



iFood CRM Data Analyst Case

By Israel Mendes

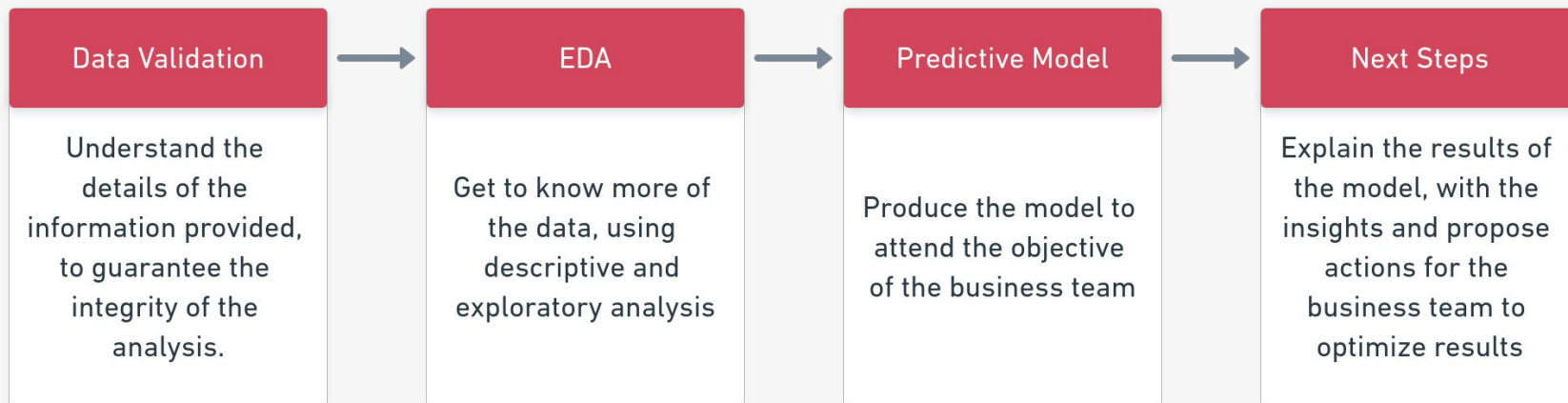
Summary

1. **Process**
2. Main Results
3. Explanation of the Model
4. Data Analysis in Resume
5. Conclusion



Process

How was it done?



Process

What was used?

- **Tools:**

- a. Databricks;
- b. Tableau;
- c. Github.

- **Details in the Process:**

- a. [Repo;](#)
- b. [Case;](#)
- c. [Dashboard;](#)
- d. [Exploratory Data Analysis;](#)
- e. [Predictive Model.](#)

Summary

1. Process
- 2. Main Results**
3. Explanation of the Model
4. Data Analysis in Resume
5. Conclusion



Premises

When evaluating the sample results, is possible to conclude a few metrics that will determine our success:

- **Campaign Rate:** the standard (from the sample) was 15%;
- **Cost per customer:** 3.03MU;*;
- **Return per customer:** 11.03MU;*;
- **R.O.I.:** -45.32%;
- **Goal campaign rate:** for a campaign to be profitable, it needs to have a higher campaign rate of +27.47%

* => Considering customers with not null values or neither outliers.

Process

Calculations with estimates

- **Model:** results here:
 - a. **Process:** using machine learning and statistics to improve the metrics;
 - b. **Campaign Rate:** 100%;
 - c. **R.O.I.:** +264,02%;
 - d. **Total return:** 1.318MU;
 - e. **Total investment:** 362MU;
 - f. **Savings:** up to 6.358MU (94.61% of the budget).
- **Insights:**
 - a. **Process:** use statistics and data analysis to produce better segmentations
 - b. **Campaign Rate:** 46,99%;
 - c. **R.O.I.:** +171.072%;
 - d. **Total return:** 1.897MU;
 - e. **Total investment:** 1.108MU;
 - f. **Savings:** up to 5.612MU (83.51% of the budget).

Summary

1. Process
2. Main Results
- 3. Explanation of the Model**
4. Data Analysis in Resume
5. Conclusion



Essencial

Model

The model is based on three steps:

1. Getting the data and guarantee the information is correct, concise, and not nullable;
2. Using the main metrics to predict if the customer will react to the campaign or not;
3. The **result will be the ID of the user + the Prediction** (with 1 if it's likely the customer purchase from the campaign and 0 if not likely).

Metrics

Model

By using statistics, was possible to find a correlation between the Response to react on the campaign vs the metrics. Here are the main metrics:

- **Positive Correlation:**

- a. *Accepted Campaign 5: +32%*
- b. *Accepted Campaign 1: +30%*
- c. *Amount Wine: +25%*
- d. *Accepted Campaign 3: +25%*
- e. *Amount Meat Products: +24%*
- f. *Amount Catalog Purchases: +22%*

- **Negative Correlation:**

- a. *Recency: -20%*
- b. *Teen in Home: -15%*
- c. *Kid in Home: -7.8%*

Summary

1. Process
2. Main Results
3. Explanation of the Model
- 4. Data Analysis in Resume**
5. Conclusion



Exploration

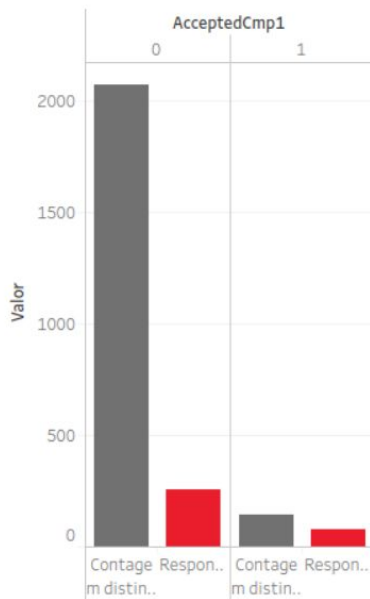
Analysis

After defining the main metrics using statistics, the exploration of each metric will help to understand better the data and provide more accurate insights.

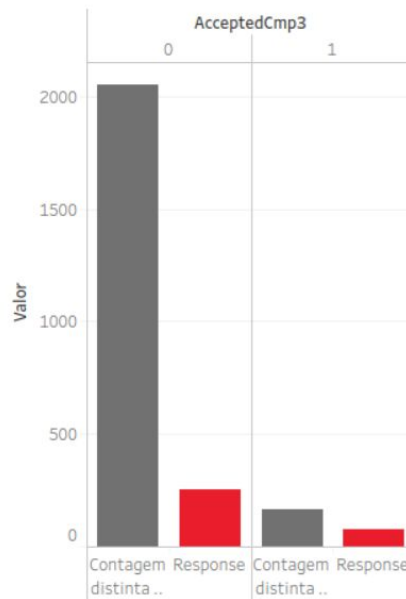
Accepted Campaign 1, 3 and 5

Analysis

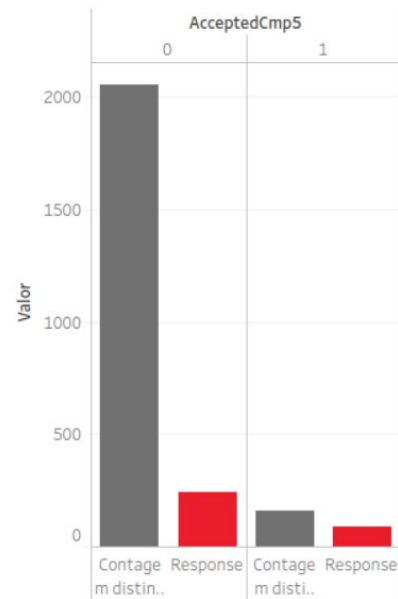
Accepted Campaign 1



Accepted Campaign 3

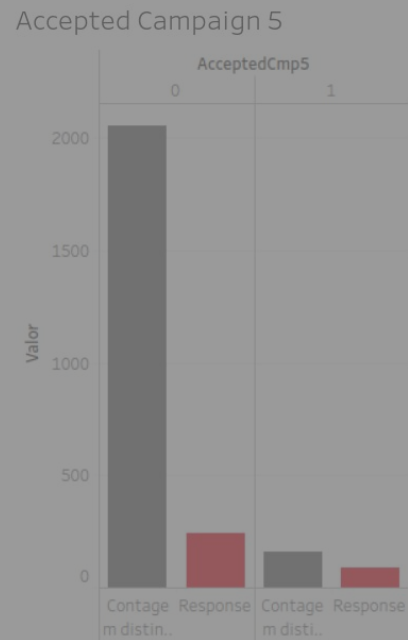
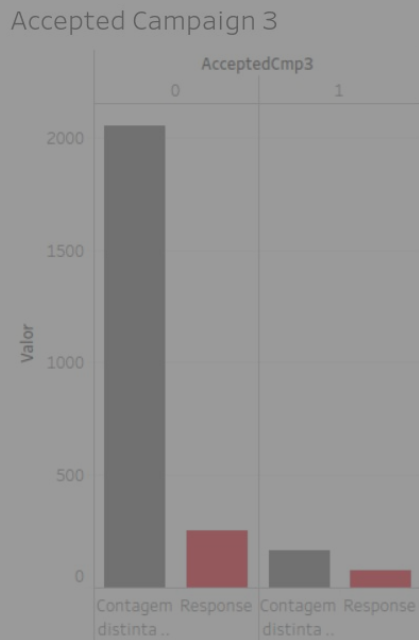
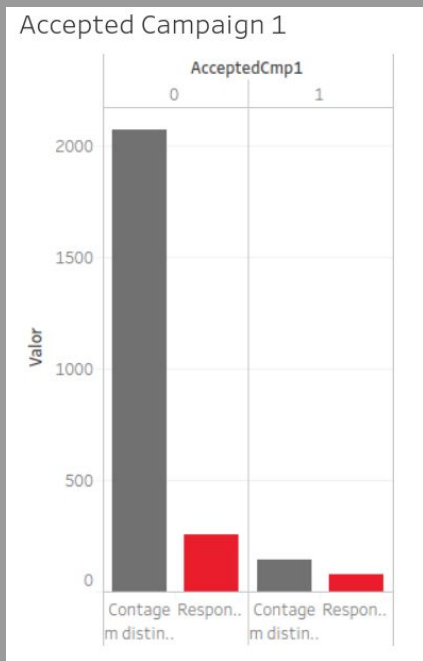


Accepted Campaign 5



Those that accepted campaign
1 vs. also reacted positively
with the Response represent
55.63% off the whole campaign
1.

