Bank Customer Churn Rate Analysis and Prediction

Lazarus Mapfurira

Introduction:

Customer churn is a major problem faced by banks across the globe, with significant impacts on revenue and growth. This analysis focuses on examining customer churn for a major retail bank that operates in Germany, France, and Spain. The objective is to identify the key drivers of churn in order to develop effective customer retention strategies.

The dataset used for this analysis includes demographic and banking indicators for a sample of 10 000 individuals across the three countries. The demographic variables capture age and gender. The banking variables provide information on channel of acquisition, credit score, account balance, number of banking products used, and tenure with the bank. Additionally, estimated salary for each customer is also provided based on algorithms used internally by the bank.

The goal is to thoroughly analyze this dataset to uncover patterns and trends that distinguish between loyal and churning customers. These insights will be used to build a predictive model that can identify high-risk customers proactively so that targeted interventions can be made to improve retention.

Methodology:

The analysis began with thorough exploratory data analysis to understand the distribution of key variables and visualize differences between churn and non-churn customers. Univariate analysis was conducted by plotting histograms and density plots. For multivariate analysis, box plots, scatter plots, and violin plots were created to study relationships between variables.

Feature engineering was then conducted to derive new attributes that could better explain churn. The tenure in years was converted to tenure buckets to capture non-linear effects.

The data was split 80:20 into train and test sets respectively. Five machine learning algorithms – logistic regression, decision tree, random forest, SVM, and KNN were trained on the dataset. Model tuning was done using randomized search and 5-fold stratified cross-validation. The top models were evaluated on the test set using AUC-ROC and precision-recall curves. The decision tree classifier was chosen as the best model for its performance.

Results:

In this exploratory data analysis, I investigated a dataset from a bank with operations in **Germany, France, and Spain**. The dataset provided insight into the bank's clients, including their tenure, age, balance, credit score, and whether or not they have churned.

The customer age distribution was similar across the three countries – majority were young and middle-aged professionals between 30-45 years.

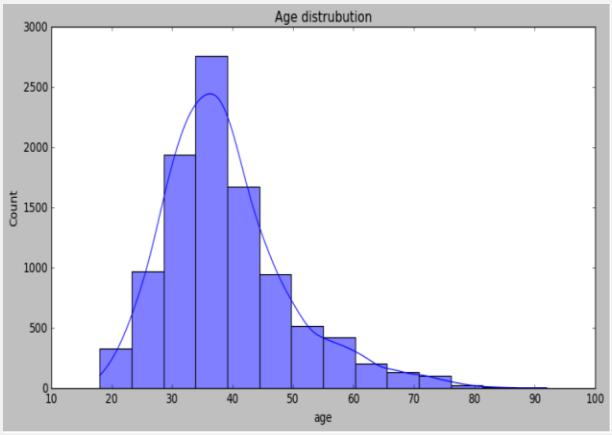


Figure 1: Age Distribution

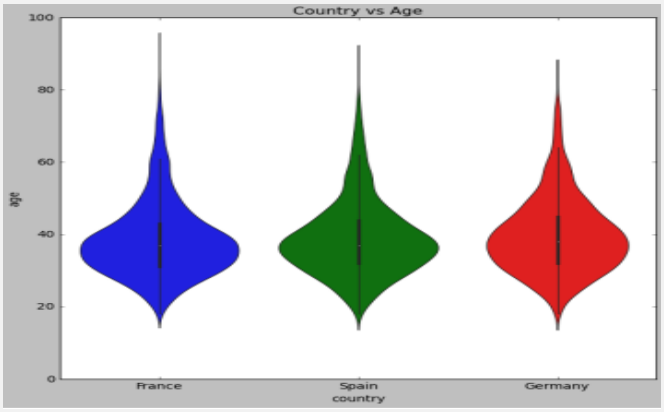


Figure 2: Violin plot of age distribution by country

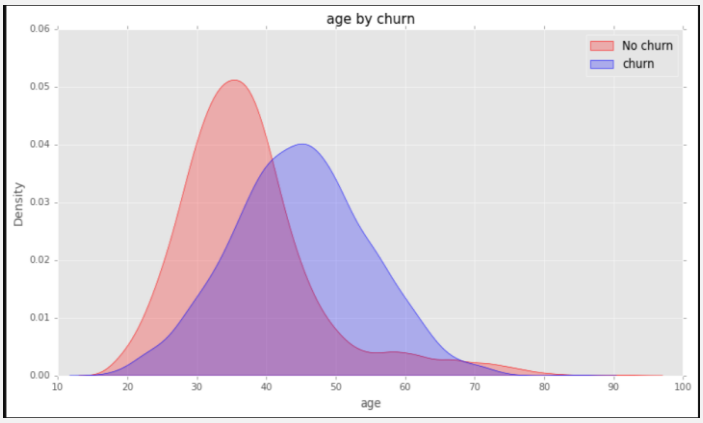
Churn is more likely to occur for clients who are in their 30s and 40s. This may be because these clients are starting to have families and other financial responsibilities, which may make them more likely to switch banks.

