

# FINAL PROJECT

Marketing Campaign



Data Science Batch 38



**MARKET  
INSIDER**



# MEET OUR TEAM



ACHMAD HILMAN SHADIQIN  
CEO OF MARKET INSIDER



ANDREAWAN SOFIAN  
DATA SCIENTIST



FIGO AKMAL  
DATA SCIENTIST



NABILAH ASTIARINI  
DATA SCIENTIST

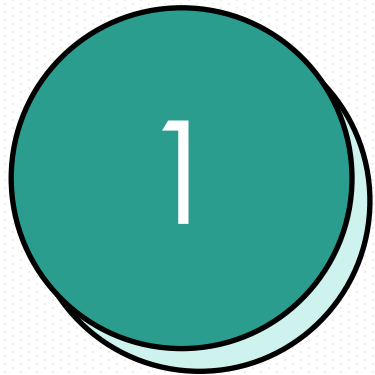


RIYAN MAULA  
DATA SCIENTIST



DZAKWAN D.  
DATA SCIENTIST

# Agenda



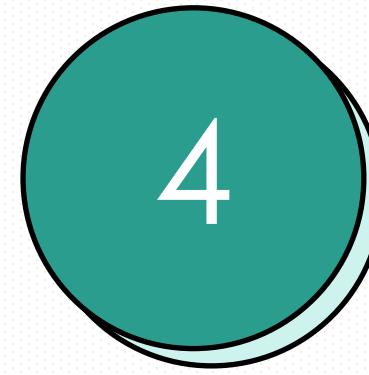
Introduction



EDA & BI



Data Preprocessing



Modelling &  
Evaluation



Business  
Recommendation

# INTRODUCTION

## Business Understanding

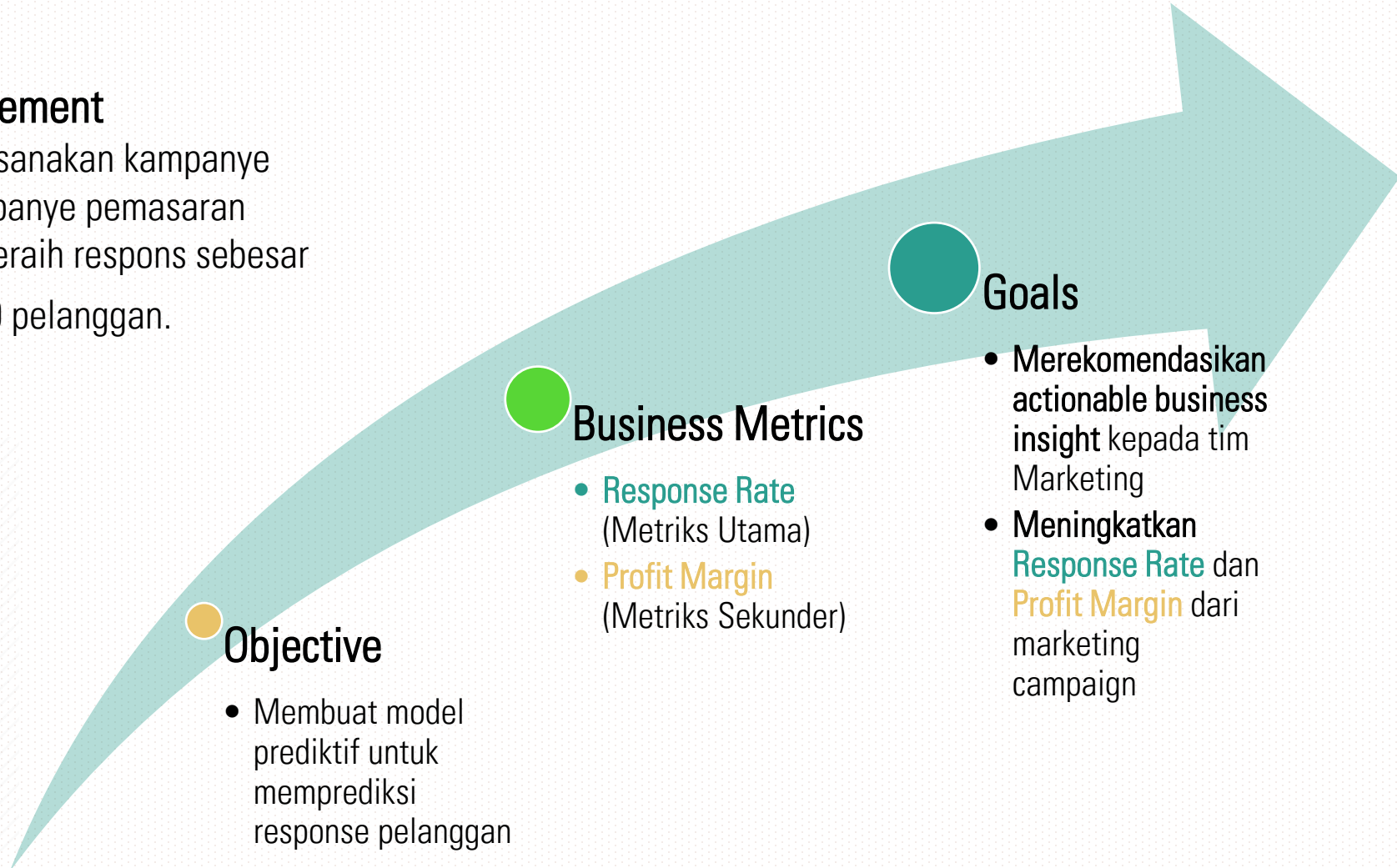
- ☐ Problem Statement
- ☐ Objective
- ☐ Business Metrics
- ☐ Goals



# Business Understanding

## Problem Statement

Sebuah perusahaan melaksanakan kampanye pemasaran, dimana kampanye pemasaran terakhir hanya berhasil meraih respons sebesar **14.91%** dari 2240 pelanggan.



# EDA & BI

- ☐ Data Overview
- ☐ Statistical Analysis
- ☐ Univariate Analysis
- ☐ Multivariate Analysis
- ☐ Business Insight



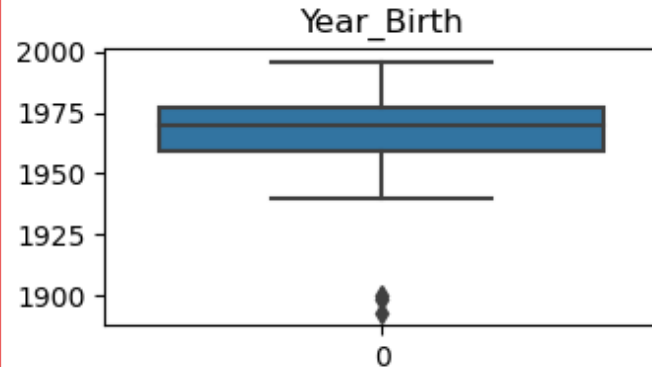


# Data Overview

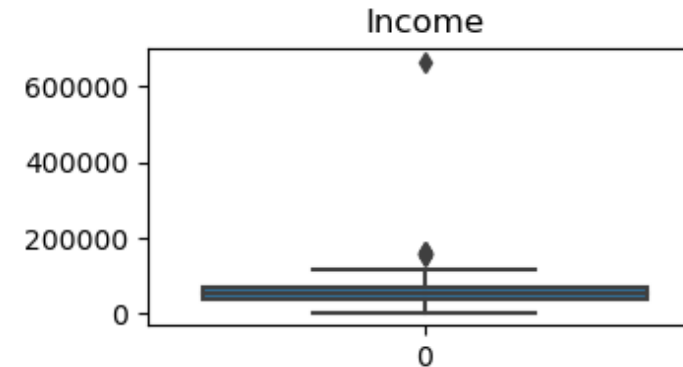
- 2240 baris, 29 kolom
- 26 feature bertipe numerik, 3 feature bertipe kategori
- 1.07% missing values dan 8.12% duplicates
- Outliers pada 'Year\_Birth' dan 'Income'
- Invalid values pada 'Marital\_Status' dan 'Education'
- 'Dt\_Customer' diubah ke tipe data datetime
- 'ID', 'Z\_CostContact', 'Z\_Revenue' bersifat terlalu unique (drop)

