

# Pilot Campaign Analysis

July, 2020

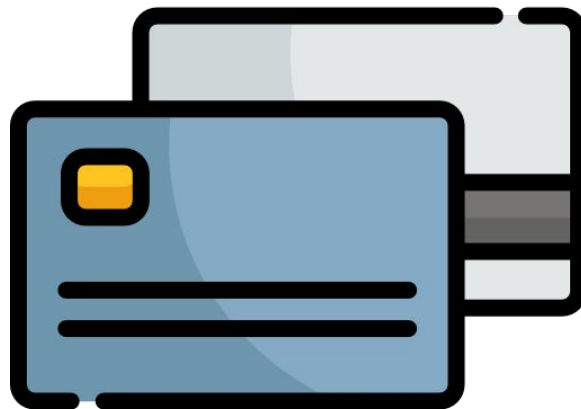




## The campaign

Improving performance of the campaign will help growing profit of the company

The new campaign, sixth, aims at selling  
the new gadget to the Customer  
Database





## Pilot results

There is a lot of room to improve the results of the pilot

**2.240** customers

**15%** success rate

**6.720 MU** of cost

**3.674 MU** of revenue

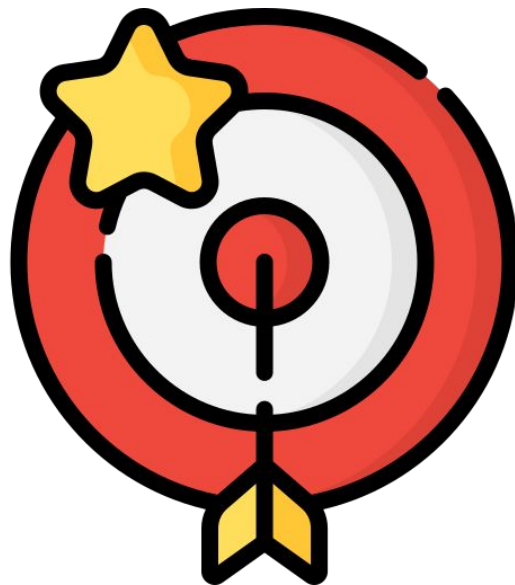




## Goal of the study

The pilot is just a study to maximize the profits for the next direct campaign

The idea of the study with the pilot campaign is to **build a predictive model that will produce the highest profit** for the direct campaign and **understand customers behaviour**



# Understanding Customer Behaviour



## Before starting the analysis

Convention used in the study

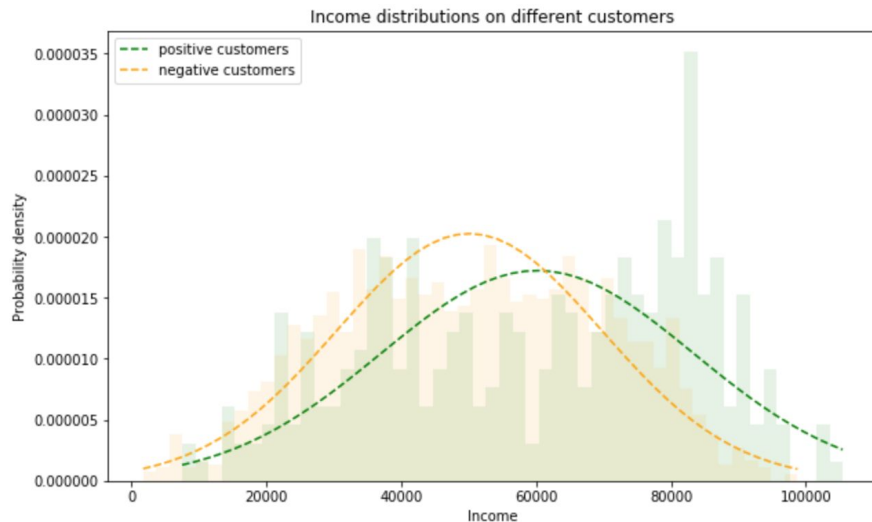
**Positive customers** = Customers that accepted the offer, that is, bought the new gadget

