Bank Churn Prediction

Project 4: Introduction to Neural Networks

06/07/24

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Appendix

Executive Summary

Overall Insight Summary-

The analysis of customer churn data revealed several key factors contributing to higher churn rates. Notably, customers around the age of 38 and mid-aged customers around the age of 50 show higher churn rates. Additionally, Identified high-risk customers, those with lower credit scores, higher balance customers, customers in Germany, female customers, customers without credit cards and those who are not active members or those with higher bank products all tend to leave more frequently. This comprehensive understanding highlights critical areas for intervention to reduce churn.

Executive Summary (cont)

Recommendations-

Targeted Retention Strategies- Develop and implement retention strategies specifically aimed at younger and mid-aged customers, such as personalized offers and loyalty programs tailored to their needs and preferences.

Geographic Focus- Initiate targeted marketing campaigns and customer service improvements in regions with higher churn rates, particularly Germany, to address region-specific issues and enhance customer satisfaction.

Active Membership Engagement- Encourage customers to become more active members through engagement programs and incentives that foster a sense of belonging and community.

Product Diversification- Promote the adoption of multiple bank products among customers by offering bundled services and benefits for multi-product holders, reducing the likelihood of churn.

Retention Strategy for High-Risk Customers- Lower Credit Scores: Offer credit-building programs. Lower Balance: Provide incentives to increase account usage and engagement.

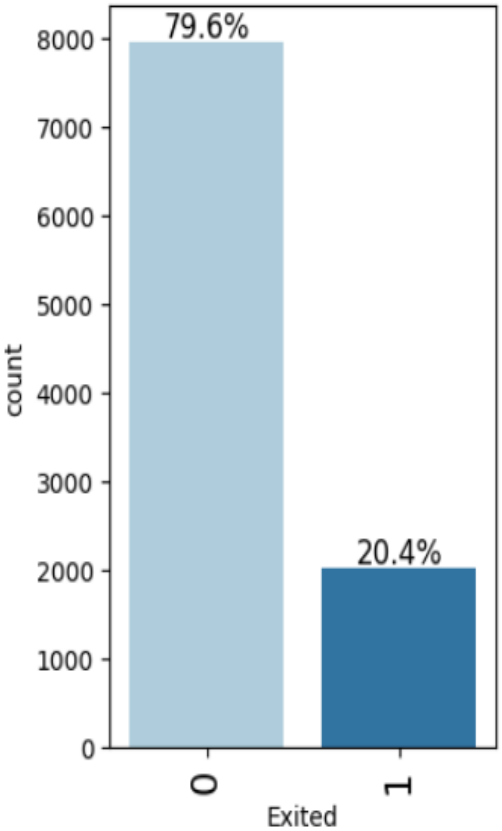
Business Problem Overview and Solution Approach

Problem Definition

Customer churn is a significant issue for businesses, especially in the banking sector, where retaining customers is crucial for long-term profitability. The problem involves **predicting which customers are likely to leave the bank within the next six months** so that proactive measures can be taken to retain them.

Churn Rate- Approximately 20.4% of customers have exited the bank, indicating a significant churn rate that needs to be addressed.

Retention Majority- A majority of 79.6% of customers are retained, which shows a solid base but also highlights the need for improved retention strategies for the 20.4% who exit.



Business Problem Overview and Solution Approach (cont)

Solution Approach / Methodology

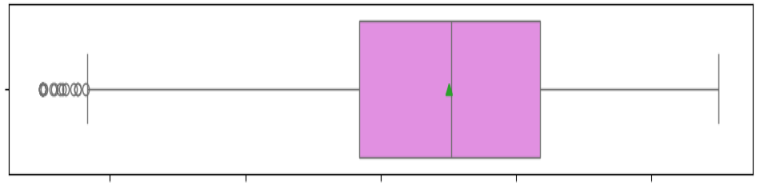
- **Data Collection-** Use sourced data on customer demographics, financial status, and account activity to understand customer behavior and factors influencing churn. 10000 rows, 14 columns.
- **Exploratory Data Analysis (EDA)-** Utilize Python libraries such as Pandas for data manipulation and Matplotlib/Seaborn for visualization. Process: Conduct a thorough analysis to identify key trends and relationships that influence customer churn.
- **Data Preprocessing-** Clean the data, handle missing values, create relevant features, and normalize the data.
- **Model Building-** Develop multiple neural network architectures using TensorFlow and Keras. Experiment with different configurations, including layers, activation functions, and optimizers to enhance model performance.
- **Model Evaluation and Selection-** Compare model performance using metrics like **recall**, precision, and accuracy, and selected the best-performing model emphasizing recall to minimize false negatives in churn prediction .
- **Implementation-** Last is deployment of the model in a production setting for churn prediction and customer retention efforts.

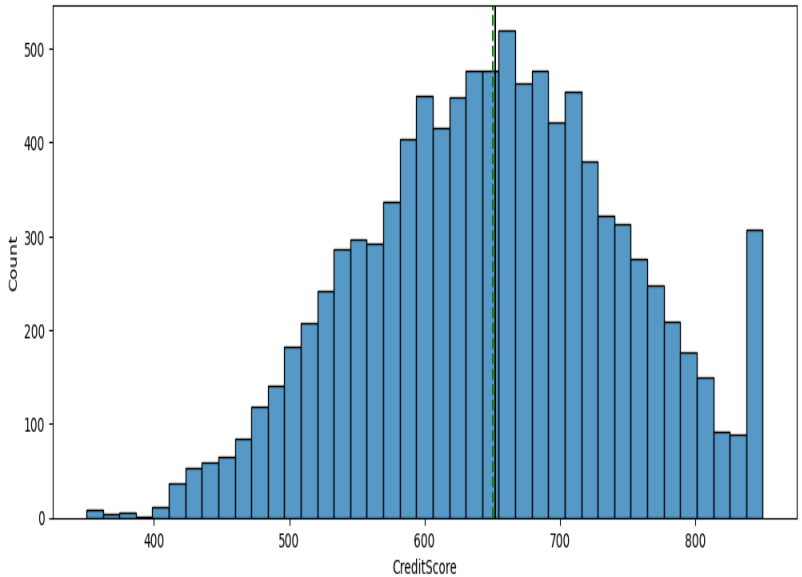
EDA Results

Observations on Credit Score

Insight:
Credit Scores Distribution- Many customers have credit scores clustered between 550-750. Around 650 customers have a score above 800.

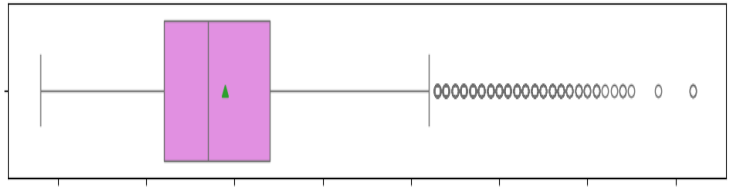
Recommendation:
Credit Score Improvement Programs- Implement credit-building programs to help customers with lower credit scores improve their financial health, reducing their likelihood of churn.

