MY BIO

EDUCATION

- MSC in Mental Health(UK)
- BN in Mental Health(UK)
- BA Statistics(SA)
- Current Clinical Facilitator with UTS and UTAS

Capstone Project

- Predicting Customer Churning (Banking Sector)
- Classification Problem

- Presented by Elijah Sabondo
- Capstone Project Institute Of Data (IOD)

Agenda

Introduction

What is customer churn?

Why is it a Business Problem?

Project Pipeline

Exploring Data Analysis

Model Designs

Model Evaluation & Choice of Model Design

Recommendations

Conclusion

Introduction- What is customer churn?

Customer churn occurs when customers or subscribers stop doing business with a company or service

For this project it occurs when customers stop doing business with the bank and leave

Customer retention, on the other hand, is generally more cost-effective as you've already earned the trust and loyalty of existing customers.

Why is it a Business Problem?

Customer Churning is a problem for businesses because of costs of getting new customers:

Less Customer Less Revenue or More Customers More Revenue. Implicit in customer Churning is

Customers Churn because of:

Bad Customer Service

Bad Onboarding

Lack of ongoing Customer Success

When customers churn a business lose both the customer and revenue.

Introduction- Dataset for Bank Customers Churning

This is a dataset in which there are details of a bank's customers and the target variable is a binary variable reflecting the fact whether the customer left the bank (closed their account) or they continues to be a customer.

Churn Prediction of bank customers | Kaggle

There are 10000 rows and 14 columns. Each row represents a bank customer, and each column contains customer's attributes.

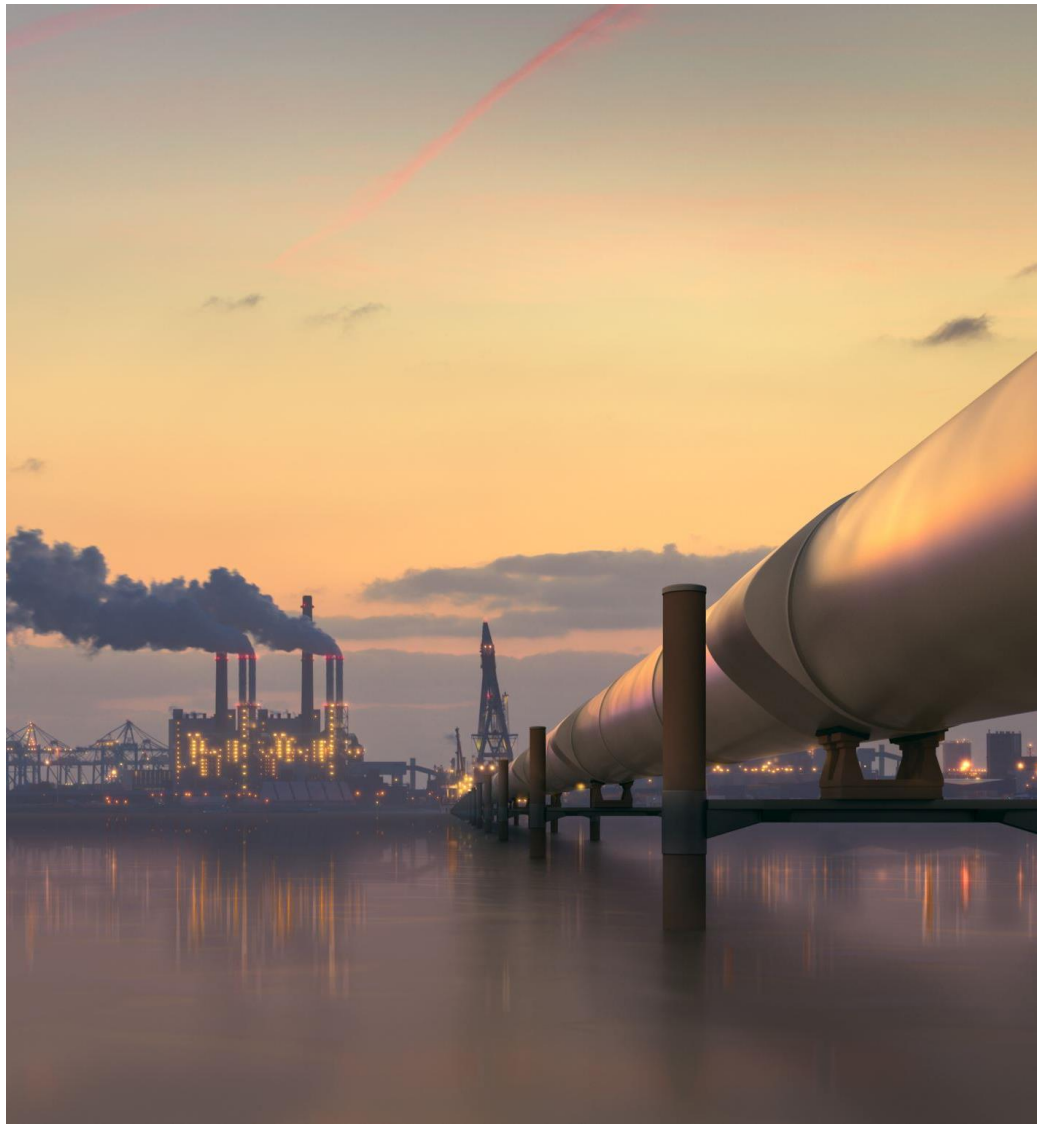
The attributes or features are:

