



DATALICIOUS

Marketing Campaign

LOTTO MART

Rakamin Academy
Data Science Batch 32

READ MORE



COMPANY BACKGROUND

Lotto Mart is a supermarket retail that sells various types of products such as Fish, Meat, Fruits, Sweet Products, Wines, and Gold Products. For the last 6 months, Marketing Team has conducted a campaign in the form of giving discount vouchers to all customers via Broadcast Messages.



MEET THE TEAM



NUR IMAM
MASRI



SITI HAJJAH
MARDIAH



PRASIDYA
BAGASKARA



RISKIYATUL
HASANAH



M. RAYHAN
AZZINDANI



ASTUTI
RAHMAWATI



M. HARWIN
PRAYOGA



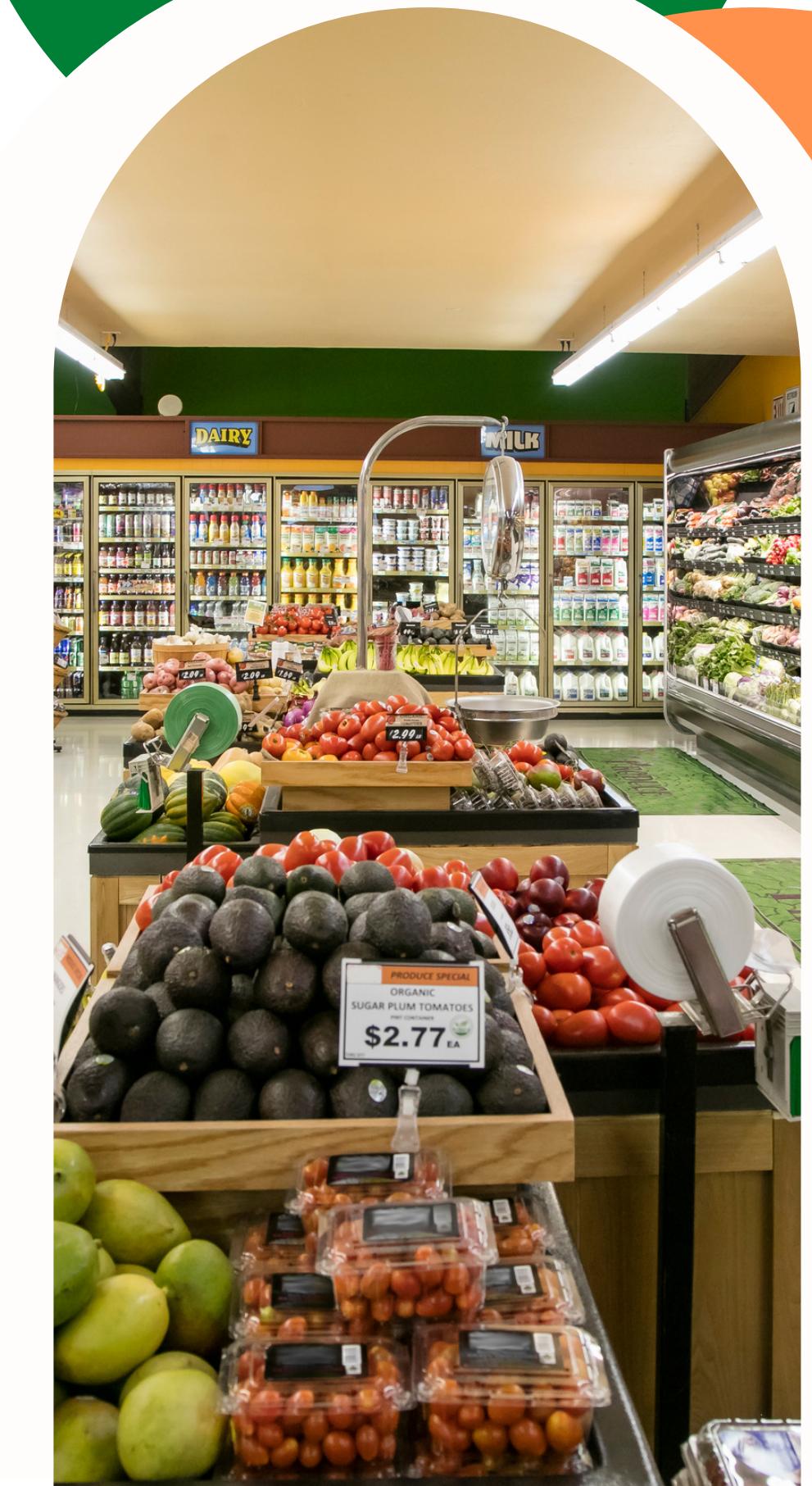
CHRISTINE



M. IFZAL
ASRIL



JOHANES
CRISTIAN



OUTLINE

**Business
Understanding**

**Exploratory Data
Analysis**

**Data
Preprocessing**

**Modelling &
Evaluation**

**Business
Recomendation**



Datalicious - Lotto Mart

Business Understanding

➤ Problem Statement

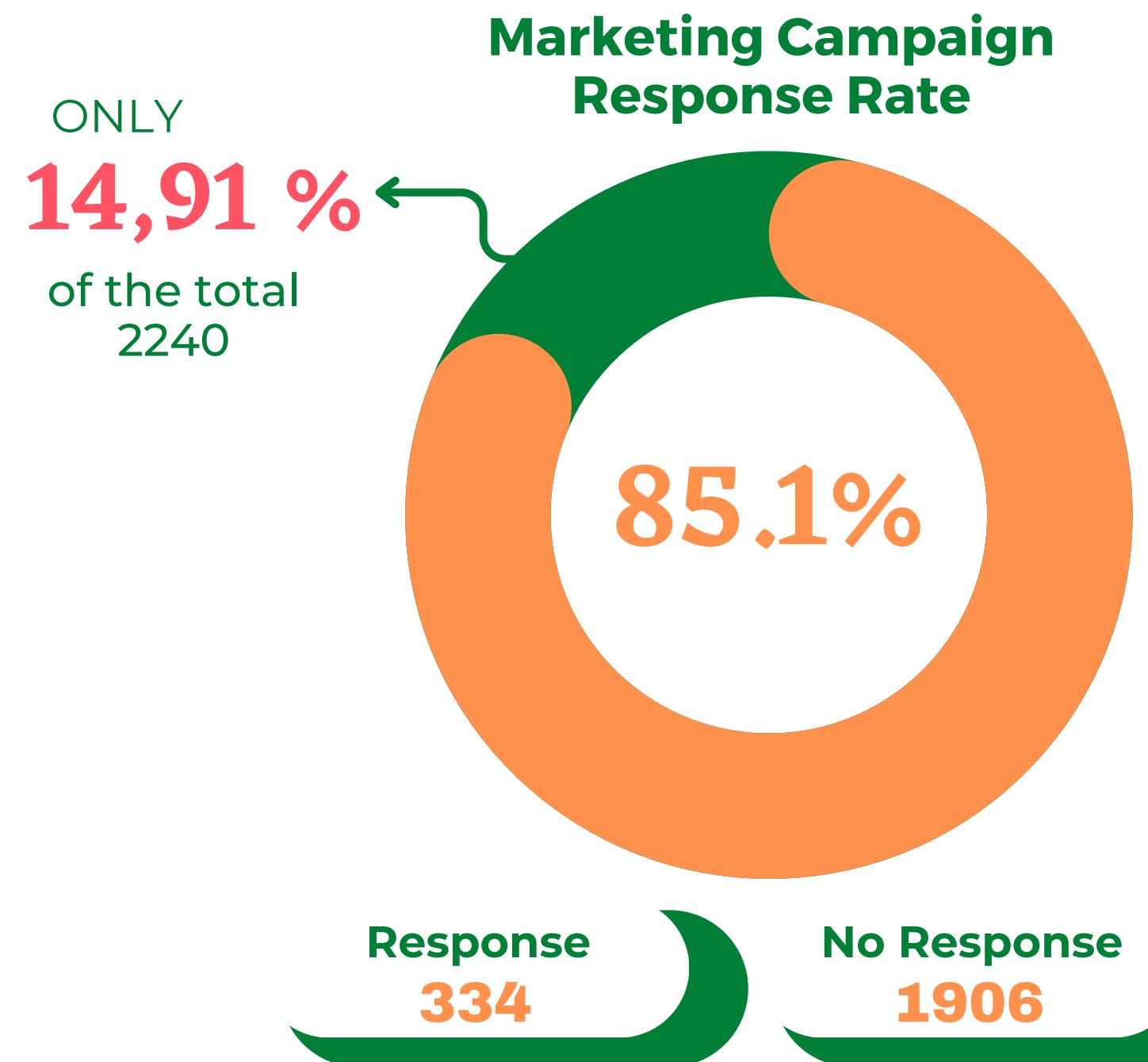
➤ Goals, Objective & Business Metrics



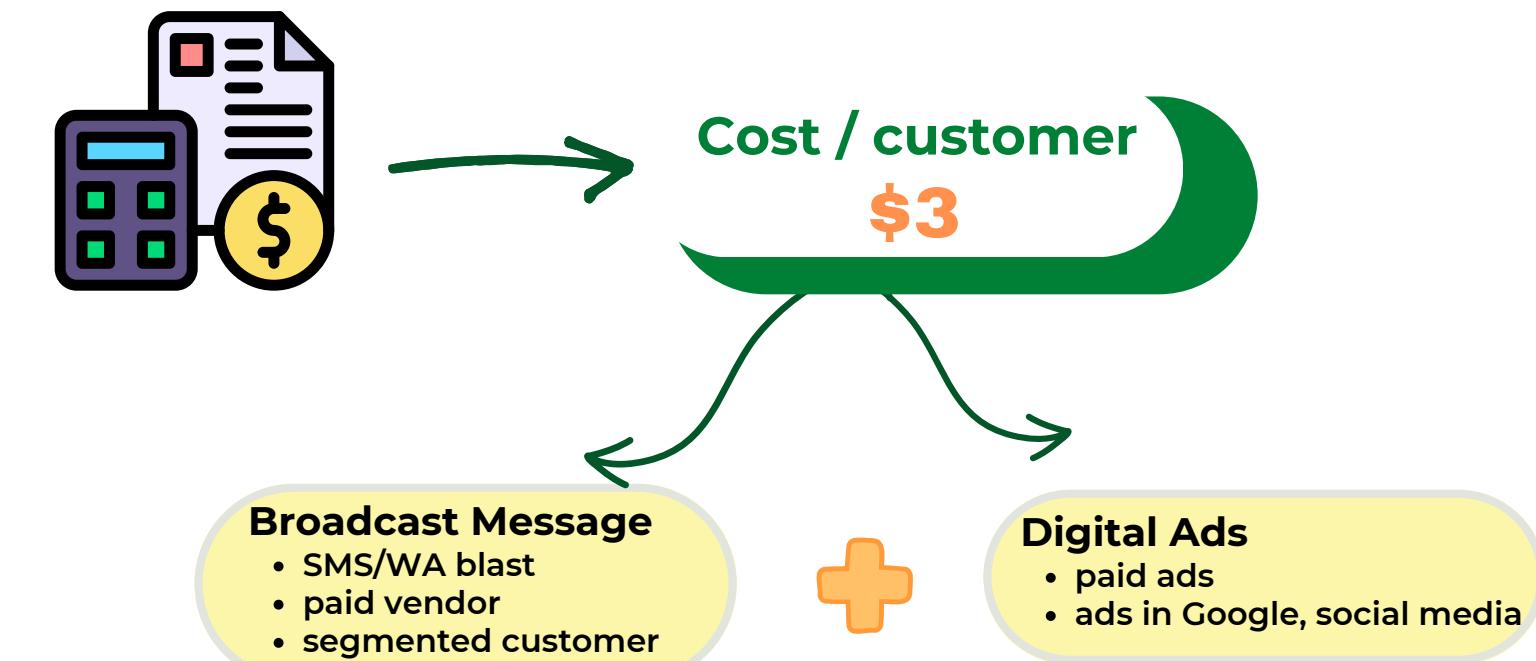
PROBLEM STATEMENT

DATALICIOUS

1. Lower Response Rate



2. Inefficient Cost



3. Profit Isn't Maximum



GOALS, OBJECTIVE & BUSINESS METRICS

Goals



Increasing response rate and minimizing marketing cost for each customers, so it can **boost profit** for the next marketing campaign

