CRM Data Analyst Case

Andressa Ribeiro
e-mail: andressaribeiro@usp.br

BACKGROUND

Company

Established player in retail food sector with hundreds of thousands of registered customers serving nearly one million consumers annually.

Products

Wines, rare meat products, exotic fruits, specially prepared fish, and sweet products, offered in both gold and regular variants.

Financial

Solid revenues and healthy bottom line in past years, but limited profit growth prospects for the next three years prompt strategic initiatives.

Sales channels: Physical stores, catalogs, and company website.

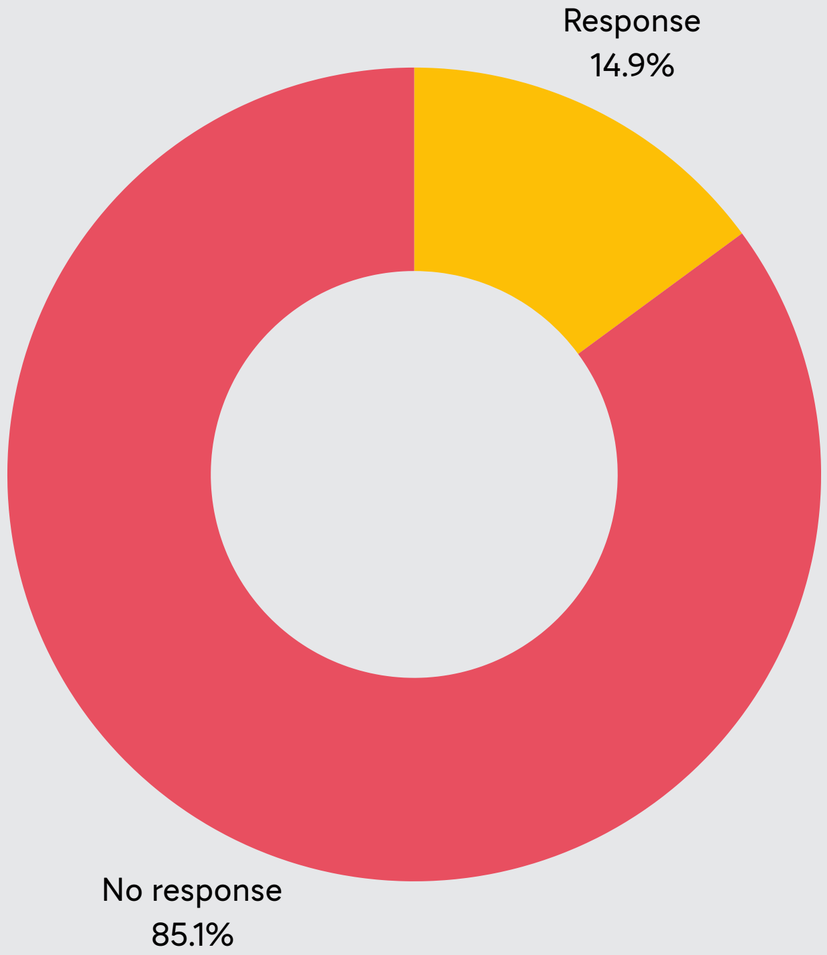
PROBLEM STATEMENT

CAMPAIGN PILOT

Involving 2,240 randomly selected customers, contacted via phone for gadget offer.

Success rate: 15%.
Financial Metrics: cost per person is 3 MU and revenue per conversion is 11 MU. Total cost of the campaign was 6,720 MU and profit was -3,046 MU.

Objective: Build a predictive model to identify potential buyers for the new gadget offer, maximizing profitability for future campaigns.



EXPLORATORY DATA ANALYSIS

CUSTOMER INFO

ID
Year_Birth
Education
Marital_Status
Kidhome
Teenhome
Income
DtCustomer
Recency

SALES PRODUCTS

MntFishProducts
MntMeatProducts
MntFruits
MntSweetProduct
MntWines
MntGoldProds

CAMPAIGNS INFO

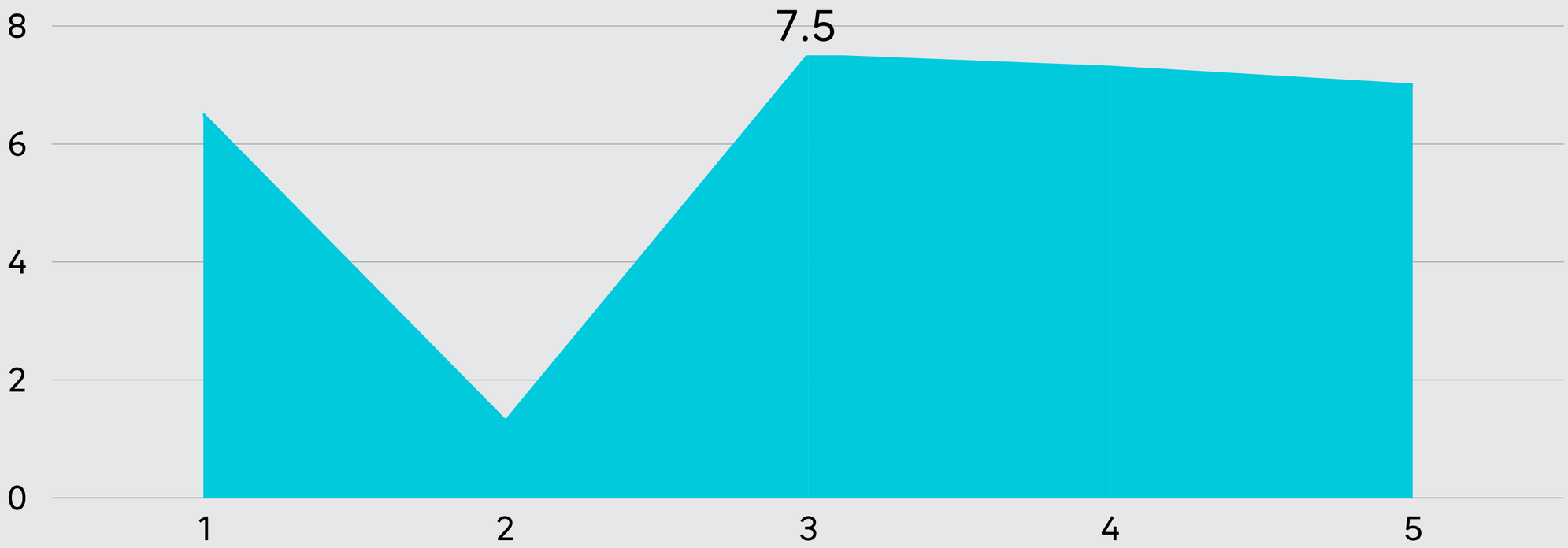
AcceptedCmp1
AcceptedCmp2
AcceptedCmp3
AcceptedCmp4
AcceptedCmp5
Response (target)
Z_CostContact

