Bank Churn Prediction

Project 4, AI with Deep Learning

By: Carter Fields

April 26, 2024

Contents / Agenda

Executive Summary

- A large amount of exits come from customers with zero balances. Focusing on predicting those customer would be worth the investment to reduce some of that loss.

- Investing in ways to make your customers more active will also reduce churn.

- Clarification or corrections about some of the data is required for further analysis. Specifically around the German customers having no zero balances and every customer with four products exiting. Proper treatment of that data would likely result in a better performing model.

- Because there is such a high percentage of customers with zero balances it would likely be beneficial to separate the data and create two different models. One for customers with a positive balance and customers with a zero balance. The models would likely produce a higher degree of accuracy.

Business Problem Overview and Solution Approach

- **Background and Context**

Businesses like banks that provide service have to worry about the problem of 'Churn' i.e. customers leaving and joining another service provider. It is important to understand which aspects of the service influence a customer's decision in this regard. Management can concentrate efforts on the improvement of service, keeping in mind these priorities.

- **Objective**

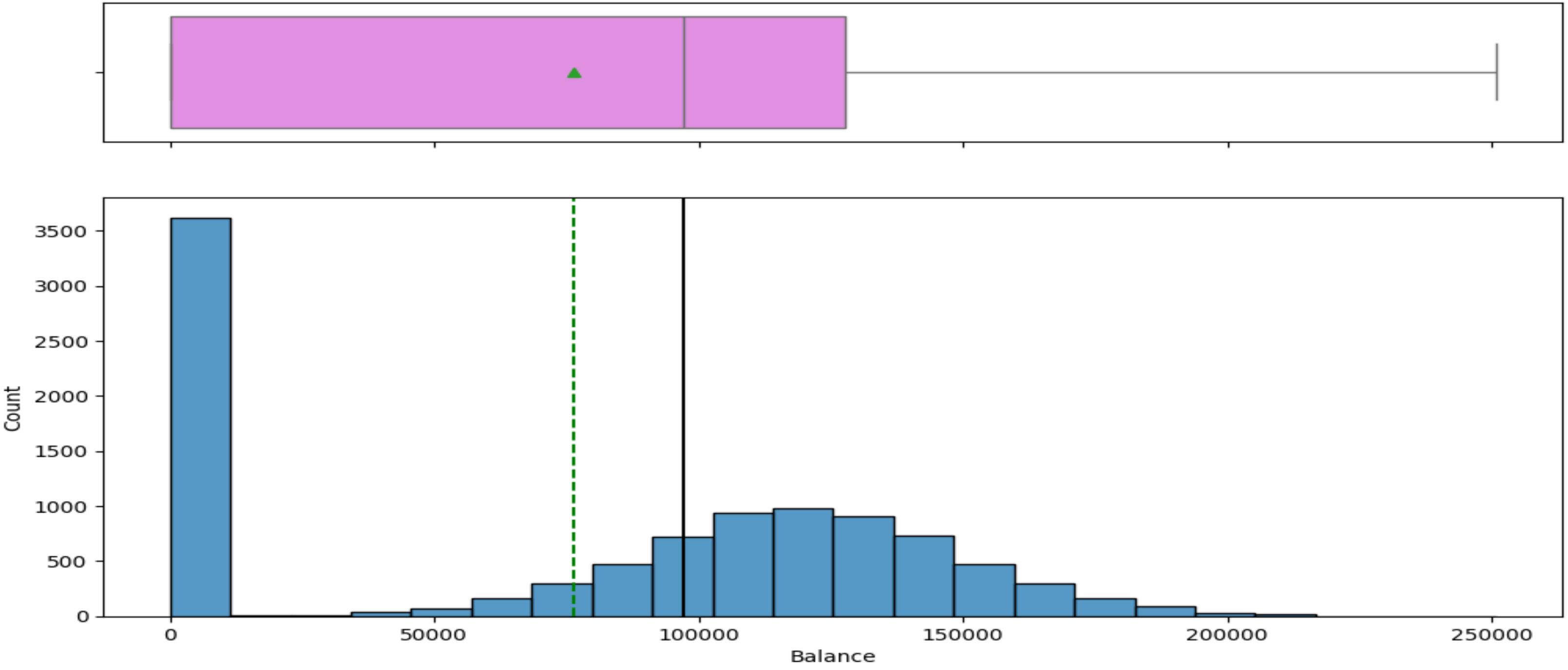
Given a Bank customer, build a neural network-based classifier that can determine whether they will leave or not in the next 6 months.