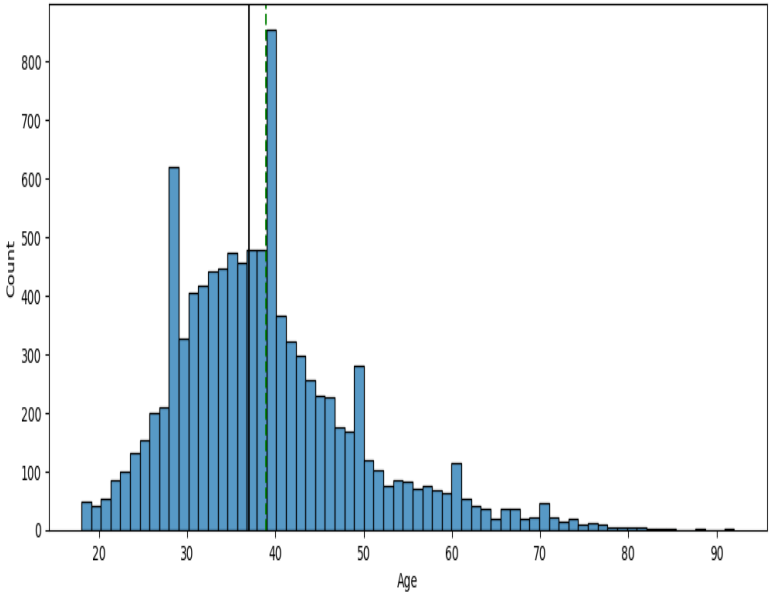




EDA Results (cont)

Observations on Age





Insight:

A normal distribution of the majority of customer showing between the ages of 25 to 50.

The mean or average shows the customers aged around 38 has a significant spike. Additional spikes are observed around ages 28 and 48.

Recommendation:

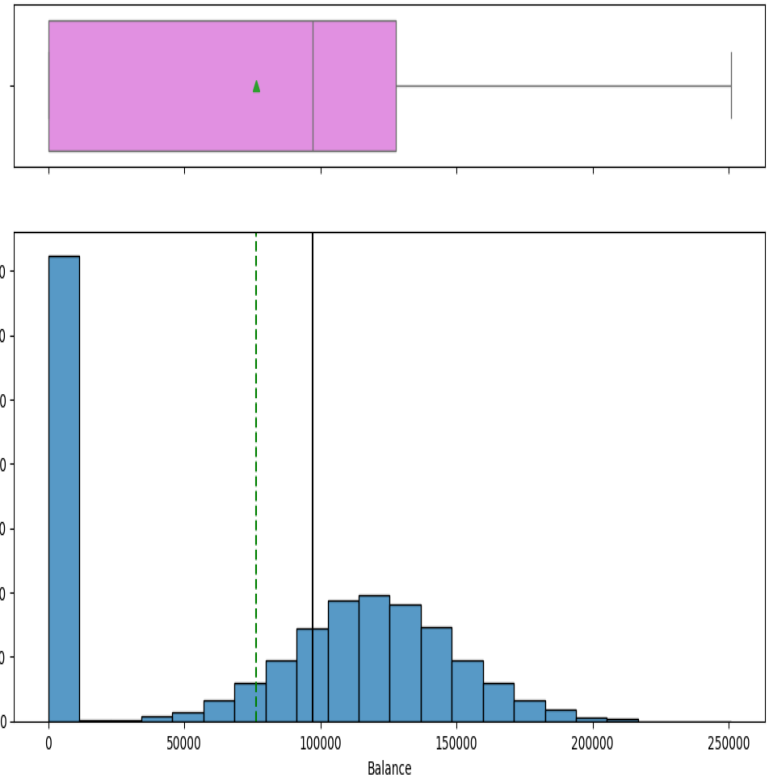
Age-Specific Programs- Implement tailored retention strategies targeting customers in their late 20s, late 30s, and late 40s. This could include personalized financial advice, life stage-specific products, and improved customer service to address their unique needs and reduce churn.

EDA Results (cont)

Observations on Balance

Insight:
Balance Distribution- A significant number of customers have a zero balance, while the rest show a normal distribution between 100K & 150K.

Recommendation:
Incentivize Maintaining a Balance- Introduce programs that encourage customers to maintain a positive balance, such as offering higher interest rates or rewards for maintaining a certain minimum balance. This could help reduce the high churn rate among customers with zero balances.

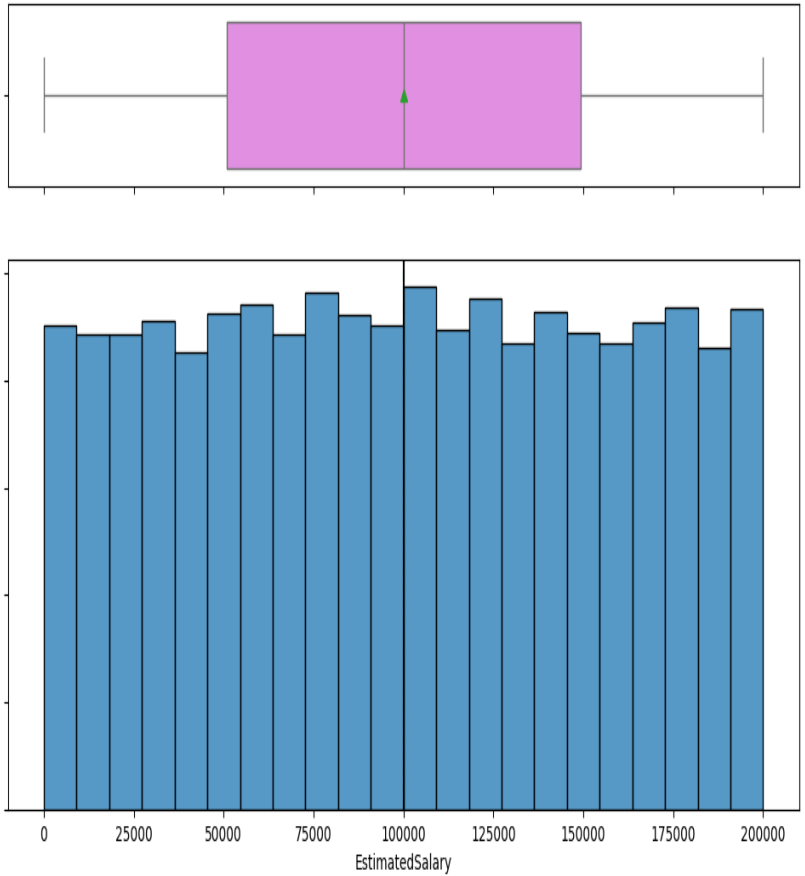


EDA Results (cont)

Observations on Estimated Salary

Insight- The distribution of estimated salaries is relatively uniform, with no significant peaks or valleys. The median estimated salary is around 100,000, and the range spans from 0 to 200,000.

Recommendation- To cater to a broad range of customers, consider offering tailored financial products and services that address the diverse income levels. Develop marketing strategies that emphasize value propositions relevant to both lower and higher-income segments to ensure inclusive financial engagement.

