affect customer attrition. Significantly, and in line with expectations, credit score and active membership have an impact on employee retention rates. Important techniques for successful customer retention include increasing credit card ownership, addressing the unique preferences of female clients, customizing marketing efforts to satisfy age specific demands and increasing the advertisement in the specific country which is Germany where people are more interested in maintaining the relation with the bank.

**Reference:**

Guliyev, Hasraddin & Yerdelen, Ferda. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning models. Journal of Applied Microeconometrics. 1. 85-99. 10.53753/jame.1.2.03.

Topre, G. (2022, August 30). *Bank Customer Churn Dataset*. Kaggle. https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset

**Appendix**

**Sas code**

%web\_drop\_table(MARKET.project);

FILENAME REFFILE '/home/u63095890/marketing/Bank Customer Churn Prediction.csv';

PROC IMPORT DATAFILE=REFFILE

DBMS=CSV

OUT=MARKET.project;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=MARKET.project; RUN;

%web\_open\_table(MARKET.project);

----------------------------------

/\*Sorting the data by time variable\*/

proc sort data=market.project;

by tenure;

run;

-----------------------------------

/\*tranformation of dataset\*/

data market.survival\_data;

set market.project;

time = tenure;

status = churn;

/\* Convert categorical variables to indicator variables \*/

if country = 'France' then country\_France = 1; else country\_France = 0;

if country = 'Spain' then country\_Spain = 1; else country\_Spain = 0;

if country = 'Germany' then country\_Germany = 1; else country\_Germany = 0;

if gender = 'Male' then gender\_Male = 1; else gender\_Male = 0;

/\* Add other categorical variables as needed \*/

run;

-----------------------------------------------

/\*cox hazard model\*/

proc phreg data= market.survival\_data;

class country gender;

model time\*status(0) = credit\_score age balance products\_number credit\_card active\_member estimated\_salary

country gender / ties=efron;

run;