Objective



1. Create a **classification model** to predict which **customer groups** will **respond** for the next marketing campaign
2. Create a **clustering model** to make it easier for companies **to determine the right target of customers**

Business Metrics



1. Response Rate (%)
2. Net Profit Margin (%NPM)
3. Return of Investment (RoI)



Datalicious - Lotto Mart

Exploratory Data Analysis

› Dataset Information

› Preliminary Insight



DATA UNDERSTANDING

Accepted/Responses
Campaign

AcceptedCmp1
AcceptedCmp2
AcceptedCmp3
AcceptedCmp4
AcceptedCmp5
Complain
Response (target)

Customer Information

ID
Year_Birth
Education
Marital
Kidhome
Teenhome
Income
DtCustomer
Recency

Z_CostContact
Z_Revenue

Sales Product Type

MntFishProducts
MntMeatProducts
MntFruits
MntSweetProduct
MntWines
MntGoldProds

Number of Purchases
per Type

NumDealsPurchases
NumCatalogPurchas
NumStorePurchases
NumWebPurchases
NumWebVisitsMonth

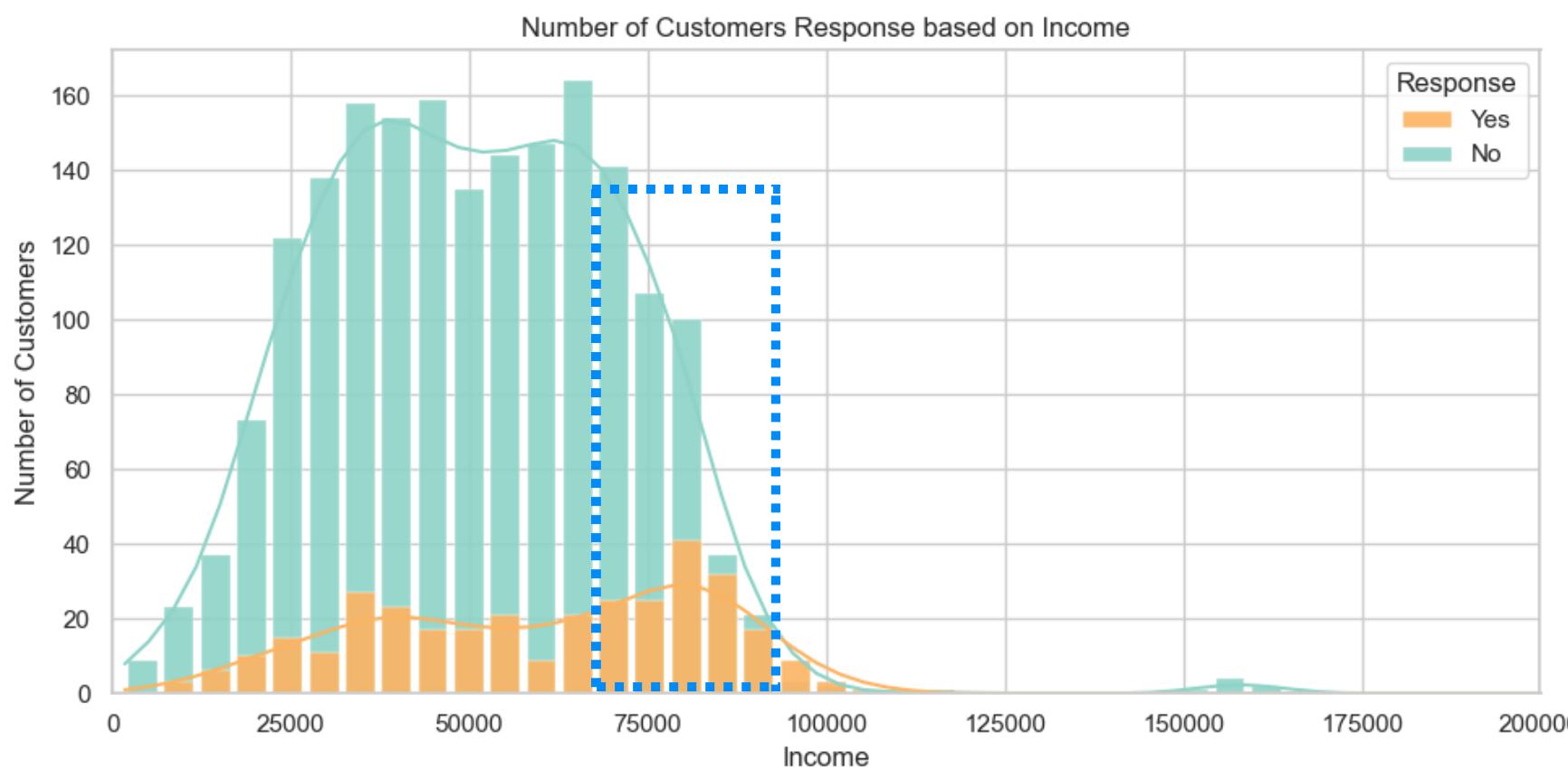
Dataset memiliki 29 kolom dan 2240 baris
Kolom income memiliki 2216 nilai non-null,dan 24 nilai null (1.07% data null)
Tidak terdapat data duplikat



Business Insights

1

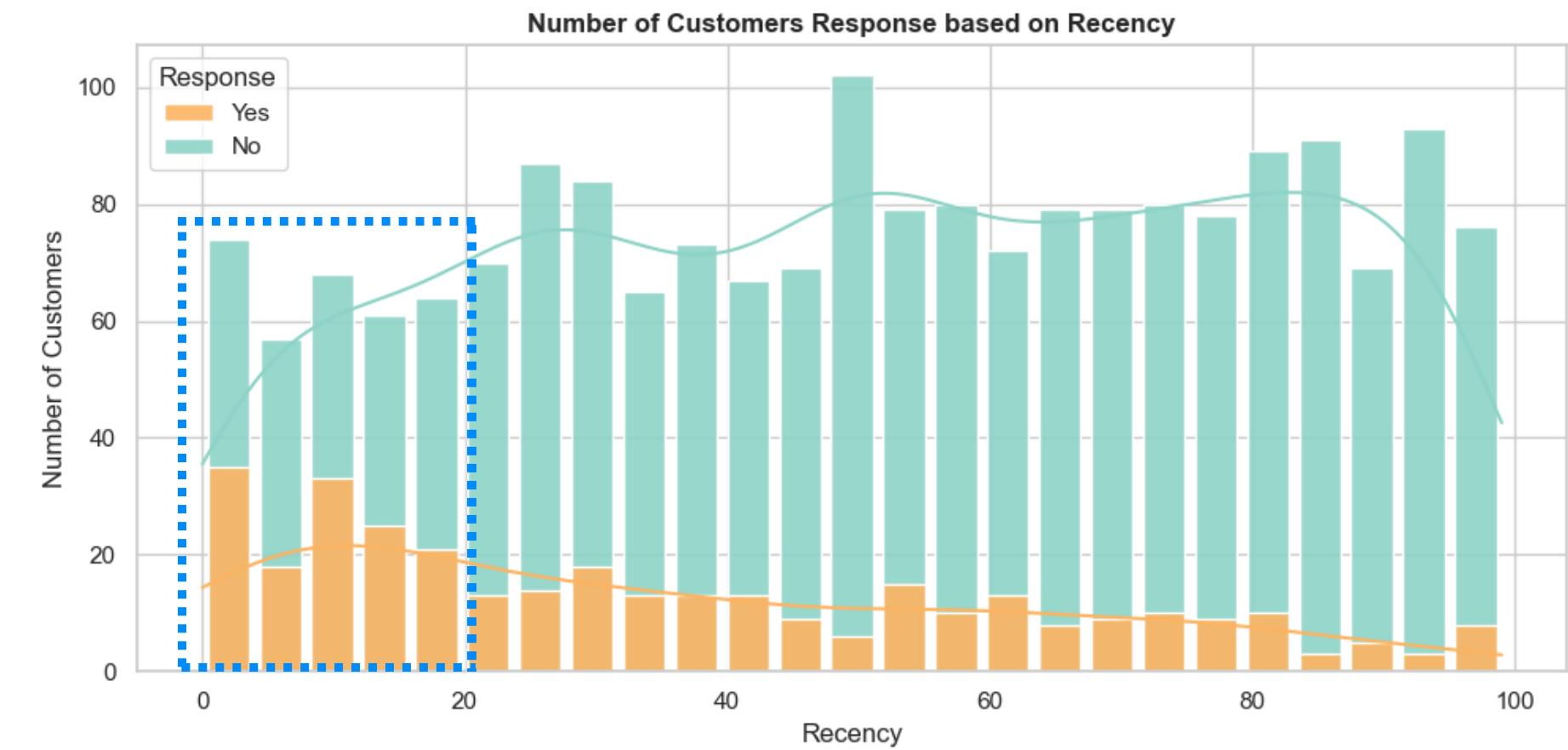
income $\geq \$75K$



From the Income visualization, it can be seen that the customers who respond the most come from customers with income $\geq \$75000$.

2

Recency = 0-20 days



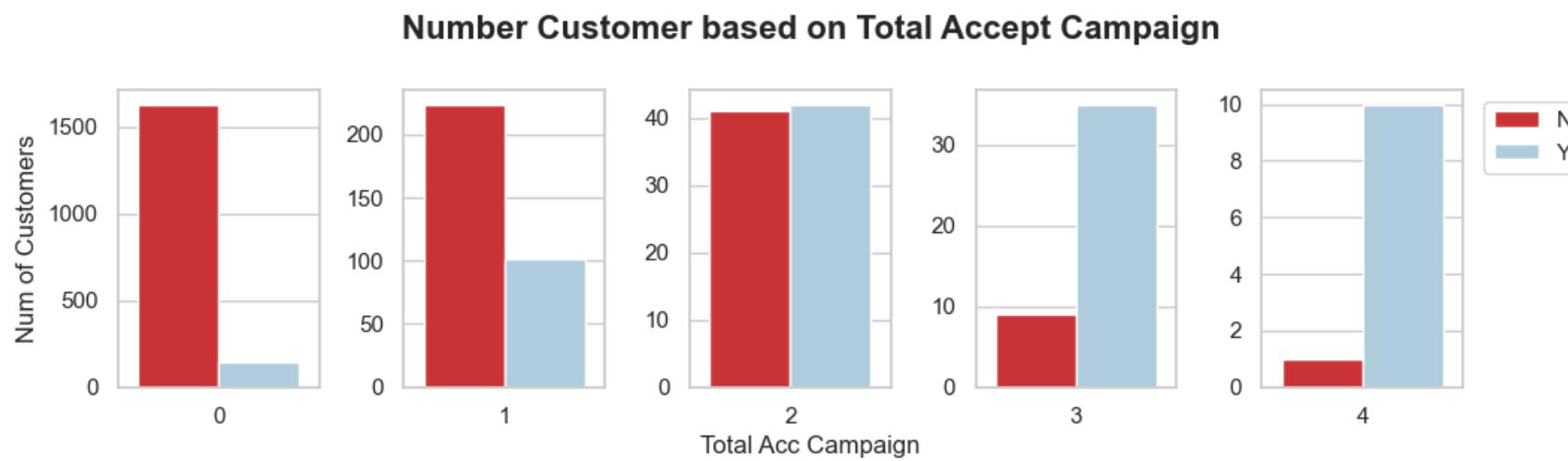
From the Recency visualization, it can be seen that customers who respond the most come from customers with low Recency.