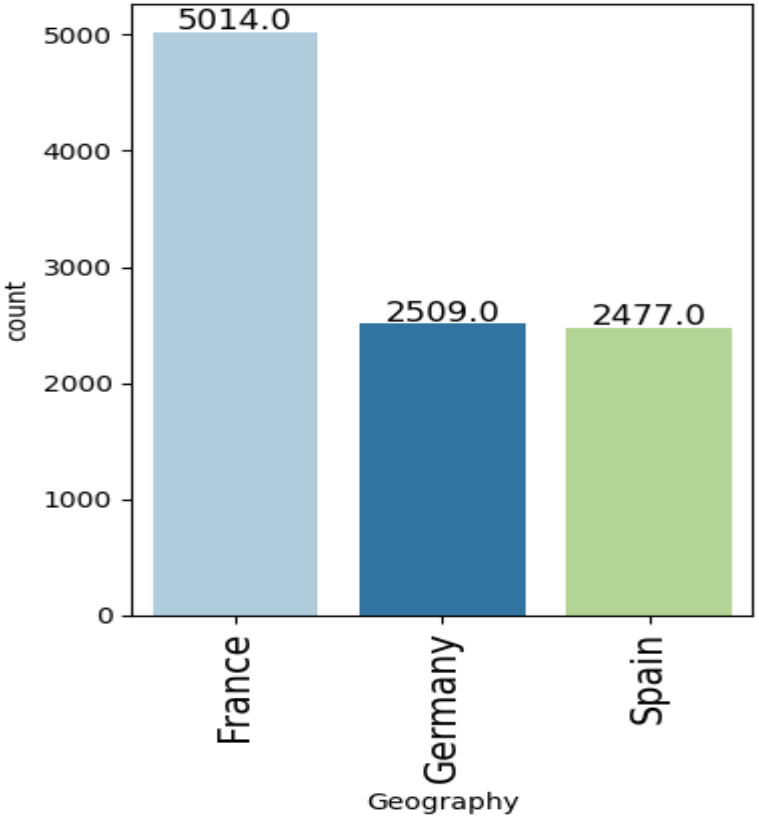


EDA Results (cont)

Observations on Geography

Insight- The customer distribution across geography shows that France has the highest number of customers (5014), followed by Germany (2509) and Spain (2477). This indicates a significant regional disparity in customer base distribution.

Recommendation- Enhance regional marketing efforts in Germany and Spain to attract more customers and balance the geographical distribution. Additionally, analyze the factors contributing to higher customer numbers in France to replicate successful strategies in other regions.

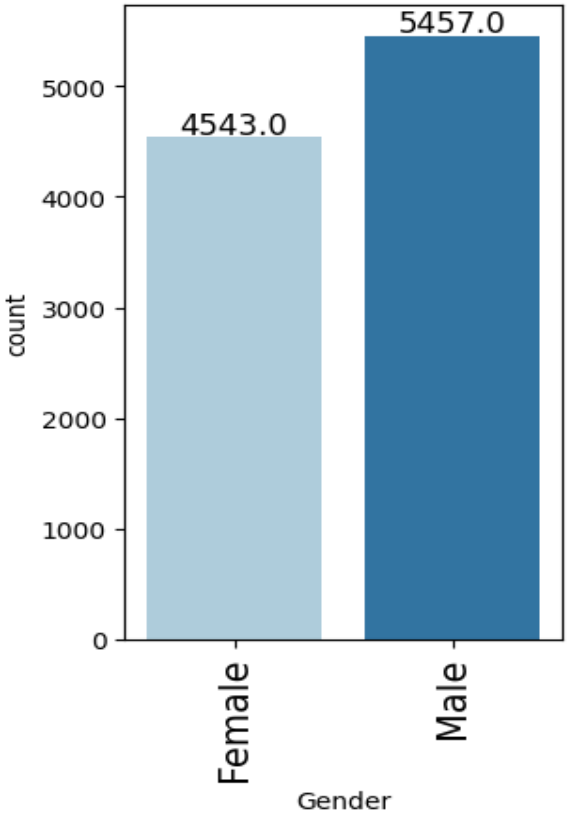


EDA Results (cont)

Observations on Gender

Insight- There are more male customers (5457) than female customers (4543). This indicates a gender imbalance in the customer base, with males making up a slightly larger portion.

Recommendation- Implement targeted marketing campaigns and services aimed at attracting more female customers. By understanding and addressing the specific needs and preferences of female customers, the bank can work towards balancing the gender distribution and potentially increasing overall customer satisfaction and retention.

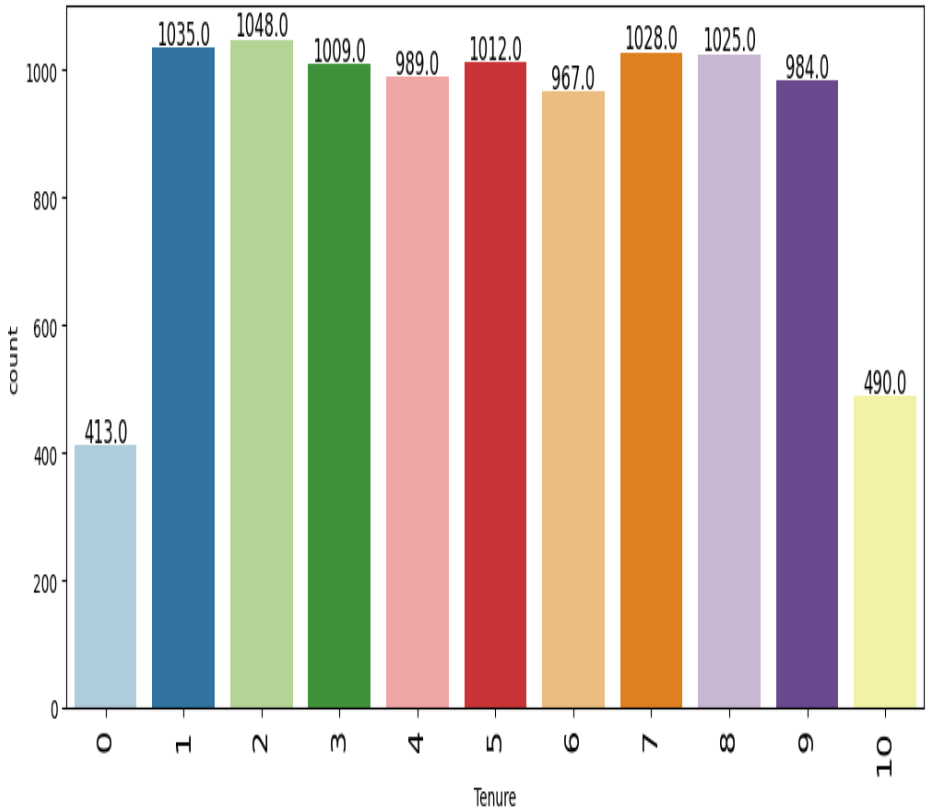


EDA Results (cont)

Observations on Tenure

Insight:
Tenure Distribution- Customers with 1 to 9 years of tenure are evenly distributed, while those with 0 and 10 years show significantly lower counts.

Recommendation:
Tenure-Based Loyalty Programs- Develop loyalty programs targeting customers with low tenure (0 years) and those nearing 10 years to improve retention rates within these specific tenure brackets.

