



Data Analyst Test

Business Report

Rodrigo Rezende dos Santos
<https://www.linkedin.com/in/rodrigorezendes/>

Objective

Understand the data, find business opportunities & insights and propose any data driven action to optimize the campaigns results & generate value to the company.

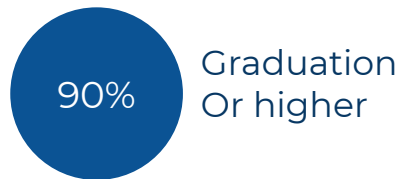
Agenda

- Customer
- Orders
- Income x Revenue
- Campaings
- RFM Analysis
- Classification Model
- Actions

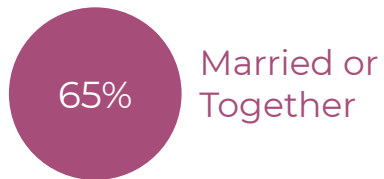
Customers



Education



Marital Status



Kids Home



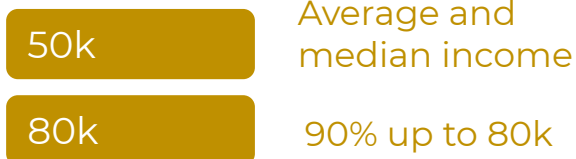
Teens Home



Age



Income



Lifetime



Orders

Revenue

1.2 Million

54% Wines

29% Meats

17% Others

Purchases

28 k

46% Stores

