BANK CUSTOMERS CHURN PREDICTION

Anugerah Radifan Kholis

EXECUTIVE SUMMARY

Business Background

ABC BANK is a banking and financial services corporation with **10000** customers.

Problems Statements

For a given period of time, ABC BANK lost **20,4%** of its customers.

Objective

- To understand why customer churn
- To predict customer's likelihood to churn

Result:

Data Analysis Results:

- There are problems with **German** customers because of its high churn rate.
- Surprisingly, having **more than 2 products** increases the likelihood to churn significantly.
- **Non active** members are more likely to churn.
- **Female** customers have higher churn rate.
- **Short term** and **long term** tenure customer are more likely to churn compared to those that are of average tenure.

Machine Learning Results:

Using Logistic Regression, we achieve 71,2% Recall Score for our model.

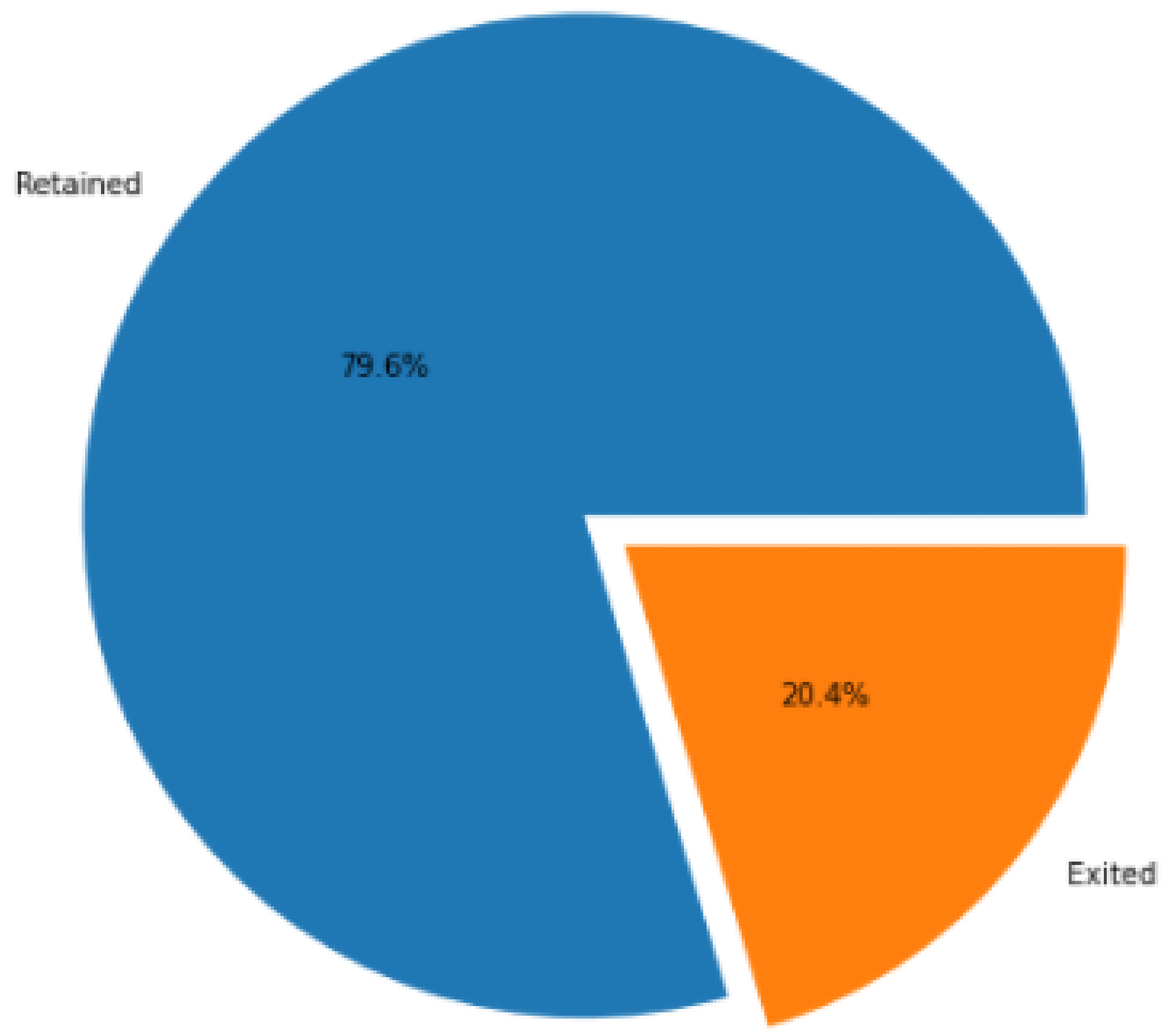
Proposed Solutions

- Use Machine Learning to predict customer's likelihood to churn.
- Apply recommendations based on our findings