Churn is less likely to occur for clients who have high credit scores. This may be because these clients are more likely to be seen as good risks by the bank, or because they are more likely to be satisfied with their financial situation.

Exploratory data analysis revealed that majority of customers are young professionals in their 30s and 40s and have a credit score between 600 and 700. However, many clients, irrespective of their country or gender, have zero balance in their accounts, which I suspected might be due to low-interest rates offered by the bank.

As age increased the possibility for churn also increased.

Figure 3: Age



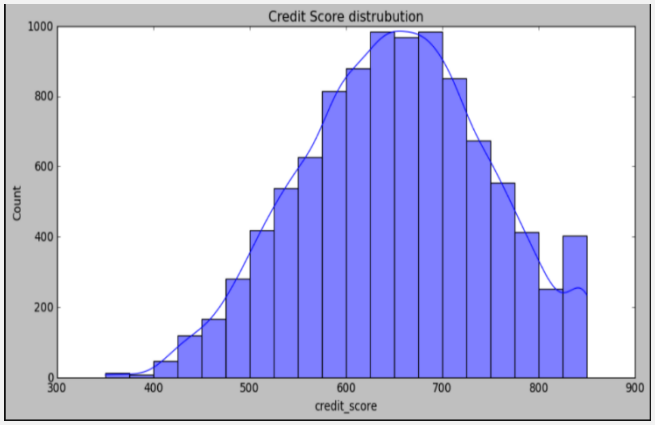


Figure 4: Credit Score

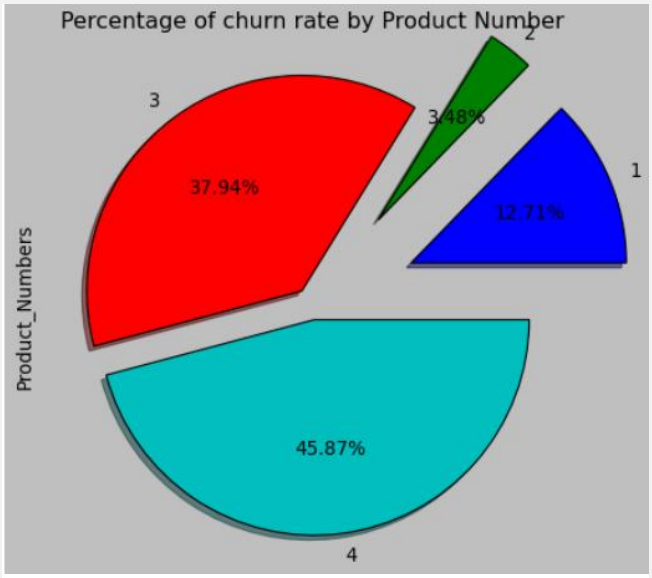


Figure 5: Product churn rate

I also observed that Germany has the highest churn rate among the three countries, even though most of the bank's financial products are offered there. The bank's product with the highest churn rate is Product 4, followed by Product 3.

