

CUSTOMER CHURN PREDICTION USING ANN

Presented by Team 12

Introduction



Banks are eager to retain as many active customers as possible. Naturally they are curious to know whether their client base needs are met or whether their clients plan to leave the company. If the bank suspects that their client would potentially lean toward another company, the bank can take measures to convince the client to stay (targeted marketing campaign, more personal attitude etc.).

Aim



Aim of this project is to find the most accurate and precise model to predict, which clients (test data) will stay and which are hesitant and might plan to leave the company. We are using dataset of bank clients (10000 rows) with attributes specified below

Dataset

Rownumber: Unique ID for every row

CustomerID: Unique ID for every client

Surname: Client's surname

CreditScore: Client's credit score

Geography: Country of client's origin

Gender: Client's gender

Age: Client's age

Tenure: Number of years for which the client has been with the bank

Balance: Client's balance on account

NumOfProducts: Number of client's products

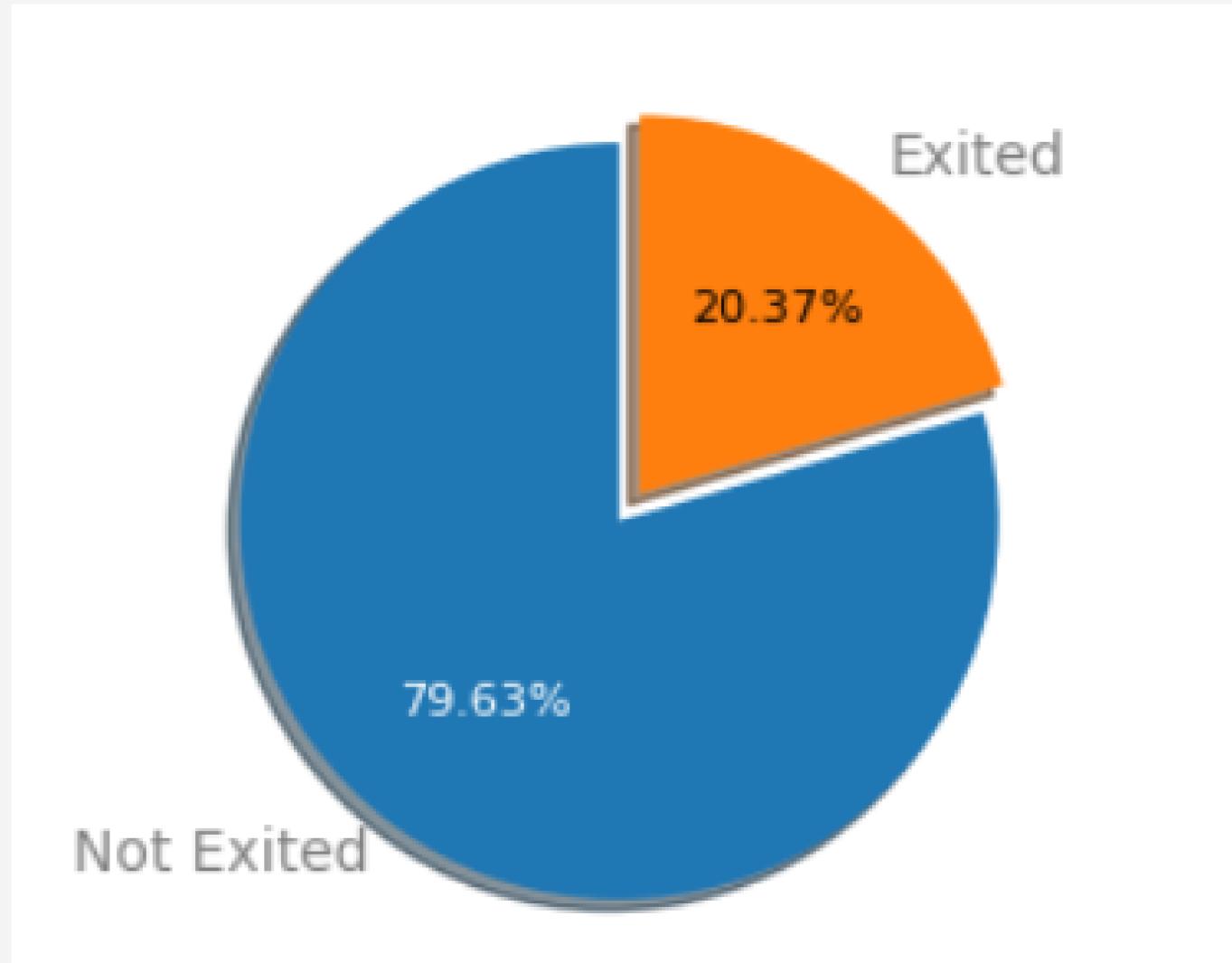
HasCrCard: Flag whether client has credit card or not

