

Bank Churn Prediction

Project #4 AIML for Business

Matthew Clark

June 8, 2024

Contents / Agenda

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

Executive Summary

- Banks are susceptible to losing customers to other banks. This is called customer churn. In order for a Bank to avoid losing customers it is important they understand how their services can effect a customers decision to leave for another Bank. This information can be used by management to make the most informed decisions on improving service that have the largest impact on customer decisions.
- The data indicates that customers that remain at the bank have higher balances. This may indicate that customers with lower balances are not be happy with the service they are receiving. We recommend you improve service levels for lower balance customers to avoid exit.
- Customers that have a credit card remain at the bank at a much higher rate. We recommend that credit cards are encouraged for all bank customers to avoid exit.
- Women are more likely to exit the bank. Improving service for Women is recommended.
- Customers that exit the bank are younger than customers that remain. We recommend regularly ensuring younger customers are receiving the same level of service as older cutomers.

Executive Summary (continued)

- Credit Scores and Estimated Earnings are not indicators that a customer may exit the bank as the numbers are very similar for both columns. We recommend that service strategies be consistent across all credit scores and levels of income.
- Member Activity is slightly less for customers that exit. We recommend closely monitoring activity and reaching out to any customer who becomes less active before they exit.
- The data shows very few customers have more than 1 or 2 bank products. We recommend service for bank customers encourages them to use more bank products.
- Customers from Germany are more likely to exit the bank. Service levels in Germany should be improved.
- The data set provided did not contain information on any rewards offered by the bank. However, rewards and incentives are an important part of encouraging customers retention.

Business Problem Overview and Solution Approach

- Banks can face churn, customers leaving and joining another bank. Banks are service providers, and different aspects of their service may influence a customer's decision to leave the bank. As a result, it is important for Banks to understand which aspects of business are influencing customer's decisions. This allows Bank Management to concentrate their efforts improving the services that have the largest effect on customer decisions.
- The Data Science team has been tasked with creating a neural network-based classification model to determine which customers are likely to leave or stay in the next six months.
- Our approach to this task will start with processing and analyzing the dataset to ensure accuracy and best modeling. Next, several different models will be constructed using different hyperparameters for the neural network models. The output of the different models will be assessed and compared using recall as the primary metric. To select the best model for this task we will tune the models that performed well initially and then compare those results to determine the best and final model for this task.
- We selected Recall as the appropriate metric because the cost of misidentifying a customer that will leave the bank is higher than misidentifying a customer that will stay with the bank.

EDA Results

- There is very little difference between the credit scores of customers that have exited the bank and customers that remain with the bank.
- There are some outliers for customers that have very low credit scores.
- Customer ages are right skewed.
- Age has outliers that seem logical as the age range for bank customers can go late in life.
- The Age of customers that have exited the bank are somewhat higher.
- The average Age of customers that have exited the bank is 45.
- The average Age of customers that remain at the bank is 36.
- Customers that have exited the bank have lower balances.
- A third of customers have a very low balance. Most of the balances for customers that remain at the bank are 40,000 and higher.
- Estimated salary is very evenly distributed among both classes of customers.

EDA Results (continued)

- Half of customers are in Germany, 25% in Spain, and the remaining 25% in France.
- More customers in Germany have exited the bank, but that also make up a majority of bank users.
- Males make up nearly 55% of bank customers.
- Females are slightly more likely to have exited the bank.
- The same number of customers that have exited have credit cards as those that are still at the bank.
- Inactivity is higher among customers that have exited the bank.
- Customers that have exited the bank have a median Tenure of 5 years.
- Customers that remain with the bank also have a median of 5 years, however the range of the Tenure is larger with more customers having both a smaller and larger range.
- Customers that have exited the bank and customers that remain both primarily have 1 or 2 products from the bank. Very few have 3 or 4 products.

Data Preprocessing

- No duplicate values were identified.
- The data contained no missing values.
- The outliers that existed in the data in Age and Credit Score were logical considering a bank has a large variety of customers.
- The CustomerID, RowNumber, and Surname columns were removed because they contained unique values.
- OneHotEncoding was used for the Geography column.
- The data was then split into training, validation, and testing before further processing in order to avoid any data leakage in the testing data.
- After splitting the data was normalized using StandardScaler.

Model Performance Summary

- We built six different neural network models using different hyperparameters, such as optimizers, dropouts, and data balancing using SMOTE.
- The models we chose were:
 - Neural Network with SGD Optimizer
 - Neural Network with Adam Optimizer
 - Neural Network with Adam Optimizers and Dropout
 - Neural Network with Balanced Data (by applying SMOTE) and SGD Optimizer
 - Neural Network with Balanced Data (by applying SMOTE) and Adam Optimizer
 - Neural Network with Balanced Data (by applying SMOTE), Adam Optimizer, and Dropout
- Different hyperparameter combinations will be tried to improve model results including epochs, batch size, initializers, optimizers, and dropout.

Model Performance Summary

- After many hyperparameter combinations were applied the models yielded these results:

Training Data	
NN with SGD	0.197086
NN with Adam	0.584356
NN with Adam & Dropout	0.483896
NN with SMOTE & SGD	0.722331
NN with SMOTE & Adam	0.845761
NN with SMOTE,Adam & Dropout	0.826923

Validation Data	
NN with SGD	0.174847
NN with Adam	0.429448
NN with Adam & Dropout	0.429448
NN with SMOTE & SGD	0.693252
NN with SMOTE & Adam	0.641104
NN with SMOTE,Adam & Dropout	0.769939

Final Model

- The models that did not have upsampling (SMOTE) performed poorly.
- The model that performed best was the Neural Network with SMOTE, Adam Optimizer, and Dropout.
- The hyperparameters for the model included a total of 5 layers, with Relu and Sigmoid activation, and 2 Dropouts with .5 and .2 rates.
- The classification reports shows the best generalizing with this model.