**Negative customers** = Customers that did not accept the offer



## Income distributions

Customers willing to buy the new gadget have higher income



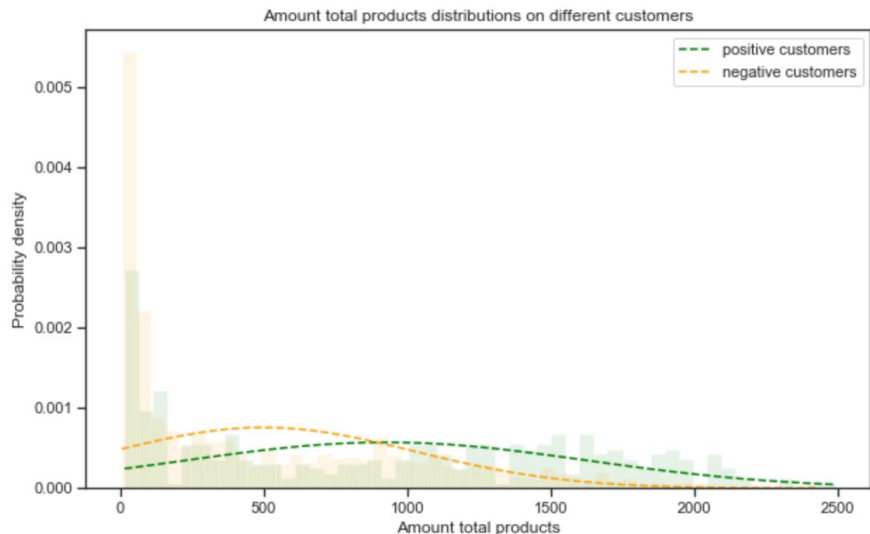
Distributions show that customers that accept the offer have higher incomes than those who do not.

There is a peak of positive customers with income above 70.000 MU yearly.



## Amount spent on products

Income has a moderate correlation to amount spent on products



Since income is a sensible data (not every customer might want to give this information), it was found a moderate correlation to the amount spent on products (fish, meat, fruits, sweet products and wine).

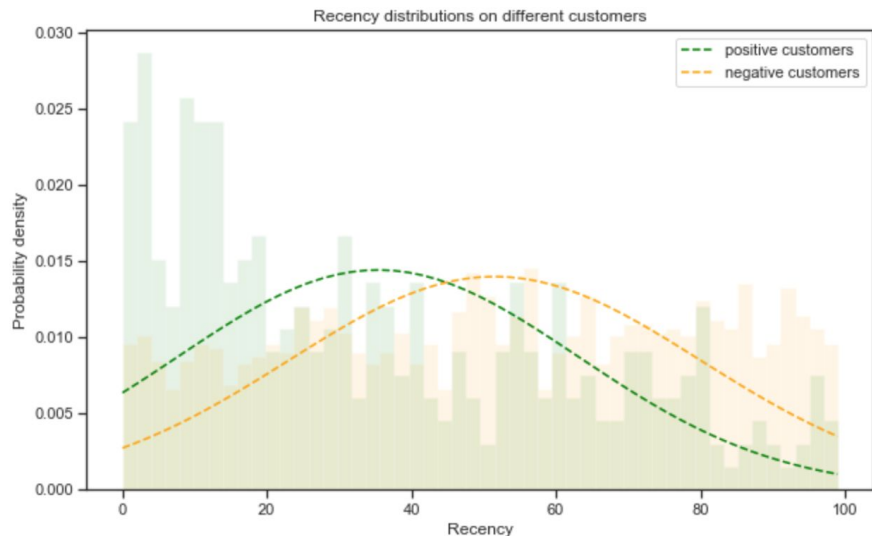
Customers who spend more on products are more willing to accept the offer.





