



Data Lovelace

FINAL PROJECT RAKAMIN BATCH 39

# BANK CUSTOMER CHURN

PREDICTION

[Sourcecode](#)





As a Data Scientist consultants, we are responsible for analyzing the factors that cause customer churn in banking companies. We will create a machine learning model that can predict customers who have the potential to churn, and make business recommendations based on the results of this analysis.

## Meet Our Team



Kevin  
Haposan  
Aripuratama



Adi Nur  
Rohkim



Shania  
Widianingrum  
Puspitasari



Mirra  
Afifah



Febi  
Maharani  
Noro Putri



Kurniasari  
Septa  
Nugrahaeni

Project Leader

Data Analyst

Data Analyst

Data Scientist

Data Scientist

Business Intelligence

# About the Client

## RakaBank Company

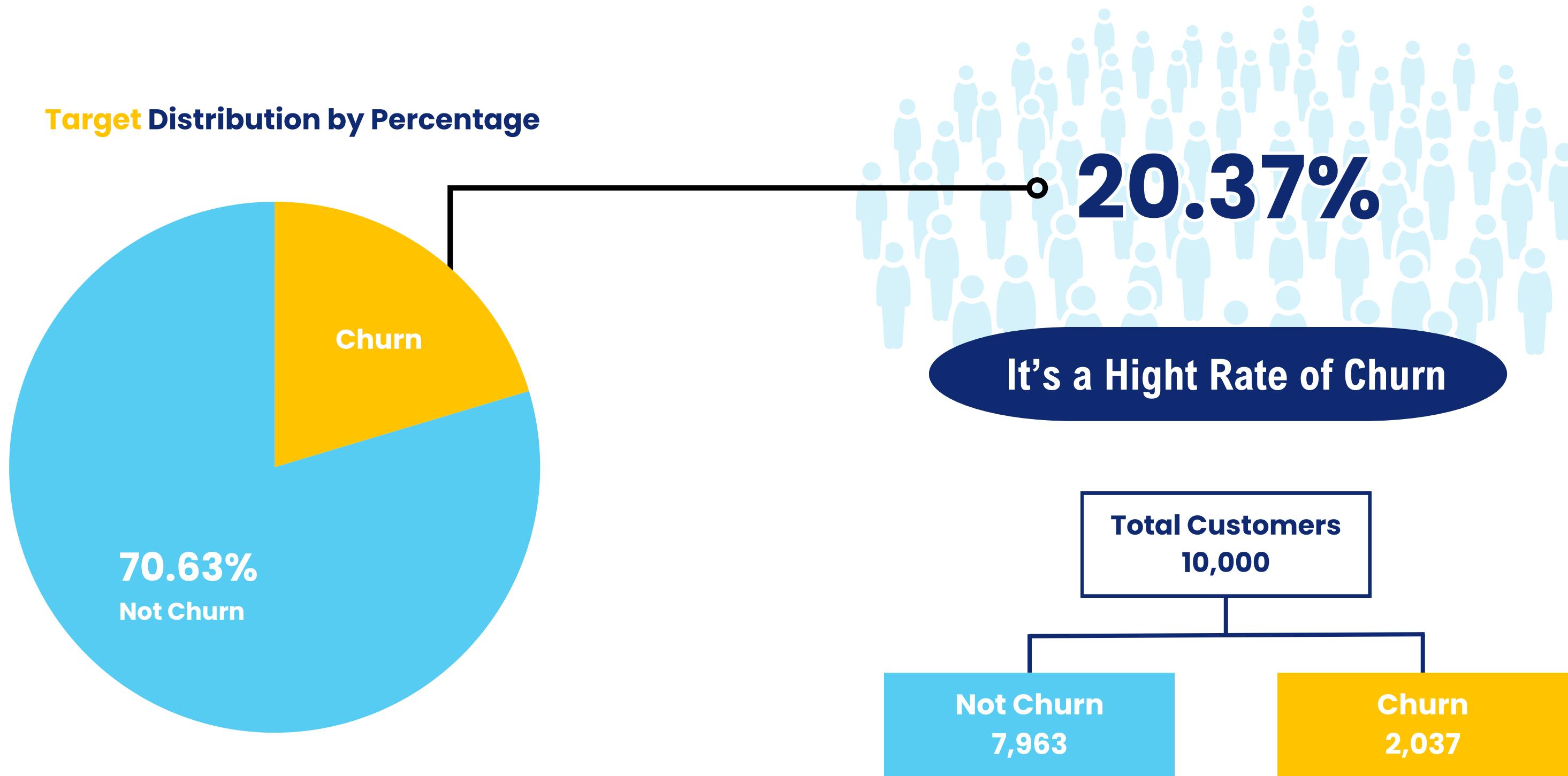
Technological developments and increasingly fierce competition are pushing companies to **not only** focus on getting new customers, but also **retaining existing customers**. **Raka Bank** is one startup company that focuses on this problem.

Based on the data the company has, **Raka Bank** wants to identify which aspects encourage **customers to churn** so that they can provide the right solution to prevent customers from churning.

**Raka Bank** asked **Data Lovelace** for help to **predict and identify factors** that encourage **customers to churn**.



# What is The Problem ?



What is a good churn rate for a startup bank company?

It is generally believed that an annual churn rate of **less 15%** is acceptable for startup banks.

Source: [www.younium.com/blog/saas-churn-rate](http://www.younium.com/blog/saas-churn-rate)

# What is The Problem ?

## Acquisition Cost Loss



## HIGH RATE = HIGH COST

The high rate of customer churn causes spending to be wasted.

What is CAC in finance? As per research, the average acquisition cost is \$200. The customer acquisition strategies across the successful banks are the diversified and personalized methods of engagement across the entire customer journey. (**Snigda Patel, 2024**)

Source: [www.revechat.com](http://www.revechat.com)



# Problem Summary

## Background

Bank company wants to know which **aspects** drive the **customer to churn** and **reduce the cost** of finding a new banking customer.