32% Web

22% Catalog

Avg price

\$45

Average price of 75% of purchases

Deals

18%

Purchases made with discount

Gold Products

8%

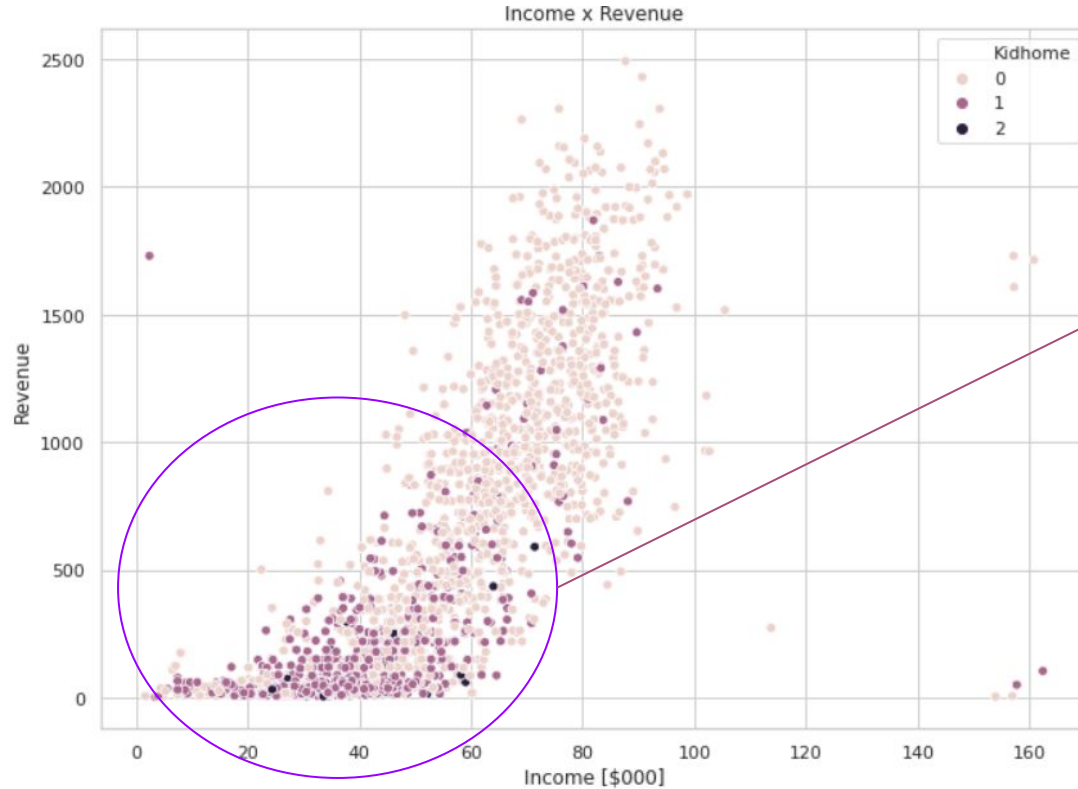
Purchases of gold products

Complains

1%

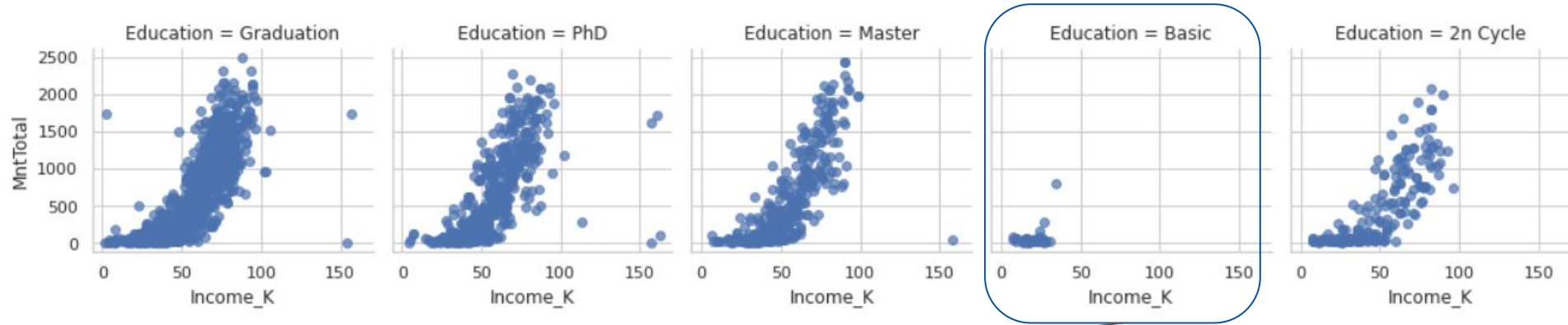
Complained last 2 years

Income X Revenue (Kids)



Customers who has at least 1 kid at home earn less and spend less.

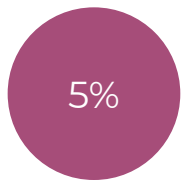
Income X Revenue (Education)



Customers who has basic education earn less and spend less.

Campaigns

Campaigns

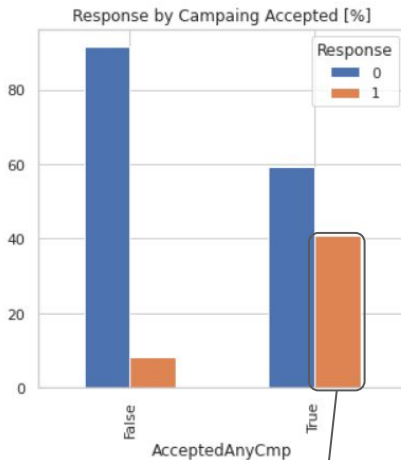


Average success rate of the five campaigns

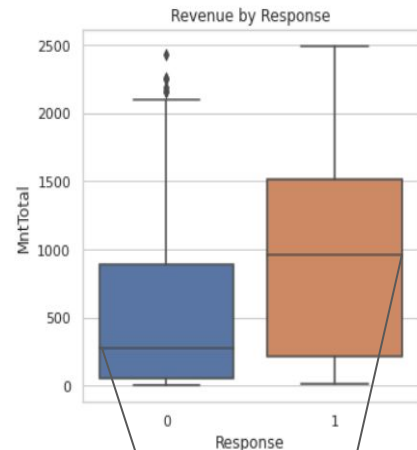
Last Campaign



Success rate of the last campaign (6th)