IsActiveMember: Flag whether client is active member of bank or not

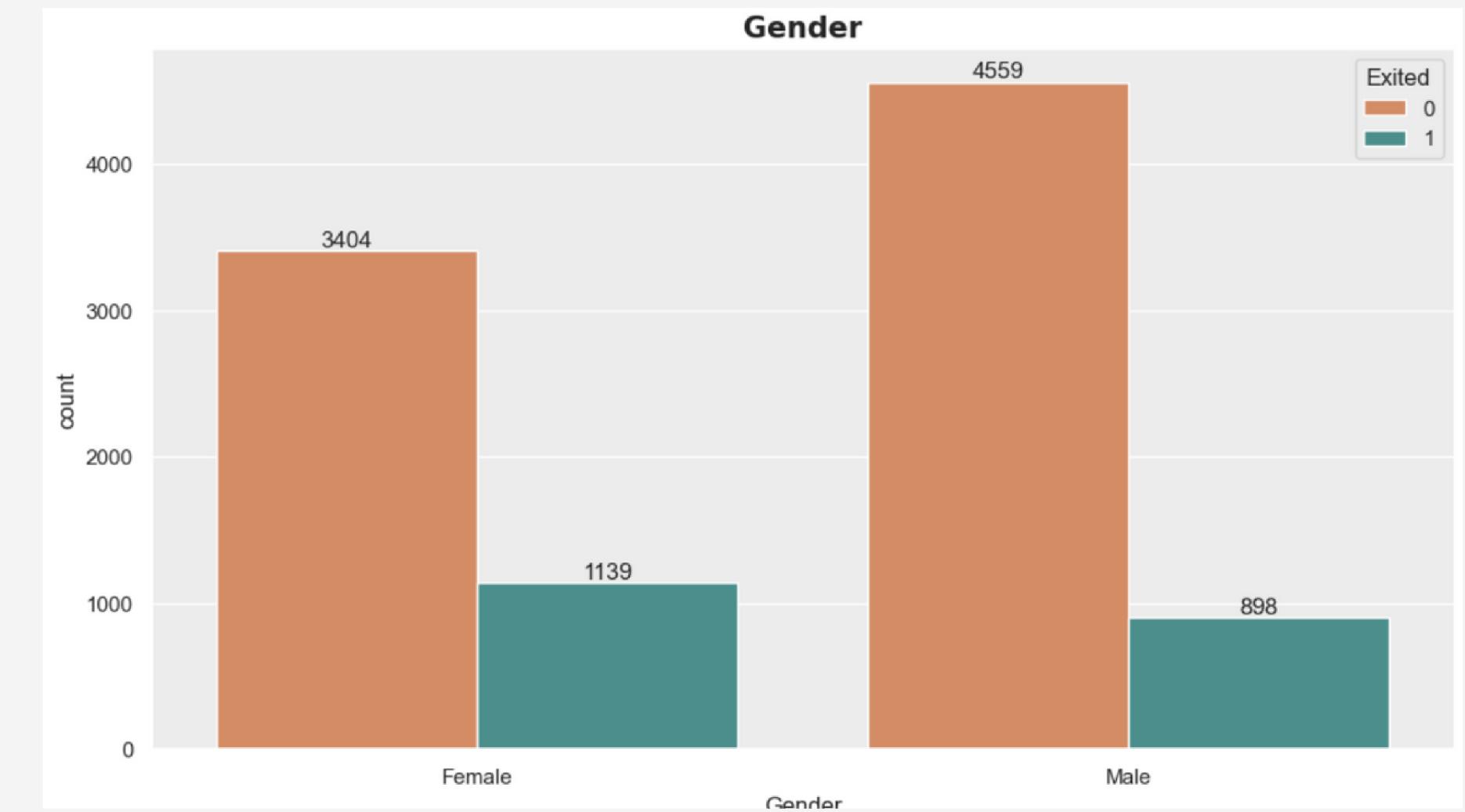
EstimatedSalary: Client's annual estimated salary in euros