## Goals

Decrease churn rate from **20.37%** to at least **under 15%** in order to reduce cost

## Objectives

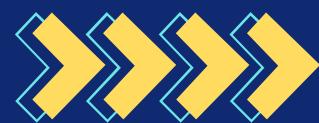
Create a model machine learning to **predict potential customer churn** and **identity the factors of customer churn**

## Business Metrics

**Churn Rate (%)**

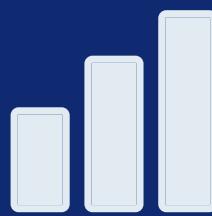
# Exploratory Data Analysis

- About Dataset
- Churn by Feature





# About Dataset



The dataset consist of 10,000 rows and 14 columns



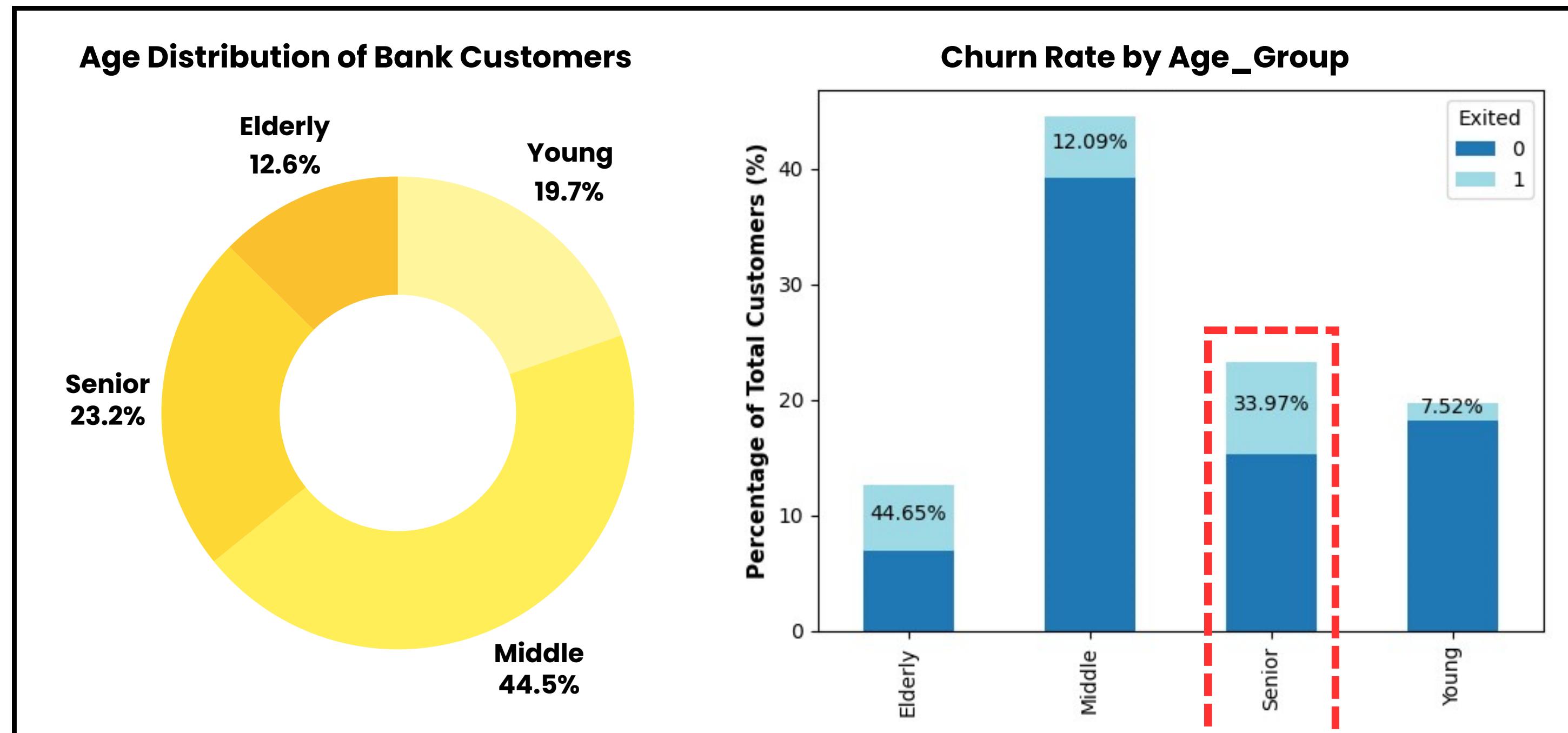
It has No Missing Value and No Duplicated