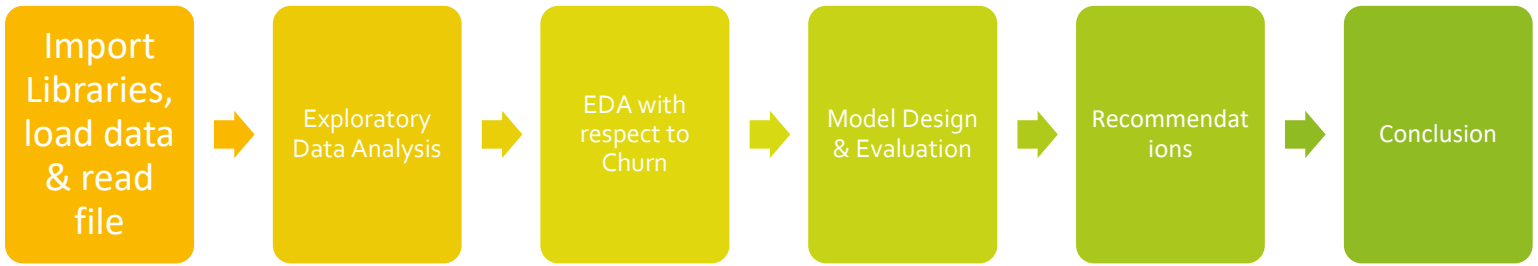
'

Dataset Shape & features

Shape: (10000 rows and 14 columns)
RowNumber
CustomerId
Surname
CreditScore
Geography
Gender
Age
Tenure
Balance
NumberOfProducts
HasCrCard
EstimatedSalary
IsActiveMember
Exited

Project Pipeline




Import Libraries, load data & read file
Exploratory Data Analysis
EDA with respect to Churn
Model Design & Evaluation
Recommendations
Conclusion

