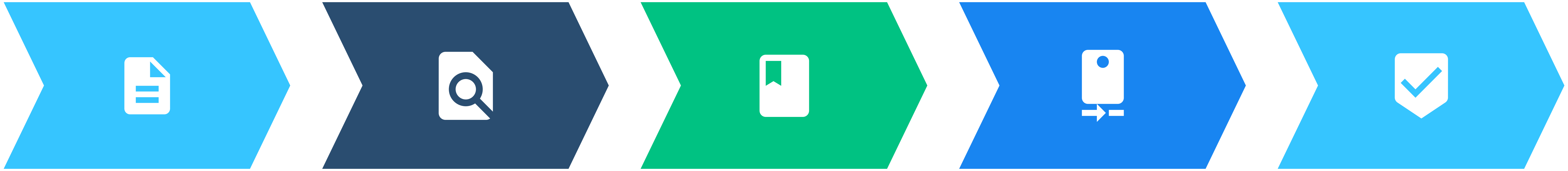
FIKA HUSNA AMALINA MUBAROK

Banking Analyze Marketing Campaign

OUTLINE PROJECT



Problem Statement

EDA

Machine Learning Model

Model Simulation

Recommendation

Problem Statement

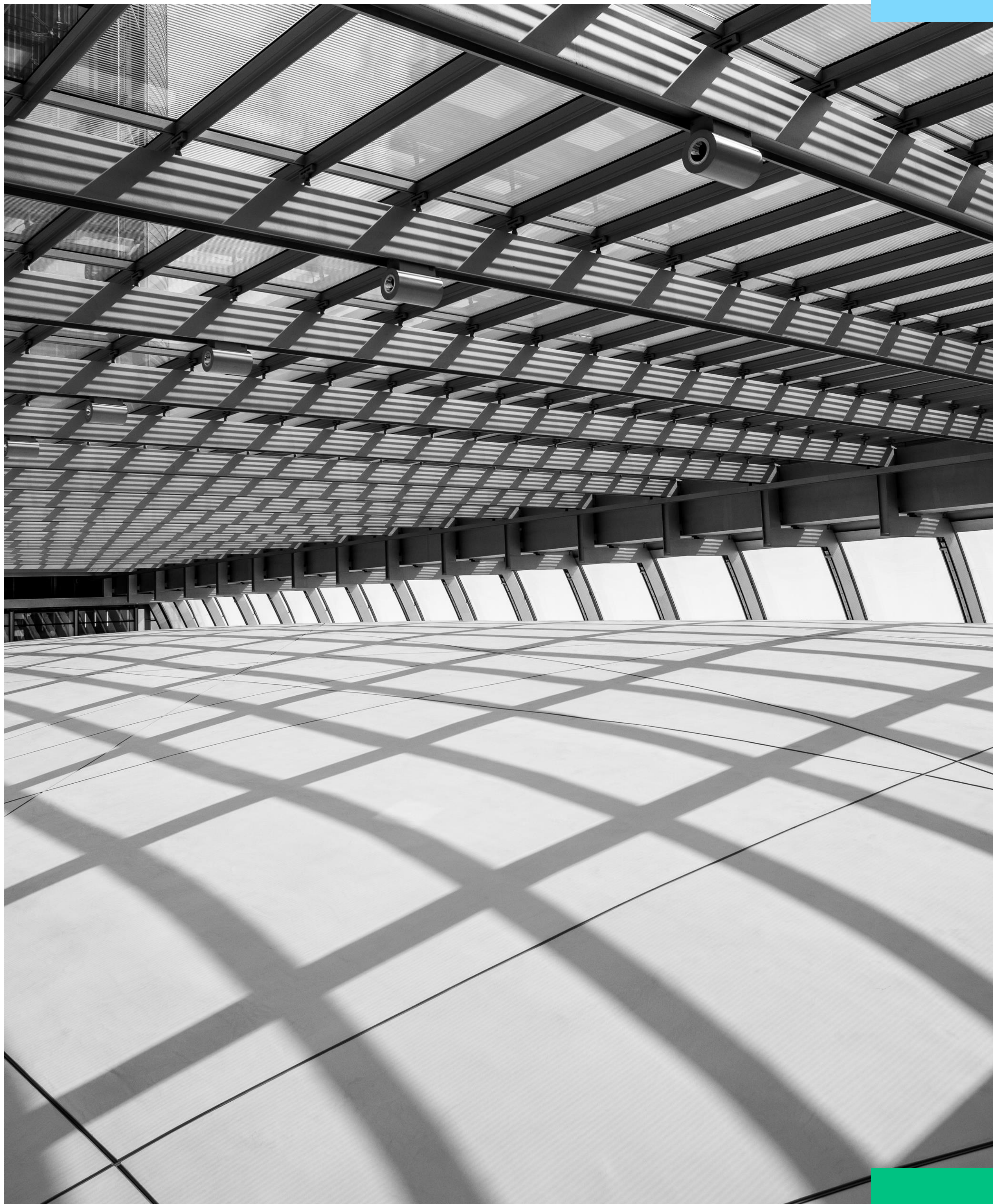
Through the Marketing Campaign strategy, this bank will provide time deposit products (by telephone).

Excellent campaign for businesses offering time deposits to customers over the phone.

Customers who approve deposits have not been maximized, accounting for only 11.7%, implying that promotions have been ineffective in attracting customers.

Goals

Predicting customers who are likely to receive campaigns so that the marketing team can be more effective and efficient in campaign execution.



Objective

Create a **Machine Learning** model that can predict which potential customers will receive the campaign..

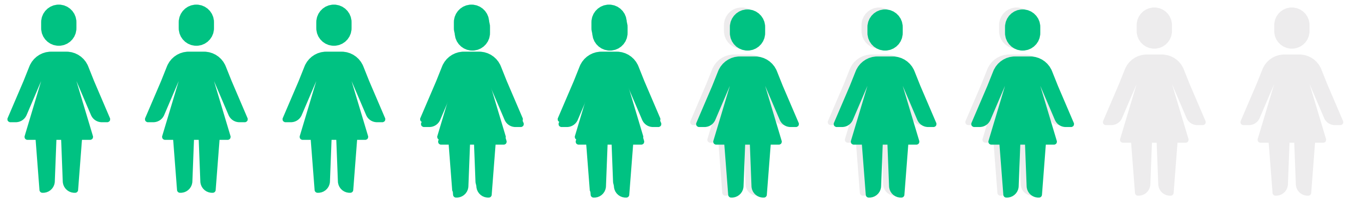
Finding important factors and main characteristics of customers who accept the campaign and agree to the deposit products offered.