## Recency

Customers who made a purchase recently are more willing to buy the new gadget



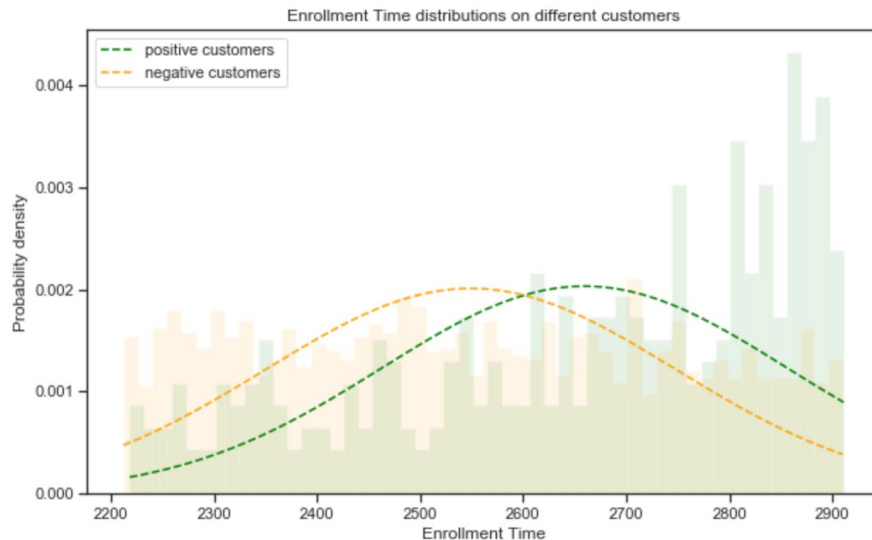
It is very clear in this graph that customers that made a purchase in the last days are more willing to accept the offer.

There is a peak of customers who bought the new gadget with a recency below 20 days.



## Enrollment time

The higher the enrollment time, more chances the customer will accept the offer



Another great variable for segmentation is  
the time the customer enrolled with the  
company.

Customers with higher enrollment time are  
more willing to accept the offer.

The histogram shows a peak of positive  
customers with enrollment time above 2700  
days.

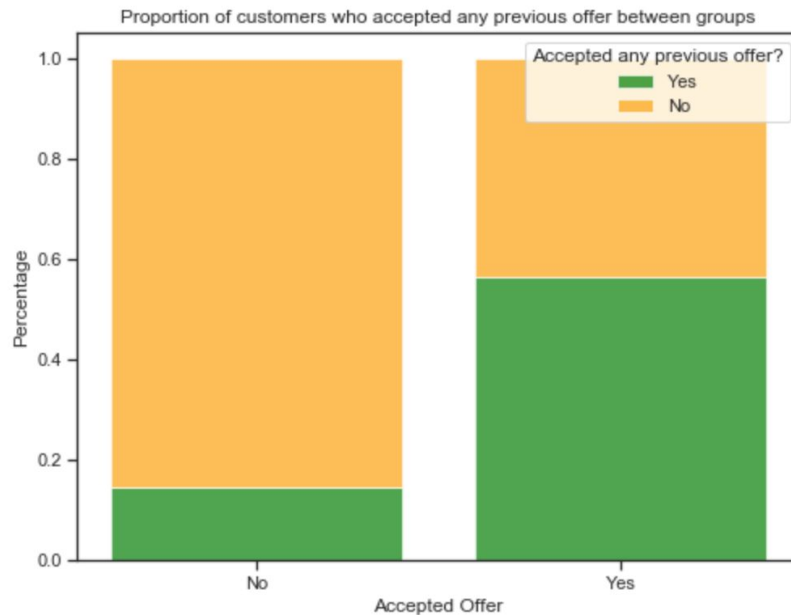


## Acceptance of previous offers

Accepting previous offers is a good sign that customer might accept the new one

Customers who accepted a previous offer are also more willing to buy the new gadget.