Target variabel is the Exited columns

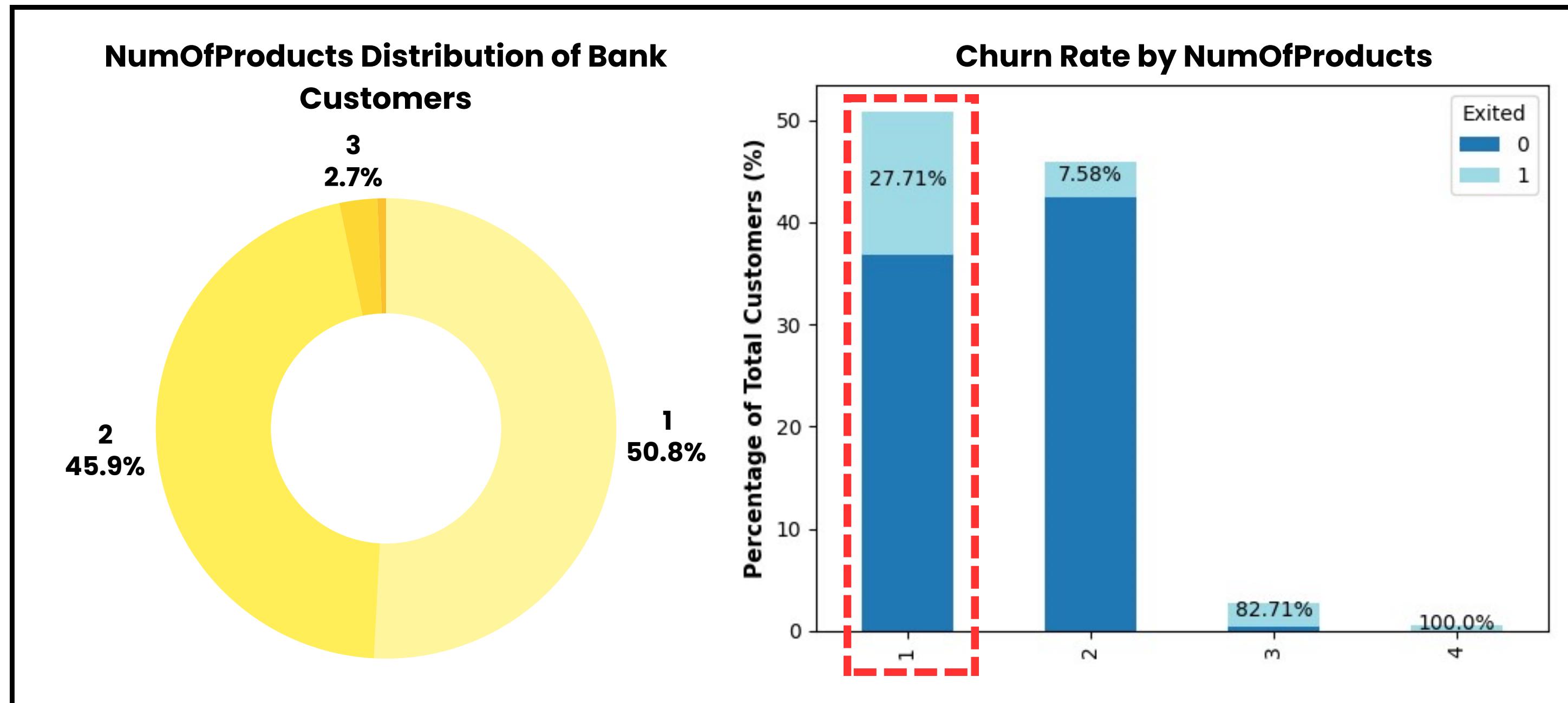
# Churn Rate by Age

The majority of customers are **Middle** age, the lowest are **Elderly**, and the highest churn rate is **Senior**.



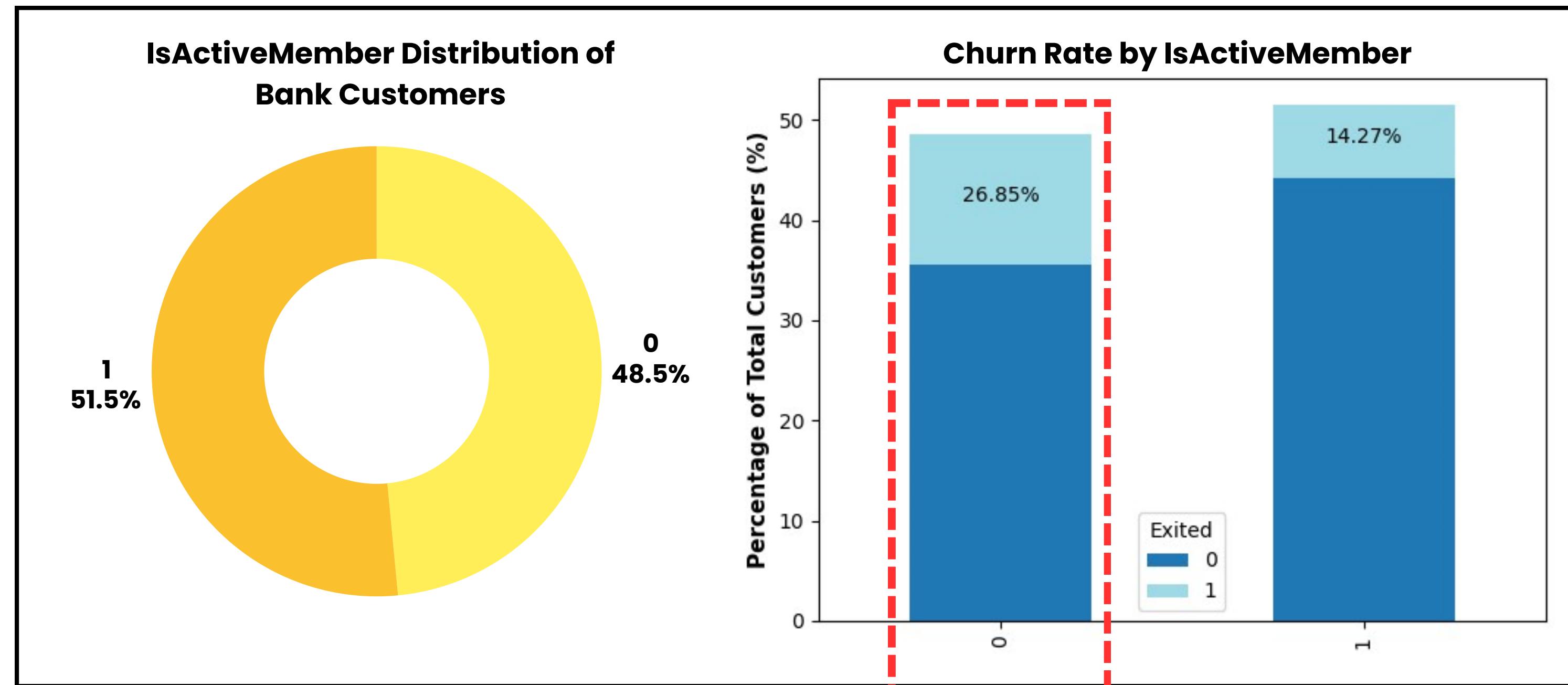
# Churn Rate by NumOfProduct

The **majority** of customers have 1 NumOfProducts and the **lowest** is 4. The **highest churn rate** is customers who have 1 NumOfProduct.