Business Insights

3



Total Accepted Campaign by Customer >=2 (from 5 campaigns)



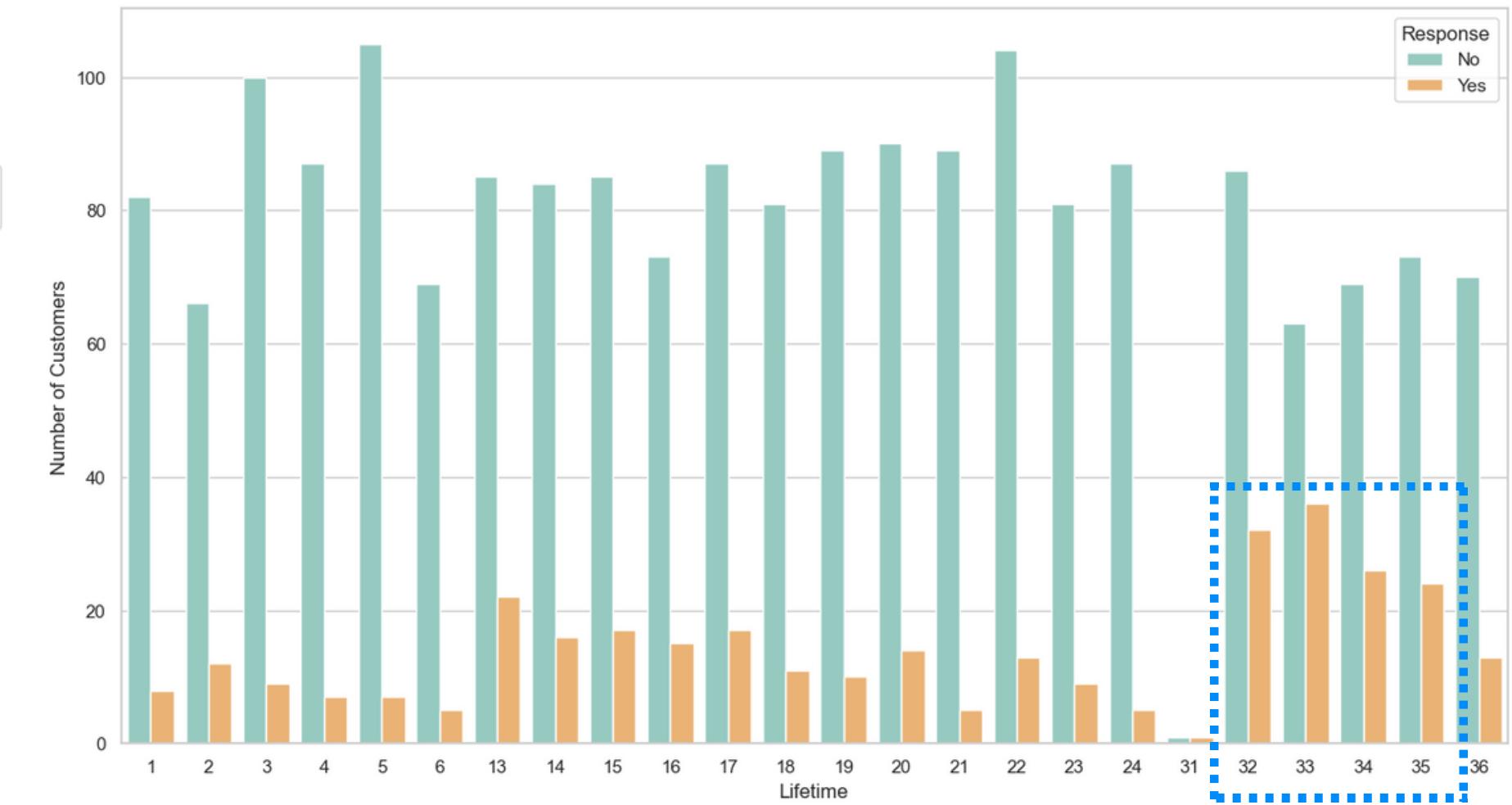
The most in our Five Campaigns is 0 (never responded), but there is a little potential in, only once (1) or twice (2) responded 325 and 83 Customers respectively.

4

Lifetime minimal 32 - 35 months



Number Customer based on Lifetime (months since the first purchase)



Customers who have a high lifetime tend to respond to the campaign

Business Insights

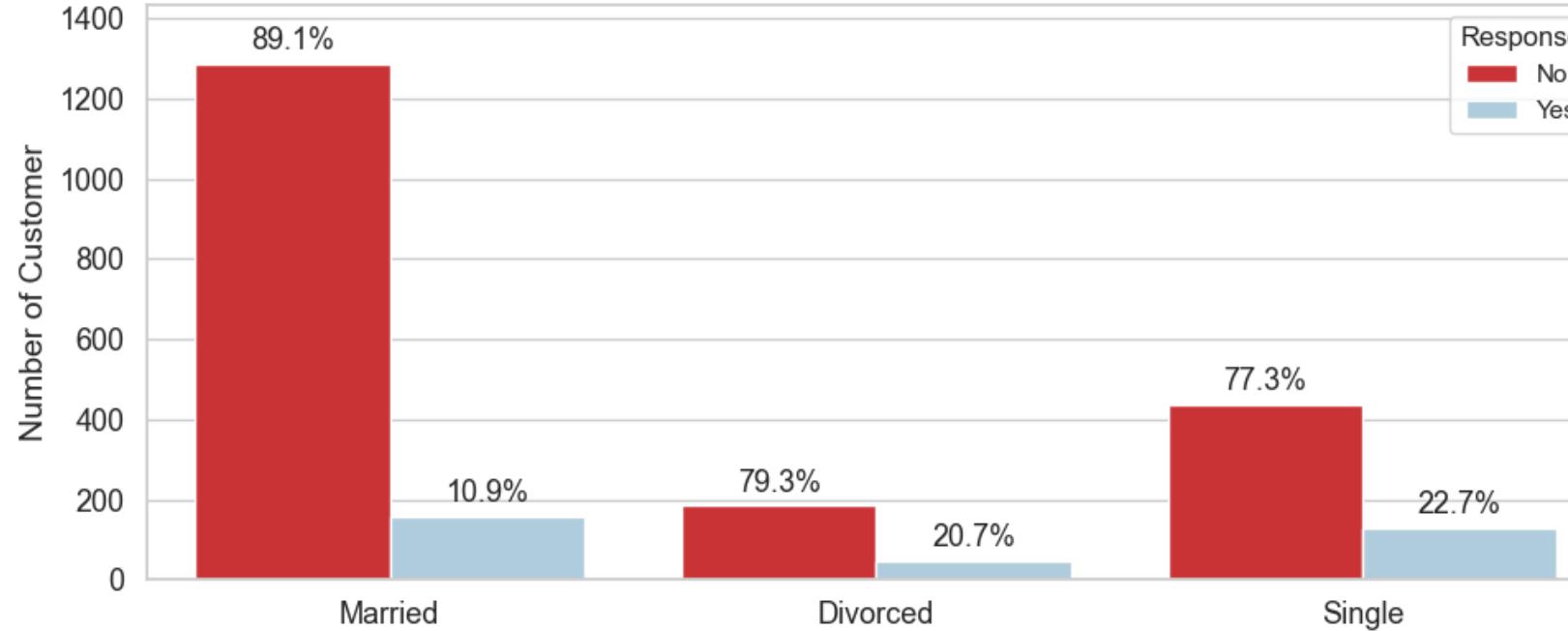
5

**Marital Status =
Married > Single > Divorced**



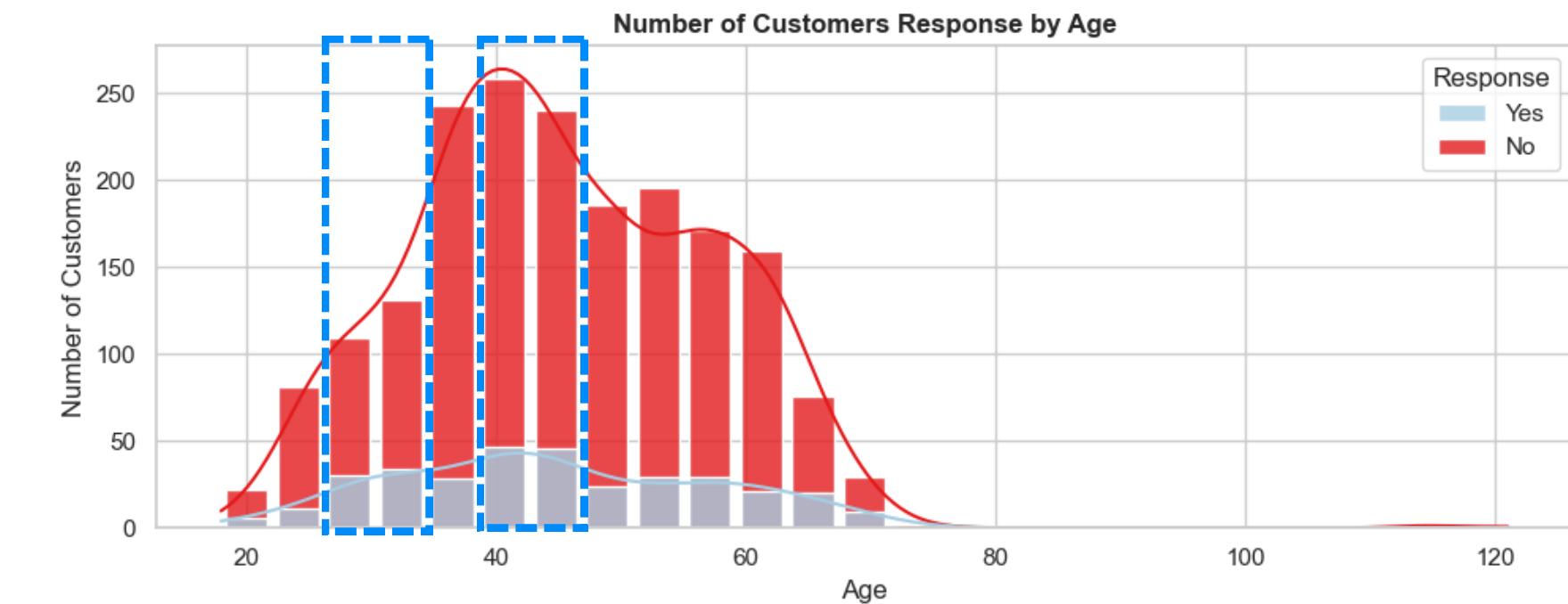
Comparison of Response Users in Marital Status

