



Term Deposit Conversion Rate Prediction & Analysis

Banking Telephonic Marketing Case.

Pegasus Data Scientist Team

Final Project Presentation – Batch 18 Rakamin Academy





Outline

Introduction



Data Overview,
EDA



Model
Evaluation



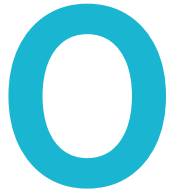
Business
Understanding



Data Cleaning
& Preparation



Business Insight &
Recommendation



O

Introduction



Who are we?

Pegasus adalah tim data science di **Bank PTG** yang bertanggung jawab untuk menyelesaikan permasalahan pada perusahaan yang berhubungan dengan data





The Pegasus Team

Ramadhan Adi

Lead Team

Muhammad Firman A

Member

Bagaskara TA

Member

Puguh Santosa D.

Member

Gerry Chandra

Mentor

Aldin Islamy

Member

Nadiana Yemin M.

Member

Dien Galuh

Member

1

Business Understanding



Background

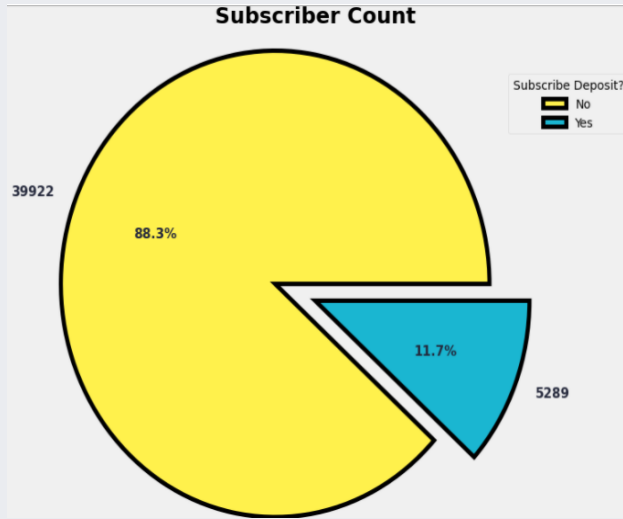
Produk deposit berjangka merupakan salah satu produk perbankan di Bank PTG yang memberikan **kontribusi besar** terhadap pendapatan bank.

Saat ini untuk menawarkan produk, metode pemasaran melalui telepon (**telemarketing**) merupakan **cara yang paling efektif** untuk menarik pelanggan.

Sayangnya, **biaya untuk metode pemasaran ini cukup tinggi**, sedangkan berdasarkan data yang ada, **conversion rate** nasabah yang mengambil produk deposito masih **relatif kecil**.



Let's check the data..



$$\text{Conversion Rate} = \frac{\text{Subscribed}}{\text{Total Call}}$$

Reference [1]: <https://www.investopedia.com/terms/t/termdeposit.asp> [2]: <https://www.nngroup.com/articles/conversion-rates/>

What is a term deposit?

A **term deposit** is a fixed-term investment that includes the deposit of money into an account at a financial institution[1].

What is a conversion rate?

The **conversion rate** is the percentage of users who take a desired action[2].

- **Dari periode 2008-2010 dengan total 45.211 data**, *conversion rate* terhadap deposit sebesar **11.7%**.



Problem Statement

Rendahnya conversion rate yaitu **11.7%**, sedangkan **biaya untuk telemarketing cukup tinggi**.

Goal

Nasabah yang ditargetkan berhasil menjadi nasabah deposito berjangka, dengan tetap mengoptimalkan marketing cost.

Objectives

- Membuat model *Machine Learning* untuk memprediksi nasabah yang berpotensi mengambil produk deposito berjangka.
- Memberikan *insight* dari dataset yang ada dengan menampilkan visualisasi.