PURCHASES INFO

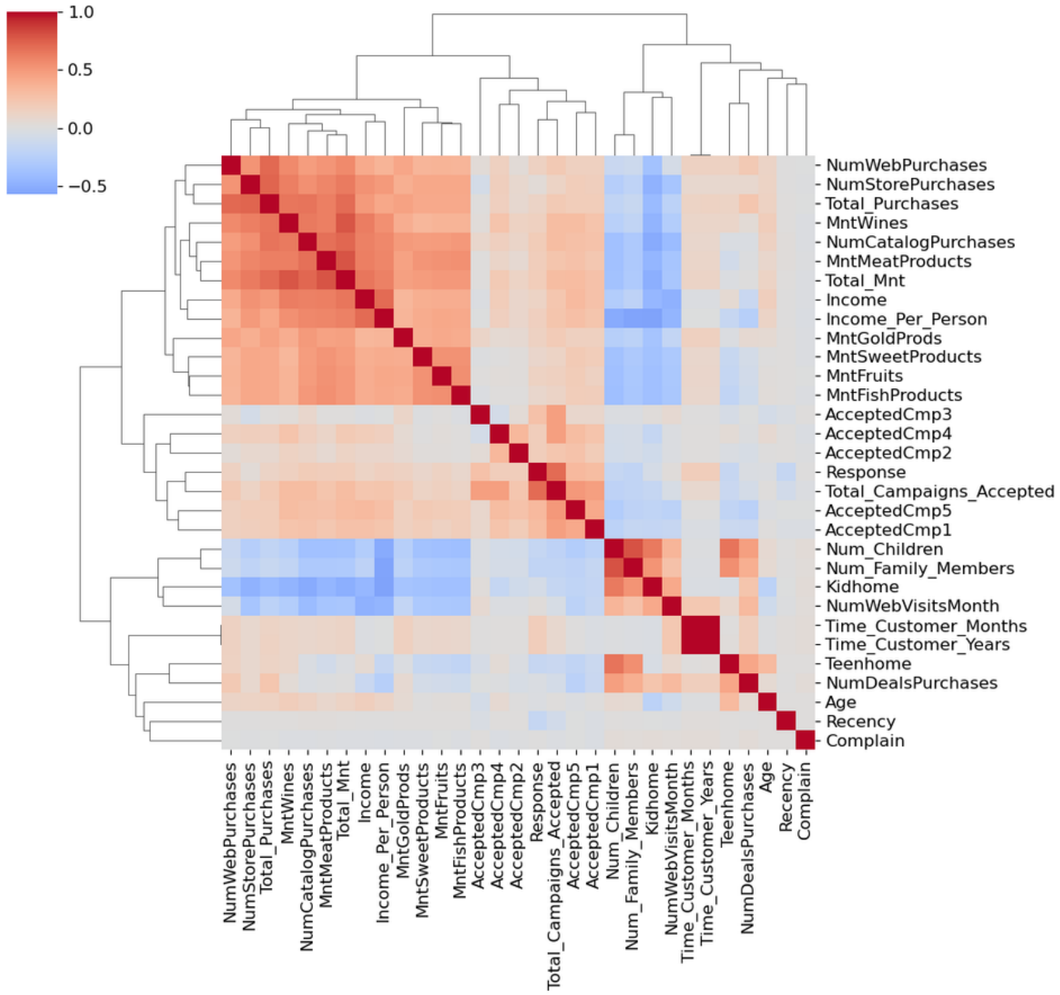
NumDealsPurchases
NumCatalogPurchas
NumStorePurchases
NumWebPurchases
NumWebVisitsMonth
Complain
Z_Revenue

EXPLORATORY DATA ANALYSIS

Acceptance of campaigns was **around 7%**, with the exception of Campaign 3, which had only 1.3% conversion.



EXPLORATORY DATA ANALYSIS



We can observe three clusters of correlated variables:

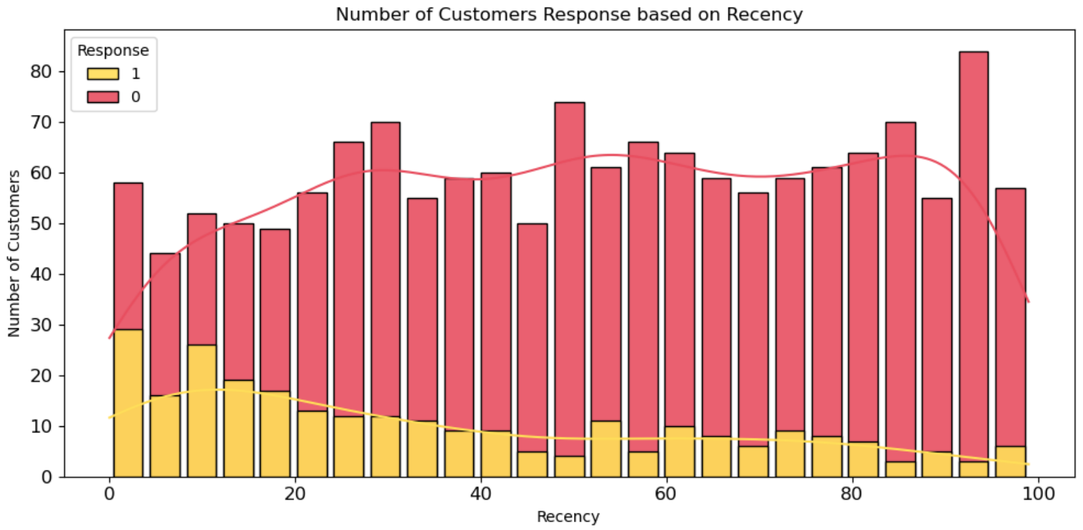
1. **HIGH INCOME**: Income, amount spent on products, and number of purchases show a moderate positive correlation.