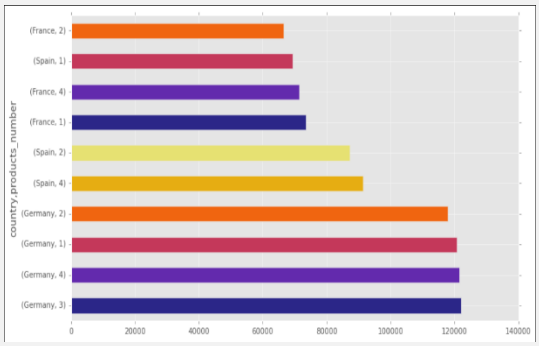
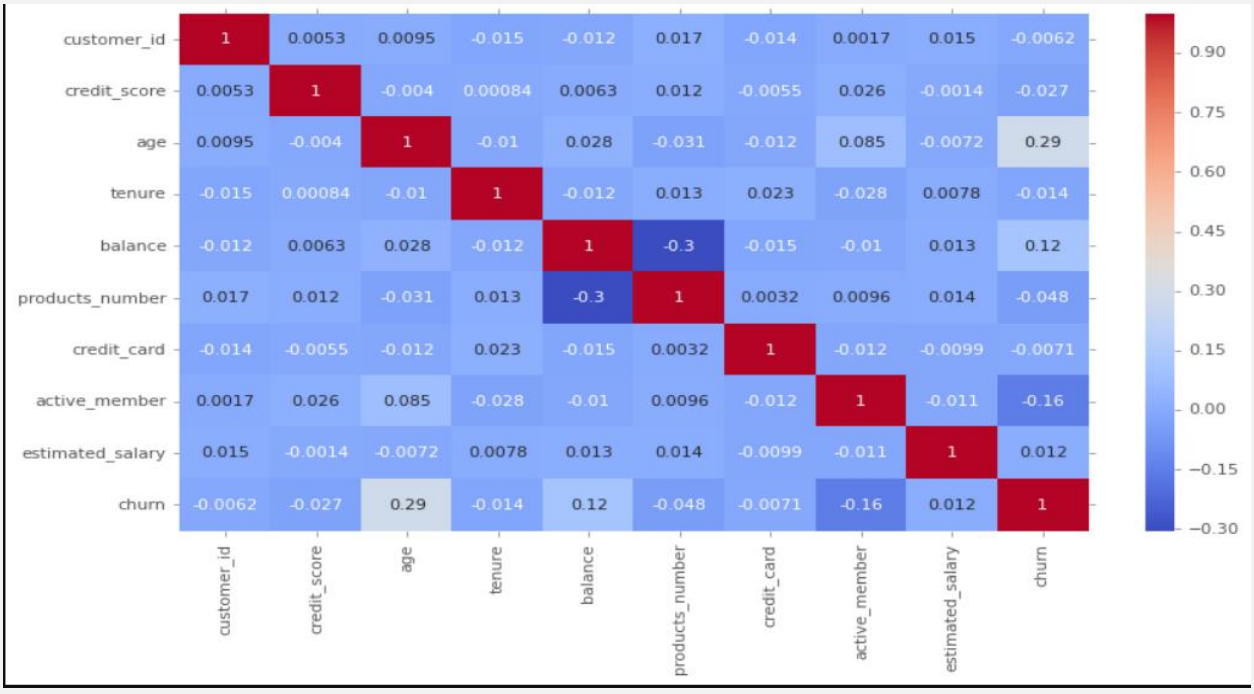


Figure 6: Country

The data also showed a correlation between client's age, balance, and estimated salary with churn rate. Interestingly, churn was negatively related to credit card usage, suggesting that clients using the bank's credit card are less likely to leave.

Figure 5: Correlation



Overall, this analysis provides a foundation for further investigation into why clients are leaving the bank and what strategies the bank could use to improve client retention. I have provided a set of 15 recommendations based on these findings for the bank to consider.

Recommendations:

1. **Focus on Client Retention:**
   Since the majority of clients are in the age group of 30-45, it's important to develop strategies targeting this group, as they seem to be the bank's primary customer base.
2. **Senior Clients:**
   Although the number of senior clients (ages 55+) is minimal, they might be more stable. The bank should consider developing special packages or products to attract more senior clients.
3. **Address Zero Balance:**
   The fact that many clients have zero balance is concerning. The bank should consider incentives to encourage clients to maintain a minimum balance.
4. **Competitive Interest Rates:**
   If the zero balance issue is due to low-interest rates, the bank should consider revising its interest rate policies to make the savings account more attractive.
5. **Credit Score Service:**
   Since the highest number of clients have credit scores between 600 and 700, the bank could offer services to help clients improve their credit scores.
6. **Gender Balance:**
   The bank should aim for a balanced gender distribution among its clientele. It could do this by creating specific financial products tailored to the needs of the underrepresented gender.
7. **Churn in Germany:**
   As Germany has the highest churn rate, the bank needs to investigate the reasons and implement strategies to improve client retention in this country.
8. **Product Promotion:**
   The bank should promote its financial products more evenly across all three countries. Currently, they are primarily offered in Germany.
9. **Churn Analysis:**
   The bank should further investigate the reasons why most clients churn when they reach the ages of 30 to 40. Understanding these reasons could help the bank develop strategies to reduce churn rate.
10. **Salary-Based Services:**
    The bank could offer more services or products based on the estimated salaries of the clients. For example, clients with higher salaries might be interested in investment or wealth management services.
11. **Loyalty Program:**
    The bank should consider implementing a loyalty program. The average estimated salary for loyal clients is quite high, suggesting that retaining these clients is beneficial.
12. **Product-Based Churn:**
    The bank should review the offering and features of Product 4 as it has the highest churn rate. Similarly, it could study and replicate the success elements of the product with the lowest churn rate.
13. **Age-Based Products:** Considering the correlation between age and churn, the bank could consider offering age-based products that cater to the specific needs of different age groups.
14. **Credit Card Utilization:** Since churn is negatively related to credit card usage, the bank could create incentives for credit card utilization to help reduce churn.

15. **Balance and Churn:** Considering the high correlation between balance and churn, the bank should investigate whether clients are churning because they are dissatisfied with the financial returns on their balances.

In addition to the above recommendations, the bank should also consider the following:

The client's overall financial situation: The client’s overall financial situation can also affect churn rates. For example, clients who are struggling financially may be more likely to churn. The bank should consider this when providing financial assistance to clients.

Conclusion:

By following these recommendations, the bank can reduce churn and improve customer retention. This will help the bank to improve its profitability and reputation.