Final Model	
Training	81
Validation	79
Testing	78

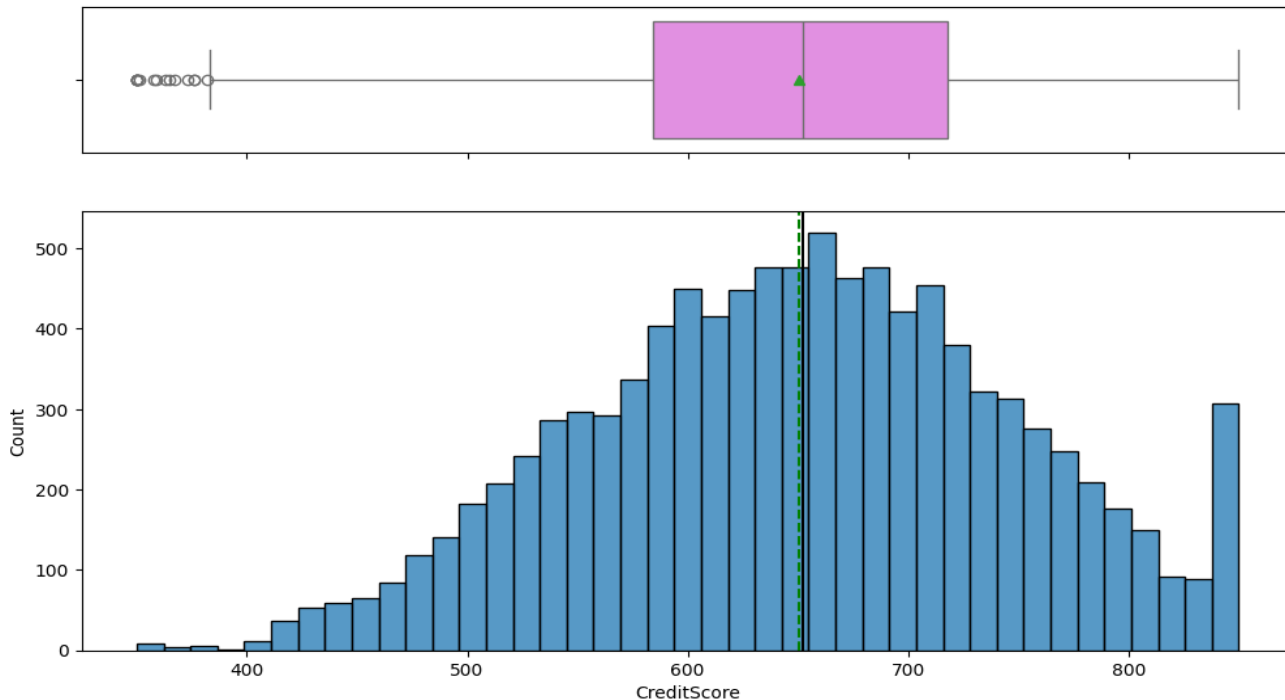
APPENDIX

Data Background and Contents

- The data set contains 10,000 samples points that started with 14 distinct features. After removing columns with unique values, the following remain:
- CreditScore: It defines the credit history of the customer.
- Geography: A customer's location
- Gender: It defines the Gender of the customer
- Age: Age of the customer
- Tenure: Number of years for which the customer has been with the bank
- NumOfProducts: refers to the number of products that a customer has purchased through the bank.
- Balance: Account balance
- HasCrCard: It is a categorical variable which decides whether the customer has credit card or not.
- EstimatedSalary: Estimated salary
- IsActiveMember: Is Estimated Salary a categorical variable which decides whether the customer is active member of the bank or not (Active member in the sense, using bank products regularly, making transactions etc.)
- Exited : whether the customer left the bank within six month.

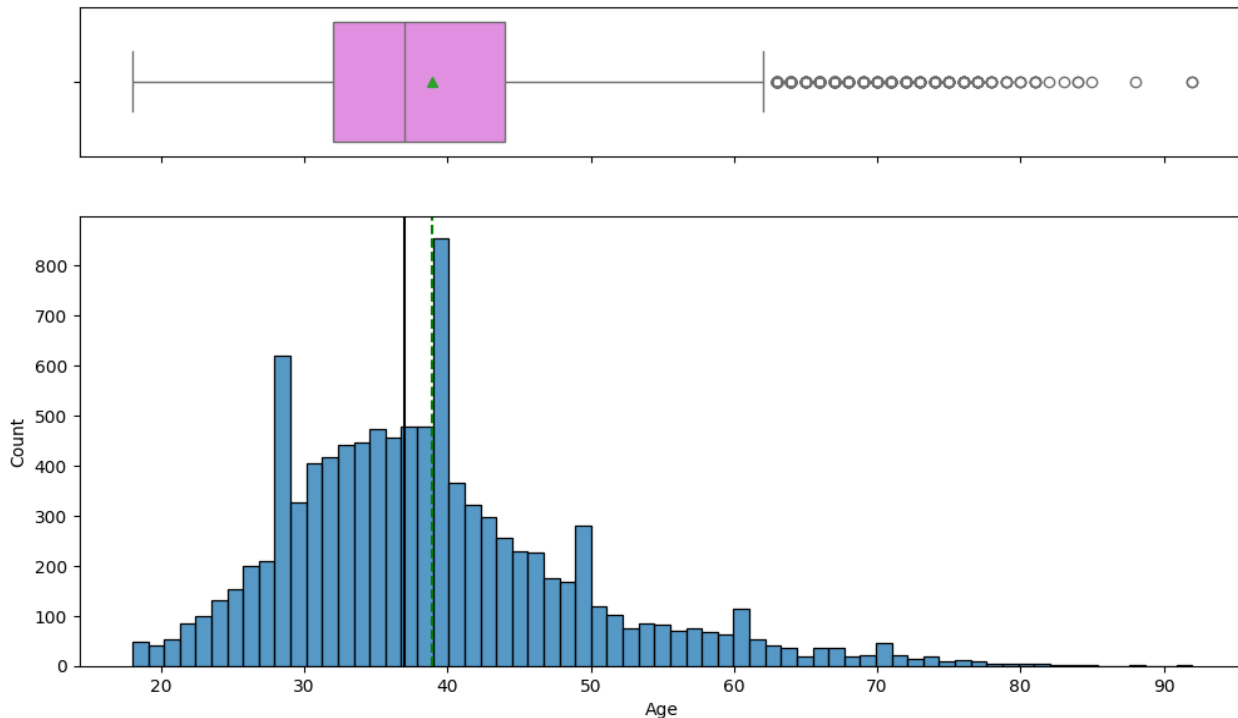
Credit Score

- Credit scores are slightly left skewed.
- There are some outliers for customers that have very low credit scores.