The older a person is, the higher the response value



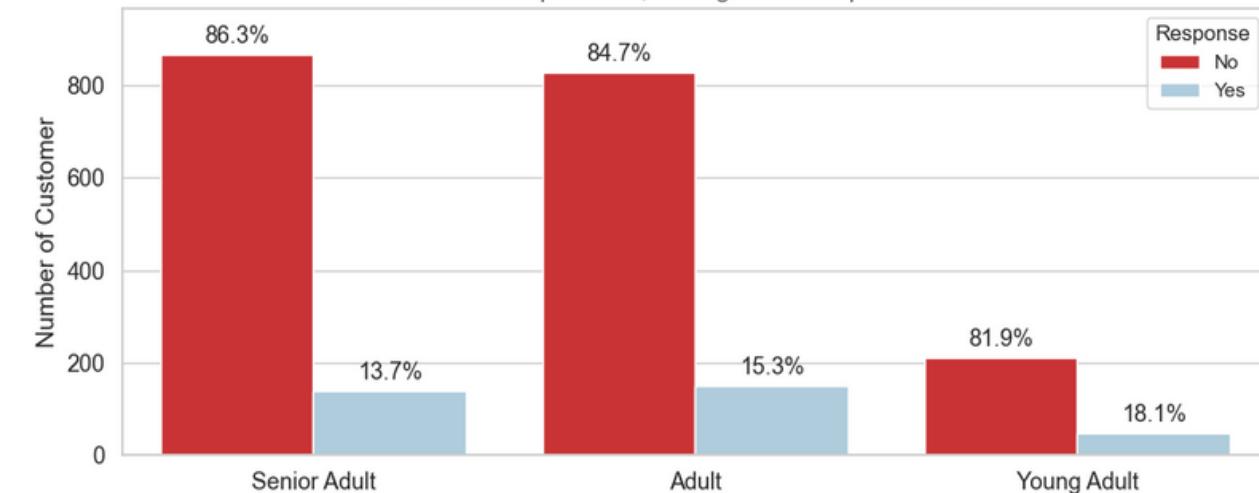
6

**Age &
Age Group (dominated by
Senior Adult group)**



Comparison of Response Users in Age Group

The older a person is, the higher the response value



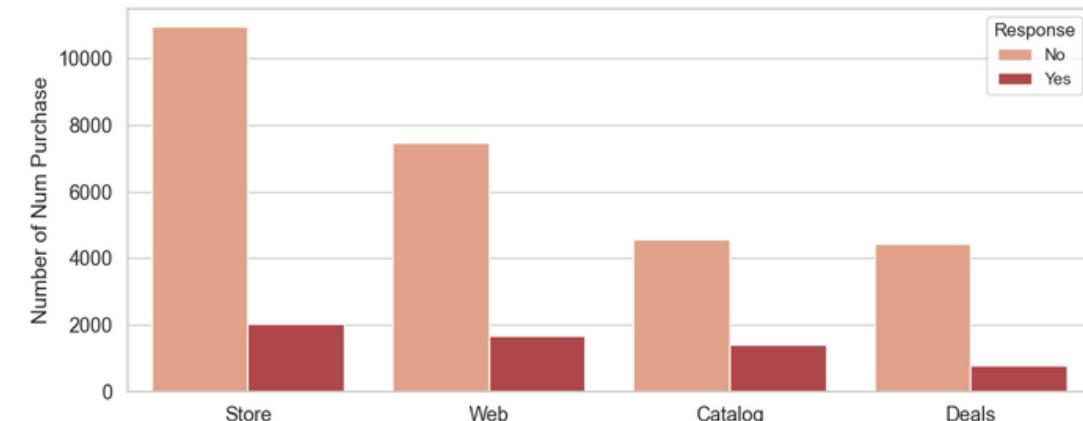
Business Insights

7

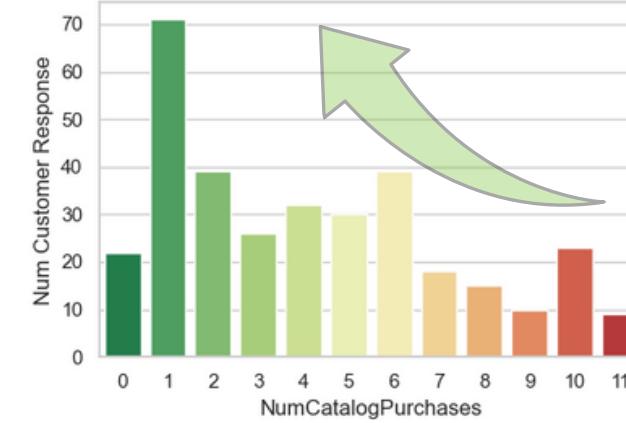
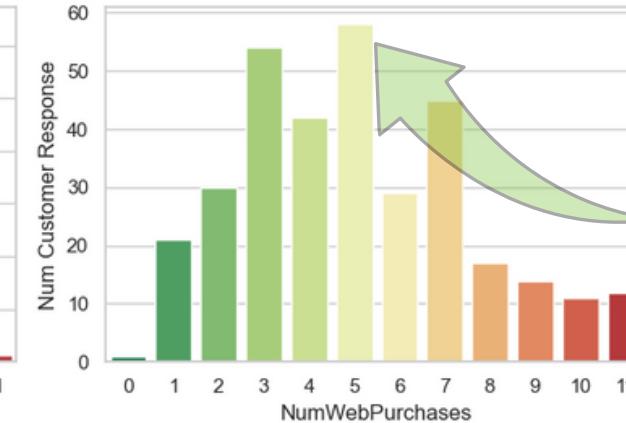
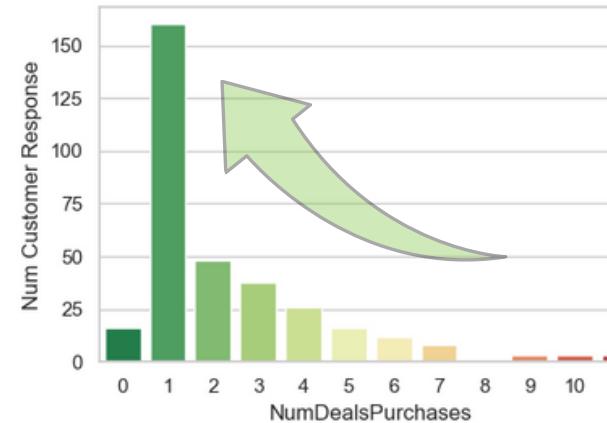
Responded Customer Purchase Types
(the fewer purchases, the more likely it is)



Comparison of Response Users in Num Purchase Type Purchase
Response customers mostly use Store and Web Purchase



Purchase Type based on Response



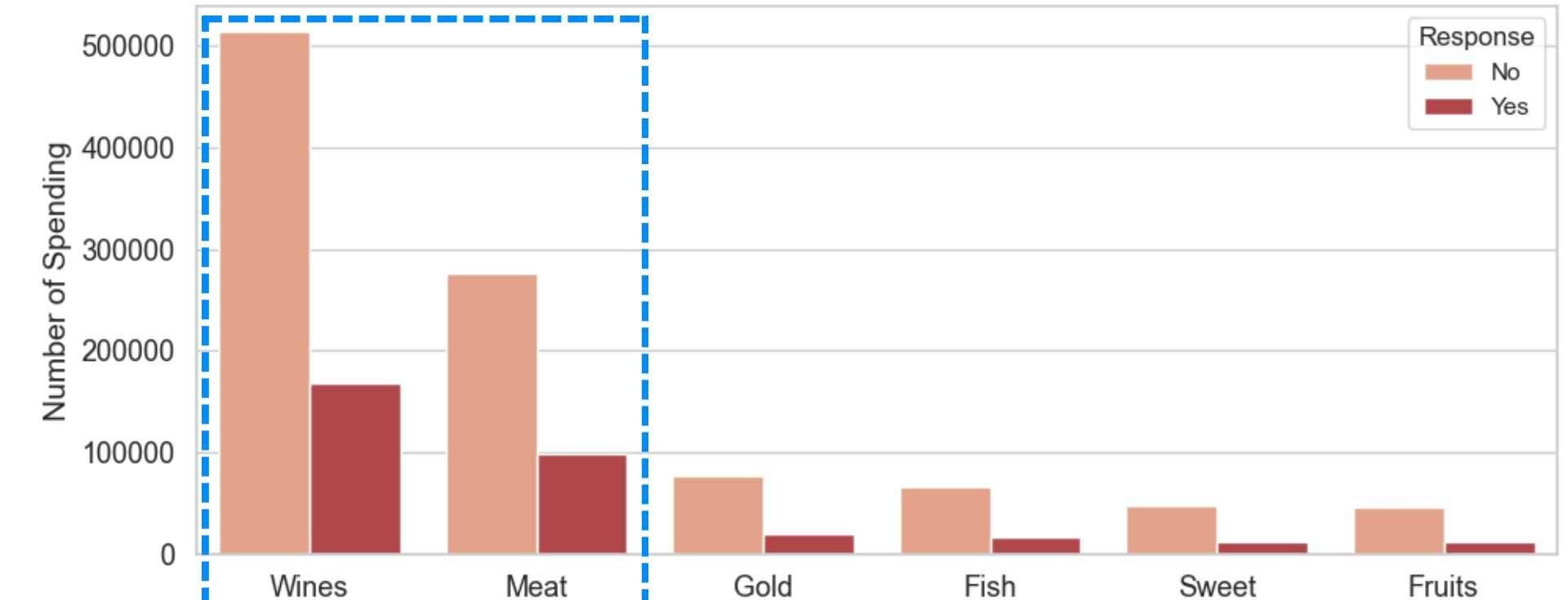
8

Types of Purchased Products
(customers tend to buy Wines and Meat)



Comparison of Response Users in Spending Product

Response customers mostly use Wines and Meat Product



Datalicious - Lotto Mart

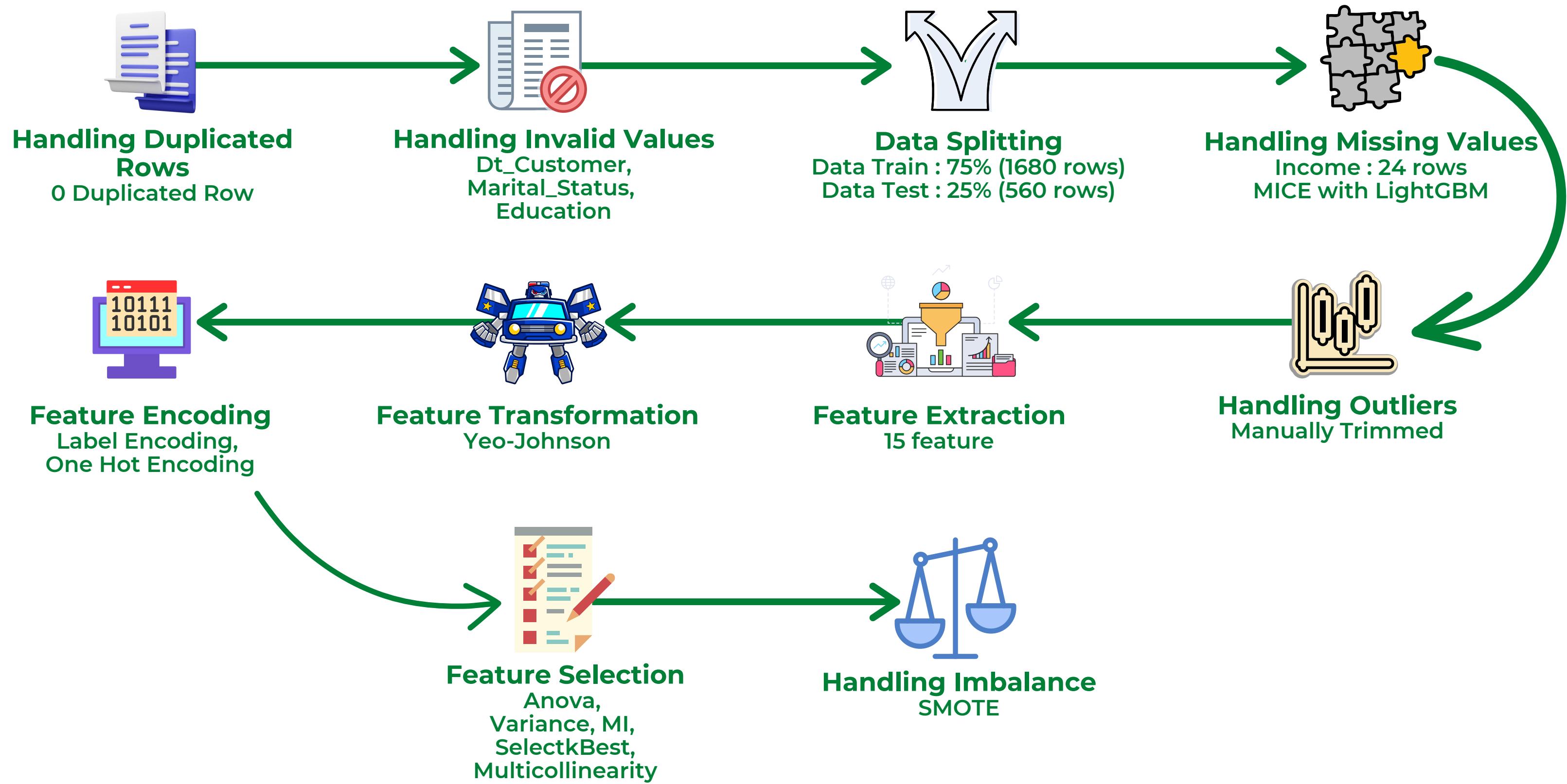
Data Preprocessing

- Handling Missing, Invalid
- Feature Transform & Engineering
- Feature Selection
- Handling Imbalance Data