As we can see, 60% of positive customers, did accept a previous offer.



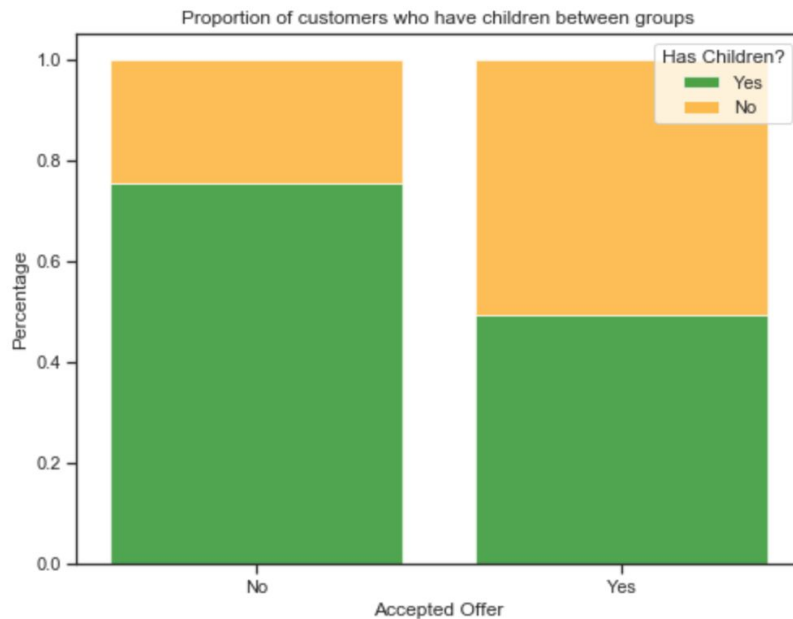


## Children at home

Having children does not have a good correlation with accepting the offer

Although the proportion of people who have children is higher for customers who did not accept the offer, still 50% of customers who did accept it have children.

Said that, having children is not a good indicator.

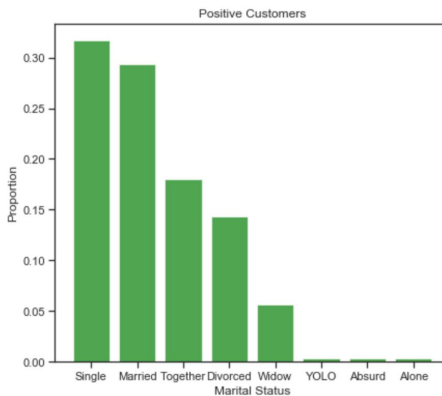




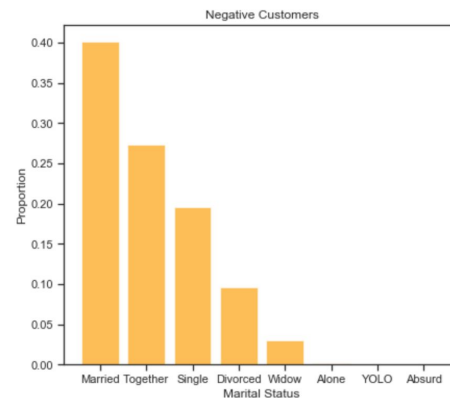
## Marital Status

It is not the best variable to segment customers

Even though the proportion of single customers is lower in customers who did not accept the offer, targeting these customers only may not be the right strategy, since it still represents 20% of this segment.