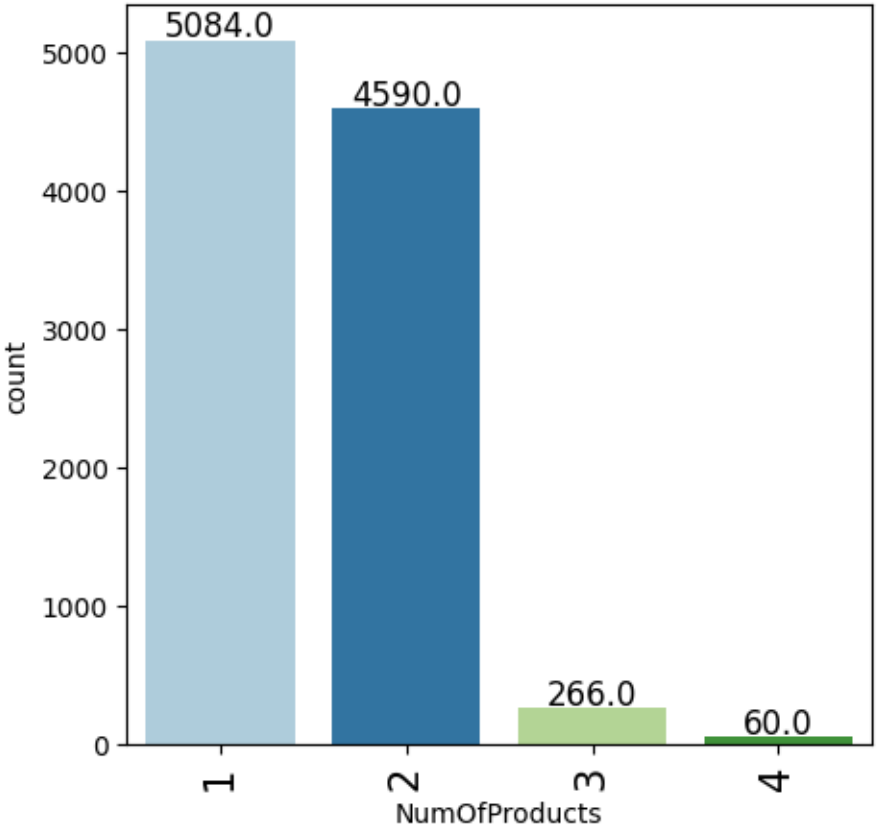


EDA Results (cont)

Observations on Number of Products

Insight- Many customers hold either 1 or 2 products with the bank. There is a significant drop in the number of customers holding 3 products, and very few hold 4 products.

Recommendation- Review and simplify the bank's product offerings. Focus on improving the value proposition of holding multiple products to ensure they meet customer needs effectively and do not lead to service complexity or customer overload.

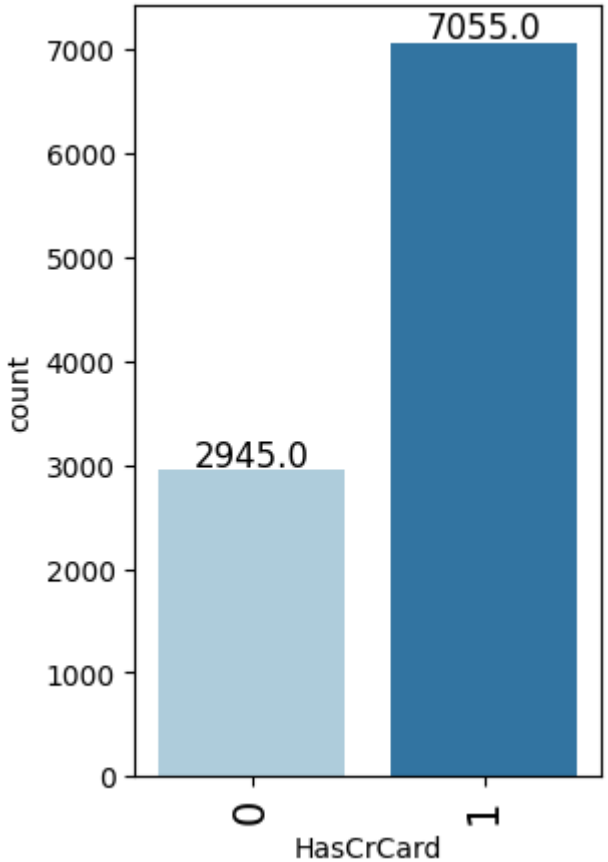


EDA Results (cont)

Observations on Has Credit Card

Insight:
Credit Card Ownership - A significant portion of customers 7055 have credit cards and 2945 do not.

Recommendation:
Credit Card Offers - Implement targeted marketing campaigns and personalized offers for customers without credit cards to encourage adoption and increase product engagement.

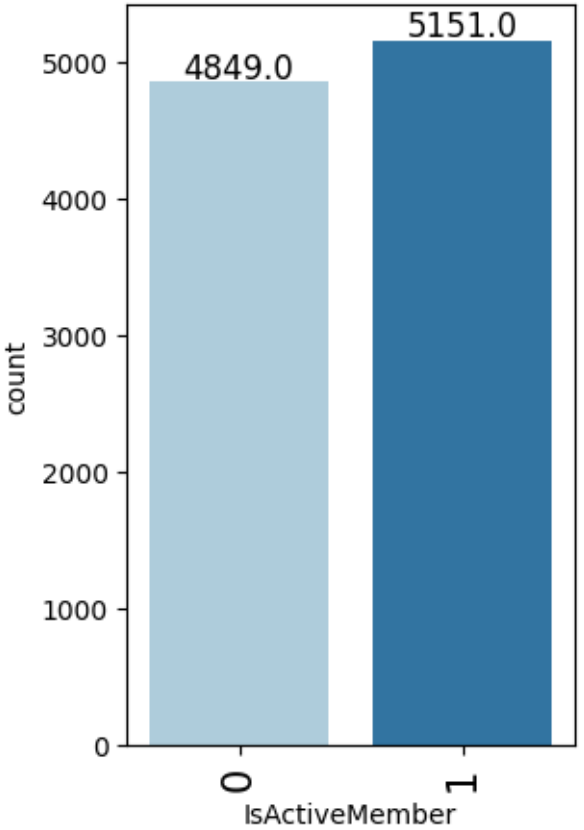


EDA Results (cont)

Observations on Is Active Member

Insight:
Slightly more active members (5151) compared to inactive members (4849).

Recommendation:
Enhanced Customer Engagement- Develop targeted engagement programs for inactive members to increase their activity levels and reduce churn.

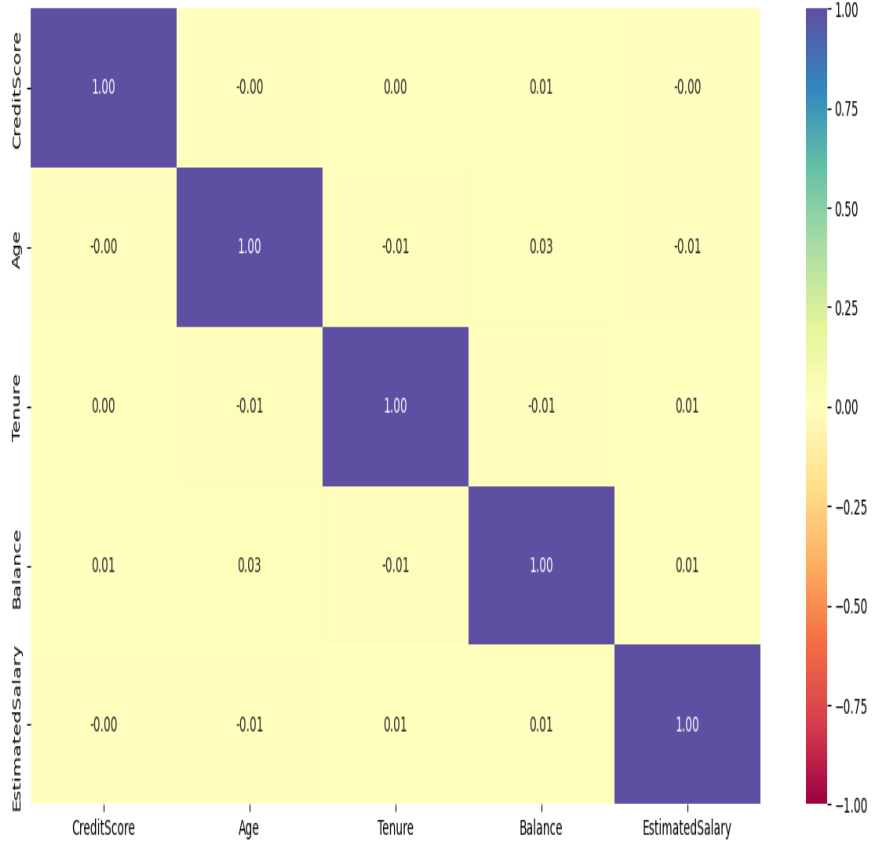


EDA Results (cont)

Correlation Plot

Insight- The correlation matrix shows very low correlation values between customer demographics, financial status, and account activity variables, indicating that these features are largely independent of each other. For instance, the highest correlation observed is only 0.03 between Age and Balance.