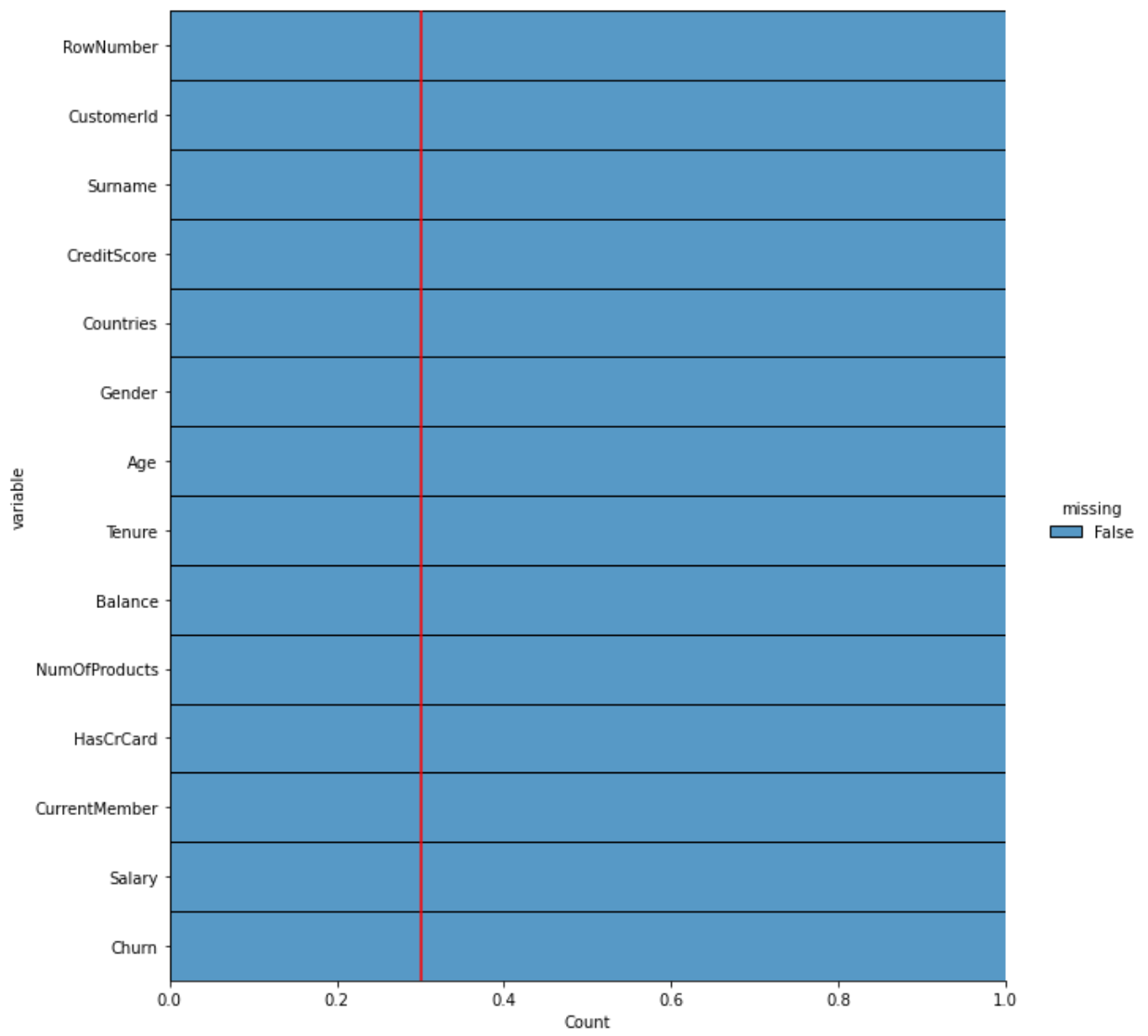
EDA

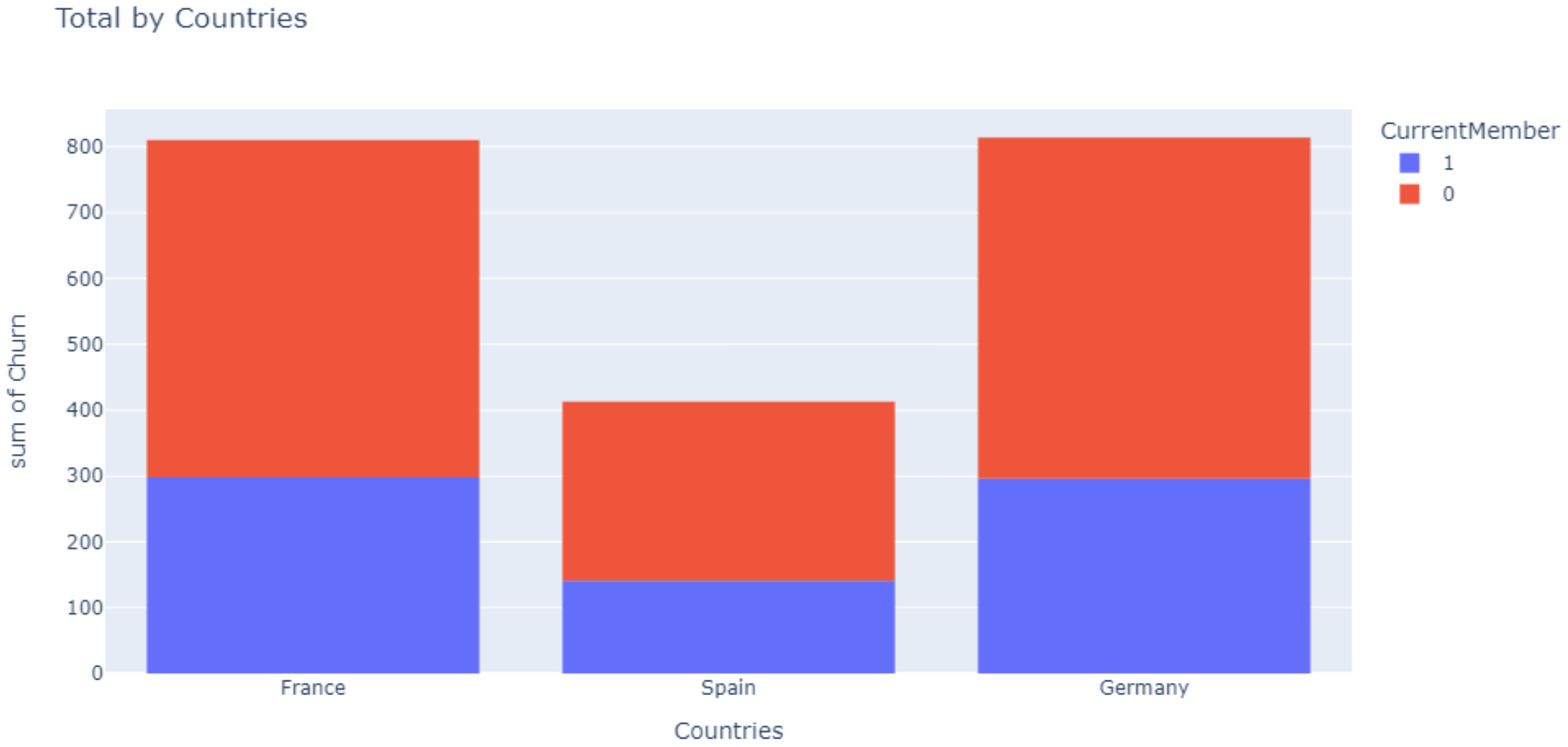
Exploratory Data Analysis