# Statistical Analysis

	count	mean	std	min	10%	20%	30%	40%	50%	60%	70%	80%	90%	max
Year_Birth	2240.0	1968.805804	11.984069	1893.0	1952.0	1957.0	1962.0	1966.0	1970.0	1973.0	1976.0	1979.0	1984.0	1996.0
Income	2216.0	52247.251354	25173.076661	1730.0	24117.5	32011.0	38198.5	44529.0	51381.5	58482.0	65247.5	71819.0	79844.0	666666.0



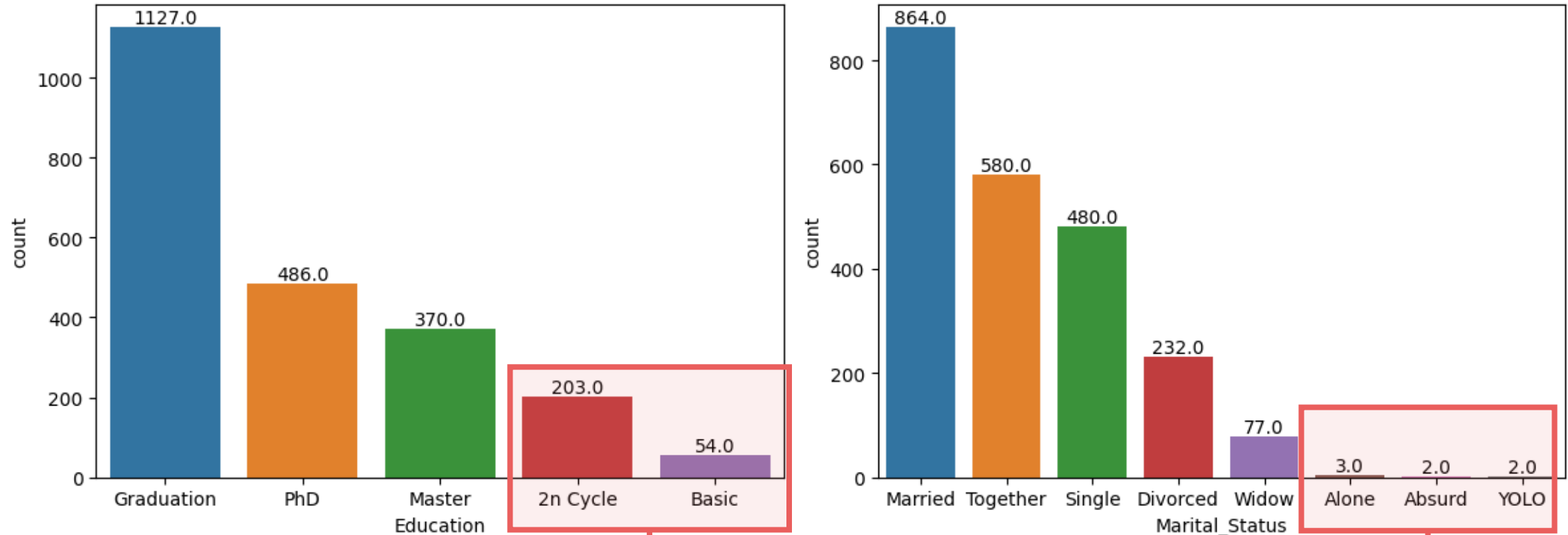
Outliers pada Year\_Birth terletak **di bawah percentile 10%** atau tepatnya di sekitar angka **1893**



Outliers pada Income terdapat **di atas percentile 90%** atau tepatnya di sekitar angka **666.666**



