Churn Prediction and Customer Retention Analysis

1. Introduction

In today's business landscape, understanding customer behavior and making informed decisions to retain valuable clients is of paramount importance. This report delves into the critical issue of customer churn and its implications for businesses. By analyzing this problem, we aim to shed light on the underlying factors that lead to customer attrition, present a predictive model, and propose actionable strategies for improving customer retention.

1.1 Justification and Importance

Predicting customer churn is crucial for businesses seeking sustainable growth. High customer attrition rates can lead to substantial revenue loss and hinder overall business success. By addressing this issue proactively, companies can optimize their resources and tailor their offerings to retain valuable customers, leading to increased customer loyalty, revenue, and market share.

1.2 Pitch to Stakeholders

Imagine being able to identify customers who are most likely to churn and offering tailored solutions to retain them. This not only enhances customer satisfaction but also maximizes profitability. By leveraging data-driven insights, we can ensure that our strategies are precise and impactful, allowing us to focus our efforts where they matter most.

1.3 Data Source

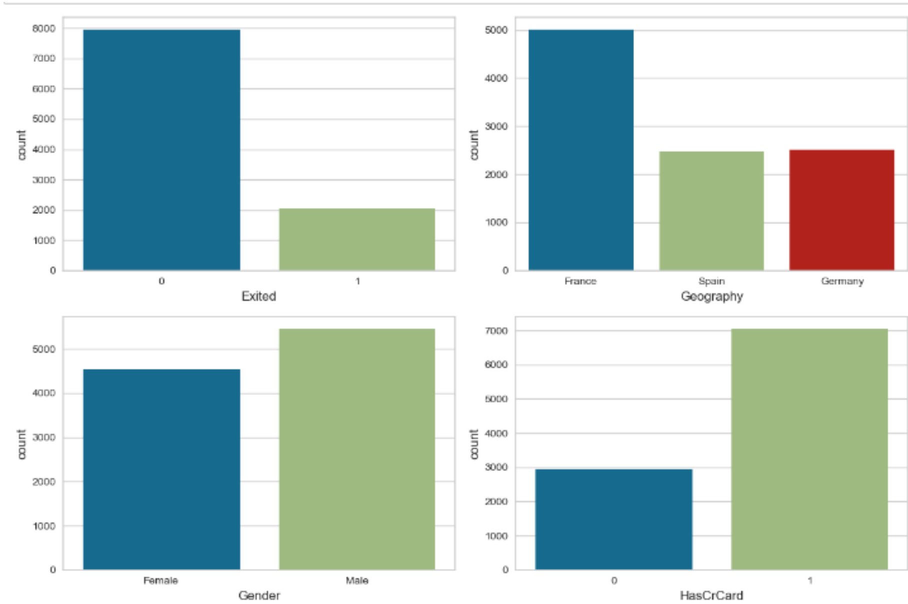
The dataset used for this analysis was obtained from kaggle, encompassing a diverse set of customer attributes, including demographics, behavior, and interactions. This rich dataset serves as a robust foundation for uncovering patterns and trends that contribute to customer churn.

2. Summary

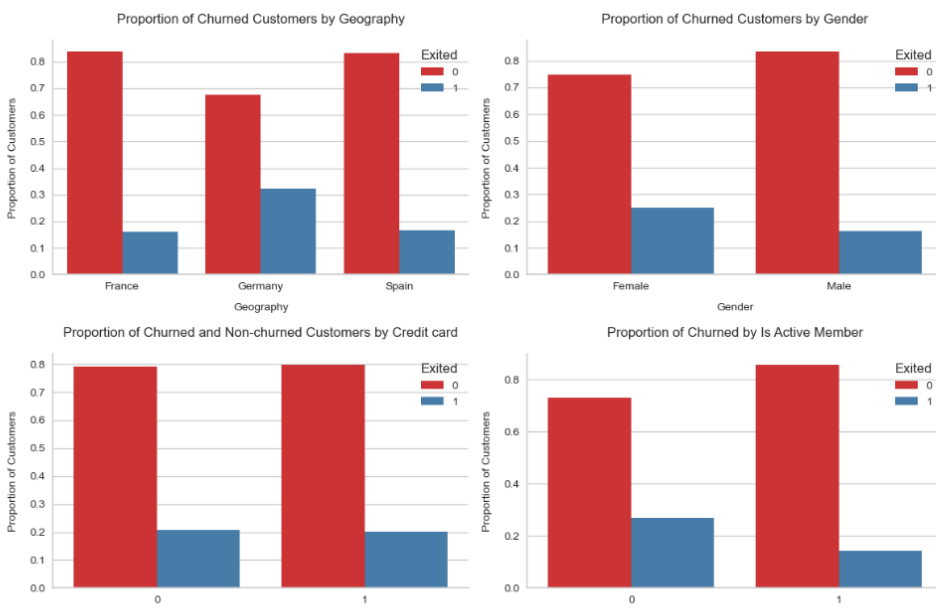
2.1 Exploratory Data Analysis (EDA)