Recommendation- Since the features are largely independent, consider focusing on a combination of variables rather than individual ones for more accurate churn prediction models.

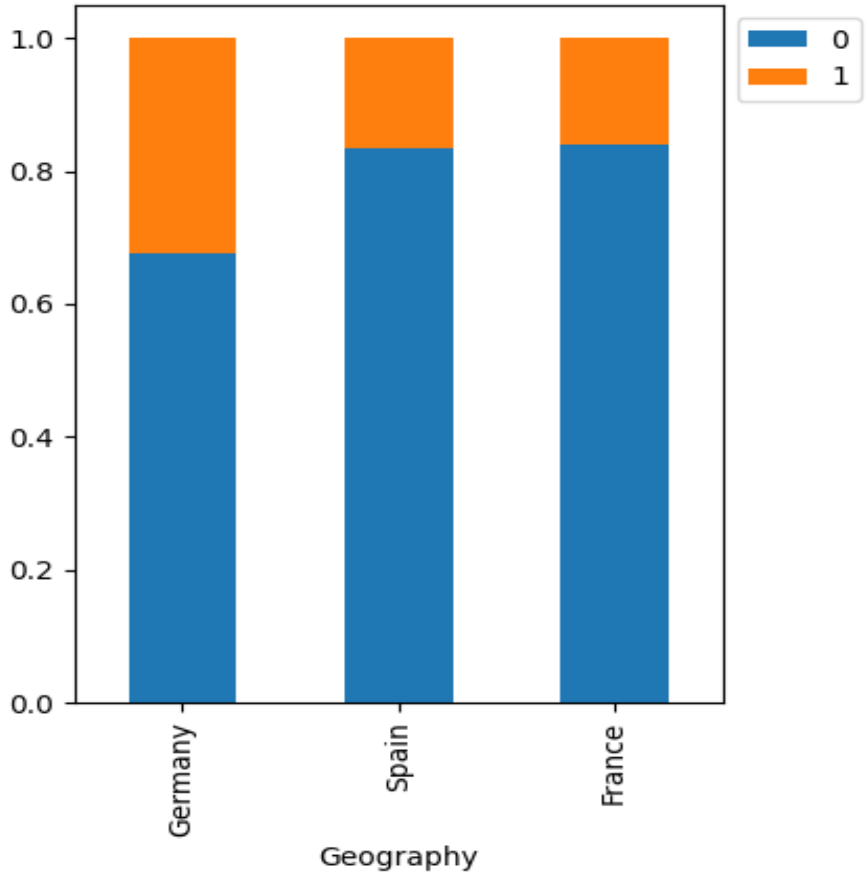


EDA Results (cont)

Exited vs Geography

Insight:
Geography and Churn- Customers in Germany have a higher churn rate compared to those in Spain and France.

Recommendation:
Geographic-Specific Strategies- Focus on regions with higher churn rates, such as Germany, to tailor marketing and service improvement efforts specific to these regions.

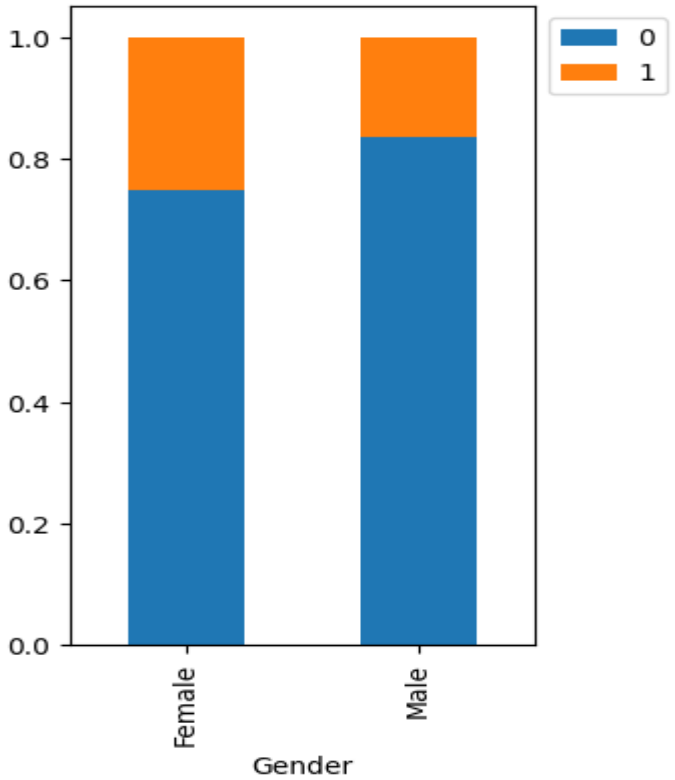


EDA Results (cont)

Exited Vs Gender

Insight:
Gender Differences- Female customers have a higher churn rate compared to male customers.

Recommendation:
Gender-Specific Initiatives- Implement targeted retention programs for female customers, such as personalized offers and services to address their specific needs and reduce churn.



EDA Results (cont)

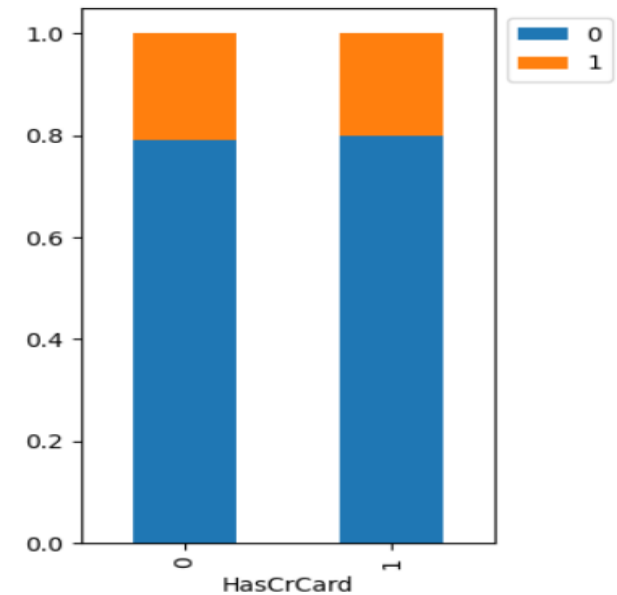
Exited Vs Has Credit Card

Insight- Among customers with a credit card (7055) approximately 79.82%, (1424) exited approximately 20.18%, whereas for those without a credit card (2945) around 79.18%, (613) exited approximately 20.82%. This shows that the churn rate is slightly higher for customers without a credit card, but the difference is minimal.