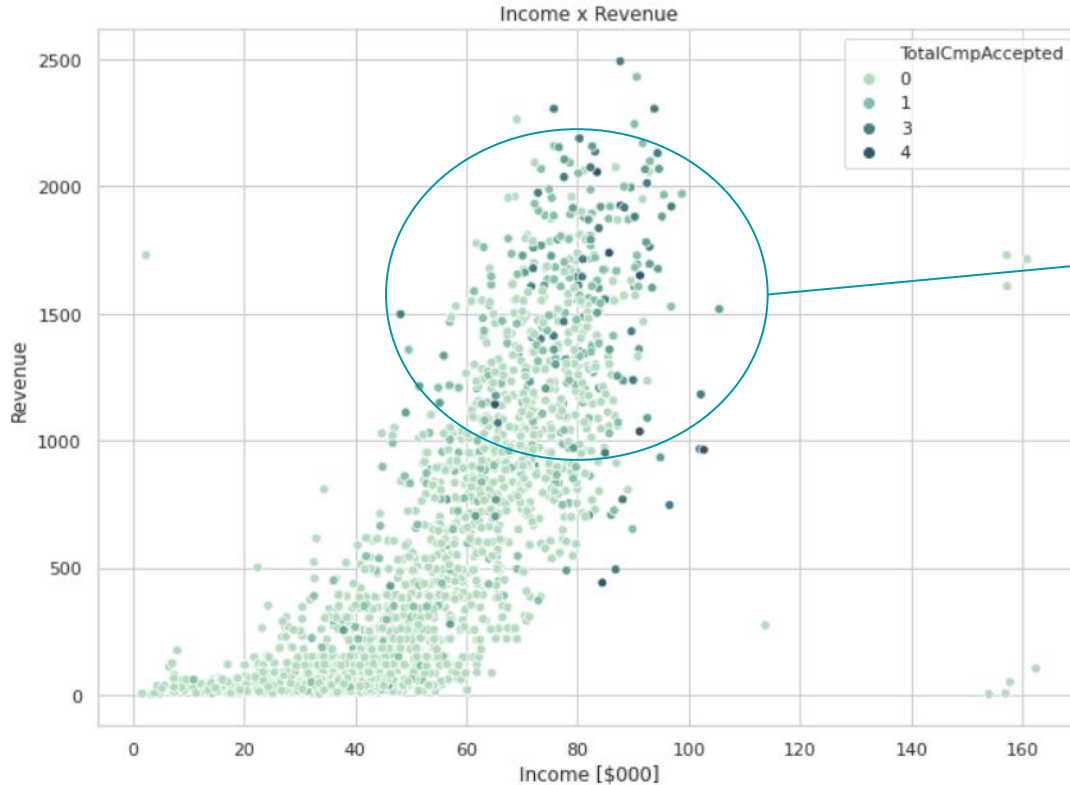


40% of those who accepted at least 1 campaign answered true in last campaign.



Median of \$300 for who NOT accepted last campaign and \$1000 for who accepted.

Income X Revenue (Campaigns)

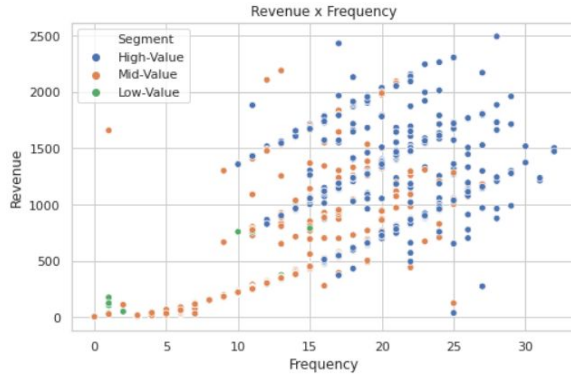


Customers who accepted more than one campaign earn more and spend more

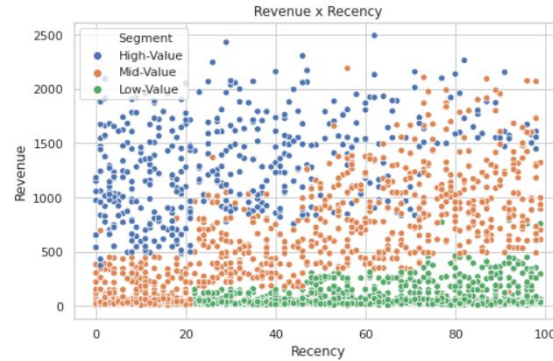
RFM Analysis



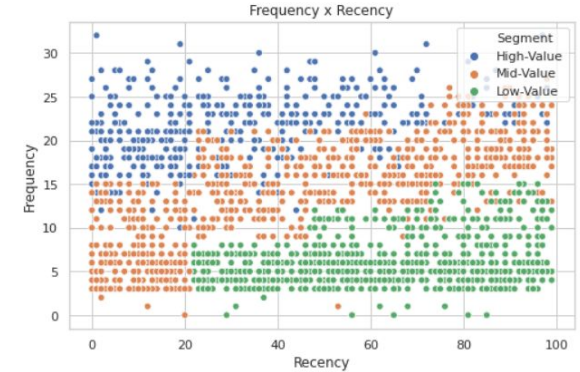
Where are the best customers?