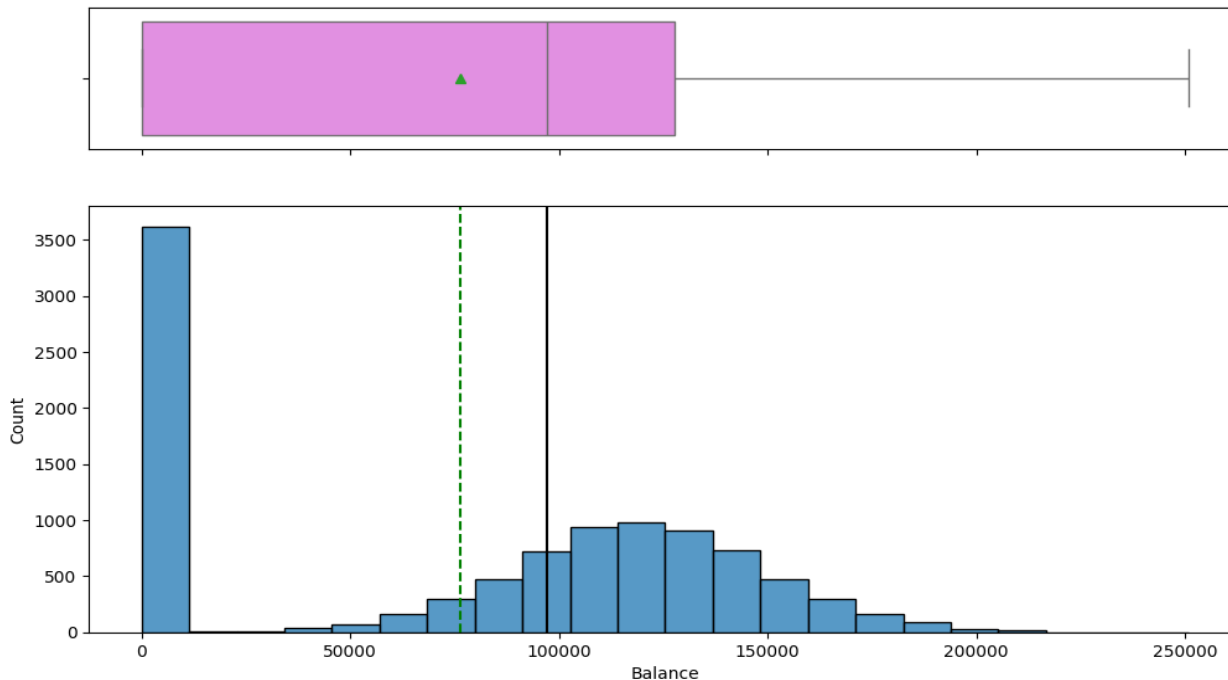
Age

- Customer ages are right skewed.
- Age has outliers that seem logical as the age range for bank customers can go late in life.



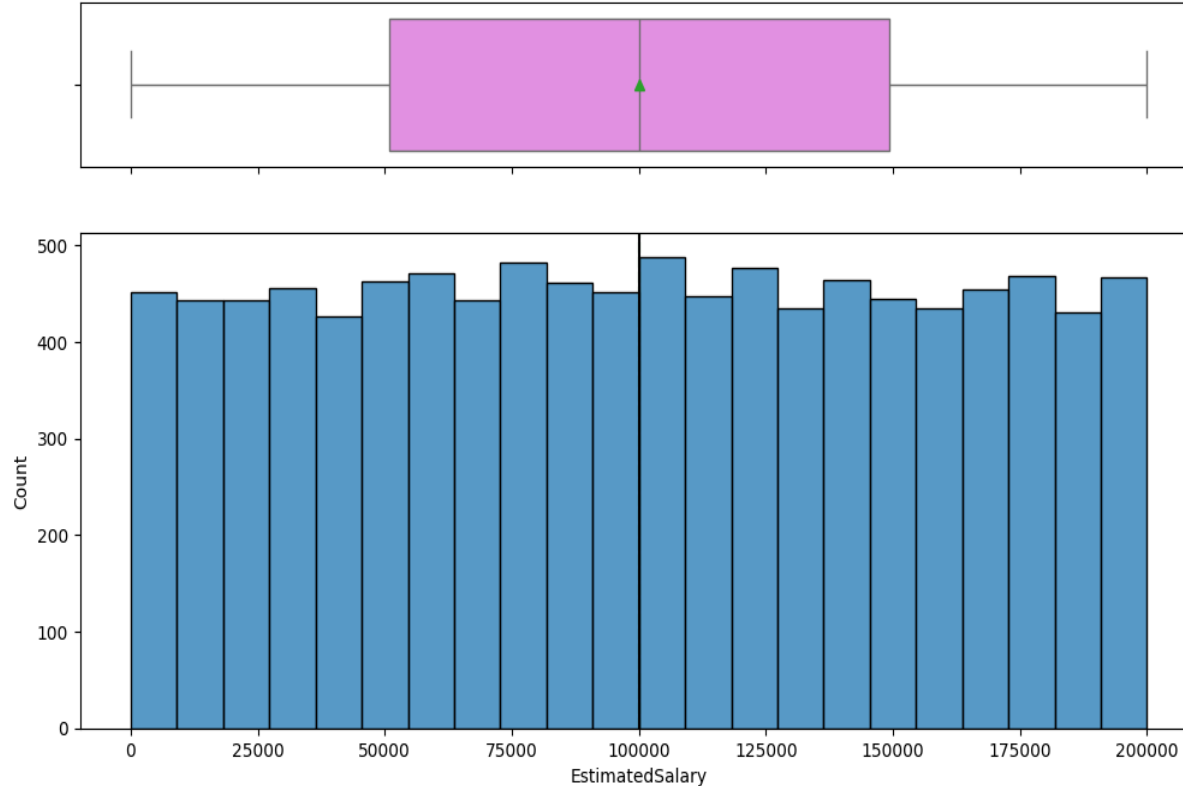
Balance

- A third of customers have a very low balance.
- The data is left skewed.



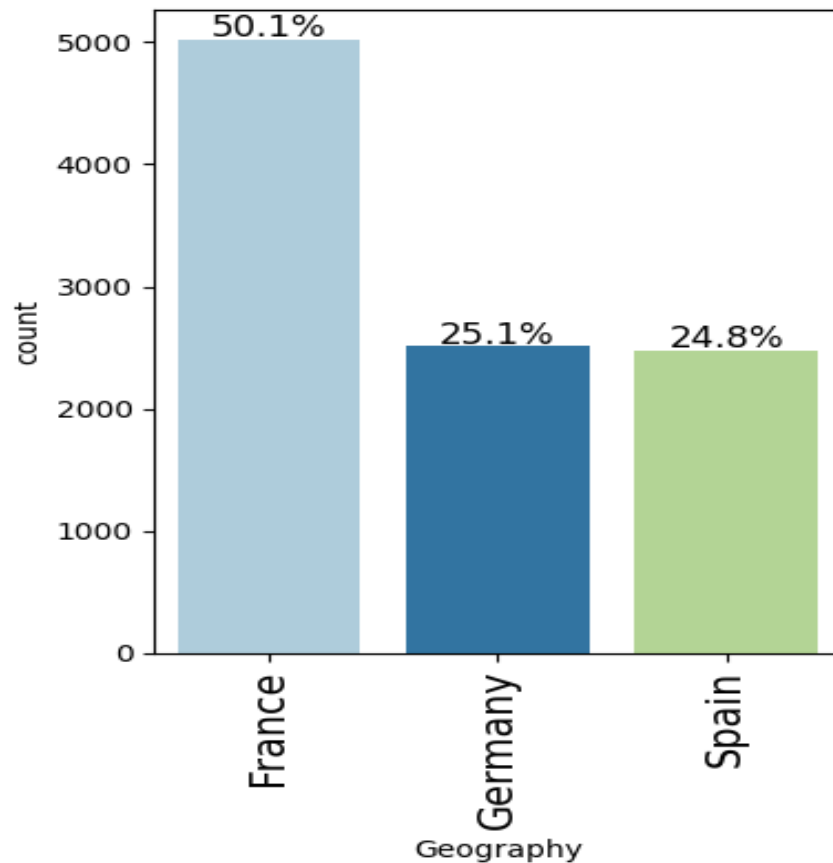
Estimated Salary

- Estimated salary is evenly distributed.