Marital status proportion



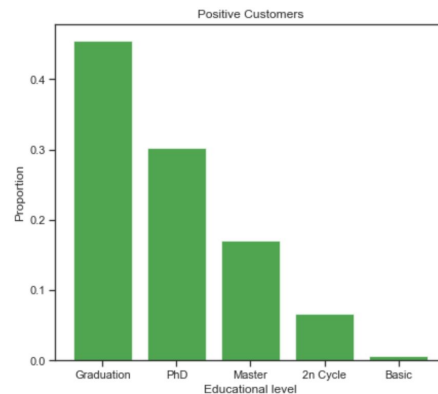


## Educational level

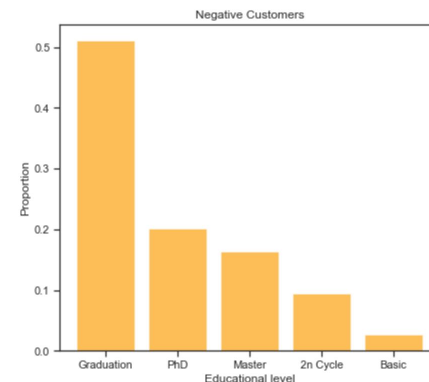
Distributions of educational level does not seem promising for segmentation

Higher educational levels were expected to be more related to positive customers, since it is common sense to correlate them to a higher income.

However, negative customers present a great proportion of highly educated people.



Educational level proportion

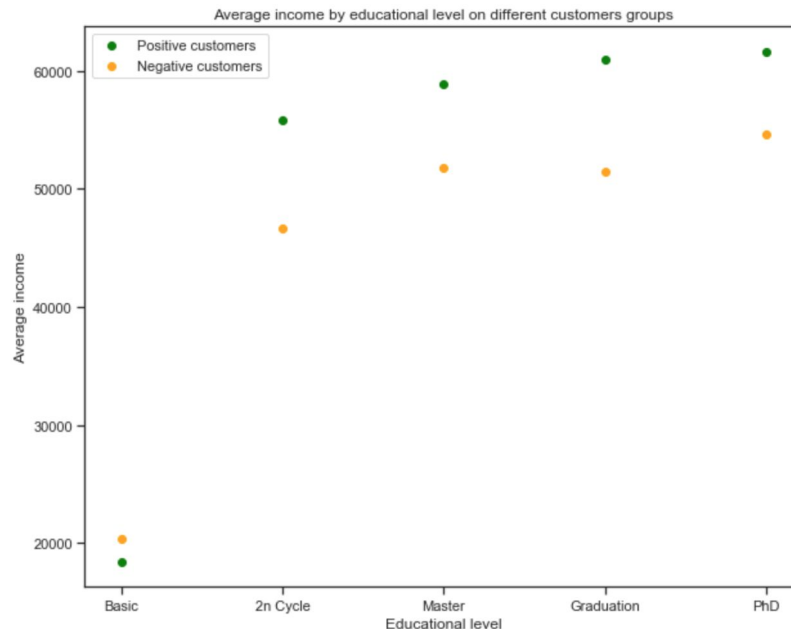




## Educational level

Income is still a better segmentation

When the average income for each educational level for positive and negative customers are compared though, it is clear that positive customers have a higher income, even with the same educational level.