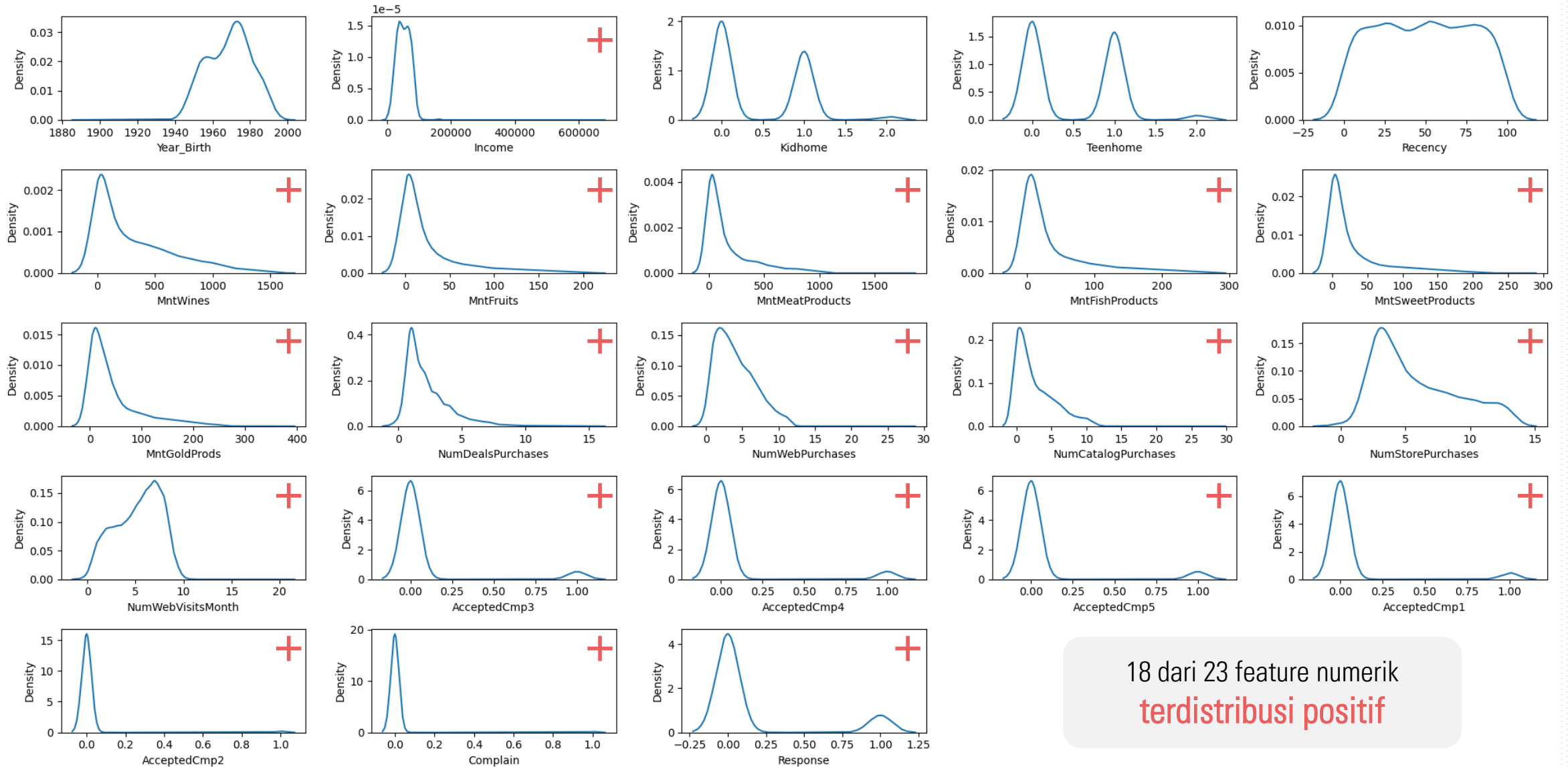
# Identifikasi Invalid Value



Basic dan 2n Cycle belum jelas tingkatan pendidikan formalnya

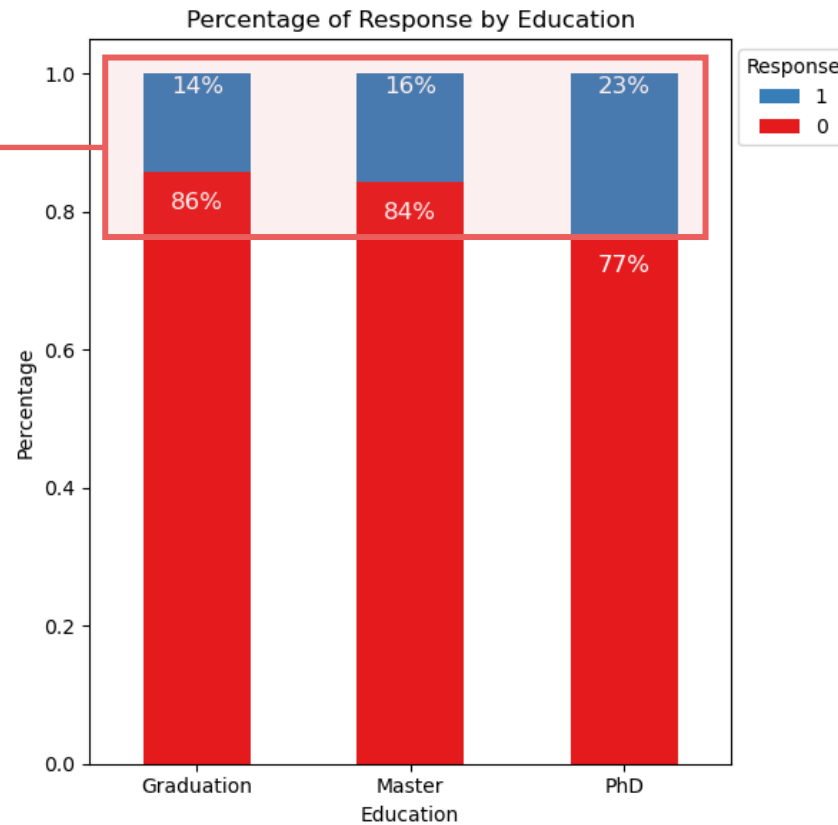
Jumlah Alone, Absurd dan YOLO dapat diabaikan.