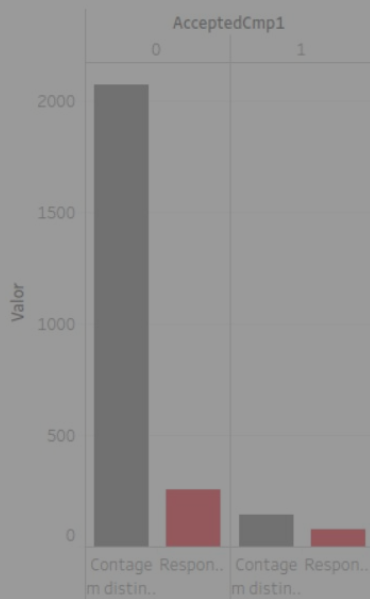
ampaign 1, 3 and 5



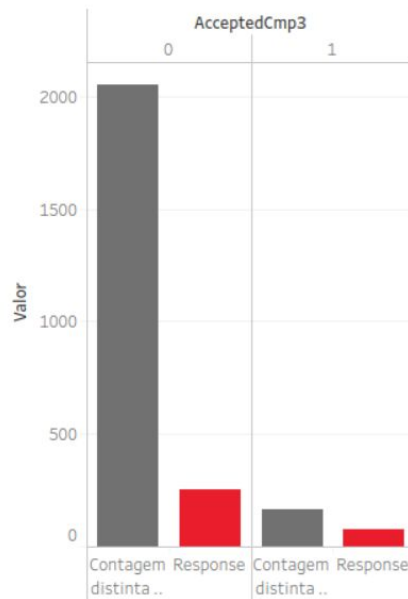
Accepted Campaign 3 and 5 Analysis

Those that accepted campaign 3 vs. also reacted positively with the Response represent **47.23%** off the whole campaign 3.

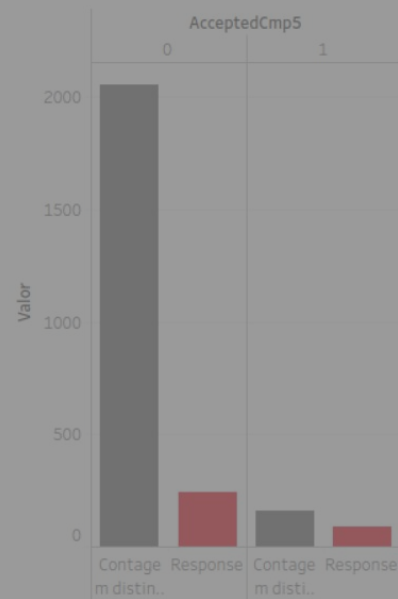
Accepted Campaign 1



Accepted Campaign 3



Accepted Campaign 5

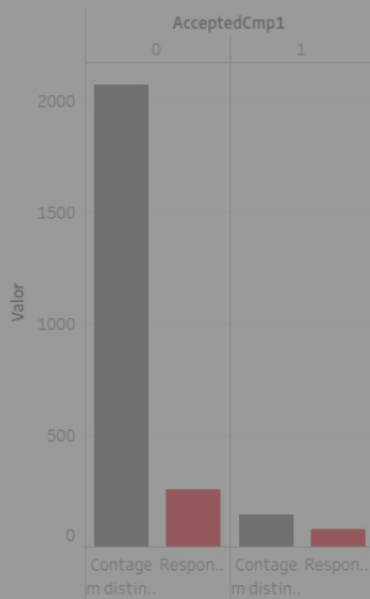


Accepted Campaign 1

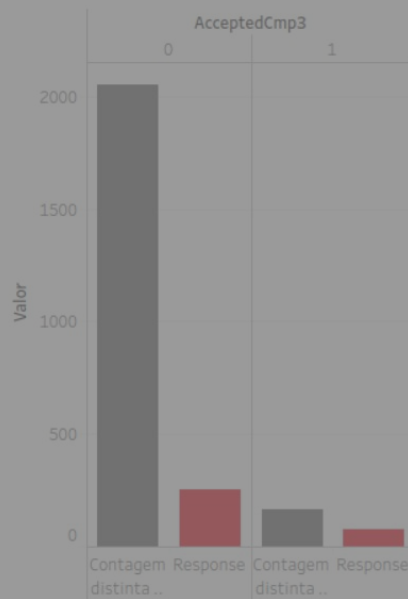
Analysis

Those that accepted campaign 5 vs. also reacted positively with the Response represent **56.52%** off the whole campaign 5.