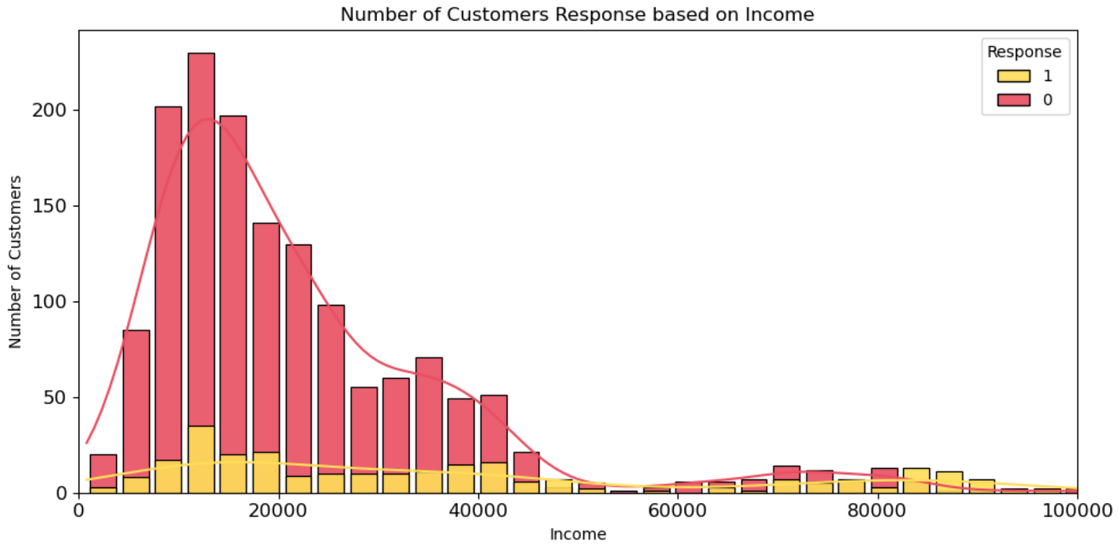
2. **CAMPAIGNS**: Acceptance of campaigns exhibits a moderate positive correlation.

3. **PARENTS**: There is a moderate negative correlation between spending (value and volume) on products and the number of children.

Anomalies: 'NumWebVisitsMonth' does not correlate with an increased number of web purchases 'NumWebPurchases', but is positively correlated with 'NumDealsPurchases', suggesting that deals are an effective way to stimulate purchases on the website.

EXPLORATORY DATA ANALYSIS

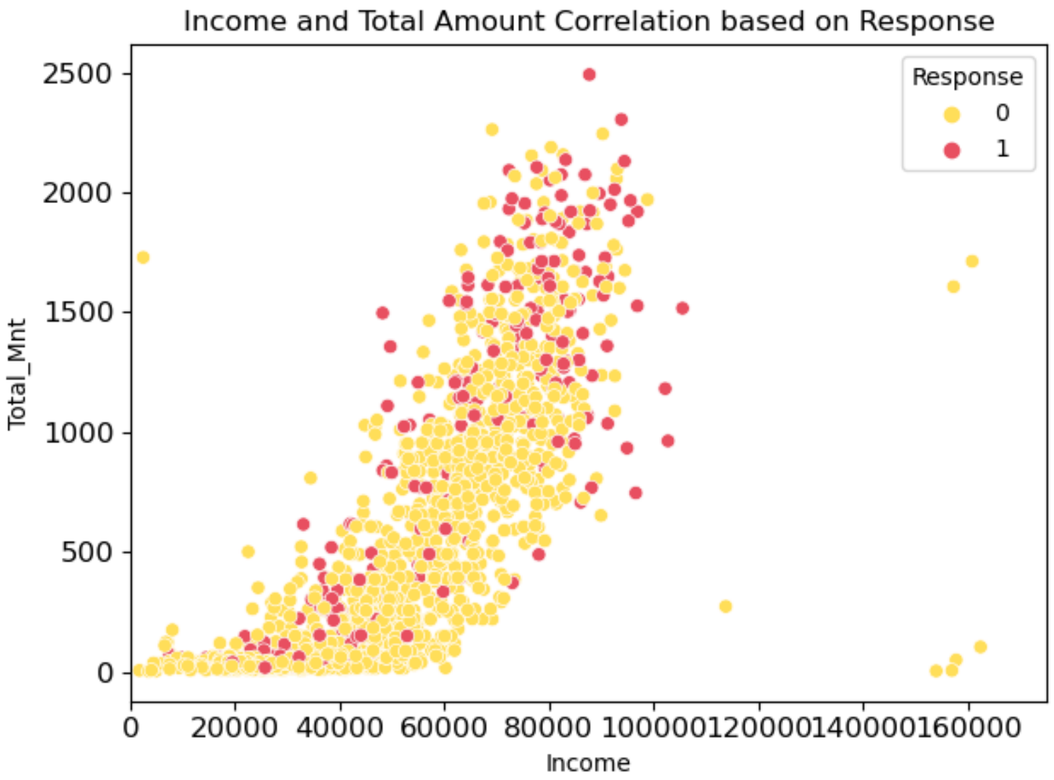




Based on recency (the time since the customer's last purchase), the **lower** the **recency**, the **greater** the likelihood that the customer will **respond positively** to the company's latest marketing campaign (Response).

Additionally, they demonstrate having higher income.

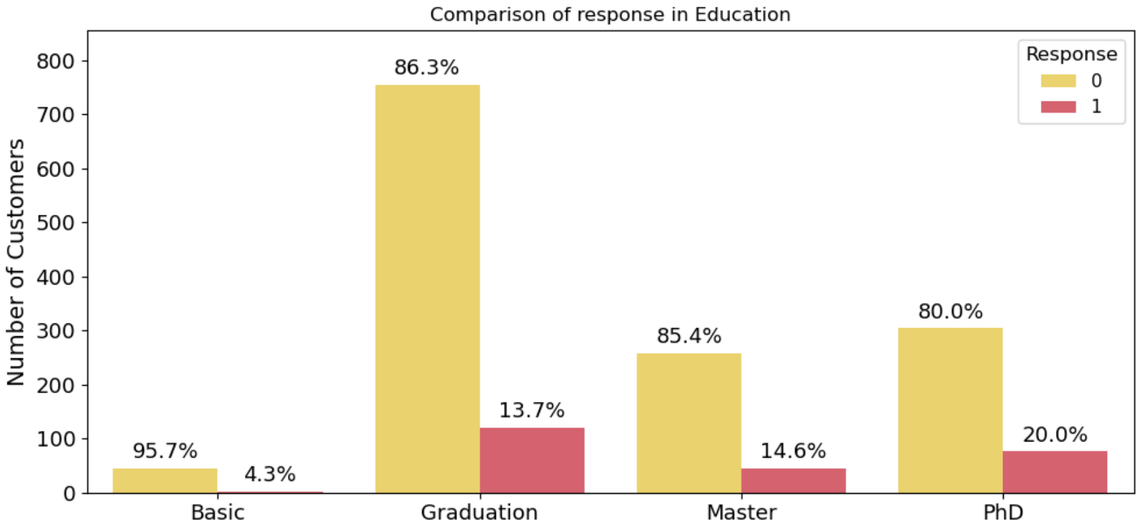
EXPLORATORY DATA ANALYSIS



Income and spending have a **positive correlation** with response, where higher income and spending values correspond to higher response rates.