EDA

There is no missing data

NO missing Data

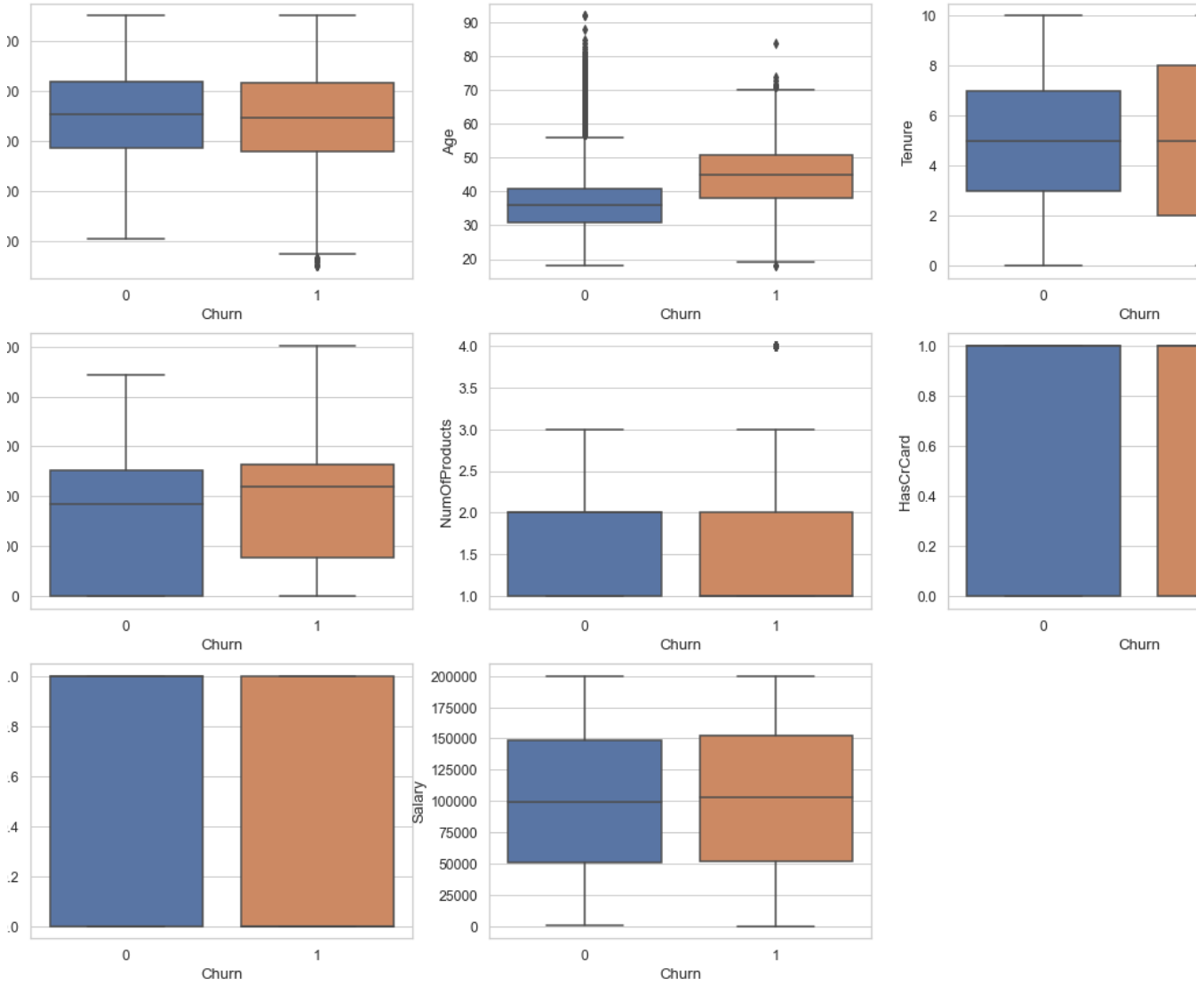



Total by Countries
sum of Churn
800
700
600
500
400
300
200
100
0
France
Spain
Germany
Countries
CurrentMember
1
0