Recommendation- Focus on more influential factors for reducing churn and consider targeted retention strategies for customers without credit cards.

```
stacked_barplot(ds, 'HasCrCard', 'Exited')
```

```
Exited          0     1    All
HasCrCard
All          7963  2037  10000
1            5631  1424   7055
0            2332   613   2945
```

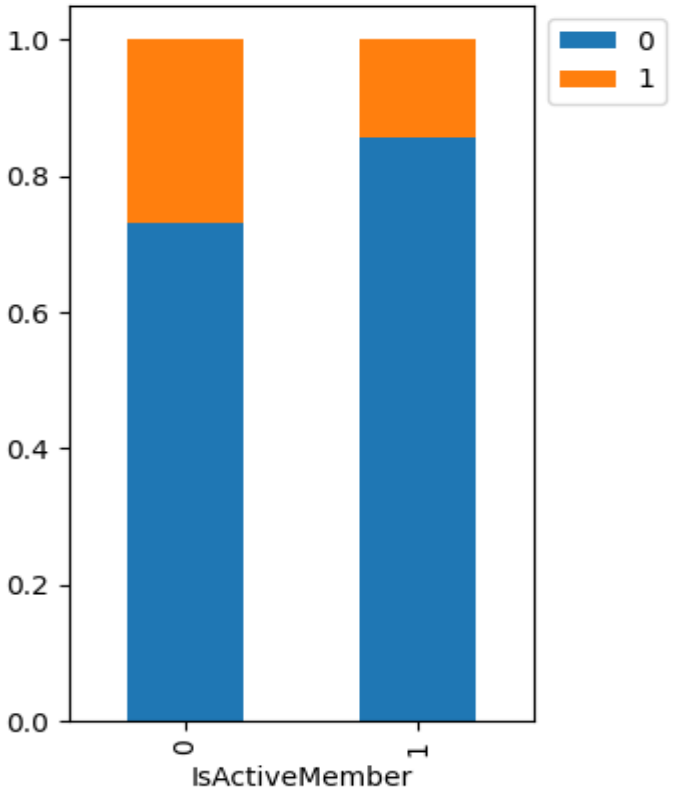


EDA Results (cont)

Exited Vs is Active Member

Insight- There is a noticeable increase in churn among inactive members compared to active ones, indicating that inactivity may contribute to higher customer turnover.

Recommendation- Implement engagement programs that specifically target inactive members, such as reactivation incentives or personalized outreach efforts, to increase their activity and reduce the likelihood of churn.

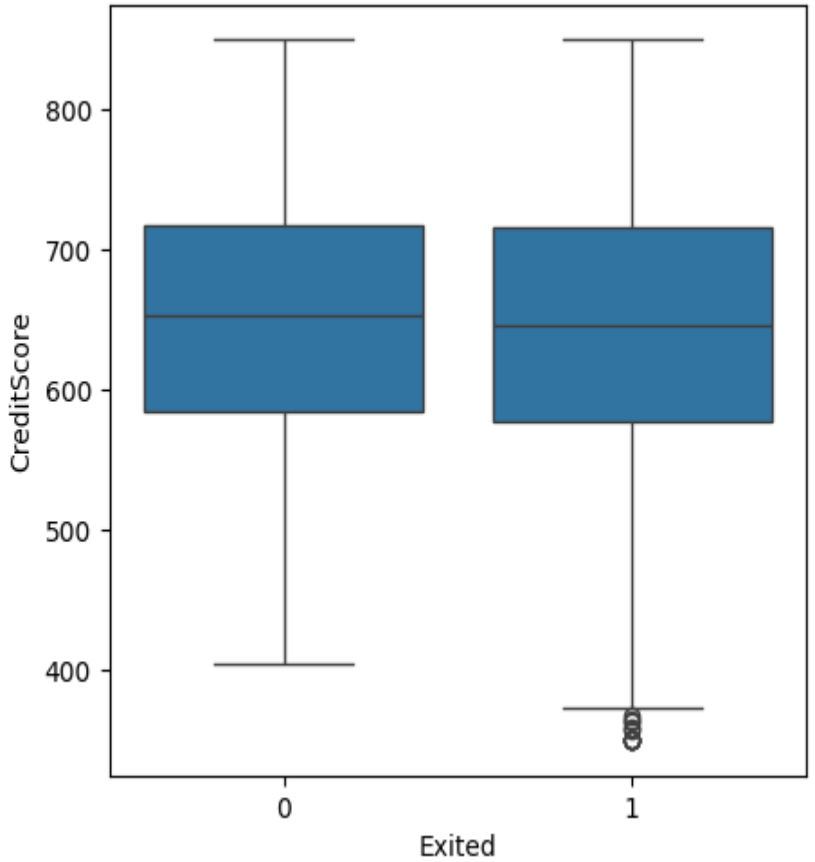


EDA Results (cont)

Exited Vs Credit Score

Insight- Customers who exited have a similar median credit score to those who did not exit, but with a slightly wider range and more outliers on the lower end of the credit score distribution.

Recommendation- Consider implementing targeted financial products or services to support customers with lower credit scores to reduce the risk of churn, as they represent a vulnerable group with higher variability in their credit scores.

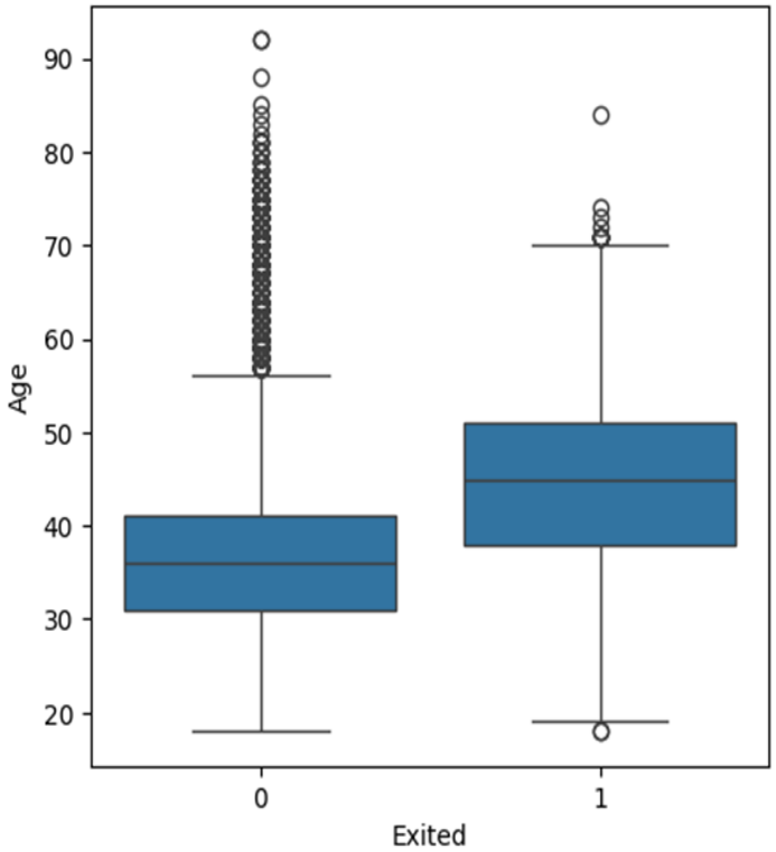


EDA Results (cont)

Exited Vs Age

Insights- Customers who have exited the bank (churned) tend to be older, with a median age of around 45 years compared to non-exited customers with a median age of around 35 years. Older customers, particularly those aged 38 and above, show a higher likelihood of churning.

Recommendation- Age-Specific Retention Strategies: Implement targeted retention strategies for older customers, including personalized financial planning services, retirement products, and loyalty programs tailored to their needs to improve satisfaction and reduce churn.