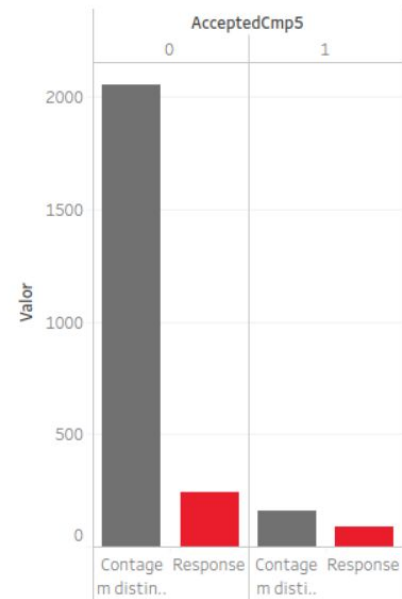
Accepted Campaign 1



Accepted Campaign 3



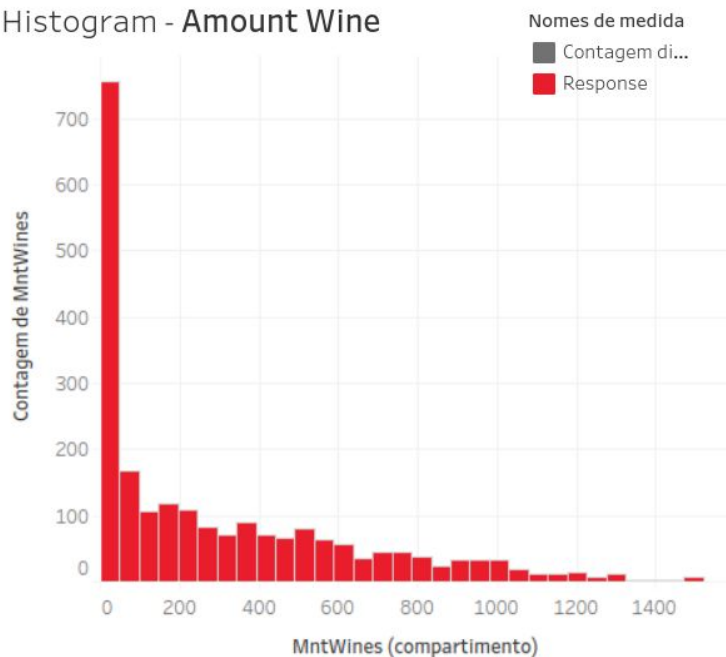
Accepted Campaign 5



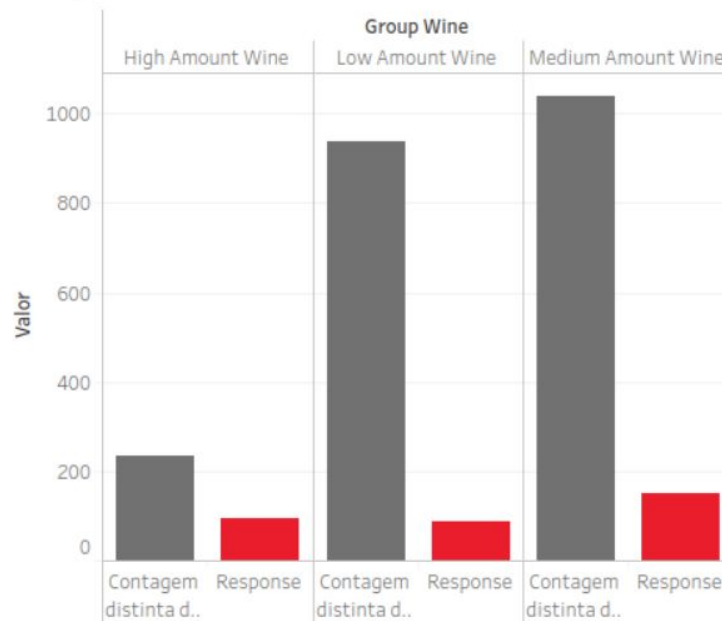
Amount Wine

Analysis

Histogram - Amount Wine

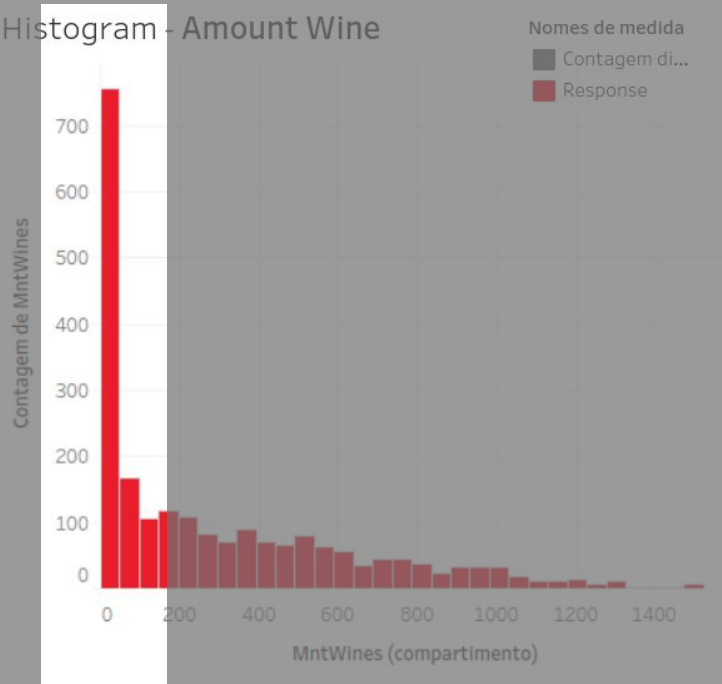


Group - Amount Wine

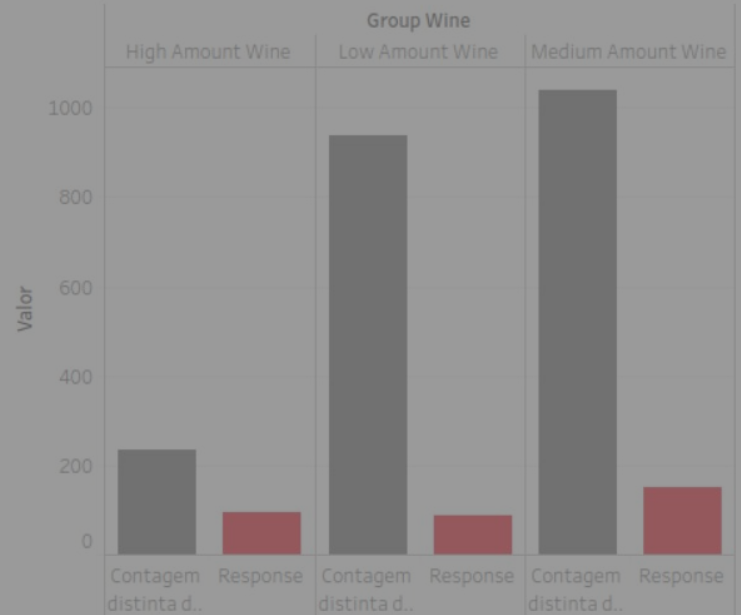


Is possible to notice
the main group is
Low Amount Wine
(0 to 100).

Histogram - Amount Wine



Group - Amount Wine

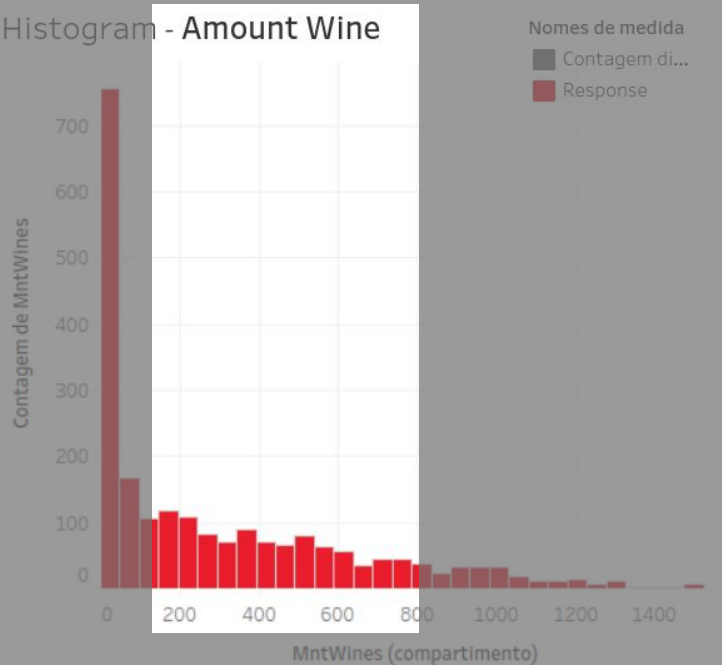


Amo e

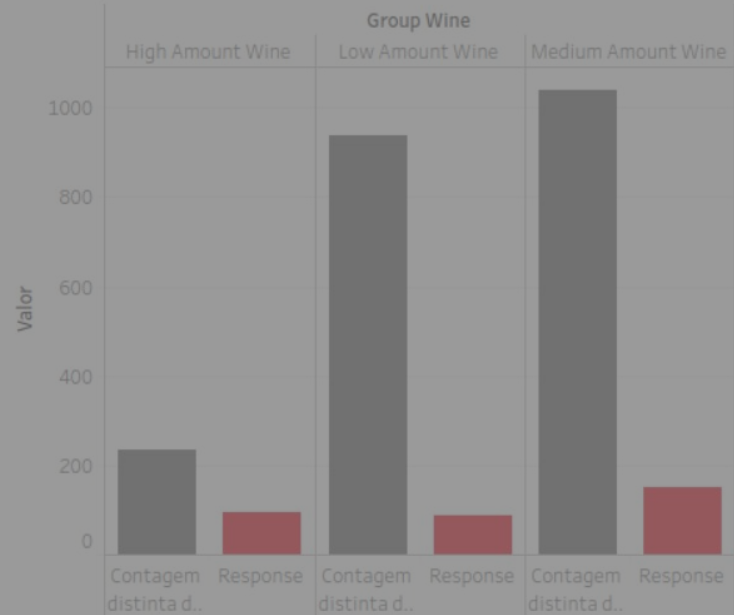
Analysis

Is possible to notice
the main group is
**Medium Amount
Wine** (101 to 800).

Histogram - Amount Wine



Group - Amount Wine

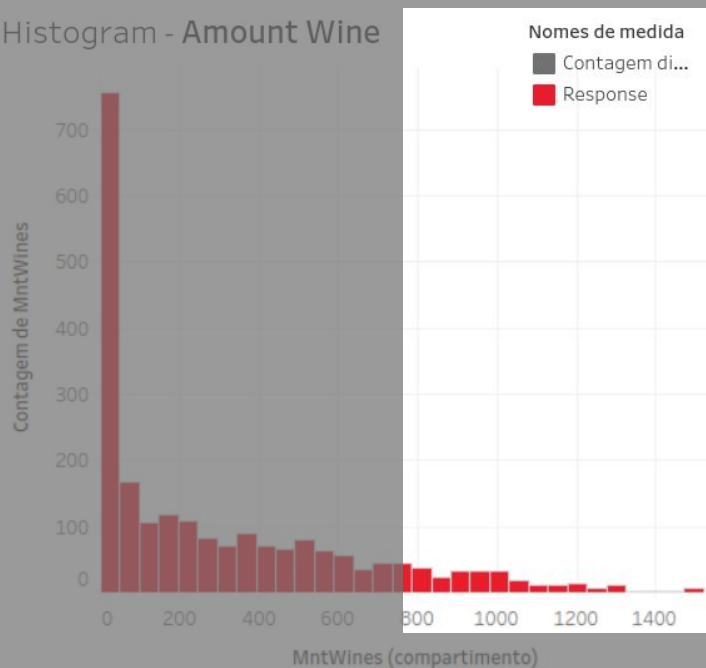


Amount Wine

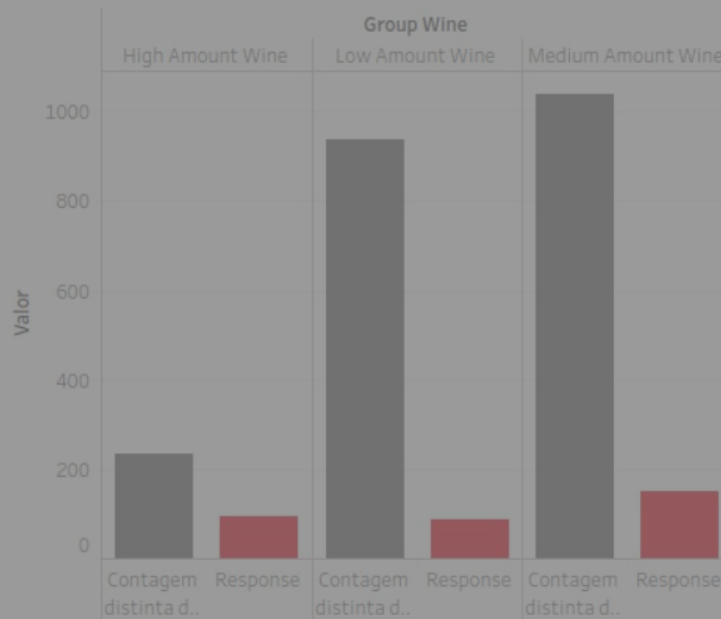
Analysis

Is possible to notice
the main group is
High Amount Wine
(+801).