Exited: Target variable, flag, whether client left the bank or not

EDA

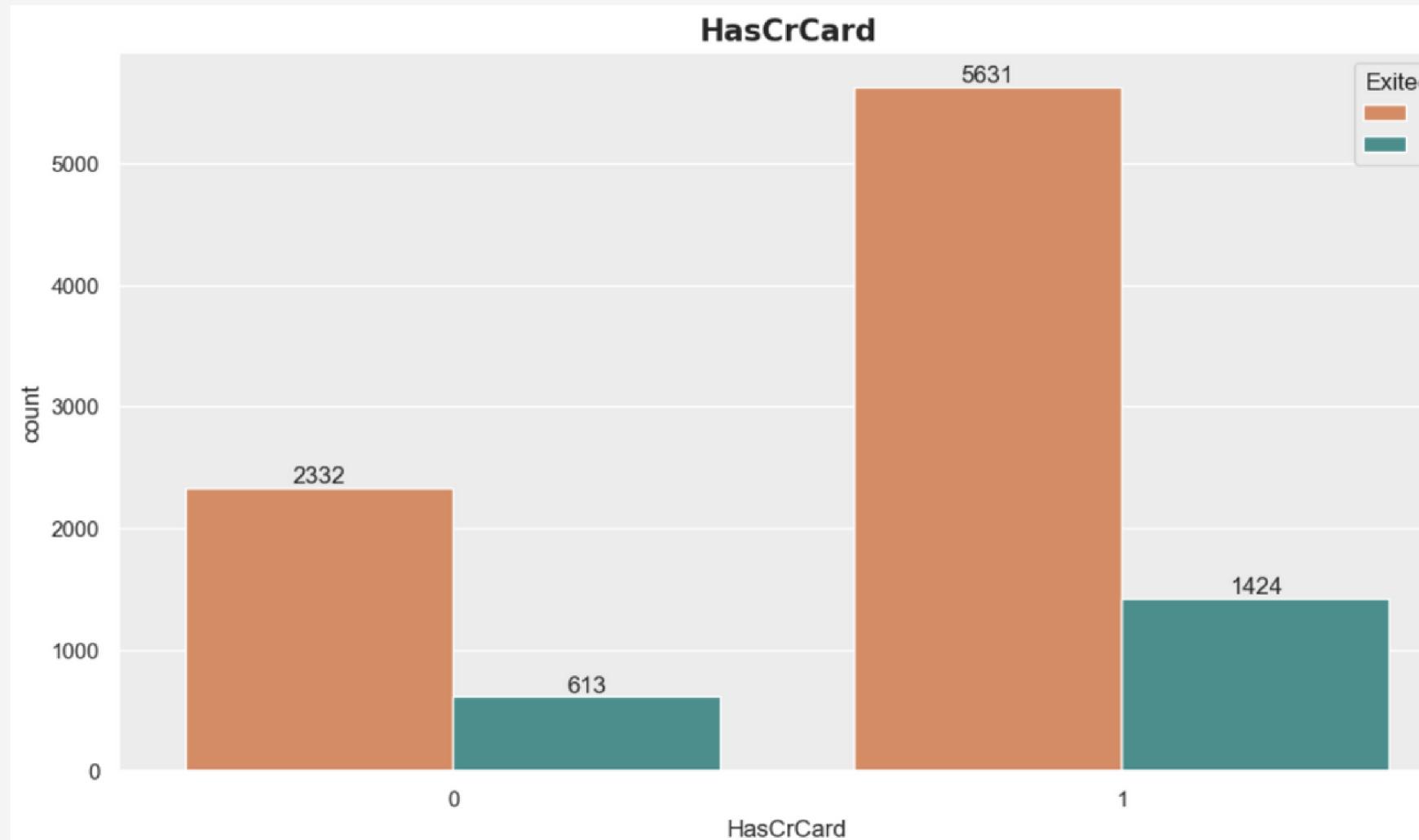


20% of the customers have churned and
80% haven't

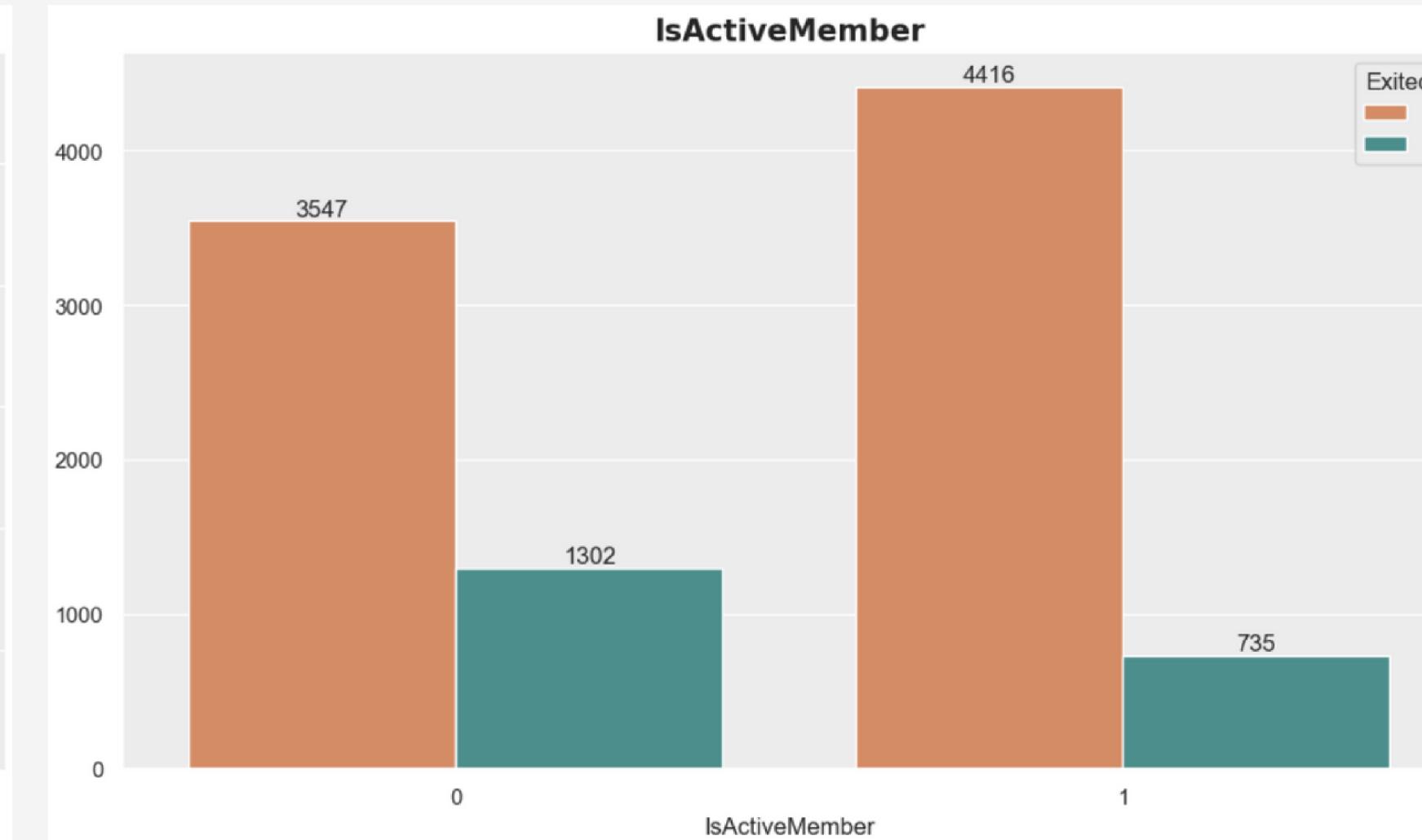


The percentage of female customers who exited is higher than the percentage of male customers who exited.