Make **recommendations** to improve the company's effectiveness in delivering campaigns to specific customers.

Exploratory Data Analysis

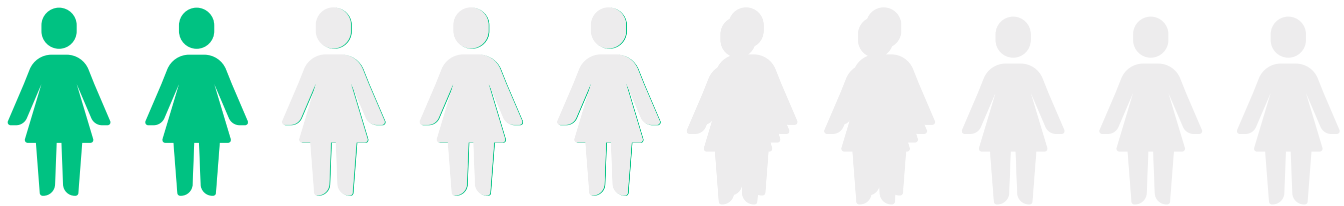
Data Overview

Customer Refuses to Make Time Deposit



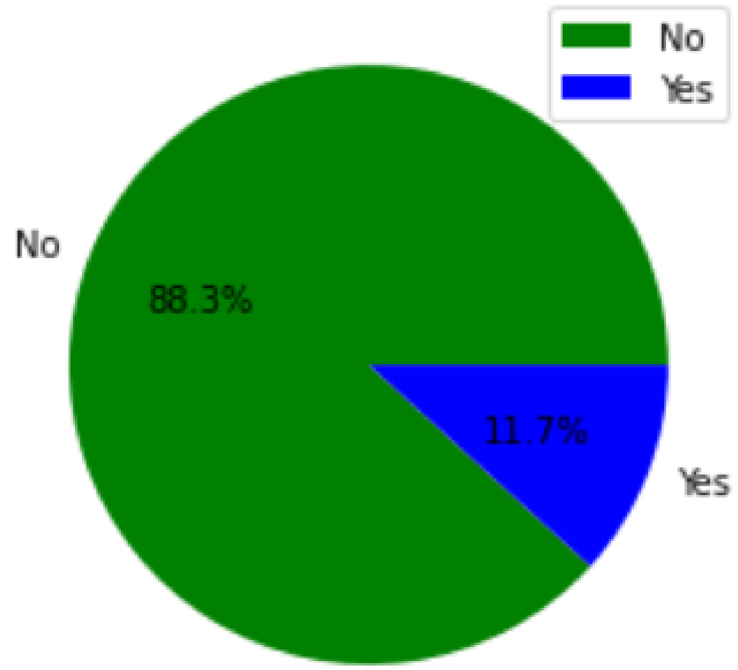
88%

Customer Approve to Make Time Deposit



12%

Baseline

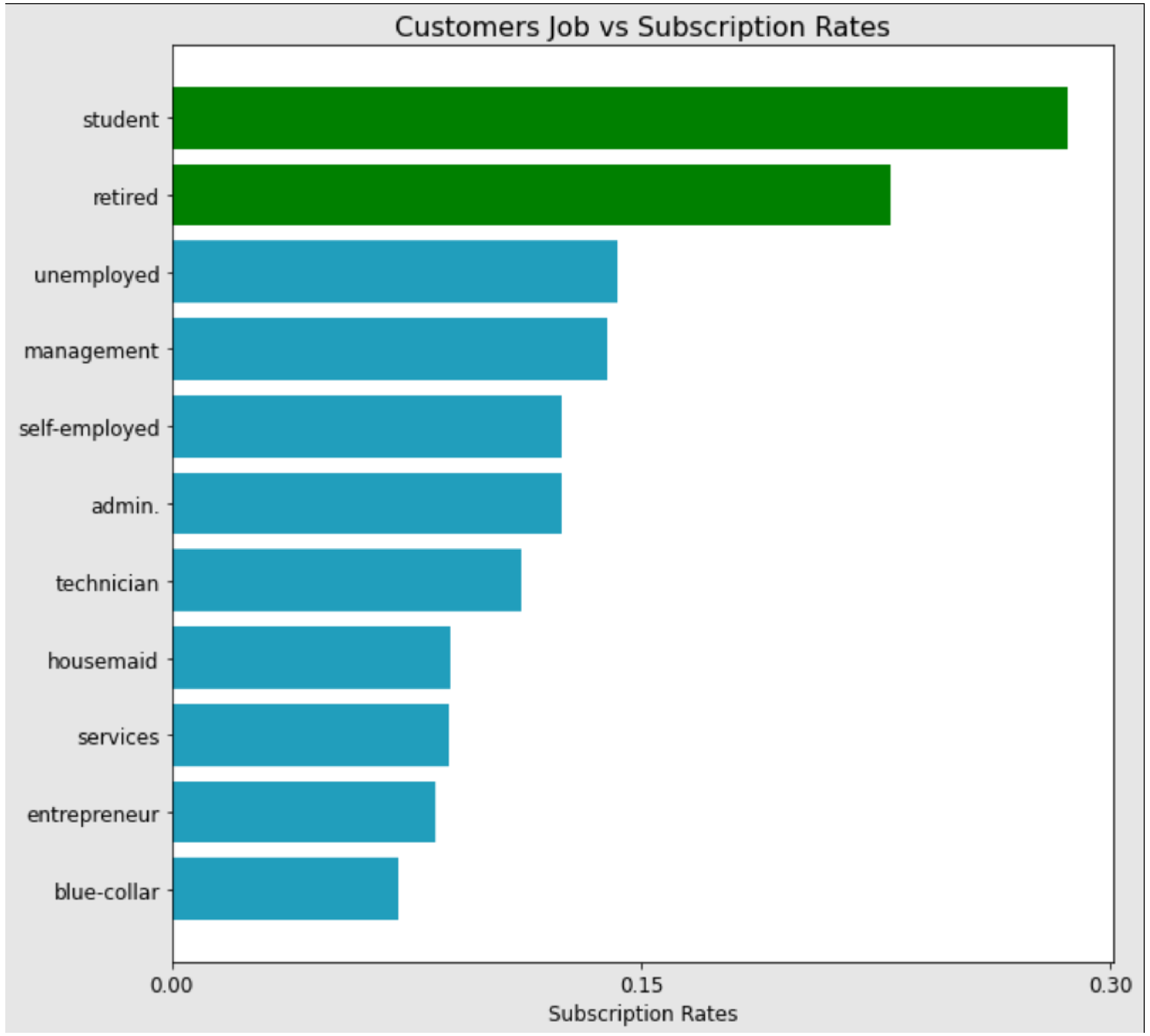


Data Understanding.