Histogram - Amount Wine



Group - Amount Wine

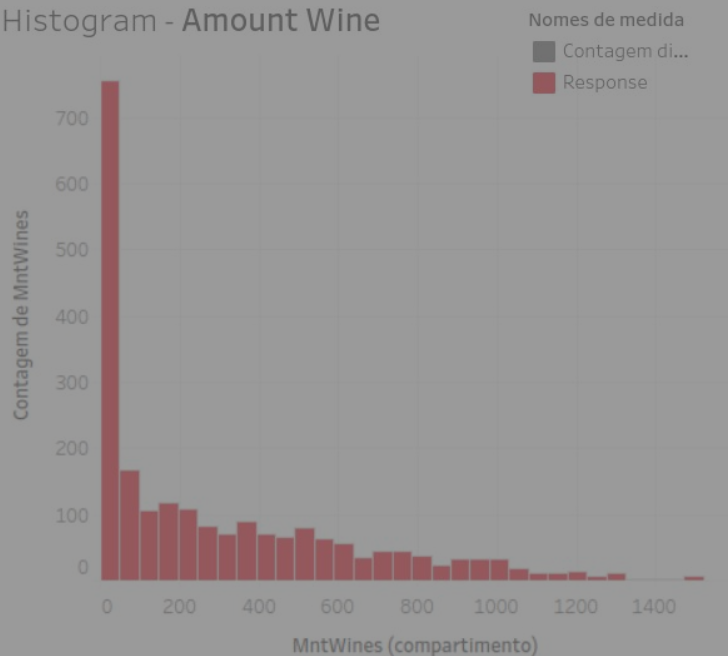


Amount Wine

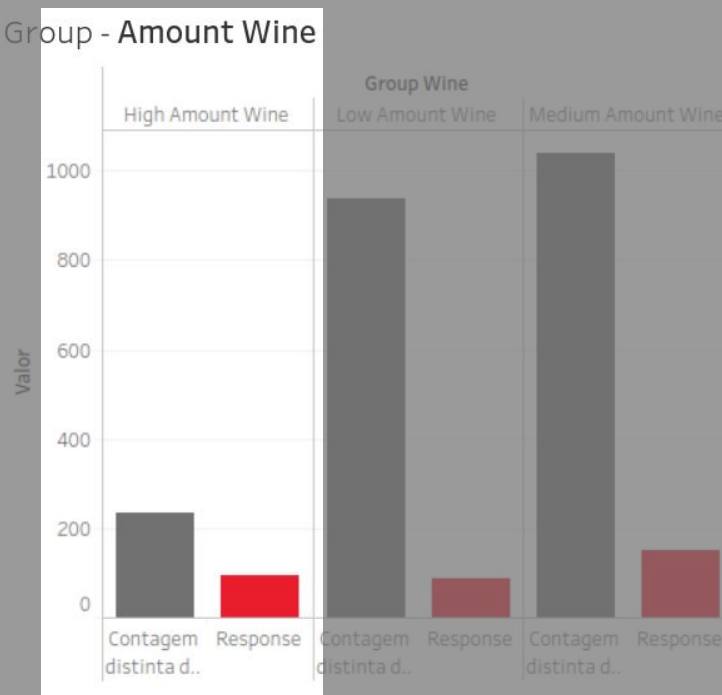
Analysis

High Amount Wine (+801):
concentrates **10.66% of the users**,
but concentrates **28.82% of the**
responses

Histogram - Amount Wine



Group - Amount Wine

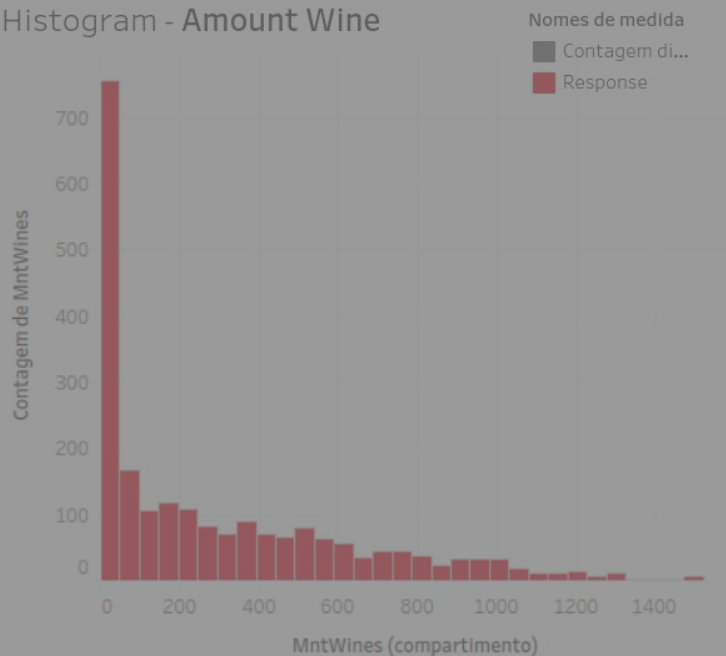


Amount Wine

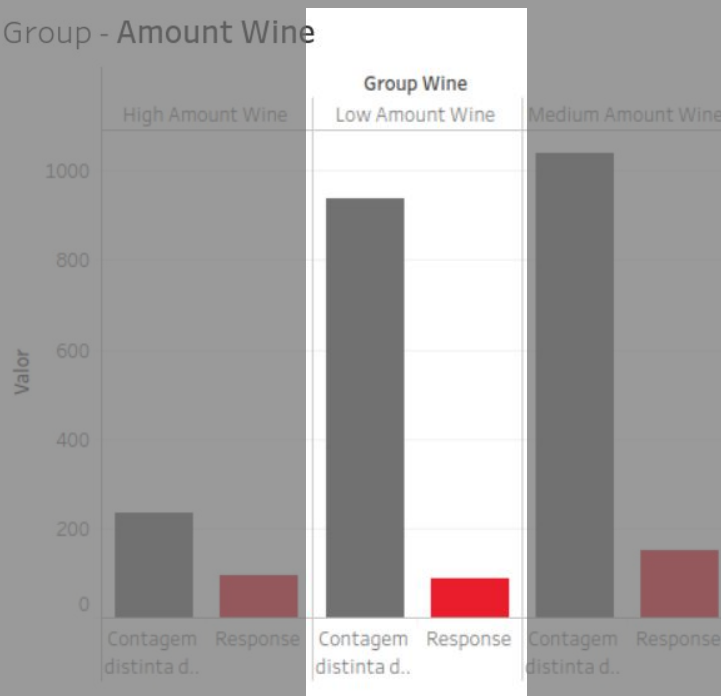
Analysis

Low Amount Wine (0 to 100):
concentrates **42.34% of the users**,
but concentrates **26.12% of the**
responses

Histogram - Amount Wine



Group - Amount Wine

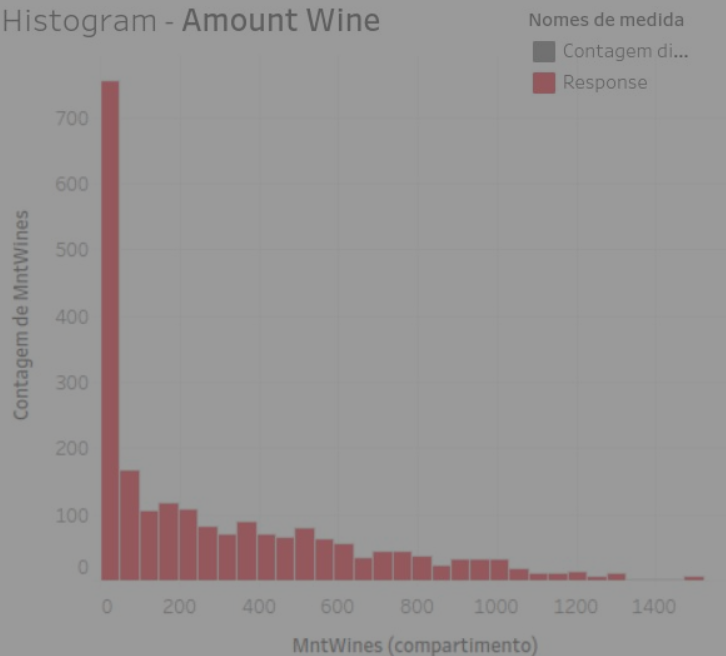


Amount Wine

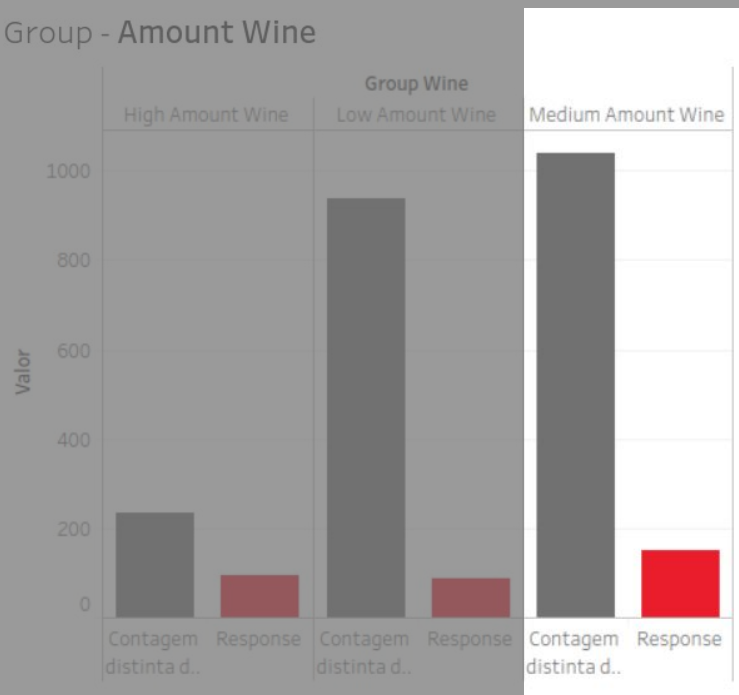
Analysis

Medium Amount Wine (101 to 800):
concentrates **46.99% of the users**,
but concentrates **45.04% of the**
responses

