Optimizing Marketing Campaigns with ML

AI Group Project

Team 6 – Section A

Annik Westermann, Carla Sureda, Vaisu Yabu Lee, Hriday Chhabria, Matteo Giardini

Jan 5th, 2022

Executive Summary

Executive Summary

Problem, Solution and Impact in a nutshell

Problem

Carrefour's current marketing campaigns have a **low response rate (~15%)**

- The campaigns are **not customized to specific customers segments**
- The ineffectiveness of the campaigns causes a **low ROI on marketing initiatives**
- **~85% of customers have been unresponsive** to the campaigns

Solution

Machine learning models shall provide a two-fold solution to Carrefour's problem.
The deployment of two ML models is recommended:

- **Supervised Learning:** to identify non-responding customers and root causes for non-response
- **Unsupervised Learning:** to tailor marketing campaigns to different non–responding customers

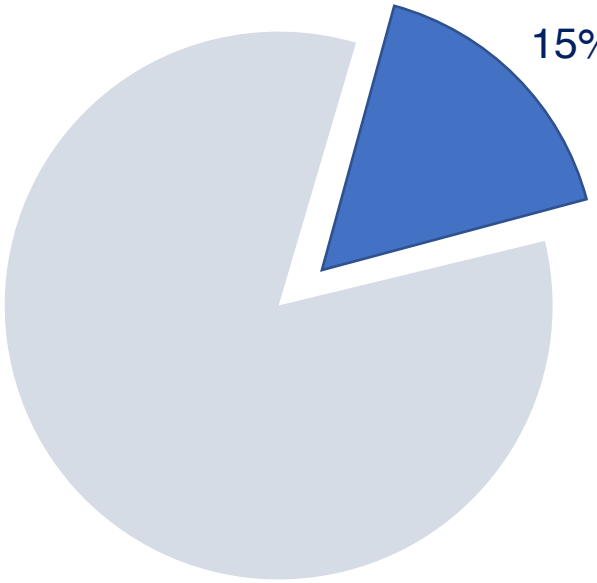
Impact

The proposed solutions will bring about substantial benefits:

- **81% increase in profit** generated by the base and the premium marketing campaign
- **Customized recommendations for premium marketing campaign** based on different customer segments.

Low response rates lead to low ROI for marketing initiatives

Problem



15% response rate on past 5 marketing campaigns

85% non-responding customers represent an opportunity to increase response rate

Root causes

i. Marketing campaigns do not target specific customer segments

ii. Marketing budget is spent ineffectively over entire customer base

resulting in

Ineffective marketing campaigns and low return on investment

Two ML models to be deployed to boost marketing efforts

Solution

	Supervised Learning	Unsupervised Learning
Type of model	*Classification*	*Clustering*
Goal	● **Identify non-responding customers** and pinpoint **reasons leading to non-response**	● **Segment non-responding customers into groups** to finetune scope of premium marketing campaigns
Output	● **Base Marketing Campaign** targeting predicted responding customers ● **Premium Marketing Campaign** targeting predicted non-responding customers	● **Cluster 0**: Medium-sized households, low income, preference for special deals. ● **Cluster 1**: Small households, high income, a high n° of purchases, low deal purchases.

Resulting in +81% in profits and refined customer segments

Impact

€€€ Financial Impact

+81%
in profits generated by marketing

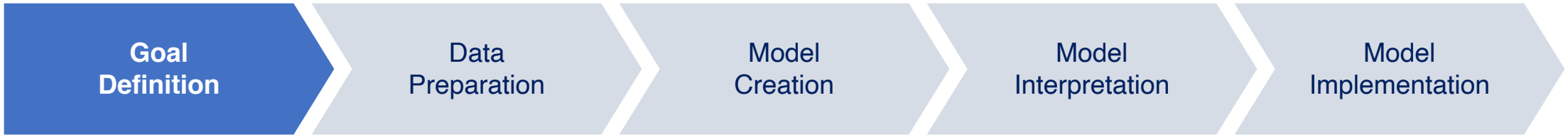
+€240k
Base Campaign

+€2M
Premium Campaign

Strategic Impact

- Customized marketing initiatives
- Clearer customer profiling & segmentation
- Higher customer engagement & satisfaction
- New product mixes and special promotions

Machine Learning Cycle



1 Goal Definition

Overall context & business problem definition

We are part of the Marketing Analytics Team from Carrefour

Company

Carrefour is:

- A French retail multinational founded in 2000.
- Main player in the food and beverage industry, covering nearly 20% of the food-producing Madrid area.

Problem Statement

- Carrefour's current **marketing campaigns lack effectiveness and customization**.
- The company targets all customers with one base campaign, however, only **~15% of customers are responding.**
- There is no alternative marketing campaign for the **~85%** of customers not responding to the base campaign.

Our role

- We are part of the Marketing Analytics Team, responsible for improving the company's marketing performance by **increasing profits and ROI.**
- By analysing customer data, we will enable Carrefour to t**arget responding customers more effectively** and to **convert non-responding customers.**

Overall context & business problem definition

Better targeting responding customers & converting current non-responding customers

Goal

i. Increase **bottom line** (i.e., profits generated by marketing) and **ROI of marketing campaigns**

ii. Produce **new, tailored marketing campaigns** to convert non-responding customers (85%)

ML Models implemented

Supervised Learning

- Analyzing and understanding variables leading to non-responding customers.
- Introducing targeted marketing campaigns to increase ROI.
- Introducing a premium campaign which targets customers that are not likely to respond to the base campaign to increase profits.

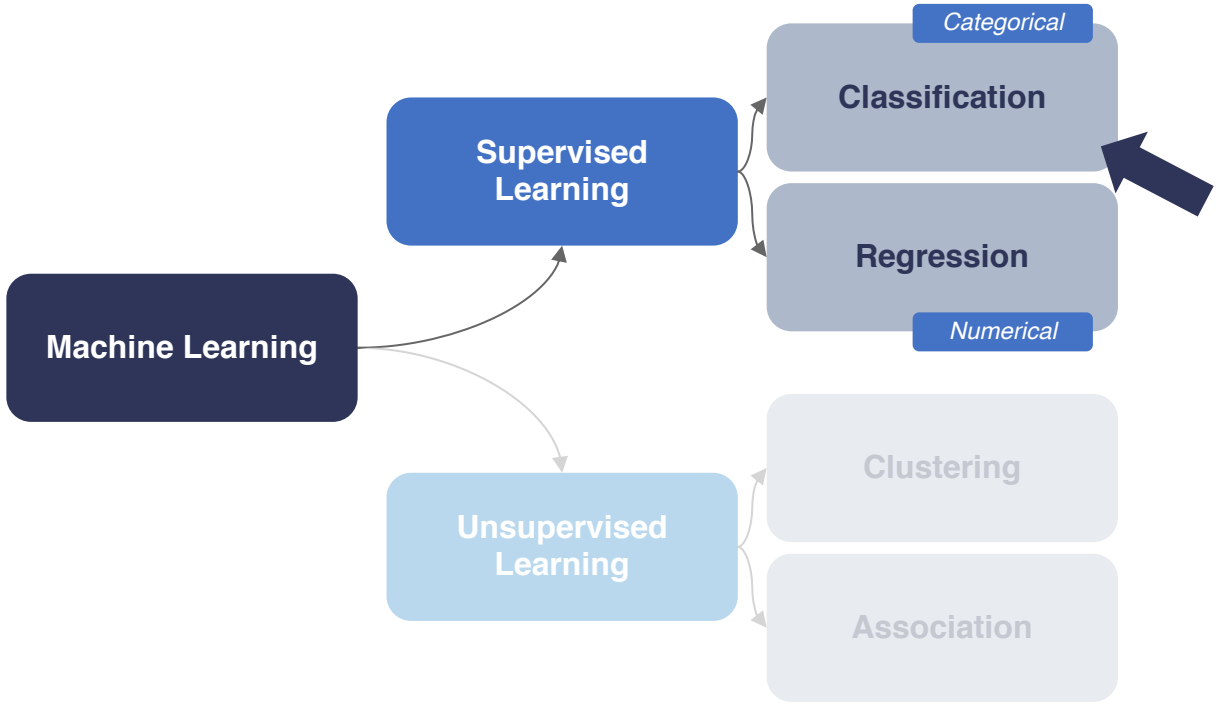
Unsupervised Learning

- Divide non-responding customers into segments to better understand the different personas & customer groups.
- Use insights to provide recommendations for designing the premium campaign which will target customers less likely to respond to the base campaign.

Relevant machine learning models to solve our problem (I/II)

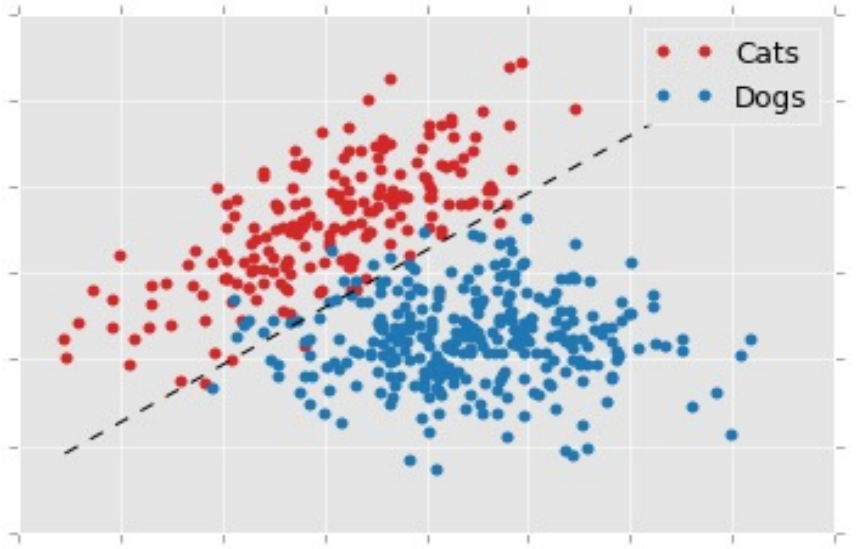
Supervised Learning to identify non-responding customers and their characteristics

Why Supervised Learning?



Why Classification?

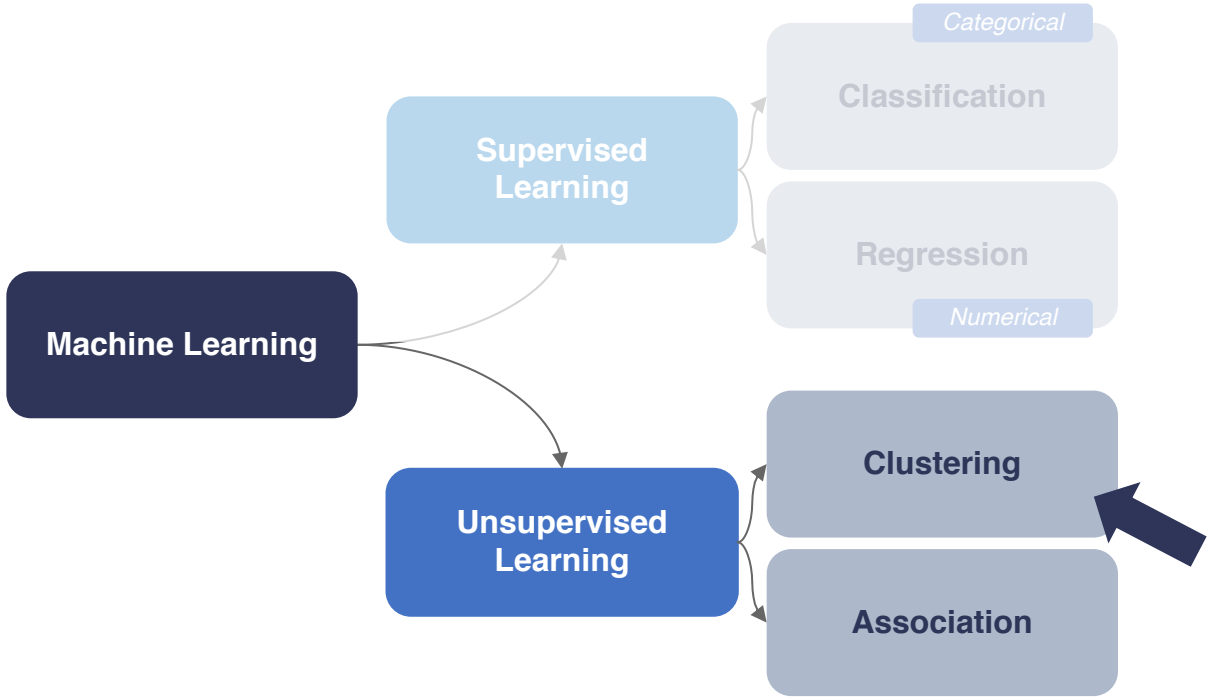
As the **dependent variable (*Response*) is categorical**, this problem is solvable through **classification** (as shown below)



Relevant machine learning models to solve our problem (II/II)

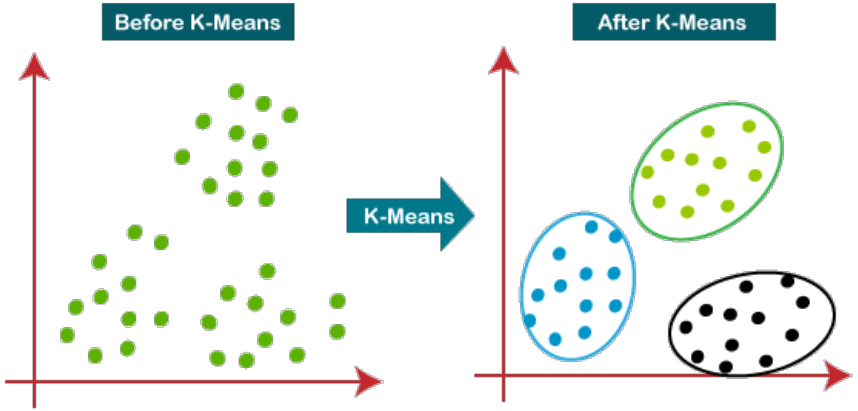
Unsupervised Learning to divide non-responding customers into groups

Why Unsupervised Learning?



Why Clustering?

Clustering is used to **determine patterns** and to, ultimately, create **customized marketing campaigns**.



Why is model interpretability important?

Keeping interpretability of the model high is essential to reach our business goal

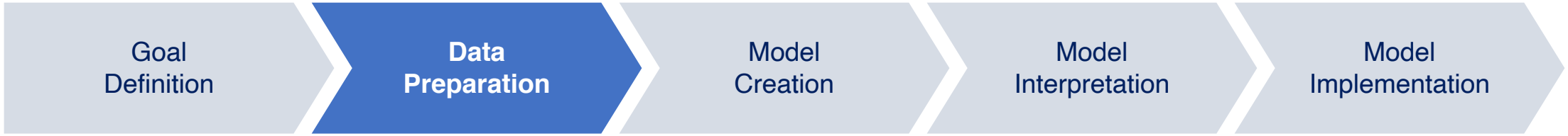
Interpretability of this Machine Learning Model is important in this specific context for **two main reasons**:

To create **appropriate suggestions and predictions**, it is preferable to **consider accurate and explainable machine learning models**.

In this analysis, it is essential to comprehend the model to **determine which variables lead to non-responding customers** and make the best possible decisions to **cater the marketing campaigns to their needs**.

For example …

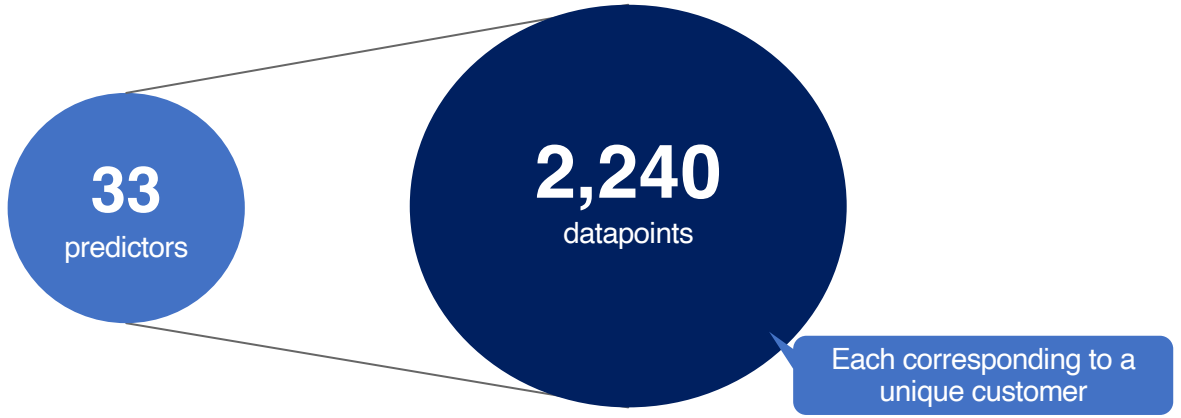
If we saw a pattern that customers with lower income are unlikely to respond, we would focus on that specific target and customize marketing campaigns for them, to increase the response rate.



2 Data Preparation

Data Exploration

Understanding the data: 4 groups of data across 33 predictors and 2,240 datapoints



Predictors can be divided into 4 main groups …

Personal Customer Data

Demographics such as Martial Status, Birth year, Teens & Kids at home etc.

Customer Purchasing Behaviour

Number of products purchased, online purchasing behaviour, Product preferences

Campaign responses

Response as the target variable for predicting whether Marketing campaign will be successful

Additional data available on the response to 5 previous marketing campaigns

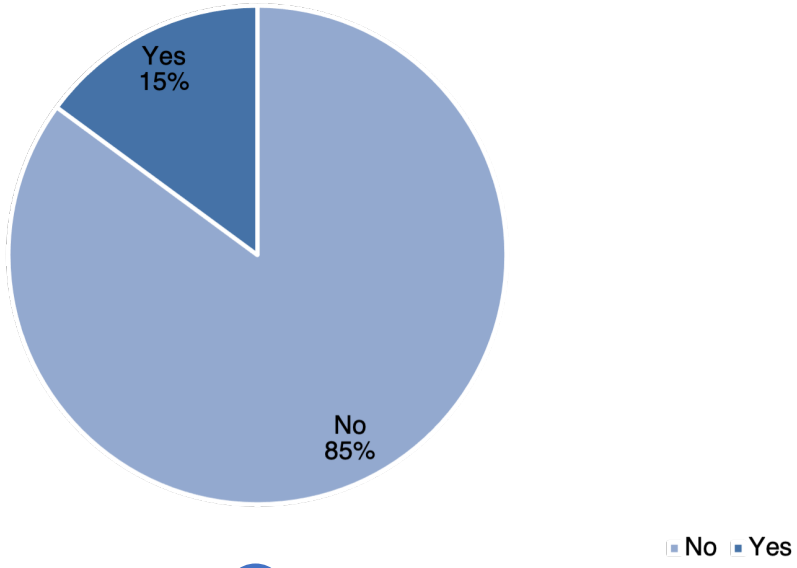
Unknown meaning

Not clear meaning or irrelevant for current business case

Data Exploration

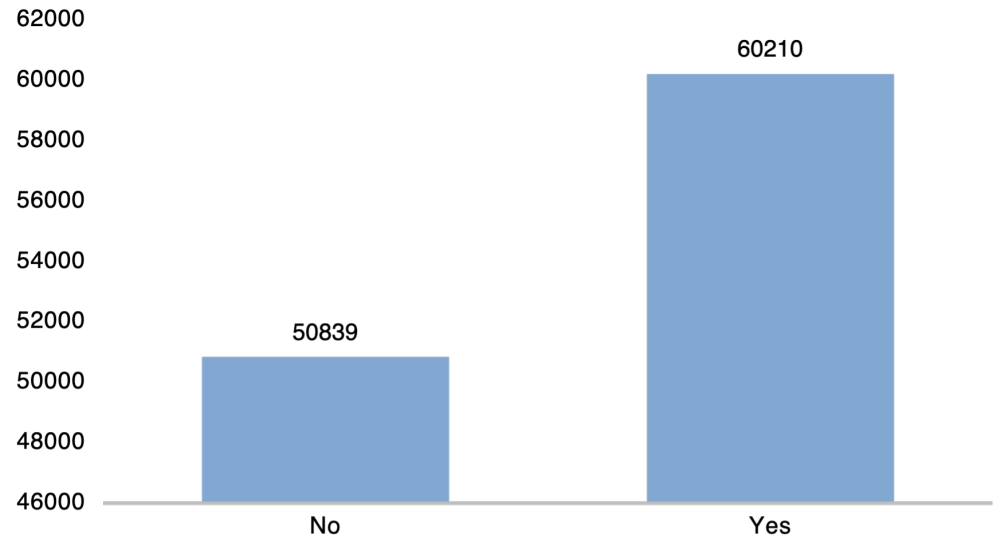
The response rate seems to be higher as income increases

Distribution of feature '*Response*'



'Response' is the target variable for our predictive modelling as it indicates whether the **marketing campaign triggered a response** from the customer.