1. **Data set consists 31647 rows and 17 Columns**
2. **No Null data**
3. **Have a target of “Yes” (Costumer Approve to Make Time Deposit) and “No” (Customer Refuses to Make Time Deposit.**

Column Name	Description
age	Age of customers
Job	Type of Job
Marital	Marital Status
Education	Type of education customers
Default	Has credit or not ?
Balance	Average yearly balance, in euros
Housing	Has housing loan ?
Loan	Has personal loan ?
Contact	Type of contact communication
Day	Last contact day of the month
Month	Last contact mount of year
Duration	Last contact duration, in seconds
Campaign	number of contacts performed during this campaign and for this client
pdays	number of days that passed by after the client was last contacted from a previous campaign
previous	number of contacts performed before this campaign and for this client
poutcome	outcome of the previous marketing campaign
Subscribed	has the client subscribed a term deposit?

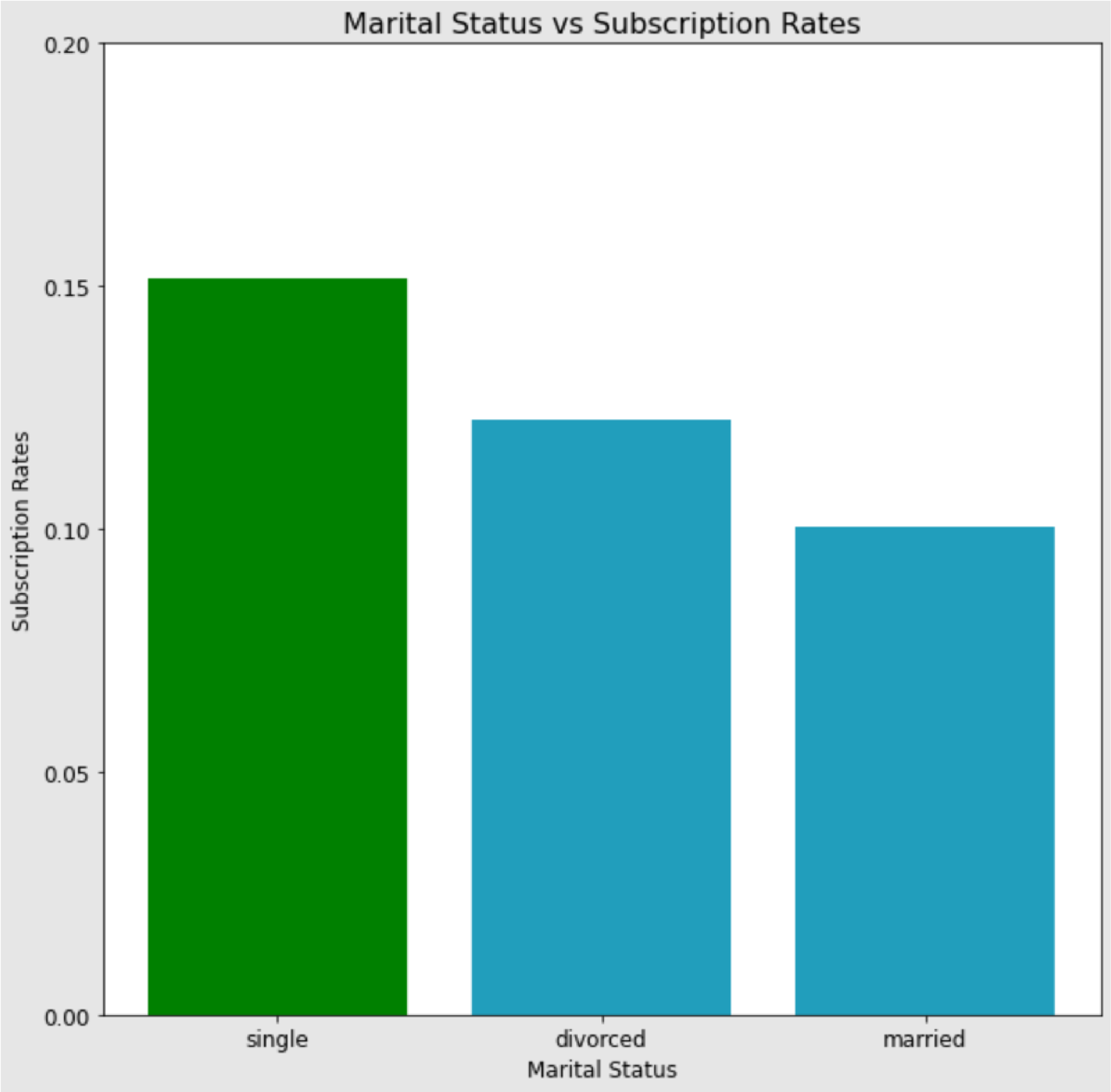
Job vs Subscription Rates



Students and retirees have a high subscription rate based on the type of work that they do, which is 0.28 for students and 0.22 for retirees.