# Distribusi Data

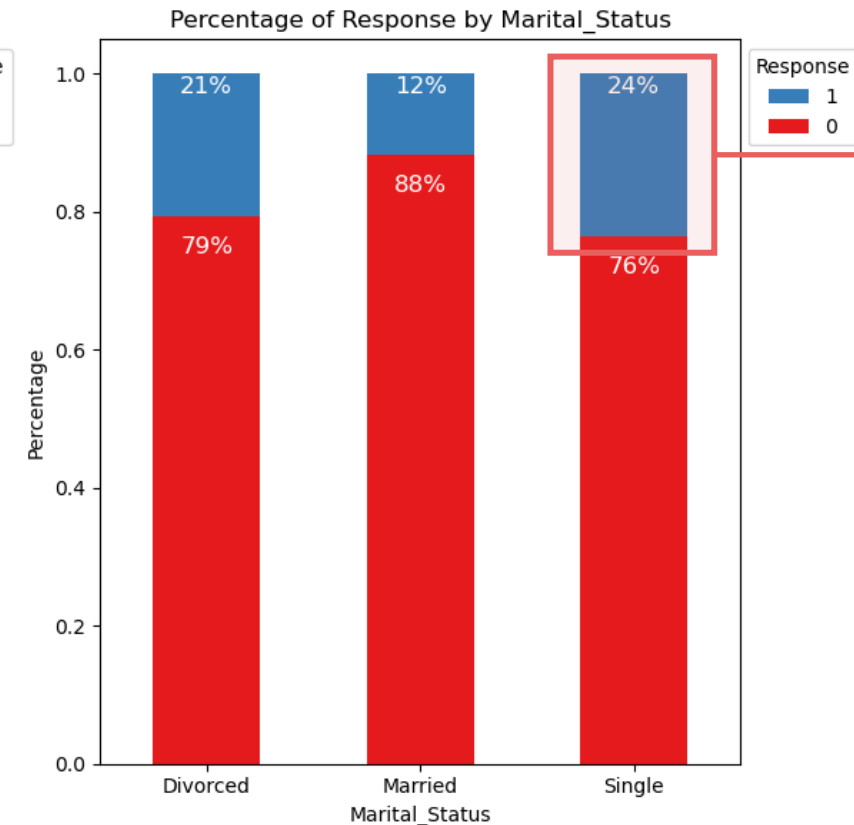


18 dari 23 feature numerik  
terdistribusi positif

# Univariate Analysis

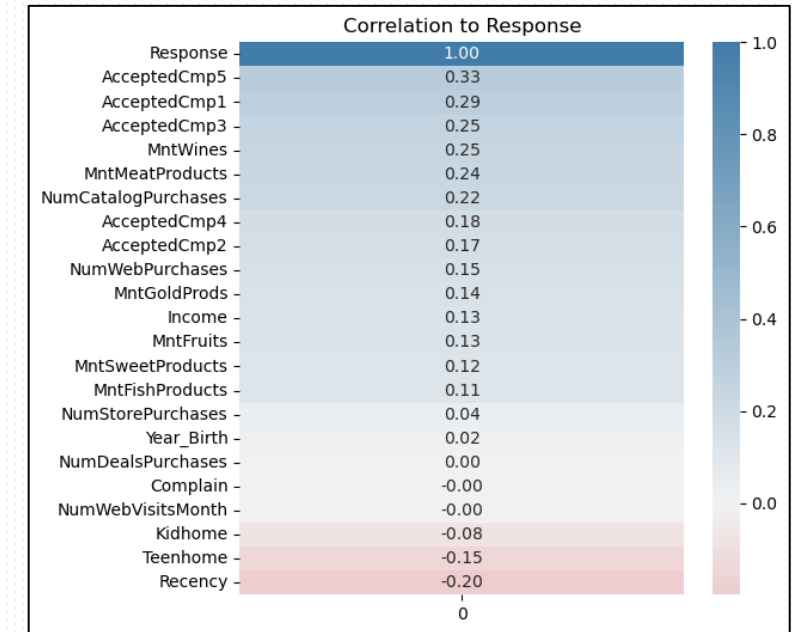
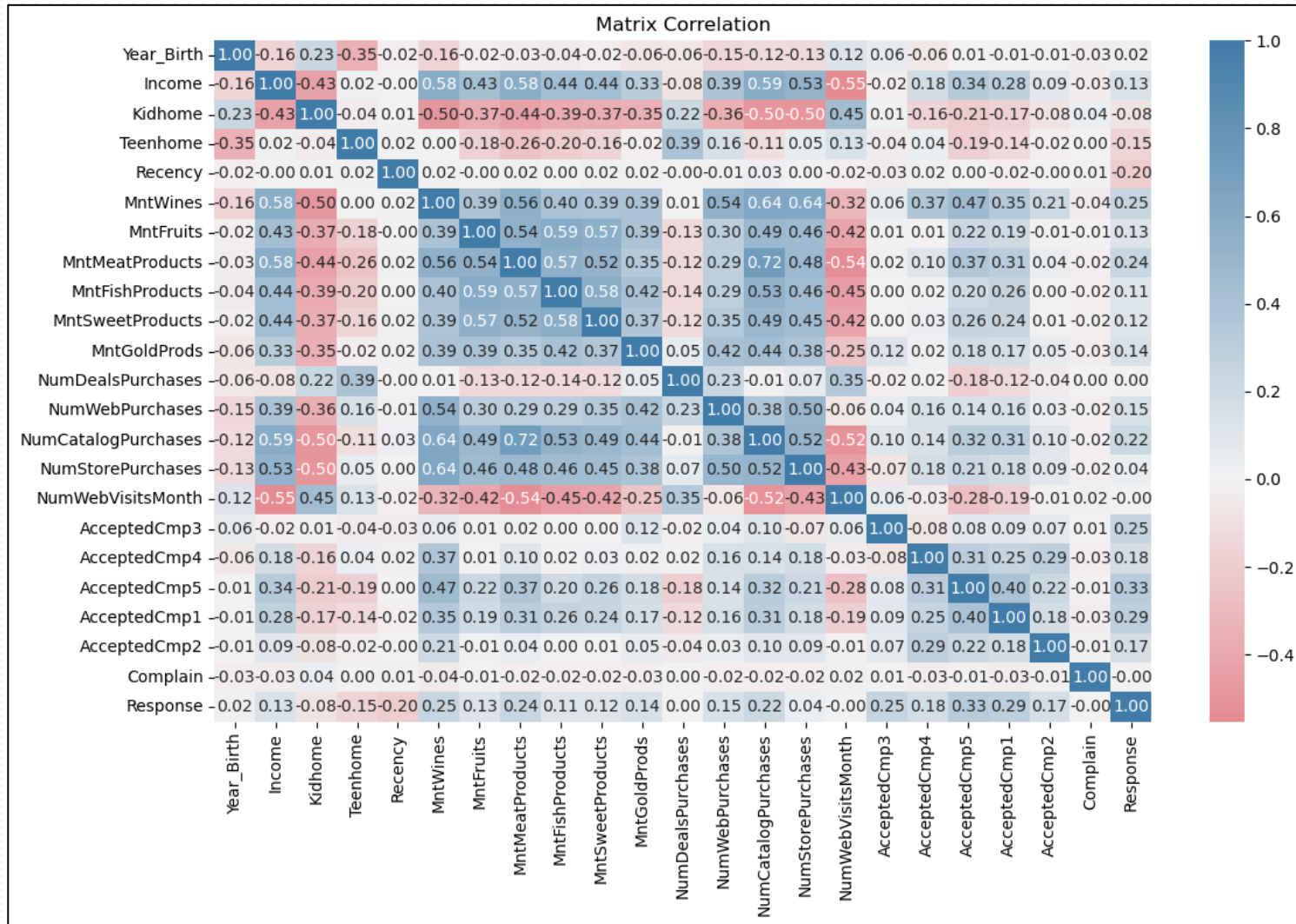


Semakin Tinggi Tingkat Pendidikan meningkatkan % response



Single mempunyai % response tertinggi

# Multivariate Analysis

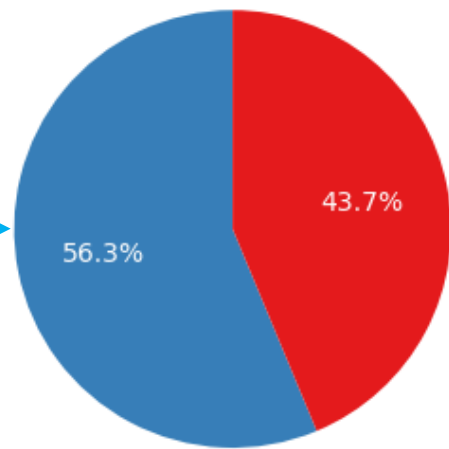


## Skor Korelasi Feature

Chanel pembelian, produk, dan campaign mempunyai **korelasi positif**, sedangkan **recency** dan **kepemilikan anak** mempunyai **korelasi negatif**

# Business Insight

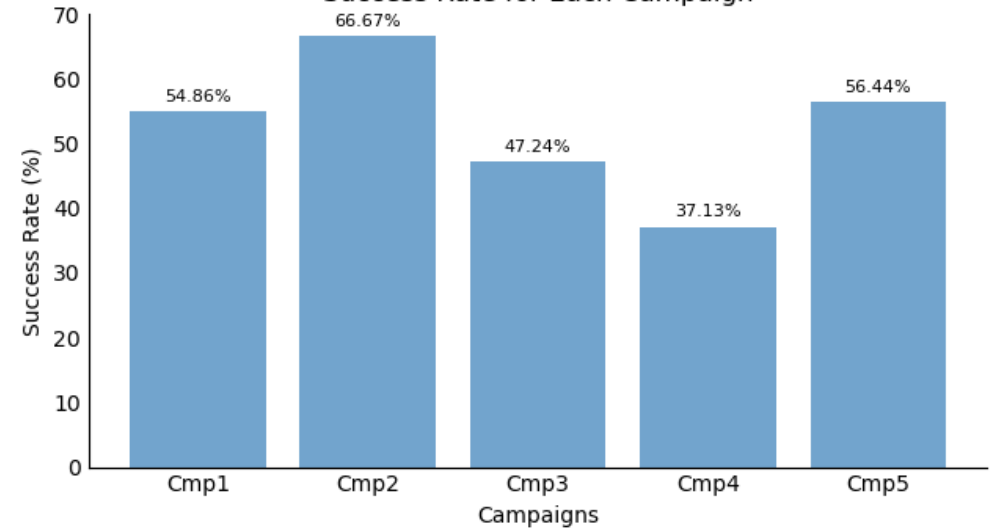
Percentage of Yes Response with and without Accepting Campaigns



Dari Campaign (Cmp1-5)

Tanpa Campaign (Cmp1-5)

