I Cloud 7 Technical Assessment

*Data Scientist Jr.*

*Presented by: Jorge Forero*

Executive Summary

Objective:

The objective of this analysis is to provide the bank with actionable insights into customer churn by identifying the key drivers that influence customer attrition. By leveraging machine learning models, the goal is to predict churn with high accuracy, enabling the bank to take proactive steps to retain at-risk customers. This report will present a data-driven strategy to improve customer retention, increase lifetime value, and reduce revenue loss.

Key Findings:

Recommendations:

* Im

1. Introduction

This report presents a comprehensive analysis of customer churn using advanced data science techniques, aimed at empowering the bank to make data-driven decisions to improve customer retention. By performing an in-depth Exploratory Data Analysis (EDA) and applying machine learning models, this report seeks to identify the key factors influencing customer attrition and develop predictive models that can accurately forecast churn risk.

* 1. Scope of Analysis

This analysis examines a dataset of 10,000 customer records from France, Germany, and Spain, indicating whether each customer is active or has churned. The focus is on exploring the data to identify churn patterns and building predictive models to classify churn. The goal is to uncover key churn drivers and provide actionable insights to help the bank reduce customer attrition.

* 1. Ethical Considerations

Several ethical considerations were taken into account during the analysis:

* Data Privacy: Since the dataset used include personally identifiable information (PII), all customer-related data was handled responsibly to ensure privacy and confidentiality. Measures were taken to anonymize any sensitive data.
* Transparency and Reproducibility: All steps of the analysis, from data preprocessing to model development, were documented clearly to ensure transparency and reproducibility. This enables other analysts to replicate the analysis and verify the results.
* Bias in Predictions: Predictive models can sometimes introduce or reinforce bias, particularly if the data contains underlying biases (e.g., Gender disparities). We made efforts to ensure the fairness of the models by carefully analyzing feature importance and ensuring balanced evaluation metrics.

1. Methodology
   1. Data Collection and Processing

Data for this analysis was provided by the client, which includes the following key fields:

CustomerId: A unique identifier for each customer

Surname: The customer's last name

CreditScore: A numerical value representing the customer's credit score

Geography: The country where the customer resides (France, Spain or Germany)

Gender: The customer's gender (Male or Female)

Age: The customer's age

Tenure: The number of years the customer has been with the bank

Balance: The customer's account balance

NumOfProducts: The number of bank products the customer uses (e.g., savings account, credit card)

HasCrCard: Whether the customer has a credit card (1 = yes, 0 = no)

IsActiveMember: Whether the customer is an active member (1 = yes, 0 = no)

EstimatedSalary: The estimated salary of the customer

Exited: Whether the customer has churned (1 = yes, 0 = no)

1. Key Findings
   1. France Branch

For this branch we could find the following information after making the ETL and themodel training

Data Analysis and Findings

* The credit score distribution is approximately normal, with most customers having scores between 600 and 700. The age distribution is skewed towards younger customers, primarily between 30 and 50, indicating a solid base of financially responsible young clients. (See Chart 1)
* The balance and estimated salary data suggest that many customers either don’t save significantly with the bank, or there may be issues with how balance is measured. It might be worth considering using an average balance over a year for better accuracy. (See Chart 1)
* There is a positive correlation between age, balance, and churn, suggesting that older customers or those with higher balances are more likely to churn. (See Chart 2)
* Customers with more products are slightly less likely to churn, which is a key correlation that will be further explored with our ML models. (See Chart 2)
* Active membership appears to play a critical role, with active customers showing lower churn risk. Similarly, customers holding only one product are more prone to churn, which follows logical expectations. (See Chart 3)

Visualizations and Data

* Chart 1: Distribution of Numerical FeaturesA graph of different types of numbers

  Description automatically generated with medium confidence
* Chart 2: Correlation Matrix France

A screenshot of a graph

Description automatically generated

* Chart 3: Churn by Categorical Variables

A screenshot of a graph

Description automatically generated

ML Processing and Model Insights

We selected and trained a ML model

* There

Visualizations and Data

* 1. Spain Branch

Using the

Analysis and Findings

●

Visualizations and Data

Visualizations and Data

* Chart 1: Distribution of Numerical Features

A group of blue and white bars

Description automatically generated

* Chart 2: Correlation Matrix France
* Chart 3: Churn by Categorical Variables

Insights

● There

Overall Insights

● To effectively

* 1. Germany Branch

“What is

Analysis and Findings

● W

Visualizations and Data

● Table 1: SQL Query Table Inflow and AHT Changes

● Table 2: SQL Query Table Issue Type Weekly

Insights

● There

Overall Insights

● To effectively

1. Recommendations
   1. Improvement Areas

* Implement strategies to manage ticket inflow more effectively while optimizing the efficiency of handling specific issue types. Based on the data, we recommend increasing self-service options. Using FAQs, knowledge bases, or interactive guides is typically suitable, but enhancing chat services with Generative AI can empower customers to find solutions independently.
* Prioritize resolving high-impact issues such as cancellations, incomplete orders, and app-related problems to reduce overall AHT and improve operational efficiency. Providing targeted training for agents or establishing a specialized team for these issues can equip agents with the skills and knowledge to handle them efficiently. Empowering agents to make decisions and take ownership of resolving critical issues can improve first-contact resolution rates.
* Implement enhanced diagnostic tools and support frameworks to aid agents in resolving issues more efficiently. For example, Clinical Decision Support (CDS) tools can assist in information gathering, facilitate cognitive processes by organizing data, and help generate differential diagnoses for each issue type.
  1. Strategic Initiatives
* Utilize new data related to the financial costs of ticket resolution and sentiment analysis to understand their impact on the business and to help find the optimal AHT value.
* Implement initiatives to improve first-time resolution rates, especially for issues that frequently require reopening. Start by motivating agents through recognition and rewards for consistently high first-time resolution rates. This can be done through incentives, awards, or bonuses.
* Design and implement training programs that focus on equipping agents with the skills needed to handle complex and high-frequency issues effectively. Begin with problem-solving and critical thinking training to help agents analyze issues, identify root causes and risks, and determine the best course of action.

1. Next Steps

* Action Items:

Immediate: Initiate the implementation of self-service options and enhance chat services with generative AI. Develop targeted training programs for agents

Within 3 Months: Implement enhanced diagnostic tools and support frameworks. Launch recognition and rewards initiatives for high first-time resolution rates.

* Future Analysis

Suggest areas for future analysis or additional data collection to have advanced analytics over the AHT and the CSAT. With more advanced techniques it is expected to find optimization options with new trends and factors impacting the customer satisfaction and operational efficiency.