Histogram - Amount Wine



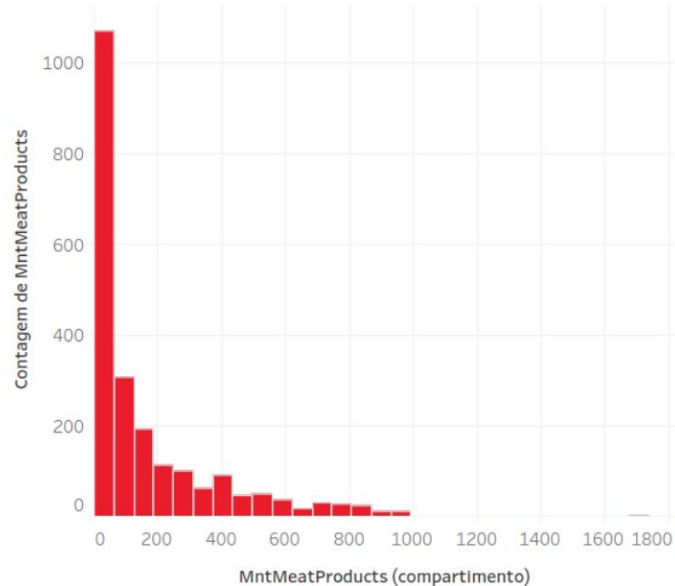
Group - Amount Wine



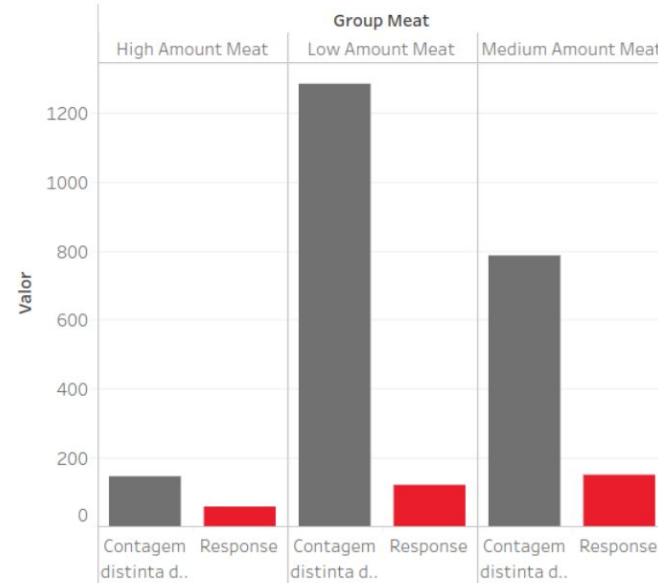
Amount Meat

Analysis

Histogram - Amount Meat

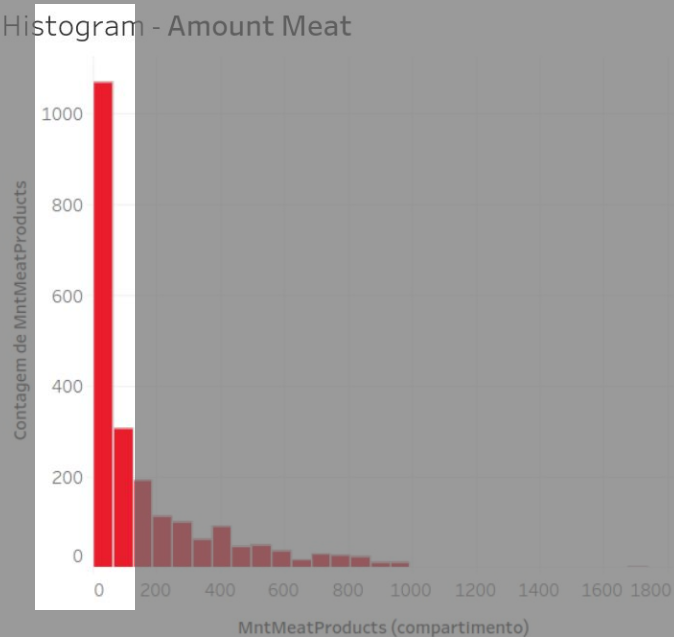


Group - Amount Meat

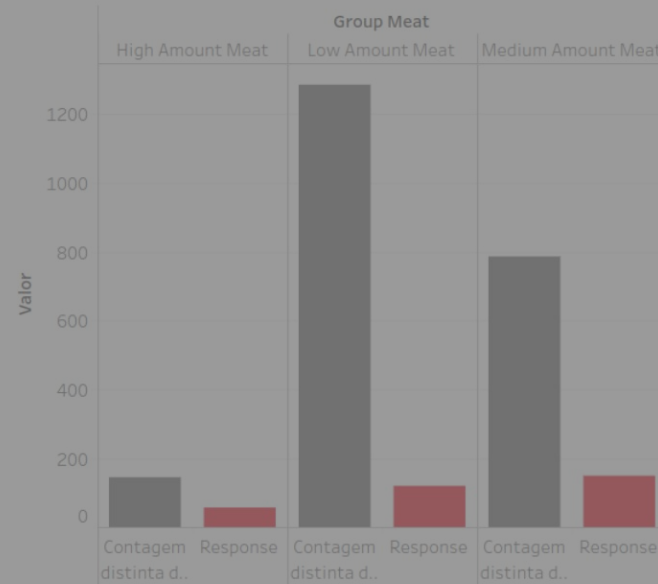


Is possible to notice
the main group is
Low Amount Meat
(0 to 100).

Histogram - Amount Meat



Group - Amount Meat

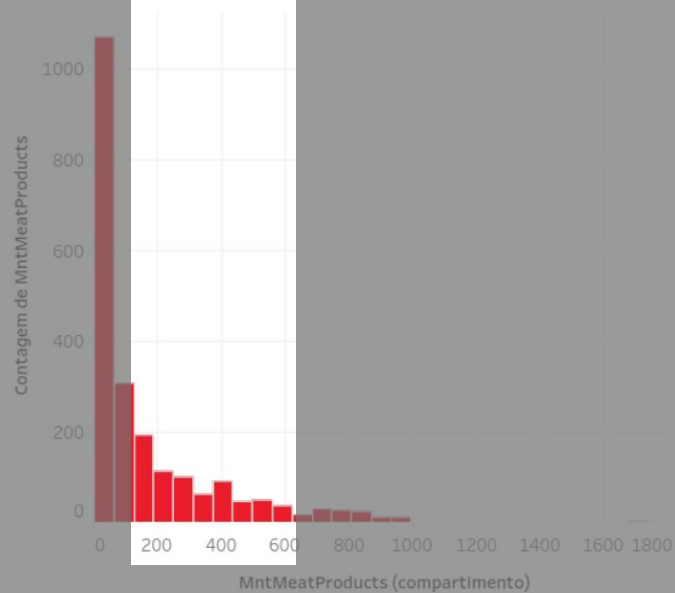


Amount Meat

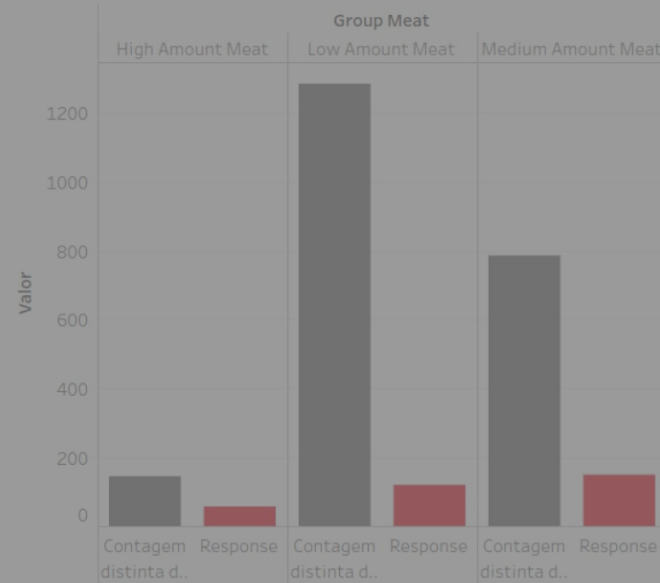
Analysis

Is possible to notice
the main group is
**Medium Amount
Meat** (101 to 600).