Business Benefit

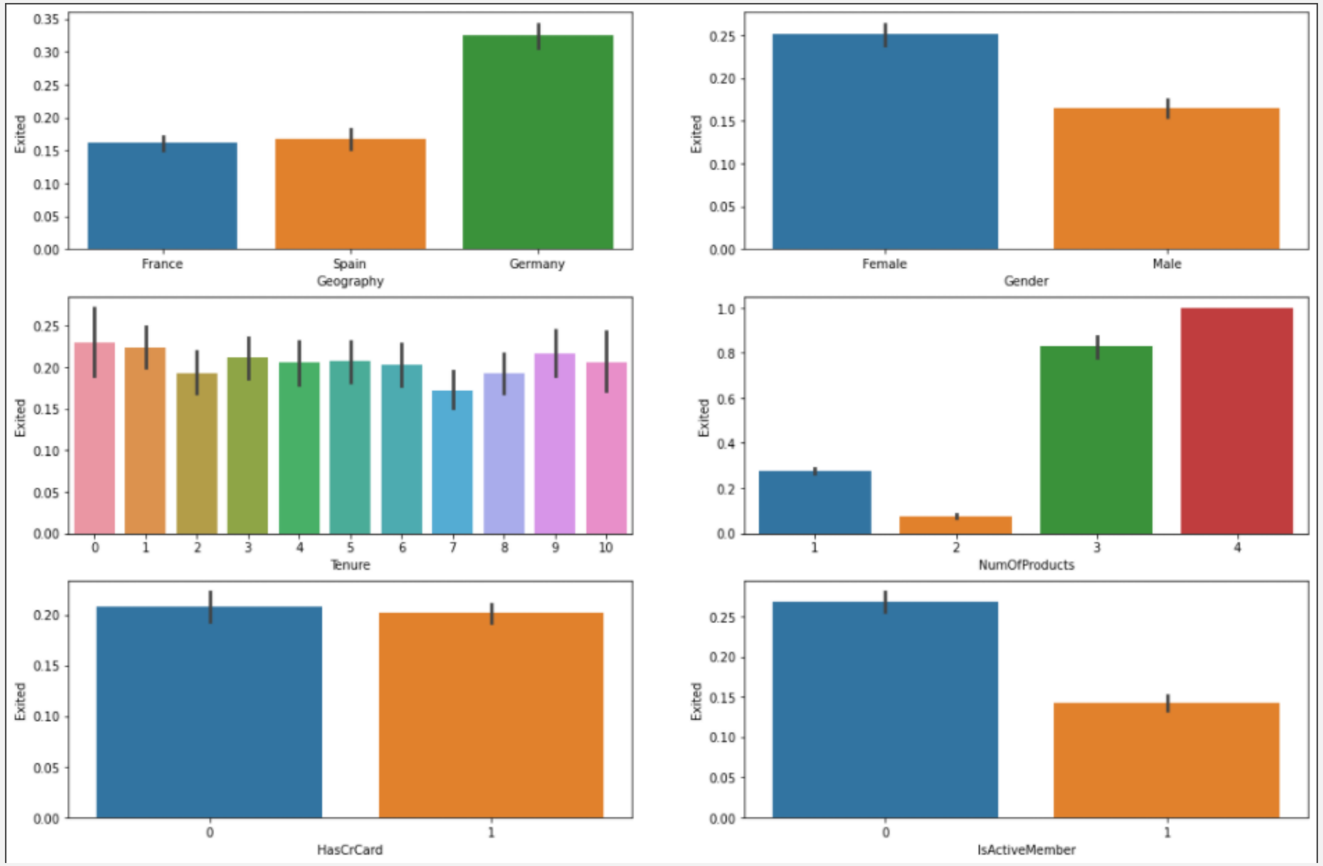
- Prevent future customer to churn by correctly classifying **306** out of **427** costumer who churned based on test data.
- Reducing the cost to attract new customer by preventing customer to churn

