00

CHURN FOR BANK CUSTOMER

15/10/2022

Present by 2pendo Group

ABOUT US

01

Rahmawati Glamindia

Nur Amilah

Archie Citra Muhammad

Bernard Kian Yuniantoro

Mercy Eunike

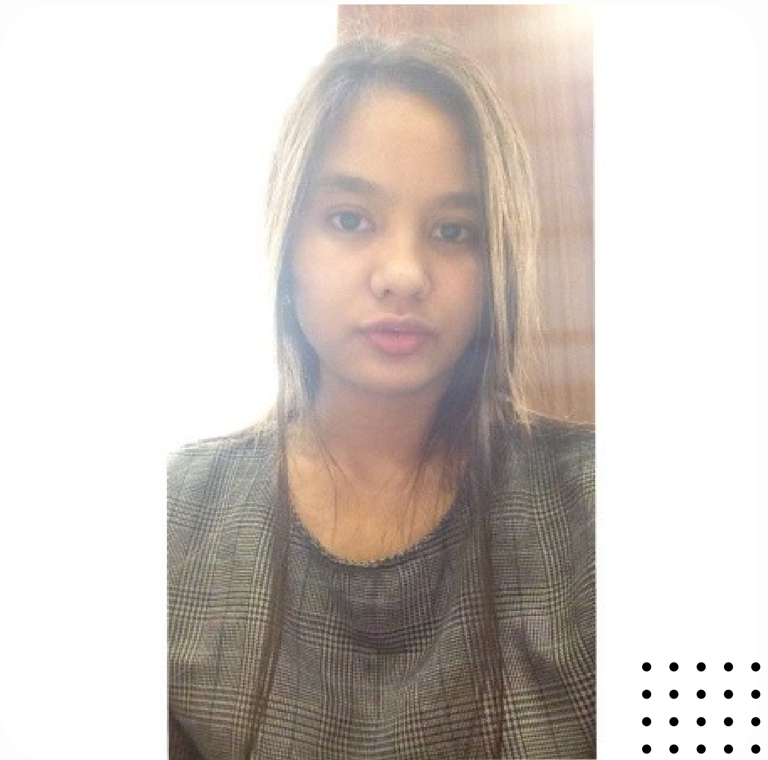


TABLE OF CONTENTS

Presentation Outline

2pendo
Final Project

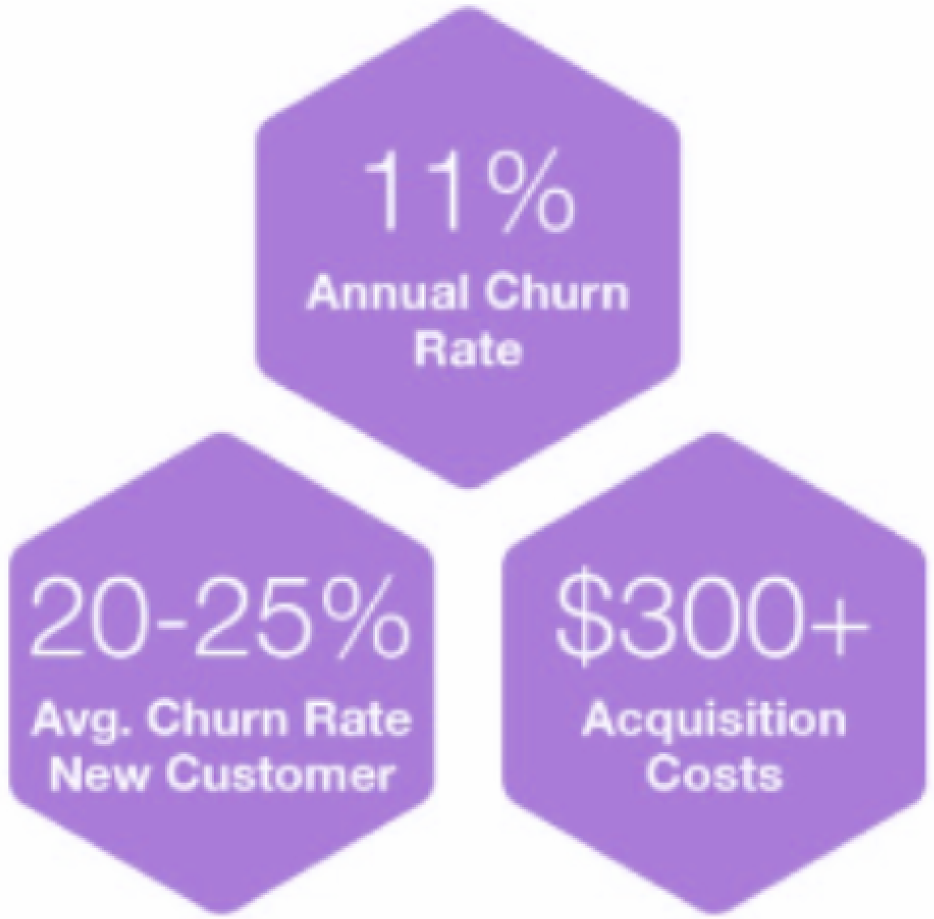
PROBLEM STATEMENT

03

From Dataset

20% CHURN

From External Data



accenture

> **Each point of attrition reflects as much as 2% of net income loss for banks.**
>
> Accenture Banking Customers tahun 2020

https://thefinancialbrand.com/news/bank-marketing/banking-customer-acquisition-attrition-growth-strategy-68371/
https://blueorange.digital/bank-customer-acquisition-three-data-driven-strategies-that-work/

Data Scientist Team

Dataset **Predicting Churn for Bank Customers**
source : **https://www.kaggle.com/datasets/adammaus/predicting-churn-for-bank-customers**

04

G

Goals

- Prediksi nasabah yang akan churn
- Memberikan rekomendasi kepada business team

O

Objective

- Membentuk sebuah model machine learning dengan false negative terkecil
- Mengidentifikasi prediktor/faktor yang berpengaruh terhadap churn rate

M

Business Metrics

Churn Rate (jumlah pelanggan yang hilang dalam suatu periode dibagi dengan jumlah total pelanggan pada awal periode tersebut)