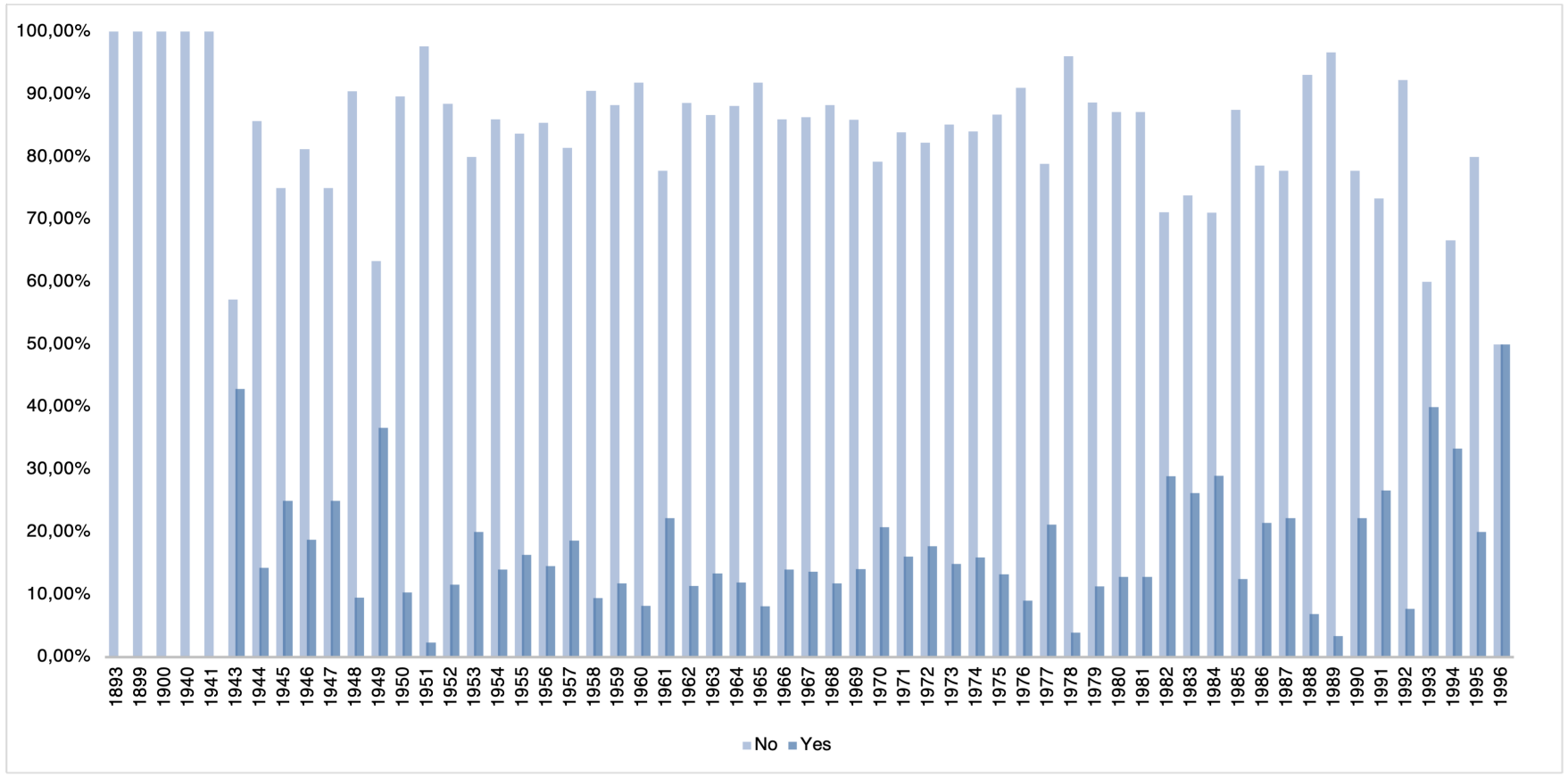
'*Average Income*' by Response



The average income of **customers responding** to the campaign is **higher** than the income of **customers not responding**.

Data Exploration

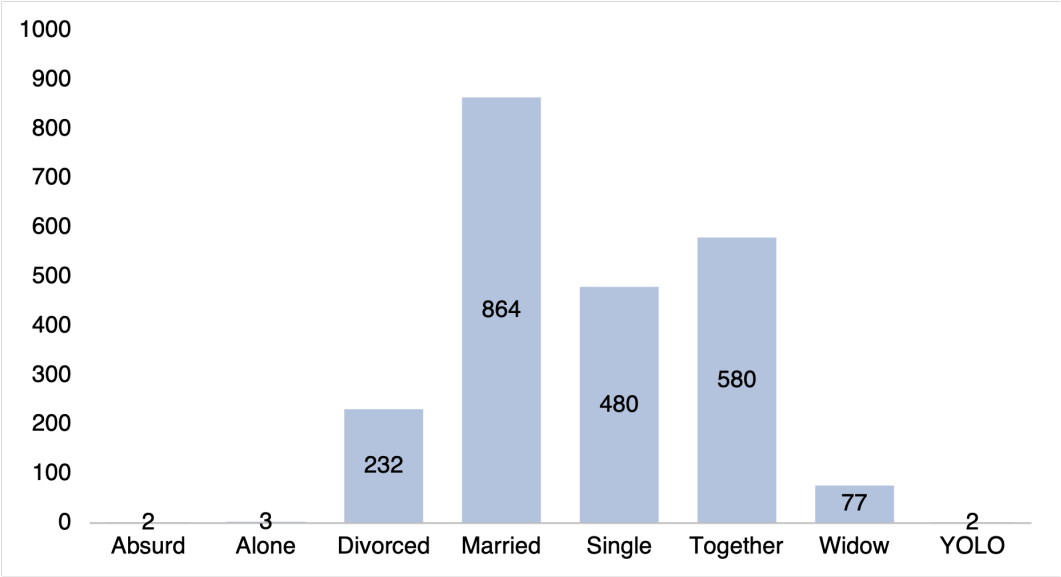
The response rate seems to be equally distributed across age groups



Data Exploration

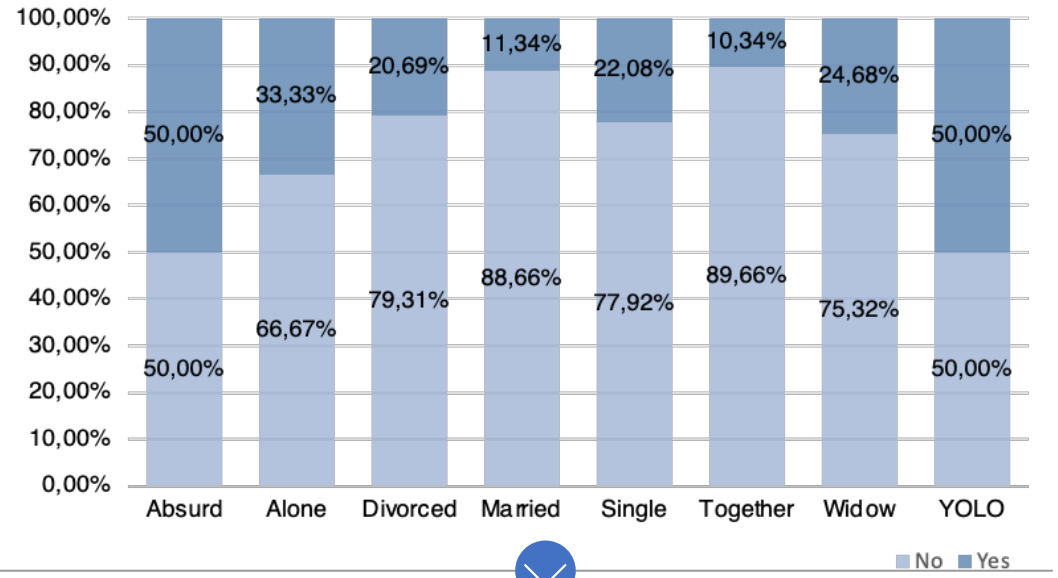
Past campaigns have attracted high positive responses from divorced and single customers

Customers by ‘Marital Status’



A **vast majority** of our customer base has a marital status of **'divorced', 'single', or 'together'.**

Response Ratio by ‘Marital Status’

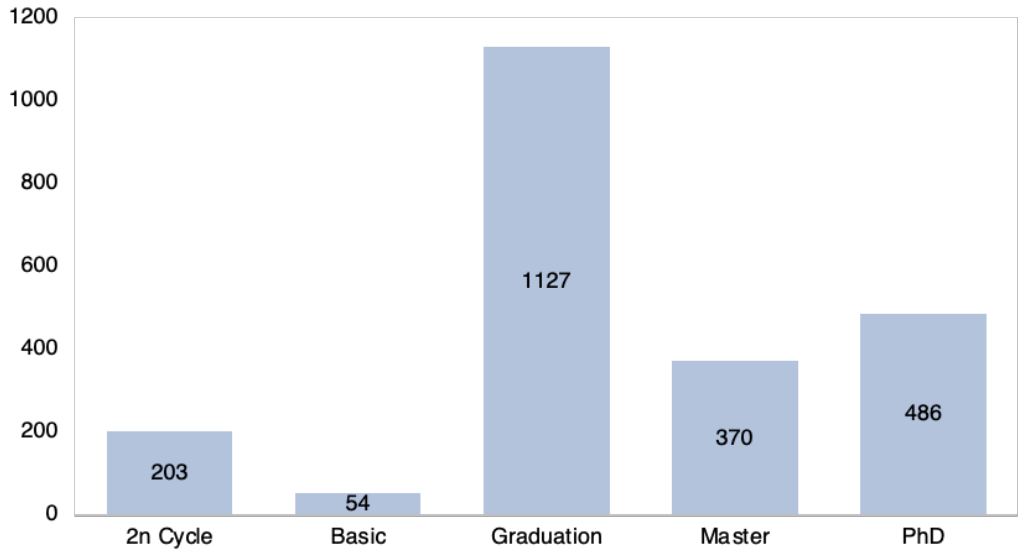


The campaign triggered a **higher response rates** for customers that are **'divorced' or 'single'** than for customers that are 'married' or 'together'.

Data Exploration

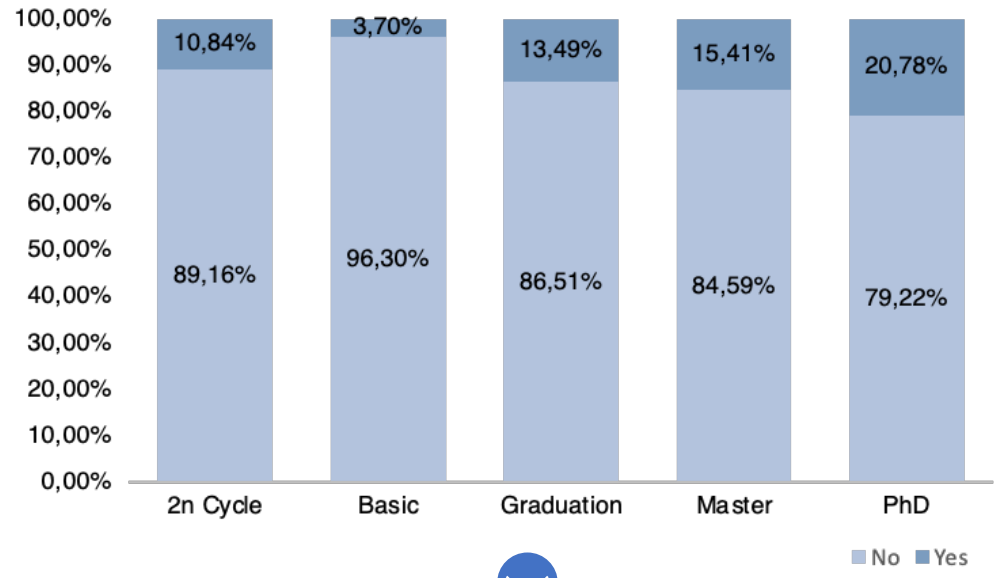
Customers with higher education were more likely to respond to past campaigns

Customers by ‘Education’



A **majority of our customer base has an education background 'Graduation'**, followed by 'PhD' and 'Master'.

Response Ratio by ‘Education’

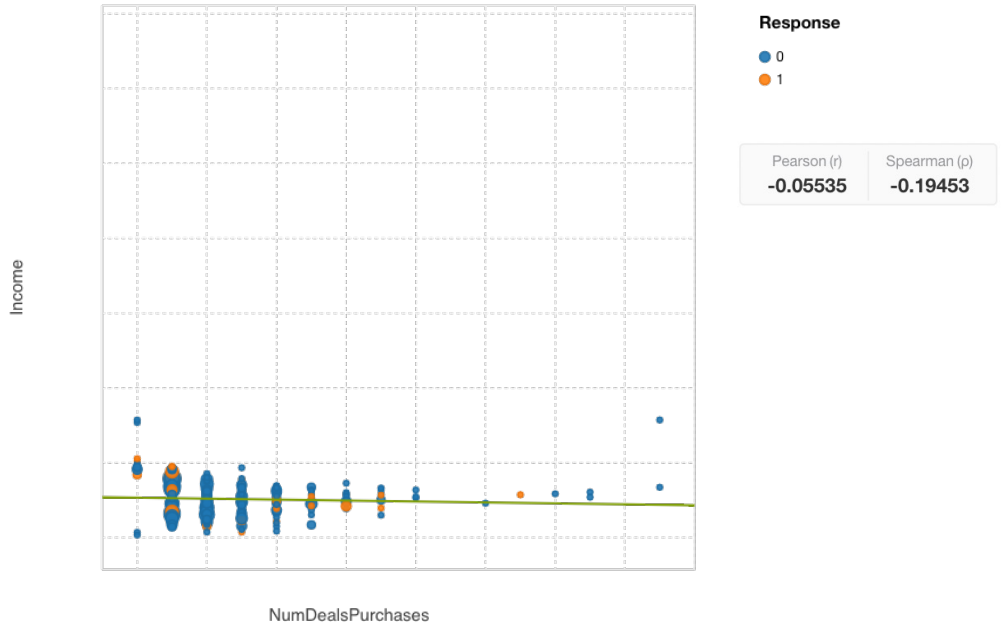


The campaign triggered a **higher response rates for customers groups that have a higher education**, such as a PhD or Master.

Data Exploration

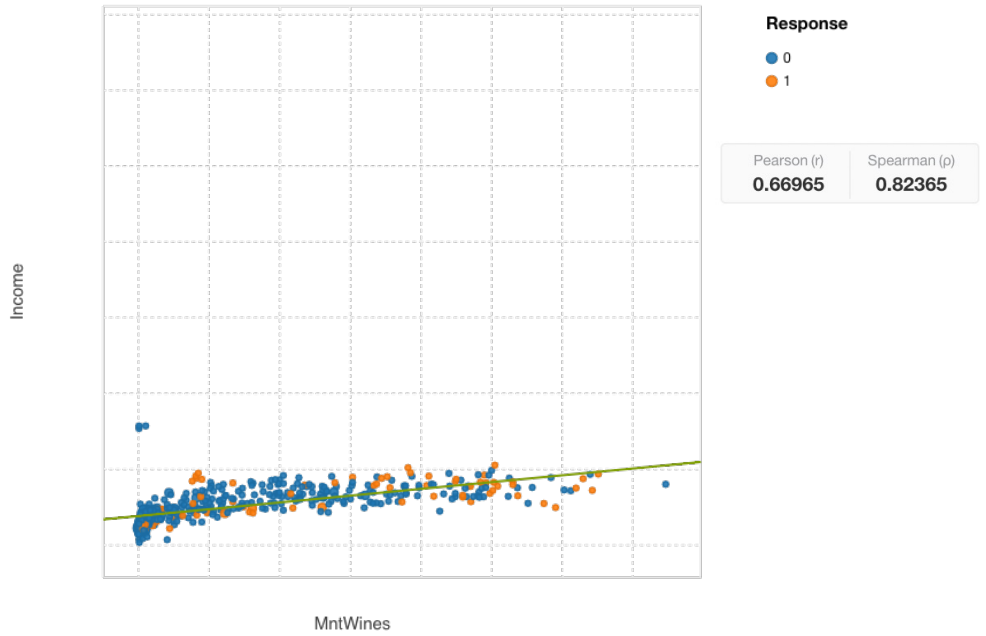
Customers with higher income are less respondent to deals but generally buy more wine

Number of Deals Purchased



Negative correlation between number of deals purchased and income.

Wine Products

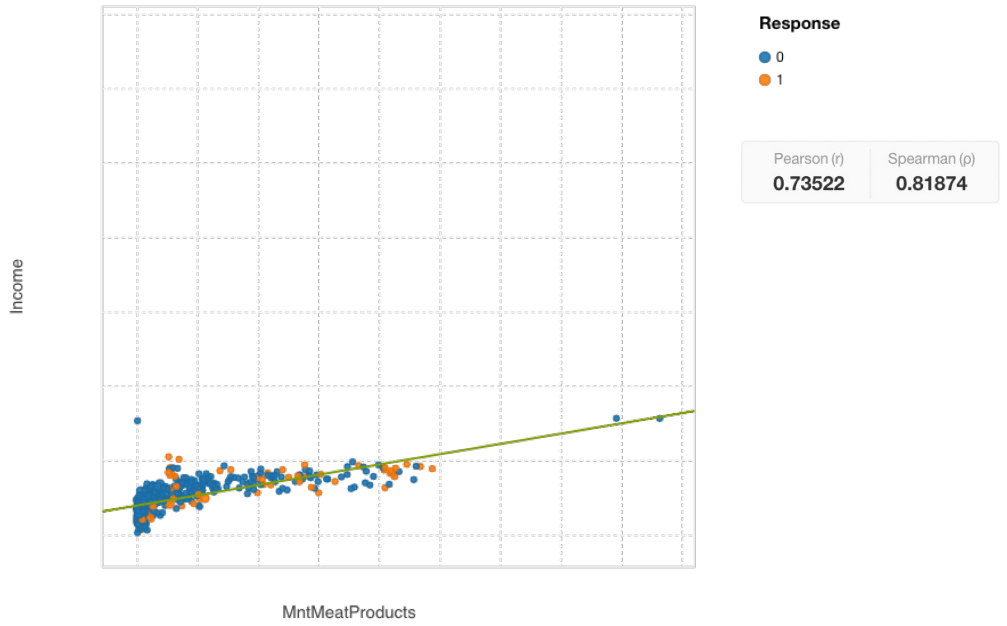


Positive correlation between number of Wines bought and income.

Data Exploration

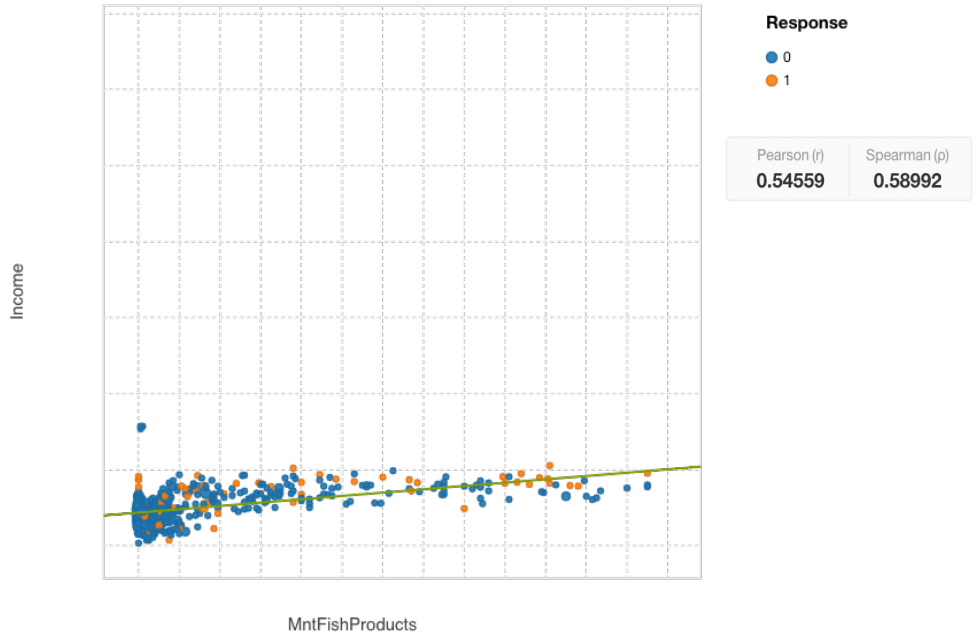
Customers with higher income are more likely to buy fish and meat products

Meat Products



Positive correlation between number of meat products bought and income.

Fish Products



Positive correlation between number of fish products bought and income.