# Churn Rate by IsActiveMember

**Inactive customers** are more likely to experience **churn** than regular customers active.





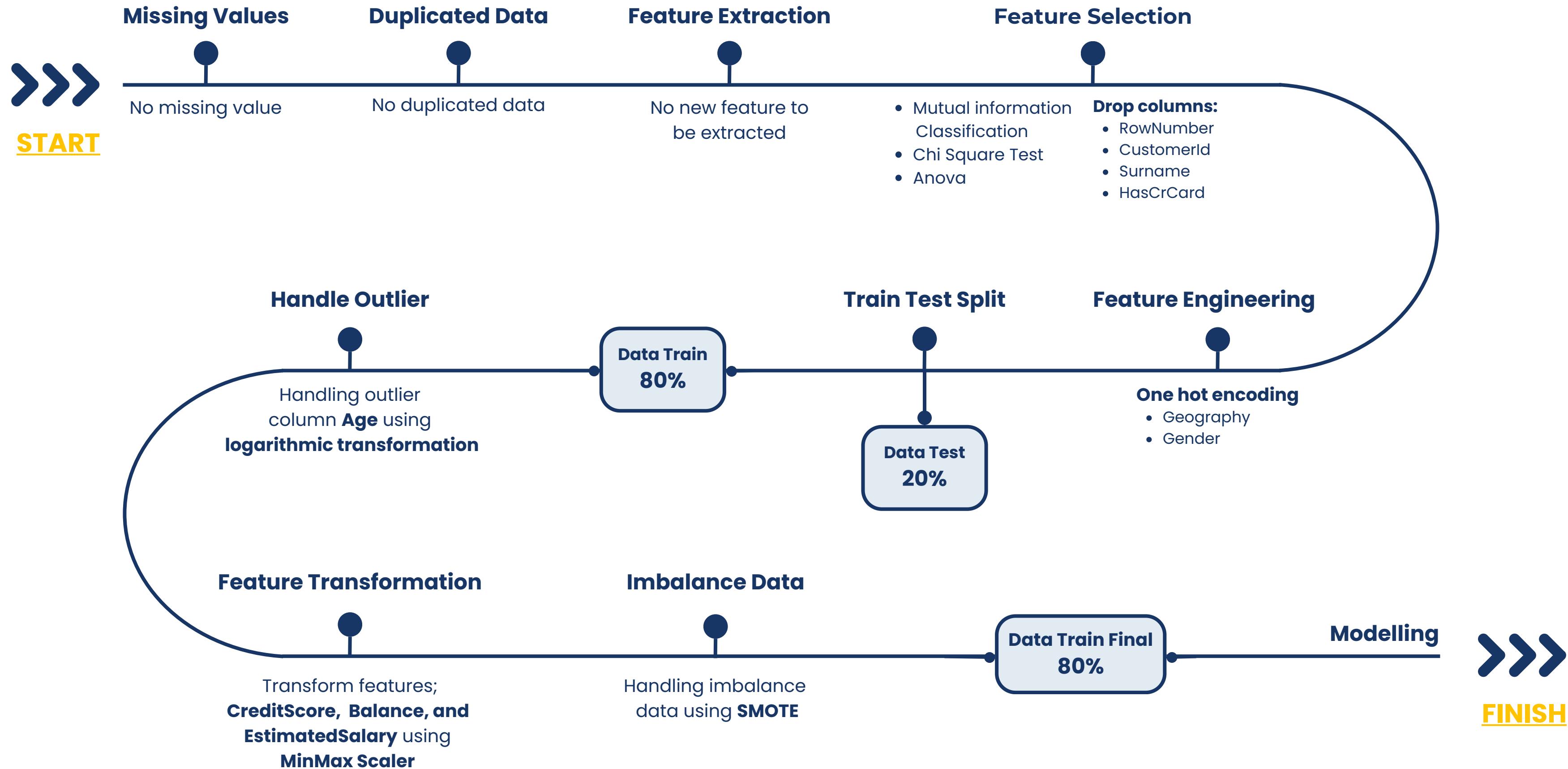
# Data Preprocessing



Preprocessing Flow