Job	Subscribed
Student	0.286614
Retired	0.229987
unemployed	0.142541
management	0.138997
self-employed	0.124666
admin	0.124484
technician	0.111928
housemaid	0.088913
services	0.088415
entrepreneur	0.084325
blue-collar	0.072489

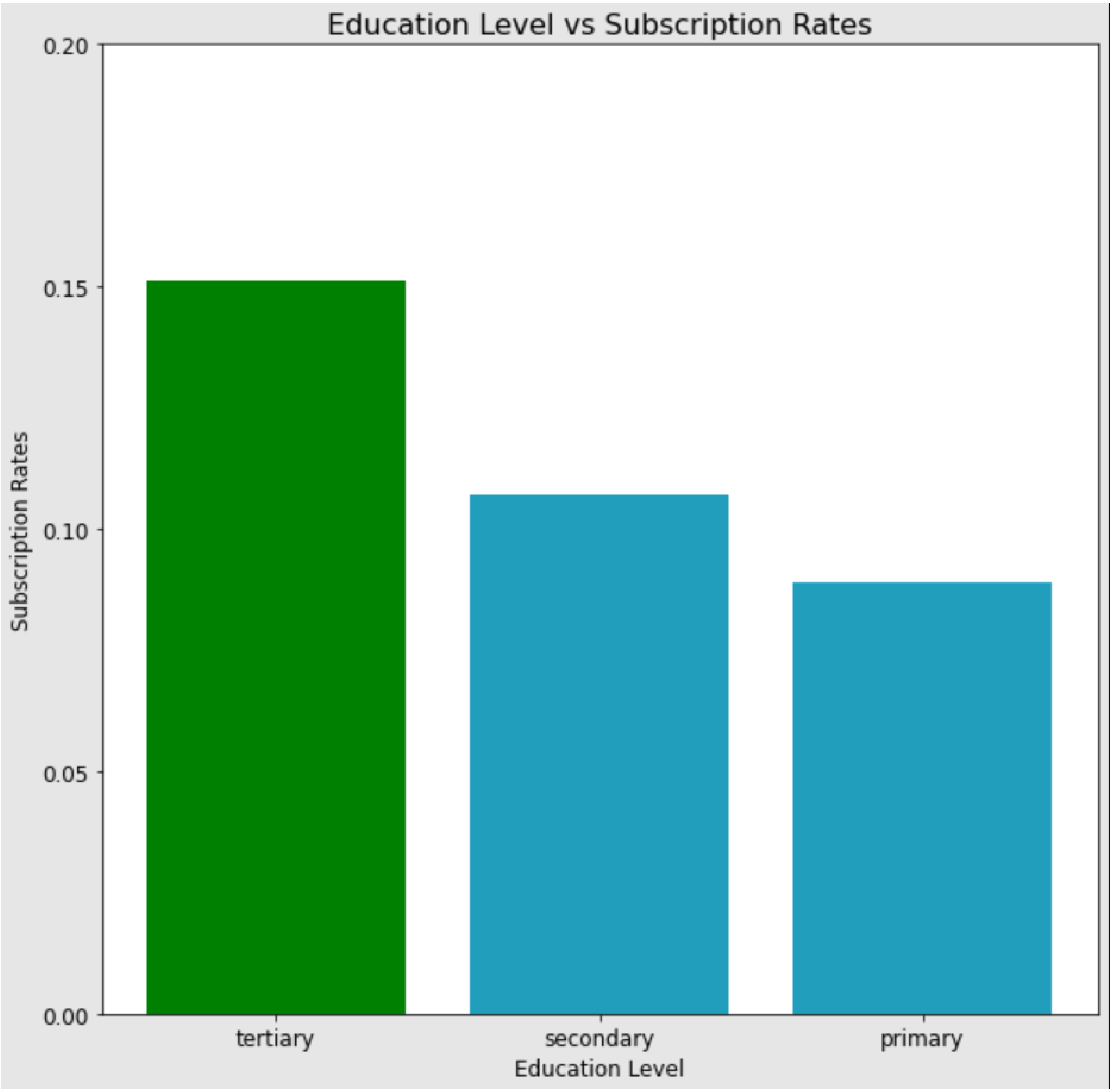
Marital Status vs Subscription Rates



Among the 31647 customers, those who are still single have a subscription rate of 0.15, which is 20% higher than those who are divorced and 33% higher than those who are married.

Marital	Subscribed
Single	0.151423
Divorced	0.122590
Married	0.100498

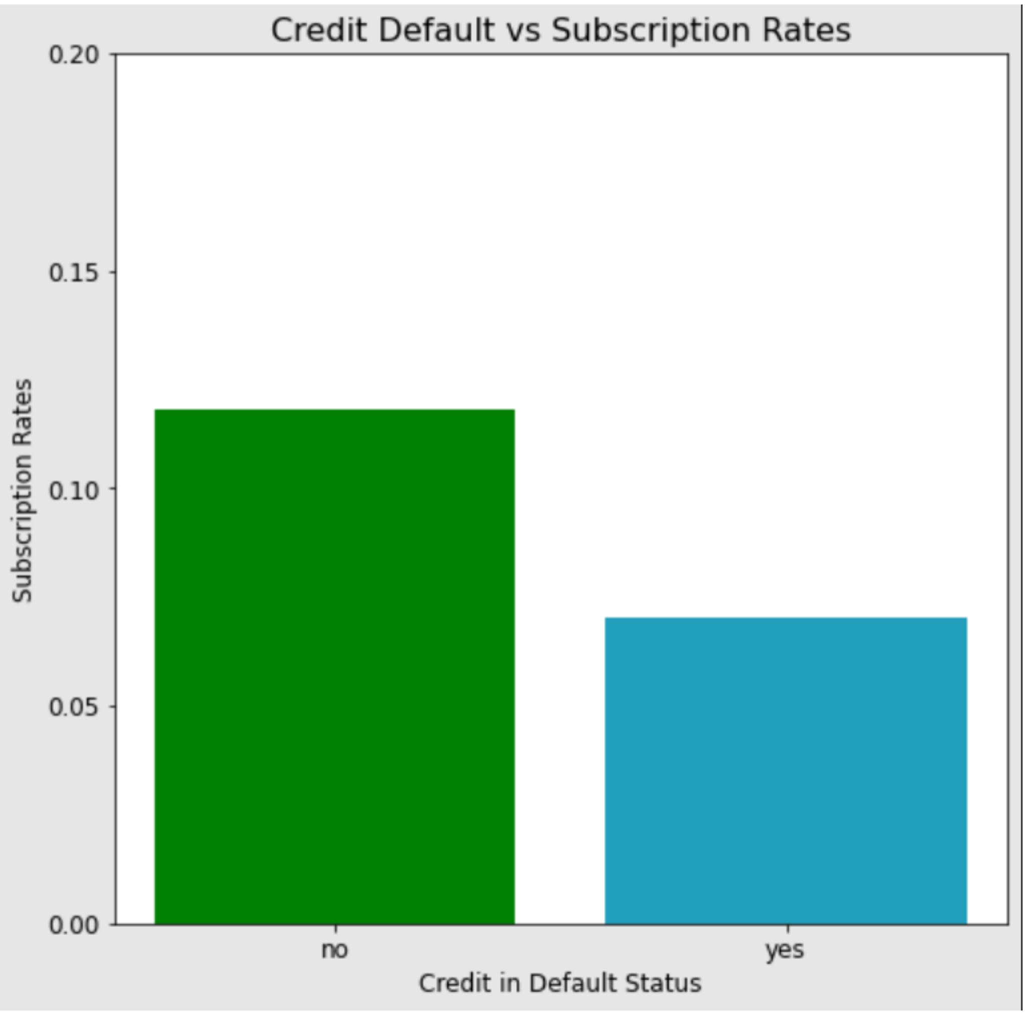
Education vs Subscription Rates



Customers with tertiary education levels have the highest subscription rate of 0.15 if we classify them by education level.