Business Metrics

- Conversion Rate
- Telephone Marketing Cost

2

Data Overview, EDA



Data Overview

No	Feature	Description
1	Age	Age of customer
2	Job	Type of job
3	Marital	Marital Status of customer
4	Education	Education level
5	Default	Whether customer has credit in default
6	Balance	Average yearly balance (in euros)
7	Housing	Whether customer has housing loan
8	Loan	Whether customer has personal loan
9	Contact	Contact communication type
10	Day	Last contact day of the month
11	Month	Last contact month of the year
12	Duration	Last contact duration (in seconds)
13	Campaign	Number of contacts performed during this campaign
14	Pdays	Number of days that passed by after the client was last contacted
15	Previous	Number of contacts performed before this campaign
16	Poutcome	Outcome of the previous marketing campaign
17	Y	Whether the client subscribed a term deposit



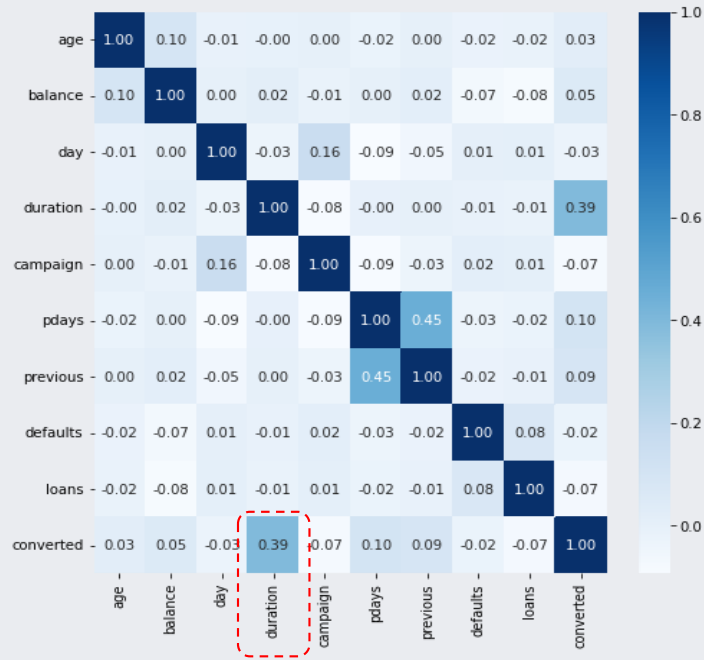
7 Numerical
9 Categorical
1 Target



Target Label
(converted)



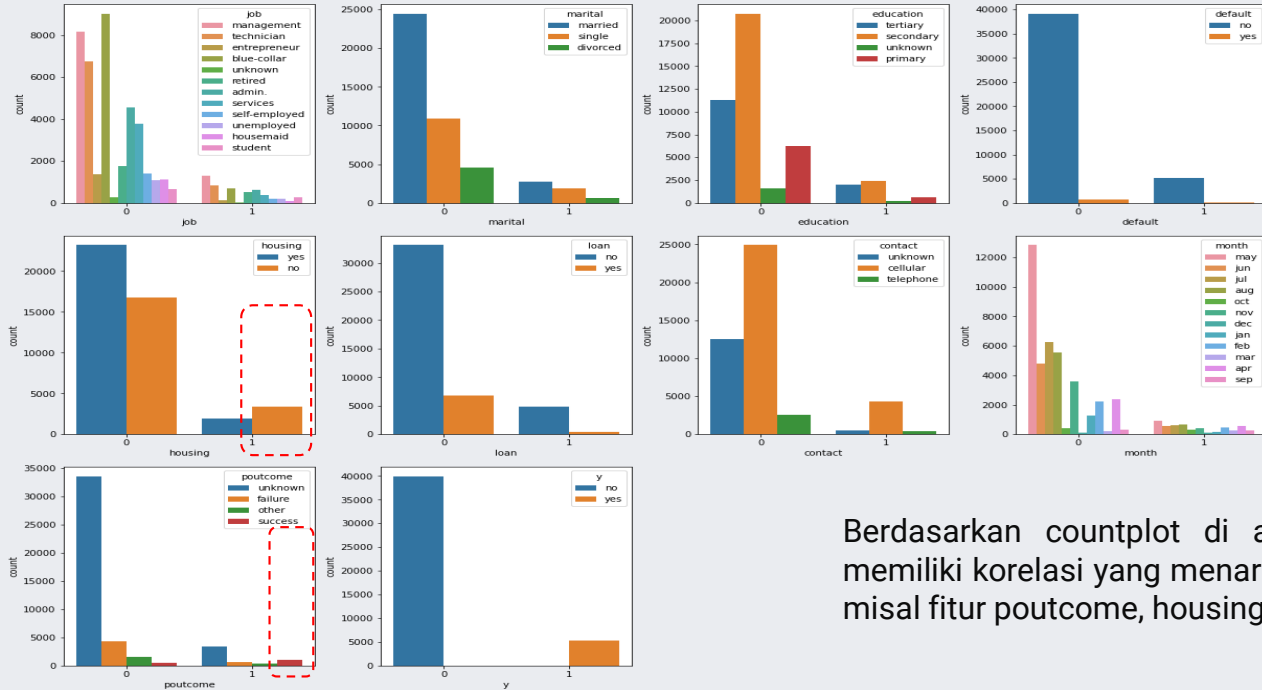
Feature Numerical



Berdasarkan heatmap correlation disamping, yang paling memiliki korelasi paling kuat dengan fitur target *converted* (*subscribed*) adalah fitur ***duration***.