Our exploration of the data revealed intriguing insights into the relationship between various features and customer churn. Visualizations, such as histograms and scatter plots, enabled us to grasp the distribution of key variables and potential correlations, highlighting the significance of factors like geography, age, and product usage. The key findings are below:

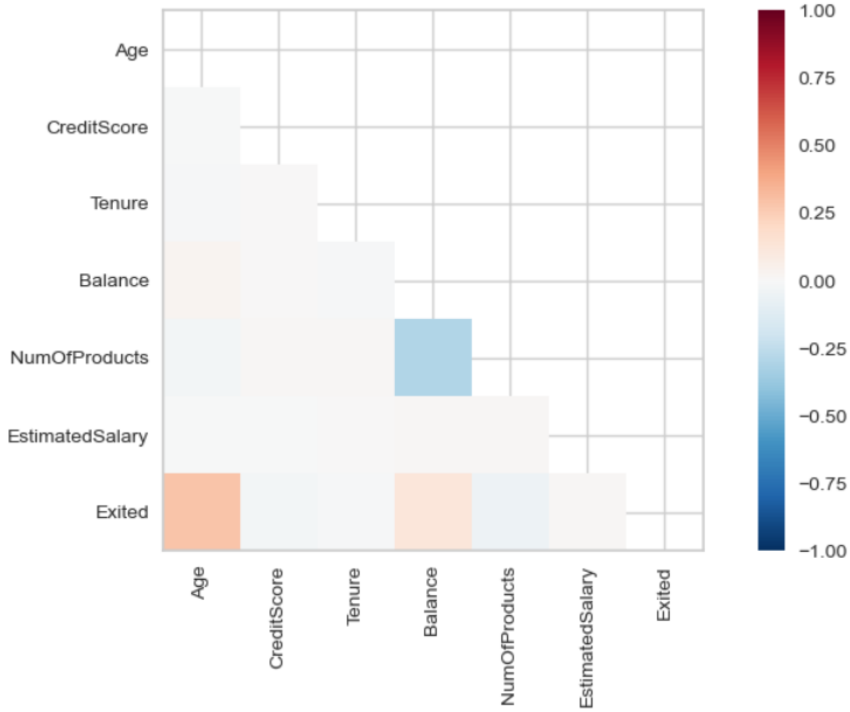
- Most of the customers in the dataset didn't churn out.
- Very few customers in the dataset churned out.
- Most of the customers in the dataset are from Germany.
- There are more male customers in the dataset as compare to female customers.
- Most of the customers in the dataset has credit card as can be shown in the figure below.



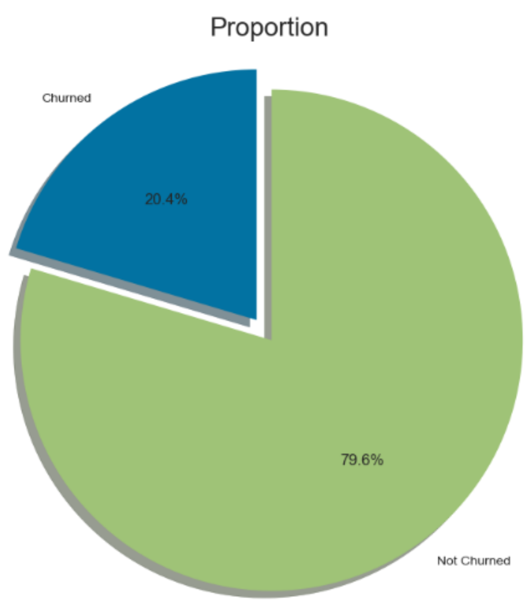
- Highest proportion of customers from Germany have churned out.
- Second highest proportion of customers from Spain have churned out.
- Proportion of churned customers varied by geographical location.
- This could play an important role for predicting whether the customer will churn out or not.
- Female customers are more likely to churn out as compare to male customers.
- Non-Active customers are more likely to churn out as compare to active customers as can be shown in the figure below.