DATA PRE-PROCESSING

CLASSIFICATION



DATA PRE-PROCESSING

CLASSIFICATION

AcceptedCmp1
AcceptedCmp2
AcceptedCmp3
AcceptedCmp4
Age 
Education
Income
Lifetime 

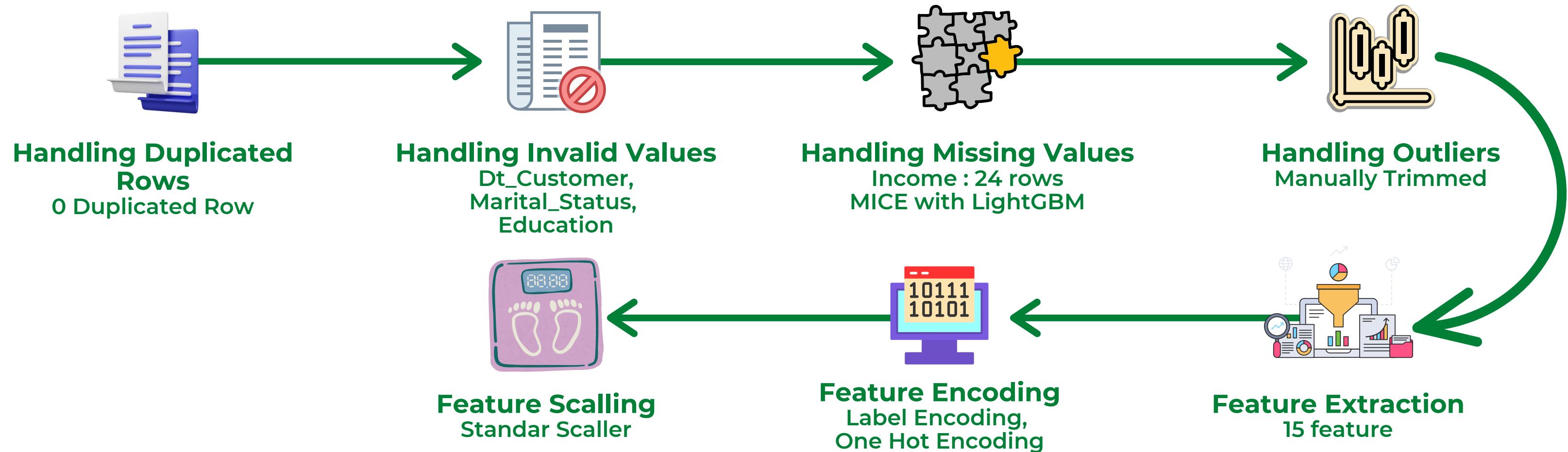
Married
MntGoldProds
NumCatalogPurchases
NumDealsPurchases
NumWebVisitsMonth
Recency
Recency_sgmt 
Teenhome
Total_cmp 

There are 17 features that will be used for modeling.
The stars is the result of Feature Engineering



DATA PRE-PROCESSING

CLUSTERING



Datalicious - Lotto Mart

Modeling and Evaluation

- Model Evaluation Parameters
- Machine Learning Techniques
- Model Comparison
- Feature Importance

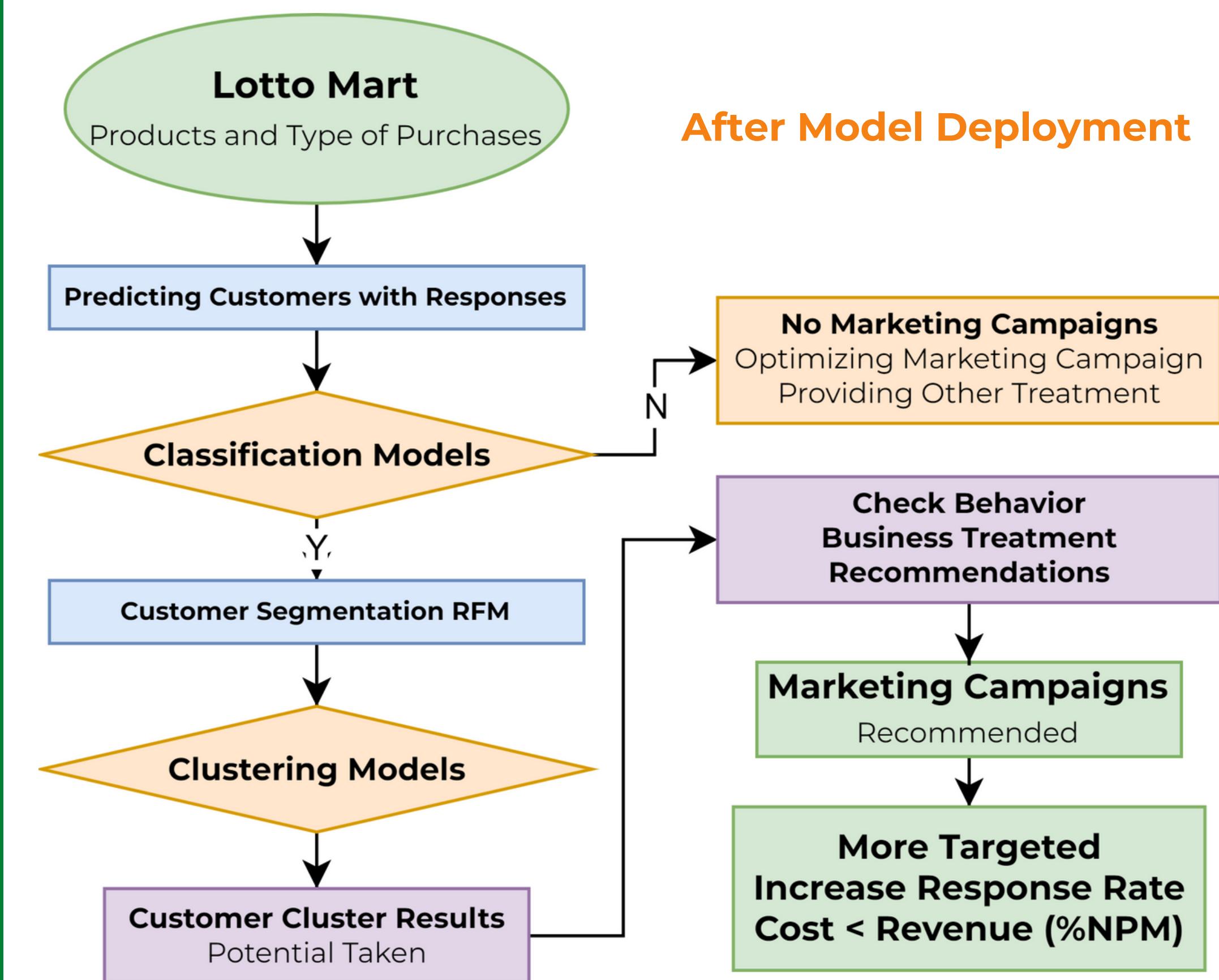


Business Flow Simulation

DATA LICIOUS



Before Model Deployment



...

Modeling and Evaluation

Model Evaluation Parameters

DATALICIOUS

Positive = Response Campaign
Negative = No Response Campaign

False Positive

Main target to be reduced

False Negative

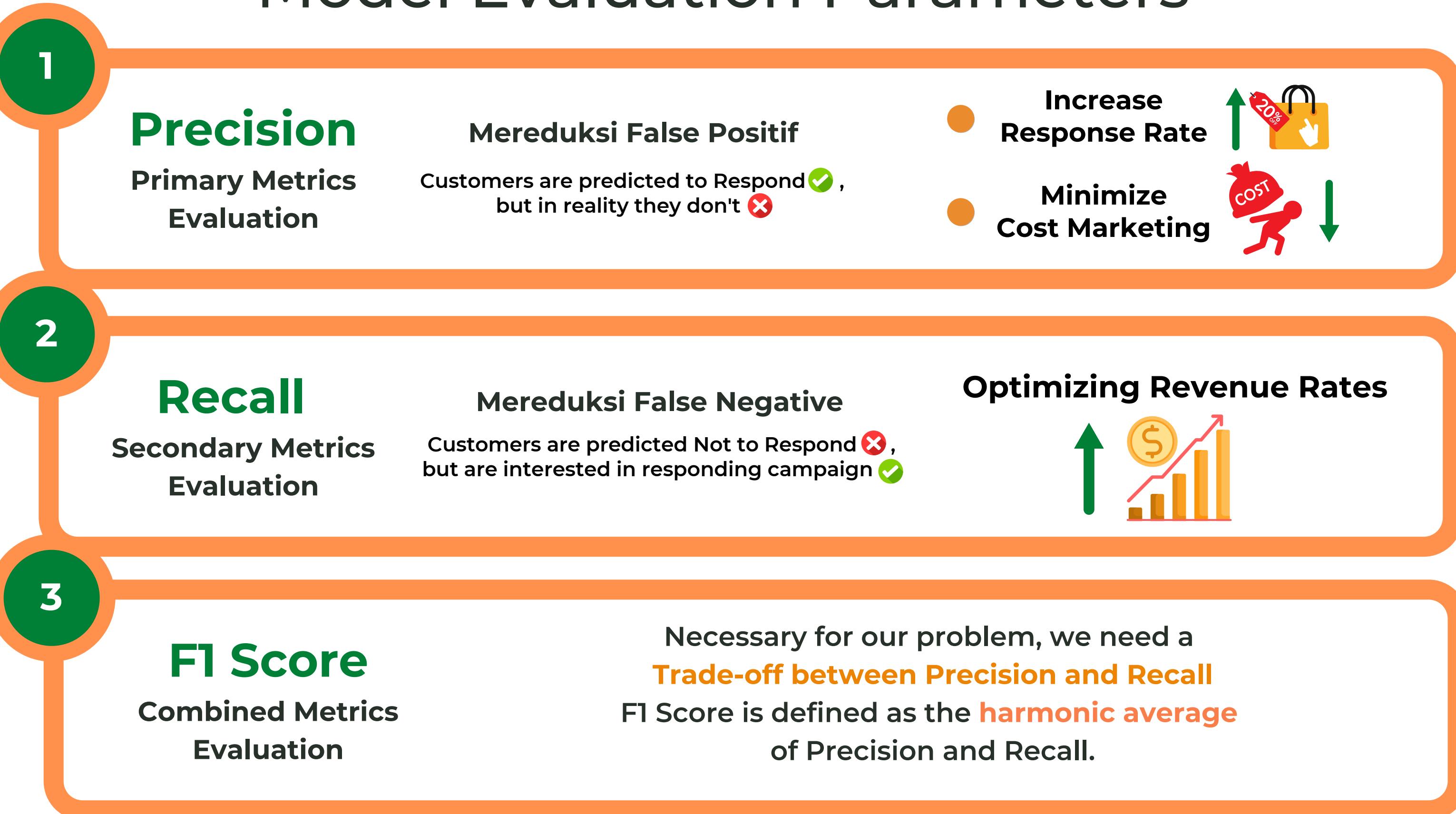
Second target to be reduced



Modeling and Evaluation

Model Evaluation Parameters

DATA LICIOUS



Modeling and Evaluation

Machine Learning Techniques

-
- 1. Decision Tree
 - 2. Random Forest 
 - 3. Logistic Regression
 - 4. Gaussian Naive Bayes
 - 5. K-Nearest Neighbor
 - 6. MLP Classifier (Neural Network)
 - 7. Adaboost Classifier
 - 8. XGBoost Classifier
 - 9. Gradient Boosting Classifier
 - 10. Support Vector Machine