Feature Categorical



0 = Not subscribed
1 = Subscribed

Berdasarkan countplot di atas, fitur yang memiliki korelasi yang menarik dengan target misal fitur poutcome, housing.

3

Data Cleaning & Preparation

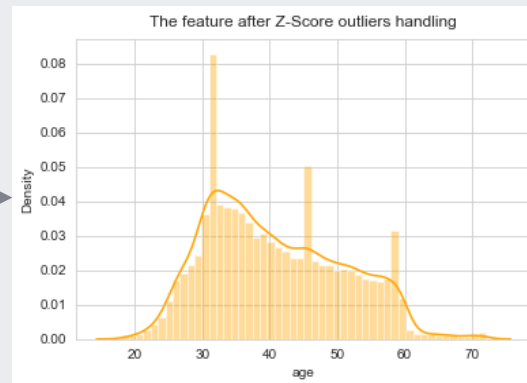
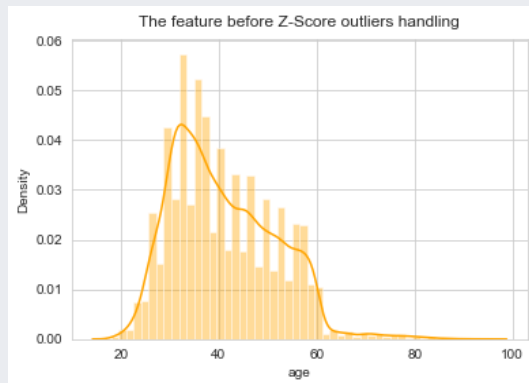


Data Cleaning

- Handling Null / Missing Values -> 0 Missing value
- Replace Issues Values with Modus
- Check & Remove Duplicates -> 4.163 Duplicates
- Handling Outliers (Z-Score Method) :

Why is it necessary?

Getting **clean data** to reveal insights is essential, as directly jumping into data analysis without proper data cleaning may lead to **incorrect results** [1]



*Data before ~ after cleaning:
45.211 → 41.048



Data Preparation

Encoding

- Label Encoding
- One Hot Encoding

Scaling & Transformation

- Log Transformation
- Standardization

Feature Selection

Analyze the feature that has possible to decrease ML Performance

Split Data Train & Test

Train : Test = 70 : 30 ratio

Imbalance Handling

Threshold-Moving
However, 0.5 is not ideal for our imbalanced datasets.

Threshold in our dataset ~ 0.2

Around **90% of the time spent** on data analytics, data visualization, and machine learning projects is dedicated to performing **data preprocessing** ^[1]