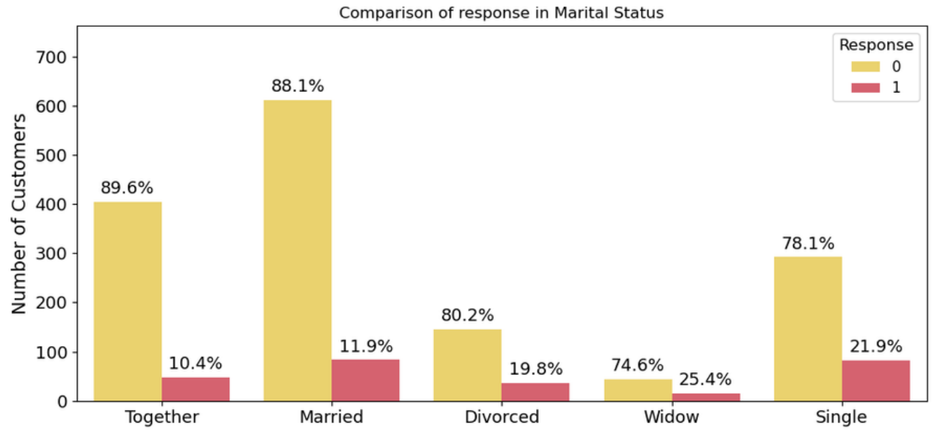
The increase in **income** also showed an increase in **campaign acceptance**.

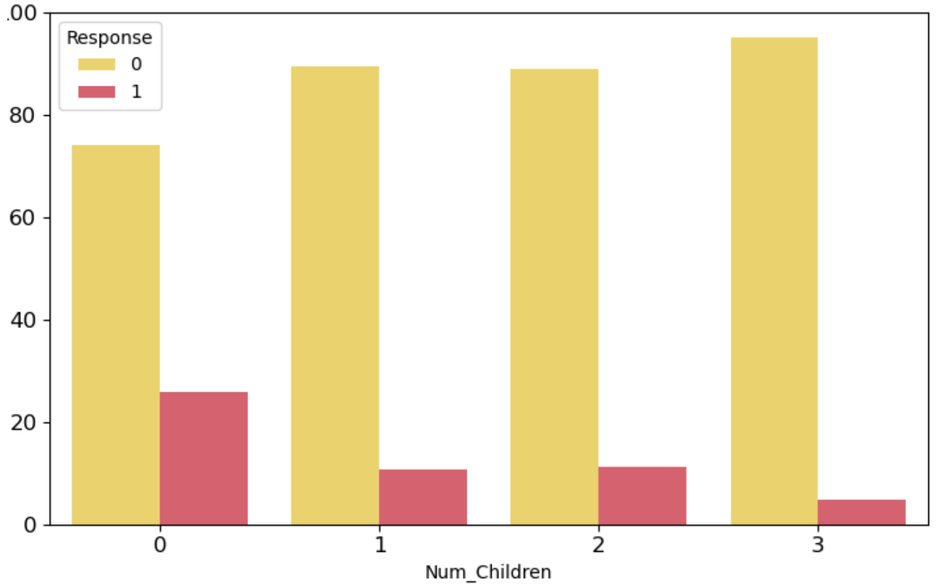
EXPLORATORY DATA ANALYSIS



Furthermore, the audience that converted has a **higher relative level of education**, which may also be related to higher income.