- The numerical variables that are correlation with Exited are Age, Balance and Number of Products as shown in the figure below.



- This pie chart shows that the actual distribution of classes is itself imbalanced for the Exited variable.
- Only ~20% of the dataset Churned as shown in the figure below.



Hence, this dataset and problem statement represent an example of Imbalanced Classification, which has unique challenges in comparison to performing classification over balanced target variables.

2.2 Data Preparation

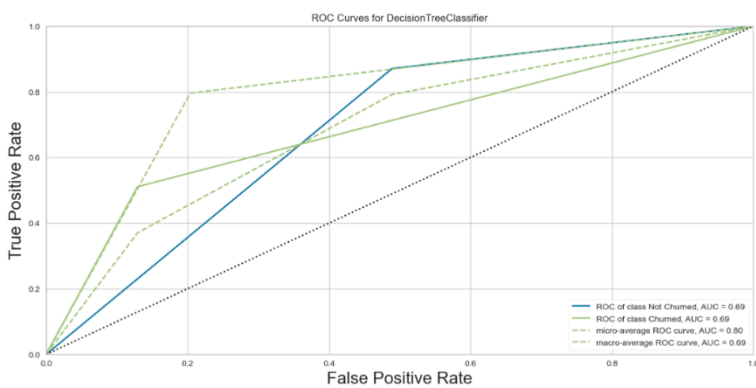
During the data preparation phase, we streamlined the dataset by removing irrelevant columns, optimizing the features for modeling. We introduced a log-transformed age variable to address distribution skewness, capturing potential non-linear age-churn relationships. Categorical variables were transformed into numerical representations via one-hot encoding, enabling machine learning algorithms to process them effectively. Furthermore, we ensured fair feature treatment by employing the StandardScaler to normalize the dataset, aligning feature scales and enhancing the model's convergence and accuracy during analysis and prediction.

Summary of Data Preparation:

- Drop unnecessary columns.
- Replace age variable with log of age.
- Split data into train and test data.
- Create dummy variables for train and test data.

2.3 Model Building and Evaluation

Employing a Decision Tree classifier, we constructed a predictive model for customer churn. The model demonstrated promising performance, achieving approximately 80% accuracy. Beyond accuracy, we extensively evaluated the model using metrics like precision, recall, F1-score, and the ROC-AUC score, providing a comprehensive understanding of its predictive capabilities as can be shown in the figure below.



3. Conclusion

3.1 Analysis and Model Insights

The analysis highlighted the critical role of factors such as geography, age, credit card usage, balance, and number of products in predicting customer churn. These insights enable us to understand customer behavior more deeply, allowing for targeted interventions and strategies.

3.2 Model Deployment and Recommendations

While the model shows strong potential, further refinement and validation are necessary before deployment in a real-world environment. To improve its reliability, incorporating more advanced techniques like ensemble methods and hyper parameter tuning can enhance its predictive power.

3.3 Recommendations

Based on our findings, we recommend implementing personalized retention strategies tailored to customer segments. By addressing specific pain points and enhancing the overall customer experience, businesses can mitigate churn and foster lasting customer relationships.

3.4 Challenges and Opportunities

While we have made significant progress, challenges remain, such as addressing class imbalance and refining the model's performance. Exploring additional data sources and feature engineering techniques could uncover deeper insights and enhance the model's robustness.

In conclusion, this analysis lays the foundation for proactive customer retention strategies. By leveraging data-driven insights and predictive modeling, businesses can navigate the complexities of customer churn and embark on a journey toward sustained growth and success.