Histogram - Amount Meat



Group - Amount Meat

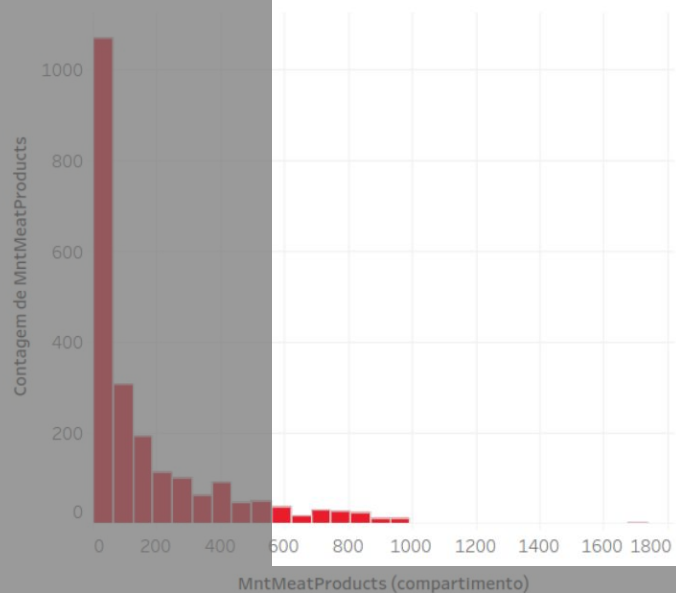


Amount Meat

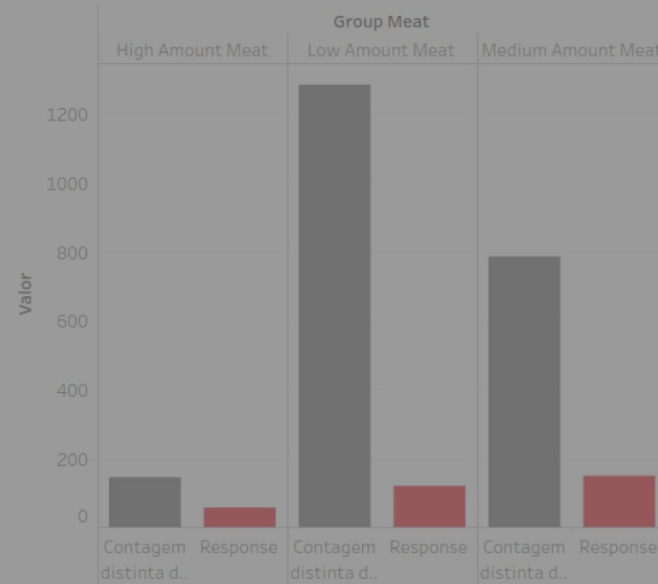
Analysis

Is possible to notice
the main group is
High Amount Meat
(+601).

Histogram - Amount Meat



Group - Amount Meat

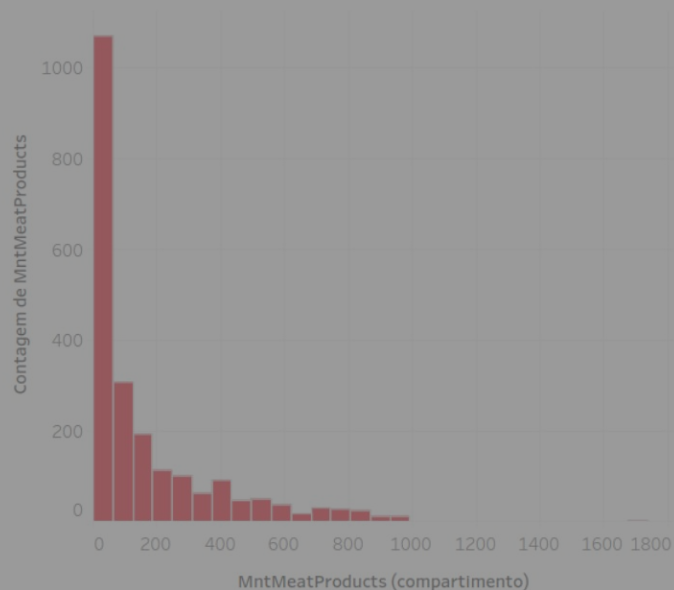


Amount Meat

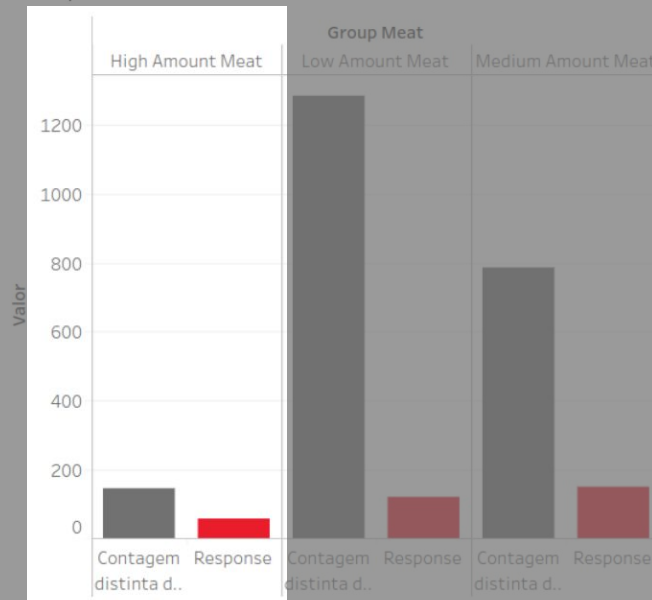
Analysis

High Amount Meat (+601):
concentrates **57.93% of the users**,
but concentrates **36.63% of the**
responses

Histogram - Amount Meat



Group - Amount Meat

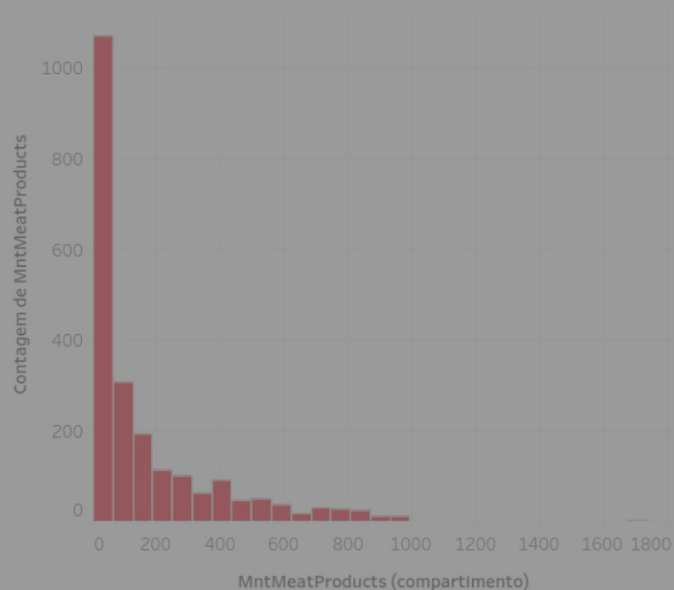


Amount Meat

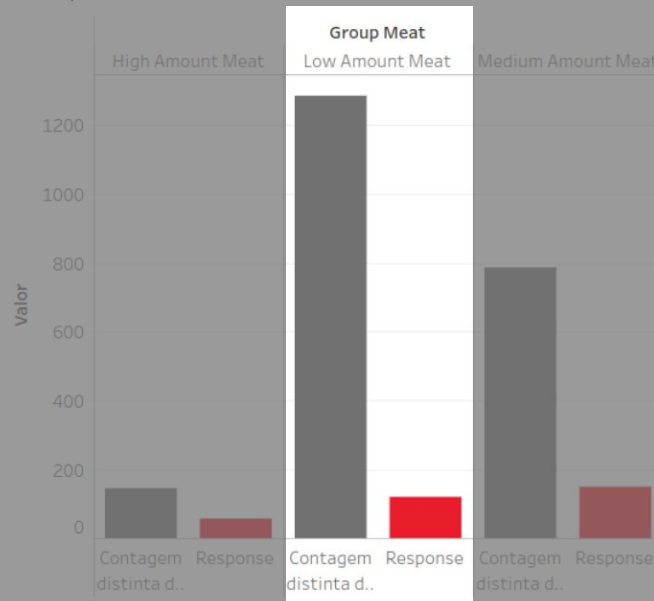
Analysis

Low Amount Meat (0 to 100):
concentrates **6.55% of the users**,
but concentrates **17.71% of the**
responses