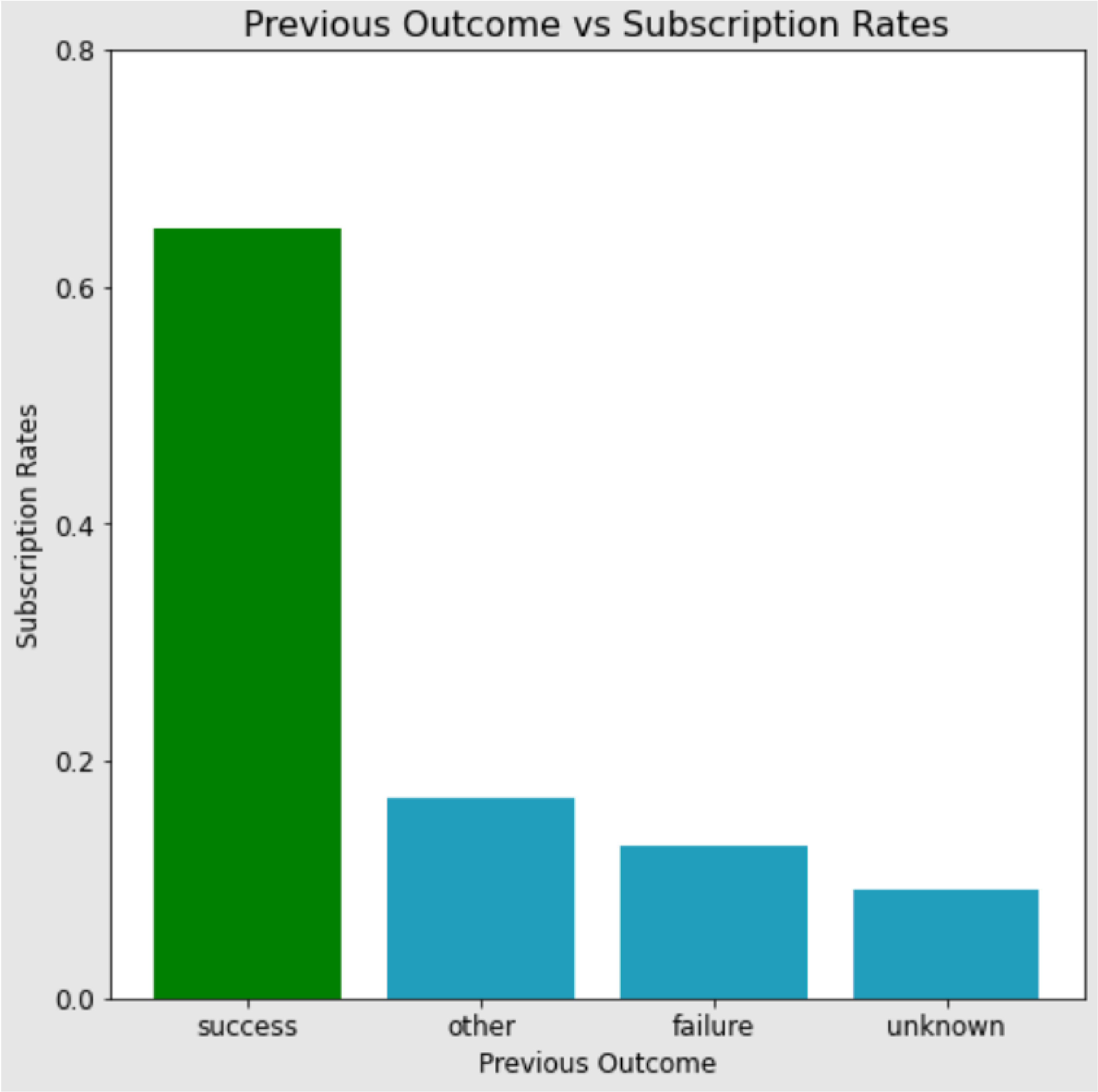
Education	Subscribed
Tertiary	0.151017
Secondary	0.106916
Primary	0.088810

Loan, Housing, Credit default vs Subscription Rates



When it comes to subscription rates, Loan, Housing, and Credit Default all have something in common. Customers who have outstanding debts with the bank are more likely to reject marketing offers to open a deposit account.