Data Cleaning

How missing values have been handled

Features	# Missing values
Income	**24** (1% of all instances)

- *'Income'* is the **only predictor with missing values**
- All '*income*' missing values will be **replaced with value 0**

Rationale

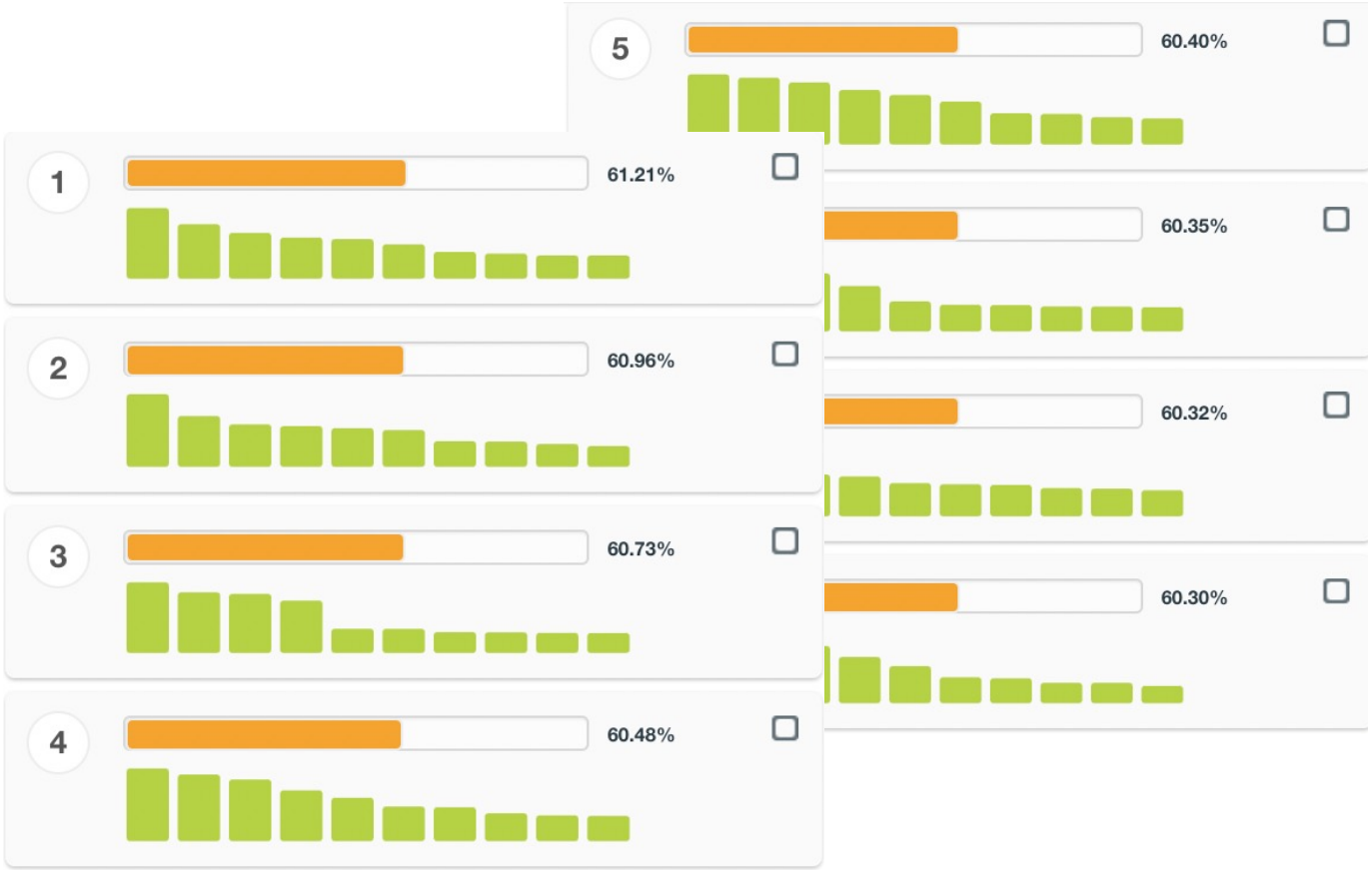
i. Other fields do not have any missing values

ii. There is only 6 instances with income '0'

iii. We can **assume that some people have kept the field empty** to indicate **zero income**.

Data Cleaning

9 detected anomalies have been removed from the dataset



Rationale

- Anomaly Detection discovered **9 instances** with an **anomaly score above 60%**
- To increase the performance of our model, **all of these 9 instances will be excluded from the analysis**.

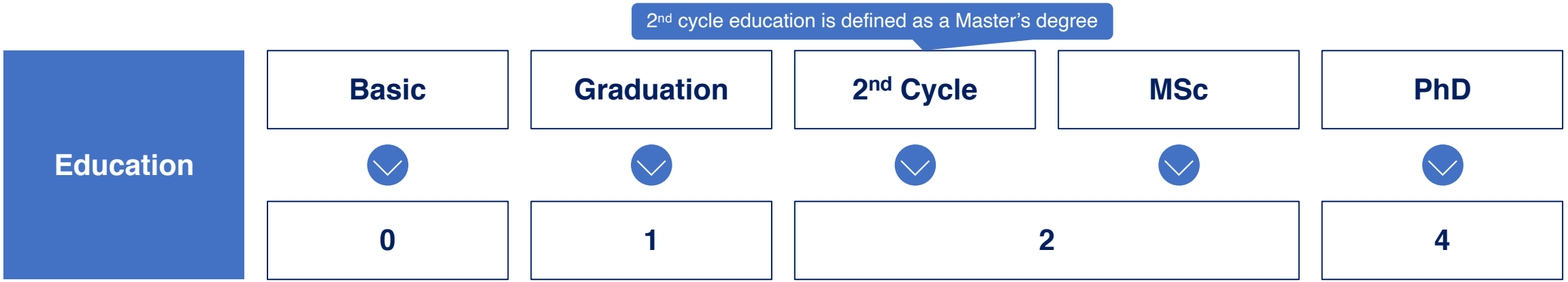
Data Cleaning

The predictor ‘Education’ has been converted from categorical to numerical

i. Since some algorithms do only work with numerical values, it is useful to **convert categorical variables into numerical variables.**

ii. In BigML this is done automatically, however, the default conversion is **one-hot encoding**

iii. For ‘education’, however, it makes sense to do **integer encoding** since the categories can be naturally ordered.



Data Cleaning

Outliers have been removed with different rationales for numerical and categorical variables

Outliers

Numerical variables

- Outliers for all numerical variables have been detected and excluded.
- Approach: The filter option in BigML has been used to remove ~1.5% of edge cases.

Categorical variables

- Outliers for all categorical variables have been detected and excluded.
- Approach: The filter option on BigML has been used to remove:
  - ***'Martial Status'***: categories with less than ~50 records ('*Alone*', '*Absurd*', *YOLO*'), keeping the 5 most frequent categories.

Data Cleaning

Removing unnecessary predictors

The following predictors have been **excluded from the dataset** for the reasons listed below:

Outliers	Rational for removal
Customer ID	• Excluding 'Customer ID' as it is a random number allocated to each individual customer. Hence, it does not have predictive power.
Dt_Customer	• Excluded since there is no interpretation available and hence, we would risk losing explainability of our model.
Z_CostContract	• Excluded since there is no interpretation available and hence, we would risk losing explainability of our model. • Furthermore, the predictors have the same value for the whole data set. Hence, the variables do not have predictive power.
Z_Revenue	

Data Cleaning

Z-score normalization is performed on all numeric features

What to do to rescale features?

i Z-score normalization has been conducted on **all continuous variables** in the dataset

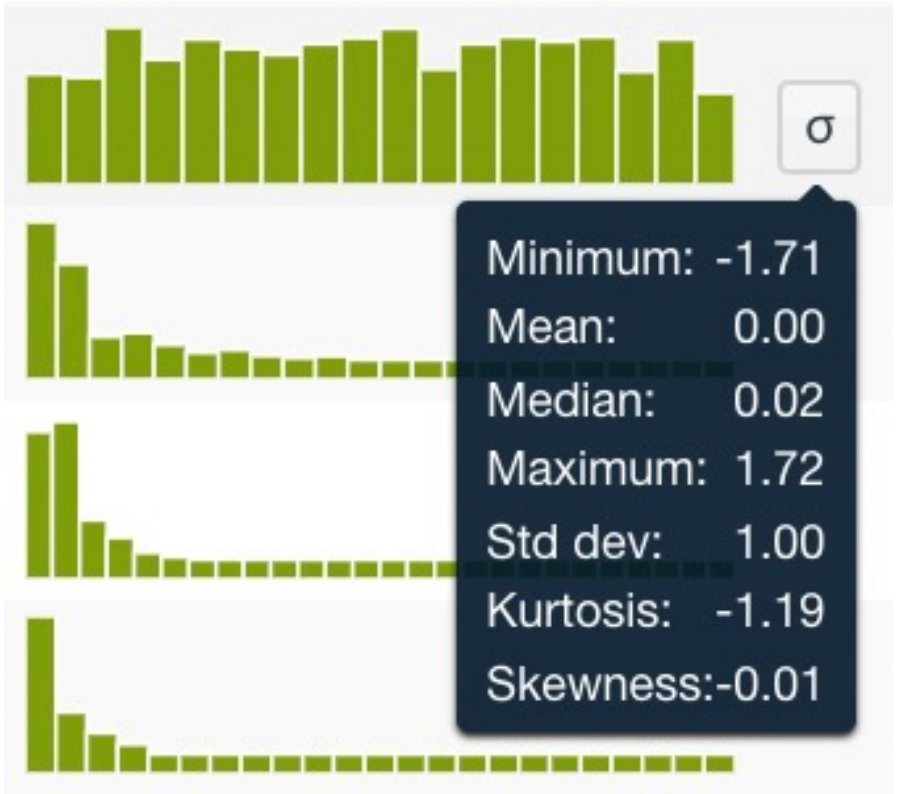
ii This centres the feature columns at **mean 0 and standard deviation 1** so that the feature columns take the form of a **normal distribution**

Why have features been rescaled?

To give **equal importance to each feature**

To facilitate ML algorithm to **process the data**

To make the algorithm **less sensitive to min-max scaling**



Data Cleaning

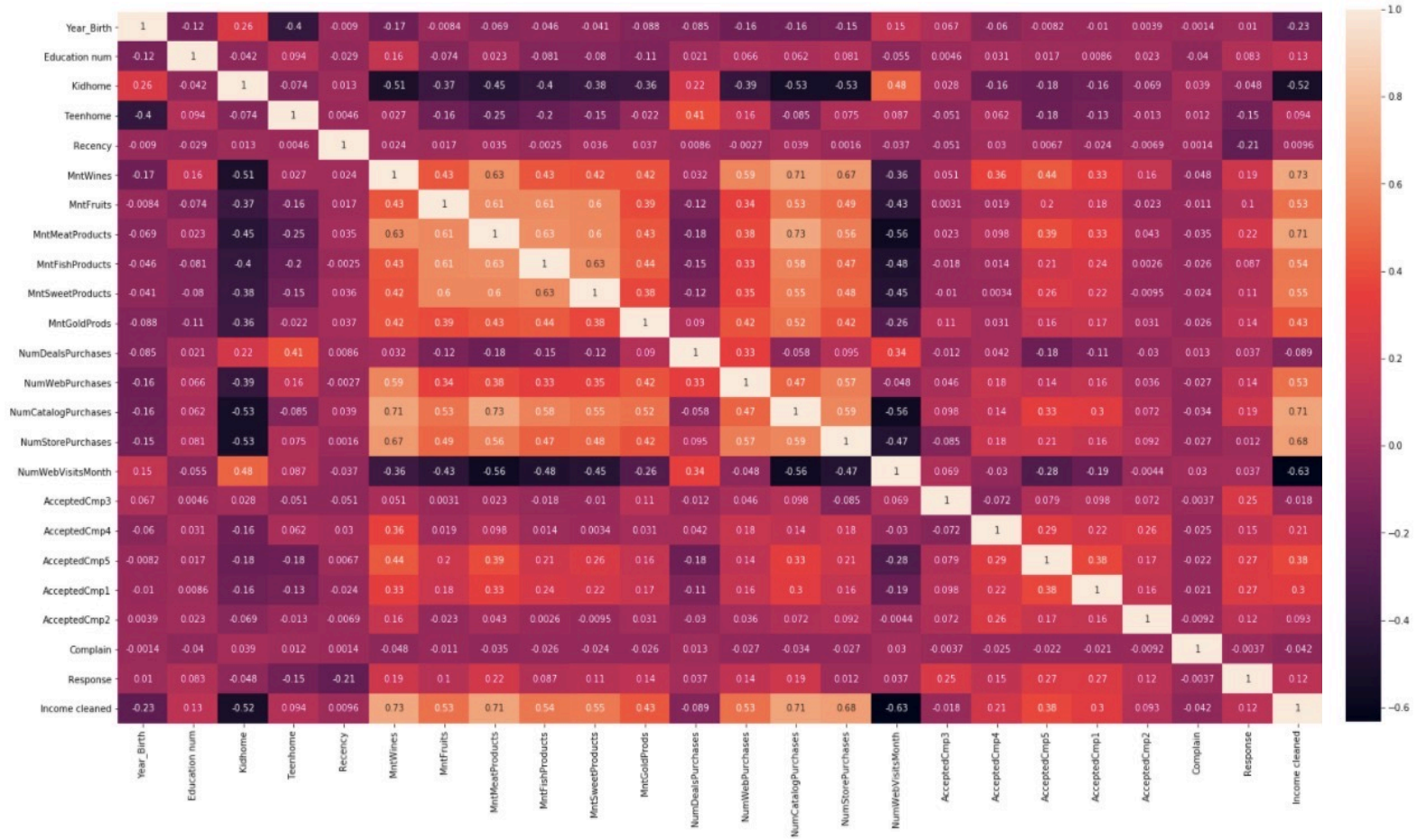
Predictors to be utilized for the analysis, post feature selection

Based on the data exploration and data cleaning, we will **move forward using the following predictors**:

Category	Features
Personal Customer Data	ID, Year_Birth, Education, Martial_Status, Income, Kidhome, Teenhome, Recency
Customer Purchasing Behavior	MntWines, MntFruits, MntMeatProducts, MntSweetProducts, MntGoldProds, NumDealsPurchases, NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth
Campaign Responses	AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5, Complain, ***Response***

Correlation matrix across all features

No feature selection to be performed prior to having a preliminary feature importance list



Feature removal

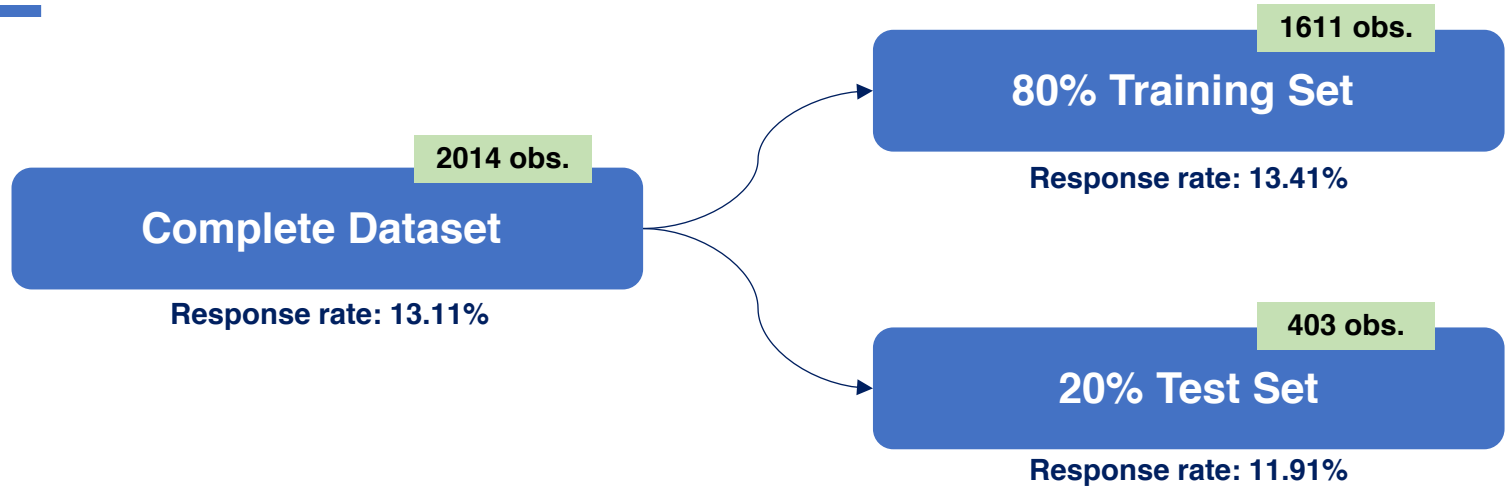
The following correlation matrix displays **moderate correlation** between:

- *NumCatalogPurchases & NumMeatProducts*: 0.73
- *NumCatalogPurchases & MntWines*: 0.71
- *Income & MntWines, NumMeatProducts, Num CatalogPurchases*: ~0.71- 0.73

- For now, no features are selected
- Feature selection will be **based on a preliminary feature importance list** !

Dataset test-train split

Performing the 80/20% split for supervised learning

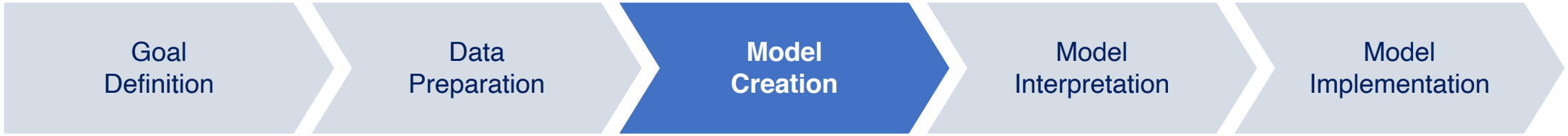


Why 80-20 Split Ratio?

i. For supervised learning, we are dividing the data into a training and test set. Thereby, we can use the **training set for creating the model** and then **evaluate it with the 'unknown data' of the test set**.

ii. The **80/20% split is a best practice** when creating and training models. Using this split on our data set, gives us 1611 units in our training and 403 units in our test split.

iii. The ratio of the **target variable, Response, is similar in the training and test set**.



3.1 Supervised Model Creation

Performance comparison of 3 models

To start off, a decision tree, an ensemble and a logistic regression have been trained

	Accuracy	F-measure	Precision	Recall	Phi-coefficient
Decision Tree	86.6%	0.9235	92.9%	91.8%	0.3841
Ensemble	88.8%	0.9394	89.9%	98.3%	0.2919
Logistic Regression	90.6%	0.9482	91.8%	98.0%	0.4578

By using the cleaned dataset to train the **three classification models** at our disposal we notice the resulting **performance metrics behave rather similarly.**

Feature engineering

Is the customer buying high end products?

Feature Engineering

1

Highend_Products = ***MntMeatProducts*** + ***MntFishProducts*** + ***MntWines***

- As identified in the heatmap, **the number of monthly meat, fish and wine products launched** are all **highly correlated** with each other as well as with the predictor **'income'**.

- Considering that all these products have a **comparable high price** and are affordable only with a **medium/ high average income**, we will explore summing them as 'Highend_Products'.

Feature engineering

Does the customer have a partner?

Feature Engineering

2 ***Partner[1]*** = ***True*** if "**Marital Status**" is '*Married*' or '*Together*'

- By creating a field called 'Partner', which takes on the value '1' when the person has a 'Martial Status' 'Married' or 'Together', and '0' otherwise, we can **indicate whether the customer has a partner or not**.

- This new feature will be **used to calculate the total 'Household size'** as seen on the following page.

Notes: Partner = if((Marital Status = 'Married' or 'Together'),1,0)

Feature engineering

How many people are part of the household?

Feature Engineering

3

Household_size = ***Partner*** + ***Kidshome*** + ***Teenshome*** + 1

- By taking the sum of the customer, partner, kids and teens at home, the **total size of the household can be calculated**.

- The assumption is that this predictor matters as the **customer would need to purchase & shop more if the household is bigger**.

Feature engineering

How much income is available for each household member?

Feature Engineering

4 ***Income_per_household_member*** = $\frac{\textit{Income}}{\textbf{Household\_size}}$

- By dividing the total ‘Income’ with the newly created feature ‘Household Size’, we can **calculate the total ‘income per household member’**.

- This matters as the **money available per person is higher if the size of the household is smaller**. In other words, this new predictor potentially has an **impact on the purchasing behaviour of the customer**.