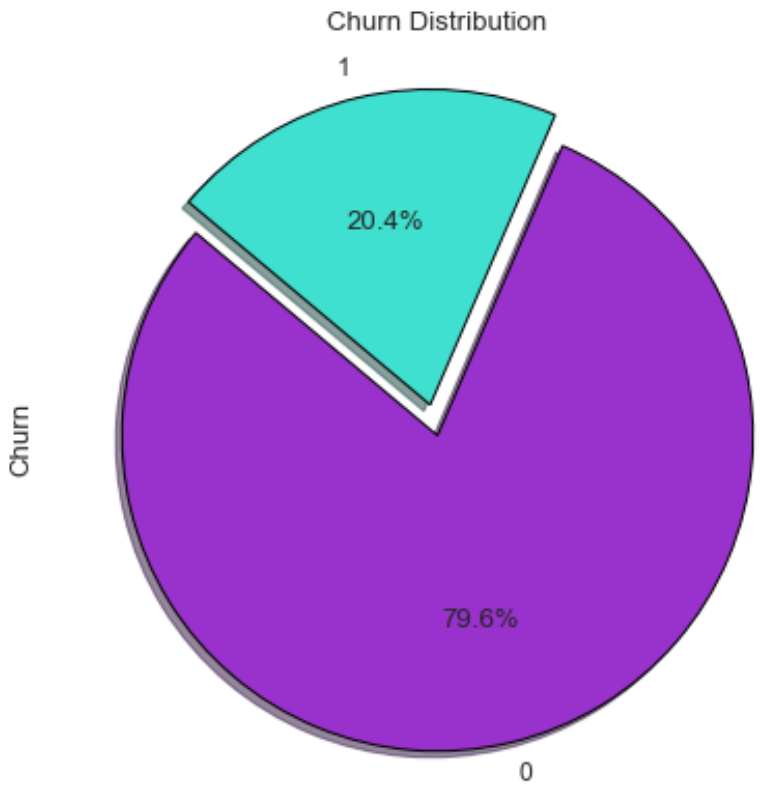
EDA WRT CHURN

- EDA with Respect to Churn

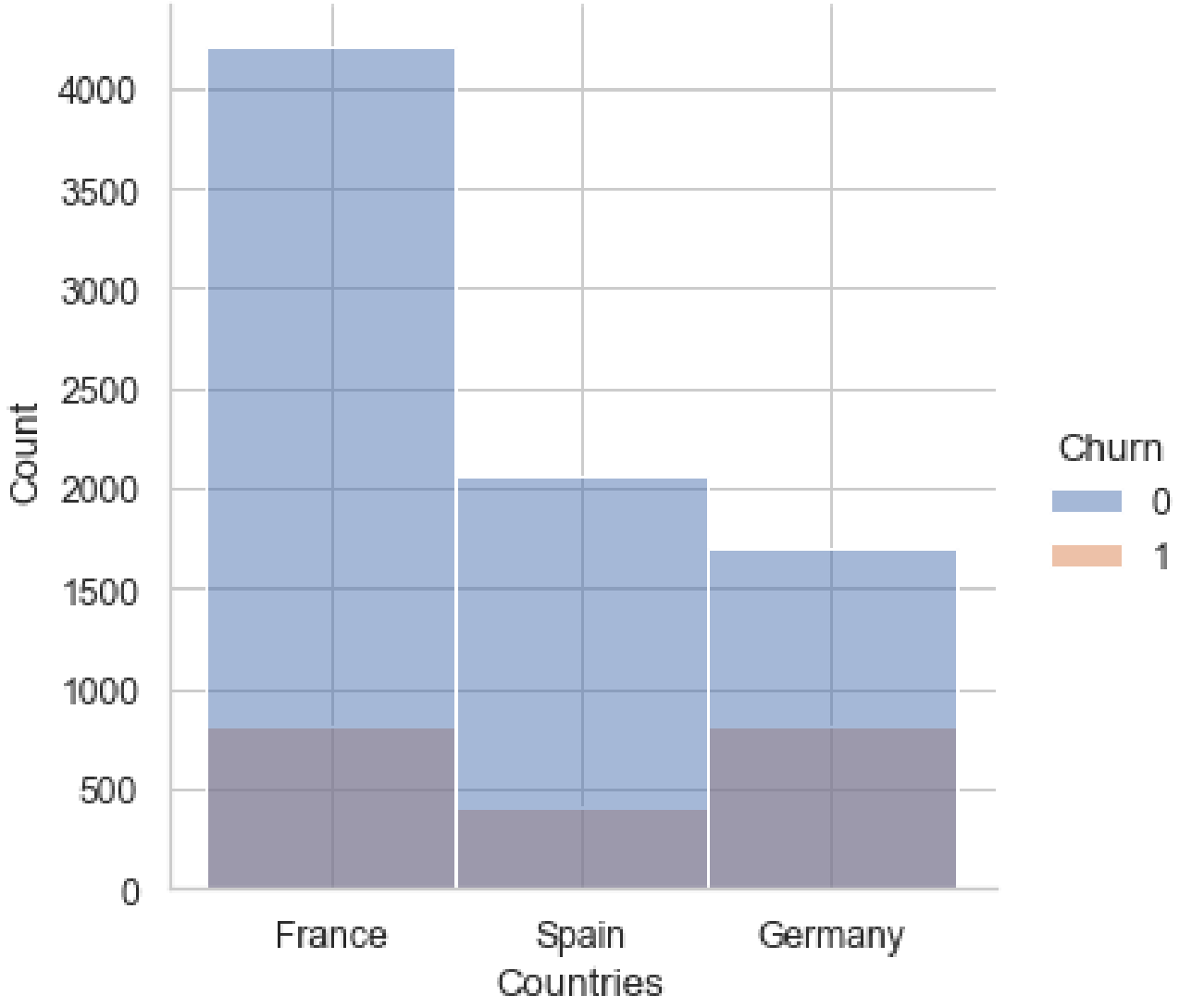
EDA



EDA with respect to Churn

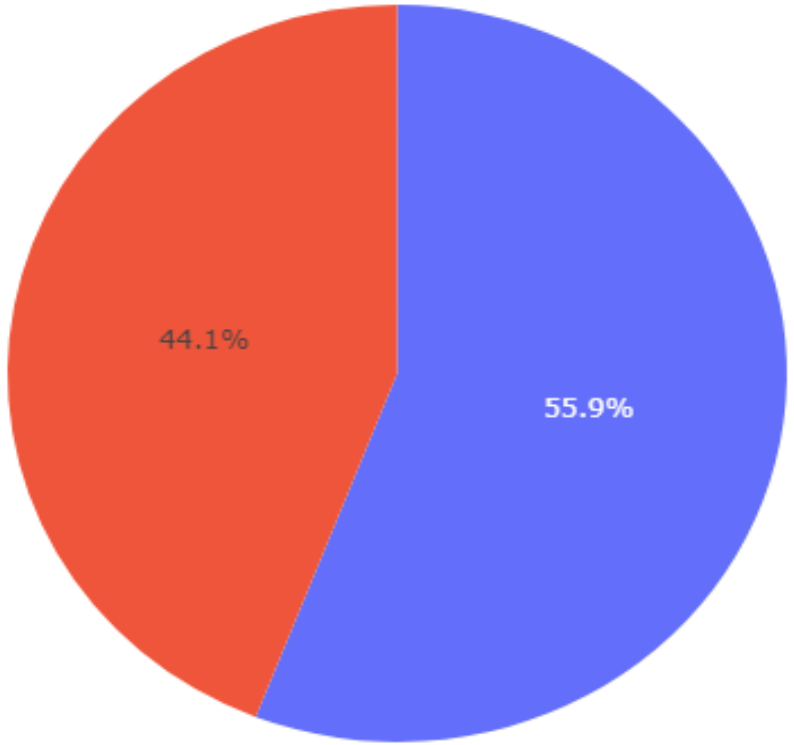


Churn by country



Churn by Gender

ıder



EDA with respect to Churn

- Insights about Target variable ("Churn")
- 2037 customers left and 7963 did not Churn
- France has the largest number of customers followed by Germany and Spain the last.
- There were 7963 non-Churn (80%) and 2037 Churn (20%)
- More females(56%) left the bank than male(44%)
- (Lost Customers ÷ Total Customers at start of period) x 100) = (2037/10000) x 100%) = 20.37%
- The data is imbalanced