4

Model Evaluation

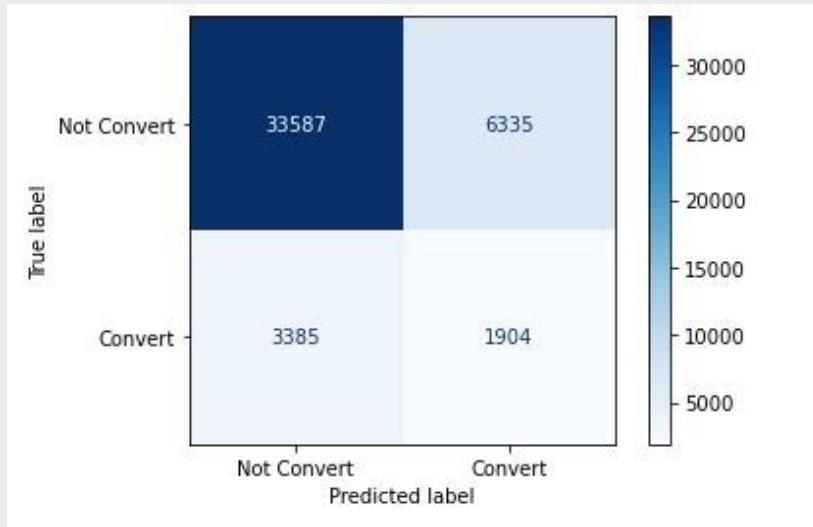


Model Evaluation After Hypertuning

Model	Precision		AUC		Accuracy		Recall		F1 Score	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	0.43	0.41	0.90	0.90	0.89	0.89	0.63	0.61	0.51	0.49
Decision Tree	0.38	0.37	0.89	0.87	0.87	0.89	0.61	0.58	0.47	0.45
KNN	0.41	0.37	0.90	0.88	0.88	0.88	0.63	0.62	0.50	0.47
Random Forest	0.56	0.44	0.94	0.91	0.92	0.90	0.67	0.66	0.61	0.53
XGBoost	0.53	0.48	0.94	0.92	0.92	0.91	0.70	0.62	0.60	0.55



Confusion Matrix (XGBoost)



$$\text{Conversion Rate} = \frac{1904}{1904 + 6335} = 23.1\%$$

1st Metrics : AUC

2nd Metrics : Precision

Karena fokus utama adalah **mengurangi cost** maka dipilih **Precision** karena tidak memperbolehkan **False Positive** yang besar.

*Nasabah yang **dihubungi** namun ternyata **tidak convert**

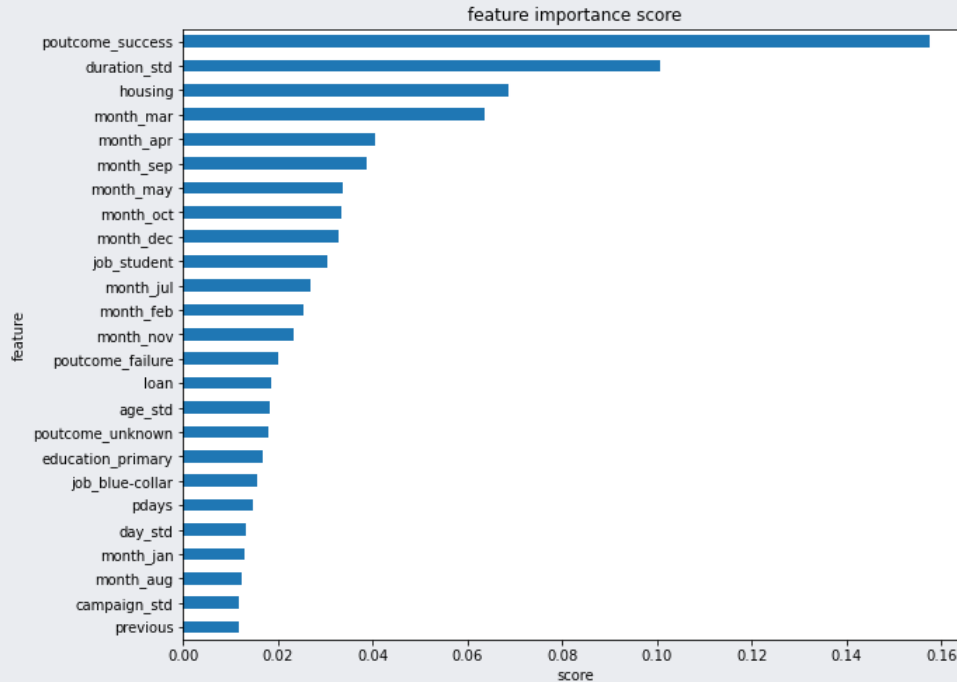
Skor **AUC** dipilih untuk mempertimbangkan agar model *machine learning* yang digunakan adalah yang ideal untuk *dataset* kami. Karena **Precision** tidak memiliki variable **True Negative** dalam kalkulasinya.

Precision = $TP / (TP + FP)$

AUC = $TPR = TP / (TP + FN)$, $FPR = FP / (FP + TN)$



Feature Importance (XGBoost)



Fitur **poutcome_success** (hasil OHE) dari data *categorical* dan fitur **duration** dari data *numerical* menjadi yang paling tinggi berdasarkan pemodelan XGBoost.