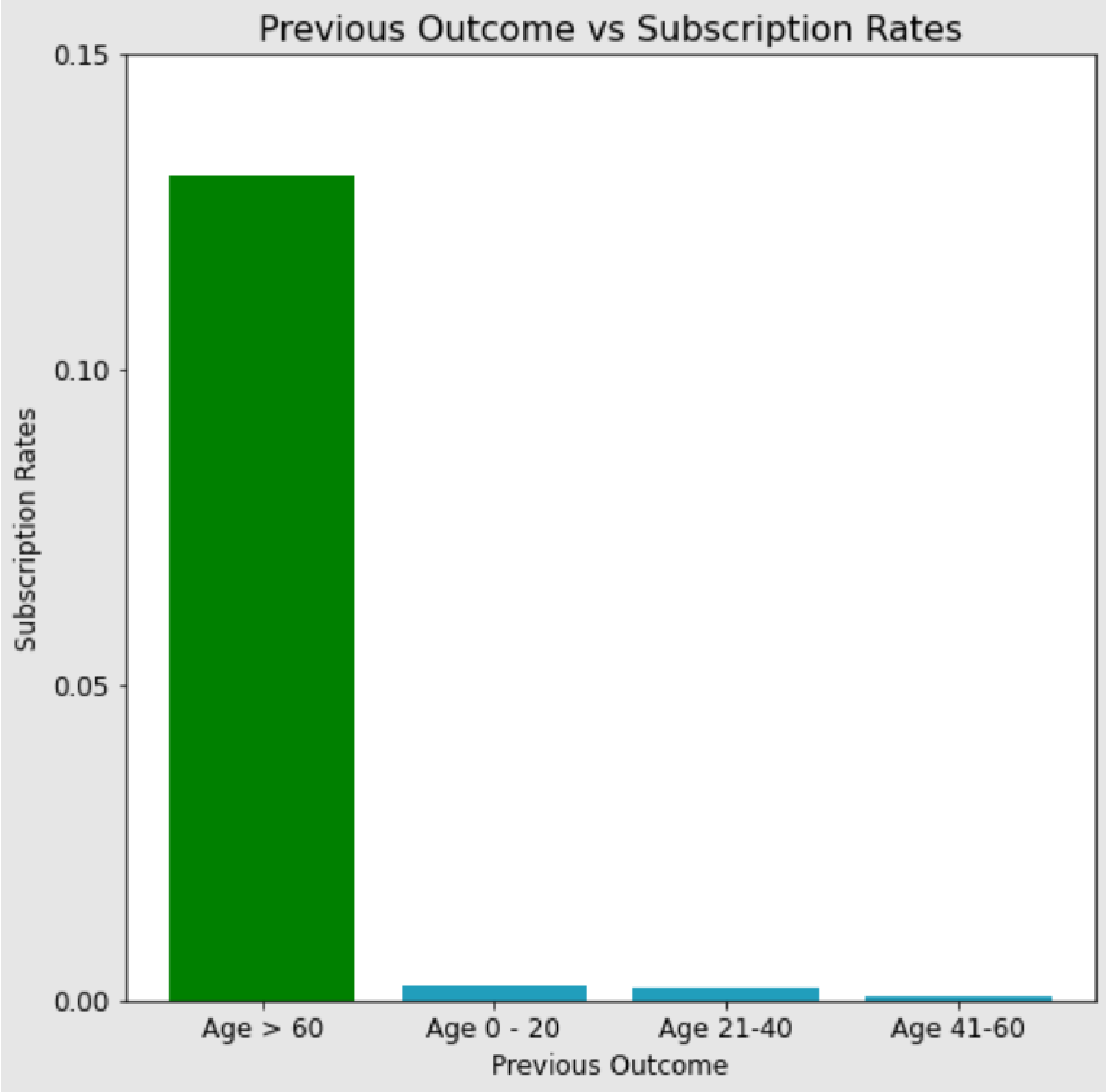
Previous outcome vs Subscription Rates



The previous campaign had a subscription success rate of 0.64, which is a significant difference of 48% when compared to the others.

Poutcome	Subscribed
Success	0.649813
Other	0.168578
Failure	0.128198
Unknown	0.091519

Previous outcome vs Subscription Rates by Age



Customer age is divided into four groups, and we can see a difference in subscription rates for customers over 60 years old based on the time spent by telemarketing making calls, which is 206 770 minutes or 3446 hours.

age_category	Subscribed
>60	0.130611
0-20	0.002478
21-40	0.002022
41-60	0.000692

age_category	durasi_telfon
>60	206770
0-20	1446
21-40	989
41-60	807

Data Preprocessing

Feature Engineering

- Label Encoding
- One-Hot Encoding

Handling Imbalanced data
With over.SMOTE

Split Data
Train 80 % dan Test 20 %

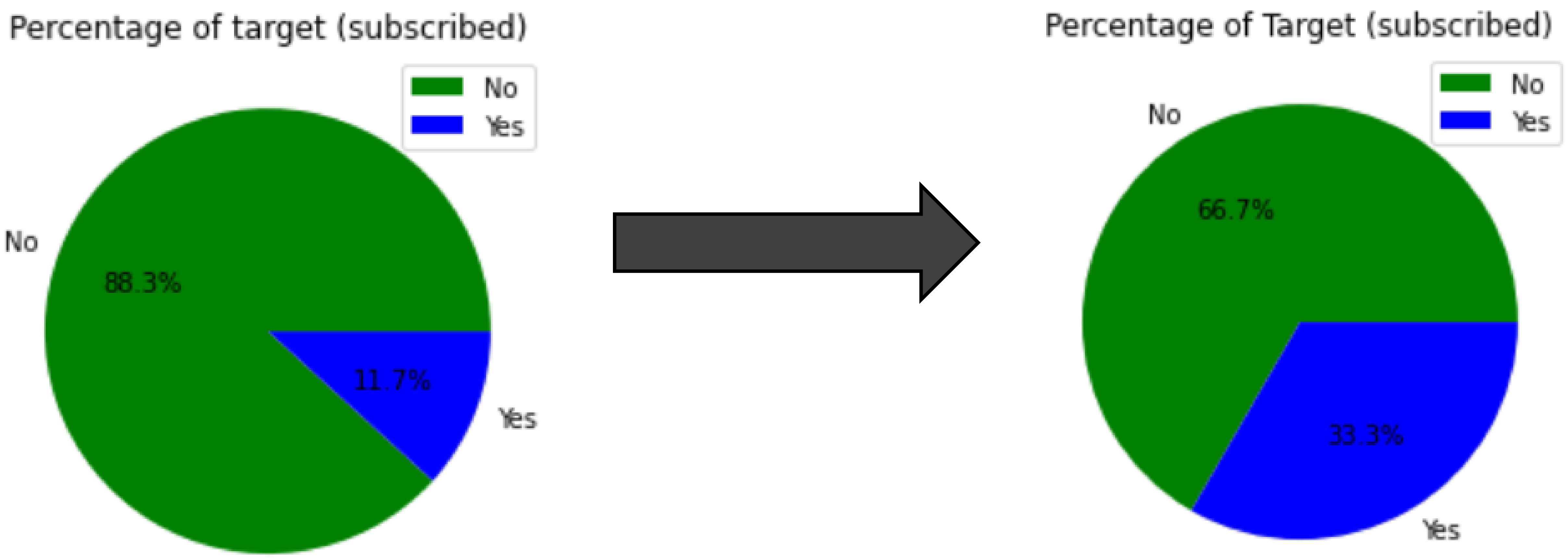
In this project, I divide the data set in an 80:20 ratio, dividing it into 80 for Train and 20 for Test, then dividing it into 16% validation data and 64% train data. This project employs four classifications: Decision Tree, KNN, Nave Bayes, and Random Forest.

Handling Data

First, I did not handle data imbalances .