EDA

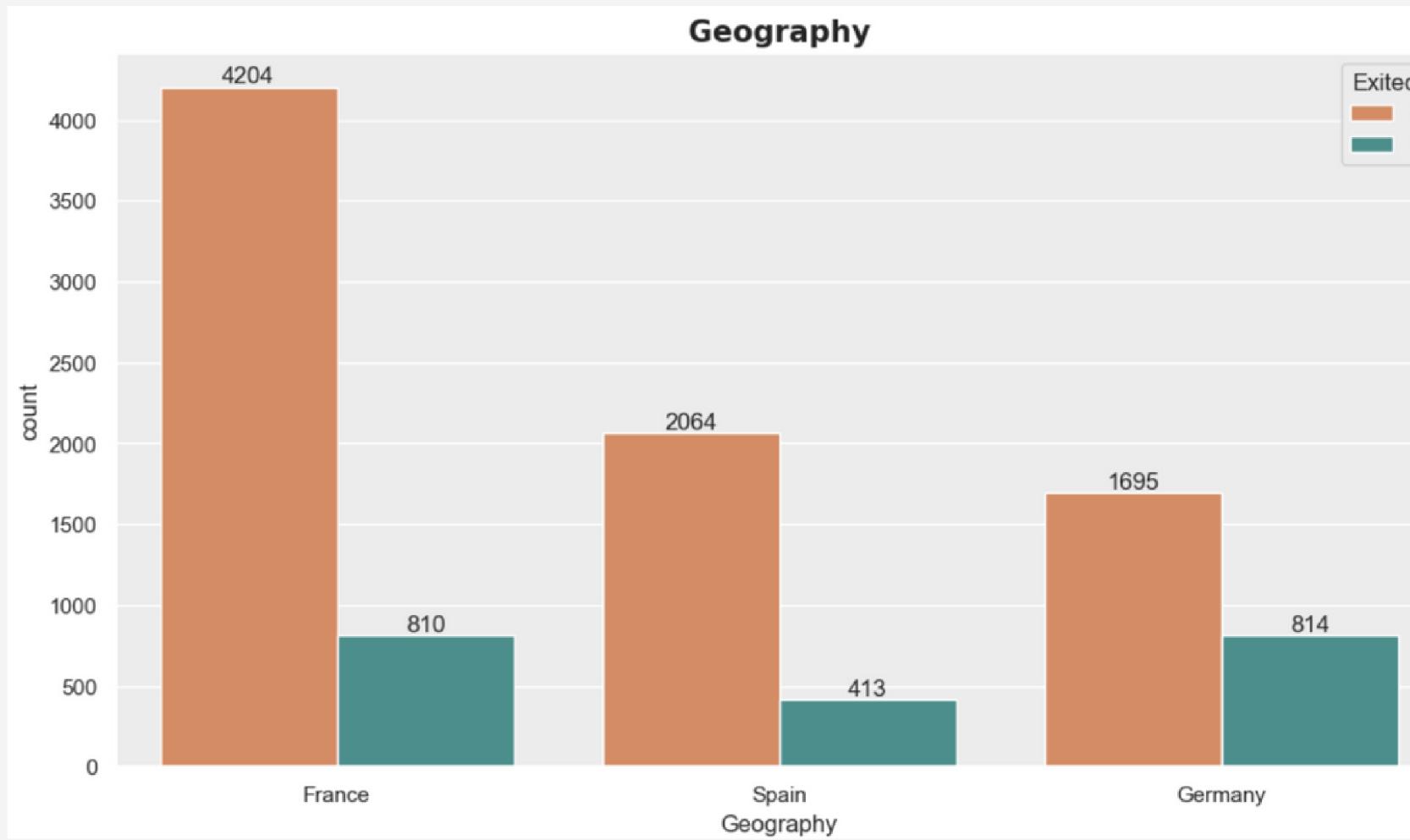


The majority of the customers that churned are those with credit cards.

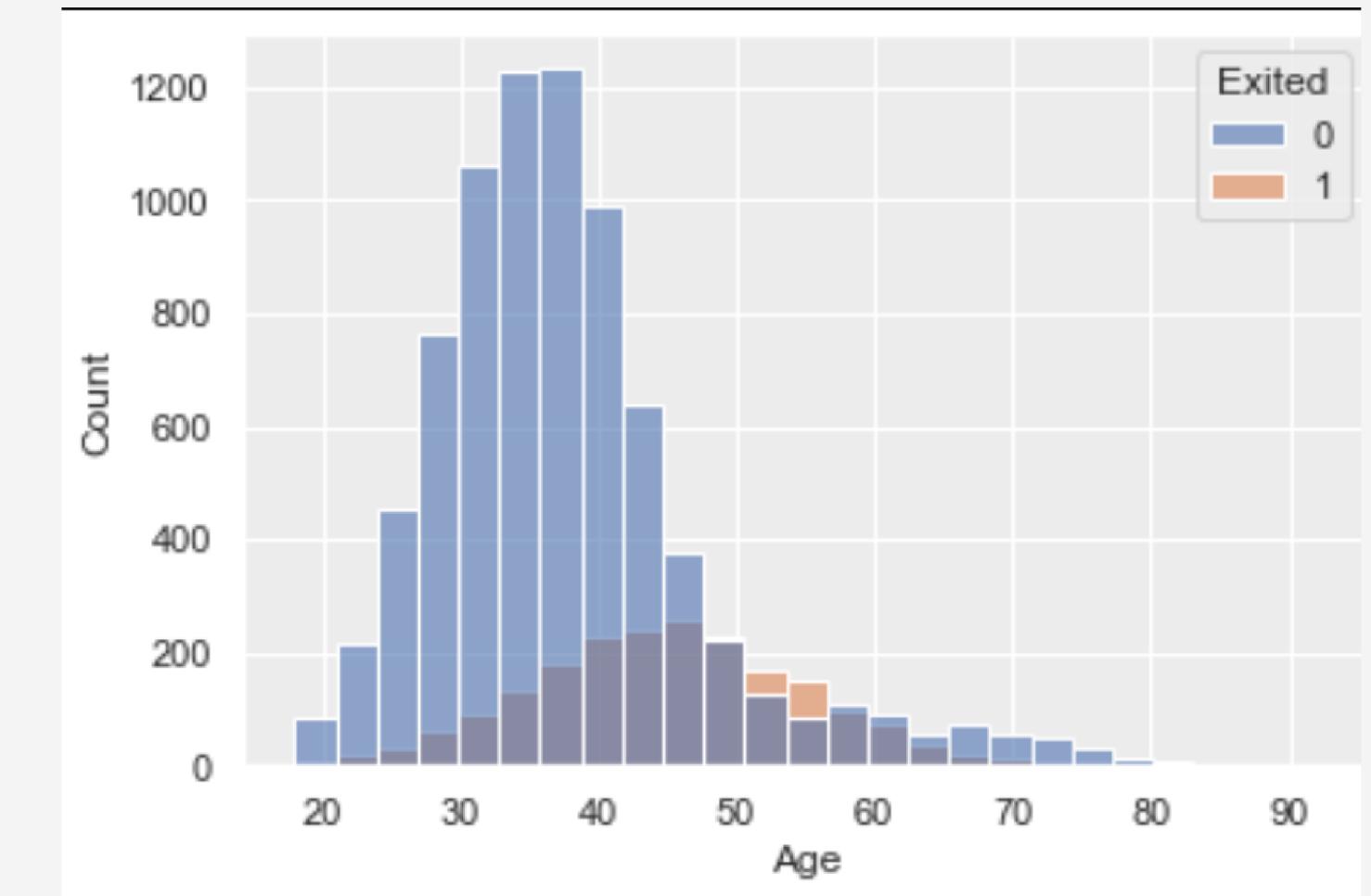


The inactive members have a greater churn.