EXPLORATORY DATA ANALYSIS

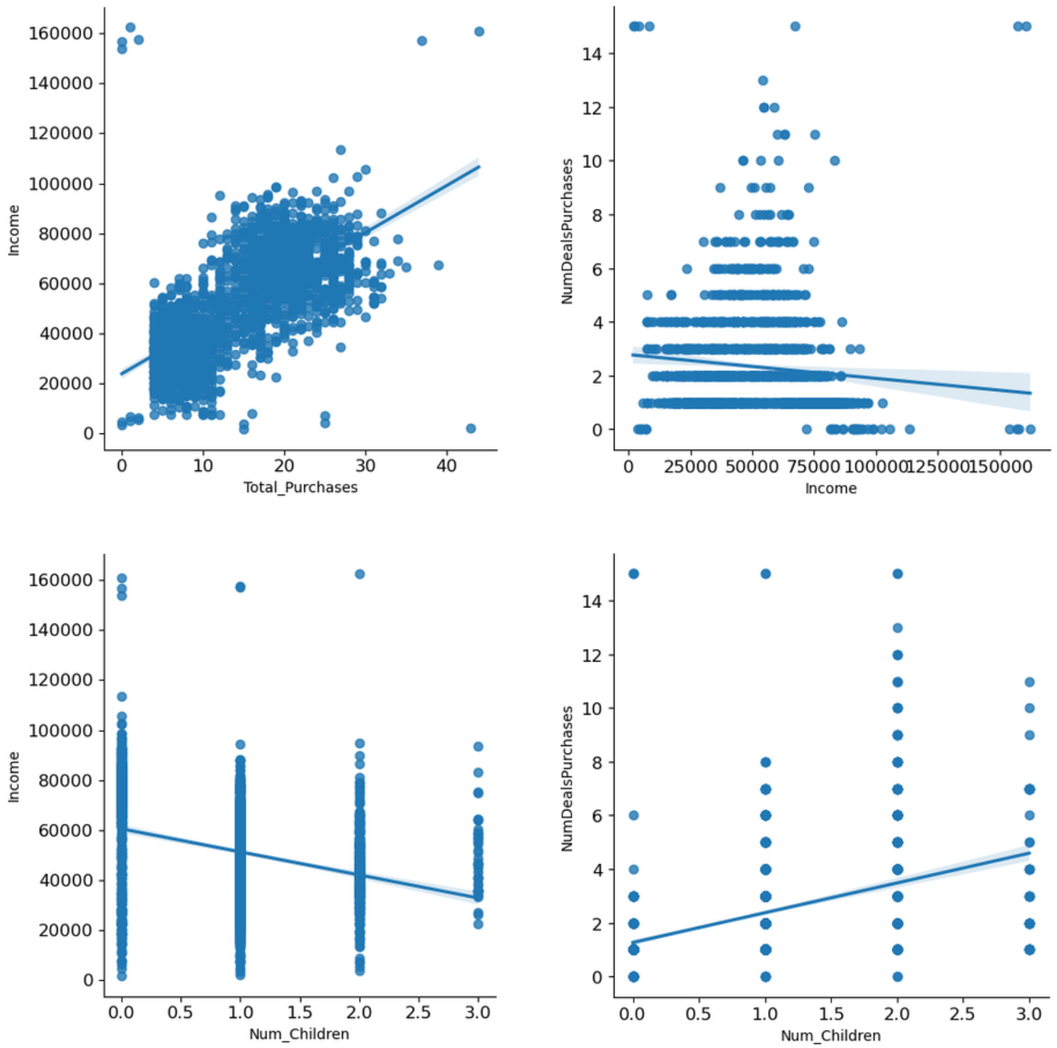




The audience that converted in the campaign has a smaller proportion of married and together individuals. Additionally, they have fewer children.

It can be said that it is a **less complex family structure**, with fewer members.

EXPLORATORY DATA ANALYSIS

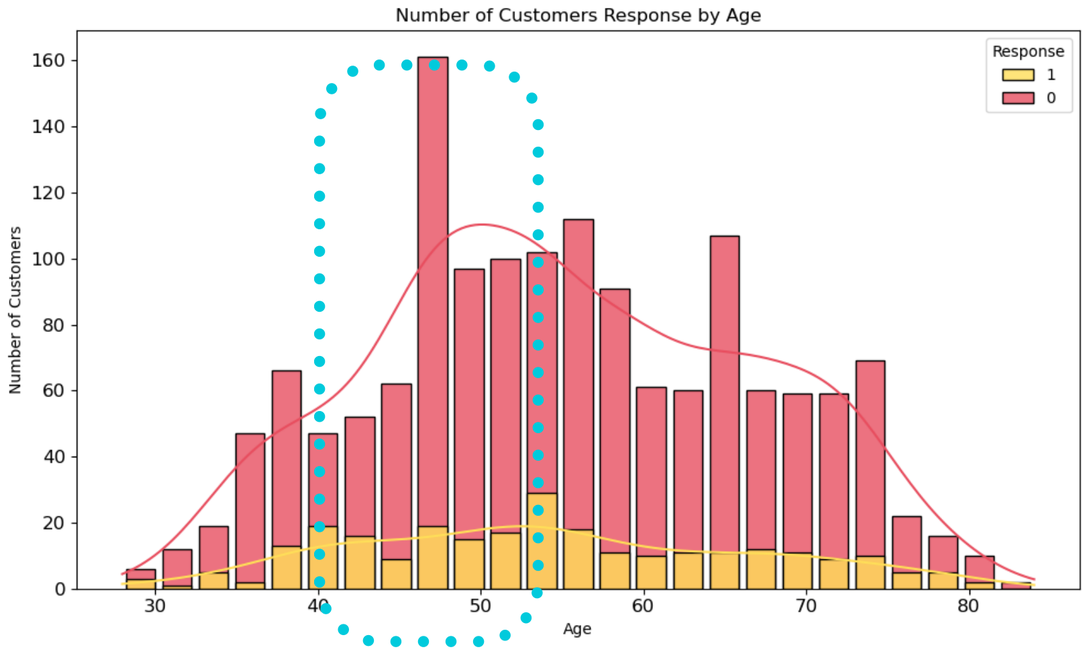


The correlation between income and purchasing behavior is evident:

Higher income corresponds to increased **purchase volume** and reduced discounted purchases.

Conversely, income tends to **decrease** with a **higher number of children**, leading to more purchases made with **discounts**.

EXPLORATORY DATA ANALYSIS



The converted audience is also dominated by an **older audience**, with **adults aged 40** and above.

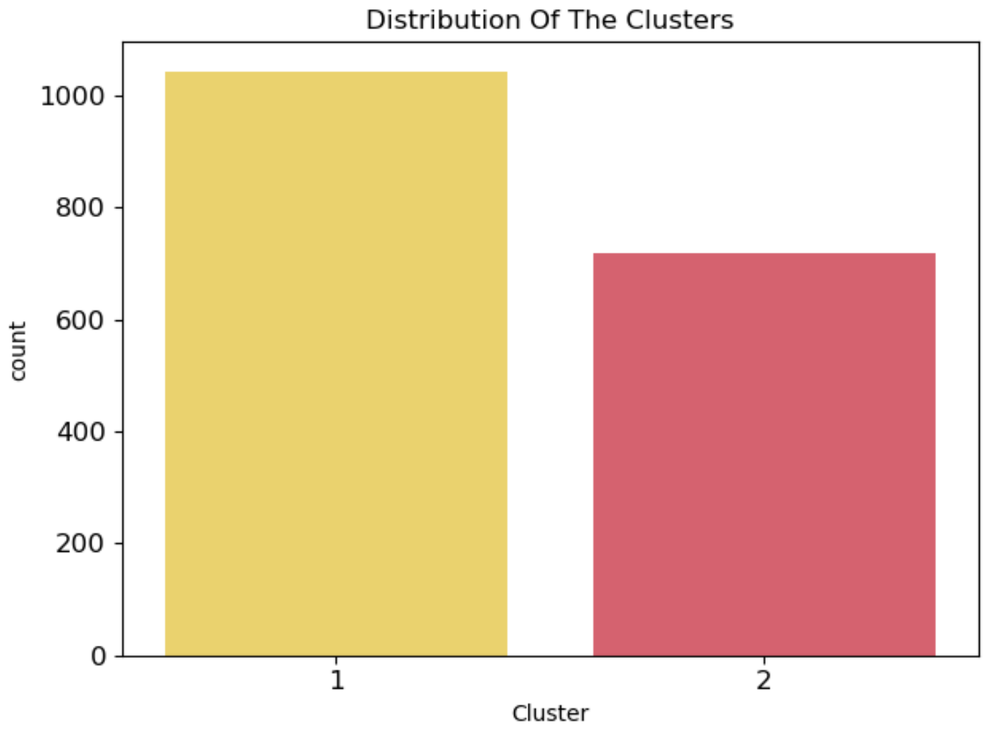
The main products purchased in terms of value are **meats** and **wines**, but it is not possible to determine if the price of the product influenced this. They spend more overall. Further analysis would be needed regarding the volume purchased per item.