# Data Preprocessing Flow



# Machine Learning Modelling

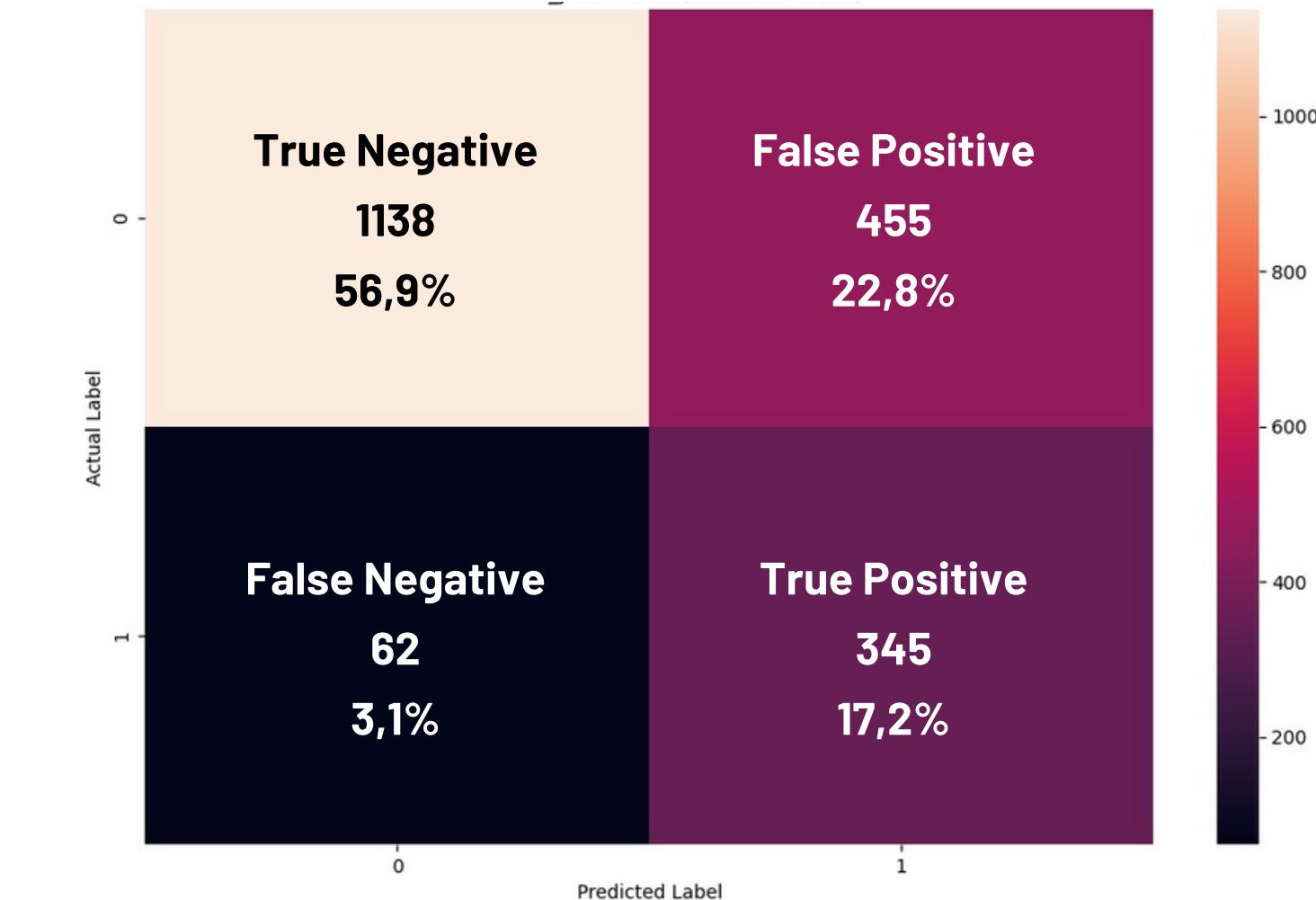
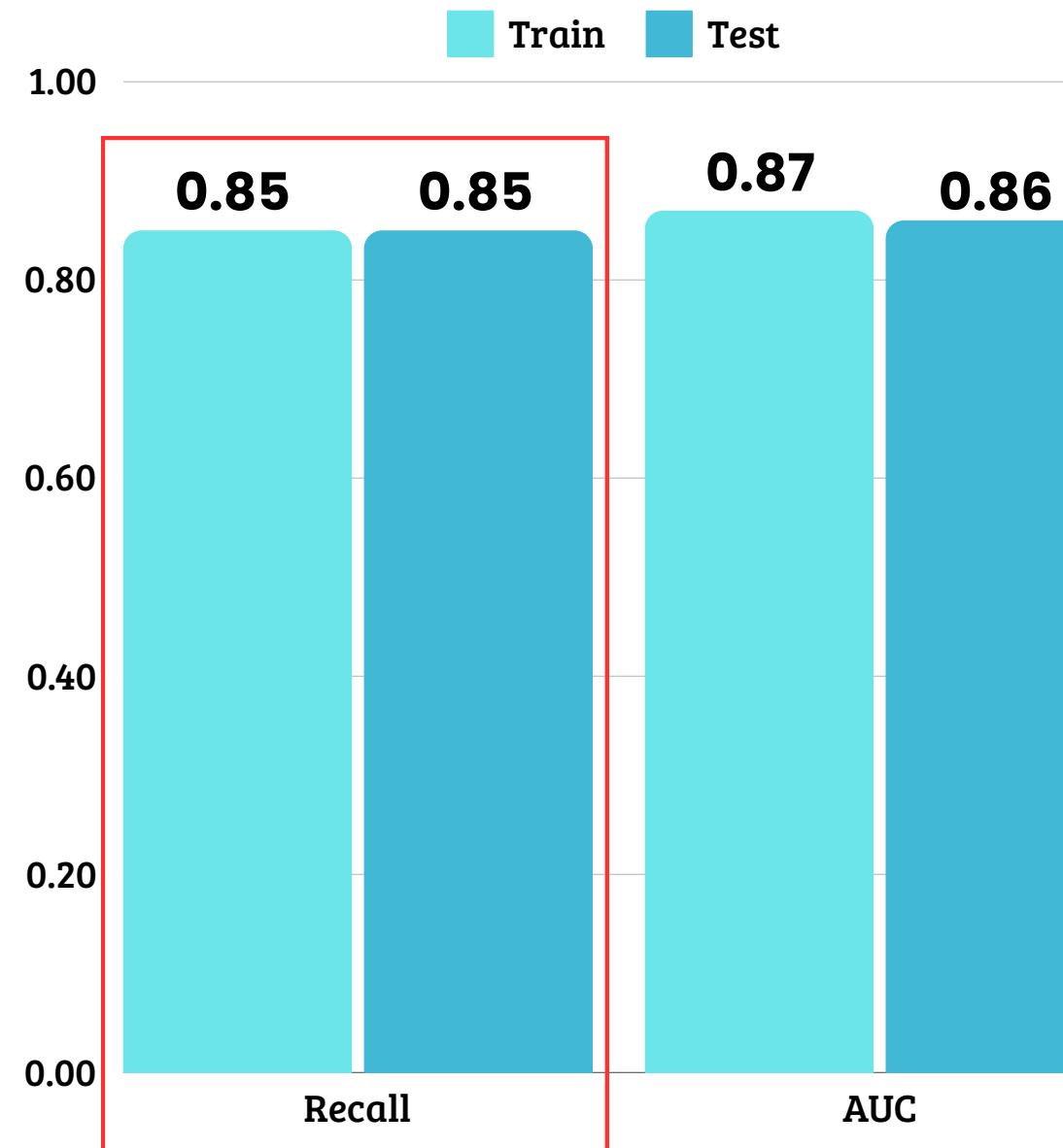
- Model Evaluation
- Feature Importance