Success Rate for Each Campaign



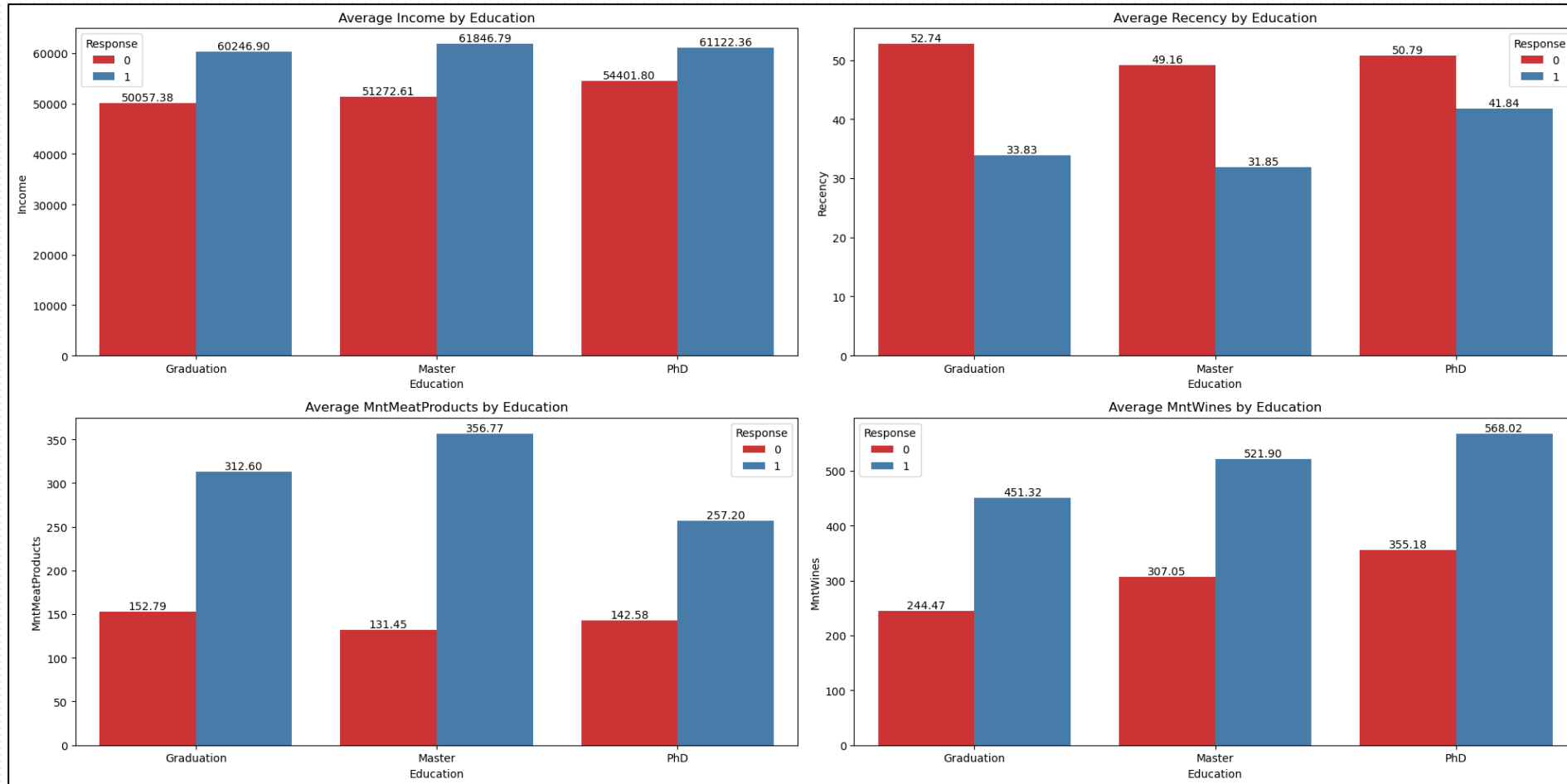
## Insight

- Semakin banyak campaign, semakin tinggi response
- Campaign ke-2 adalah yang terbaik

## Rekomendasi

- Perbanyak campaign
- Gunakan strategi pada Campaign ke-2
- Targetkan Customer yang belum menerima campaign sama sekali

# Business Insight



## Kelompok Customer Response

- Income tinggi (±60000)
- Recency rendah (±30 hari)
- Meat tinggi (250-350)
- Wine tinggi (450-650)

## Kelompok Customer Tidak Response

- Income rendah (55000)
- Recency tinggi (±50 hari)
- Meat rendah (130-150)
- Wine rendah (245-355)

## Rekomendasi Target Customer

- Income minimal 60000
- Recency maksimal 40 hari
- Meat minimal 250
- Wine minimal 450

# Data Preprocessing

## Data Cleansing

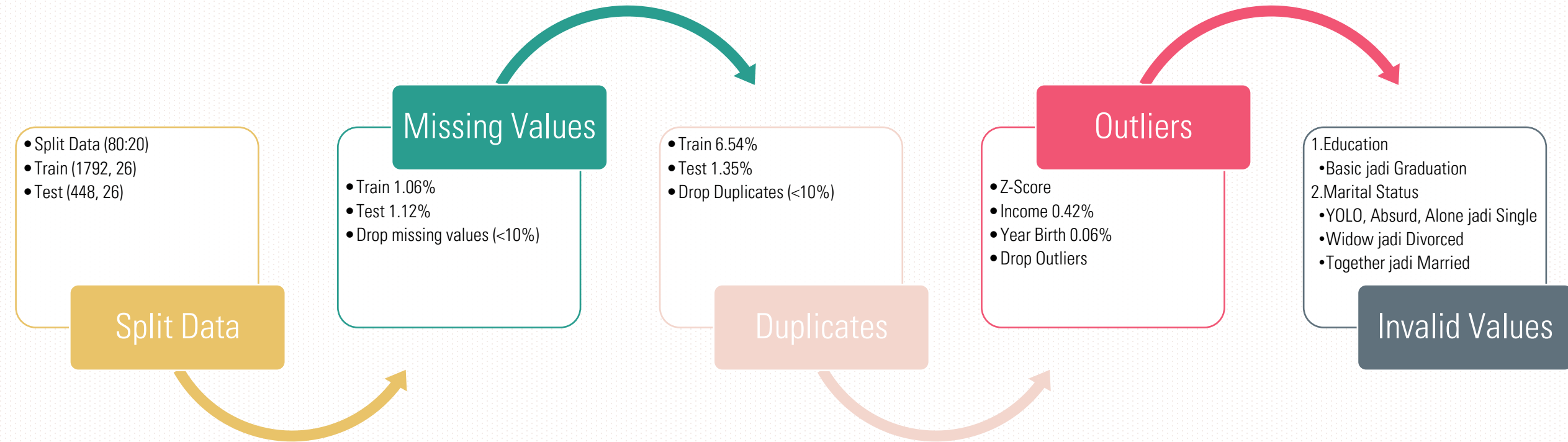
- ☐ Handling Missing Values
- ☐ Handling Duplicates
- ☐ Handling Outliers
- ☐ Handling Invalid Values

## Feature Engineering

- ☐ Feature Extraction
- ☐ Feature Encoding
- ☐ Feature Transformation
- ☐ Feature Selection
- ☐ Imbalance Handling