CLUSTERING

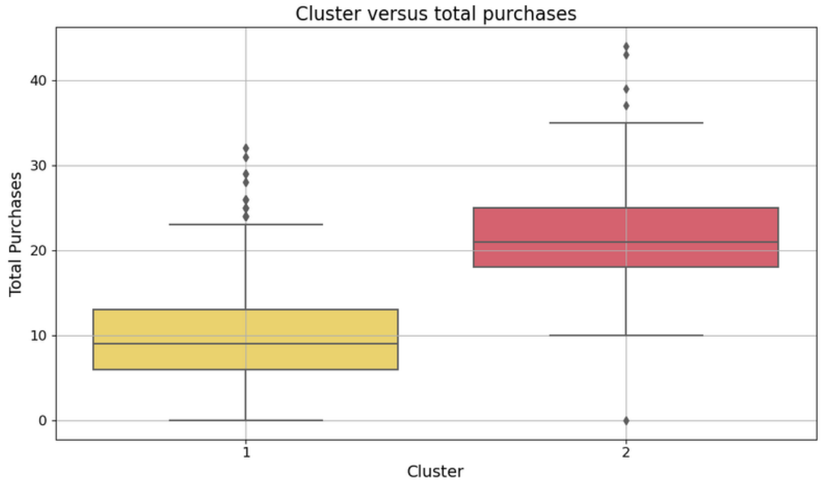
Two clusters were found through k-means.
They are:

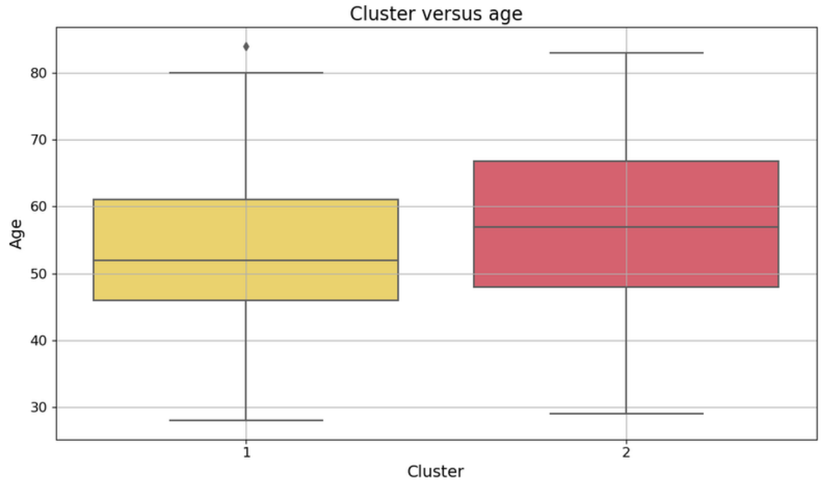
High income per person, higher expenditure on products, with larger volume of purchases and greater acceptance of campaigns, relatively older, smaller family structure.

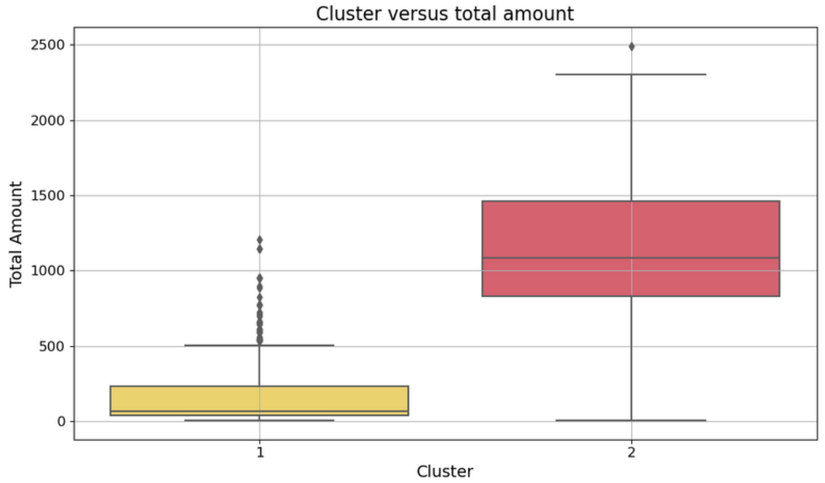
Low income, less expenditure on products and lower volume of purchases, higher number of children, slightly lower educational level.