- **Solution Approach / Methodology**

I will start by doing a deep dive of the data to fully understand all the variables and correlations. I will process the data as needed. Then I will make several models, from simple to complex. Evaluate the results of the models and choose the model with the best results. I will then tune that model to get the best results possible.

EDA Results

- There is a high number of customers that have a zero balance.

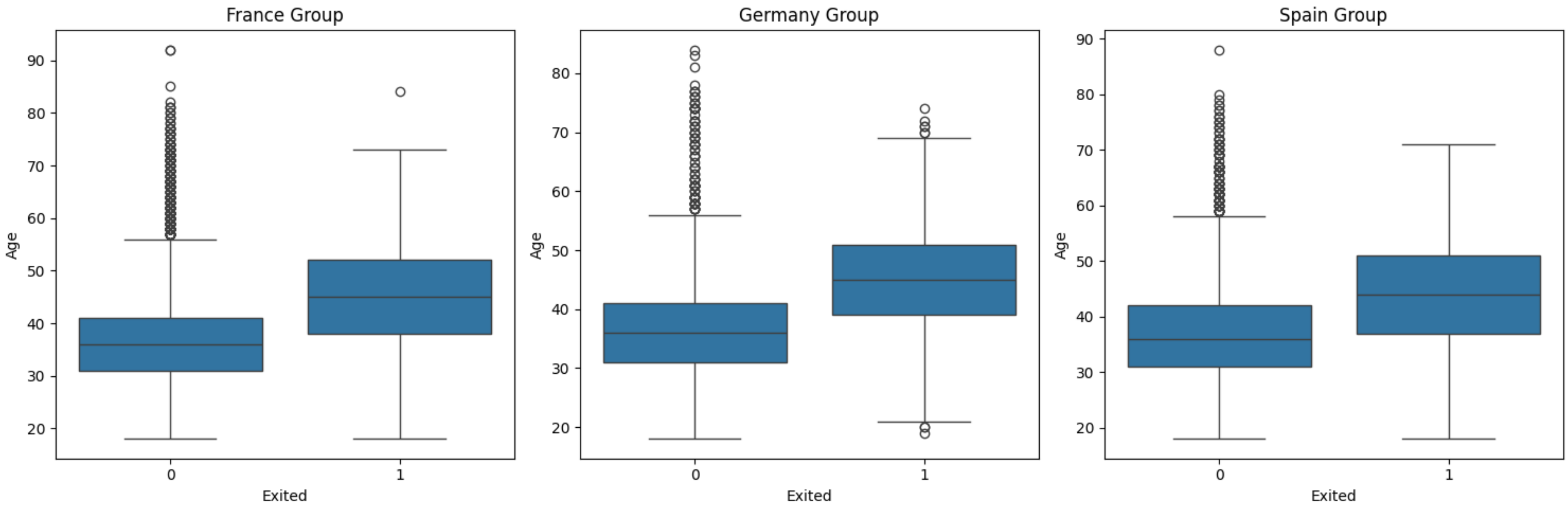


EDA Results

- Because there was a high number of customers with zero balance in their account, I spent a lot of time looking into all the colorations with zero balance, exited and other variables.
- 25% of all exited have a zero balance
- 14% of all exited have a zero balance and only one product
- 17% of exited from France had a zero balance
- 8% of exited from Spain had a zero balance