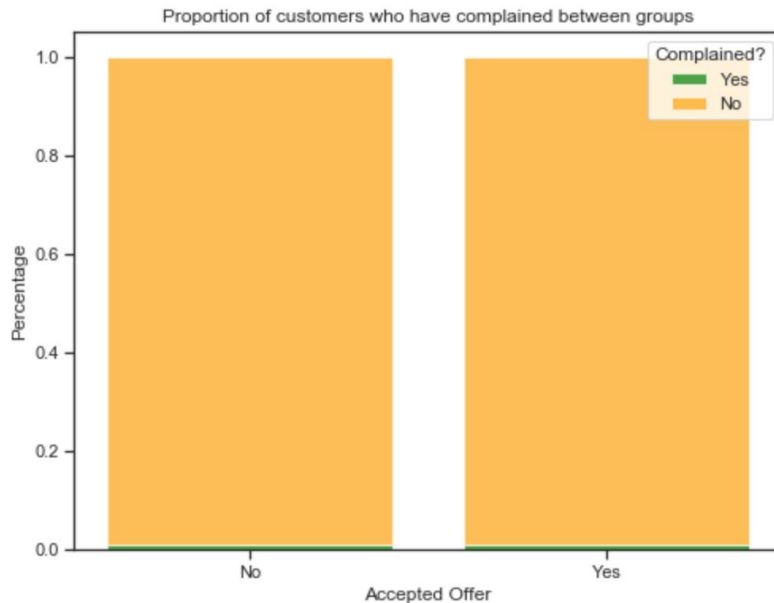




## Complaint

There is a very small rate of customers who made a complaint

Most customers did not make any complaint in the last 2 years, making this variable a bad choice for segmentation.





Predictive Model

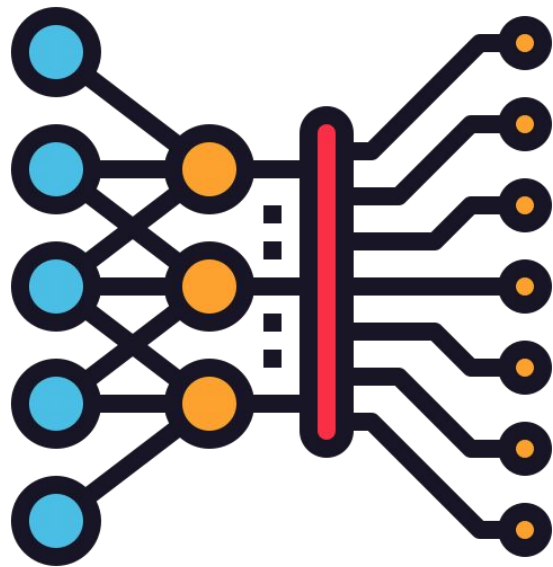


## Classification algorithms

The accuracy of each model was tested for better results

Three different algorithms were chosen to create a predictive model:

- Decision Tree
- Random Forest
- Logistic Regression



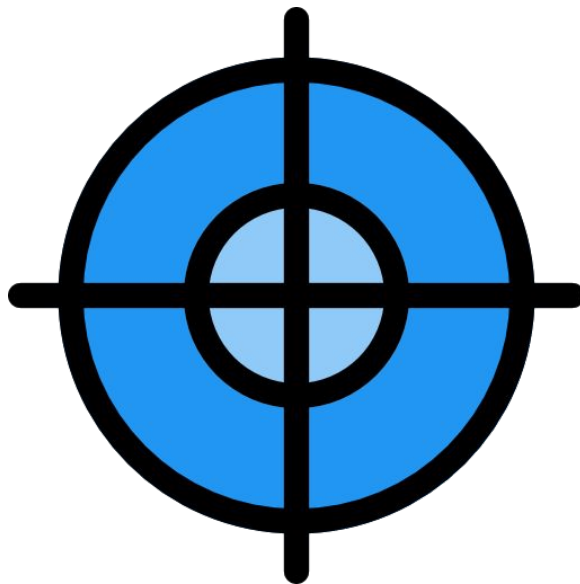


## Cross-validation

Given the conditions this technique was chosen to measure the quality of the models

Since pilot data is very small, separating the dataset into training and validation sets would not be the best option.

It was chosen a technique called cross-validation to evaluate the accuracy of the models



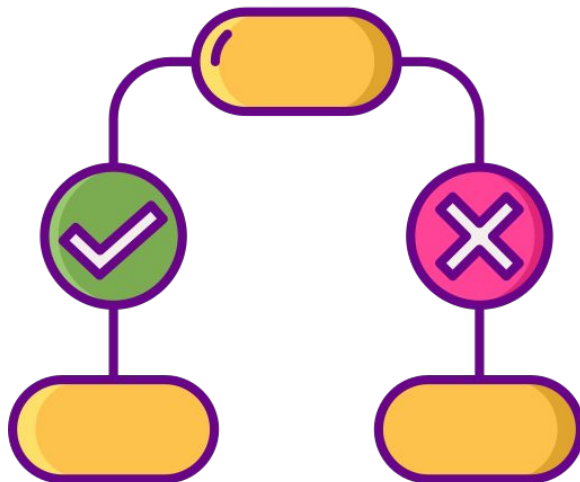


## Decision tree

Needs the adjustment of hyperparameter to have a better accuracy

After some adjustments to avoid overfitting, the accuracy of the algorithm was 86.8%.