Histogram - Amount Meat



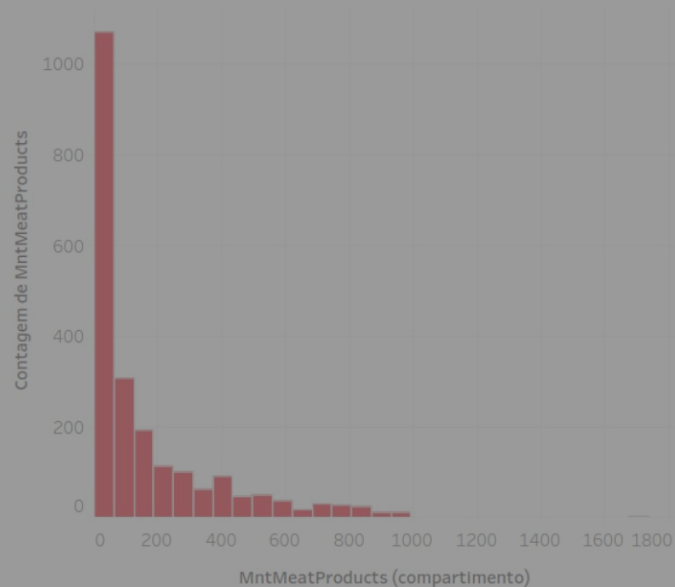
Group - Amount Meat



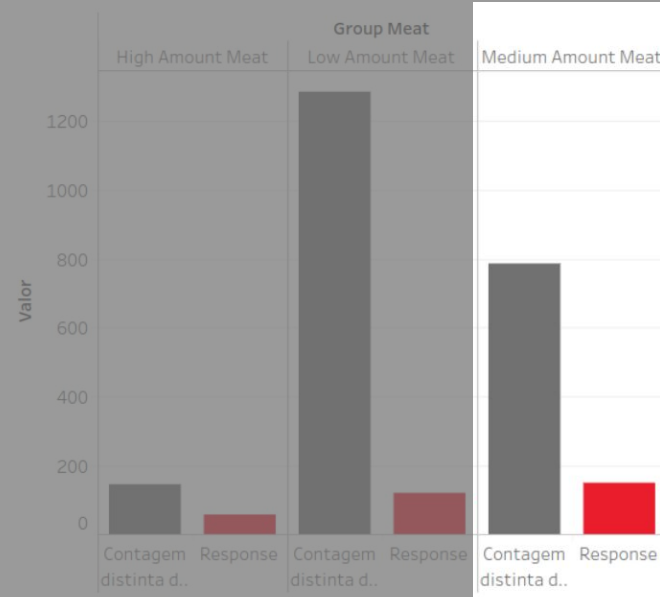
Amount Meat

Analysis

Histogram - Amount Meat



Group - Amount Meat

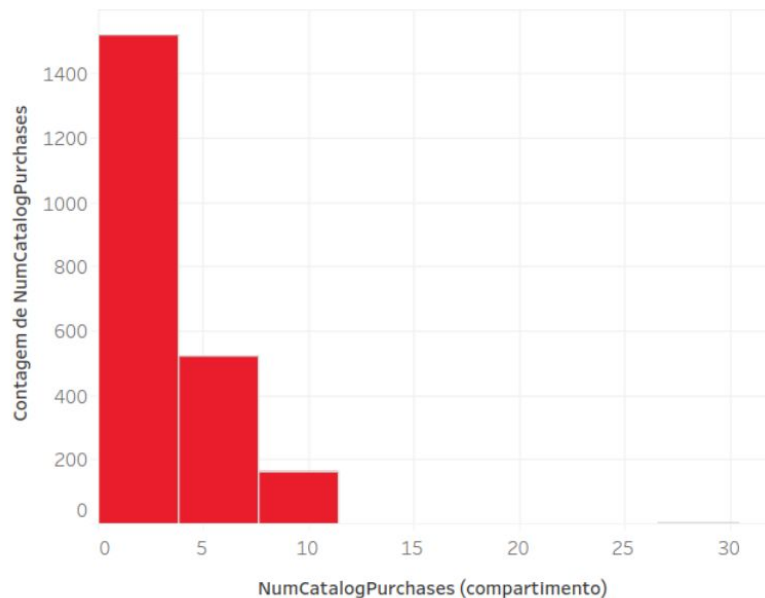


Low Amount Meat (0 to 100):
concentrates **35.51% of the users**,
but concentrates **45.64% of the responses**

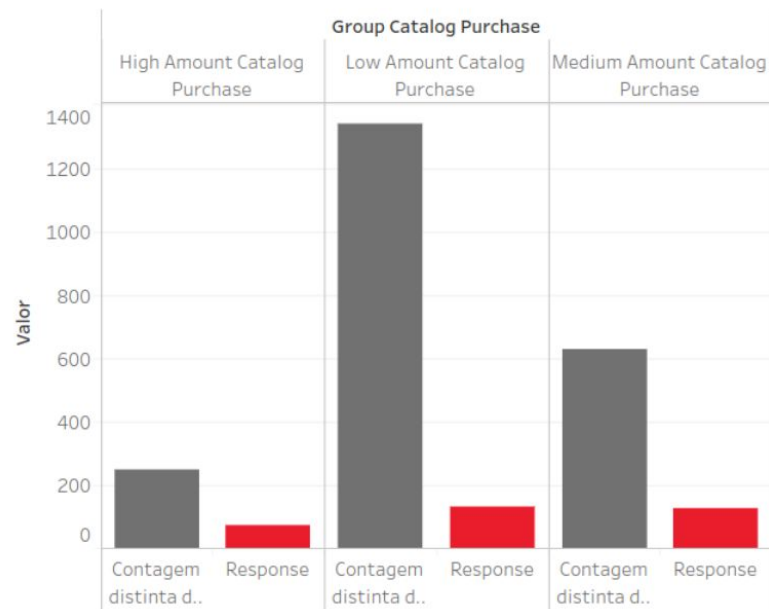
Amount Catalog Purchase

Analysis

Histogram - Catalog Purchase



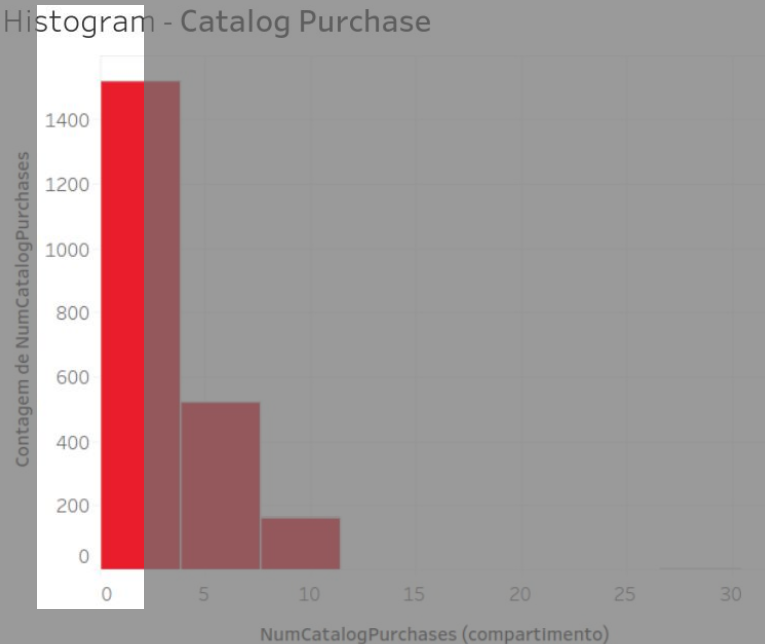
Group - Catalog Purchase



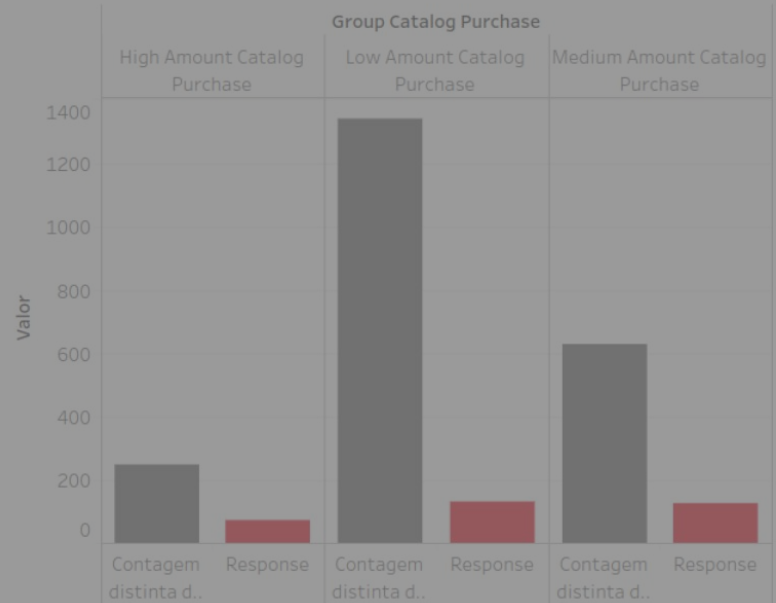
Is possible to notice
the main group is
**Low Catalog
Purchases** (0 to 2).