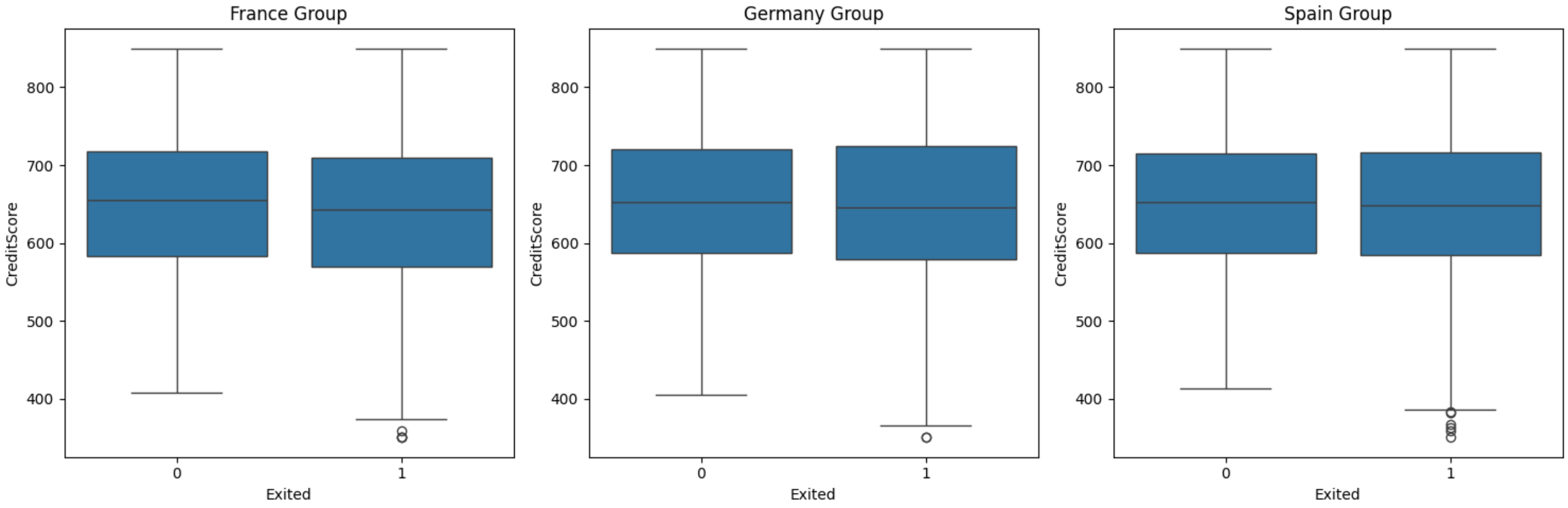
EDA Results

- The average age of customers that exited is higher than that of customers that did not exit.