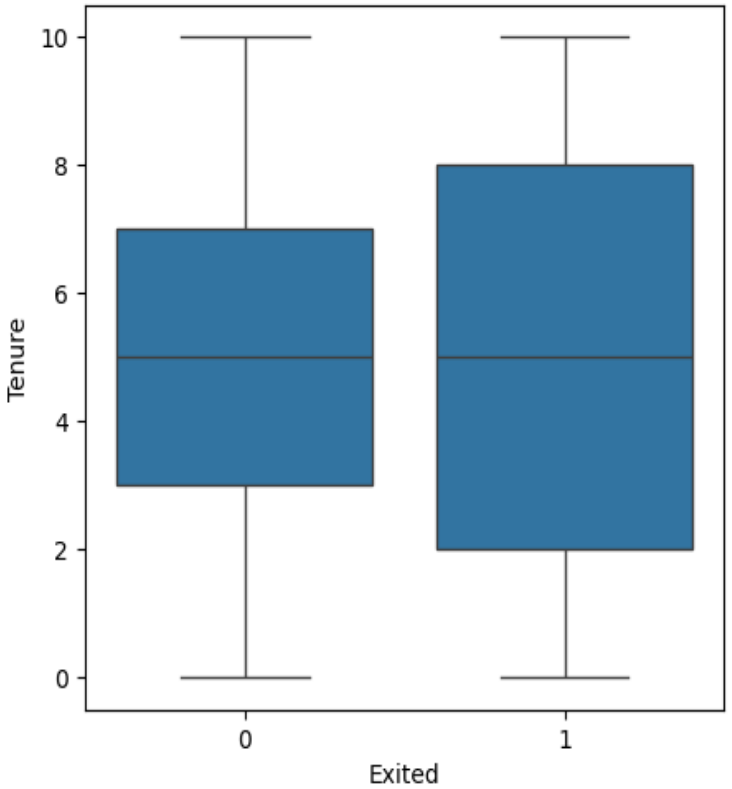


EDA Results (cont)

Exited Vs Tenure

Insight- Both exited and non-exited customers have a median tenure of around 5 years. However, exited customers generally have a wider range of tenure years (2 to 8 years) compared to those who stayed (3 to 7 years).

Recommendation- Focus retention efforts on customers with tenure between 2 to 8 years by providing personalized engagement strategies and loyalty programs. Address specific needs and concerns through regular feedback mechanisms to reduce churn and maintain customer satisfaction. This targeted approach can help in retaining a higher percentage of at-risk customers.

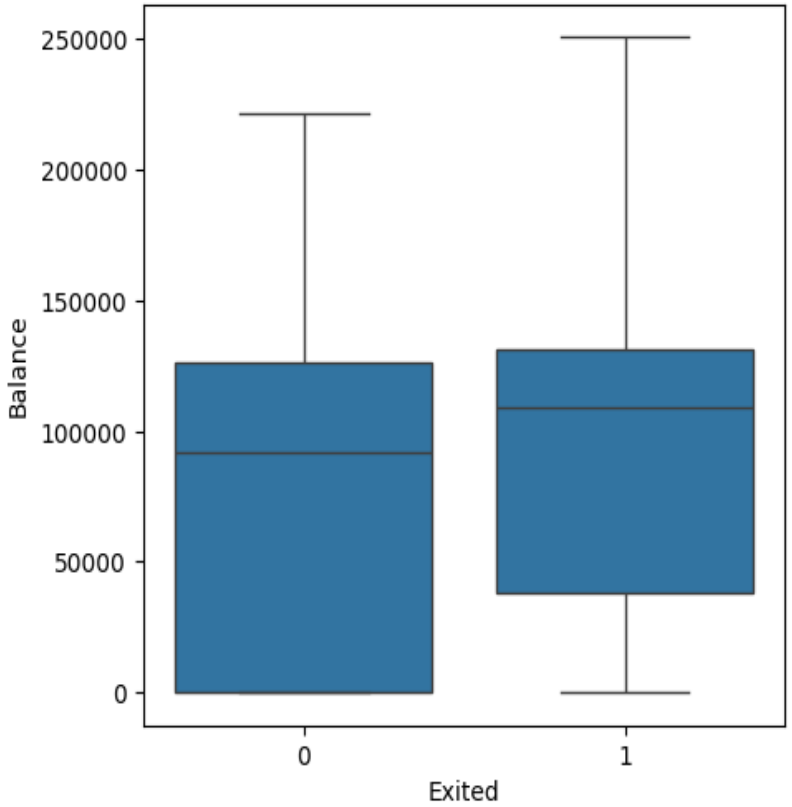


EDA Results (cont)

Exited Vs Balance

Insight- Customers who exited have a slightly higher median balance compared to those who stayed, indicating that higher-balance customers are also at risk of churn.

Recommendation- Implement proactive engagement strategies for higher-balance customers, such as exclusive offers or premium services, to enhance customer satisfaction and loyalty, reducing the risk of churn.

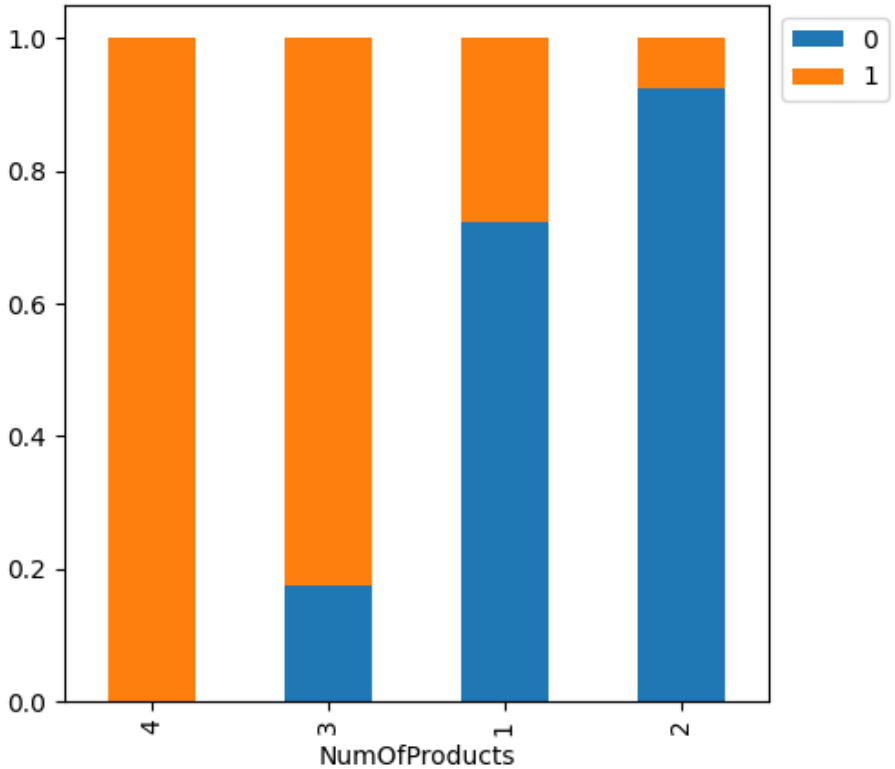


EDA Results (cont)

Exited Vs Number of Products

Insight- There is a significantly higher churn rate for customers with 3 or 4 products compared to those with 1 or 2 products. This indicates that customers with more products are more likely to leave the bank.

Recommendation- Develop targeted retention strategies for customers with multiple products by offering exclusive benefits and personalized services to enhance their satisfaction and reduce churn.



