Churn Analysis of XYZ Bank Clients

Intro

The following represents a customer churn analysis of a bank regarding customers leaving its credit card services. The bank is concerned that more and more customers are leaving and wants to understand the main reasons and recommendations to avoid this. The dataset includes 10127 data points across 21 features. These features include information about the customers (e.g. age and gender), their relationship with the bank (e.g. attrited/ existing, last 12 months contacts with the bank), and use of bank services (e.g. nr. of products held, total transaction amount, card category). Of the dataset, 84.3% are existing and 15.7% are attrited customers.

We first performed a Summary Statistics analysis, which included calculating basic statistics and detecting outliers. This helped us understand the distribution of the data and identify any issues that could affect the results. Next, we cleaned the data which involved examining (e.g. for errors, inconsistencies, and completeness), cleaning, and transforming the data. Overall, the dataset was of good quality which required limited data wrangling. As a next step, we analyzed the distribution of each feature in the dataset to gain further insights. Finally, we conducted a Cross-Correlation analysis to understand the relationships between different features and identify any trends or patterns.

Conclusions

The following shows the Top 5 conclusions we gathered from the analysis:

1. **Avg. utilization ratio – customers with lower ratios have higher churn rates**: Customers with lower avg. utilization ratios are more likely to churn (attrited 0.16% vs. existing 0.30%). This is consistent across all dimensions of the features gender, education level, marital status, age groups, income category, and card category. Comparing the ratio against credit limit shows a negative regression meaning customers with a lower credit limit are using the credit card less.
2. **Contacts count last 12 months – customers who were contacted less have lower churn rates**: Values for contact counts range from 0-6 with a mean being 2.5 and a median of 2. Attrited customers were on average more contacted than existing customers (2.97 vs. 2.36). This is consistent across all dimensions of the features gender, education level, marital status, age groups, income category, and card category. In addition, as the number of contacts with customers increases, the percentage of customers leaving the services increases (Contacts:1x 4% vs. 6x 100%). Starting from contacts of 3 and higher we have a higher churn rate than the overall total churn rate. Unclear is if customers were contacted more often before they planned to churn or after.
3. **Total relationship count – customers having more bank products have lower churn rates**: Existing customers use on average 2.91 vs. attrited customers 3.28 products. Again, this is consistent across all dimensions of the features mentioned before (with the exception being the Gold and Platinum card). Churn rates fall substantially when customers use 3 or more products and customers with only 1 or 2 products represent the smallest group of customers.
4. **Card category – higher card categories have higher churn rates**: The cards with the highest status have the highest churn rate (Platinum: 25%, Gold: 18%). They also represent the cards with the highest average transaction amounts and counts. Yet, it is important to note that Gold cards represent only 1.1% and Platinum only 0.2% of all customers, resulting in the statistical significance being questionable.
5. **Education level – highest education levels have the highest churn rates**: Customers with the highest education levels, Doctorate and Post-Graduate, have the highest churn rate with 21.06% and 17.83%. They present 4.5% and 5.1% of the data respectively. A deeper analysis across demographics gives no clear additional insight.

Recommendations

It is important to note that the conclusions are based solely on indications given by the data and require further analysis (correlation does not mean causation). Based on the Top 5 conclusions we have the following recommendations:

1. **Avg. utilization ratio:** As customers with a lower average utilization ratio are more likely to churn, initiatives (e.g. marketing, promotions) to nudge more usage could be considered and analyzed. Regarding the negative correlation with the credit limits, it could be considered increasing limits for clients that are considering churn (however as a higher utilization ratio and credit limit increase default risk and as no such information was given, any further analysis should include this).

2. **Contacts count last 12 months:** Deeper analysis using the means (e.g. telephone, email, personal, etc.) and reasons for contact to get additional insights.
3. **Total relationship count:** Further analysis regarding the type of products existing vs. attrited customers have and potentially increase the marketing of these products or combine them with the credit card service.
4. **Card category:** Additional analysis using info on the different card categories (e.g. (un)used features, fees, interest rates, loyalty programs, customer acquisition costs) to further analyze the churn across card categories.
5. **Education level:** Additional analysis using acquisition costs and methods vs. potential profit & features used across the various education levels for more defined insights.

Finally, an analysis of the main reasons for churning, external factors such as competitors (e.g. new entrants, new offerings), and the overall macroeconomic environment (e.g. interest rates, unemployment) potentially impacting the churn rate would enrich the analysis.

Appendix

Link to the Tableau dashboard