EDA



Majority of the customers are from France but most customers which churned are from Germany.

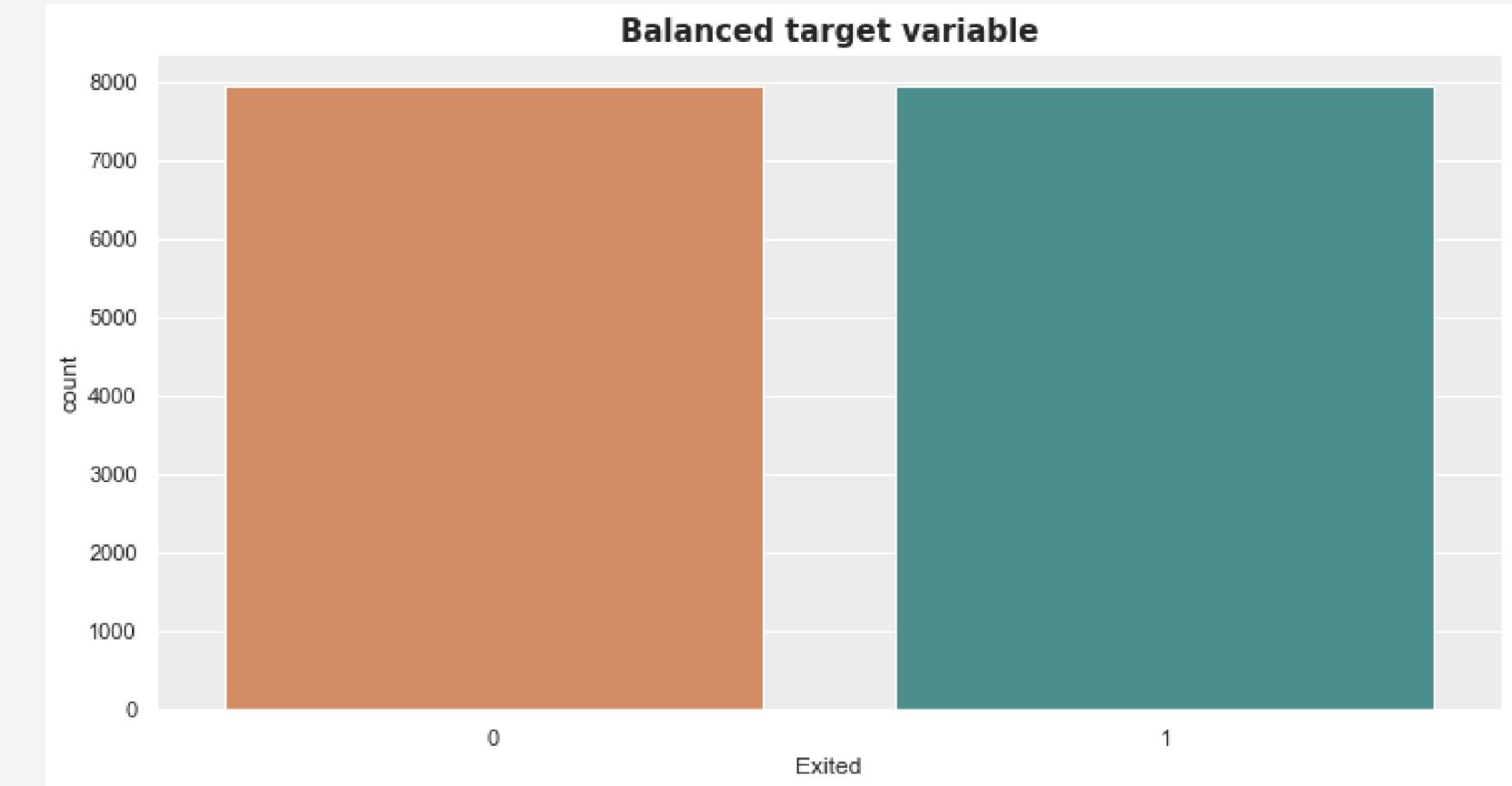
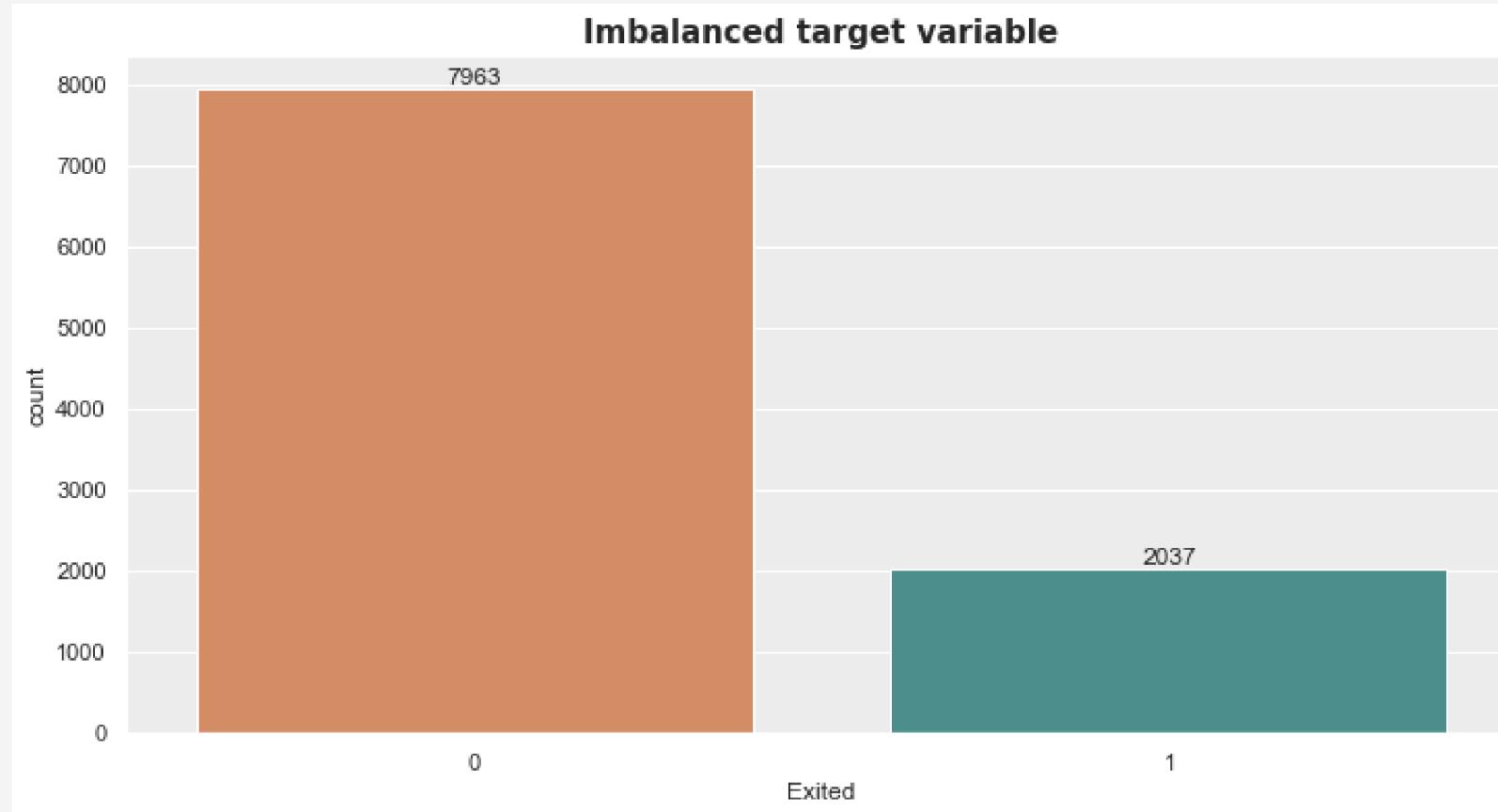