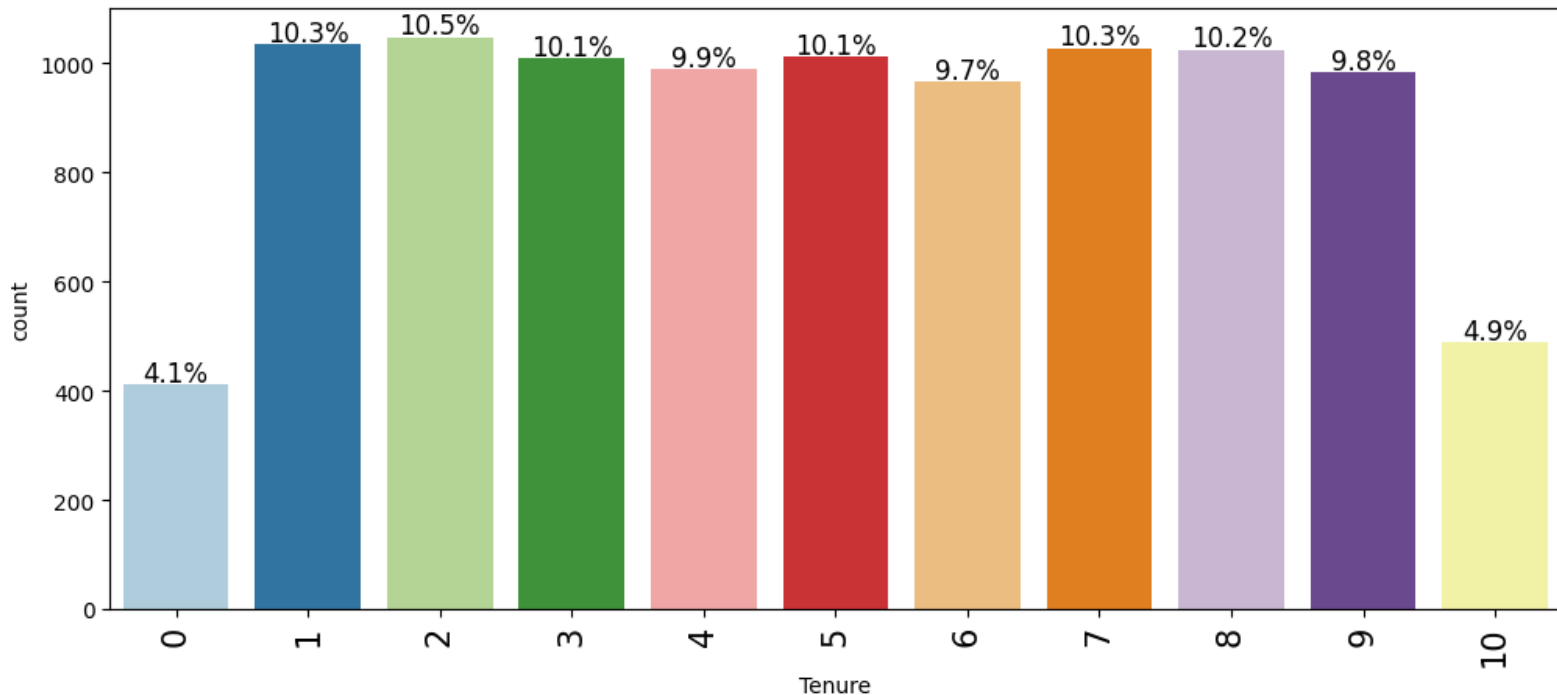
Geography

- 50% of customers are in France.
- 25% of customers are in Germany
- 25% of customers are in Spain.



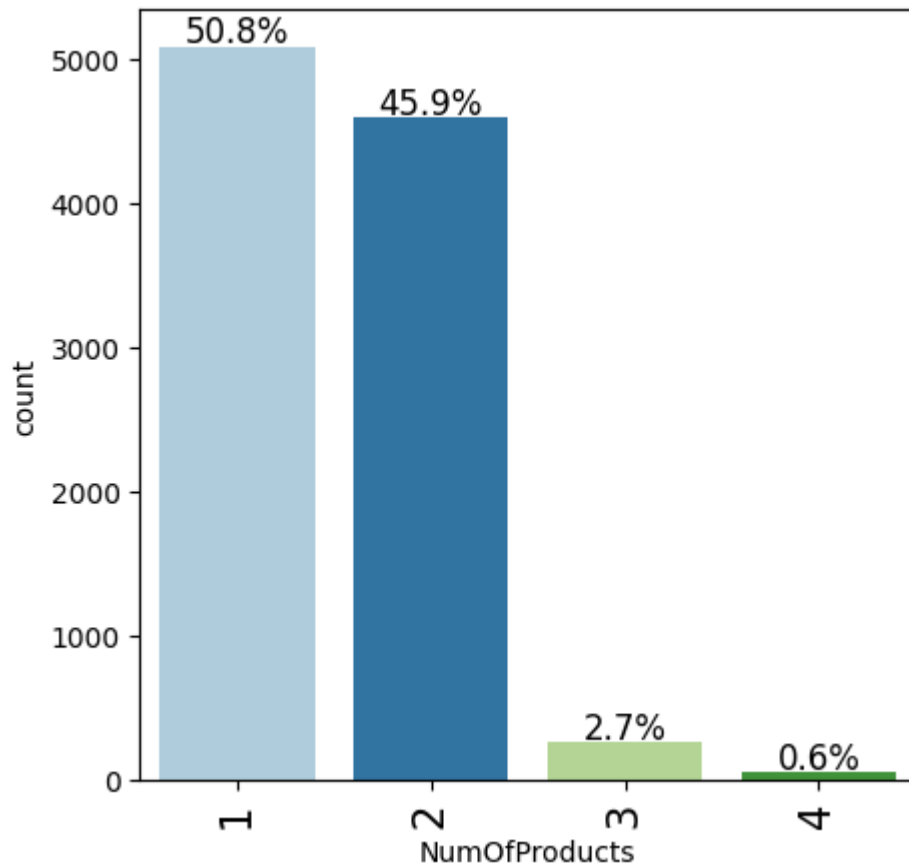
Tenure

- Tenure is evenly distributed among bank customers with about 10% of customers having tenure for each year.