# Data Cleansing





# Feature Extraction & Encoding

## Feature Extraction

1. RFM Cat
2. Customer Lifespan
3. Total Purchase
4. Total Spending
5. Total Offers, dan 10 lainnya.

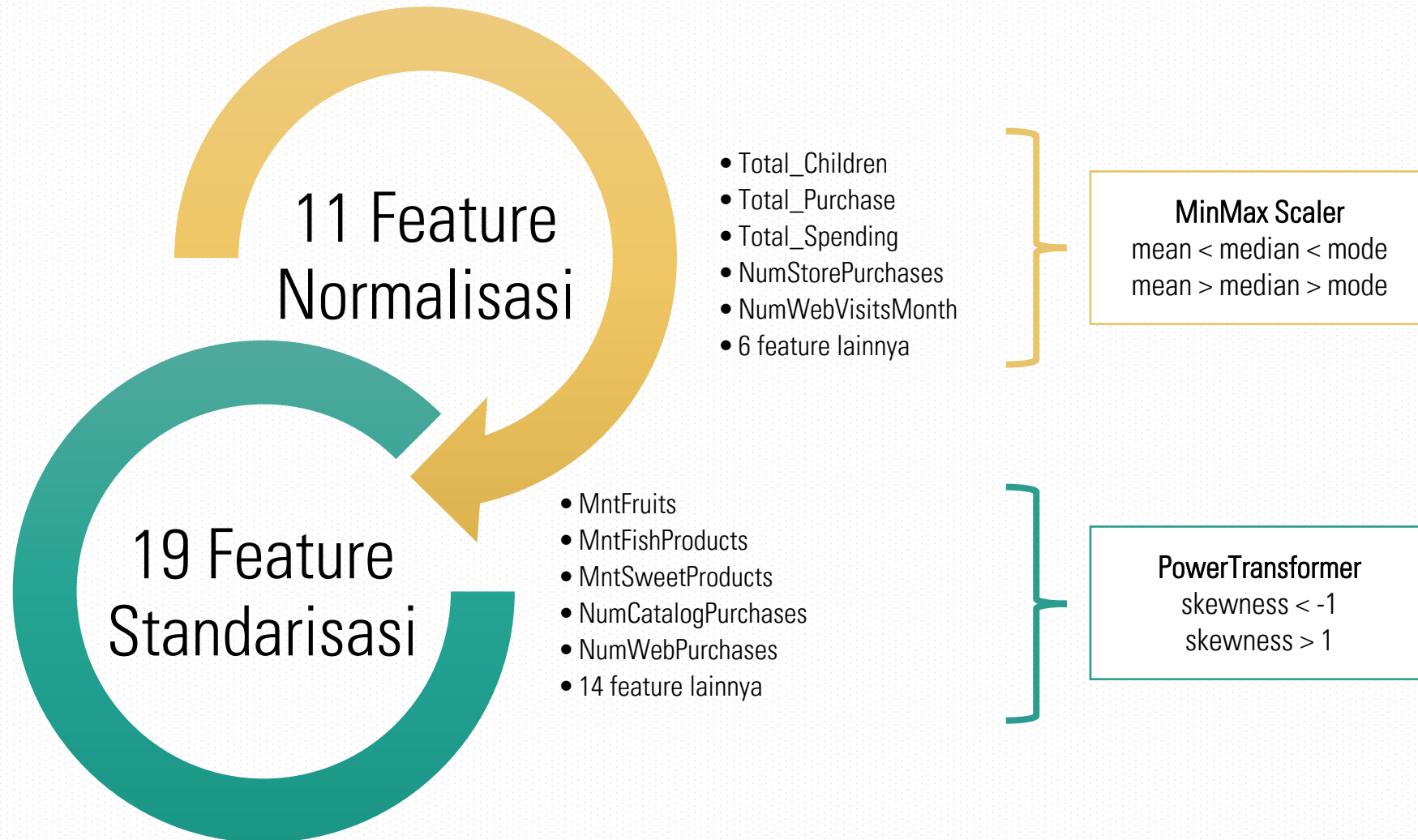
## Encoding Categorical

1. Ordinal Encoding
  - Education
  - Marital Status
2. Manual Encoding
  - Relationship Status
  - Primary Needs
  - RFM Cat

**RFM Cat**, segmentasi customer (champions, loyal customer, at risk customer, dan new customer).  
– Recency, Total Purchase, Total Spending

**Customer Lifespan**, jumlah hari saat menjadi customer sampai data dikumpulkan (Juli, 2014).  
– Dt Customer

# Feature Transformation



# Feature Selection

## Uji Anova

### Numerical - Categorical

Before: 33 Feature  
After: 28 Feature

## Uji Chi-Squared

### Categorical - Categorical

Before: 5 Feature  
After: 3 Feature

## Redundancy Analysis

### Redundancy Feature

Before: 31 Feature  
After: 15 Feature

\*Confidence Level 95% (Anova, Chi-Squared)

## Feature Selected

1. Recency
2. Customer\_Lifespan
3. RFM\_Cat
4. Education
5. MntFruits
6. MntFishProducts
7. MntSweetProducts
8. MntGoldProds
9. NumWebPurchases
10. NumCatalogPurchases
11. AcceptedCmp1
12. AcceptedCmp2
13. AcceptedCmp3
14. AcceptedCmp4
15. AcceptedCmp5



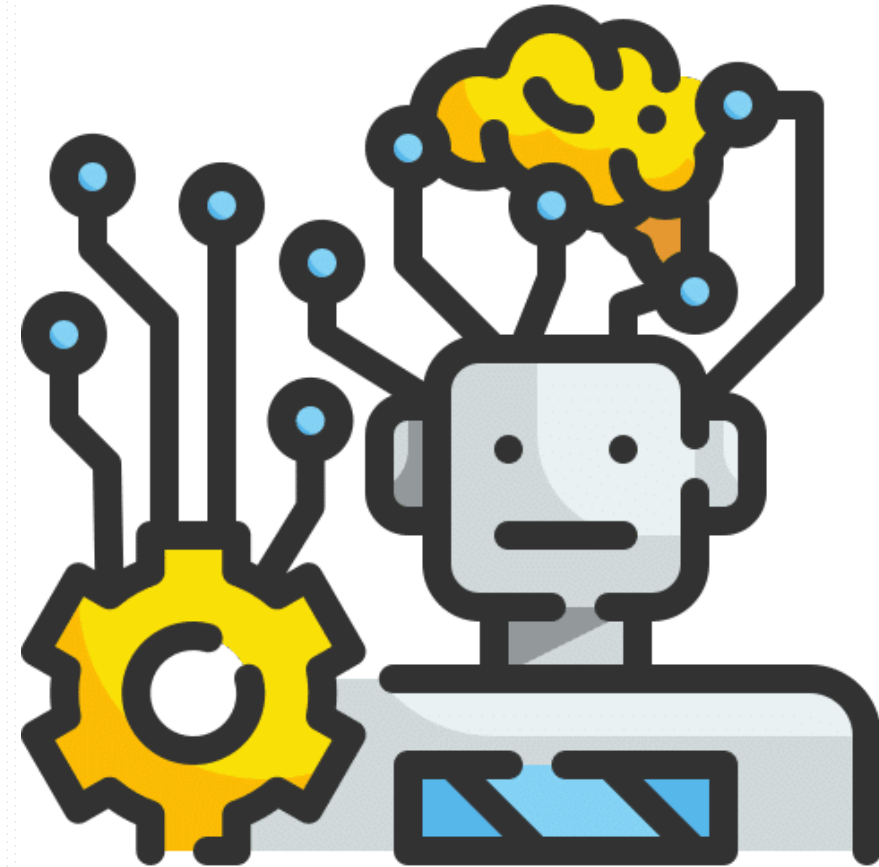
# Modeling & Evaluation

## Modelling & Evaluation

- ☐ Modelling
- ☐ Evaluation
- ☐ Feature Importance

## Business Insight & Recommendation

- ☐ Business Simulation
- ☐ Business Insight & Recommendation



# Modelling

1

## Precision

Metrik Primer

**Mengurangi False Positive** yaitu situasi ketika model memprediksi customer untuk merespon kampanye padahal realitanya tidak.