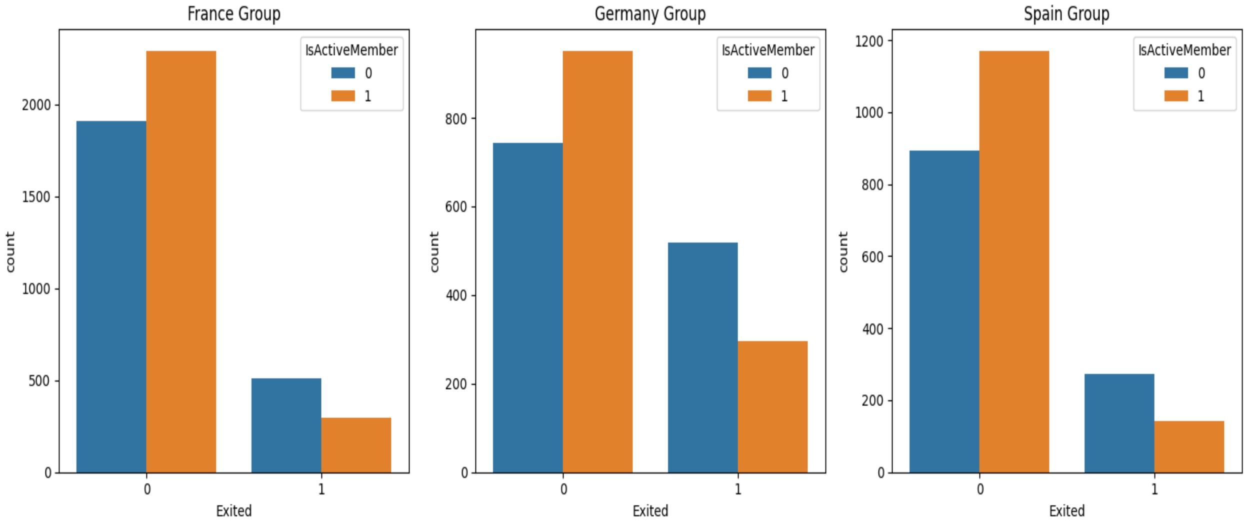
EDA Results

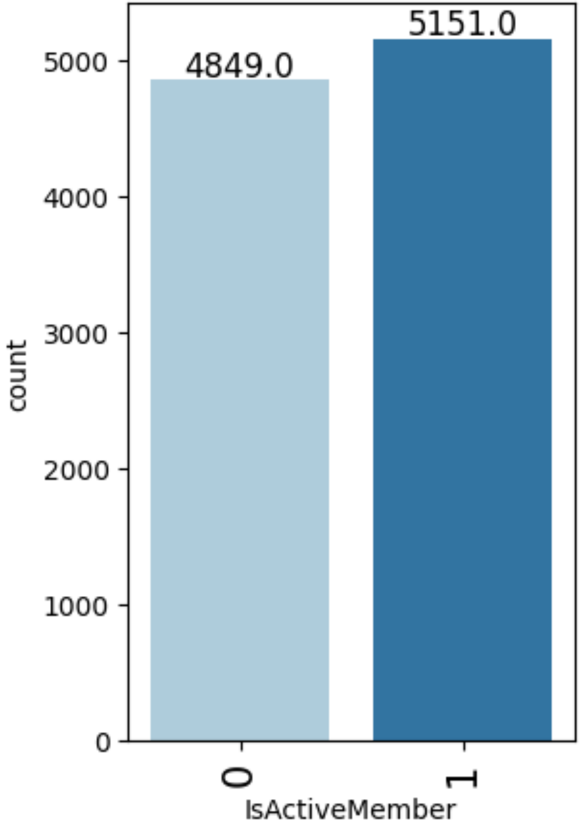
- Customers with very low credit scored were more likely to exit.



EDA Results

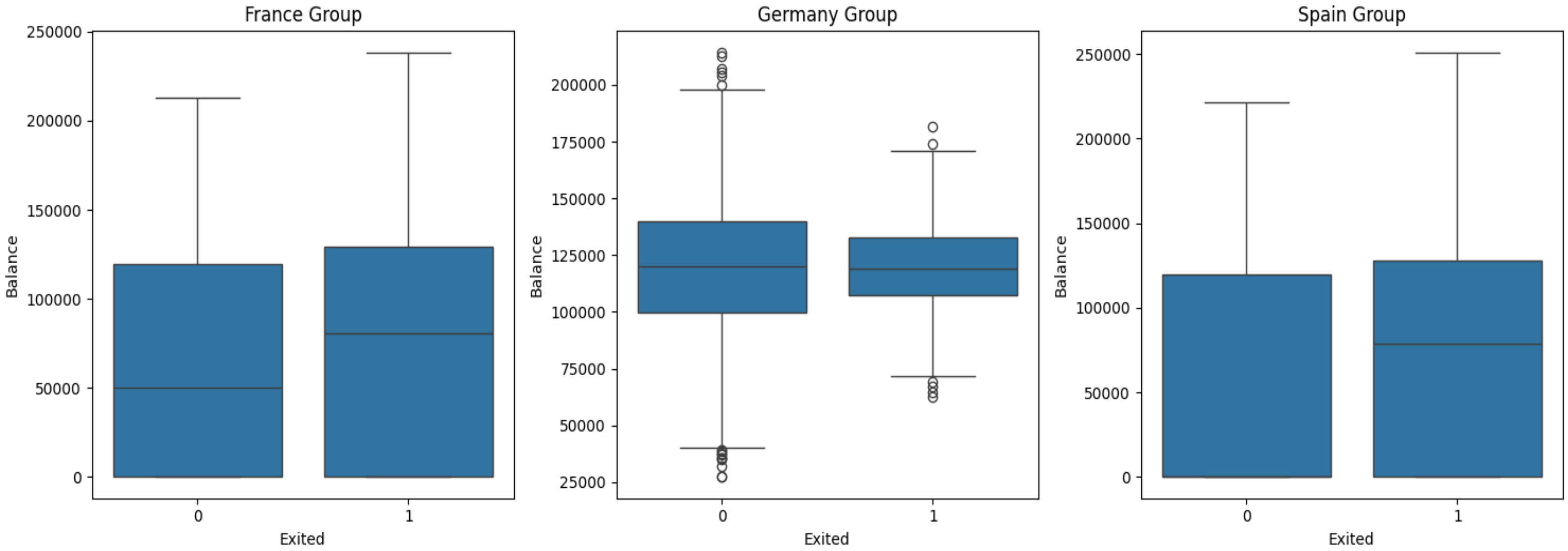
- Active members make up a little over 51% of customers.
- Inactive members are more likely to exit.