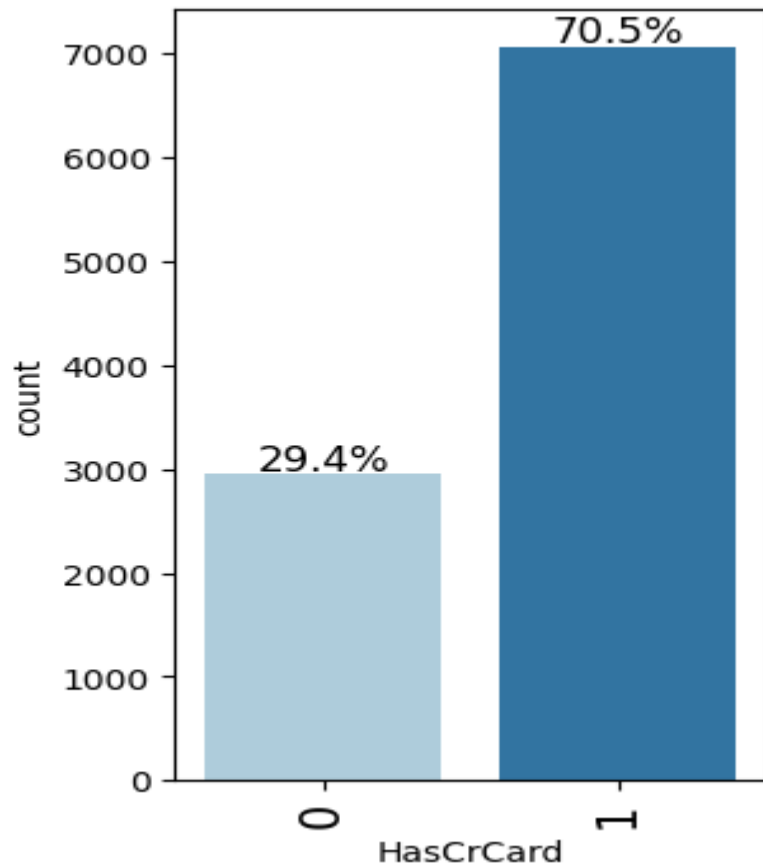
Number of Products

- Only a small number of customers have more than 2 bank products.
- Half of customers have only 1 bank product.



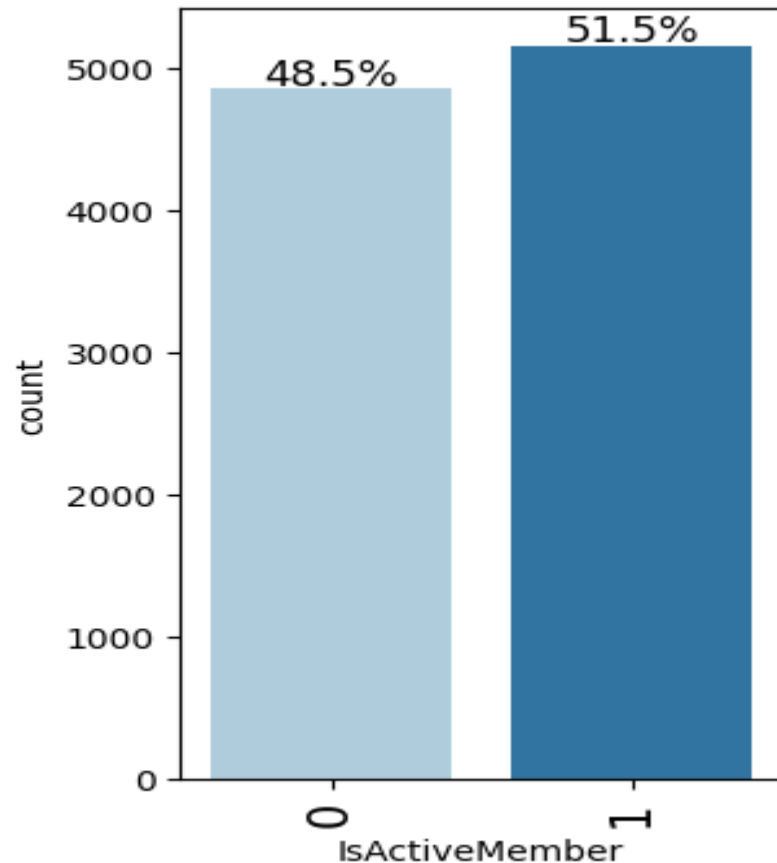
Has Credit Card

- 70% of customers have a Credit Card.



Active Members

- 51% of customers are active members.
- Active members regularly use bank products and make transactions.



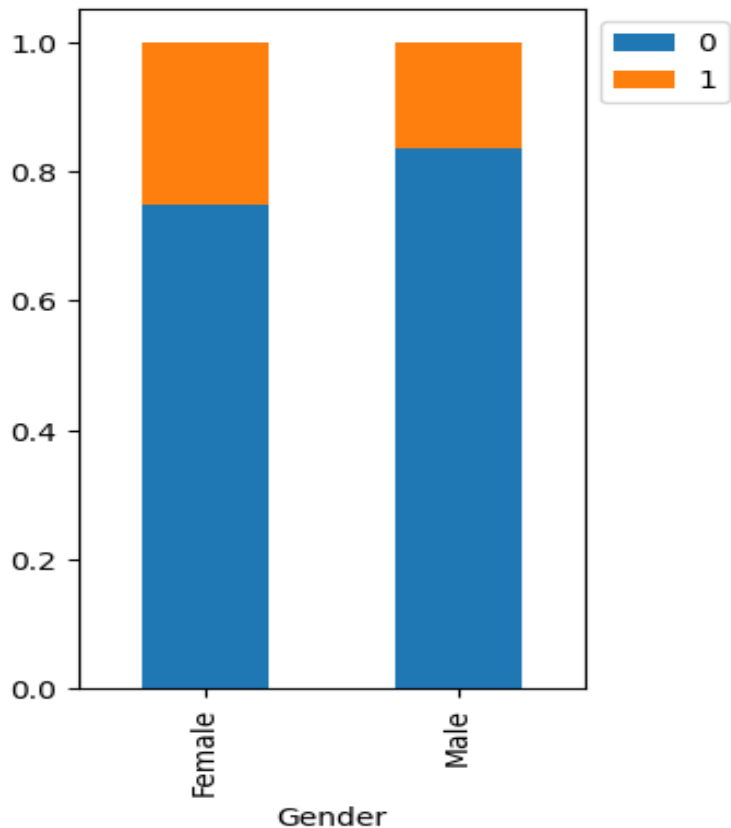
Correlations

- There are no strong correlations in the data set.
- The largest correlations is between IsActiveMember and Age.