EDA Results (cont)

Insight-Based Questions

Why do customers from Germany have a higher churn rate? Possible reasons could be competitive banking products in Germany or dissatisfaction with the services provided.

Why do churn rates spike for Ages above 35? This might be due to major life transitions such as career changes, family growth, or financial instability.

What factors contribute to the higher churn rates among customers with zero balance? Understanding the financial behaviors and needs of these customers can help develop targeted retention strategies.

Why is there a higher churn rate among inactive members? Identifying the reasons behind inactivity could provide insights into improving engagement and reducing churn.

EDA Summary

Key Observations and Insights

Age- Churn rates are notably higher among customers around ages 30 and 40, pointing to potential life transitions impacting their banking needs.
Balance- There are a significant number of customers with zero balances. Customers who exited have a slightly higher median balance compared to those who stayed.
Credit Score- Lower credit scores correlate with higher churn, suggesting financial health as a key factor in retention.
Products- Customers with a greater amount of bank products tend to leave more frequently, underscoring the need to understand why that is.
Active Membership- Inactive members show higher churn rates, indicating the importance of regular engagement.
Geography- Customers in Germany exhibit higher churn rates compared to those in Spain and France, which may be due to competitive market dynamics or service dissatisfaction.

Data Preprocessing

Data Preprocessing

Duplicate Value Check: Ensured there were no duplicate entries in the dataset.

Missing Value Treatment: Handled missing values by either imputing or removing them. Missing values in columns like EstimatedSalary, Balance, Tenure and Age were particularly addressed.

Outlier Check: Identified outliers, particularly in financial metrics like balance and salary. Extreme values in Balance and EstimatedSalary were analyzed using boxplots and histograms.

Feature Engineering: Created new features that might help in predicting churn, such as interaction terms or aggregated features. For example, created dummy variables for categorical features like Geography and Gender.

Data Preparation for Modeling: Normalized the data to bring all features to the same scale for 'CreditScore', 'Age', 'Tenure', 'Balance' and 'EstimatedSalary'. Split the data into training, validation, and test sets to ensure robust model evaluation.

Model Performance Summary

Overview of Model and Its Parameters

Model 0: Neural Network with SGD Optimizer

Model 1: Neural Network with Adam Optimizer

Model 2: Neural Network with Adam Optimizer and Dropout

Model 3: Neural Network with Balanced Data (SMOTE) and SGD Optimizer

Model 4: Neural Network with Balanced Data (SMOTE) and Adam Optimizer

Model 5: Neural Network with SMOTE, Adam Optimizer, and Dropout

Model Performance Summary (cont)

Summary of the Final Model for Prediction

The final model selected is **Model 3: Neural Network with Balanced Data (SMOTE) and SGD Optimizer** achieving the best balance between training, validation, and test performance assuring the lowest overfitting possible.

Key Performance Metrics for Training, Validation and Test Data

Model	Training Recall	Validation Recall	Test Recall
NN with SGD	0.00	0.00	0.00
NN with Adam	0.54	0.47	0.47
NN with Adam & Dropout	0.56	0.48	0.48
NN with SMOTE & SGD	**0.73**	**0.68**	**0.68**
NN with SMOTE & Adam	0.97	0.53	0.53
NN with SMOTE, Adam & Dropout	0.87	0.67	0.67

APPENDIX

Data Background and Contents

Data Background and Contents

The dataset contains 10,000 rows and 14 columns, including customer demographics, account information, and whether the customer has exited the bank.

The key columns include:

- **CreditScore**: Credit history score
- **Geography**: Location of the customer
- **Gender**: Gender of the customer
- **Age**: Age of the customer
- **Tenure**: Number of years with the bank
- **Balance**: Account balance
- **NumOfProducts**: Number of products purchased from the bank
- **HasCrCard**: Whether the customer has a credit card
- **IsActiveMember**: Whether the customer is an active member
- **EstimatedSalary**: Estimated salary
- **Exited**: Whether the customer left the bank within six months

Solution Approach in Further Detail

1. Data Collection:
Objective: Gather relevant data to understand customer behavior and factors influencing churn.

Data Sources: The data was sourced from the bank's internal databases, including customer demographics, account information, transaction history, and churn status.

Key Columns: The dataset includes 10000 rows and 14 columns, with key columns such as Customer ID, Surname, Credit Score, Geography, Gender, Age, Tenure, Balance, Number of Products, Has Credit Card, Is Active Member, Estimated Salary, and Exited.

Solution Approach in Further Detail (cont)

2. Exploratory Data Analysis (EDA):

Objective: Understand the data's structure, identify patterns, and detect anomalies.

Techniques Used:

Descriptive Statistics: Calculated mean, median, standard deviation, etc., to summarize the data.

Visualization: Created histograms, boxplots, and bar charts to visualize distributions and relationships.

Correlation Analysis: Identified correlations between features and the target variable (Exited).

Solution Approach in Further Detail (cont)

3. Data Preprocessing:
Objective: Clean and prepare the data for modeling.

Steps:

Duplicate Value Check: Ensured there were no duplicate entries to maintain data integrity.

Missing Value Treatment: Handled missing values using appropriate methods (e.g., imputation, removal). Missing values in critical columns like Estimated Salary, Balance, Tenure, and Age were specifically addressed.

Outlier Check: Identified outliers, particularly in financial metrics like balance and salary, using boxplots and histograms.

Solution Approach in Further Detail (cont)

Feature Engineering: Created new features and transformed categorical variables into dummy variables. Examples include:

- **Interaction Terms:** Created interaction terms between features that might have a combined effect on churn.
- **Dummy Variables:** Converted categorical features such as Geography and Gender into numerical format using one-hot encoding.

Data Preparation for Modeling:

- **Normalization:** Scaled numerical features to bring them to the same range, which is essential for neural network models.
- **Data Splitting:** Split the data into training (70%), validation (15%), and test (15%) sets to ensure robust model evaluation and prevent overfitting.

Solution Approach in Further Detail (cont)

4. Model Building:

Objective: Develop and train machine learning models to predict customer churn.

Models Developed:

Neural Network Models: Several configurations of neural networks were built using different techniques.

Model 0: Neural Network with SGD Optimizer.

Model 1: Neural Network with Adam Optimizer.

Model 2: Neural Network with Adam Optimizer and Dropout.

Model 3: Neural Network with Balanced Data (SMOTE) and SGD Optimizer.

Model 4: Neural Network with Balanced Data (SMOTE) and Adam Optimizer.

Model 5: Neural Network with SMOTE, Adam Optimizer, and Dropout.

Solution Approach in Further Detail (cont)

5. Model Evaluation and Selection:

Objective: Evaluate the performance of the models and select the best one.

Metrics Used:

Recall: Measured the ability of the model to correctly identify churned customers.

Precision: Assessed the accuracy of the churn predictions.

Accuracy: Evaluated the overall correctness of the model.

Model Selection:

Compared models based on their performance on the validation and test sets.

Selected Model **Model 3:** Neural Network with Balanced Data (SMOTE) and SGD Optimizer as it achieved the best balance between training, validation, and test performance.

Solution Approach in Further Detail (cont)

6. Implementation:

Objective: Deploy the final model in a production environment to predict churn and support customer retention efforts.

Steps:

Model Deployment: Integrated the final model into the bank's existing systems to start predicting customer churn in real-time.

Monitoring and Maintenance: Set up processes to monitor the model's performance over time and update it as needed to maintain accuracy and relevance.

Great Learning

Happy Learning !