Business Metrics Evaluation

Parameter	Before	After
Total Customer	45,211	45,211
Subscribed customer	5,289	1,904
Total telephone	45,211	8,239
Conversion Rate (%)	11.7	23.1
Total Telephone Duration(sec)	29,208,626	4,143,950
Marketing Cost (Rp)	590,987,866	83,707,790

Increment of Conversion Rate

Save Cost

11.4%

Rp 506,306,455

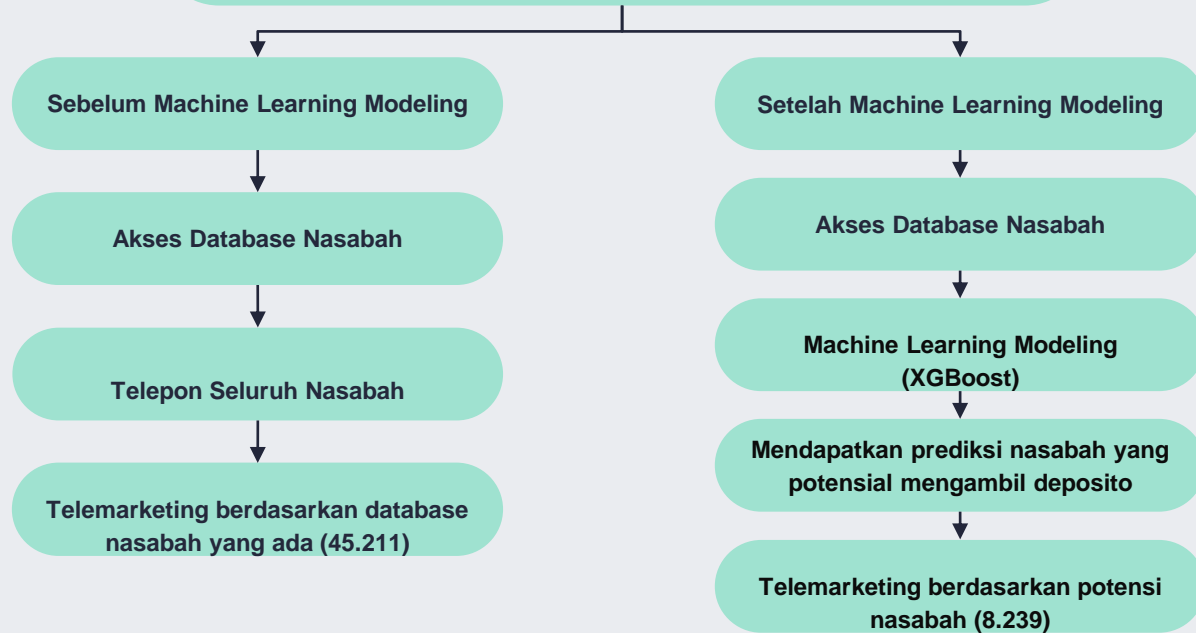
Asumsi:

- Total nasabah pada periode selanjutnya sama dengan total nasabah periode sebelumnya
- Tarif telepon yang digunakan^[1]
Rp 578/30s - Rp636/30s

Conversion Rate meningkat
~2x Lipat dari sebelum
adanya model.