Performance comparison of 3 models post feature engineering

The three models have been re-trained with the new features

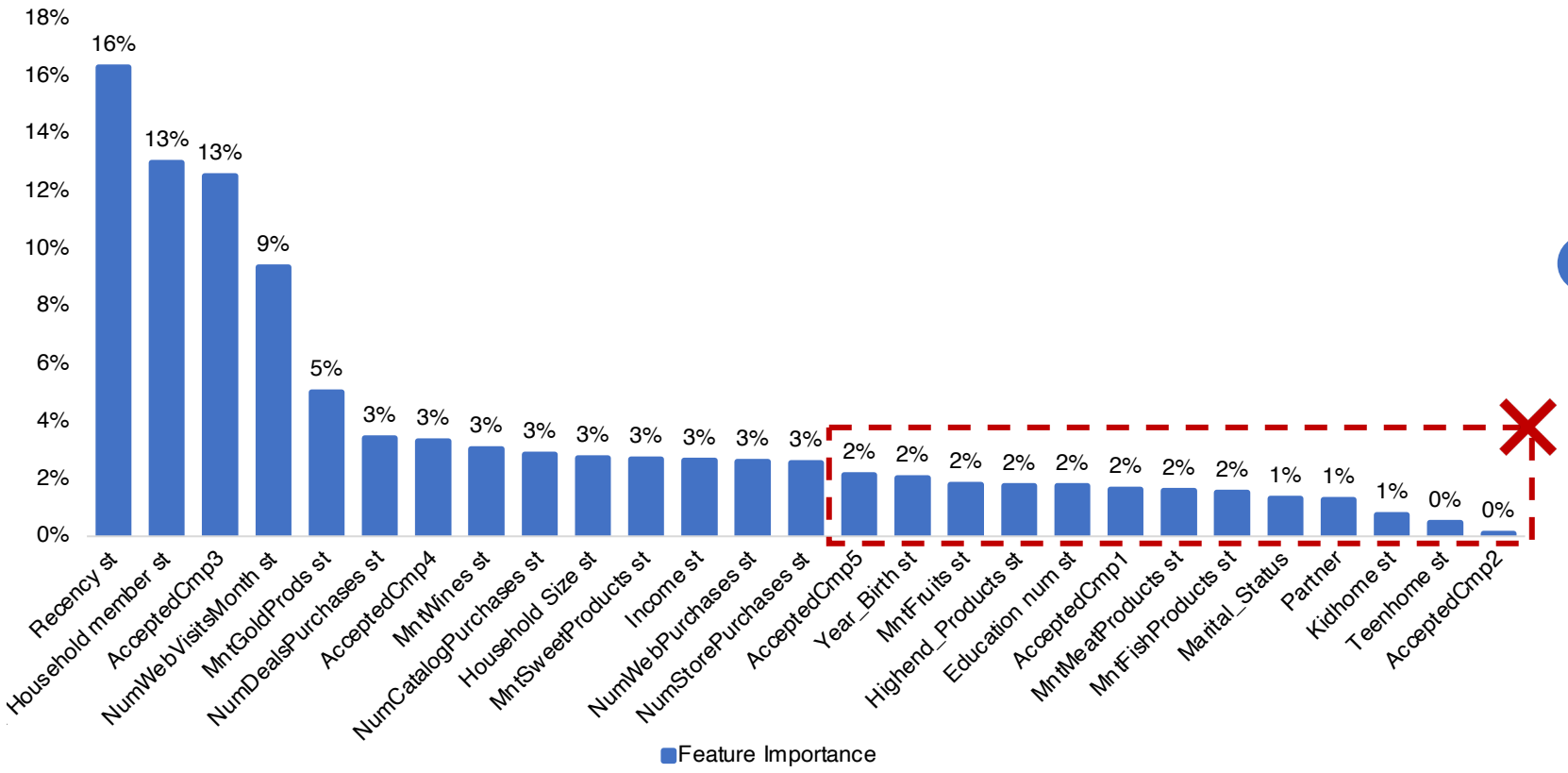
	Accuracy	F-measure	Precision	Recall	Phi-coefficient
Decision Tree	86.8% ↑	0.9255 ↑	92.4% ↑	92.7% ↑	0.3676 ↓
Ensemble	89.6% ↑	0.9423 ↑	92.0% ↓	96.6% ↓	0.4120 ↑
Logistic Regression	91.1% ↑	0.9506 ↑	92.8% ↑	97.5% ↓	0.5086 ↑

- Re-training the models after feature engineering has **improved most performance metrics**
- The **logistic regression** seems to be the **best performing model,** however, given its higher interpretability and its good performance, the **ensemble will be utilized** to continue the analysis.

Finetuning

Feature selection - All variables with less that 2% of predictive power have been removed

Feature Importance – Ensemble



- i Increases **robustness** of the model
- ii Reduces **noise** within the dataset
- iii Improves **performance** metrics

Finetuning

Feature selection has improved the model performance, making it more robust

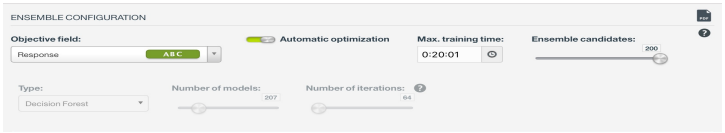
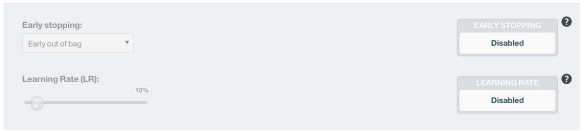
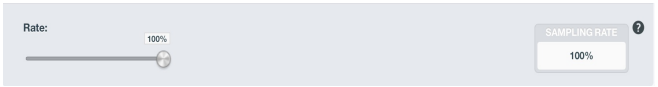
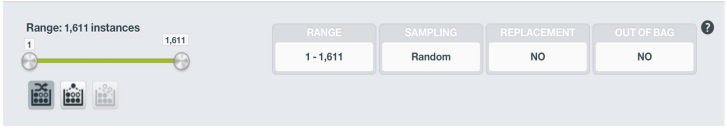
Accuracy	F-measure	Precision	Recall	Phi-coefficient
89.6%	0.9423	92.0%	96.6%	0.4120
90.1% ↑	0.9455 ↑	91.6% ↓	97.7% ↑	0.4254 ↑

After performing feature selection, **all performance metrics** (apart from precision) **have improved**.

Accuracy	• Accuracy was already satisfactory and increased even further, by 0.5%
Precision	• Precision slightly decreased by 0.4%, representing a rather insignificant change overall
Recall	• Recall was increased by 1.1% - feature selection has strengthened our model

Finetuning

Tweaking ensemble hyperparameters did not significantly improve the model's performance

Hyperparameters changed		Rationale
Model Type & Iterations		• Automatic optimization was chosen after boosted tree and decision forest did not gain better results
Boosting		• Boosting techniques are disabled when automatic optimization is activated
Weights		• Additional weight is applied on recency as it is the feature with highest predictive power according to slide 37
Sampling		• Data sampling rate is at 100%
Advanced Sampling		• Random sampling is used and out of bag option is disabled since sampling rate is already 100%

3 out of 5 metrics have improved, yet it is not sufficient to conclude the model is better

Accuracy	F-measure	Precision	Recall	Phi-coefficient
89.8% ↑	0.9436 ↓	92.2% ↑	96.6% ↓	0.4401 ↑

K-fold cross validation

Cross validation allows efficient data usage to overcome the overfitting limitation

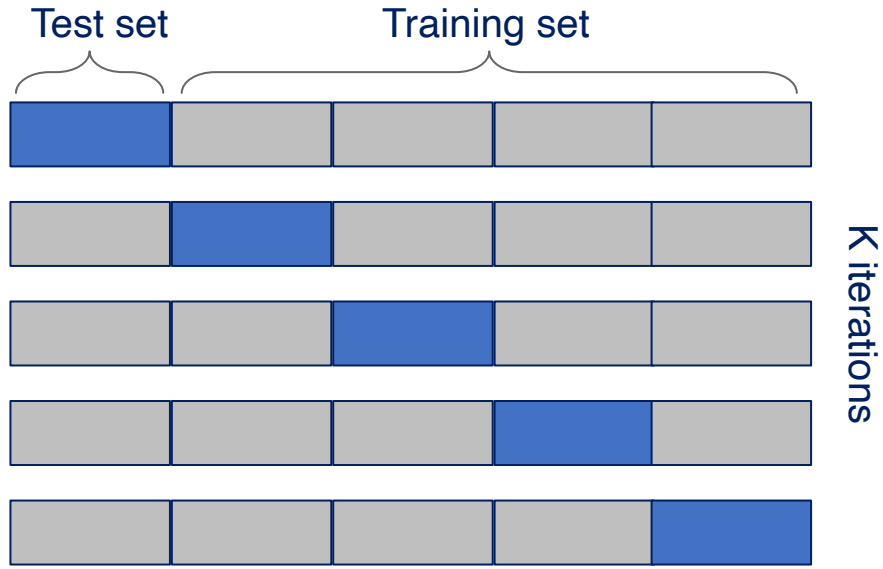
What is k-fold cross validation?

Problem:

When splitting the dataset between test and train, there is always a trade-off between the amount of data included in one dataset or the other.

Solution:

Through k resampling iterations, cross validation allows using the entire dataset both for testing and training.



K-fold cross validation

Cross validation performs worse on average but confirms the robustness of our model

Avg Accuracy

85.32% ⬇

STD = 0.753

Avg Precision

91.28% ⬇

STD = 0.676

Avg Recall

91.88% ⬇

STD = 1.259

Avg F-measure

0.9157 ⬇

STD = 0.004

Avg Phi

0.3405 ⬇

STD = 0.061

Insights

i. 5-fold cross validation is performed to evaluate how well the model is able to **predict given an unseen dataset.**

ii. On average, **all performance measures decrease**, however the change with respect to the best performing model is **not significant.**

iii. It may be concluded that the model has **satisfactory generalizability** and **sufficient robustness.**

Overfitting

Training set versus test set performance as a simple way to check overfitting

How well is my model responding to unseen data?

Potential problems	Solution
i Algorithm is too closely aligned with data set	i Reduce the number of predictive variables
ii There are too many variables	ii Shorten training time
iii The model 'memorizes' the training set and performs poorly with unseen data	iii Prune the decision tree

! To check for overfitting, the **model will be evaluated against the training set** !

Overfitting

Is our model overfitted?

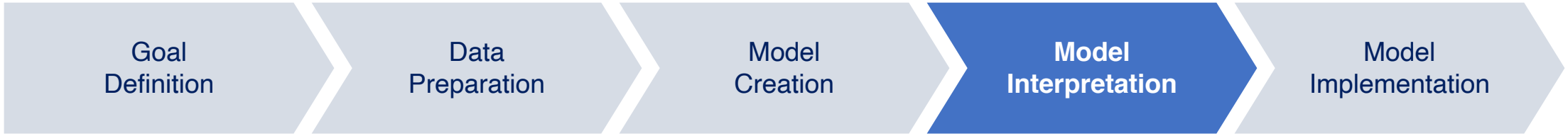
	Accuracy	F-measure	Precision	Recall	Phi-coefficient
Evaluation vs Test	89.8%	0.9436	92.2%	96.6%	0.4401
Evaluation vs Training	99.2% ⬆	0.9946 ⬆	99.2% ⬆	99.7% ⬆	0.9595 ⬆

Insights and conclusion on overfitting

i. Our **training set does perform better than test set**, which is typical in machine learning.

ii. Although metrics from training set are all close to 100%, evaluation with test set shows **our model only falls behind by mere 3% to 10%**, showing signs of **robustness.**

iii. Therefore, we can conclude that our **model is not overfitted.**



3.2 Supervised Model Interpretation

Metrics interpretation

How is our model behaving with respect to our business goal?

Positive Class: *Response = False*

ACTUAL VS. PREDICTED	0	1	ACTUAL	RECALL
0	347	8	355	97.75%
1	32	16	48	33.33%
PREDICTED	379	24	403	65.54% AVG. RECALL
PRECISION	91.56%	66.67%	79.11% AVG. PRECISION	90.07% ACCURACY

TP	Customer does not respond, as predicted
TN	Customer responds, as predicted
FP	Customer will respond, contrary to prediction
FN	Customer will not respond, contrary to prediction

Accuracy	F-measure	Precision	Recall	Phi-coefficient
90.1%	0.9455	91.6%	97.7%	0.4254

	Technical Interpretation	Business Interpretation
Accuracy	Percentage of correctly predicted instances of total prediction	How likely the model makes mistakes in predicting (non-)responses
F-measure	Balanced combination of precision and recall	Useful to compare the performance of different models
Precision	Correct predictions over predicted instances in positive class	How trustable is the model in predicting (non-)responses
Recall	Correct predictions over actual instances in positive class	How well is the model effectively predicting (non-)responses
Phi-coefficient	Correlation between predicted and actual values	How close are the predictions to reality

Most relevant metrics

Compatibly to our business problem, our main goal is to maximize recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

Minimize False Negatives to avoid unnecessary loss of potential revenues

Why is recall relevant?

i. **False Negatives** represent customers whom our model has predicted would respond to the marketing campaign but, in reality, would not respond.

ii. Our **goal is to capture as many non-respondents as possible,** turn them into respondents and increase the response rate.

iii. Missing non-responding customer means **lowering our chance to increase revenue** through a more **effective marketing campaign.**

! Minimizing false negatives, thus maximizing recall is essential to **increase the potential financial upside.** !

Features with highest predictive power (I/II)

Why the 6 most important features hold most of the predictive power of the model

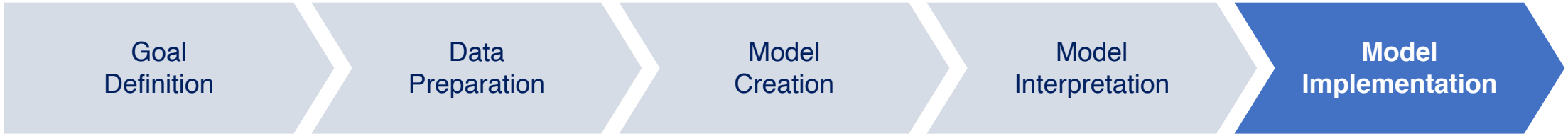
Features	Importance (%)	Possible Explanation
Recency st	**18.9%**	Number of days since the last purchase; customers with a **high recency** tend to be **more likely to not respond** to the campaign.
Income per Household member	**14.5%**	The amount of money available per person; customers with a **low income** per household member tend to be **more likely to not respond** to the campaign.
NumWebVisits Month st	**10.8%**	Number of visits to company's web site in the last month; customers with a **low** number of web visits per month tend to be **more likely to not respond** to the campaign.
AcceptedCmp3	**10.2%**	1 if customer accepted the offer in the 3rd campaign, 0 otherwise; customers which did **not accept** the 3rd campaign tend to be **more likely to not respond** to the campaign.

Features with highest predictive power (II/II)

Why the 6 most important features hold most of the predictive power of the model

Features	Importance (%)	Possible Explanation
MntGoldProds st	7.7%	Amount spent on gold products in the last 2 years; customers with a **low** MntGoldProds tend to be **more likely to not respond** to the campaign.
MntWines st	6.8%	Amount spent on wine products in the last 2 years, customers with a **high** MntWines tend to be **more likely to not respond** to the campaign.

The top **6 most important features** account for **68.9%** of the predictive power of the model.



3.3 Supervised Model Implementation

Assumptions

Implementing a base & premium marketing campaign to target respondents & non-respondents

Items	Figures	Explanation
Total number of customer	**560.000**	Assuming 17.5% market share of Madrid's market (3.2 million total population)
Success rate of premium campaign	**20%**	Percentage of TP & FP (*non-responding customers*) our premium marketing campaign will be able to capture, turning them into responding customers
Basic marketing campaign cost per customer	**€6**	Cost of marketing initiative towards TN and FN – one-time cost (*predicted responding customers*)
Premium marketing campaign cost per customer	**€10**	Cost of marketing initiative towards TP and FP – one-time cost (*predicted non-responding customers*)
Revenue per customer	**€55**	Revenue a responding customer brings thanks to the marketing campaign (for the basic and premium campaign)

Economic valuation of predicted responding customers

Predicted responding customers will be targeted with the basic marketing campaign

Financial impact of **basic marketing campaign**

Explained more in details in following slides

Financial impact per customer

Actual \ Predicted	Non-Response	Response
Non-Response	+€1	-€6
Response	+€45	+€49

Rationale

TP	=€55 of revenues x 20% - €10 of premium marketing
TN	=€55 of revenue - €6 of basic marketing costs
FP	=€55 of revenue - €10 of premium marketing costs
FN	=€6 of basic marketing costs

Impact of basic campaign on revenues and costs

Additional revenues and costs incurred due to the basic marketing campaign

Financial impact of **basic marketing campaign**

Model predictions for entire customer base

Actual \ Predicted	Non-Response	Response
Non-Response	482,184	11,117
Response	47,246	19,545

All 560,000 customers have been distributed along the confusion matrix

Changes in financial figures

i Cost of Basic Campaign

The campaign in total will cost **€183,424**
= (11,117 + 19,545) x €6

ii Cost of False Negatives

€66,702 will be wasted on False Negatives
= 11,117 x €6

iii Revenue by True Negatives

True Negatives will bring **€1,069,975** in revenue
= 19,545 x €55

Impact on bottom line of the basic marketing campaign alone

By only targeting predicted responding customers with a basic campaign profit increases by 27%

Current scenario without prediction model	
Revenue	
Revenue for responding customers[1]	€4,004,000
Costs	
Base marketing campaign[2]	(€3,360,000)
Net impact	**€644,000**

Based on historical data, 13% response rate (13% x 560,000 = 72,800)

Future scenario with model prediction *only basic marketing campaign*	
Revenue	
Revenue from TN[4]	€ 1,069,975
Costs	
Base marketing campaign[2]	(€183,424)
Net impact	**€ 886,551**

+€242,551

increase in profits thanks to **basic marketing campaign**

Calculations: **1**: 13% x 560,000 x €55 I **2**: 13% x 560,000 x €6 I **3**: 482,184 TP x 20% x €55 I **4**: 19,454 TN x €55 I **5**: 47,246 FP x €55 I **6**: (482,184+47,246) x €10

Economic valuation of predicted non-responding customers

Predicted non-responding customers will be targeted with the premium marketing campaign

Financial impact of **premium marketing campaign**

Explained more in details in following slides

Financial impact per customer

Actual \ Predicted	Non-Response	Response
Non-Response	+€1	-€6
Response	+€45	+€49

Rationale

TP	=€55 of revenues x 20% - €10 of premium marketing
TN	=€55 of revenue - €6 of basic marketing costs
FP[1]	=€55 of revenue - €10 of premium marketing costs
FN	=€6 of basic marketing costs

1: **False Positives** are customers who are unnecessarily targeted with a premium marketing campaign, whereas a basic marketing campaign would have been enough to make them respond

Impact of premium campaign on revenues and costs

Additional revenues and costs incurred due to the basic marketing campaign

Financial impact of **premium marketing campaign**

Model predictions for entire customer base

Actual \ Predicted	Non-Response	Response
Non-Response	482,184	11,117
Response	47,246	19,545

All 560,000 customers have been distributed along the confusion matrix

Changes in financial figures

i Cost of Premium Campaign

The campaign in total will cost **€5,294,292**
= (482,184 + 47,246) x €10

ii Revenue by True Positives

True Positives will bring **€5,304,019** in revenue
= 482,184 x 20% x €55

iii Revenue by False Positives

False Positives will bring **€2,598,511** in revenue
= 47,246 x €55

Impact on bottom line of the premium marketing campaign alone

If 20% of TP can be converted with a premium marketing campaign, profit increases by ~€2M

Current scenario without prediction model	
Revenue	
Revenue for responding customers[1]	€4,004,000
Costs	
Base marketing campaign[2]	(€3,360,000)
Net impact	**€644,000**

Based on historical data, 13% response rate
(13% x 560,000 = 72,800)

Future scenario with prediction model	
Revenue	
Revenue from TP[3]	€ 5,304,019
Revenue from FP[5]	€ 2,598,511
Costs	
Premium marketing campaign[6]	(€5,294,292)
Net impact	**€2,608,238**

+€1,964,238
increase in profits thanks to
premium marketing campaign

Calculations: **1**: 13% x 560,000 x €55 I **2**: 13% x 560,000 x €6 I **3**: 482,184 TP x 20% x €55 I **4**: 19,454 TN x €55 I **5**: 47,246 FP x €55 I **6**: (482,184+47,246) x €10

Overall (per customer) economic valuation of model outcomes

Financial impact of basic and premium marketing campaign (per customer)

Financial impact per customer

Actual \ Predicted	Non-Response	Response
Non-Response	+€1	-€6
Response	+€45	+€49

Rationale

TP	=€55 of revenues x 20% - €10 of premium marketing
TN	=€55 of revenue - €6 of basic marketing costs
FP	=€55 of revenue - €10 of premium marketing costs
FN	=€6 of basic marketing costs

Overall financial impact of model deployment

With the two suggested marketing campaigns, profits could be increased by 81.6%

Current scenario without prediction model	
Revenue	
Revenue for responding customers[1]	€4,004,000
Costs	
Base marketing campaign[2]	(€3,360,000)
Net impact	**€644,000**