**Meningkatkan Tingkat Respon Customer**

2

## Accuracy

Metrik Sekunder

**Meningkatkan tingkat kebenaran (True Positive + True Negative)** prediksi dari suatu model.

**Meningkatkan revenue stream** dari setiap kampanye.

# Evaluation

## Metrics Precision

Model	CV Precision	CV Precision HT	Fit Rate (%)	Fit Rate (%) HT
SVC	0.799378	0.873749	2.25	2.79
XGBClassifier	0.847233	0.912467	4.29	11.30
<b>AdaBoostClassifier</b>	<b>0.872463</b>	<b>0.908271</b>	<b>6.47</b>	<b>8.20</b>

## Metrics Accuracy

Model	CV Accuracy	CV Accuracy HT	Fit Rate (%)	Fit Rate (%) HT
SVC	0.832502	0.834291	12.87	13.28
XGBClassifier	0.840023	0.879048	6.63	7.92
<b>AdaBoostClassifier</b>	<b>0.861142</b>	<b>0.871889</b>	<b>7.18</b>	<b>7.34</b>

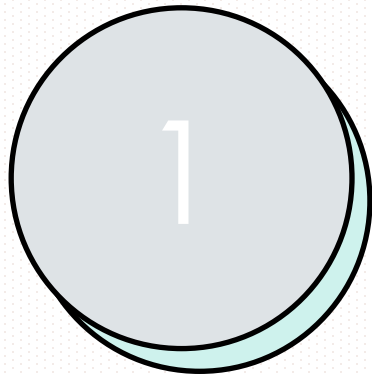
\*Fit Rate = Selisih nilai presisi pada data train-data test

### Efek Tuning

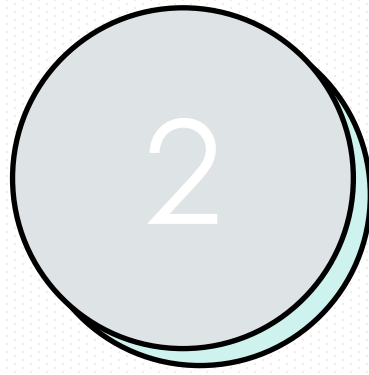
Hyper Parameter tuning  
meningkatkan performa model  
menuju best-fit

Nilai presisi juga terlihat meningkat  
setelah di-tuning

# Pemilihan Model Machine Learning



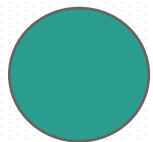
SVC



XGB Classifier



AdaBoost Classifier



**Model Rekomendasi  
Market Insider**

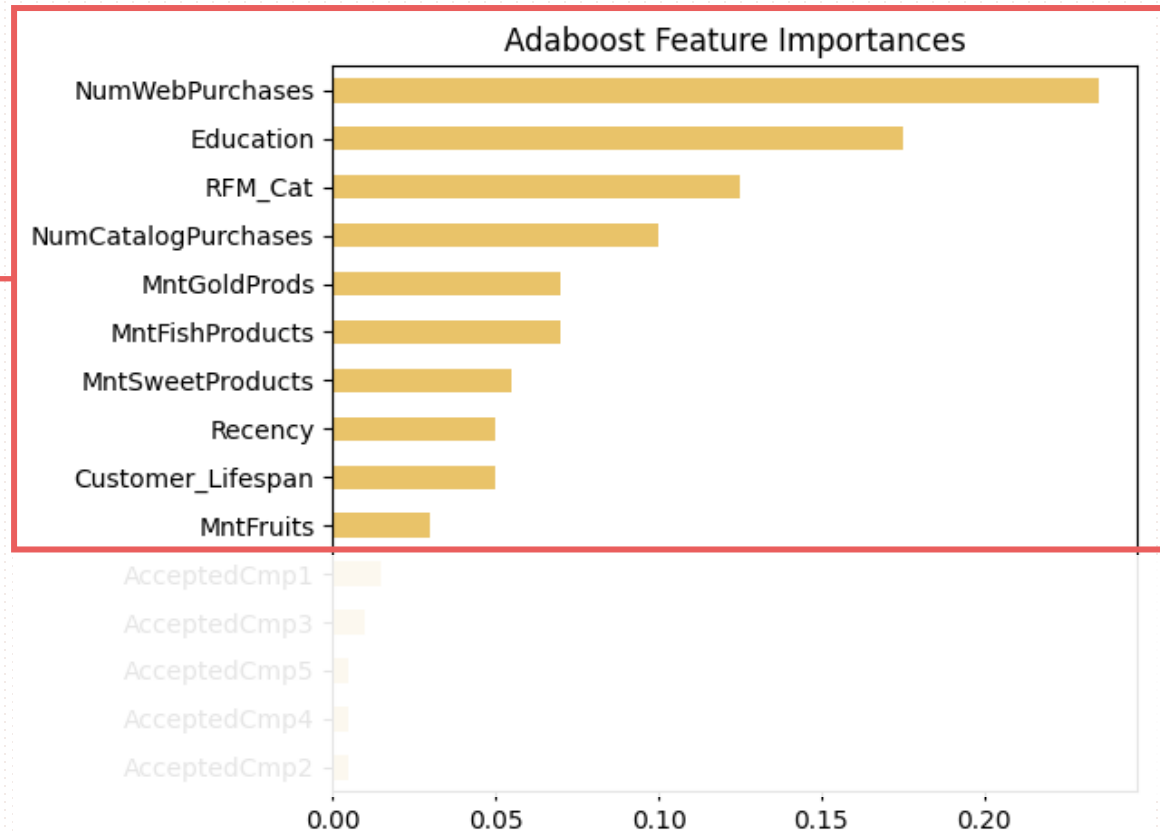
Model AdaBoost memiliki nilai precision dan Fit Rate yang optimal diantara lainnya

Model	Precision (%)	Accuracy (%)	Fit Rate HT (%)
SVC	87.49	83.42	2.7
XGB	91.24	87.90	11.3
AdaBoost	90.82	87.18	8.2

\*Fit Rate = Selisih nilai presisi pada data train-data test



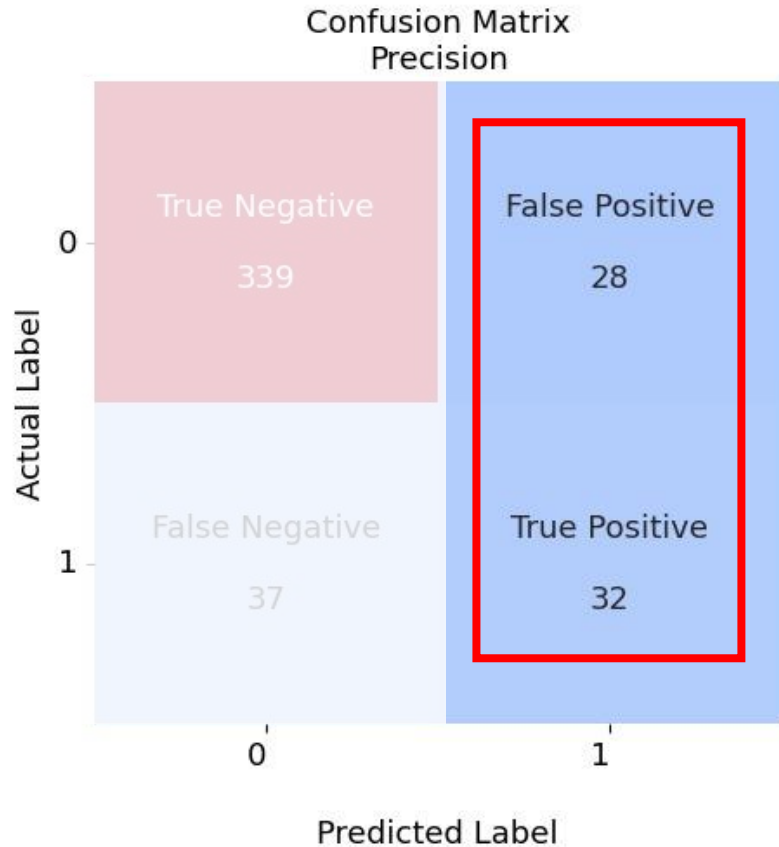
# Feature Importance



Feature	Importance
NumWebPurchases	0.235
Education	0.175
RFM_Cat	0.125
NumCatalogPurchases	0.100
MntFishProducts	0.070
MntGoldProds	0.070
MntSweetProducts	0.055
Customer_Lifespan	0.050
Recency	0.050
MntFruits	0.030