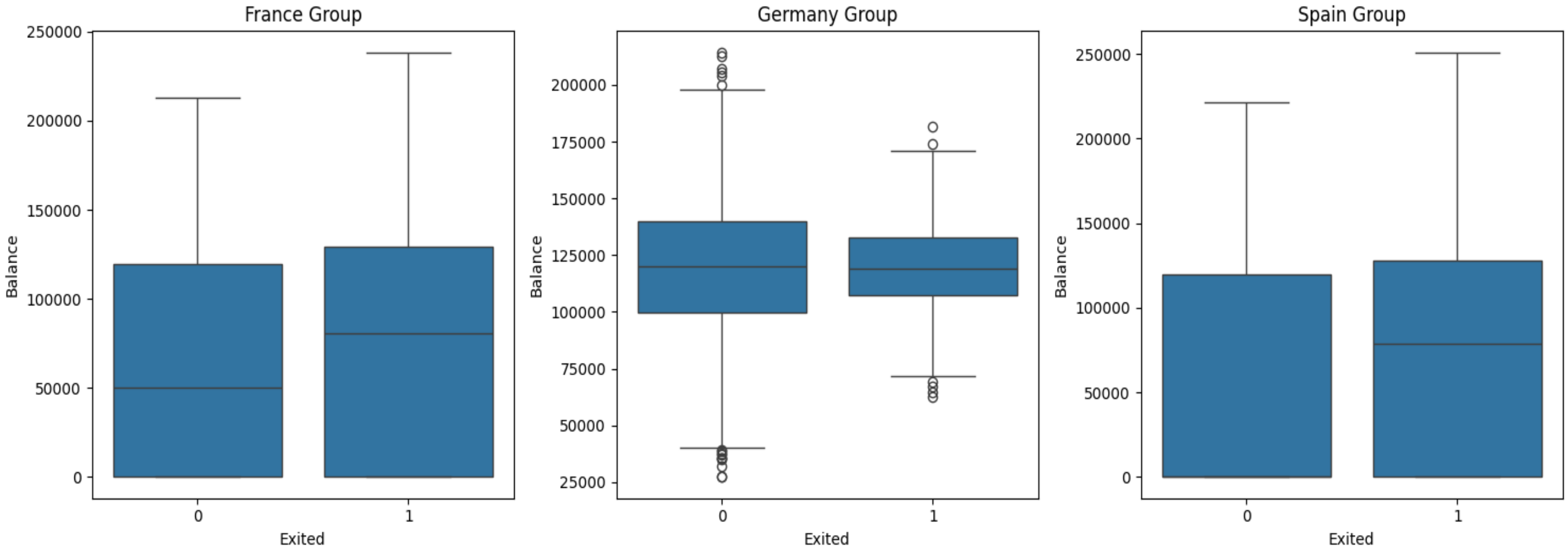
EDA Results

- German customers have a higher mean and median balance.
- No German customers have a zero balance.