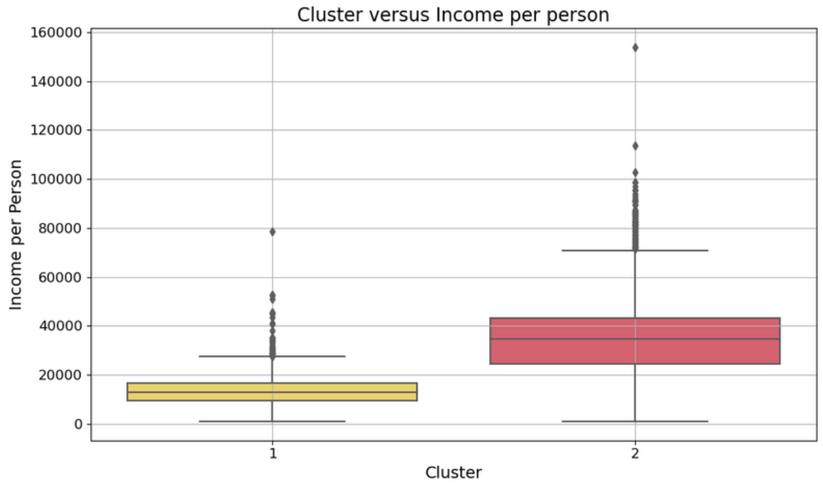


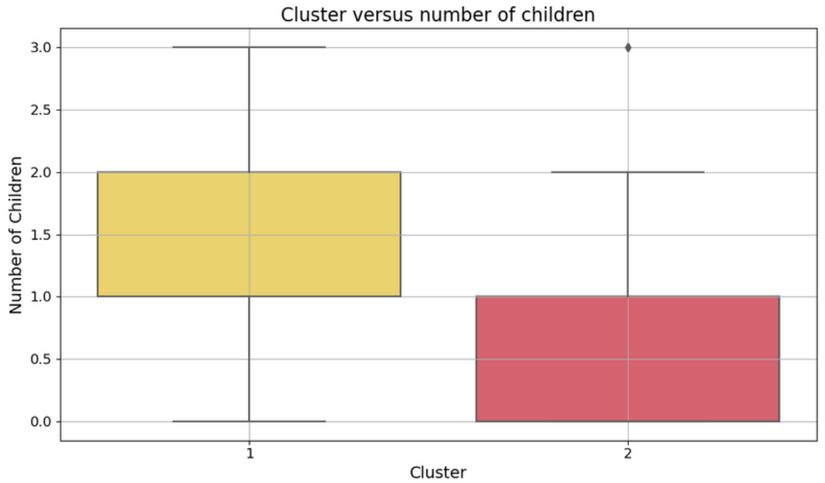
MAIN CHARACTERISTICS

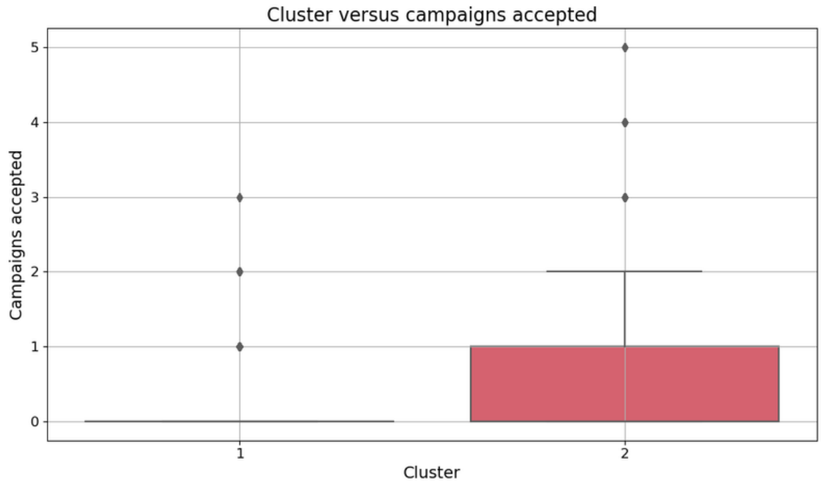












PREDICTING

MODEL PARAMETERS

PRECISION	Customers are predicted to respond, but they don't **(FALSE POSITIVE)**
RECALL	Customers are predicted to respond, but they don't **(FALSE POSITIVE)**
F1- SCORE	**(2 x Precision x Recall)/(Precision + Recall)**