Based on historical data, 13% response rate (13% x 560,000 = 72,800)

Future scenario with prediction model	
Revenue	
Revenue from TP[3]	€ 5,304,019
Revenue from TN[4]	€ 1,069,975
Revenue from FP[5]	€ 2,598,511
Costs	
Base marketing campaign[2]	(€183,424)
Premium marketing campaign[6]	(€5,294,292)
Net impact	**€3,494,789**

+€2,850,789
increase in profits

+81%

Calculations: **1**: 13% x 560,000 x €55 I **2**: 13% x 560,000 x €6 I **3**: 482,184 TP x 20% x €55 I **4**: 19,454 TN x €55 I **5**: 47,246 FP x €55 I **6**: (482,184+47,246) x €10

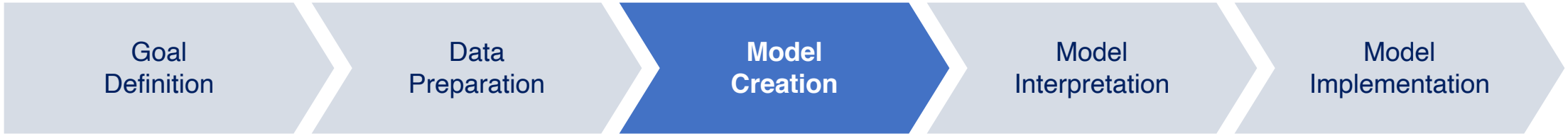
The two marketing campaigns should target two customer types

Details on customer personas for different marketing campaigns

Preliminary recommendation

Predictors	Basic Marketing Campaign	Premium Marketing Campaign
Purchase Recency	• Target frequent and loyal customers with **less than 50 days since last purchase**	• Target occasional customers with **more than 50 days since last purchase**
Income	• Target top 10% most wealthy customers with **income above €75,000**	• Target 60% of customers with **income between €30,000 and €75,000**
Web visits	• Target customers which visit the company website **more than 8 times a month**	• Target customers which visit the company website **less than 8 times a month**
Previous Responses	• Target customers who **have responded to previous campaigns** (especially the 3rd)	• Target customers who have been **less responsive to previous campaigns**
High-end products	• Basic campaign should revolve around more expensive products such as **wine, cheese** and the so-called **gold products**	• Premium marketing campaign should focus more on low-end products such as **fruits and sweets**

- Unsupervised learning will be utilized to cluster target customers for Premium Marketing Campaign
- Customer personas will be analysed based on different customer segments



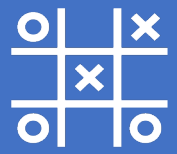
4.1 Unsupervised Model Creation

Clustering as a tool to create effective premium campaigns

Goal & Approach

Goal

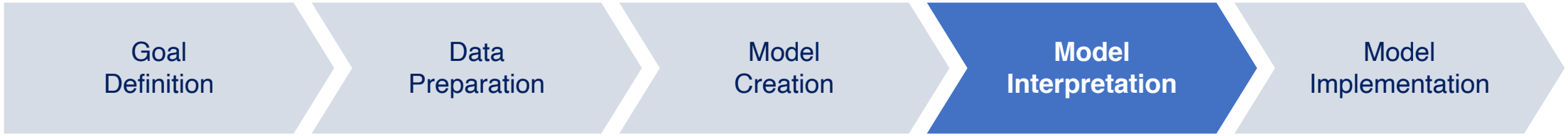
- Gain a better understanding of customers which are not responding to the base campaign.
- Based on these insights, design **premium marketing campaigns for different customer groups.**



Approach

- **Focusing only on customers not responding** to the base campaign (Response = False)
- To ensure equal contribution, Clustering is conducted based on **standardized predictors.**
- By scaling the original (non-standardized) features to zero, hence including them in the Clustering, we can keep a **high interpretability of the Clusters.**

! To choose a relevant and useful k for our Clustering, we are creating a Clustering based on **k = 2 and k = 3 and compare their interpretability and usefulness.** !



4.2 Unsupervised Model Interpretation

Clustering k-means model interpretation (I/II)

k = 3 clusters described along the 8 most relevant features

k = 3

By evaluating the three different clusters based on differences in their feature values, we found that they can best be **differentiated based on the following 8 features**:

	Number of Deals Purchased	High-end Products	Household Size	Income per household member
0	**High** (3.22)	**Medium** (655.97)	**Low/ Medium** (2.79)	**Medium** (22886.49)
1	**Low** (1.16)	**High** (1086.33)	**Low** (1.82)	**High** (45773.55)
2	**Medium** (2.13)	**Low** (84.46)	**Medium** (2.95)	**Low** (12813.41)

Clustering k-means model interpretation (II/II)

k = 3 clusters described along the 8 most relevant features

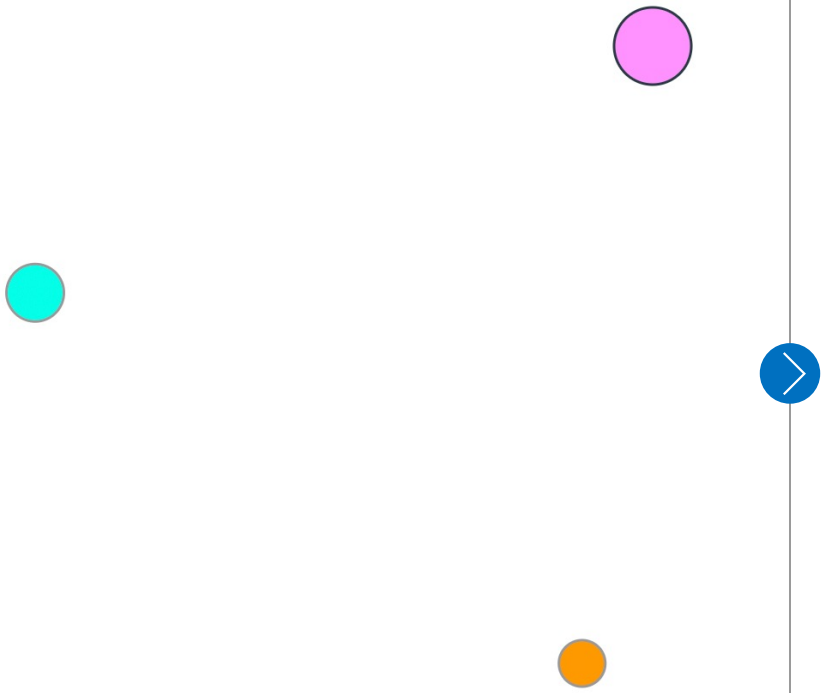
k = 3

By evaluating the three different clusters based on differences in their feature values, we found that they can best be **differentiated based on the following 8 features**:

	Number of Catalog Purchases	Number of Store Purchases	Mnt Wines	Kidhome
0	**Medium/ High** (3.35)	**Medium/ High** (8.33)	**Medium/ High** (469.12)	**Low** (0.17)
1	**High** (5.81)	**High** (8.7)	**High** (550.83)	**Low** (0.03)
2	**Low** (0.57)	**Low/ Medium** (3.42)	**Medium** (51.55)	**Medium** (0.77)

Clustering k-means model interpretation

Summary of the k = 3 clusters generated by the k-means algorithm



Cluster 0 (963 units)

- Small to medium-sized households with a medium income per person, a medium to high number of purchases, and a high preference for special deals.

Cluster 1 (491 units)

- Small households with a high income per person, a high number of purchases and specifically high-end purchases, and a low amount of deal purchases.

Cluster 2 (296 units)

- Medium-sized households with a low income per person, a medium number of purchases and little high-end purchases.

Clustering k-means model interpretation (I/II)

k = 2 clusters described along the 8 most relevant features

k = 2

By evaluating the two different clusters based on differences in their feature values, we found that they can best be **differentiated based on the following 8 features**:

	Number of Deals Purchased	High-end Products	Household Size	Income per household member
0	**Medium/ High** (2.42)	**Low** (188.17)	**Medium** (2.92)	**Low** (14750.90)
1	**Low** (1.82)	**High** (997.95)	**Low** (2.15)	**High** (37975.90)

Clustering k-means model interpretation (II/II)

k = 2 clusters described along the 8 most relevant features

k = 2

By evaluating the two different clusters based on differences in their feature values, we found that they can best be **differentiated based on the following 8 features**:

	Number of Catalog Purchases	Number of Store Purchases	Mnt Wines	Kidhome
0	**Low/ Medium** (1.06)	**Medium** (4.37)	**Low/ Medium** (131.51)	**Medium/ High** (0.64)
1	**High** (5.28)	**High** (9.06)	**High** (565.06)	**Low** (0.06)

Clustering k-means model interpretation

Summary of the k = 2 diverse clusters generated by the k-means algorithm

Cluster 0 (1217 units)

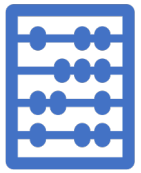
- Medium-sized households with a comparably low income per person, and a preference for special deals.

Cluster 1 (533 units)

- Small households with a high income per person, a high number of purchases, and a low amount of deal purchases.

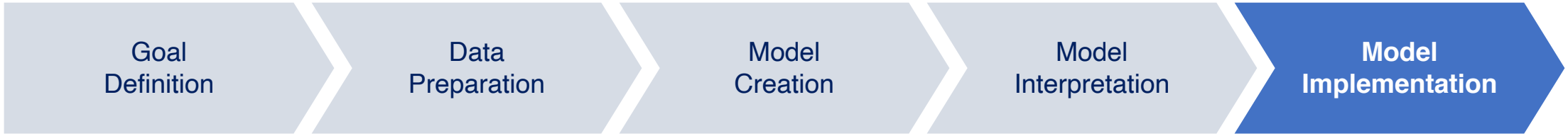
Choosing the right k

Reasoning for prioritizing k = 2 over k = 3



We chose to follow a simple policy of "**Less is More**" and hence use **k = 2** for implementation.

This is especially helpful because it **increases interpretability** while simultaneously **decreases complexity**, which is especially helpful in the analysis of business cases.



4.3 Unsupervised Model Implementation

Cluster 0: Premium Campaign Recommendations

How will our premium marketing campaign target Cluster 0?

Preference for special deals

Implications: Marketing campaigns in which special products are reduced in price or cheaper in bulk will be especially interesting for customers of Cluster 0.

Limited interest in high-end products

Implications: Measure whether interest in high-end products would be higher if the price of high-end products is reduced as part of a campaign. If not, focus on lower-end products in marketing campaigns.

Medium household size

Implications: Including big product sizes into campaigns might be of great interest for customers from Cluster 0.

Cluster 0: Premium Campaign Recommendations

How will our premium marketing campaign target Cluster 0?

Low income per Household member

Implications: Marketing campaigns should focus on price reductions.

Low to medium number of catalogue & store purchases

Implications: A low number of purchases implies that customers from Cluster 0 are mostly purchasing essential products.

Low/ medium number of wine purchases and medium/ high number of kids at home.

Implications: For customers with kids, wine should not be promoted as there is a high negative correlation between these two features.

Cluster 0: Premium Campaign Recommendations

Targeting Cluster 0 in summary

Special deals

Large packaging sizes/ high quantities

Focus on low pricing

Promote essential products

Insights

i. For customers of Cluster 0, we would recommend to **focus on essential products with a medium to big packaging size** while **promoting the price**.

ii. For example, having a **special offer to get three, essential items for the price of two**.

Cluster 1: Premium Campaign Recommendations

How will our premium marketing campaign target Cluster 1?

Low preference for special deals

Implications: Marketing campaigns which focus on special deals are not very interesting to Cluster 1 customers. Instead, campaigns which focus on product launches might be of greater interest.

High interest in high-end products

Implications: High-end products are especially interesting for customers in Cluster 1. Hence, campaigns should focus on wine, meat, and fish.

Low household size

Implications: Small product sizes are, most likely, more interesting for customers in Cluster 1.

Cluster 1: Premium Campaign Recommendations

How will our premium marketing campaign target Cluster 1?

High income per Household member

Implications: Campaigns should focus on the product itself more than reductions in product prices. This is mainly because as customers are, most likely, not as price sensitive.

High number of catalogue & store purchases

Implications: Customers of Cluster 1 are more likely to spend money on premium/ non-essential products. Hence, campaigns should focus on non-essential products.

High number of wine purchases and low number of kids at home

Implications: Wine should be promoted for customers with no kids. Since the average number of kids at home for Cluster 1 customers is low, wine is a promising product for marketing campaigns.

Cluster 1: Premium Campaign Recommendations

Targeting Cluster 1 in summary

Focus on high-end products

Small packaging sizes

Focus on the product itself

Promote non-essential products

Insights

i. For customers of Cluster 1, we would recommend to **focus on non-essential, high-end products with small packaging sizes.**

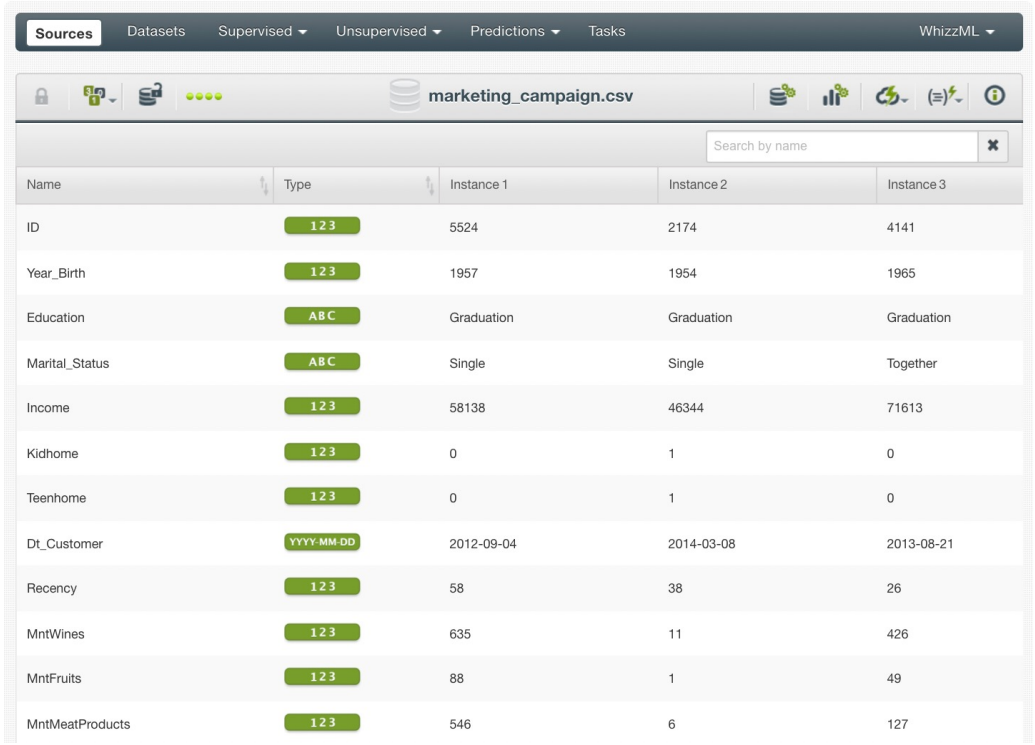
ii. For example, promoting **a newly launched wine** with a focus on the product itself and its quality rather than the price.

Technical Appendix

Appendix: Data Preparation

Understanding the Data

Configuring Source (https://www.kaggle.com/rodsaldanha/arketing-campaign)



Name	Type	Instance 1	Instance 2	Instance 3
ID	123	5524	2174	4141
Year_Birth	123	1957	1954	1965
Education	ABC	Graduation	Graduation	Graduation
Marital_Status	ABC	Single	Single	Together
Income	123	58138	46344	71613
Kidhome	123	0	1	0
Teenhome	123	0	1	0
Dt_Customer	YYYY-MM-DD	2012-09-04	2014-03-08	2013-08-21
Recency	123	58	38	26
MntWines	123	635	11	426
MntFruits	123	88	1	49
MntMeatProducts	123	546	6	127

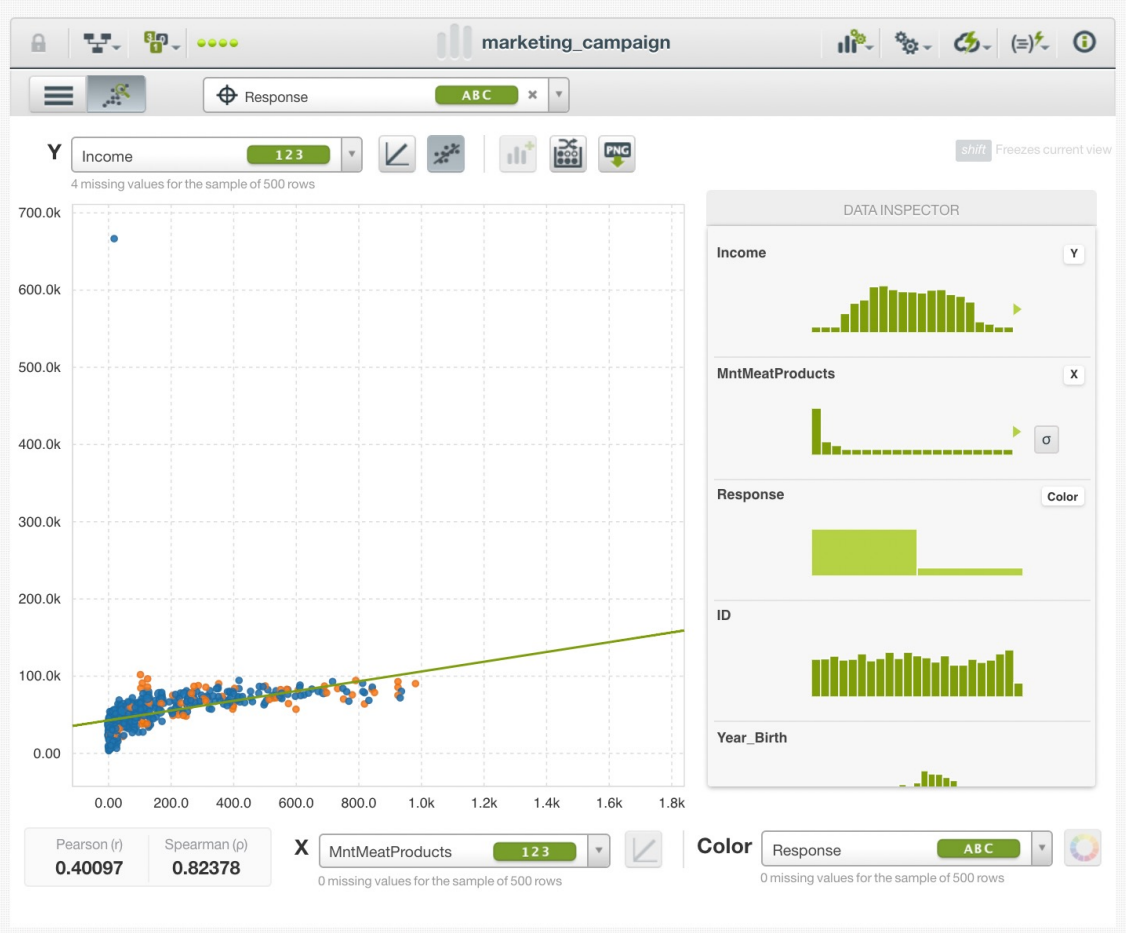
AcceptedCmp4	ABC	0	0	0
AcceptedCmp5	ABC	0	0	0
AcceptedCmp1	ABC	0	0	0
AcceptedCmp2	ABC	0	0	0
Complain	ABC	0	0	0
Z_CostContact	123	3	3	3
Z_Revenue	123	11	11	11
Response	ABC	1	0	0
Dt_Customer.year	YYYY-MM-DD	2012	2014	2013
Dt_Customer.month	YYYY-MM-DD	September	March	August
Dt_Customer.day-of-month	YYYY-MM-DD	4	8	21
Dt_Customer.day-of-week	M T W T F S S	Tuesday	Saturday	Wednesday