PERBEDAAN SEBELUM DAN SETELAH MODELING

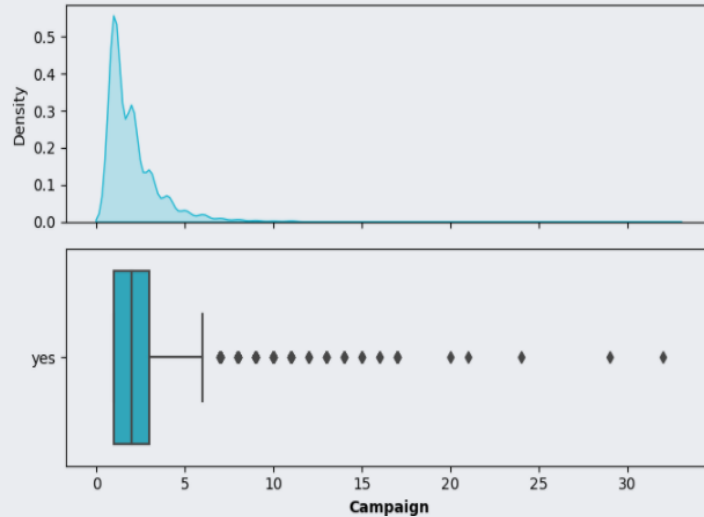


5

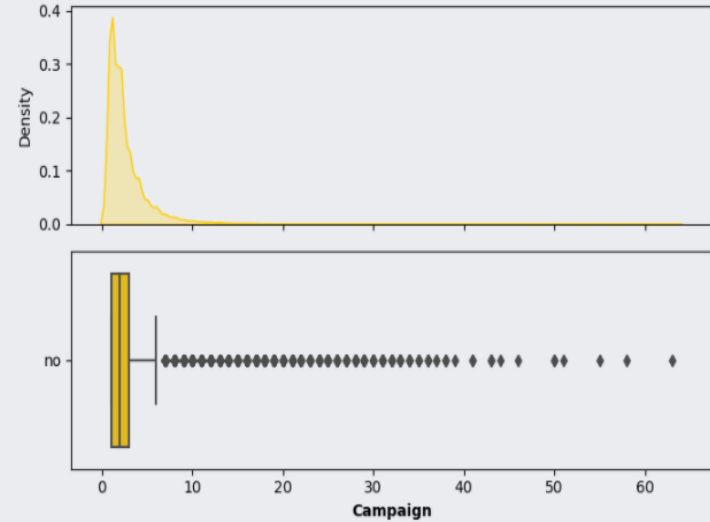
Business Insight & Recommendation



Distribution and Boxplot of Campaign for Subscribed Customer



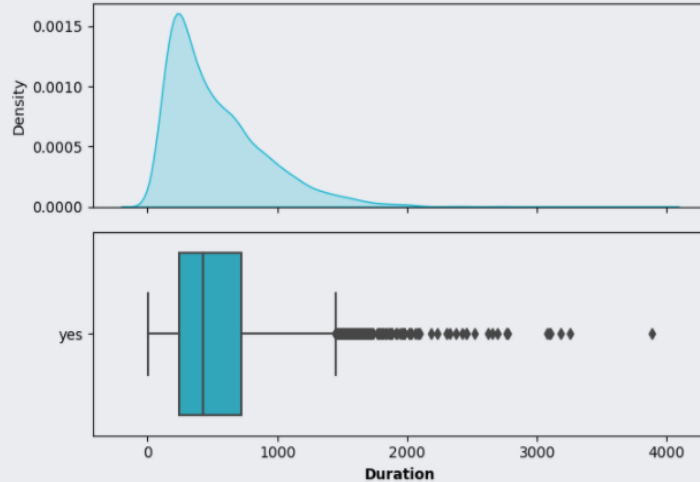
Distribution and Boxplot of Campaign for non-Subscribed Customer



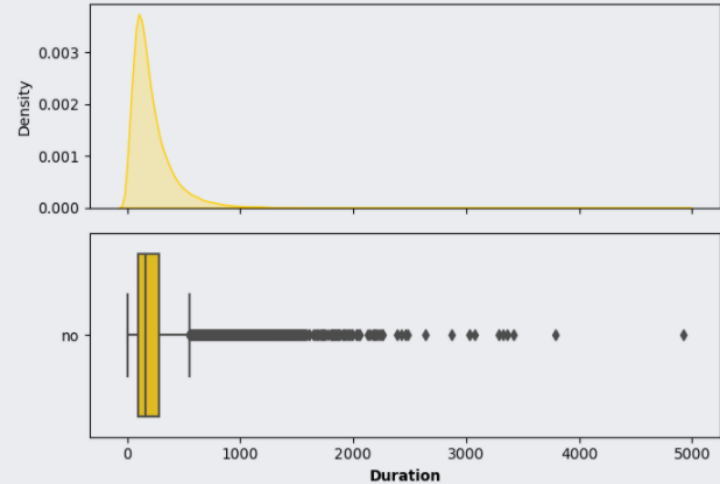
- Jika dilihat dari distribusinya, **Semakin tinggi jumlah campaign menurunkan jumlah nasabah untuk memutuskan mengambil deposito atau tidak mengambil deposit**
- Perlu menurunkan jumlah campaign untuk menurunkan biaya marketing, dimana **jumlah campaign 1–5 memiliki konsentrasi tertinggi nasabah yang subscribe deposit**
- **Rekomendasi**, tim sales **perlu membuat metode telemarketing yang menarik** dan yang cukup untuk dilakukan maksimal 5 kali sehingga dapat membuat nasabah tertarik untuk subscribe deposit.