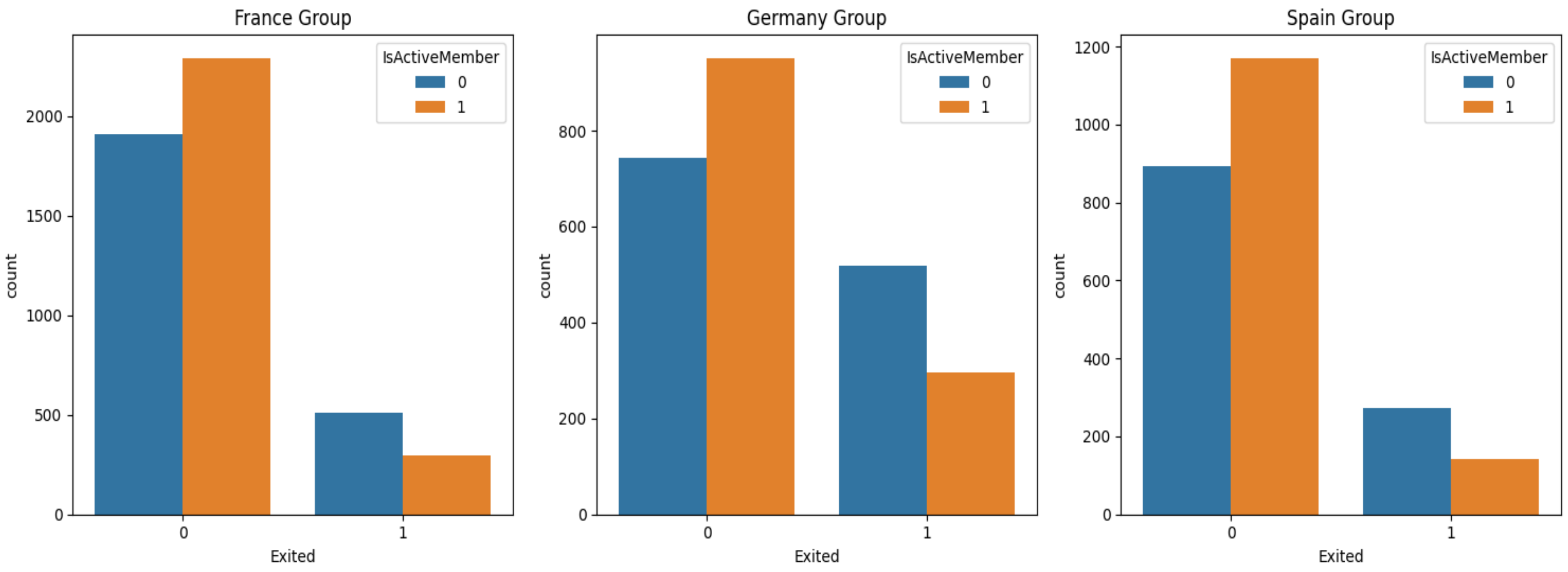
EDA Results

- 40% of all exited are from France.
- 40% of all exited are from Germany.
- 20% of all exited are from Spain