2pendo
Final Project

EXPLORATORY DATA ANALYSIS (EDA)

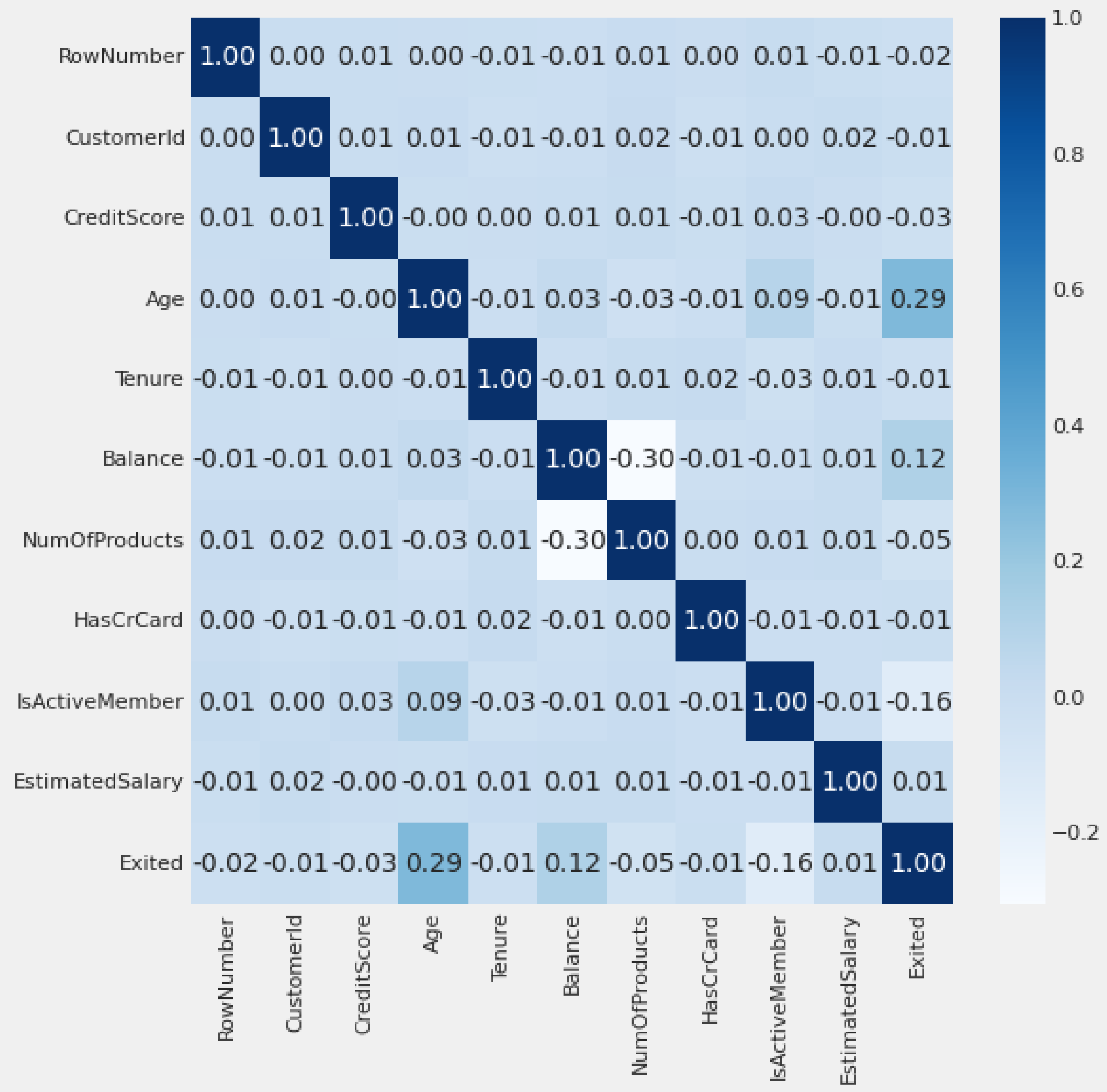
05

Statistic Description

```
[ ]  df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   RowNumber        10000 non-null  int64
 1   CustomerId       10000 non-null  int64
 2   Surname          10000 non-null  object
 3   CreditScore      10000 non-null  int64
 4   Geography        10000 non-null  object
 5   Gender           10000 non-null  object
 6   Age              10000 non-null  int64
 7   Tenure           10000 non-null  int64
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64
 10  HasCrCard        10000 non-null  int64
 11  IsActiveMember   10000 non-null  int64
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
```

Correlation Feature & Target



`Age`
`(0.29)

` IsActiveMember`
`(-0.16)