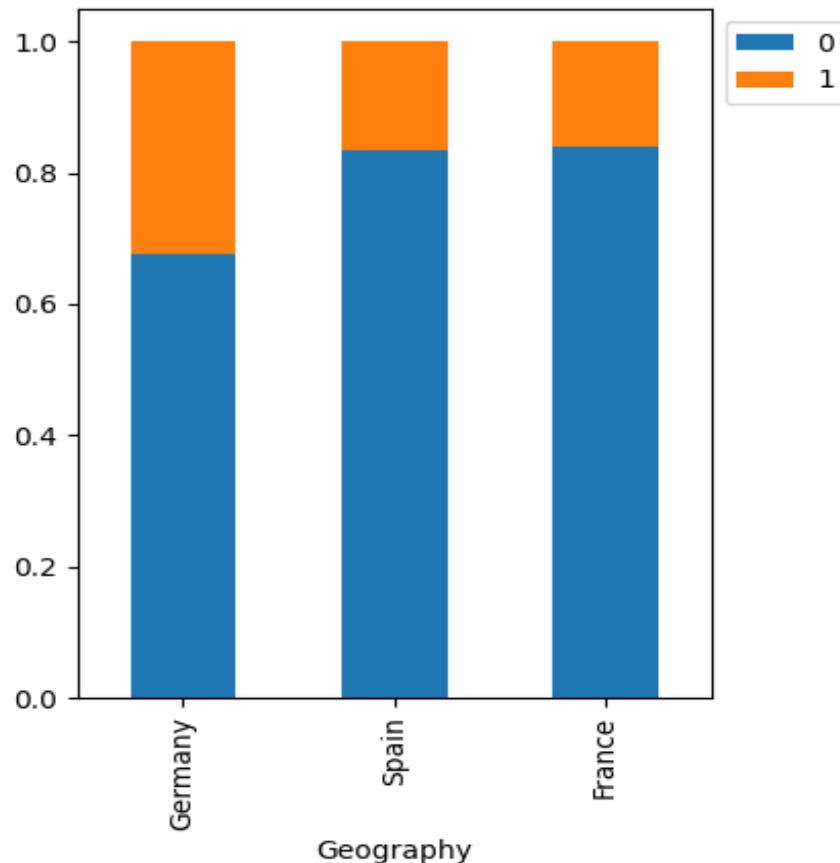
Exited vs Gender

- Females are slightly more likely to have exited the bank.



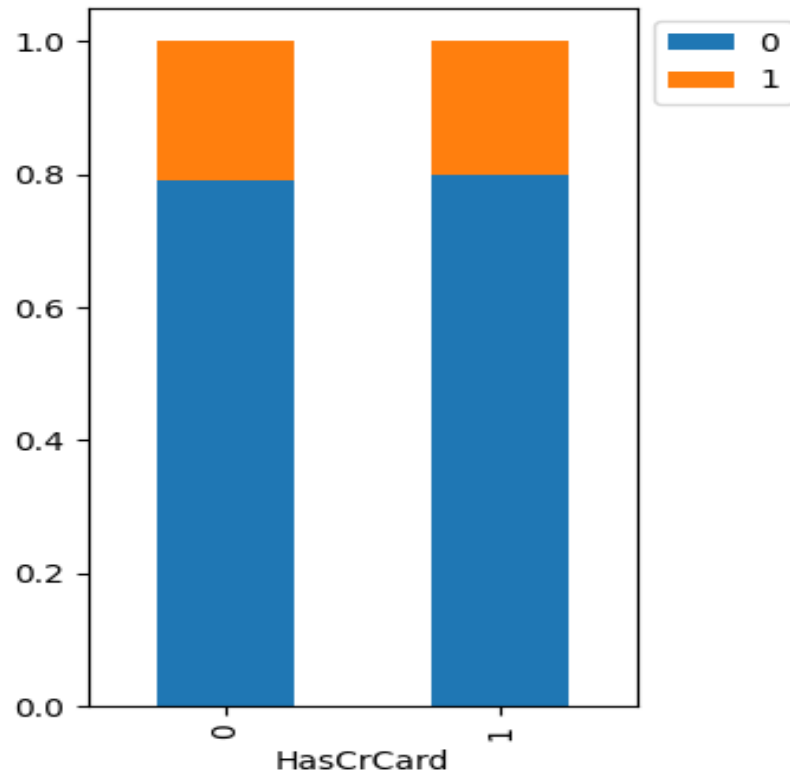
Exited vs Geography

- More customers from Germany have exited the bank. Germany also makes up a majority of bank customers.
- Spain and France are equal in the number of customers have left the bank.



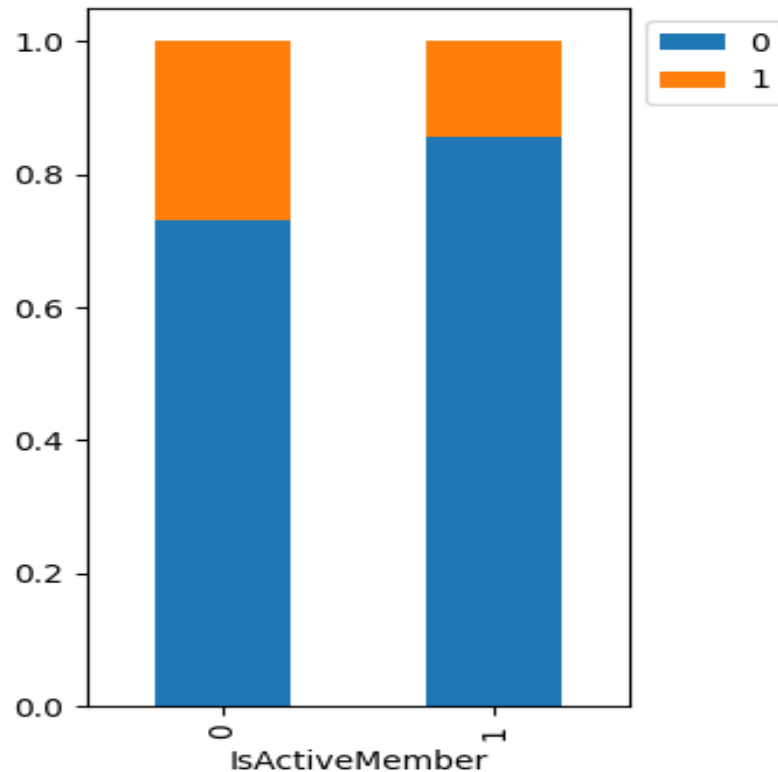
Exited vs Has Credit Card

- The same number of customers that have exited have credit cards as those that are still at the bank.