Show 100 fields 1 to 33 of 33 fields

Appendix: Data Preparation

Understanding the Data

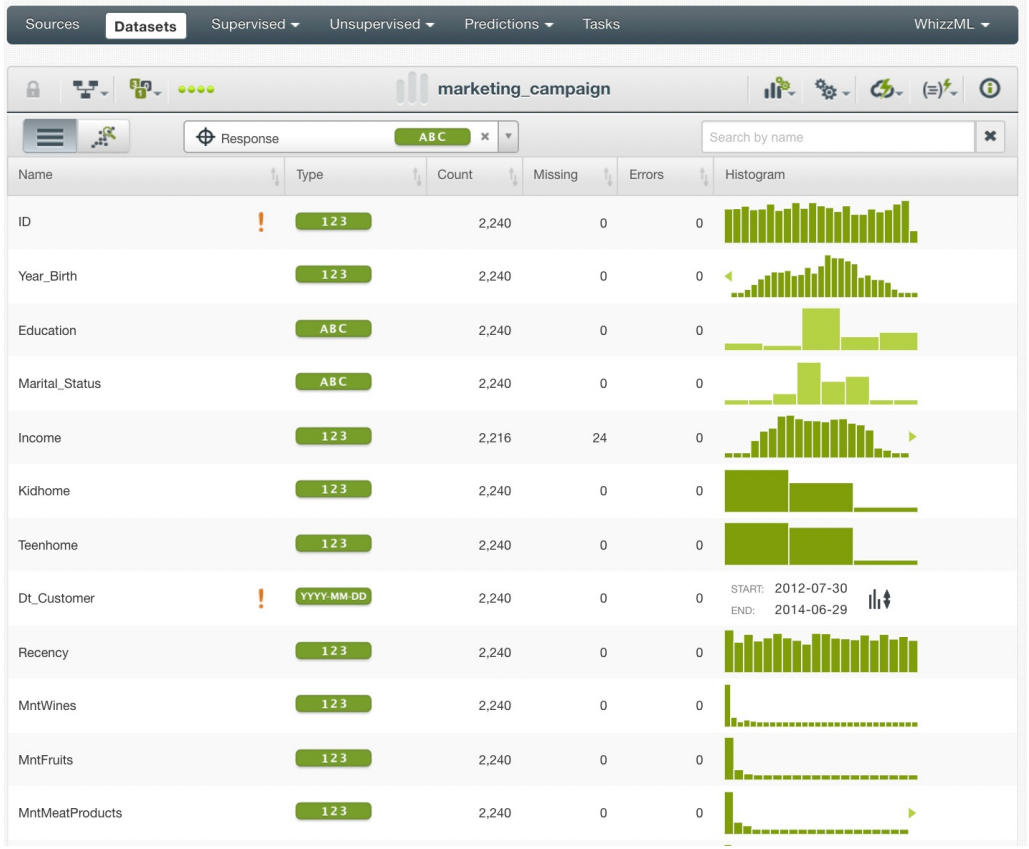
Data Exploration



Appendix: Data Preparation

Understanding the Data

Setting "Response" as the Target Variable



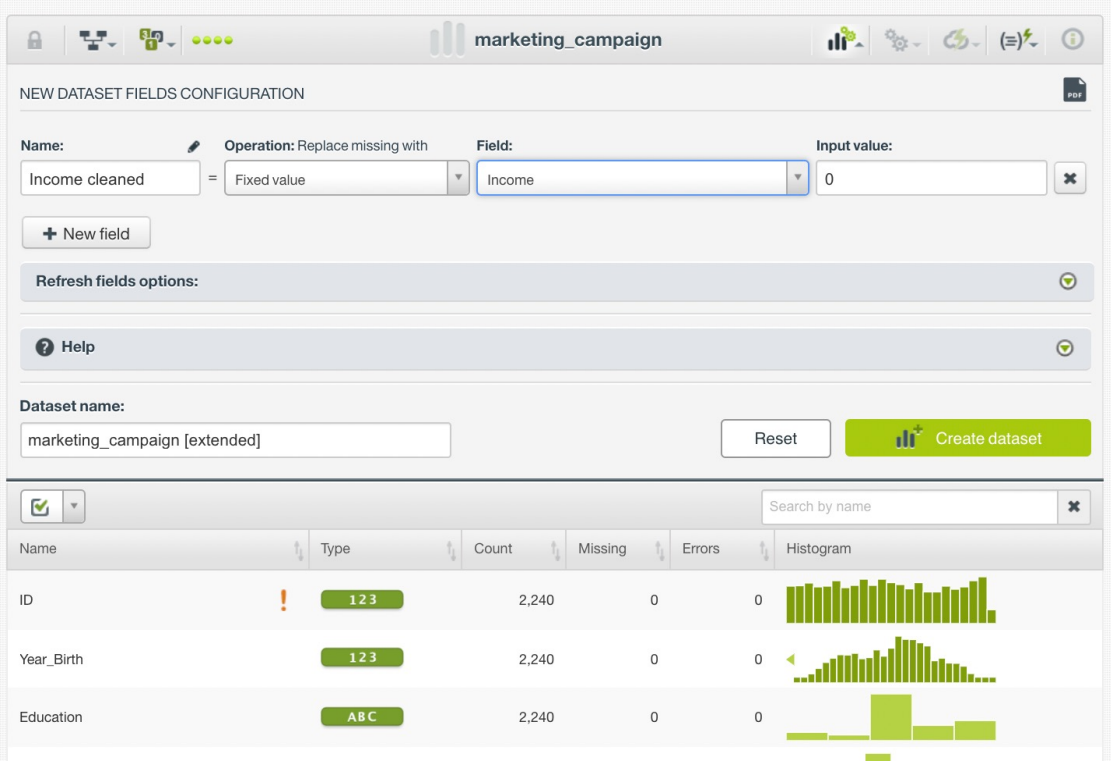
AcceptedCmp4	ABC	0	0	0
AcceptedCmp5	ABC	0	0	0
AcceptedCmp1	ABC	0	0	0
AcceptedCmp2	ABC	0	0	0
Complain	ABC	0	0	0
Z_CostContact	123	3	3	3
Z_Revenue	123	11	11	11
Response	ABC	1	0	0
Dt_Customer.year	YYYY-MM-DD	2012	2014	2013
Dt_Customer.month	YYYY-MM-DD	September	March	August
Dt_Customer.day-of-month	YYYY-MM-DD	4	8	21
Dt_Customer.day-of-week	M T W T F S S	Tuesday	Saturday	Wednesday