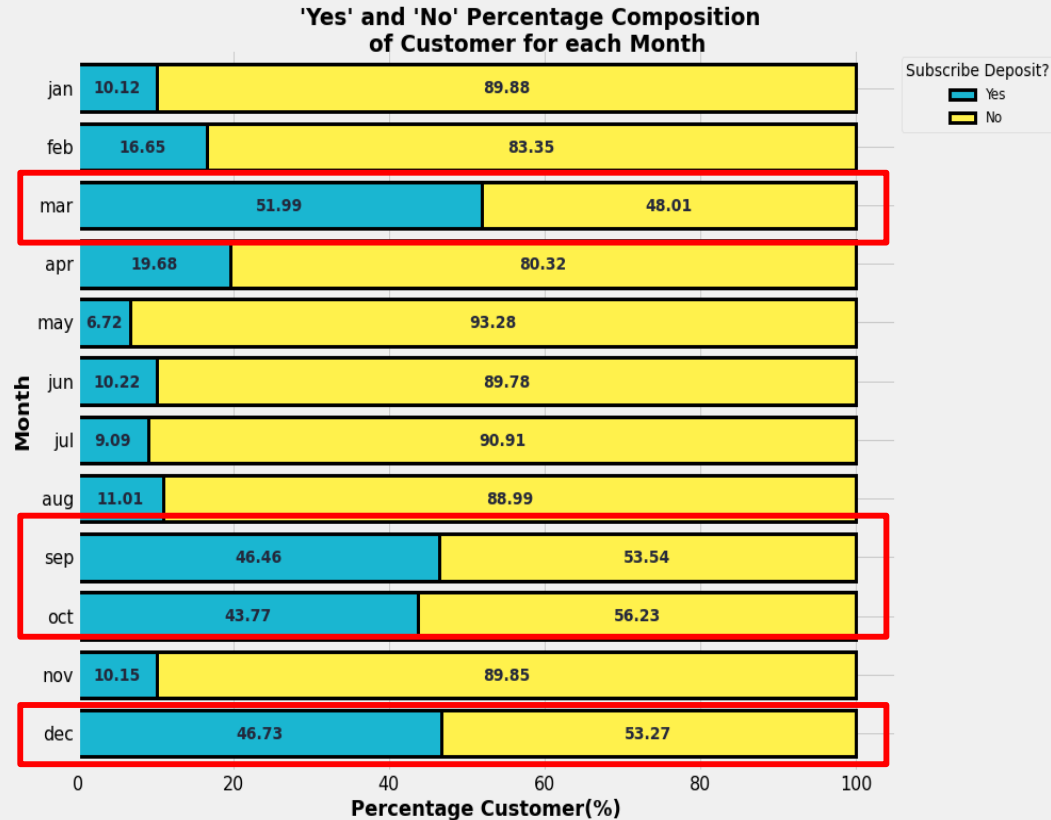
Distribution and Boxplot of Duration for Subscribed Customer



Distribution and Boxplot of Duration for non-Subscribed Customer



- Jika dilihat dari distribusinya, **Semakin lama durasi telepon akan menurunkan jumlah customer untuk memutuskan mengambil deposito atau tidak mengambil deposito**
- Perlu menurunkan durasi telepon untuk menurunkan biaya marketing, dimana durasi telepon **<500 detik memiliki konsentrasi tertinggi customer yang subscribe deposito**
- **Rekomendasi**, tim sales **perlu membuat metode telemarketing yang menarik** dan yang cukup untuk dilakukan dalam <500 detik durasi telepon sehingga dapat membuat nasabah tertarik untuk subscribe deposit.