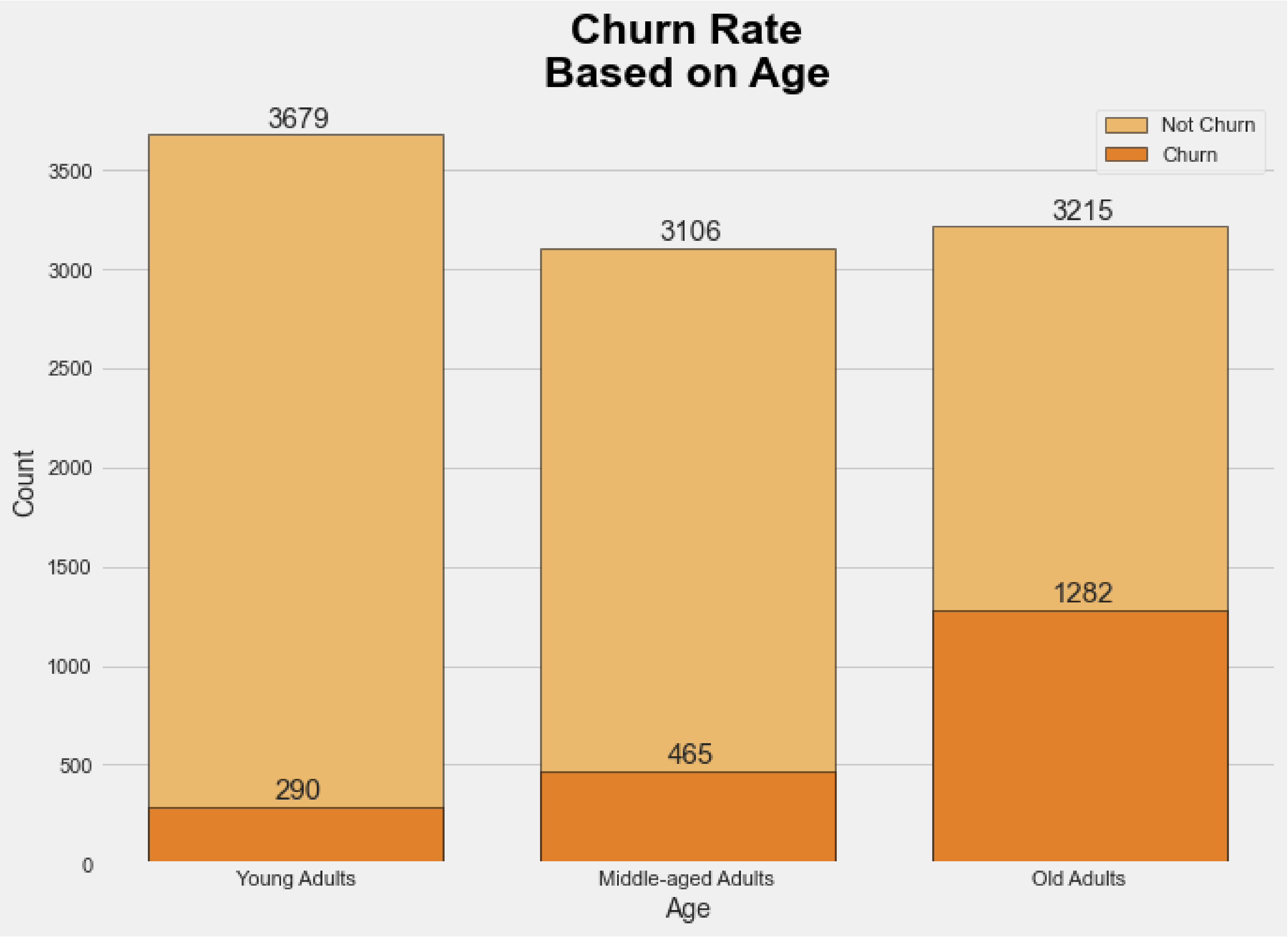
`NumOfProducts`
`(-0.05)

correlation with target **Exited** is worth to be reviewed.
But we will be focus on **churn**

INSIGHT
From Dataset

AGE

06



Insight :
1. \`Age\` Old Aldults (41-90 th) memiliki nilai churn tertinggi sebesar 39.88% dari total nasabah
2. \`Age\` yang memiliki nilai churn terendah adalah Young Adults (17-34 th) dengan persentase churn terendah sebesar 7.88% dari total nasabah

Semakin bertambah \`Age\` maka semakin bertambah churn rate

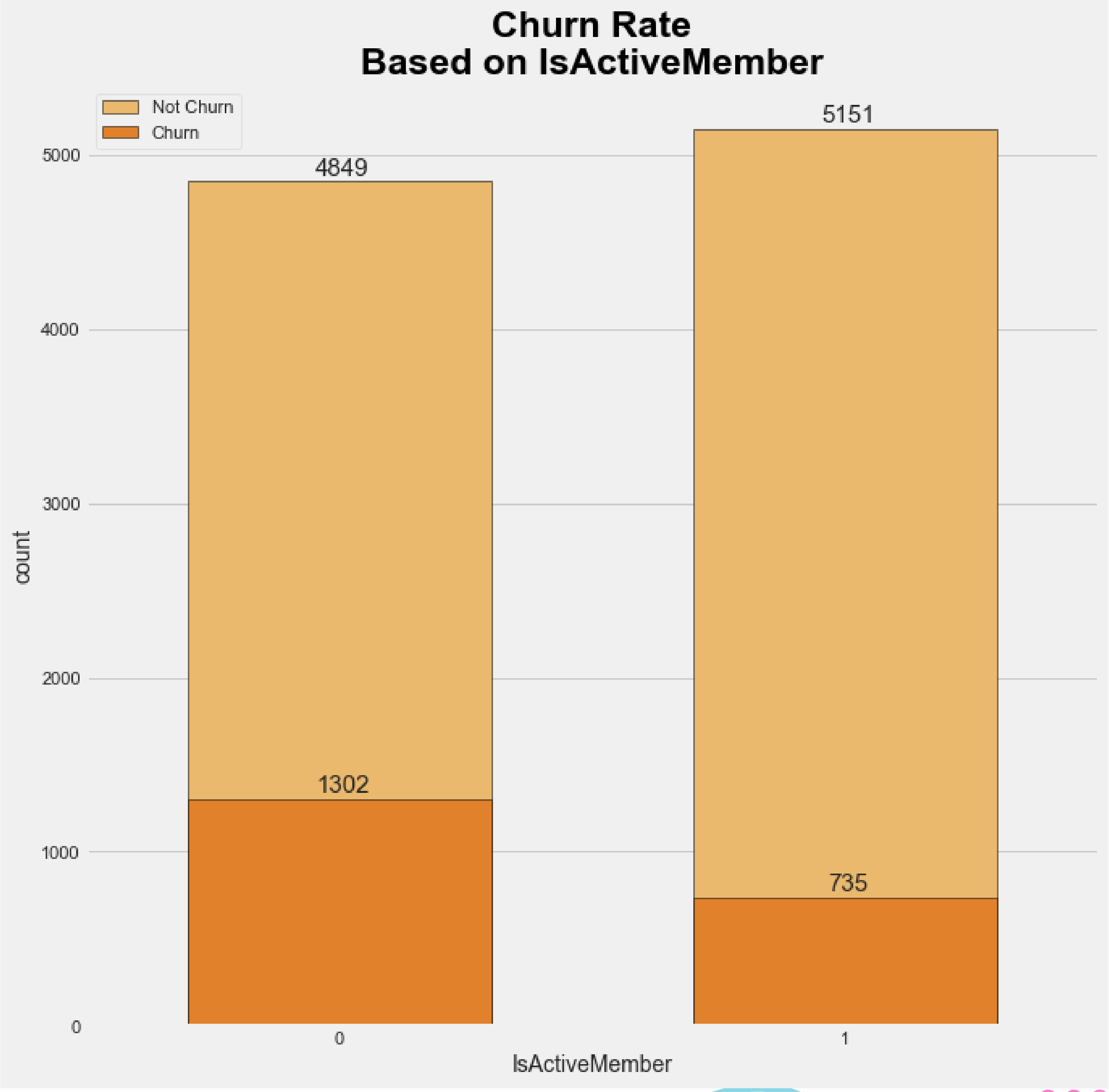
INSIGHT
From Dataset

07

ISACTIVE MEMBER

Insight :