 **Random Forest** 
Best Fit Models



Modeling and Evaluation

Model Comparison

Model (Test)	Accuracy	Precision	Recall	F1 Score	Cross Val F1 (k=5)	ROC AUC	Cross Val ROC AUC (k=5)
Random Forest	0.909	0.786	0.530	0.633	0.519	0.905	0.896
XGBoost Classifier	0.889	0.652	0.542	0.592	0.525	0.895	0.887
Logistic Regression	0.861	0.523	0.675	0.589	0.510	0.900	0.885
Support Vector Machine	0.859	0.520	0.614	0.564	0.448	0.895	0.884
K-Nearest Neighbors	0.845	0.482	0.639	0.549	0.400	0.842	0.810
Adaboost Classifier	0.855	0.511	0.578	0.542	0.539	0.892	0.885
MLP Classifier	0.846	0.483	0.518	0.500	0.558	0.830	0.887
Gradient Boosting Classifier	0.846	0.482	0.494	0.488	0.487	0.725	0.752
Naive Bayes	0.830	0.441	0.542	0.486	0.451	0.838	0.824
Decision Tree	0.838	0.456	0.494	0.474	0.488	0.699	0.718

The **Precision, Recall and F1 Score** values on the best evaluation results are produced by

Random Forest, XGBoost Classifier, dan Logistic Regression



Modeling and Evaluation

Model Comparison

Models	Precision (Train)	Precision (Test)	Recall (Train)	Recall (Test)	F1 Score (Train)	F1 Score (Test)	Total Diff
Logistic Regression (HT)	0.755	0.528	0.660	0.675	0.704	0.593	0.323
Random Forest (HT3) ✓	0.867	0.702	0.584	0.482	0.698	0.571	0.394
Decision Tree (HT2)	0.719	0.441	0.574	0.542	0.639	0.486	0.463
XGBoost Classifier (HT2)	0.826	0.595	0.699	0.566	0.757	0.580	0.541
K-Nearest Neighbors (HT)	0.763	0.465	0.871	0.639	0.814	0.538	0.806
XGBoost Classifier (HT)	0.976	0.710	0.920	0.530	0.947	0.607	0.996
Random Forest (HT2)	0.972	0.724	0.912	0.506	0.941	0.596	0.999
Random Forest (HT1)	0.997	0.772	0.993	0.530	0.995	0.629	1.054
Decision Tree (HT)	1.000	0.484	0.990	0.554	0.995	0.517	1.430

After doing **Hyperparameter Tuning**

The **Most Stable Total Diff (Gap) Value** in the train and test dataset
with **Fairly High Metrics** is **Random Forest**

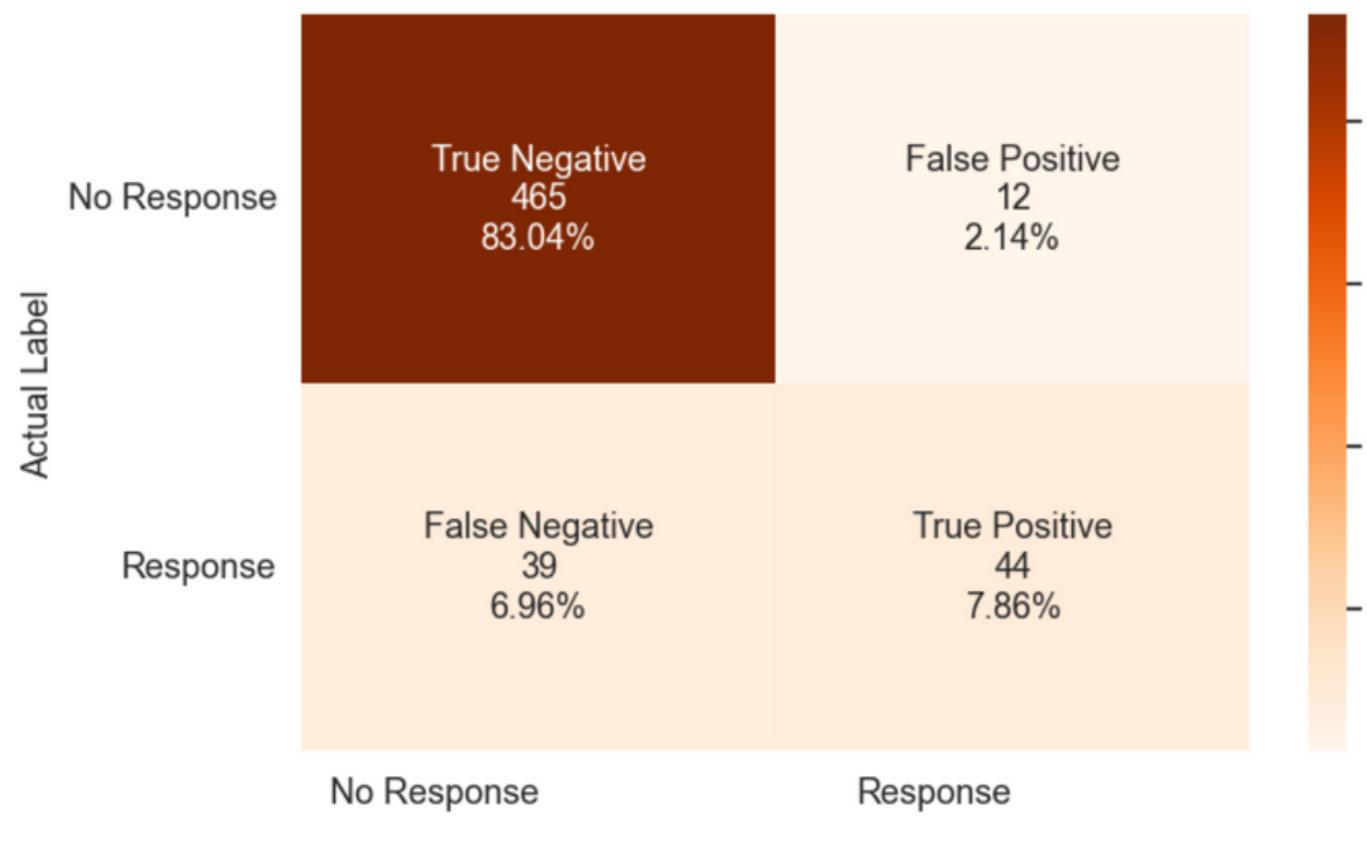
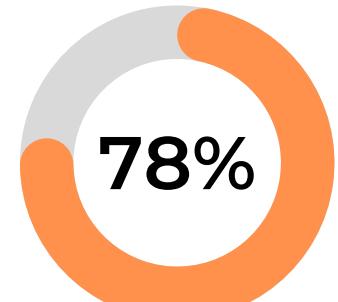
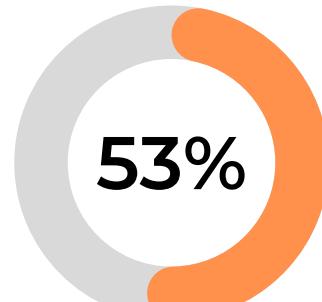
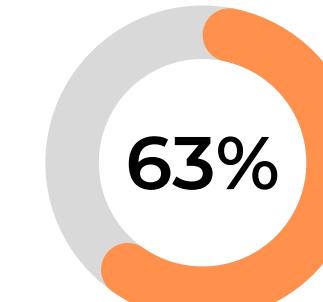


Modeling and Evaluation

Model Selection

DATA LICIOUS

Confusion Matrix for Testing Model (Random Forest)

**Precision****Recall****F1 Score**

True Positif	Predicted Respond ✓	True, Response ✓
True Negative	Predicted Not to Respond ✗	True, No Response ✗
False Positif	Predicted Respond ✓	False, No Response ✗
False Negative	Predicted Not to Respond ✗	False, Response ✓

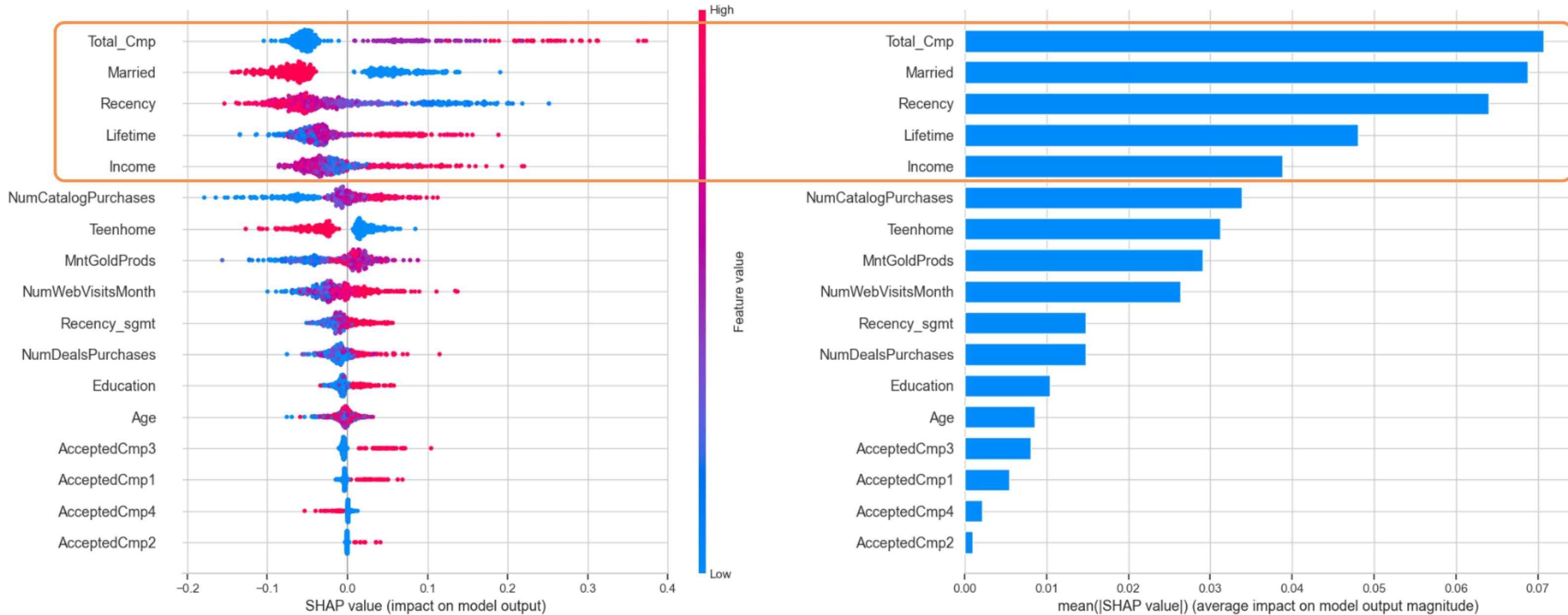


Modeling and Evaluation

Feature Importance

...