That means that the algorithm can predict 86.8% right if a customer will accept the offer or not.





## Random forest

Better accuracy without adjusting hyperparameters

Random forest consists in evaluating several different decision trees and come up with a prediction.

The accuracy for this model was 87.7%.



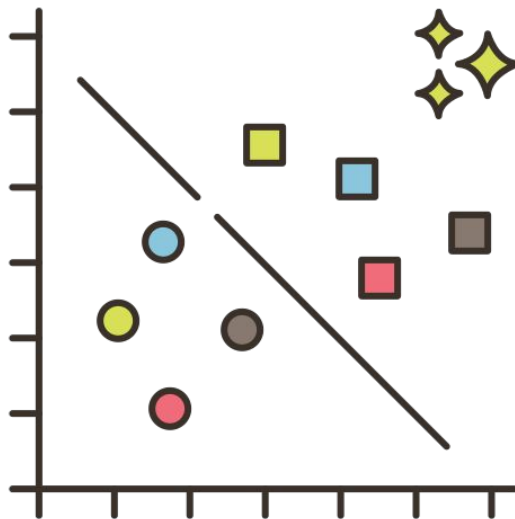


## Logistic Regression

Robust algorithm with higher accuracy

Logistic regression demands more data treatment, like one-hot encoding and standardization.

However it presented the highest accuracy, reaching a value of 88.5%.





## What does it mean?

Much better numbers for the campaign performance

Choosing logistic regression as the predictive model algorithm, these are the results we would get for a campaign the size of the pilot:

- 2.240 customers
- 88.5% success rate
- 6.720 MU of cost
- 21.802 MU of revenue





## Profits

There is a huge improvement in profits numbers

Profits for the pilot campaign would go from -3.046 MU to **+15.082 MU**.





# Recommendations



## Segmentation

Simple segmentation orientations

Behaviour of customers that are willing to accept the offer:

- Higher yearly income (peak above 70k)
- Lower recency (peak below 20 days)
- Higher enrollment time (peak above 2700 days)
- Accepted previous offer (60% of the customers who bought the new gadget).

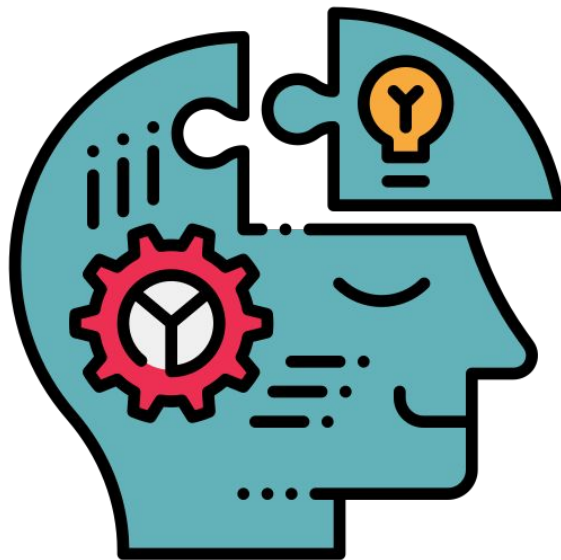




## Predictive Model

The use of a classification algorithm brings huge impact to the business

The choice of logistic regression will give the campaign a success rate around 88% and huge increase in profits.



# Thanks

Rafael Faria Castelhão  
[rafaelfariac@gmail.com](mailto:rafaelfariac@gmail.com)