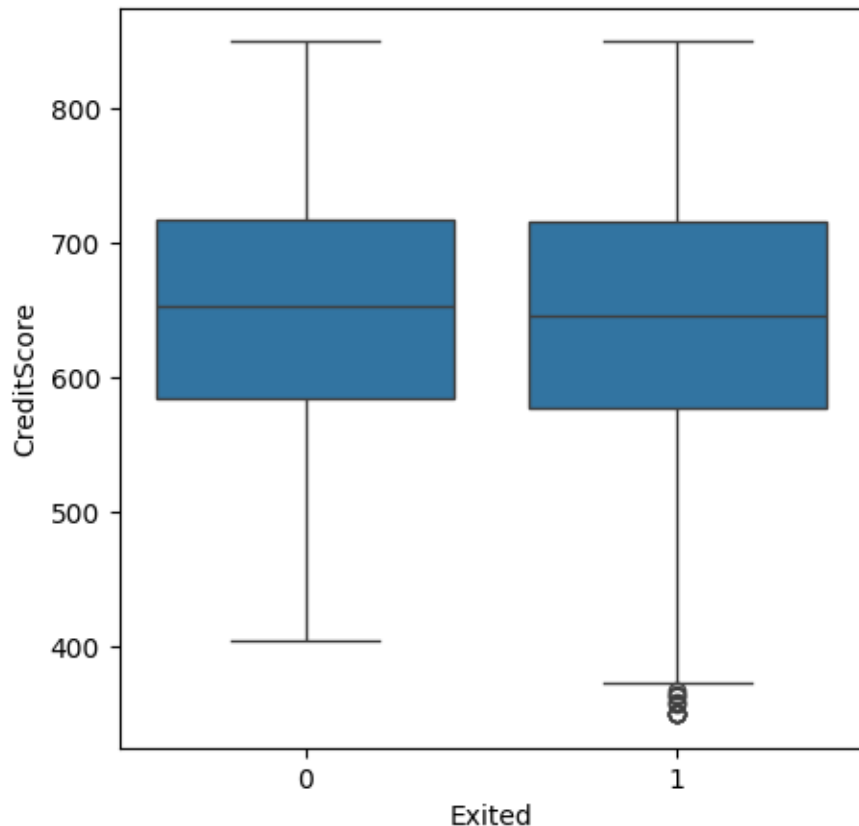


Exited vs Active Member

- Inactivity is higher among customers that have exited the bank.

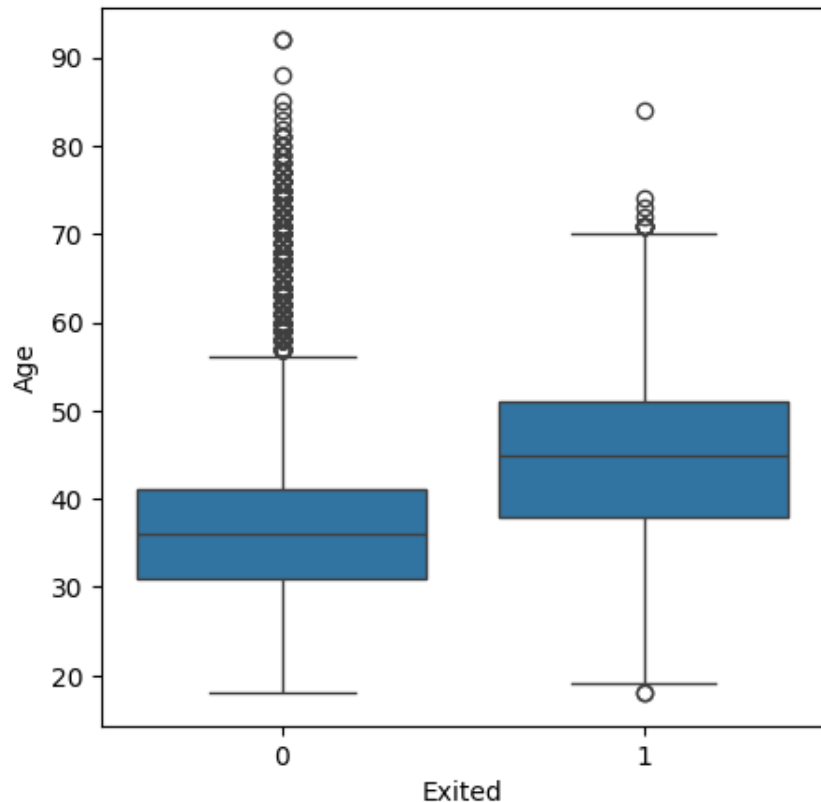


Exited vs Credit Score



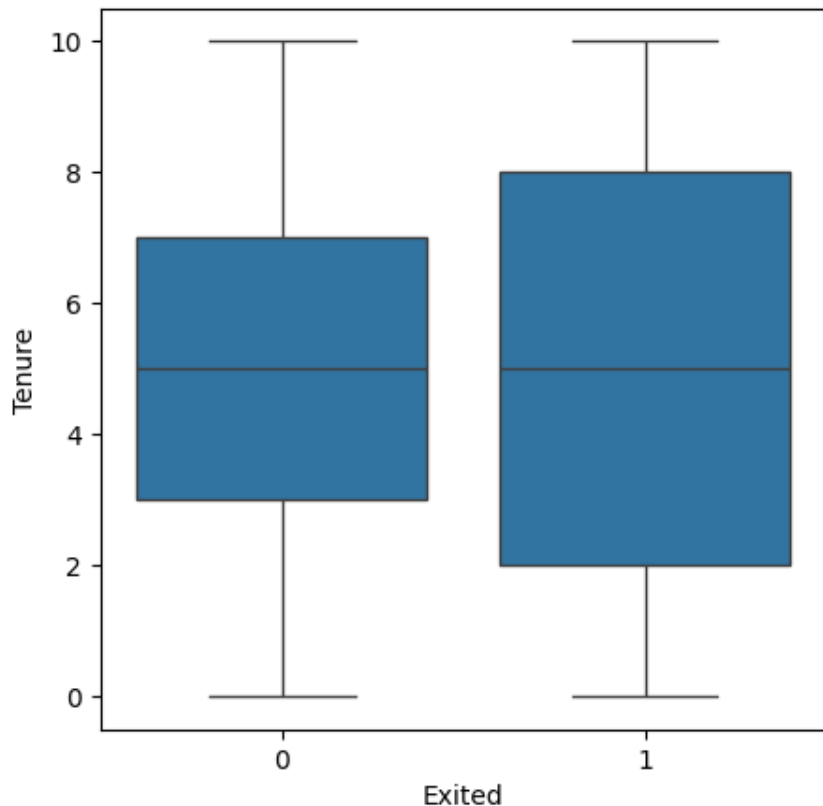
- There is very little difference between the credit scores of customers that have exited the bank and customers that remain with the bank.