The average age of customers who do not leave the bank is 30-40 years old, while the average age of customers who leave the bank is 40-50 years old

Conclusions from the Dataset

- The age of a customer has the strongest correlation with whether they have left the bank, with a coefficient of 0.29.
- Credit score distribution doesn't significantly differ between churned and non-churned customers.
- Older customers are more likely to churn than younger ones.
- The bank loses customers with substantial account balances.
- Estimated Salary doesn't strongly influence churn likelihood.
- Female clients are slightly more likely to switch banks than male clients.
- Churn rates are lower in the first and tenth years of tenure.
- Most clients own one or two products, and clients with more than three products rarely leave.
- Churn rates are similar for credit card owners and non-owners.
- Clients with higher account balances tend to churn more frequently.

Balancing the dataset



The people who have not exited is comparatively higher than the people who have exited. Hence we use the SMOTE Technique to balance the dataset

ANN

1. Definition:

- ANNs are computer programs inspired by the brain's functioning.
- They process information through interconnected "neurons"
- to solve complex problems.

1. Key Components:

- Neurons: Tiny decision-makers that process data.
- Layers: Multiple layers of neurons with specific roles.
- Activation Functions: Rules for neuron decisions.
- Training: Learning and improvement through examples.

2. Applications:

- Used in image and speech recognition.
- Predicting trends and aiding self-driving cars.