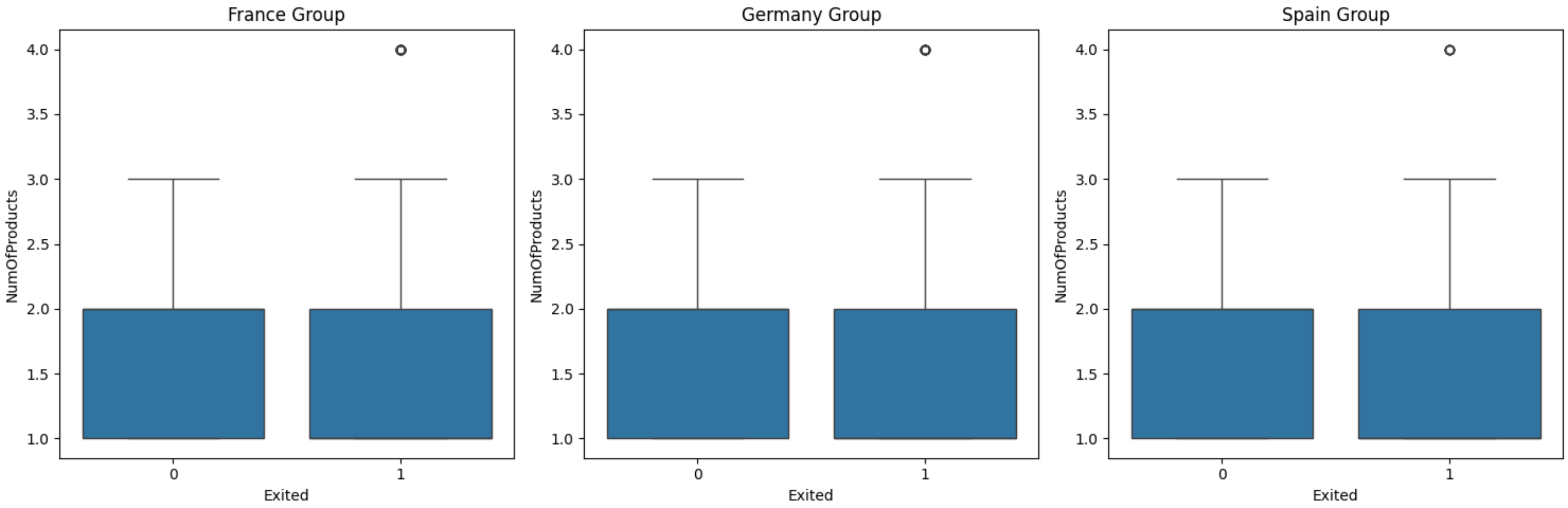
EDA Results

- Customers that are not active members are more likely to exit.



EDA Results

- All customers have roughly the same average number of products.
- Every customer with 4 products exited.



Data Preprocessing

- There are no missing values in the dataset.
- There are few outlier’s and no treatment was done.
- RowNumber, CustomerID, and Surname columns were dropped.
- The dataset was split into train, val, test datasets using 90:10 for test and 70:30 for train and val.
- Dummy variables were created for Geography then converted to Boolean columns.
- Then the data was normalized by standard scaler.

Model Performance Summary

- I tested the following six different models:
- Neural Network with SGD Optimizer
- Neural Network with Adam Optimizer
- Neural Network with Balanced Data (SMOTE) and SGD Optimizer
- Neural Network with Balanced Data (SMOTE) and Adam Optimizer
- Neural Network with Balanced Data (SMOTE), Adam Optimizer and Dropout

- The model that performed the best was the Neural Network with Balanced Data (SMOTE), Adam and Dropout.

Model Performance Summary

Training Performance Comparison	Recall
NN with SGD	0.169914
NN with Adam	0.593920
NN with Adam & Dropout	0.546376
NN with SMOTE & SGD	0.986845
NN with SMOTE & Adam	0.992426
NN with SMOTE, Adam and Dropout	0.906119

Model Performance Summary

Validation Comparison	Recall
NN with SGD	0.134545
NN with Adam	0.429091
NN with Adam & Dropout	0.465455
NN with SMOTE & SGD	0.565455
NN with SMOTE & Adam	0.554545
NN with SMOTE, Adam and Dropout	0.630909

APPENDIX

Data Background and Contents

- The dataset was spreadsheet of bank customer information, with 14 columns and 10,000 rows. The columns were, RowNumber, CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, Estimated Salary, Exited.

Great Learning

Happy Learning !