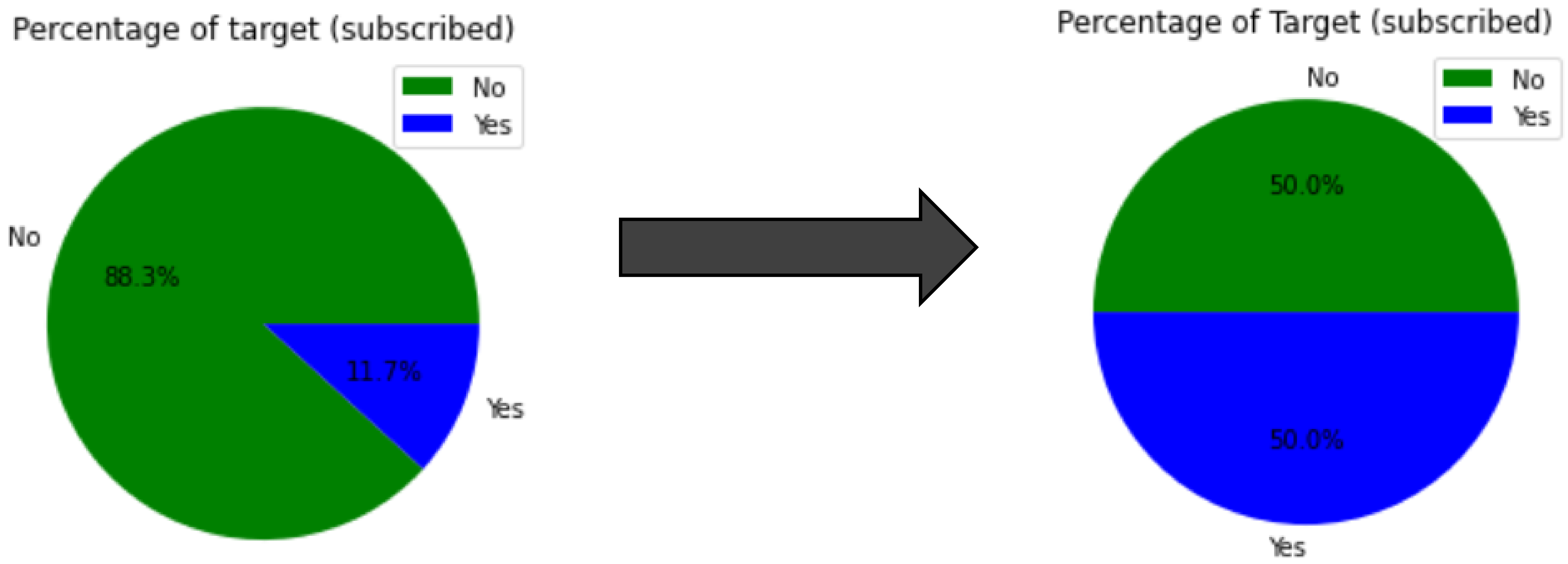
Handling Data

Target imbalance should be addressed first in order to improve Machine Learning performance.
UnderSampling is what I use.



Handling Data

Target imbalance should be addressed first in order to improve Machine Learning performance.
Over.SMOTE is what I use.