DATA LICIOUS



Positive = Response Campaign

Negative = No Response Campaign

Datalicious - Lotto Mart

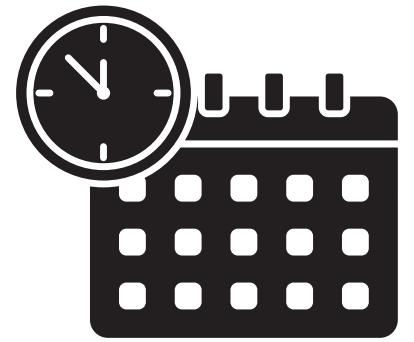
Modeling and Evaluation

- > Clustering with RFM Analysis
- > Clustering Results



Modeling and Evaluation

Clustering with RFM Analysis



Recency

The freshness of customer activity



Resources : the number of days since the last purchase



Frequency

The frequency of the customer transactions



Resources : the number of transactions made from various types of purchases



Monetary

The intention of customer to spend



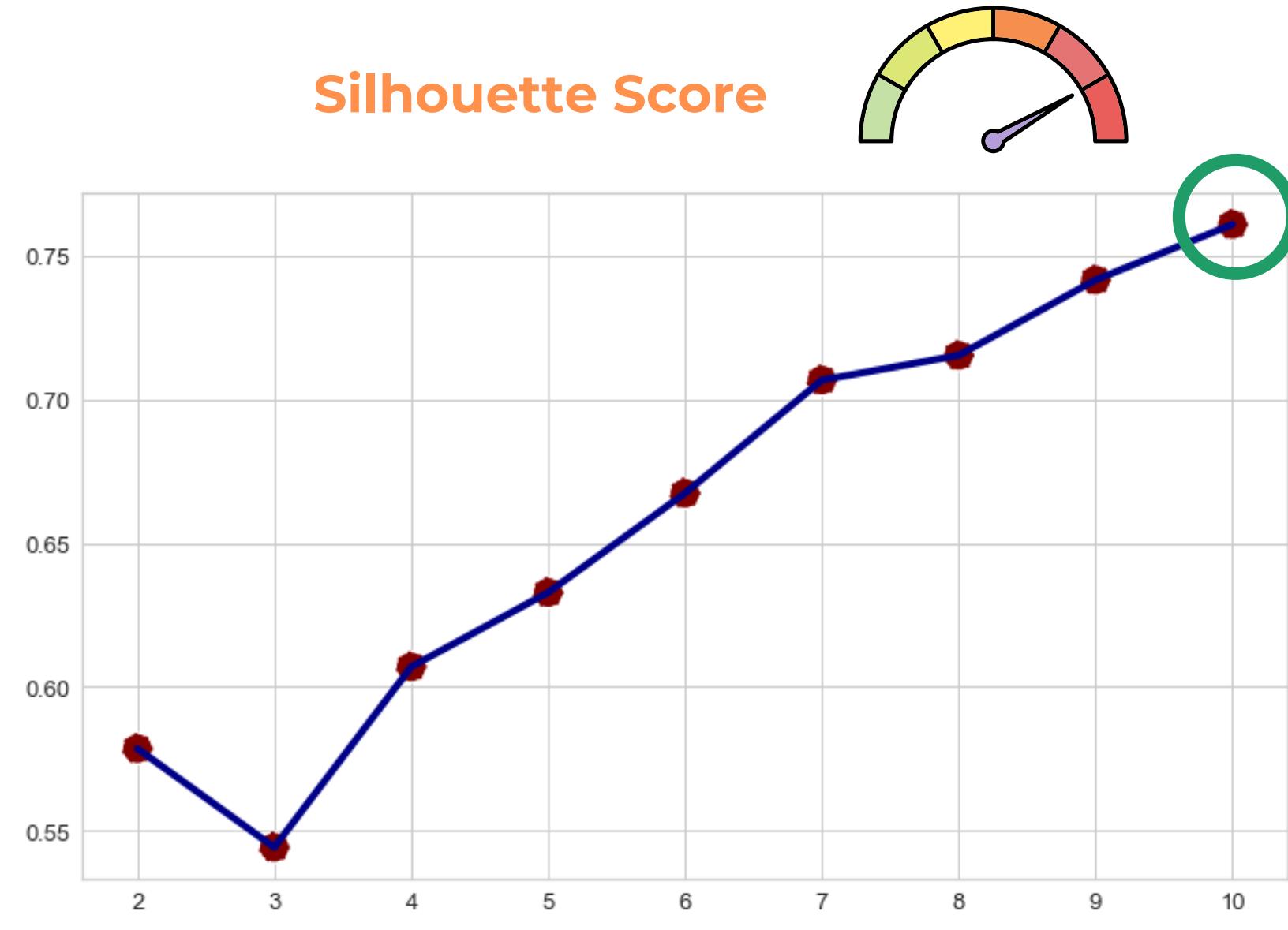
Resources : the amount spent on various products

Modeling and Evaluation

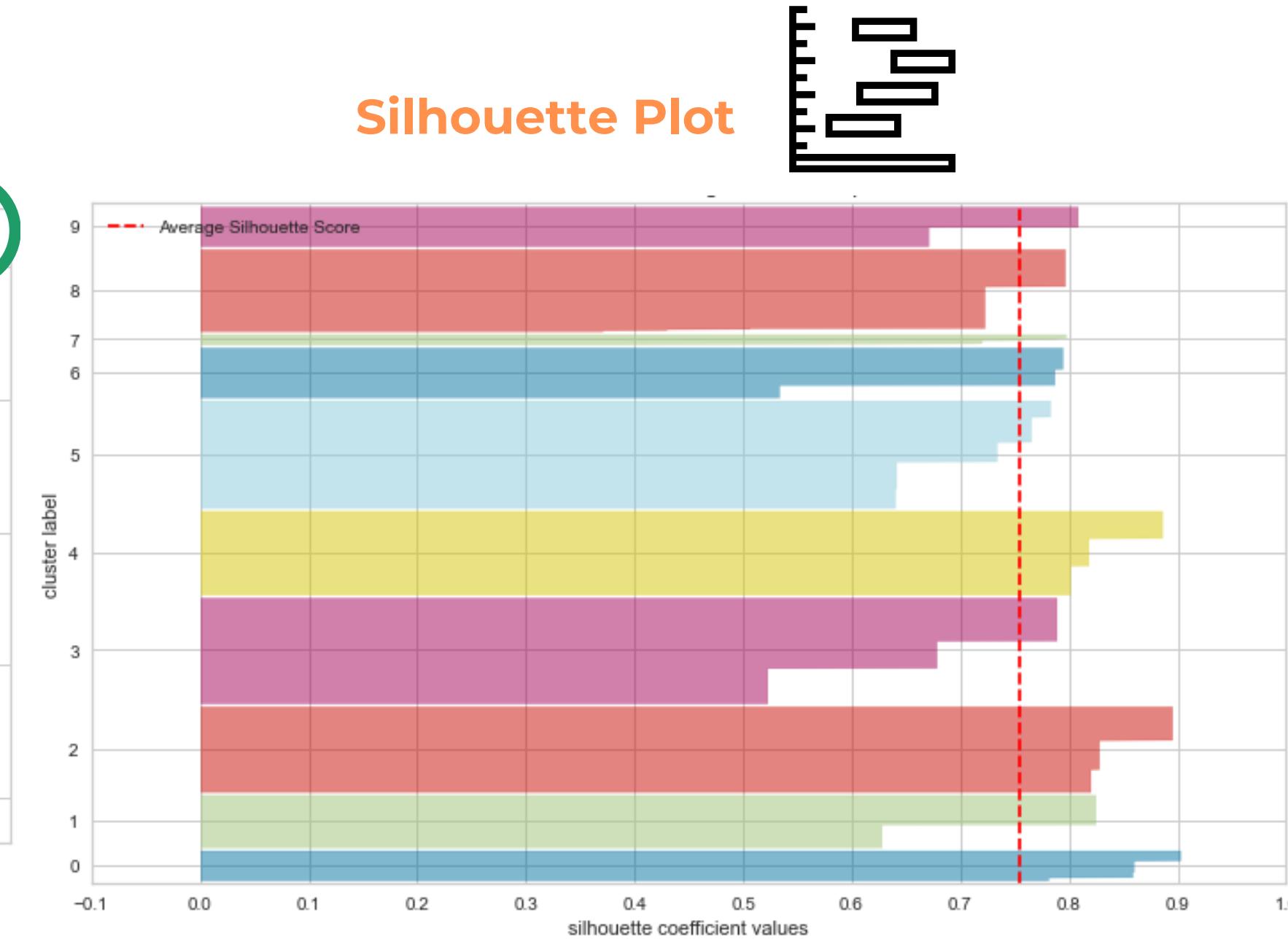
Clustering with RFM Analysis

DATALLICIOUS

Silhouette Score



Silhouette Plot



Customers will be divided into 10 clusters with the average silhouette score is 0.76



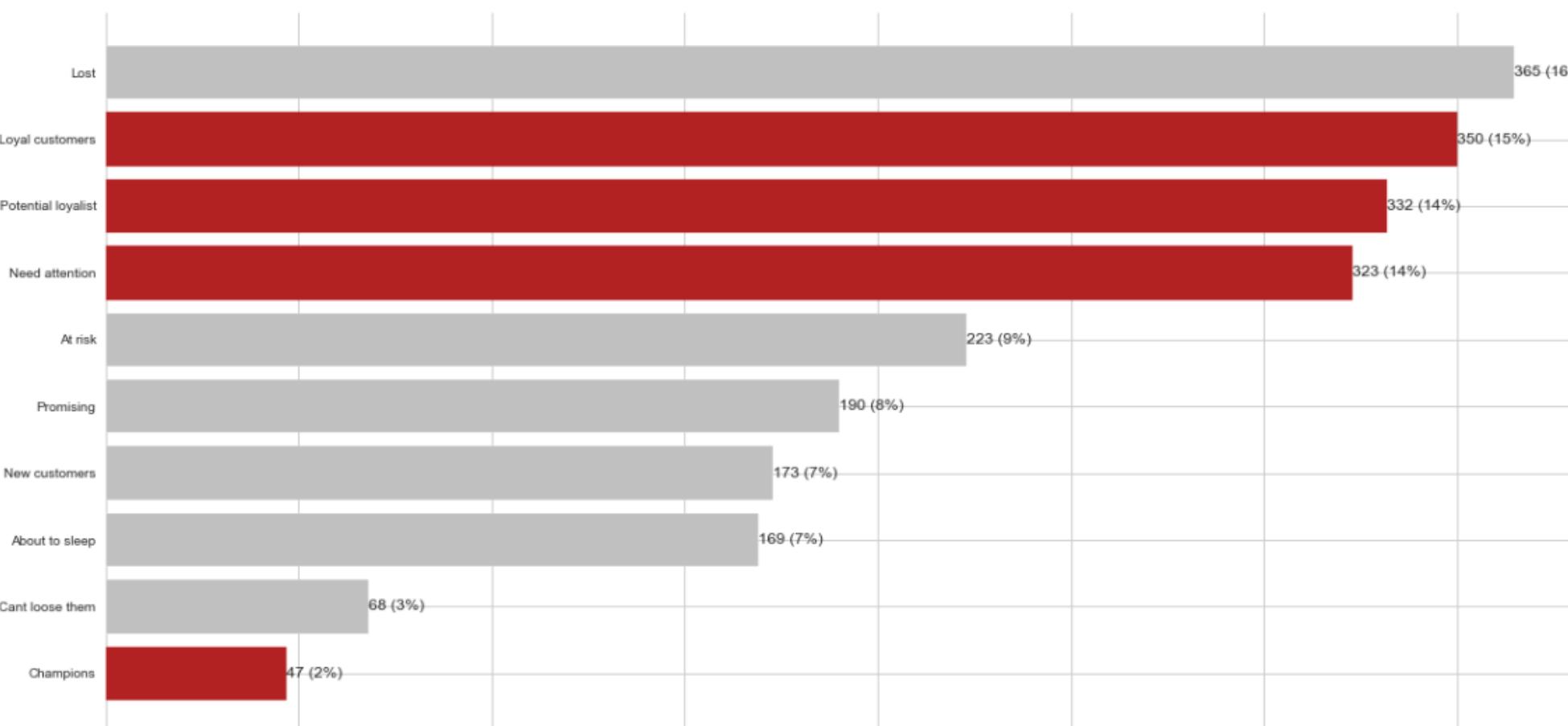
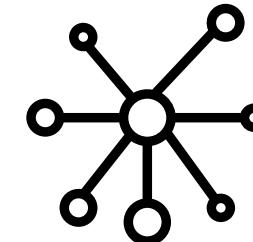
Modeling and Evaluation



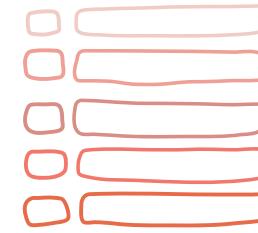
Clustering Results

DATA LICKIOUS

Cluster Distribution



Cluster Type



Segment	Recency	Frequency	Monetary
About to sleep	medium	low	low
At risk	high	medium	high
Cant loose them	high	high	high
Champions	low	high	medium
Lost	high	medium	medium
Loyal customers	medium	high	high
Need attention	medium	medium	medium
New customers	low	low	low
Potential loyalist	low	medium	medium
Promising	medium	low	low