# Model Evaluation

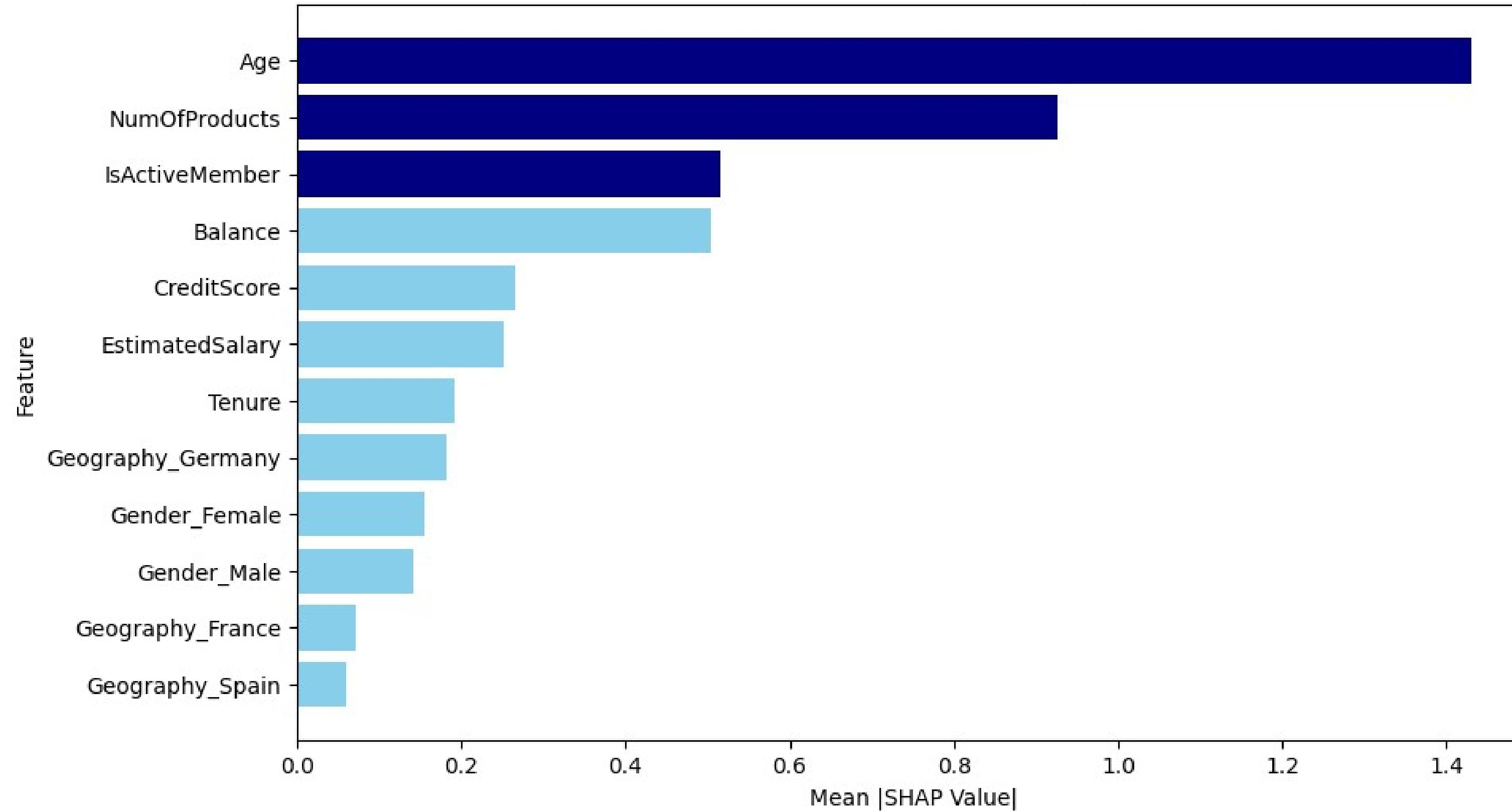
MODELS	ACCURACY (TRAIN)	ACCURACY (TEST)	RECALL (TRAIN)	RECALL (TEST)	AUC (TRAIN)	AUC (TEST)
LOGISTIC REGRESSION	0.71	0.70	0.70	0.65	0.77	0.77
DECISION TREE	0.79	0.79	0.65	0.66	0.79	0.79
ADABOOST	0.79	0.79	0.73	0.76	0.84	0.83
KNN	0.78	0.72	0.81	0.67	0.88	0.77
RANDOM FOREST	0.73	0.73	0.68	0.66	0.78	0.78
XGBOOST	0.74	0.74	0.85	0.85	0.87	0.86

# Model Evaluation



Recall **measures the ratio** between the number of customers that are **predicted to leave** who actually left the bank company, and the number of customers that are **predicted to stay** but **actually** left the bank company. **Maximizing recall** means **minimizing** the number of customers who are **incorrectly predicted** to stay.

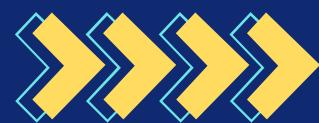
# Feature Importance





# Business Recommendation

- Recommendation
- Business Impact



# Recomendation for Business Team

## Age



**33.97% of Senior customers (40 - 50 years old) are predicted to churn, so we recommend:**

- Offer retirement preparation products with high returns.

## Bundling Products



**27.71% of Customers with 1 products predicted to churn, it's higher than customers with 2 products. So, It is necessary to evaluate and treat customers with 1 product by promoting cross selling (i.e offer bundling products).**

# Recomendation for Business Team

## Active Member



Customer who is not an active member are predicted to churn around 26,85 %. It is around 12 percent higher than churn in active member.

To increase the activity of inactive member, we recommend to offer low interest and low deposit products to the inactive members.

The offering to the inactive members can be displayed through several channels such email, sms, social media.

# Business Impact

**345** number of customers we correctly predict to churn (True Positive)  
**62** number of customer we misclassified as people who stay (False Negative)  
**2000** total customers in data test