1. `IsActiveMember` yang memiliki nilai churn tertinggi adalah non active member (0) dengan persentase churn sebesar 26.85% dari total nasabah
2. `IsActiveMember` yang memiliki nilai churn terendah adalah active member (1) dengan persentase churn sebesar 14.27% dari total nasabah



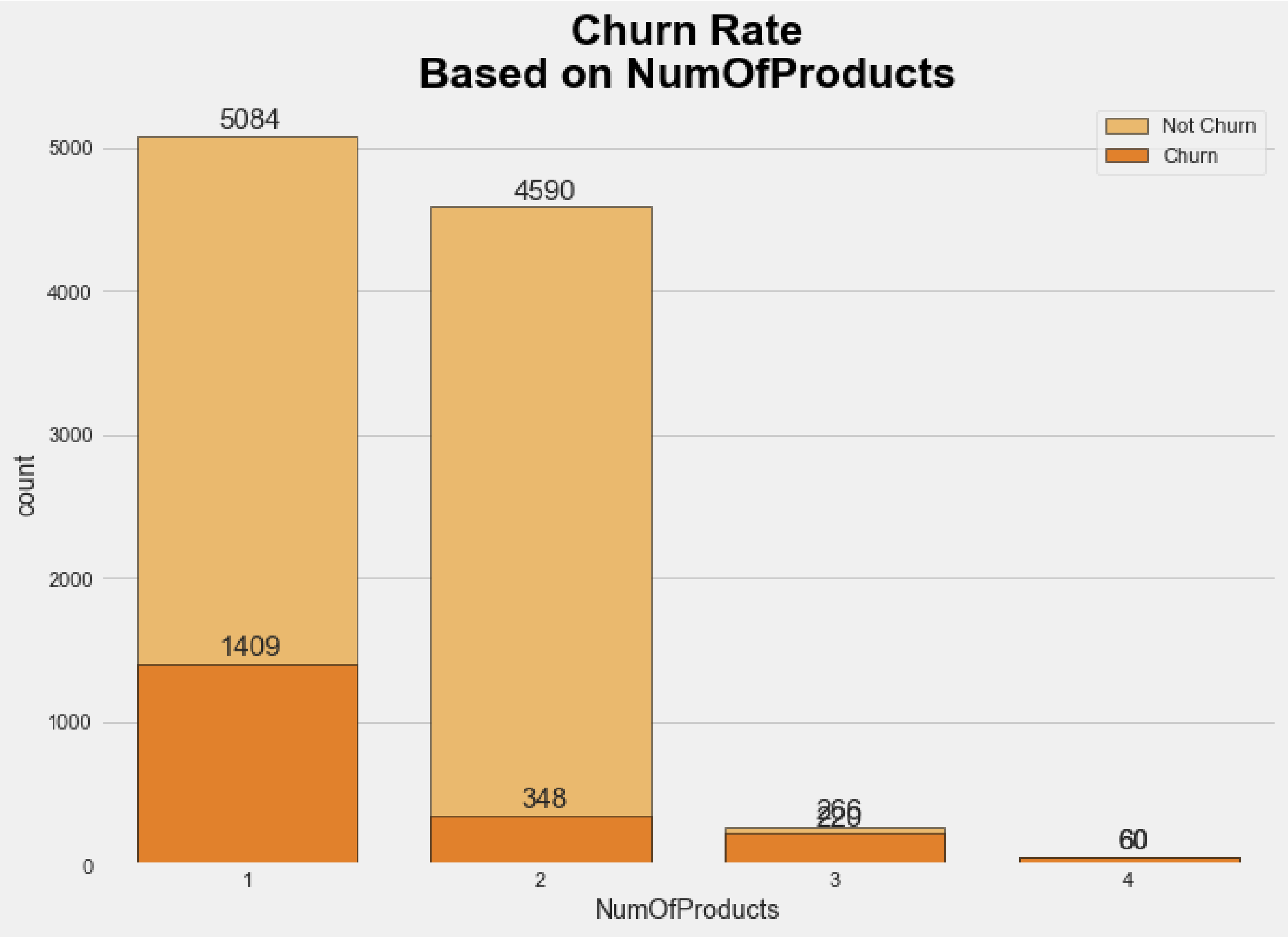
Semakin aktif member semakin rendah churn rate

INSIGHT

From Dataset

08

NUM OF PRODUCTS



Insight :

1. `NumOfProducts` 4 memiliki nilai churn tertinggi sebesar 100% dari total nasabah
2. `NumOfProducts` 2 memiliki nilai churn terendah sebesar 7.58% dari total nasabah