EXPLORATORY DATA ANALYSIS



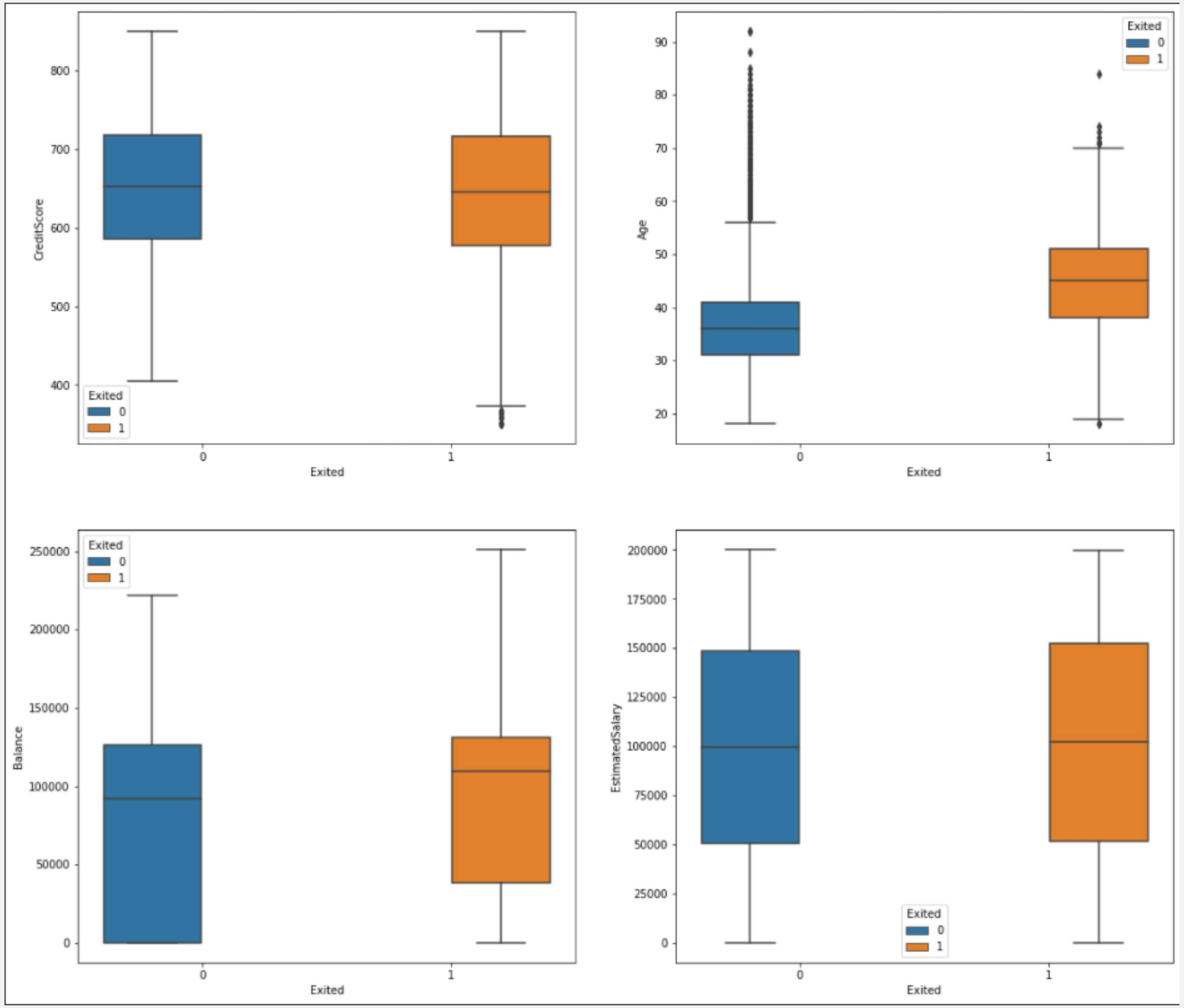
EXPLORATORY DATA ANALYSIS

- The majority of the customers is from France. However, the churning proportion is higher in Germany.
- The proportion of female customers churning is higher than male customers.
- Customers who are more likely to churn are those who are relatively new and long time customers
- Having more than 2 products increases the likelihood to churn significantly.
- The majority of customers who churned have credit card.
- Inactive members are more likely to churn.



EXPLORATORY DATA ANALYSIS

- Credit score has no significant difference for customer
- Older customer are more likely to churn
- Customers with higher bank balances are churning
- Salary has no significant difference for customer



FEATURE ENGINEERING

- Dropping RowNumber, CustomerId, and Surname since they are specific to each customer.
- Based on our EDA, we are dropping EstimatedSalary, and HasCrCard because they have no impact on the churn rate.
- Adding new feature that is likely to impact churn rate, Tenure/Age and CreditScore/Age.

MODEL FITTING

- Since there is imbalance in the dataset for the target, we need to balance the data and use sampling techniques. For this dataset, we are using SMOTE and RandomUnderSampler.
- For our classification algorithm, we chose Logistic Regression and K-Nearest Neighbors.
- We are using recall as our scoring metric because correctly classifying as many as possible customers who churned is more important for the bank.

MODEL EVALUATION

Before Feature Engineering

Logistic Regression:

- Recall: 0.707
- F1-Score: 0.503

KNN:

- Recall: 0.51
- F1-Score: 0.489

After Feature Engineering

Logistic Regression:

- Recall: 0.712
- F1-Score: 0.496

KNN:

- Recall: 0.564
- F1-Score: 0.52

ACTION AND RECOMMENDATION

- The bank might need to review or change their policy so that more customer will become active again.
- The bank might need to create a new program for new customer and long time customer.
- The bank need to create a new strategy for German customers.
- The bank need to create a new strategy to meet the requirement of older customer.
- Regarding Number of Products, the bank might need to improve customer support for customer with more products.

THANK YOU

Any Questions?