Building ANN

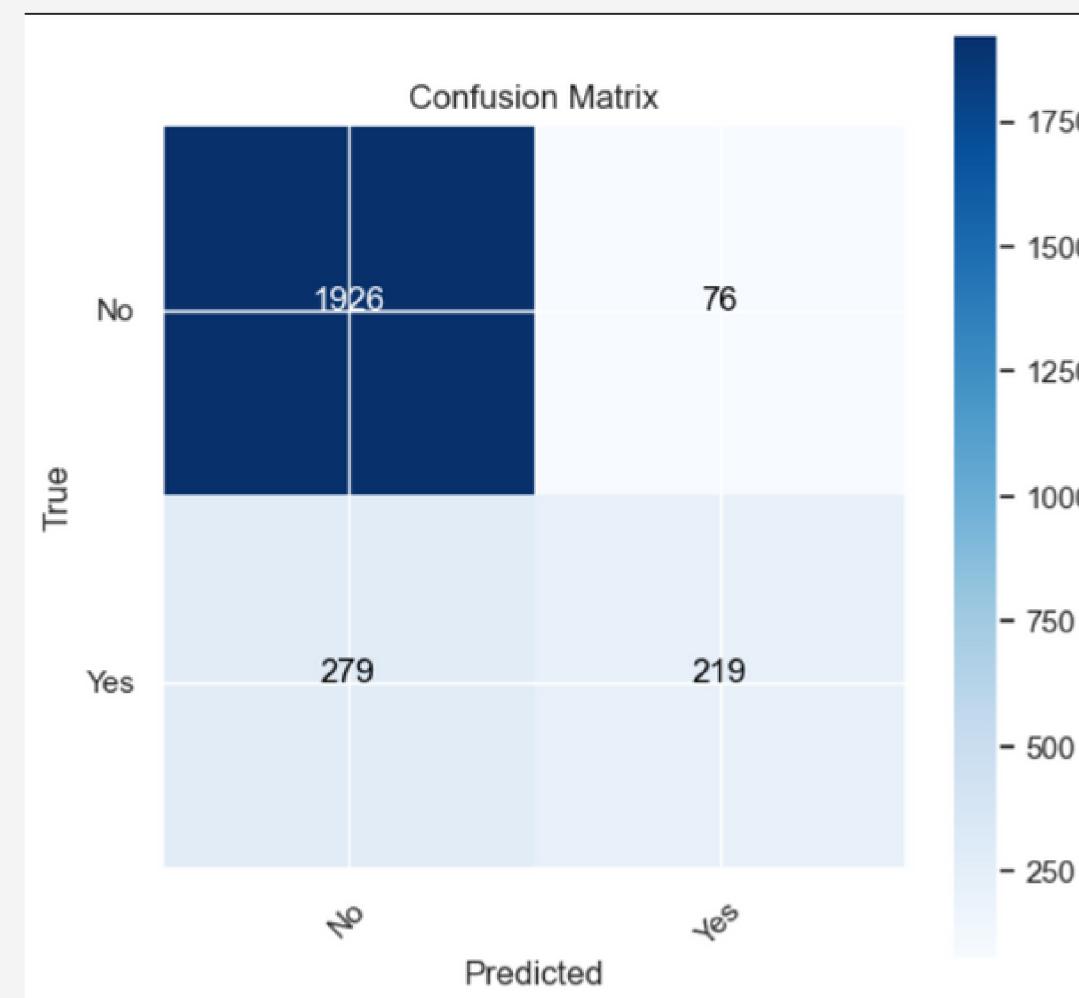
```
Model: "sequential"
-----  
Layer (type)        Output Shape       Param #  
-----  
dense (Dense)      (None, 10)          110  
dropout (Dropout)  (None, 10)          0  
batch_normalization (Batch Normalization) (None, 10)          40  
dense_1 (Dense)    (None, 7)           77  
dropout_1 (Dropout) (None, 7)           0  
batch_normalization_1 (Batch Normalization) (None, 7)          28  
dense_2 (Dense)    (None, 2)           16  
-----  
Total params: 271 (1.06 KB)  
Trainable params: 237 (948.00 Byte)  
Non-trainable params: 34 (136.00 Byte)
```

- Type: Sequential
- Layers:
 - 1st: 10 neurons, ReLU,
 - 2nd: 10% dropout
 - 3rd (Batch Norm): 10 neurons
 - 4th: 7 neurons, ReLU,
 - 5th: 10% dropout
 - 6th (Batch Norm): 7 neurons
 - 7th (Output): 2 neurons, Sigmoid
-

79/79 [=====] - 0s 3ms/step - loss: 0.3594 - accuracy: 0.8488
Accuracy of model is 0.848800003528595

Results

Confusion Matrix



Classification Report

	precision	recall	f1-score	support
No	0.87	0.96	0.92	2002
Yes	0.74	0.44	0.55	498
accuracy			0.86	2500
macro avg	0.81	0.70	0.73	2500
weighted avg	0.85	0.86	0.84	2500

Model is moderately good at predicting people who left(exited), with precision of 74%, recall of 44%, and F1 score of 55% and almost correctly predicts the people who stayed(not exited) with precision of 87%, recall of 96% and f1 score of 92%

THANK YOU!