Semakin banyak NumOfProducts semakin besar churn rate

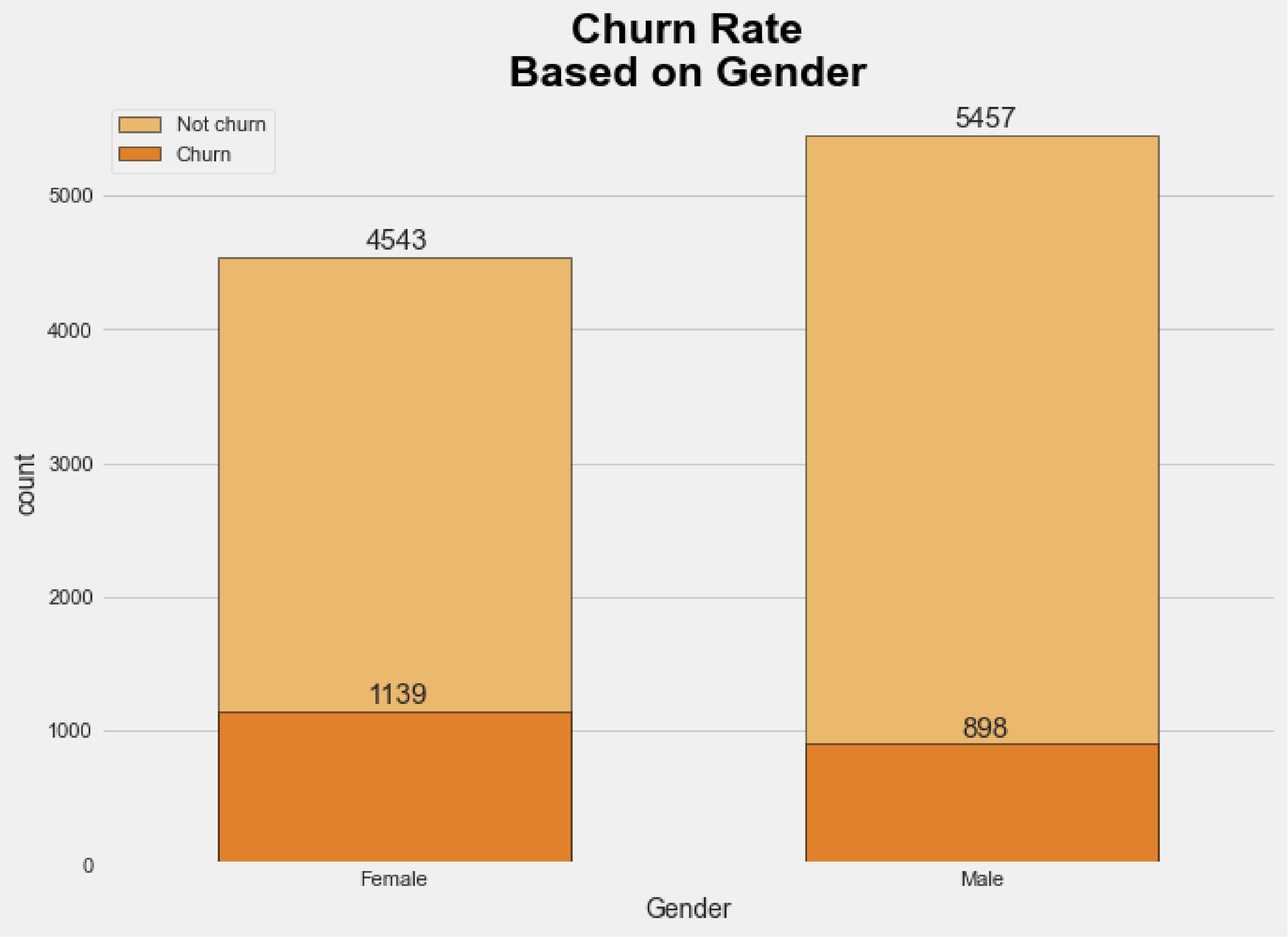
INSIGHT
From Dataset

GENDER

09

Insight :
Gender` yang memiliki churn tertinggi adalah Female dengan persentase churn sebesar 25.07% dari total nasabah

Customer Persona yang memiliki churn terendah adalah Male



INSIGHT

From External Data

THE BOSTON CONSULTING GROUP

> Our research also revealed that the main factors driving attrition aren't always what banks expect. Pricing and market conditions, such as an economic slowdown, are often assumed to be the primary reasons for revenue attrition.

> But our analysis found that 60% to 80% of retention problems stem from service issues, product gaps, and other factors, most of which are internal bank concerns and within the bank's purview to control.

https://www.bcg.com/publications/2017/financial-institutions-marketing-sales-how-banks-close-back-door-attrition

PRE-PROCESSING

01 FEATURE SELECTION

- Missing Data
- Duplicated Data

02 SPLIT DATA

- Split data
- Oversampling

03 OUTLIERS HANDLING

- Log Transform
- Z-score Method

04 FEATURE TRANSFORM

- StandardScaler

05 FEATURE ENCODING

- One Hot Encoding

06 DROP COLUMN OUTDATED

Drop column : Age, Geo, Gender

DATA MODELLING

DECISION TREE

# Hyperparameter Tuning
max_features = ['auto', 'sqrt', 'log2', 'none']
criterion = ['gini', 'entropy']

RANDOM FOREST

Hyperparameter Tuning
criterion': ['entropy', 'gini']
'max_features': ['auto', 'sqrt'],

LOGISTIC REGRESSION

# Hyperparameter Tuning
penalty = ['l1','l2','none']
C = [int(x) for x in np.linspace(0.0001, 1)]
solver = ['newton-cg', 'saga']