Assumption 50%

$$\text{Churn Rate} = (0.5(\text{TP}) + \text{FN})/\text{Total}$$

**Before Model**

**20.37%**

**2.037**



**Decrease of**

**8.65%**

**865**

**42% ↓**

**After Model**

**11.72%**

**1.172**

# Business Impact

Assumption average **acquisition cost** is \$200

Assumption **marketing cost** is \$100

**Before model**

$$2.037 \times 200 =$$

**407.400 USD**

**After model**

$$1.173 \times 200 =$$

**234.600 USD**

**Marketing Cost**

$$800 \times 100 =$$

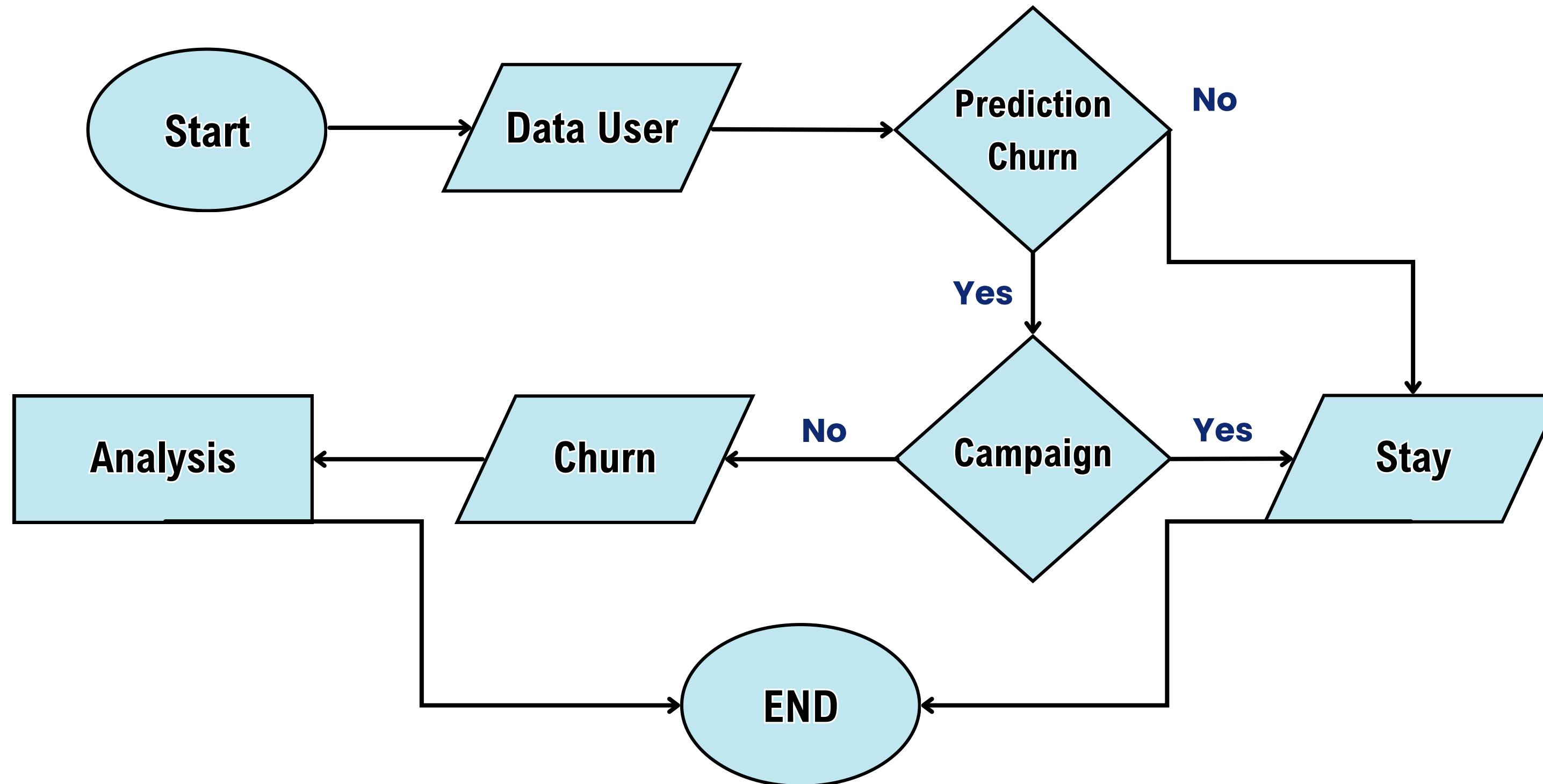
**80.000 USD**

---

**92.800 USD**

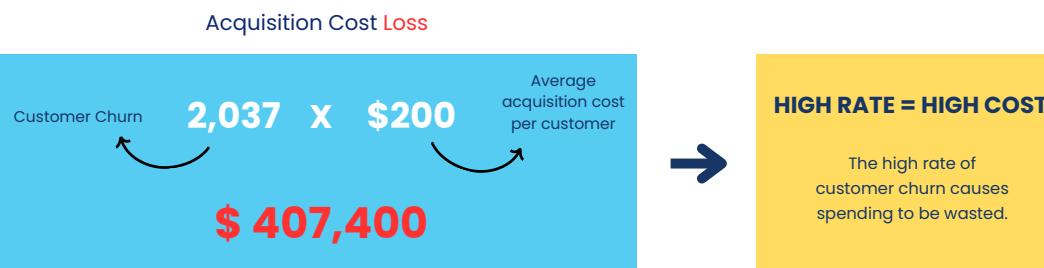
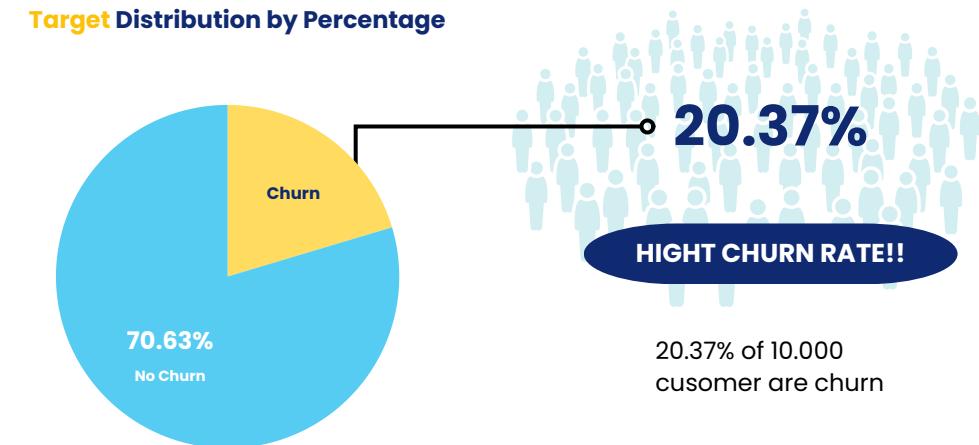
**Saving Money**

# Model Implementation



# SUMMARY

## Business Problem



### Objective

Create a model machine learning to **predict potential customer churn** and **identity the factors of customer churn**

### Goals

- Decrease churn rate from **20.37%** to at least **under 15%** in order to reduce cost.