Exited vs Age



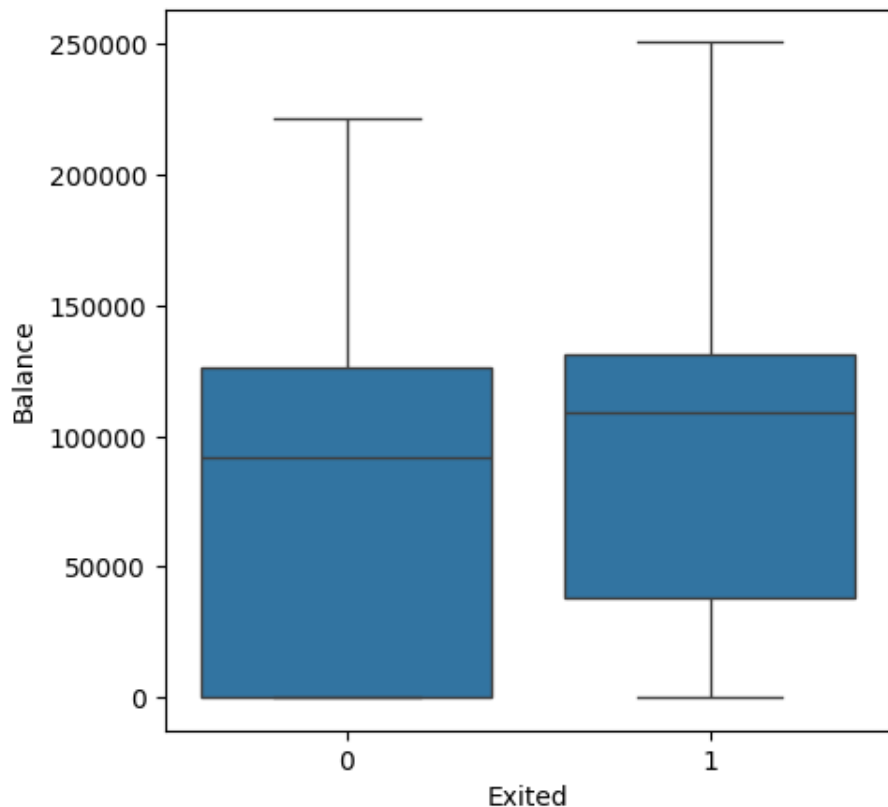
- The Age of customers that have exited the bank are somewhat higher.
- The average Age of customers that have exited the bank is 45.
- The average Age of customers that remain at the bank is 36.
- The majority of outliers among Age are customers that remain with the bank, indicating that the bank has many older customers.

Exited vs Tenure



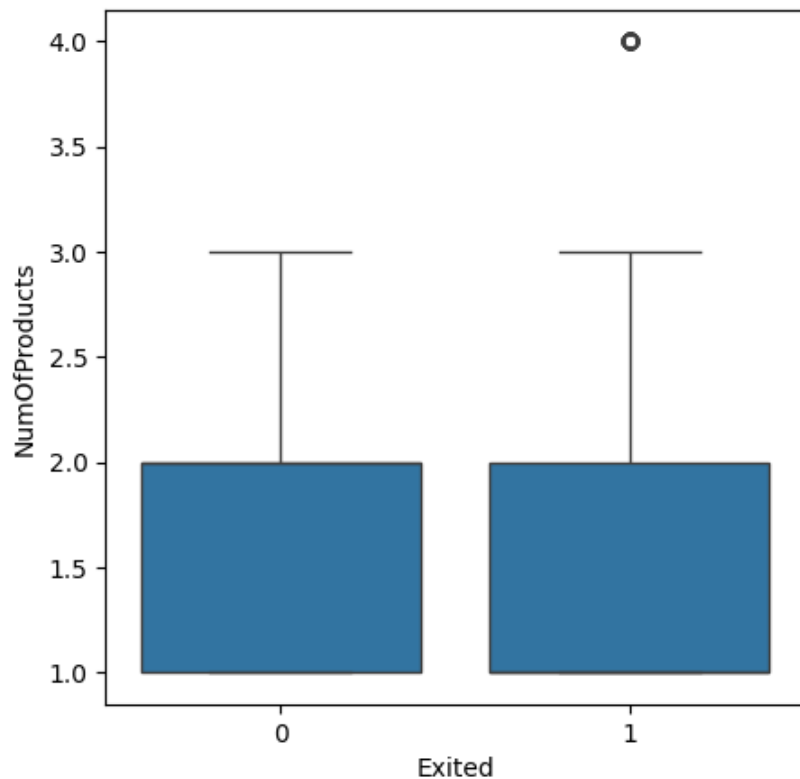
- Customers that exit the bank have a median Tenure of 5 years.
- Customers that remain at the bank also have a median of 5 years, however the range of the Tenure is larger with more customers having both a smaller and larger range.

Exited vs Balance



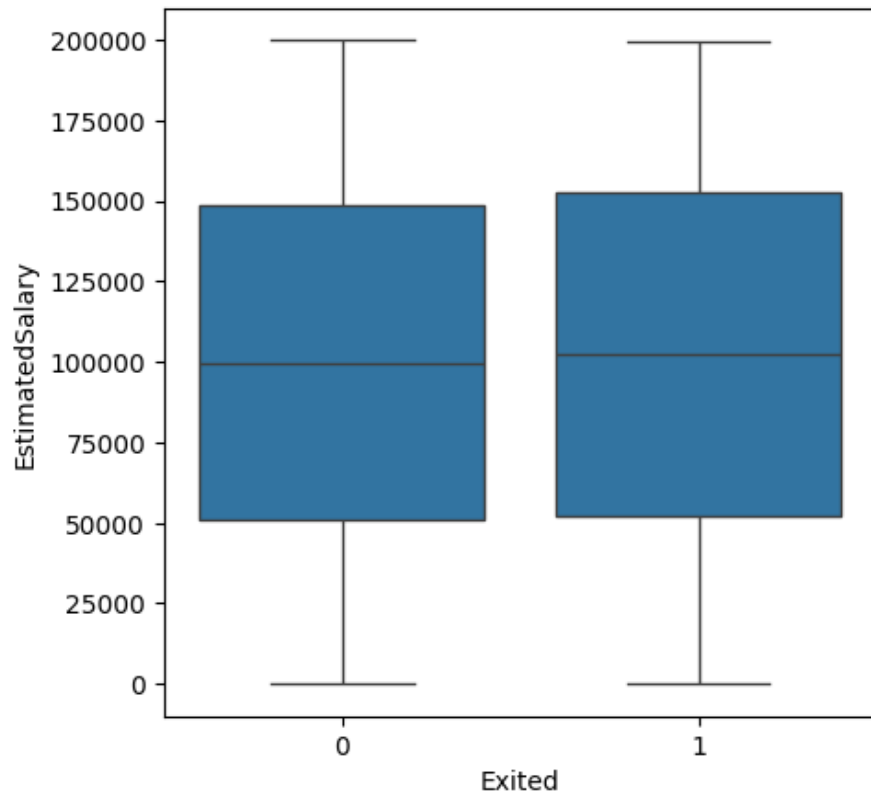
- Customers that exit the bank have lower balances.
- Customers that exit the bank have median balance of 90,000.
- Most of the balances for customers that remain at the bank are 40,000 and higher.
- The median balance for customers that remain at the bank is 110,000.

Exited vs Number of Products



- Customers that have exited the bank and customers that remain both primarily have 1 or 2 products from the bank.

Exited vs Estimated Salary



- The Estimated Salary for customers that have exited the bank is nearly identical to customers that remain at the bank.