MODELS

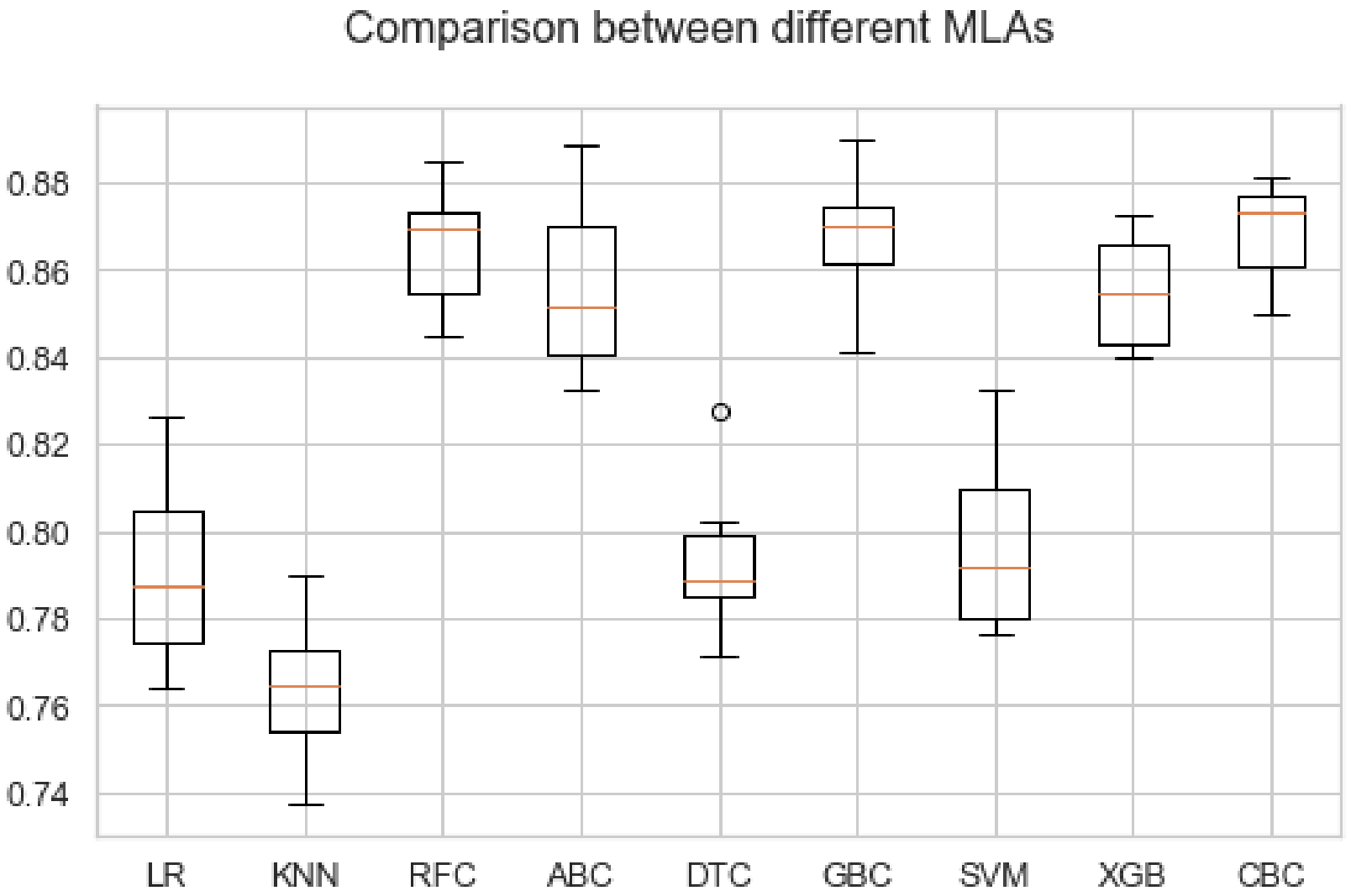
- MODELS

Which model is performing the best?

```
LR| Mean=0.789800 STD=0.006645
RFC| Mean=0.862500 STD=0.008559
KNN| Mean=0.765000 STD=0.009088
DTC| Mean=0.789400 STD=0.008126
ABC| Mean=0.858300 STD=0.006435
SVC| Mean=0.796300 STD=0.010479
GBC| Mean=0.865100 STD=0.009628
XGB| Mean=0.853900 STD=0.006503
CBC| Mean=0.864100 STD=0.007049
```

Which model is performing best

CBC: 0.868500 (0.010980)



Score	LR	SVC	DT	RF	GB	XGB	ABC	CBC	Best_Score
Accuracy	0.790300	0.7963	0.792300	0.855500	0.864200	0.865000	0.858000	0.865700	CBC
Precision	0.397260	0.0000	0.490383	0.849198	0.772858	0.758051	0.733353	0.764820	RF
Recall	0.057431	0.0000	0.518896	0.353956	0.472261	0.495828	0.476677	0.493369	DT

Recommendations & Conclusions

CONCLUSION

Of the ML alogorithms used CatBoostClassifier topped the rank, followed by Random Forest, and Decision tree in third place.

Data was imbalanced and used SMOTETokem to balance the data

Applying balanced data on Catboost Algorithm to improve the precision and recall scores returned a slight reduced Accurancy, Precision , but improved recall by about 16%

CatBoost alogorithym can still predict Customer Churn with 84% Accuraracy, Pecision 61% and Recall 59% and with 86% of the data as area under the curve(i.e. 86% of the data is included).

CONCLUSION

Consideration

- Exit Interviews/ surveys

Any Questions