High frequency and
high revenue



Low recency and
high revenue

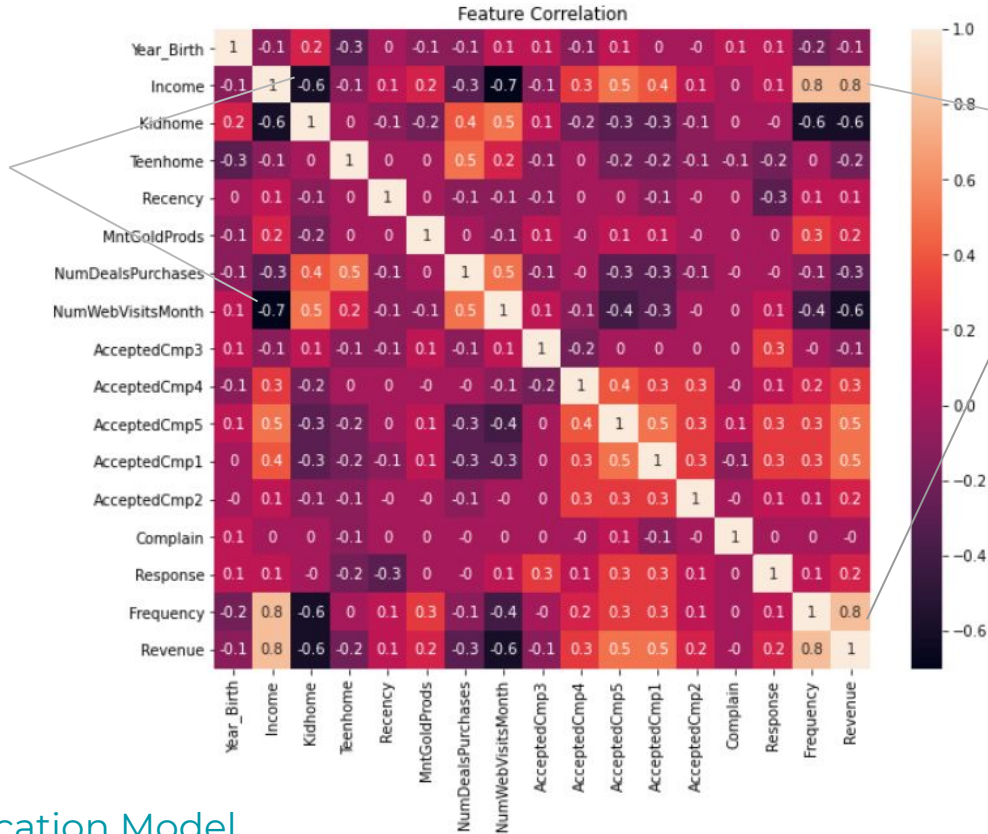


Low recency and
high frequency

Classification Model

Correlation

Strong negative correlation



Strong positive correlation

Classification Model

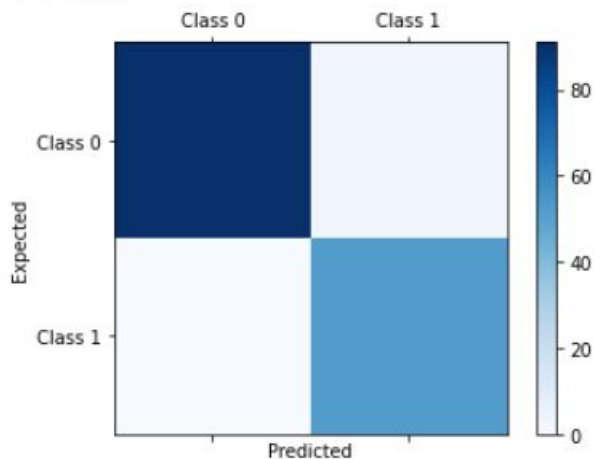
Confusion Matrix

Executed tests with five models and selected the Random Forest.



Confusion matrix:

```
[[91  3]
 [ 0 52]]
```



Random forest

	precision	recall	f1-score	support
0	1.00	0.97	0.98	94
1	0.95	1.00	0.97	52
accuracy	0.98			146
macro avg	0.97	0.98	0.98	146
weighted avg	0.98	0.98	0.98	146

```
[[91  3]
 [ 0 52]]
```

Accuracy of 98%

Classification Model

Feature Importance and probability

	feature	importance
0	Teenhome	0.119247
1	Recency	0.233745
2	Frequency	0.132581
3	Revenue	0.245307
4	NumWebVisitsMonth	0.130309
5	AcceptedCmp3	0.138810

Importance of each
selected feature in
model

Probability

A campaign for customers without teens at home, that did last purchase in 30 days, frequency of 10, revenue of 1000, who did 10 webvisits in last two months and that accepted the campaign 3

**has 90% chance of
succeeding!**

Actions

- Define the customer profile to make other campaigns
- Deploy the model to predict the 6th campaign
 - Adjust model to get 100% of Accuracy

Source Links

- [iFood Exploratory Analysis](#)
- [iFood RFM Analysis](#)
- [iFood Classification Model](#)

Support Links

- [Segmentation by RFM clustering](#)
- [Decision Support System using ML Algorithms to Predict the Success of Marketing Campaign](#)
- [Predicting Kickstarter Campaign Success with Gradient Boosted Decision Trees: A Machine Learning Classification Problem](#)
- [Simple guide to confusion matrix terminology](#)
- [8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset](#)
- [Resampling strategies for imbalanced datasets](#)



Questions?

Rodrigo Rezende dos Santos
<https://www.linkedin.com/in/rodrigorezendes/>