Data Encoding

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31647 entries, 0 to 31646
Data columns (total 18 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   ID         31647 non-null  int64
 1   age        31647 non-null  int64
 2   job        31647 non-null  object
 3   marital    31647 non-null  object
 4   education  31647 non-null  object
 5   default    31647 non-null  object
 6   balance    31647 non-null  int64
 7   housing    31647 non-null  object
 8   loan       31647 non-null  object
 9   contact    31647 non-null  object
 10  day        31647 non-null  int64
 11  month      31647 non-null  object
 12  duration   31647 non-null  int64
 13  campaign   31647 non-null  int64
 14  pdays      31647 non-null  int64
 15  previous   31647 non-null  int64
 16  poutcome   31647 non-null  object
 17  subscribed 31647 non-null  object
dtypes: int64(8), object(10)
memory usage: 4.3+ MB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55864 entries, 0 to 55863
Data columns (total 20 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   age                  55864 non-null  int64
 1   default              55864 non-null  int64
 2   balance              55864 non-null  int64
 3   housing              55864 non-null  int64
 4   loan                 55864 non-null  int64
 5   duration             55864 non-null  int64
 6   campaign             55864 non-null  int64
 7   pdays                55864 non-null  float64
 8   previous             55864 non-null  int64
 9   marital_divorced     55864 non-null  uint8
 10  marital_married      55864 non-null  uint8
 11  marital_single       55864 non-null  uint8
 12  education_primary    55864 non-null  uint8
 13  education_secondary  55864 non-null  uint8
 14  education_tertiary   55864 non-null  uint8
 15  poutcome_failure     55864 non-null  uint8
 16  poutcome_other       55864 non-null  uint8
 17  poutcome_success     55864 non-null  uint8
 18  poutcome_unknown     55864 non-null  uint8
 19  subscribed           55864 non-null  int64
dtypes: float64(1), int64(9), uint8(10)
memory usage: 4.8 MB
```

I label encode some features (yes and no) and use one hot encoding for features with multiple values; the one on the right is the result of over. SMOTE .

Machine Learning Model

Model Evaluation

Actual

Prediction

	Tidak Setuju	**Setuju**
Tidak Setuju	True Negative	**False Negative**
Setuju	**False Positive**	True Positive

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2 \text{ x } (\text{Precision x Recall})}{\text{Precision + Recall}}$$

F1 Score

Considering the value of Precision and Recall in order to reduce the risk of the value of **False Positive** (Prediction agrees but refuses) and **False Negative** (Prediction refuses, but has a high potential to agree).

Modeling Result (Not Handling)

Model	Precision		Recall		F1 Score		Auc	
	Train	Test	Train	Test	Train	Test	Train	Test
Decission Tree (Base Model)	-	0.36	-	0.41	-	0.38	-	0.65
Decission Tree	0.41	0.39	0.43	0.41	0.42	0.41	0.67	0.67
KNN	0.42	0.42	0.22	0.21	0.29	0.28	0.59	0.58
Naïve Bayes	0.41	0.40	0.41	0.46	0.41	0.43	0.67	0.68
Random Forest	0.60	0.60	0.36	0.38	0.45	0.46	0.66	0.67
Random Forest (After Tuning)	0.80	0.79	0.15	0.15	0.25	0.25	0.57	0.57

Modeling Result (UnderSampling)

Model	Precision		Recall		F1 Score		Auc	
	Train	Test	Train	Test	Train	Test	Train	Test
Decission Tree (Base Model)	-	0.36	-	0.40	-	0.38	-	0.65
Decission Tree	0.63	0.62	0.66	0.64	0.65	0.63	0.73	0.72
KNN	0.64	0.60	0.50	0.50	0.56	0.55	0.67	0.67
Naïve Bayes	0.69	0.68	0.52	0.54	0.59	0.60	0.70	0.71
Random Forest	0.72	0.72	0.70	0.72	0.71	0.72	0.78	0.79
Random Forest (After Tuning)	0.77	0.76	0.60	0.59	0.67	0.66	0.75	0.75

Modeling Result (Over.Smote)

Model	Precision		Recall		F1 Score		Auc	
	Train	Test	Train	Test	Train	Test	Train	Test
Decission Tree (Base Model)	-	0.37	-	0.41	-	0.39	-	0.66
Decission Tree	0.90	0.90	0.91	0.91	0.90	0.91	0.90	0.91
KNN	0.78	0.78	0.92	0.91	0.84	0.84	0.83	0.83
Naïve Bayes	0.74	0.74	0.88	0.88	0.80	0.81	0.79	0.79
Random Forest	0.94	**0.94**	0.92	**0.92**	0.93	**0.93**	0.93	**0.93**
Random Forest (After Tuning)	0.89	0.89	0.84	0.84	0.87	0.86	0.87	0.87

Cross Validation (F1 Score)

Model	Train			Test			Val		
	1	2	3	1	2	3	1	2	3
Decission Tree (Base Model)	-	-	-	-	-	-	-	-	-
Decission Tree	0.89	0.90	0.89	0.87	0.88	0.88	0.88	0.87	0.87
KNN	0.80	0.81	0.80	0.73	0.74	0.76	0.74	0.75	0.75
Naïve Bayes	0.78	0.78	0.81	0.82	0.81	0.81	0.81	0.80	0.82
Random Forest	**0.93**	**0.93**	**0.93**	**0.92**	**0.92**	**0.92**	**0.91**	**0.91**	**0.91**
Random Forest (After Tuning)	0.88	0.87	0.87	0.87	0.87	0.86	0.87	0.87	0.87

Random Forest Model (Over.Smote)

Model	Precision		Recall		F1 Score		Auc	
	Train	Test	Train	Test	Train	Test	Train	Test
Random Forest	0.94	0.94	0.92	0.92	0.93	0.93	0.93	0.93

The F1 score produced by the model test results using test data is **93%.**

Cross Val (F1 Score)

Model	Train			Test			Val		
	1	2	3	1	2	3	1	2	3
Random Forest	0.93	0.93	0.93	0.92	0.92	0.92	0.91	0.91	0.91

Actual			
	Tidak Setuju	**5.310**	293
	Setuju	416	**5154**
		Tidak Setuju	**Setuju**
		Prediction	

Random Forest Model (Over.Smote)

Actual

0	**5.310**	293
1	416	**5154**
	0	**1**

Prediction

1. **<u>F1- Score </u>**: Considering the value of Precision and Recall in order to reduce the risk of the value of **False Positive** (Prediction agrees but refuses) and **False Negative** (Prediction refuses, but has a high potential to agree).

2. **<u>Roc-Auc</u>** : Machine learning models accurately distinguish between 1 and 0 predictions. If the ROC and AUC are large, the prediction accuracy can be calculated from the confusion matrix prediction results.

3.**Confusion Matrix :**
True Negative : I believe the prediction is based on 0 and the actual 0 the actual number of subscribers is 5310.
True Positive : I believe the prediction is based on 1 and the actual 1 the actual number of subscribers is 5154.

Random Forest Model (Over.Smote)

Actual		Prediction 0	Prediction 1
	0	**5.310**	293
	1	416	**5154**
		0	**1**

Prediction

1. Customers assume a deposit of 10 million, which means that with **TP** 5154, they will receive 10 million x 5154 customers = 51.54 billion.

2. **TN** will not deposit because there is no additional money in the bank.

3. **FP** : we predict this person will put a deposit but the fact is not. If the assumption is that 1 person is 10 million, then in FP there are 293 customers, as a result it should be the bank balance gets 10 million x 293 customers = 2.93 billion but in fact there is no balance (loss).

4. **FN**: we predict he will not deposit but in fact he will deposit if the assumption is 1 customer is 10 million x 416 customer = 4.16 billion.

At the very least, we **get 51.54 billion**, and at the very most, we get **51.54 billion + 4.16 billion = 55.70 billion**.

Business Simulation

Business Problem

Bank

Deposito

Random Customer

Marketing

- Just 11,7 % = 3.715 Customer

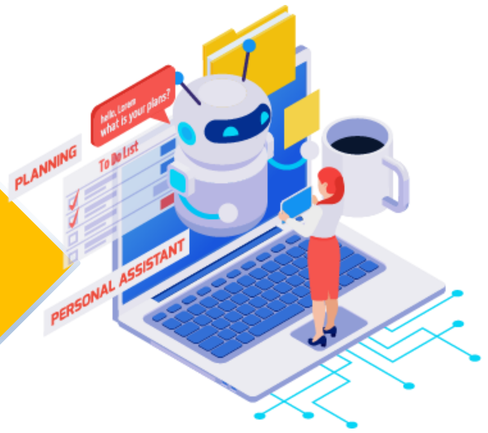
Business Simulation

<u>Machine Learning Model</u>

- Selected Customer
- Best Recommendation

Bank

Deposito



Marketing

- Up to 93 % = 29.431 Customer

Business Recomendation

Business recommendation

Some investment platforms typically offer cashback to those who subscribe to time deposits; we may do the same for student/student customers. It will have a positive financial impact. It is also necessary to attract a large number of new customers by offering special cash back promotions or discounts.

This bank is **making offers to customers** who do not have loans, housing, or credit defaults.

Targeting the duration of the call to approach the customer, because it has a strong correlation with customer attraction success**.**

Customers should be avoided if their decisions are unclear or limited**.**

-

THANKS

Email : fikahusnaamalinamubarok@gmail.com

Linkedin : www.linkedin.com/in/fikahusnaamalinamubarok

FIKA HUSNA AMALINA MUBAROK