SVC

# Hyperparameter Tuning
C = [0, 1]
kernel = ['rbf']
gamma = ['scale']

XGBOOST

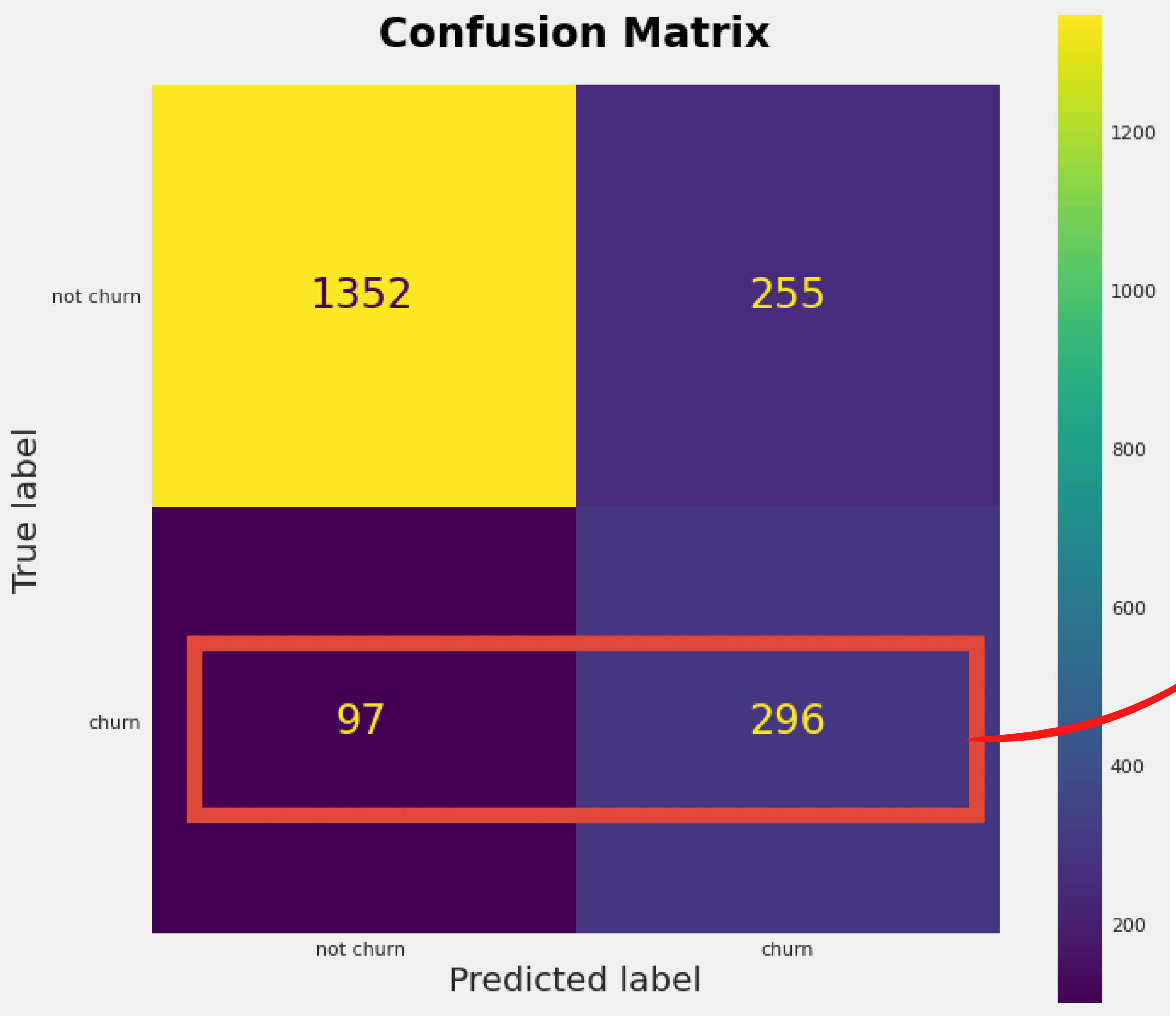
Hyperparameter Tuning
'gamma': [0],
'learning_rate': [0.01],
'max_depth': [3],

EVALUATION MODELS

13

AUC score data test

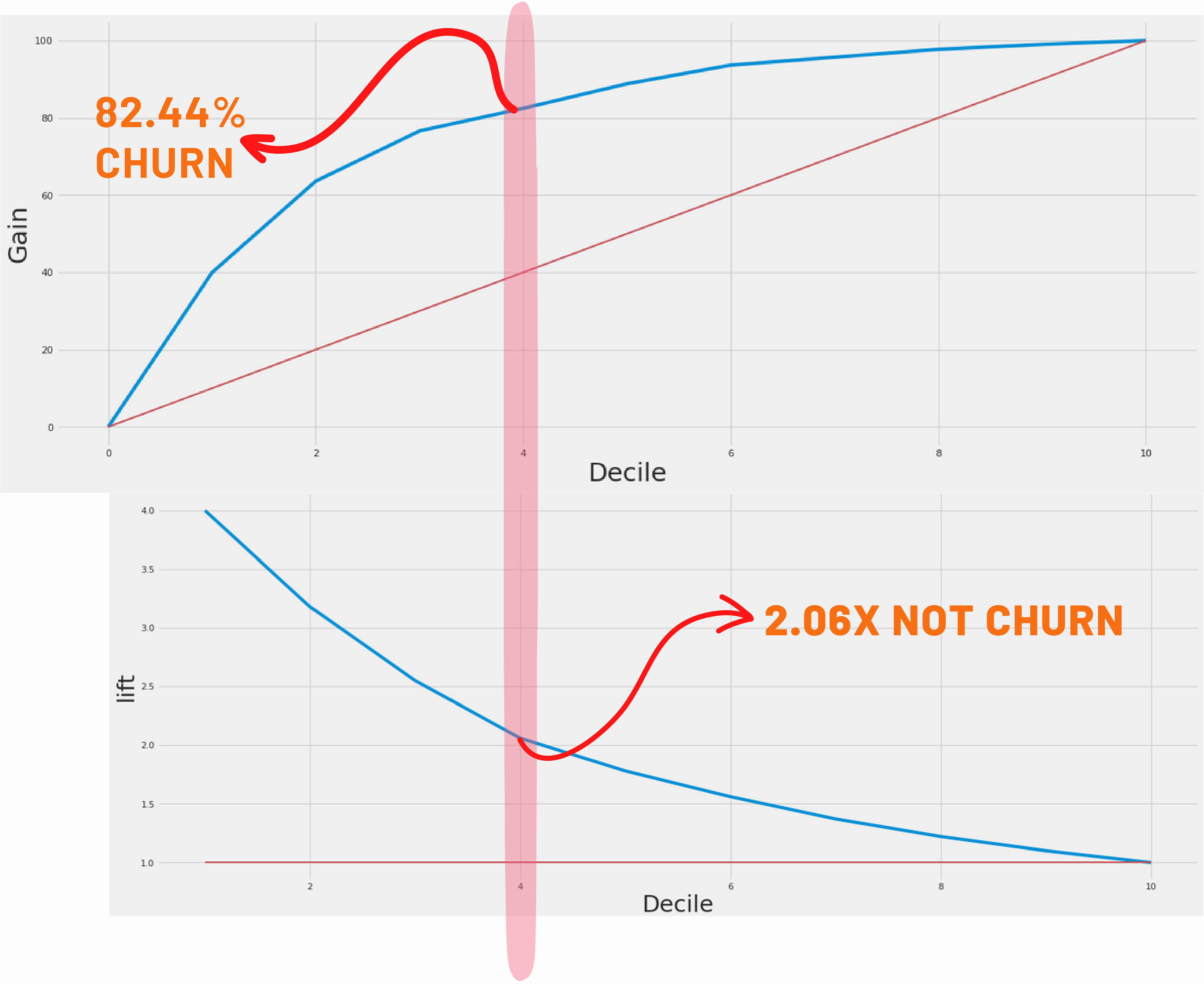
Models	AUC Score	
	No SMOTE	SMOTE
Logistic Regression	0.78	0.78
Decision Tree	0.81	0.75
SVC	0.82	0.85
Random Forest	0.85	0.85
XGBboost	0.87	0.86