PREDICTING

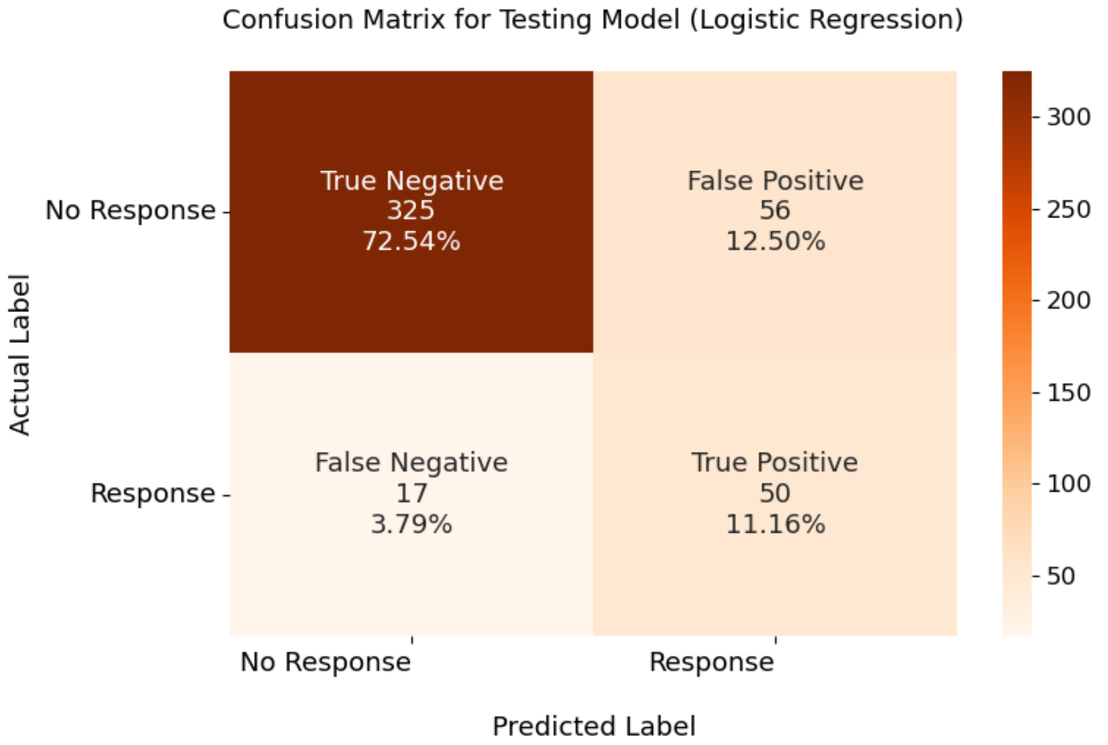
F1-Score: 58%

True positive (TP): 50
False positive (FP): 56

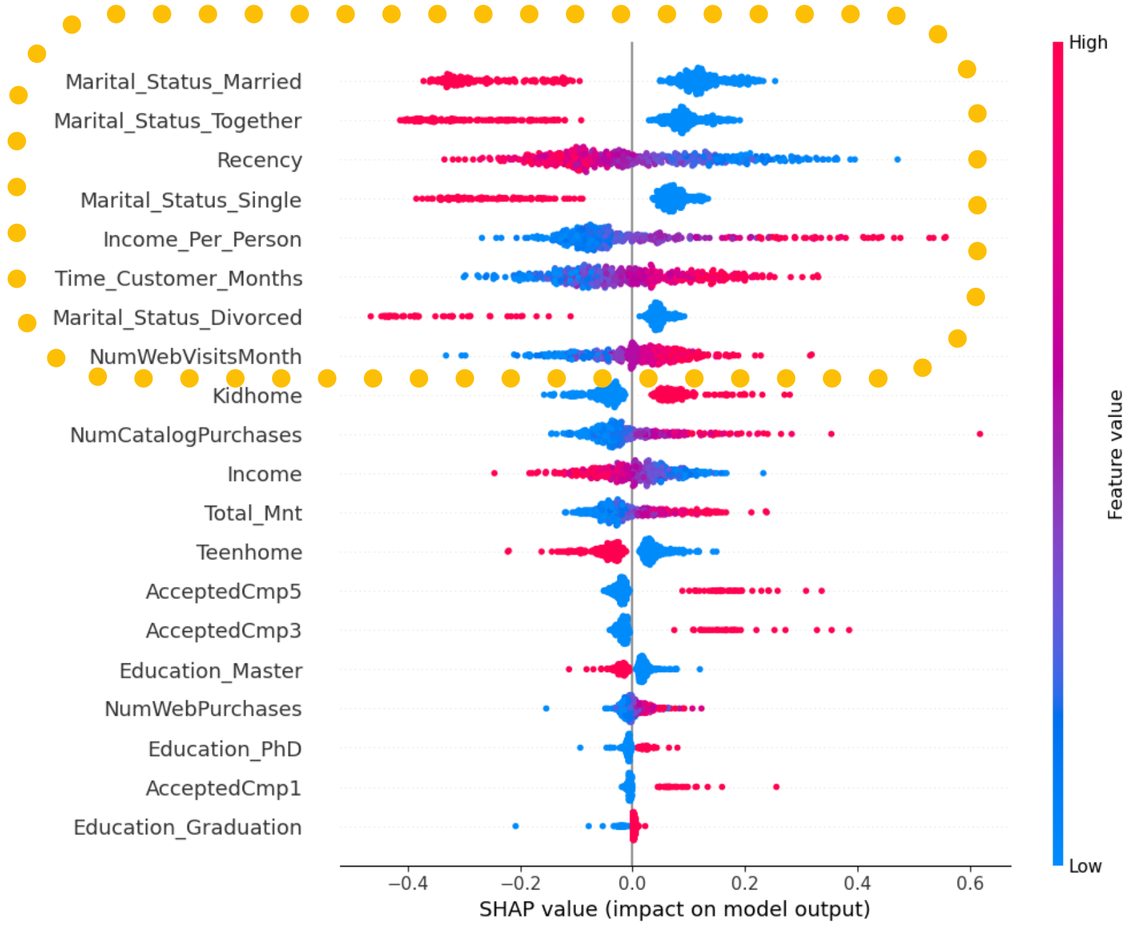
Response rate: TP/(TP+FP)

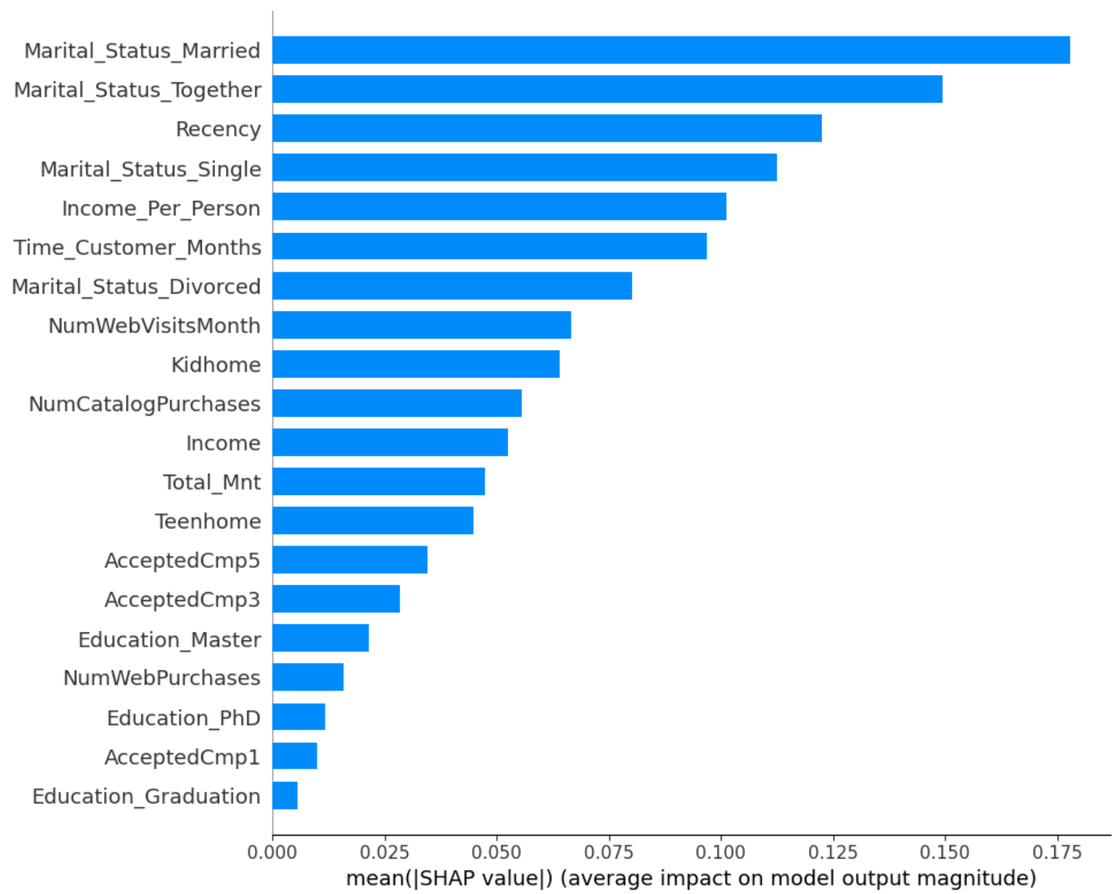
Before model: 14,91%
After model: 47,19%

+32,28 p.ps



PREDICTING





Response

No response

FINANCIAL RESULTS

NET PROFIT MARGIN (NPM)

Total response: 334	
Total campaign: 2240	
Total cost	-6720
Total revenue	3674
Total profit	-3046
NPM (%)	**-82,91**

+125,09 p.ps

Total response: 50	
Total campaign: 106	
Total cost	-318
Total revenue	550
Total profit	232
NPM (%)	**42,18**

THANK YOU!

Andressa Ribeiro
e-mail: andressaribeiro@usp.br