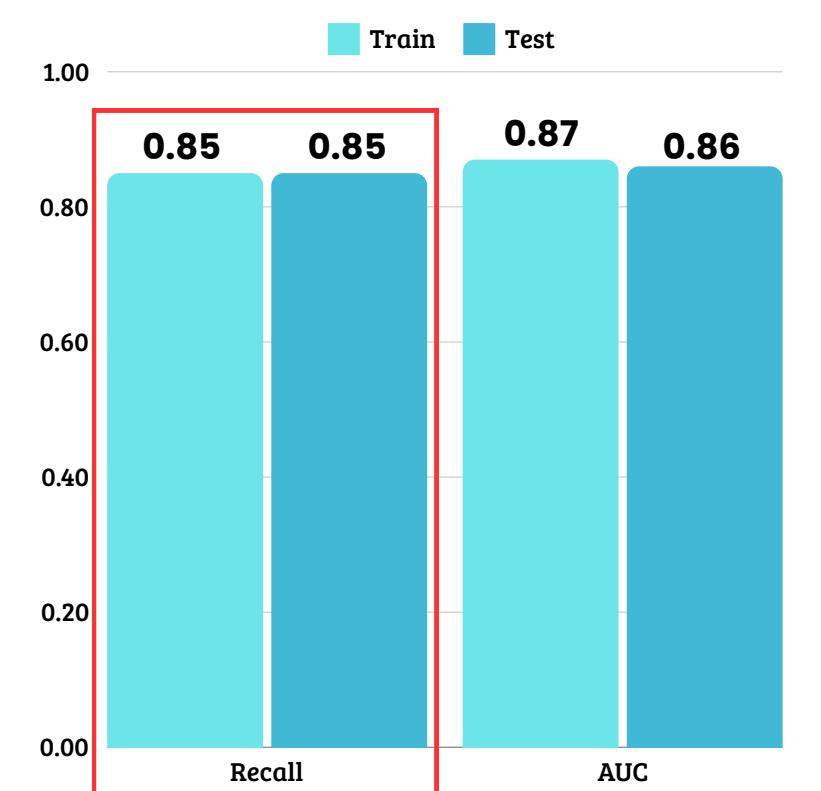
### Business Metrics

Churn Rate (%)

## Modelling

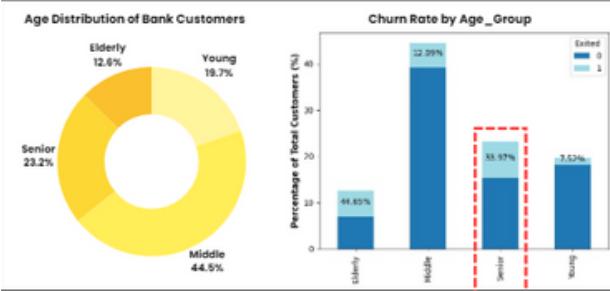
### XGBoost

- Logistic Regression
- Decision Tree
- KNN
- Random Forest
- AdaBoost



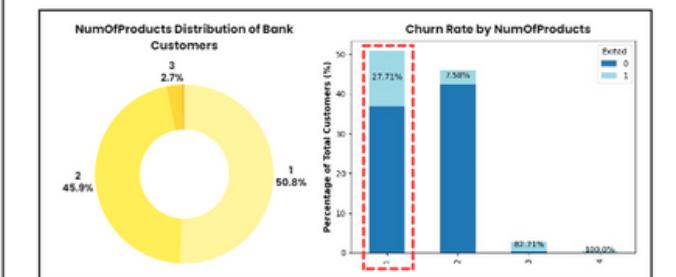
## Churn Rate by Age

The majority of customers are Middle age, the lowest are Elderly, and the highest churn rate is Senior.



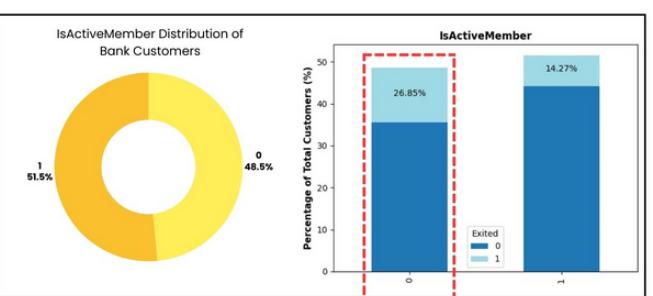
## Churn Rate by NumOfProduct

The majority of customers have 1 NumOfProducts and the lowest is 4. The hight churn rate is customers who have 1 NumOfProduct.



## Churn Rate by IsActiveMember

Inactive customers are more likely to experience churn than regular customers active.



## Recommendation

### Age

33.97% of Senior customers (40 – 50 years old) are predicted to churn, so we recommend :

- Offer retirement preparation products with high returns.

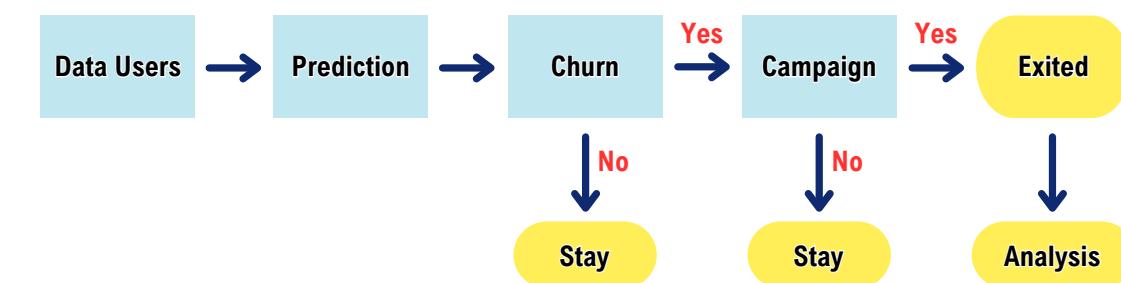
### NumOfProducts

27.71% of Customers with 1 products predicted to churn, it's higher than customers with 2 products. So, It is necessary to evaluate and treat customers with 1 product by promoting cross selling (i.e offer bundling products).

### Active Member

To increase the activity of inactive member, we recommend to offer low interest and low deposit products to the inactive members, can be displayed through several channels such email, sms, social media.

## Model Implementation



Decrease

**8.64%**

Saving Cost

**92.800 USD**

**Churn Rate**

**20.37%**

↓  
**11.73%**



Data Lovelace

# Thank You For Your Attention