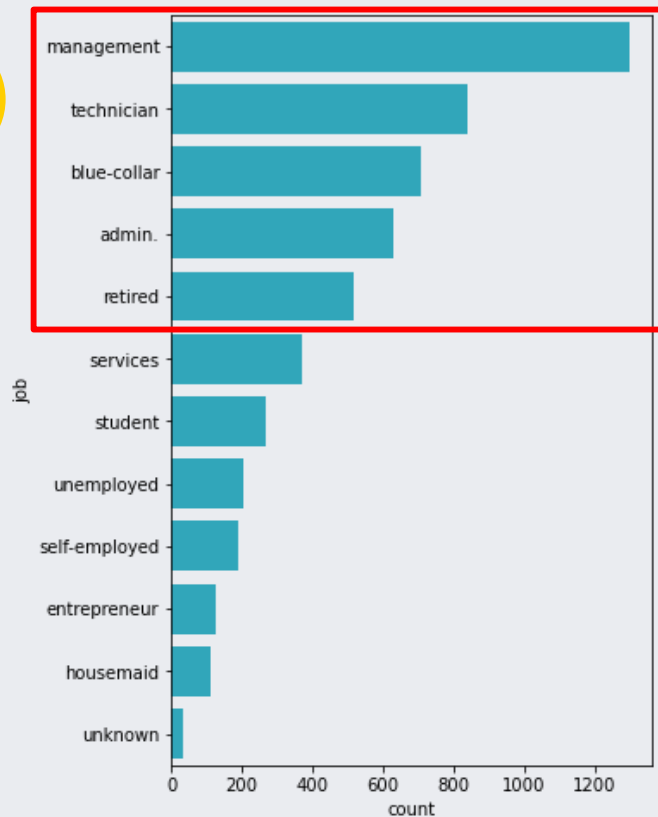


Terlihat bahwa persentase tertinggi nasabah yang *subscribe* deposito terjadi pada telemarketing yang dilakukan di bulan:

1. **Maret(51.99%)**
2. **Desember(46.73%)**
3. **September(46.46%)**
4. **Oktober(43.77%)**

Rekomendasi

Tim sales dapat mengencarkan telemarketing pada bulan-bulan tersebut

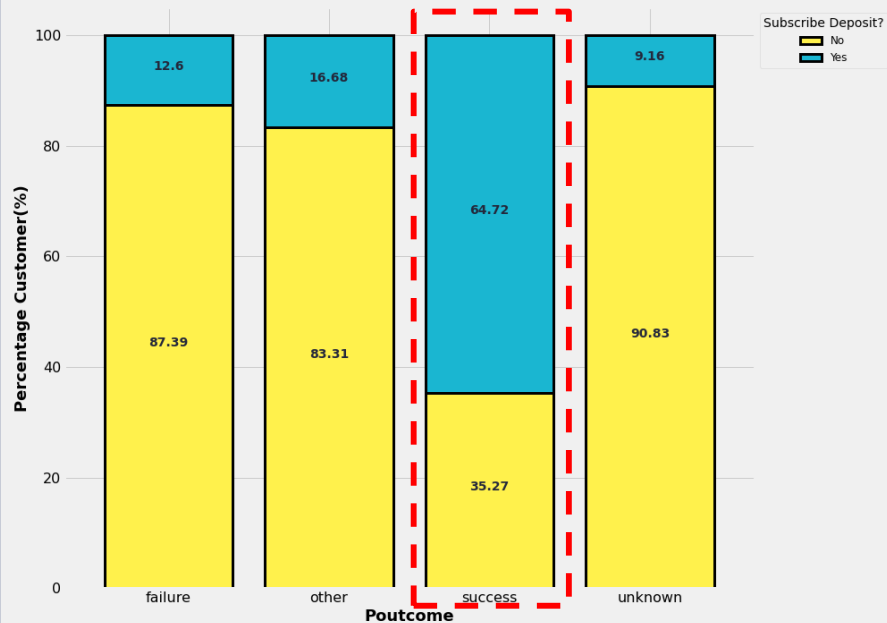


Insight:

Pada fitur **Job** diketahui jumlah nasabah yang *subscribed* deposito berdasarkan tipe pekerjaan.

Nasabah yang memiliki pekerjaan management, technician, blue-collar, admin dan retired lebih banyak yang *subscribed* deposito berjangka dibanding pekerjaan lain

Rekomendasi: Tim Sales dapat menargetkan nasabah yang memiliki pekerjaan-pekerjaan pada peringkat **Top 5** di samping.



Insight:

"Poutcome" : diketahui sebagai status campaign sebelumnya pada nasabah tersebut.

Seperti hasil importance feature dari XGBoost adalah fitur dengan pengaruh tertinggi, lalu pada EDA di samping nasabah yang *subscribed* lebih banyak berasal dari nilai poutcome-nya *success* daripada yang tidak.

Rekomendasi:

So.. Tim marketing sebaiknya mempertimbangkan hal ini, yang *success* pada campaign sebelumnya untuk diprioritaskan terlebih dahulu untuk ditelpon karena mempunyai potensi *subscribed*.

Thank you 😊