Show 100 fields 1 to 33 of 33 fields

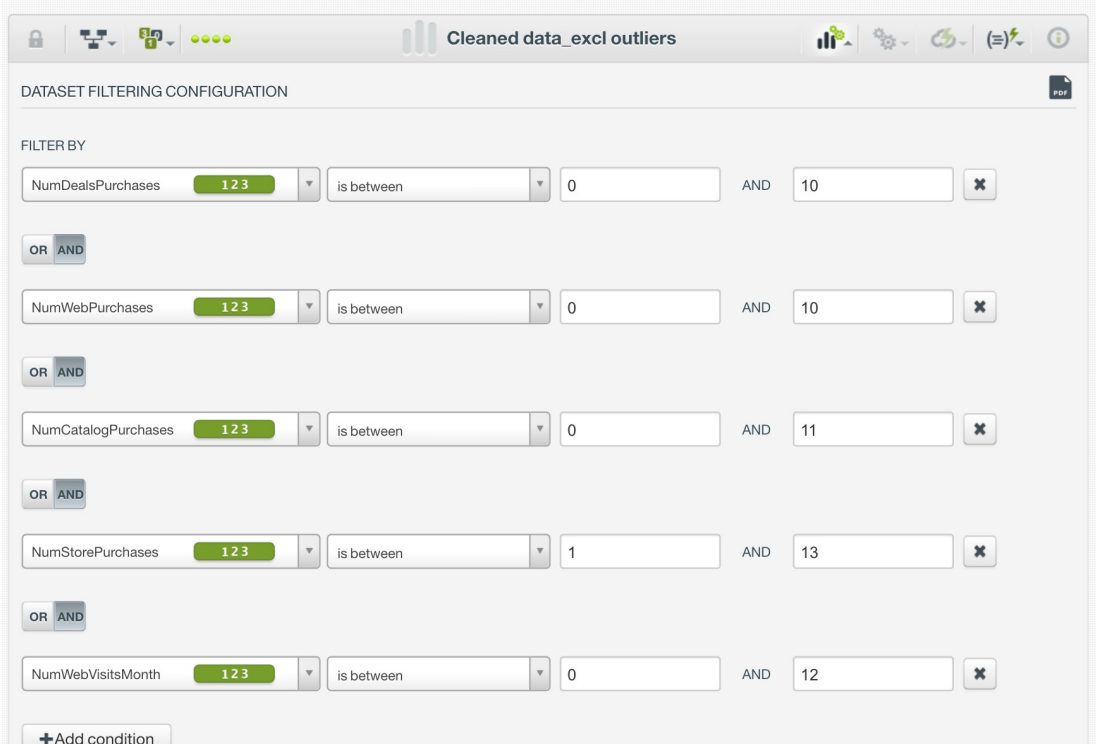
Appendix: Data Cleaning

Missing Values

Missing Values



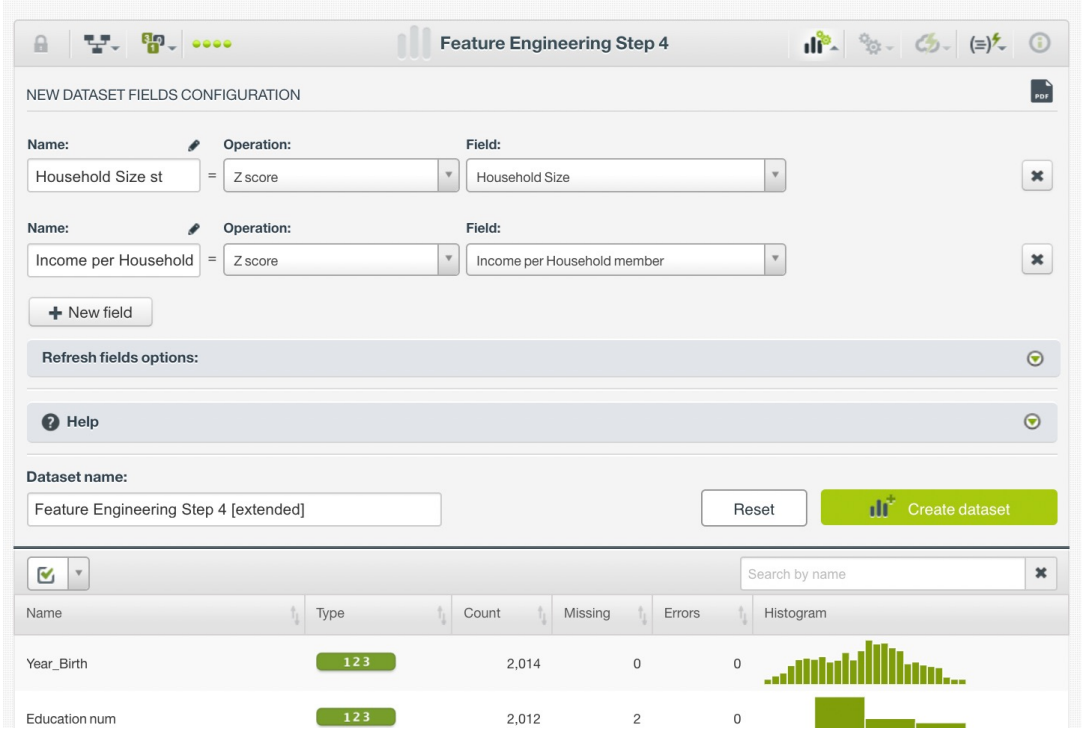
Outliers

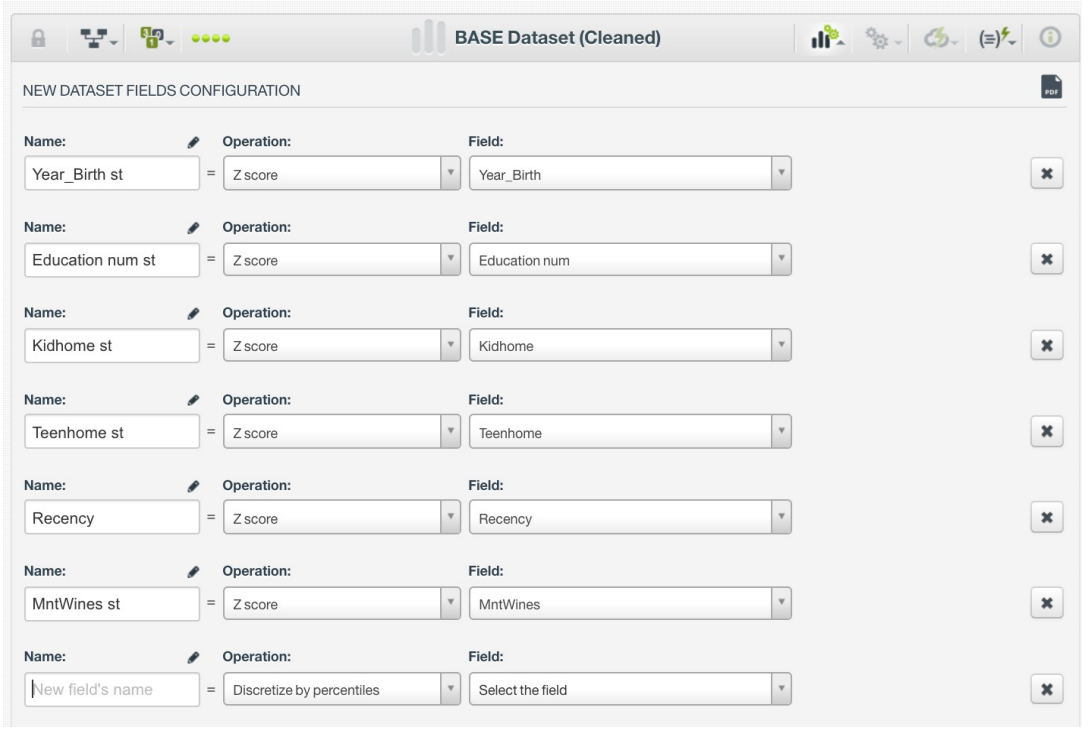


Appendix: Data Cleaning

Using Z-Score to Standardize Data

Z-Score

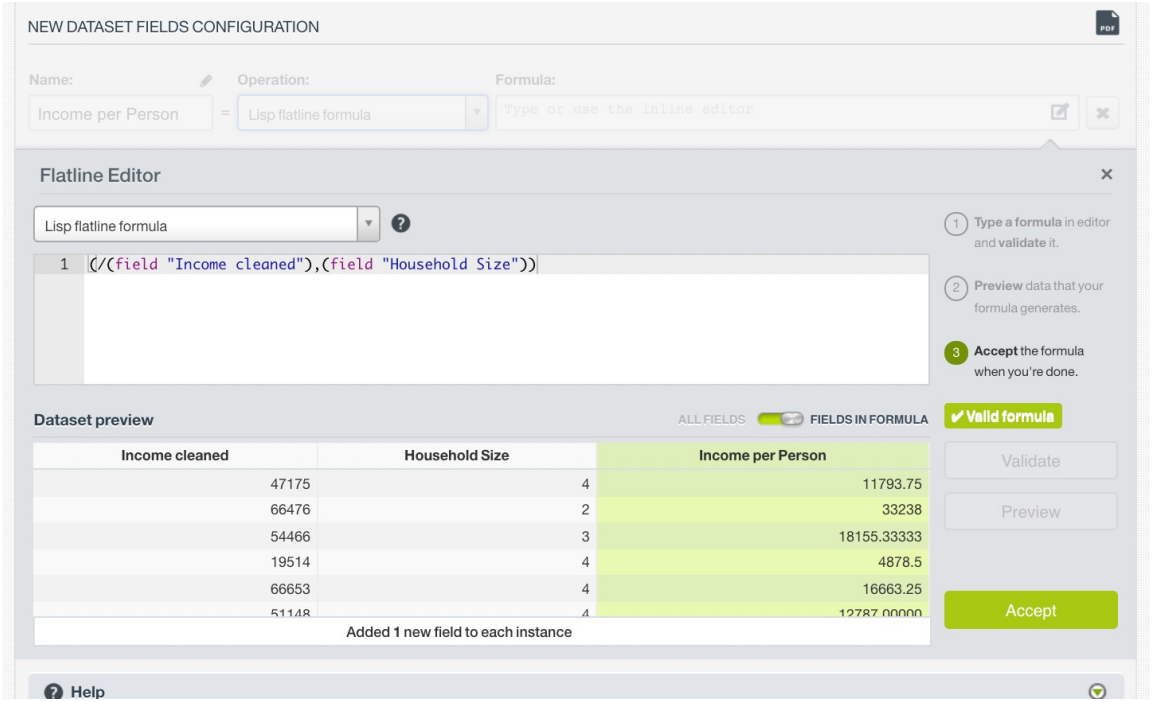


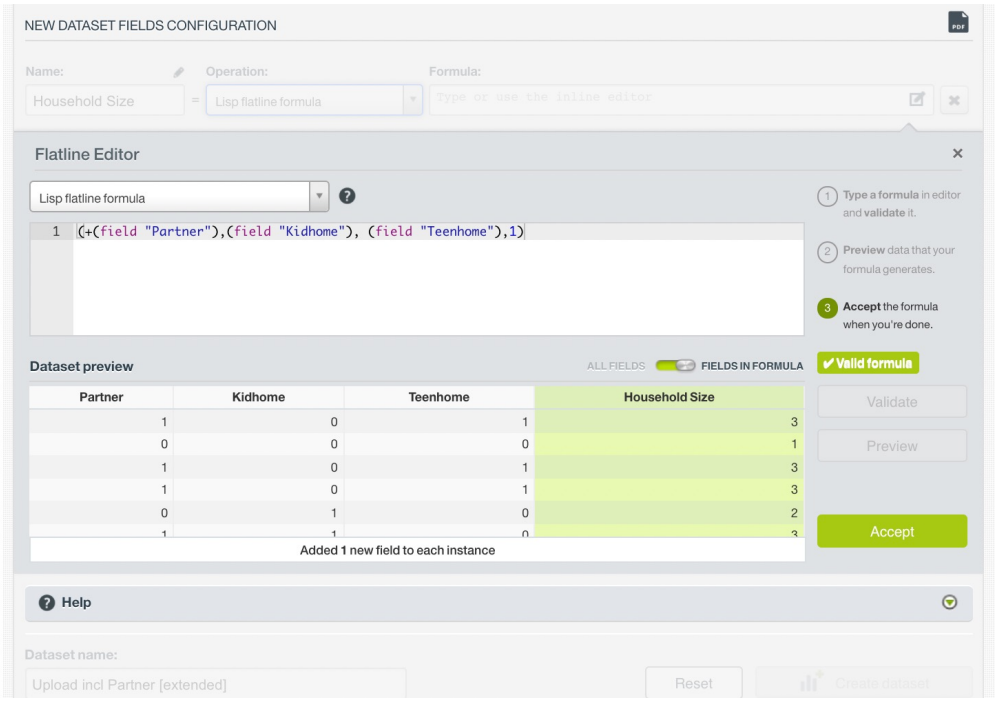


Appendix: Feature Engineering

Feature Engineering on Original Dataset

Lisp-Line Formula

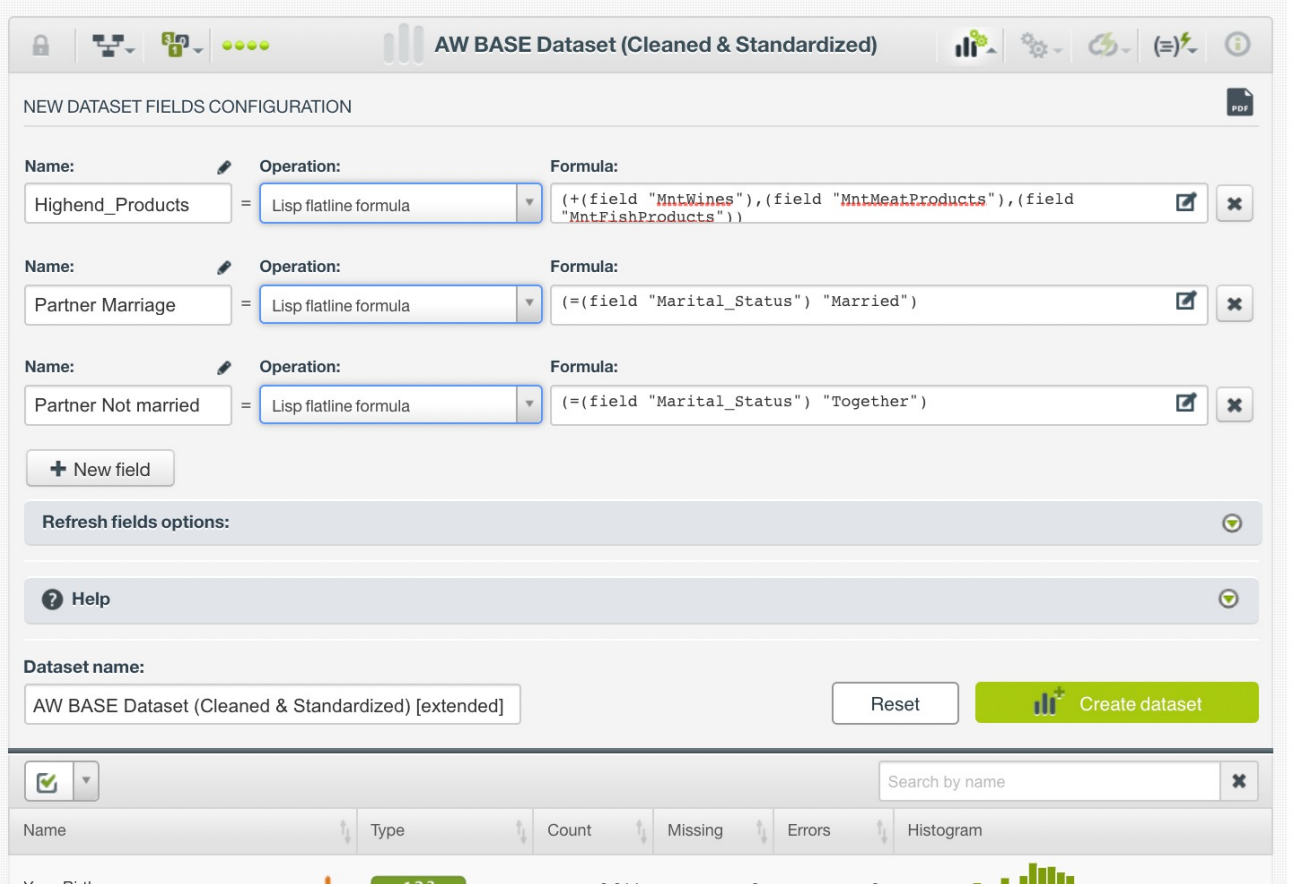




Appendix: Feature Engineering

Feature Engineering on Original Dataset

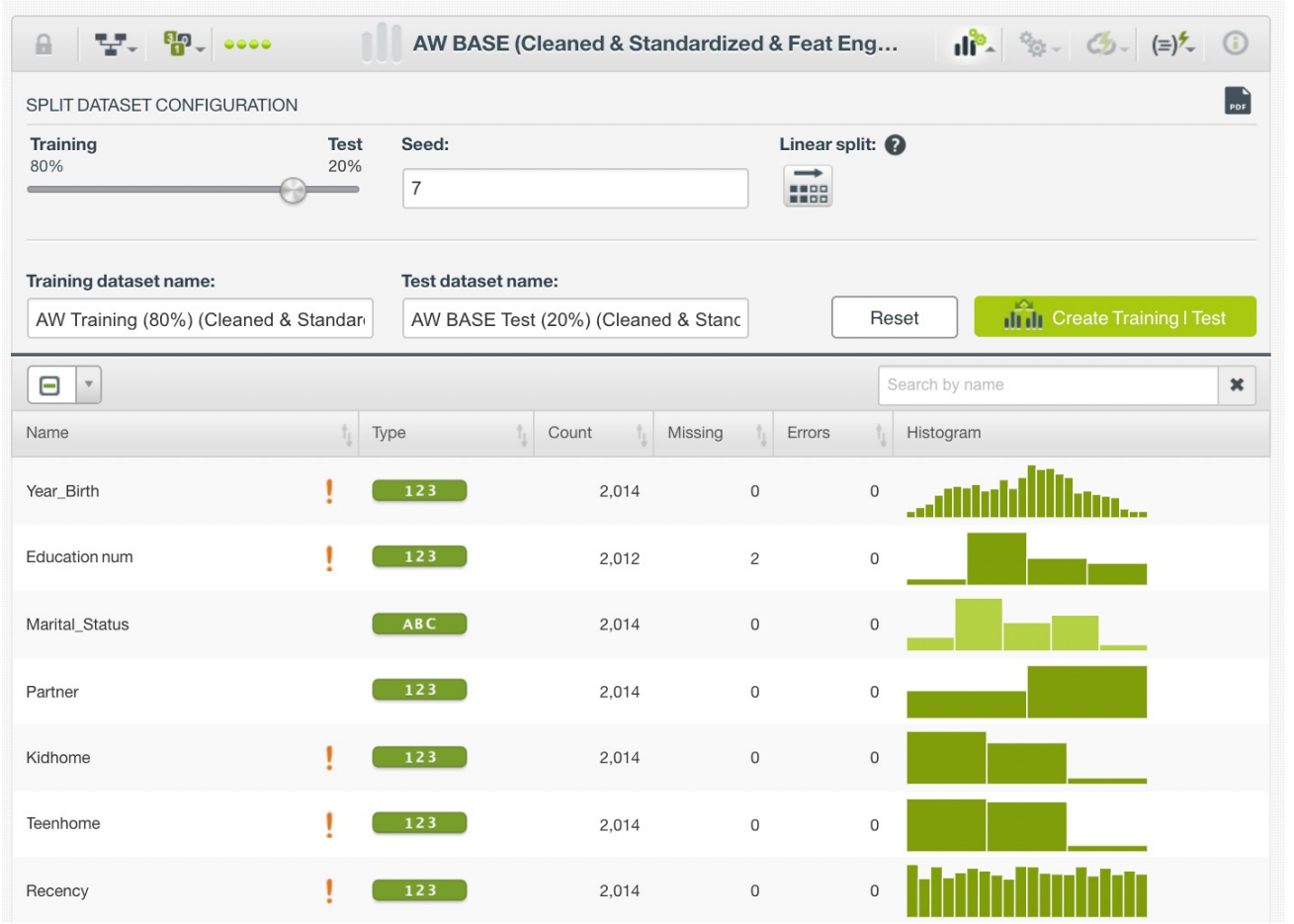
Lisp-Line Formula



Appendix: Training vs Test Split

Performing the 80/20% Split

80/20% Data Split



Appendix: Model Creation

Decision Tree

Initial Decision Tree Evaluation

TP FN FP TN

ACTUAL VS. PREDICTED	0	1	ACTUAL	RECALL	F	Phi
0	326	29	355	91.83%	0.92	0.38
1	25	23	48	47.92%	0.46	0.38
PREDICTED	351	52	403	69.87% AVG. RECALL	0.69 AVG. F	0.38 AVG. Phi
PRECISION	92.88%	44.23%	68.55% AVG. PRECISION	86.60% ACCURACY		

86.6% Accuracy	0.9235 F-measure

92.9% Precision	91.8% Recall	0.3841 Phi coefficient

Appendix: Model Creation

Ensemble

Initial Ensemble Evaluation

Appendix: Model Creation

Logistic Regression

Initial Logistic Regression Evaluation

Appendix: Feature Importance

Field importance of the Best Ensemble Model

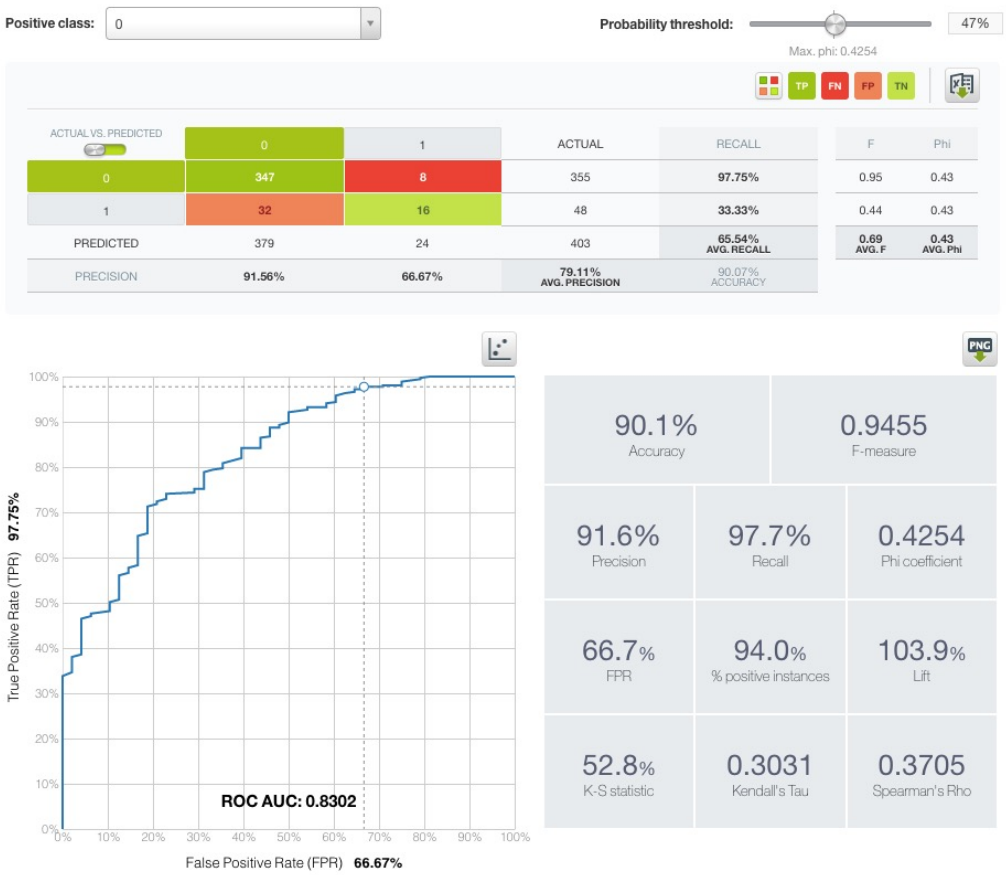
Feature Importance – Ensemble

Field	Feature Importance
Recency st	16.29%
Income per Household member st	12.99%
AcceptedCmp3	12.53%
NumWebVisitsMonth st	9.35%
MntGoldProds st	5.01%
NumDealsPurchases st	3.41%
AcceptedCmp4	3.31%
MntWines st	3.04%
NumCatalogPurchases st	2.84%
Household Size st	2.71%
MntSweetProducts st	2.67%
Income st	2.63%
NumWebPurchases st	2.60%
NumStorePurchases st	2.54%
AcceptedCmp5	2.13%
Year_Birth st	2.03%
MntFruits st	*1.80%*
Highend_Products st	*1.76%*
Education num st	*1.75%*
AcceptedCmp1	*1.63%*
MntMeatProducts st	*1.59%*
MntFishProducts st	*1.53%*
Marital_Status	*1.31%*
Partner	*1.27%*
Kidhome st	*0.74%*
Teenhome st	*0.47%*
AcceptedCmp2	*0.08%*

Appendix: Results of model post Feature Selection

Interpreting the Metrics of the Model

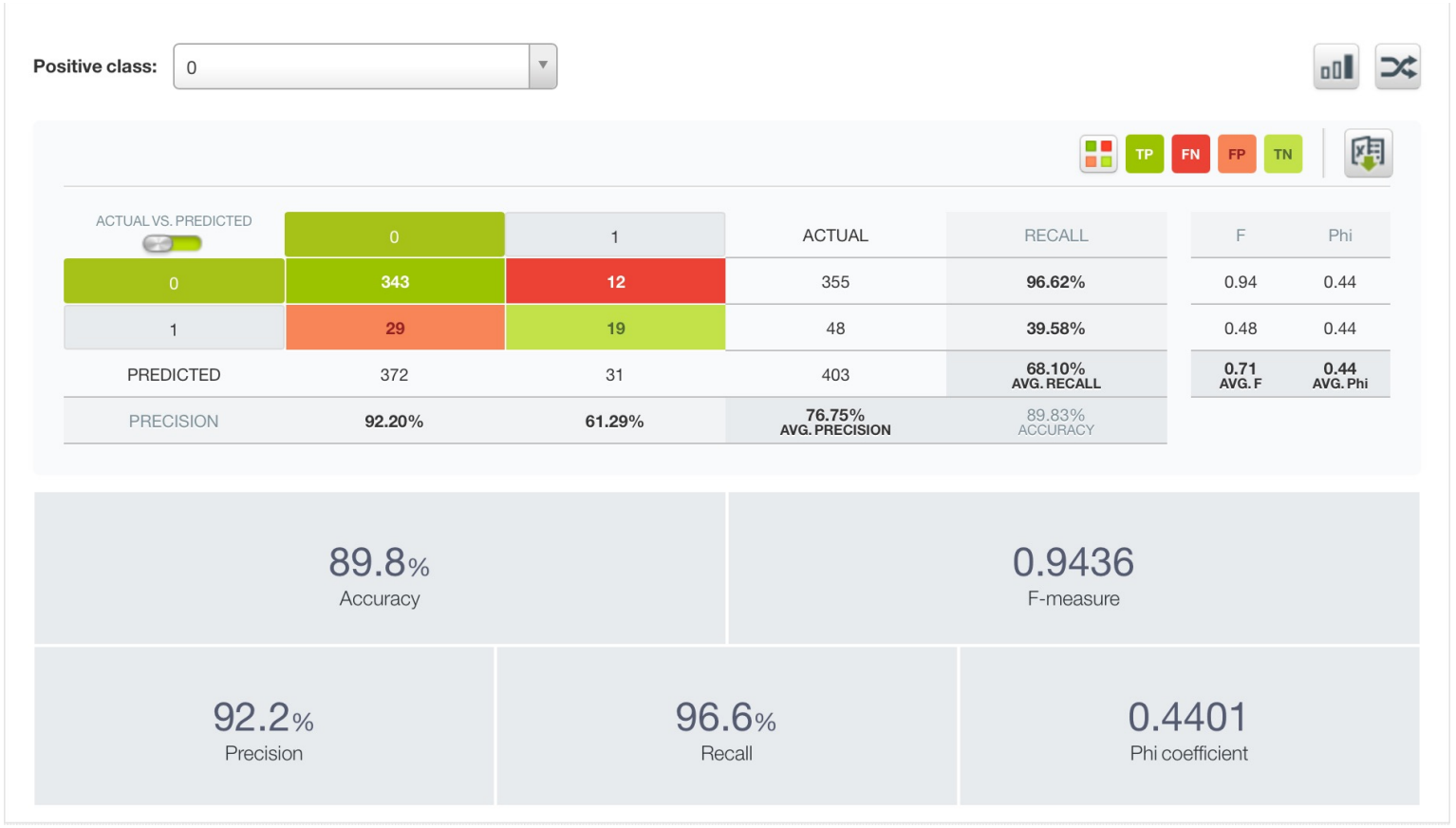
Ensemble model Evaluation



Appendix: Evaluating model against test dataset

Is our Model Overfitted?

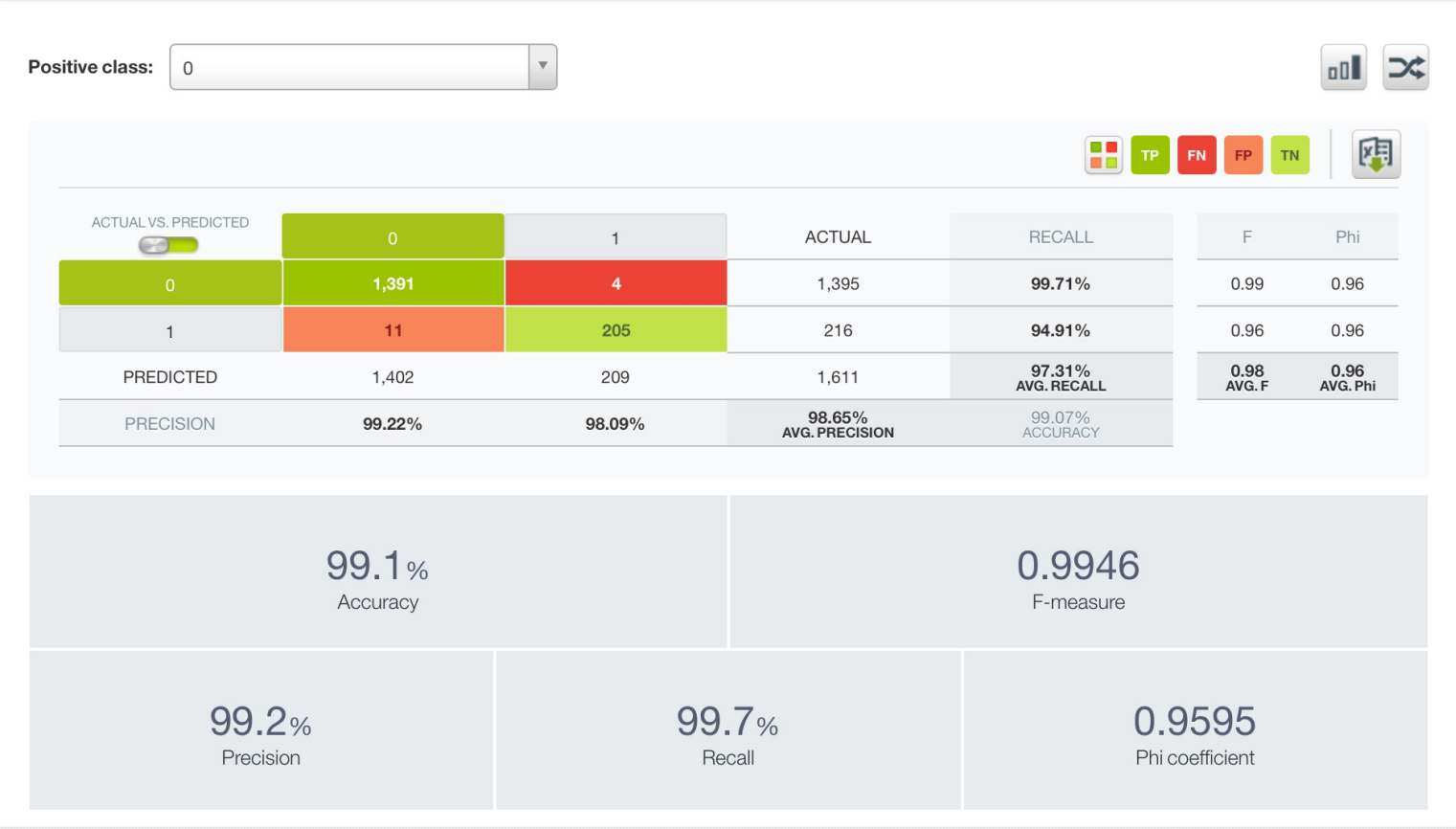
Confusion Matrix of model against test set



Appendix: Evaluating model against training dataset

Is our Model Overfitted?

Confusion Matrix of model against training set



Appendix: Confusion Matrix

How can we quantify our predictions?

Confusion Matrix in Excel

AW BASE (C&S&FE&FS) | Ensemble - Model 11

ACTUAL VS. PREDICTED

Threshold: 46%

True positive	False positive
True negative	False negative

Confusion matrix - Model against test set

Actual \ Predicted	Non-response	Response	ACTUAL	RECALL
Non-response	347	8	355	97.75%
Response	34	14	48	29.17%
PREDICTED	381	22	403	63.46%
PRECISION	91.08%	63.64%	77.36%	89.58%

F	Phi
0.9429	0.3837
0.4000	0.3837
0.6715	0.3837

	Non-response	Response	ACTUAL	RECALL
Non-response	86%	2%	0.8808933	97.75%
Response	8%	3%	0.1191067	29.17%
PREDICTED	0.945409429	0.054590571	1	63.46%
PRECISION	91.08%	63.64%	77.36%	89.58%

All customers distributed along matrix according to predicitons

	Non-response	Response	ACTUAL
Non-response	482,184	11,117	493,300
Response	47,246	19,454	66,700
PREDICTED	529,429	30,571	560,000

Financial Impact per customer (only base marketing campaign)

	Non-response	Response
Non-response	€ -	€ -
Response	€ 45.00	€ 49.00

Financial Impact per customer (base + premium campaign)

	Non-response	Response
Non-response	€ 1.00	€ (6.00)
Response	€ 45.00	€ 49.00

Overall Financil Impact

	Non-response	Response
Non-response	€ 482,184	€ (66,700)
Response	€ 2,126,055	€ 953,251

Net impact € 3,494,789

Check TRUE

Appendix: Financials

What is the financial benefit of our model?