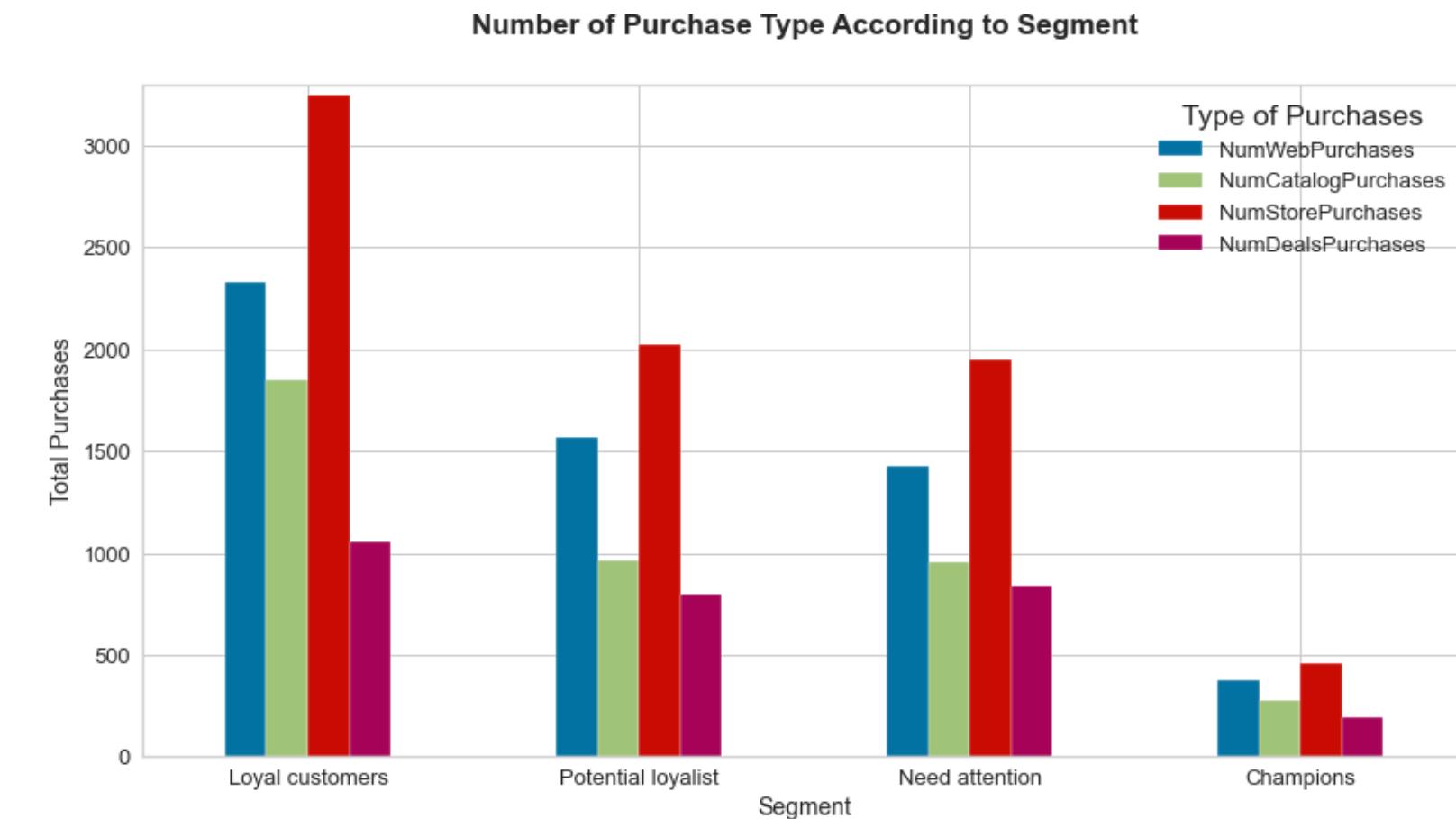
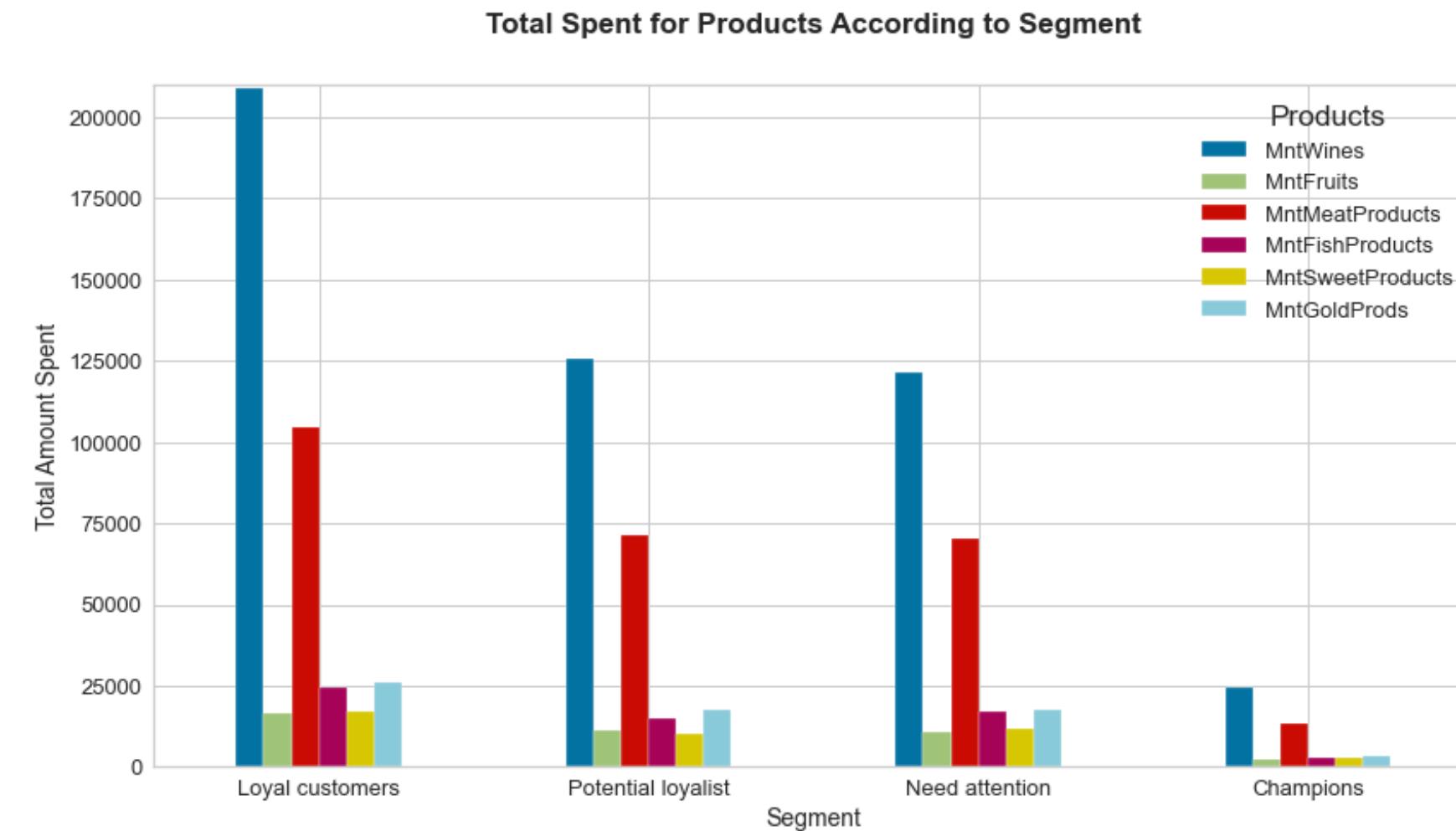
In supporting goals to get business recommendations, we focus on 4 customer segments with a total of 1052 customers.



Modeling and Evaluation

Clustering Results

DATALLIOUS



The most purchased product is wines and The most frequent type of purchase is store purchases



Datalicious - Lotto Mart

Business Recommendation

- > Recommendation
- > Business Impact
- > Simulation



Business Recommendations



Discount/Flash Sale

price reduction with a certain period of time

especially **for customers who have never response a campaign or 1x response the campaign only**

Goals:

to **increase customers shopping/purchasing interest**



Vouchers/Rewards

especially **for customers who have response the campaign at least 2 campaigns**

customers will get coupons and can be exchanged for discounts or other benefits such as points, free shipping, free gifts, etc.

Goals:
maintain customer's interest and retention



Promo Bundling/Special Offer

can implement **for all customers**

combining the main product with supporting products for more economical price for customers such as Wines with Meats, Wines with Gold Products, etc.

Goals:
encourage customers to buy more products

Business Recommendations





Business Simulation

Business metrics are quantifiable measures used to track business processes to judge the performance level of your business.

Response Rate

Net Profit Margin

Return of Investment (RoI)

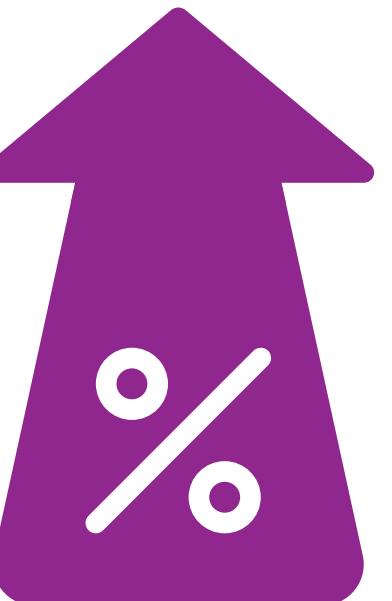


Response Rate

14,91%

before model

63,66



78,57%

after model

True Positive (TP)= 44
False Positive (FP)= 12

response rate = TP/(TP+FP)

Net Profit Margin / NPM

-82,91%

Before Modelling

Total Response (1=accept response)

334

Total Campaign (all customers)

2240

Total Cost(\$)

= Total campaign*cost
= $2240 * 3$

6720

Total Revenue(\$)

= Total response * revenue
= $334 * 11$

3674

Total Profit(\$)

= Total Revenue - Total Cost
= $3674 - 6720$

-3046

NPM (%)

= (Total Profit/ Total Revenue) *100

-82.91

148,2%



a customer

Revenue (\$) **11**

Cost (\$) **3**

Broadcast Message
• SMS/WA blast
• paid vendor
• segmented customer

Digital Ads
• paid ads
• ads in Google, social media

65,29%

After Modelling

Total Campaign (TP+FP)

56

Total Cost(\$)

= Total campaign*cost
= $56 * 3$

168

Total Revenue(\$)

= TP * revenue
= $44 * 11$

484

Total Profit(\$)

= Total Revenue-Total Cost
= $484 - 168$

316

NPM (%)

= (Total Profit/ Total Revenue) *100

65.29

Return of Investment

Average Order Size (AOS)	= Total Revenue/Total Order = $(2240*11)/2240$ = $24640/2240$	11
Average Order Frequency (AOF)	= Total Order/Total Customer = $2240/2240$	1
Average Customer Value (ACV)	= AOS/AOF	11
Average Cust. Llifetime (ACL)(Year)	= first order date-last order date	1
Customer Lifetime Value (CLV) (\$)	= ACL*ACV	11

ROI = ~3.0x

- Good ROI (Manzer 2017)
- The company receives \$3 per \$1 spent to acquire the customer

Before Modelling		
CLV	number of new customers	334
CAC	= Total Marketing Cost/Number new Customers = $(2240*3)/ 334$	20.12
ROI	= CLV : CAC	0.55
After Modelling		
CLV	number of new customers	44
CAC	= Total Marketing Cost/Number new Customers = $((TP+FP)*cost) / 44$ = $(56*3) / 44$	3.82
ROI	= CLV : CAC	2.88

Thank You

END OF SLIDE >