75%

PREDIKSI BENAR

EVALUATION MODEL WITH GAIN & LIFT



40% NEW CUSTOMER

https://towardsdatascience.com/model-benefit-evaluation-with-lift-and-gain-analysis-4b69f9288ab3

Calculation

Without Model

- Acquisition Cost (new customer)
  $300 * (2000) = $ 600,000.00

With Model

- Saving from Acquisition Cost
  $300 * (2000*60%) = $ 360,000.00
- Add potensial revenue bank from treatment customer churn
  $150 * (2000*40%*82%* 75% *70%) = $ 51,600.00
- Total Saving with model--------------> $411,600.00
- Percentage Saving with model
  $411,600/$ 600,000 *100 = 69%

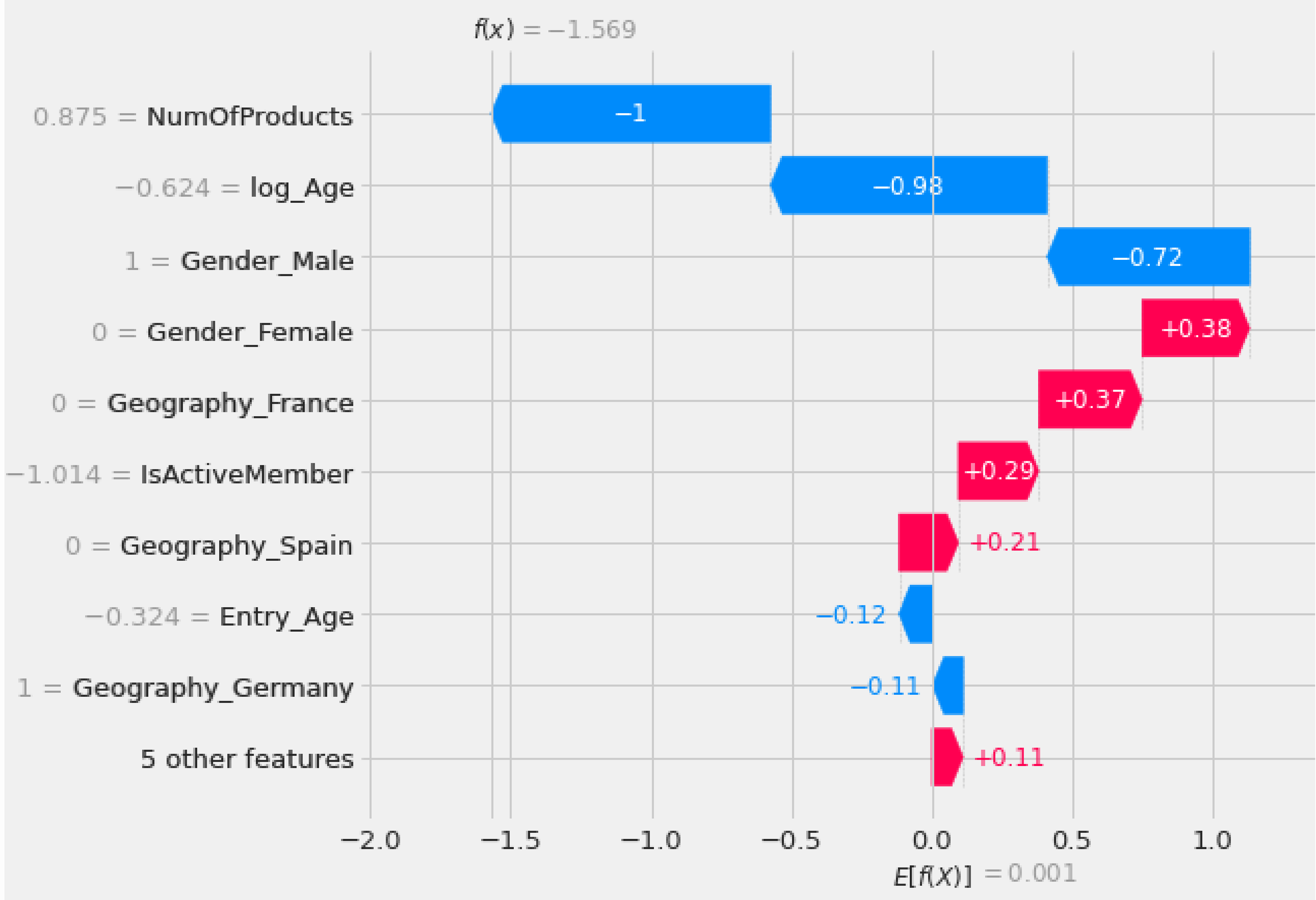
https://blueorange.digital/bank-customer-acquisition-three-data-driven-strategies-that-work/
https://www.reviewtrackers.com/blog/bank-customer-retention/

2pendo
Final Project

FEATURE IMPORTANCE WITH SHAP
(SHAPLEY ADDITIVE EXPLANATIONS)

Globe SHAP

Local SHAP



Conclusion

Log_Age, NumOfProduct, Gender_Male, Geography_France, IsActiveMember have big contribution in predict

https://medium.com/dataman-in-ai/explain-your-model-with-the-shap-values-bc36aac4de3d

RECOMMENDATION FROM DATA VIZ AND SHAP

1

2

Num Of Products

Encouraging customers to use max 2 products such as savings and mobile banking

Age

Targeting campaigns, products and services aged 17-34 years old

Gender & Geography

Creating campaigns, products and services that fit with Gender Male and Geography France.