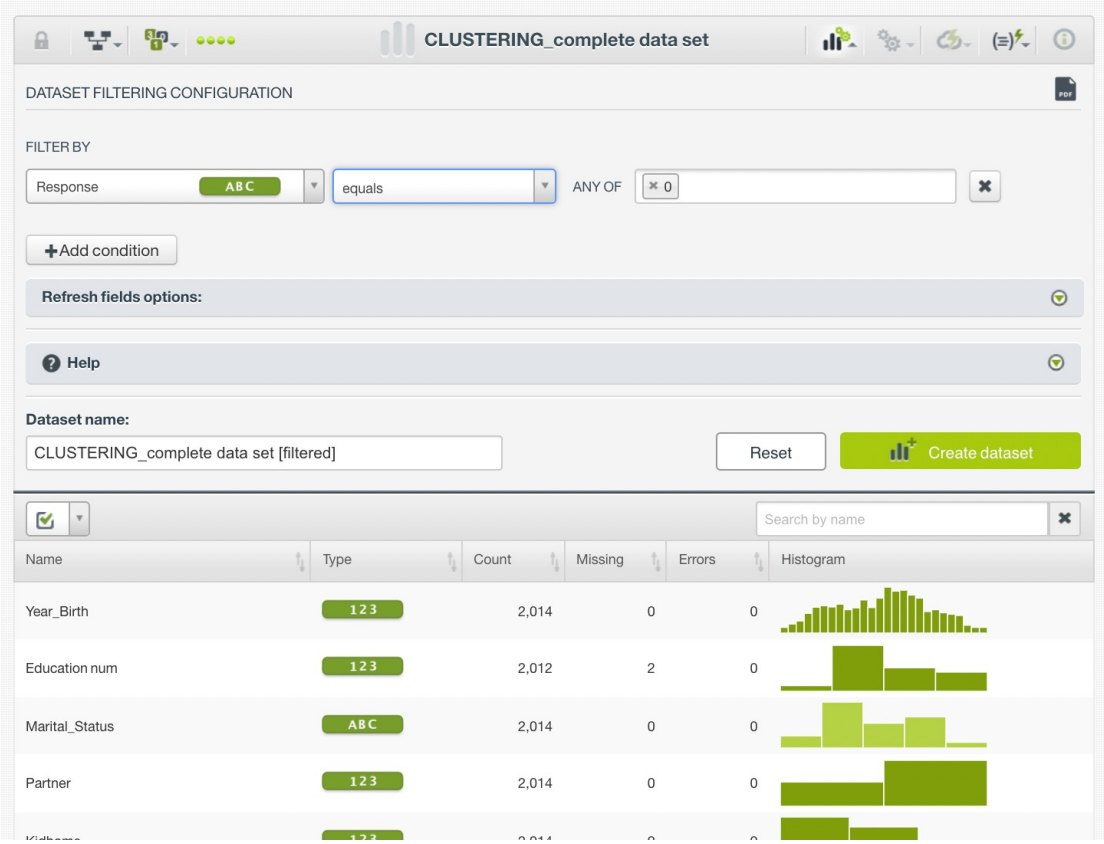
Financials in Excel

18										
19										
20		**Future Scenario - with model predictions**								
21		482,184	*True positive*							
22		19,454	*True negative*							
23		47,246	*False positive*							
24		11,117	*False negative*							
25	€	55.00	*Revenue per customer*							
26	€	6.00	*Marketing cost per customer (base)*							
27	€	10.00	*Marketing cost per customer (premium)*							
28		100%	*Success rate of base campaign*							
29		20%	*Success rate of premium marketing campaign*							
30										
31										
32		**Revenue**								
33		Revenue from True Positives				€	5,304,019.85			
34		Revenue from True Negatives				€	1,069,975.19			
35		Revenue from False Positives				€	2,598,511.17			
36										
37		**Costs**								
38		Base Marketing Campaign Total Costs		€	(183,424.32)					
39		Premium Marketing Campaign Total Costs		€	(5,294,292.80)					
40										
41		**Net Impact - Future scenario**				€	3,494,789.08			
42										
43										
44		**Profit change**				**€**	**2,850,789.08**			
45		*in %*					**81.57%**			
46										
47										
48										
49										

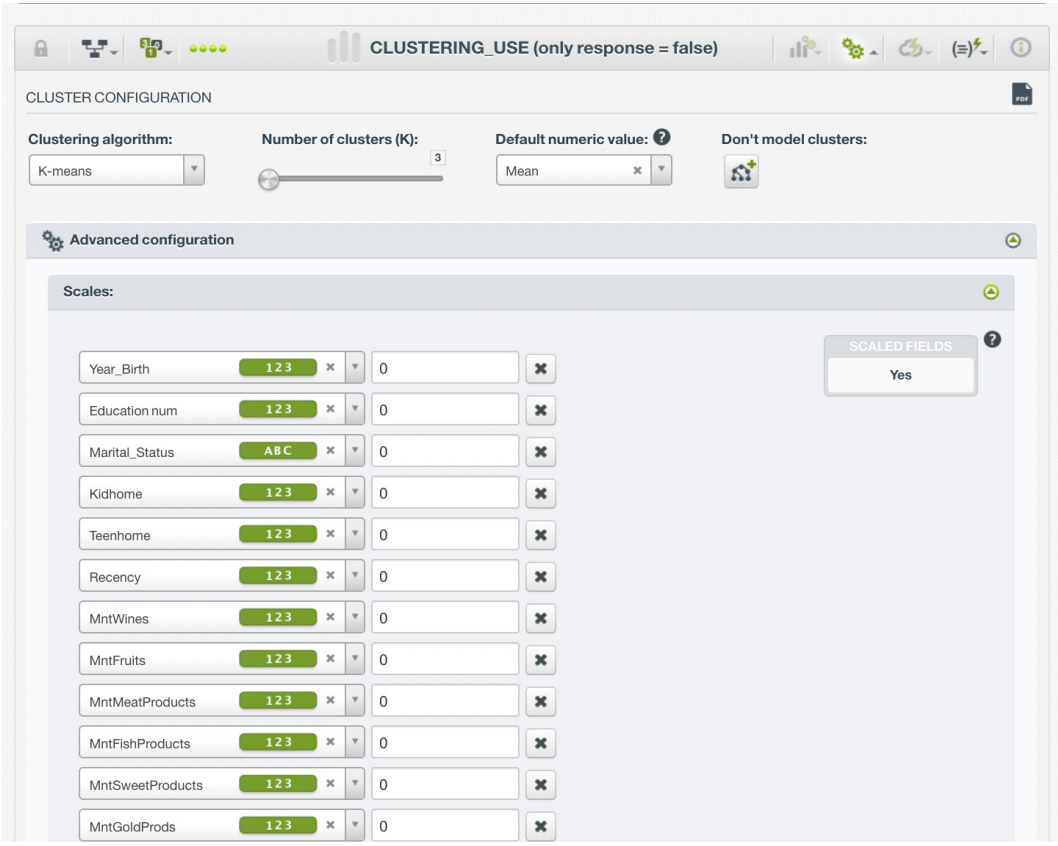
Appendix: Clustering

Modifying the Dataset for Clustering

Excluding Response = True



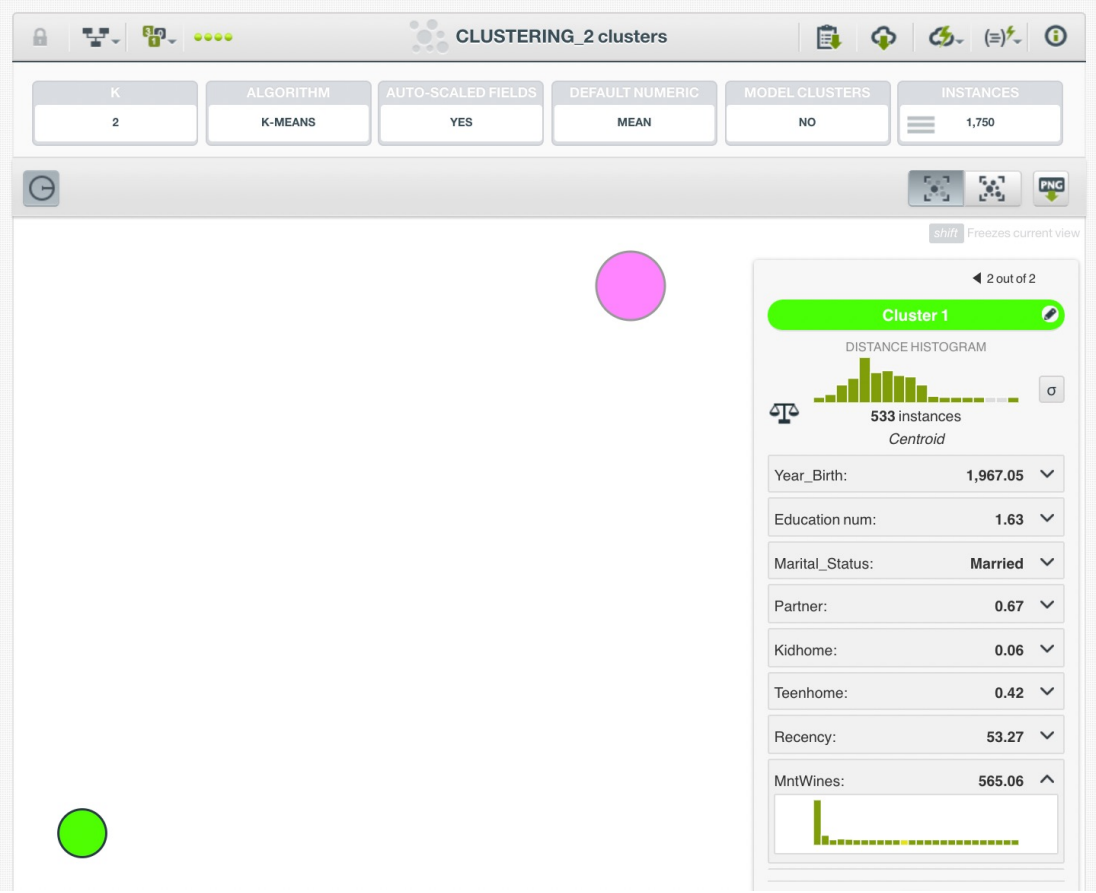
Including non-standardized feature for Interpretation (with 0 scaling)

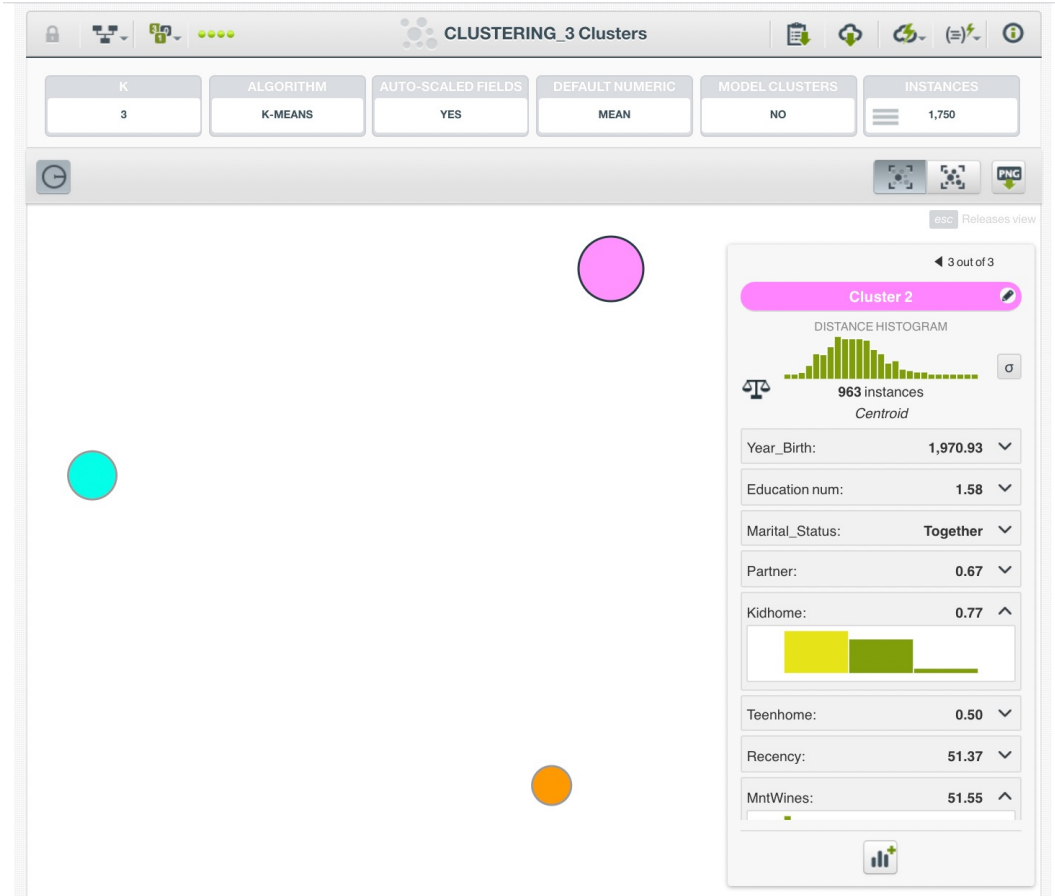


Appendix: Clustering

K = 3

K = 3

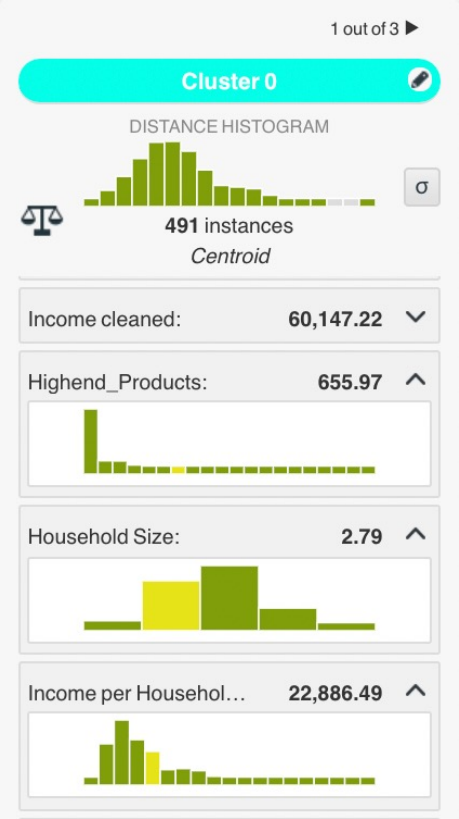




Appendix: Analyzing Clusters

Main differences & Business Meaning of Clusters

Details Cluster 0



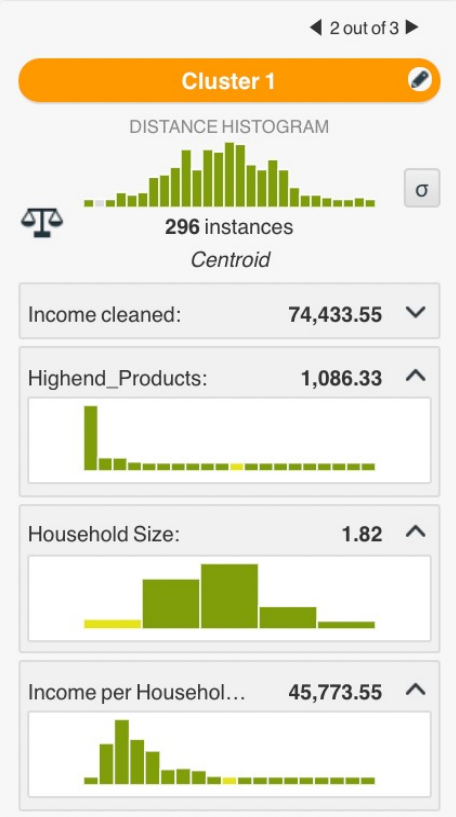
- Kidhome: Low (0.17); Teenhome: Medium (0.91)
- Number of Deals Purchases: High (3.22)
- Number of Catalog Purchases: Medium/ High (3.35)
- Number of Store Purchases: High (8.33)
- Highend_Products: Medium/ High (655.97)
- Household Size: Low/ Medium (2.79)
- Income per Household: Medium/ high (22886.49)

Summary: Small to medium households with a medium/ high income per person and a medium/ high number of purchases.

Appendix: Analyzing Clusters

Main differences & Business Meaning of Clusters

Details Cluster 1



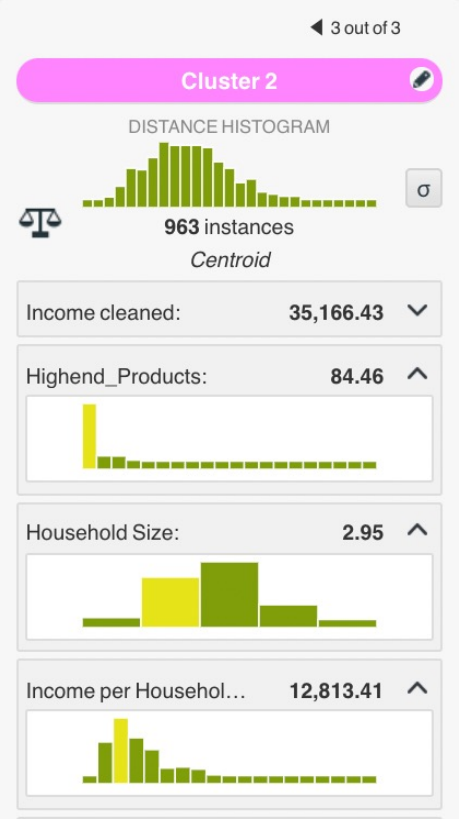
- Kidhome: Low (0.06), Teenhome: Low (0.13)
- Number of Deals Purchases: Low (1.16)
- Number of Catalog Purchases: High (5.81)
- Number of Store Purchases: High (8.70)
- Highend_Products: High (1086.33)
- Household Size: Low (1.82)
- Income per Household: Very high (45773.55)

Summary: Very small households with a very high income per person, a high number of purchases and a low amount of deal purchases.

Appendix: Analyzing Clusters

Main differences & Business Meaning of Clusters

Details Cluster 2



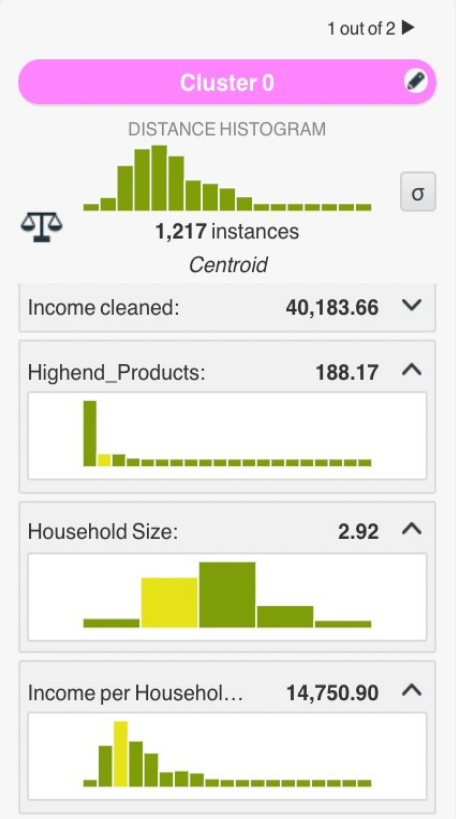
- Kidhome: Medium (0.77), Teenhome: Medium (0.5)
- Number of Deals Purchases: Medium (2.13)
- Number of Catalog Purchases: Low (0.57)
- Number of Store Purchases: Medium (3.42)
- Highend_Products: Low (84.46)
- Household Size: Medium (2.95)
- Income per Household: Low (12813.41)

Summary: Medium-sized households with a low income per person, a medium number of purchases and little high-end purchases.

Appendix: Analyzing Clusters

Main differences & Business Meaning of Clusters

Details Cluster 0



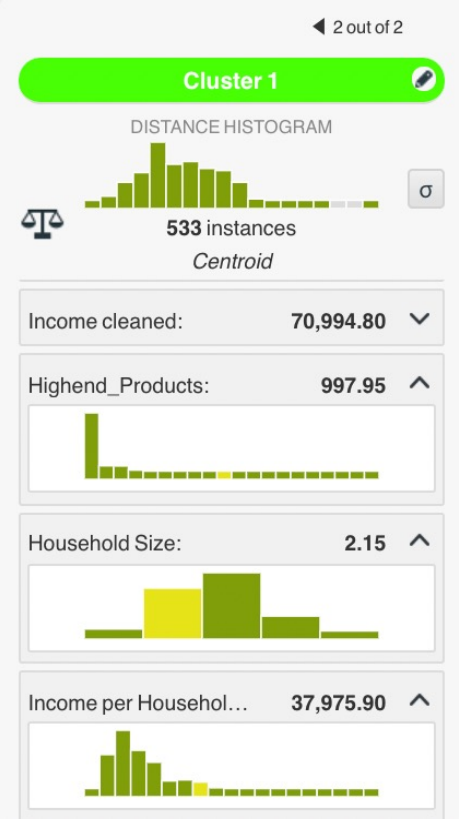
- Kidhome: Medium/ High (0.64)
- Number of Deals Purchases: Medium/ High (2.42)
- Number of Catalog Purchases: Low/ Medium (1.06)
- Number of Store Purchases: Low/ Medium (4.37)
- Highend_Products: Low/ Medium (188.17)
- Household Size: Low/ Medium (2.92)
- Income per Household: Low (14750.90)

Summary: Medium-sized households with a comparably low income per person and a preference for special deals.

Appendix: Analyzing Clusters

Main differences & Business Meaning of Clusters

Details Cluster 1



- Kidhome: Low (0.06)
- Number of Deals Purchases: Low (1.82)
- Number of Catalog Purchases: High (5.28)
- Number of Store Purchases: High (9.06)
- Highend_Products: High (997.95)
- Household Size: Low (2.15)
- Income per Household: High (37975.90)

Summary: Small Households with a high income per person, a high number of purchases and a low amount of deal purchases.