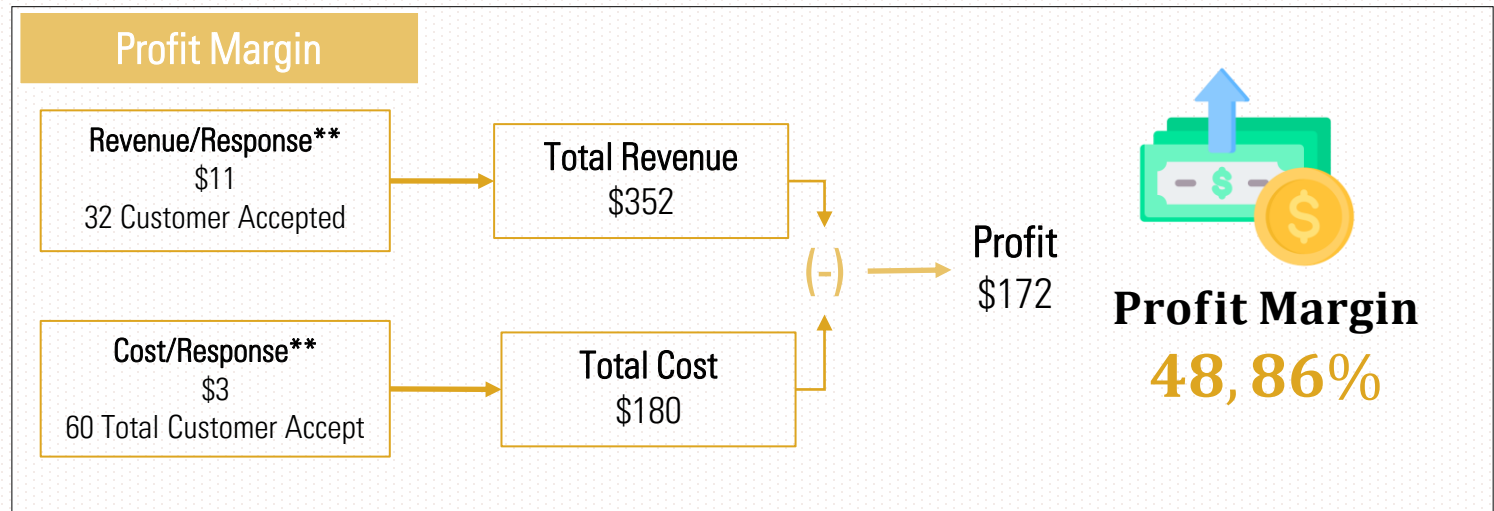
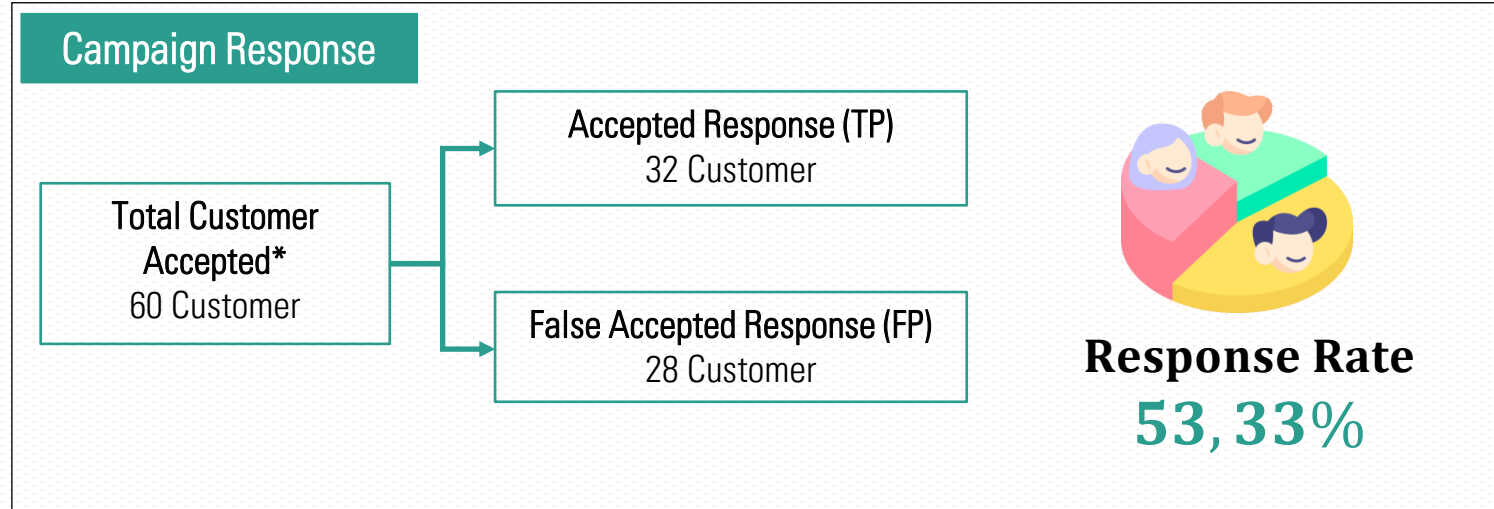
Top 10 feature Importance menjadi acuan strategi pada  
**Business Recommendation!**

# Business Simulation



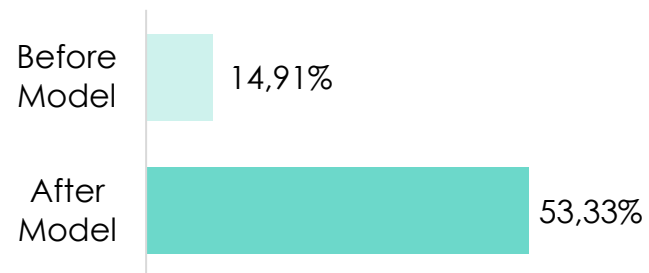
\*Hitungan bersifat simulasi karena menggunakan data testing

\*\*Nilai Awal Revenue dan Cost merupakan asumsi tim DS



# Business Simulation

## Response Rate



Response Rate mengalami peningkatan sebesar  
**38,42%**

## Profit Margin



Profit Margin mengalami peningkatan sebesar  
**131,77%**

# Business Recommendation

Sebagai acuan rekomendasi bisnis, kami akan fokus pada profil pelanggan berikut.



## Education

Level Graduation



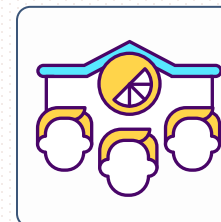
## RFM Category

Loyal Customer  
(frekuensi belanja & monetary tinggi)



## Recency

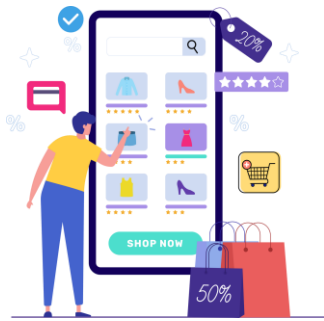
32 - 46 hari



## Lifespan

393 - 495 hari

# Business Recommendation



Memberikan **diskon** harga pada produk terpilih (**produk emas, ikan, makanan manis atau buah**) untuk pembelian melalui **website & katalog** dengan tujuan meningkatkan ketertarikan pelanggan dan peningkatan penjualan.



Memberikan **voucher** dan mengadakan program **cashback** yang dapat diredeem oleh pelanggan terpilih saat bertransaksi dengan tujuan mempertahankan pelanggan.



Mengadakan **promo bundling** untuk pembelian produk utama (emas atau ikan) dan produk pendukung lainnya (makanan manis atau buah) untuk meningkatkan ketertarikan pelanggan membeli lebih banyak produk.

# Thank you

Market Insider

Data Science Batch 38



**Rakamin Academy**  
School for Career Acceleration



**MARKET  
INSIDER**