3

4

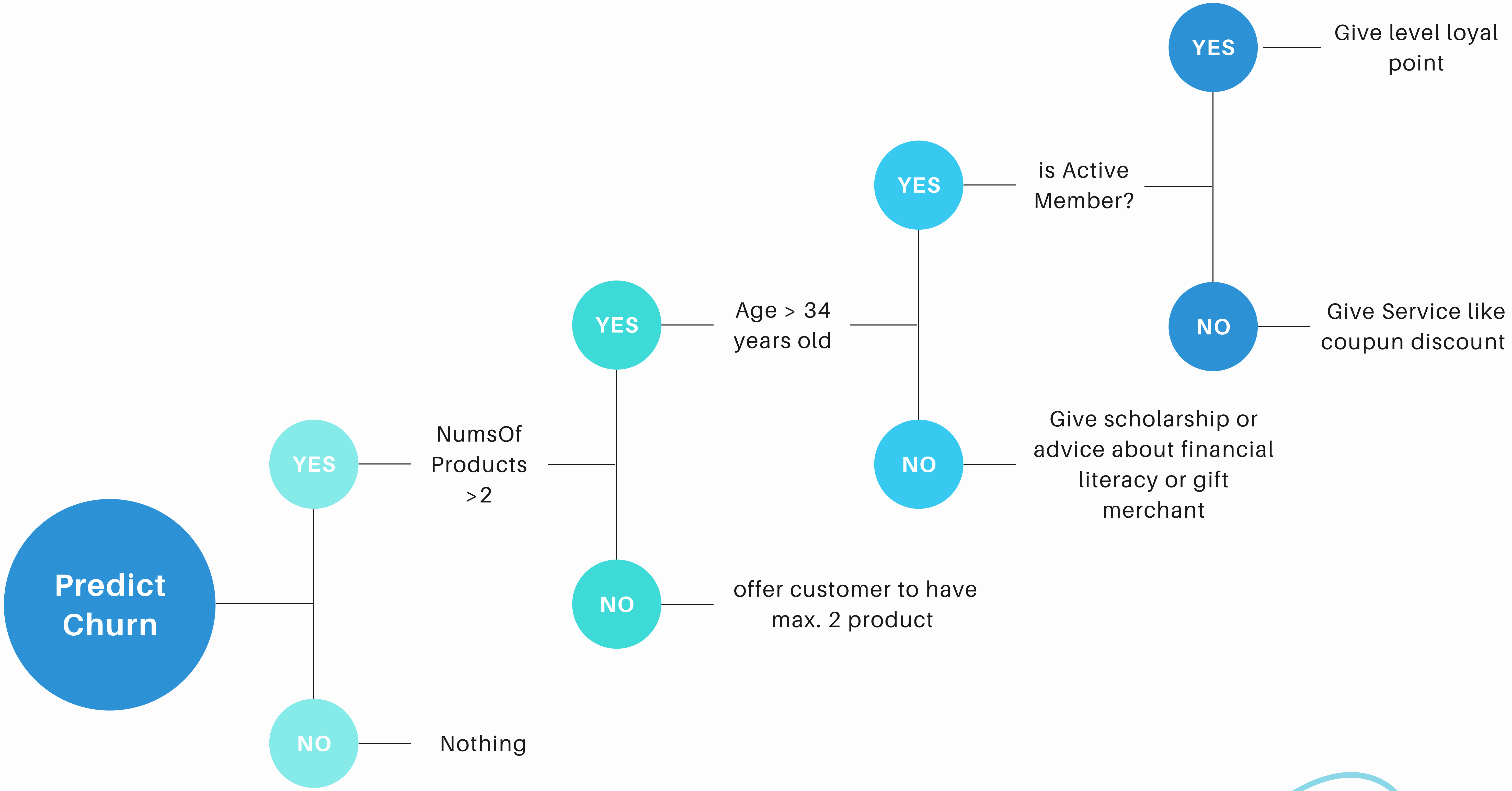
Active Member

Provide forex trading services, stock exchange, crypto market or online shopping discount coupons on mobile banking

https://www.bcg.com/publications/2017/financial-institutions-marketing-sales-how-banks-close-back-door-attrition
https://www.fpsc.com/BOC_email/November2017.htm
https://explodingtopics.com/blog/customer-retention-rates

CUSTOMER CHURN TREATMENT

2pendo
Final Project



https://document360.com/blog/reduce-customer-churn
https://towardsdatascience.com/swaying-the-persuadables-predict-marketing-uplift-in-python-c5e9fd855b80

2pendo
Final Project

CONCLUSION

18

Customer indicated churn :

- Customers who have more than 2 products
- Customers over 34 years old
- Customers with inactive members
- Gender Female

Recommendation business team :

- Give scholarship or advice about financial literacy or gift merchant
- Give Service like coupun discount

Evaluation Model:

- Selected model : XGBoost dengan SMOTE, AUC score = 0.86
- Recall from model : 75%
- Percentage Saving cost with model : 69%

2pendo Group

THANK YOU!

PEMBAGIAN TUGAS

Nur Amilah

STAGE 0
STAGE 1 (Insight)
STAGE 2
STAGE 3 (Model Machine Learning)

- SVM

Archie Citra Muhammad

STAGE 0
STAGE 1 (EDA)
STAGE 2
STAGE 3 (Model Machine Learning)

- XGBoost

Rahmawati Glamindia

STAGE 0
STAGE 1 (Visualisasi)
STAGE 2
STAGE 3 (Model Machine Learning)

- Random Forest

Mercy Eunike

STAGE 0
STAGE 1 (Insight)
STAGE 2
STAGE 3 (Model Machine Learning)

- Logistic Regression

Bernard Kian Yuniantoro

STAGE 0
STAGE 1 (Visualisasi)
STAGE 2
STAGE 3 (Model Machine Learning)

- Decision Tree

All & Mentor

Diskusi setiap yang di kerjakan dan di ambil kesimpulan mana yang tepat dari pengerjaan masing - masing individu