Catalog Purchase

Histogram - Catalog Purchase



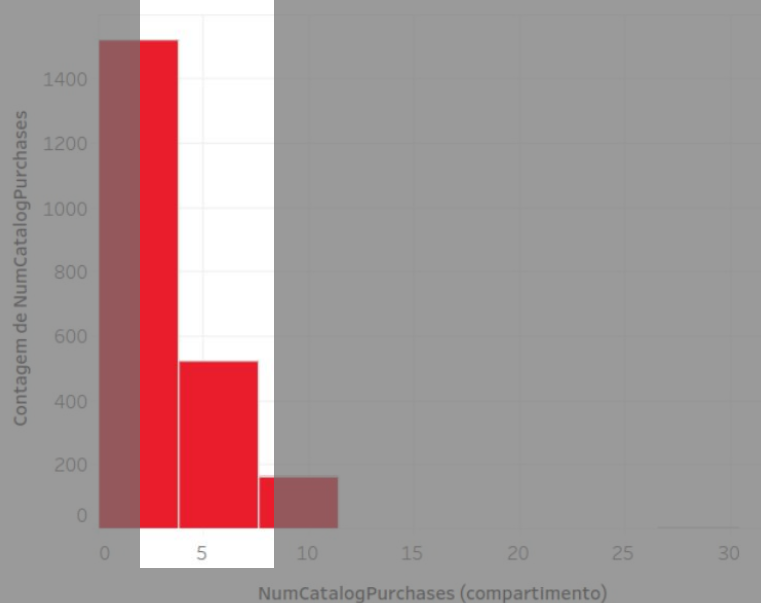
Group - Catalog Purchase



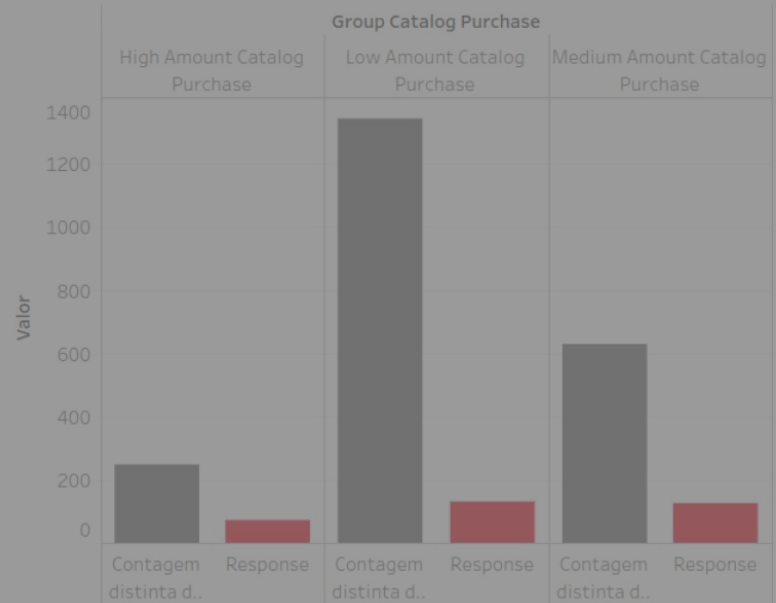
Is possible to notice the
main group is
**Medium Catalog
Purchases** (2 to 6).

Analysis of Catalog Purchase

Histogram - Catalog Purchase



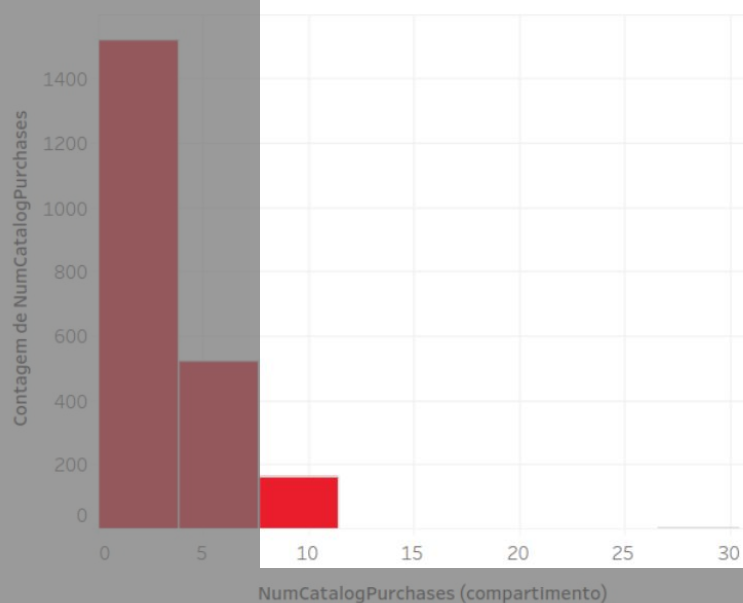
Group - Catalog Purchase



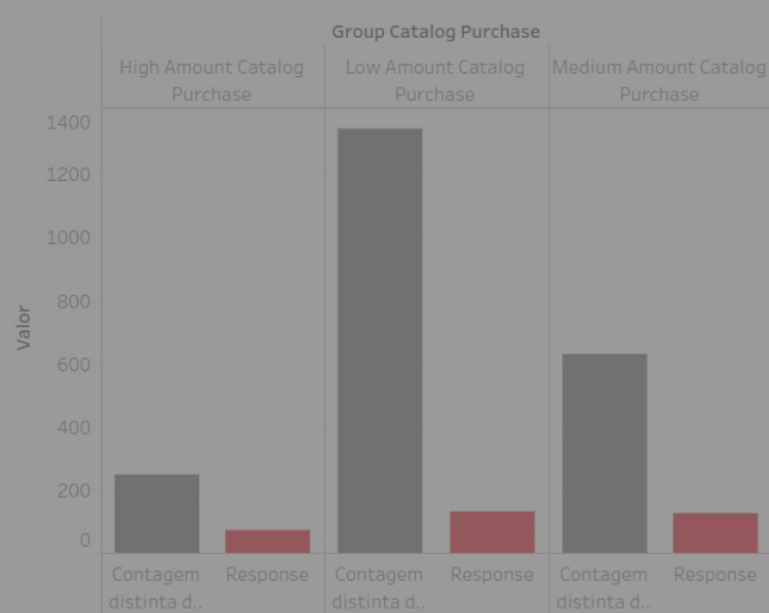
Amount of Purchase Analysis

Is possible to notice
the main group is
**High Catalog
Purchases (+6).**

Histogram - Catalog Purchase



Group - Catalog Purchase

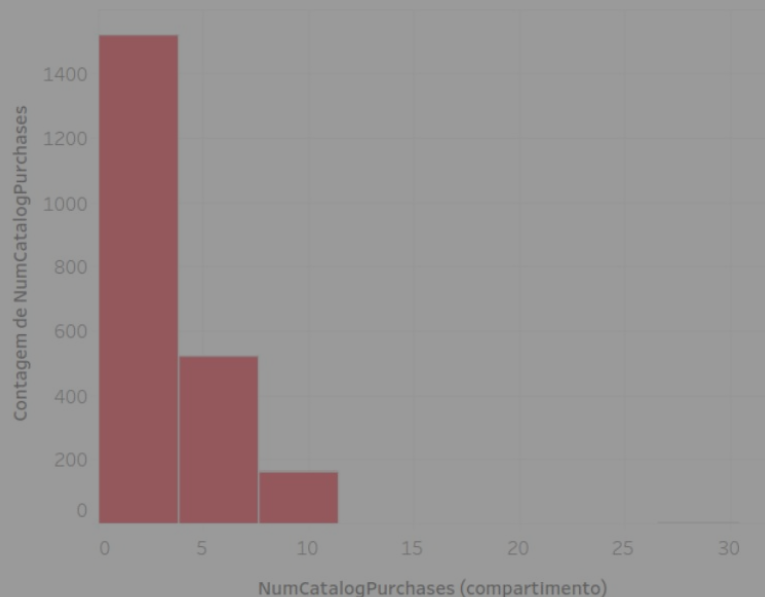


Amount Catalog

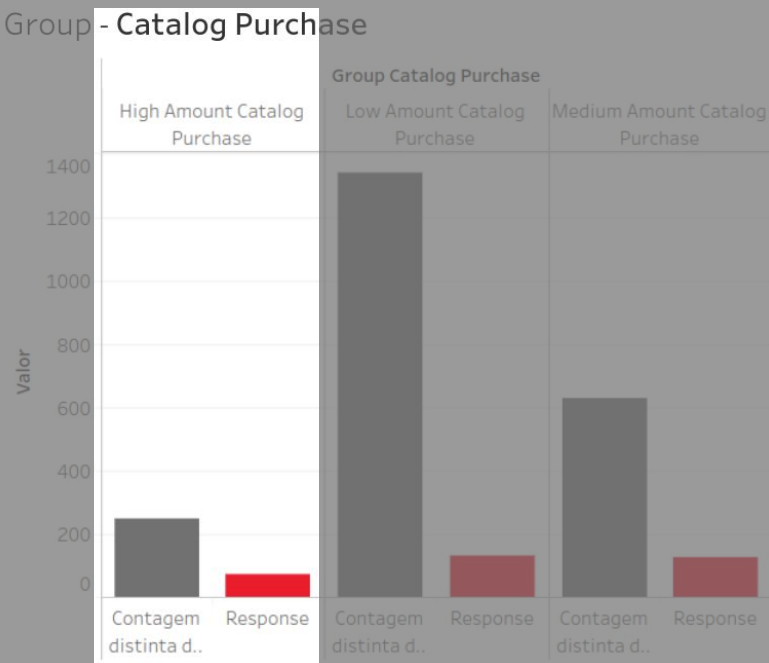
Analysis

High Catalog Purchases (+6):
concentrates **11.11% of the users**,
but concentrates **22.52% of the**
responses

Histogram - Catalog Purchase



Group - Catalog Purchase

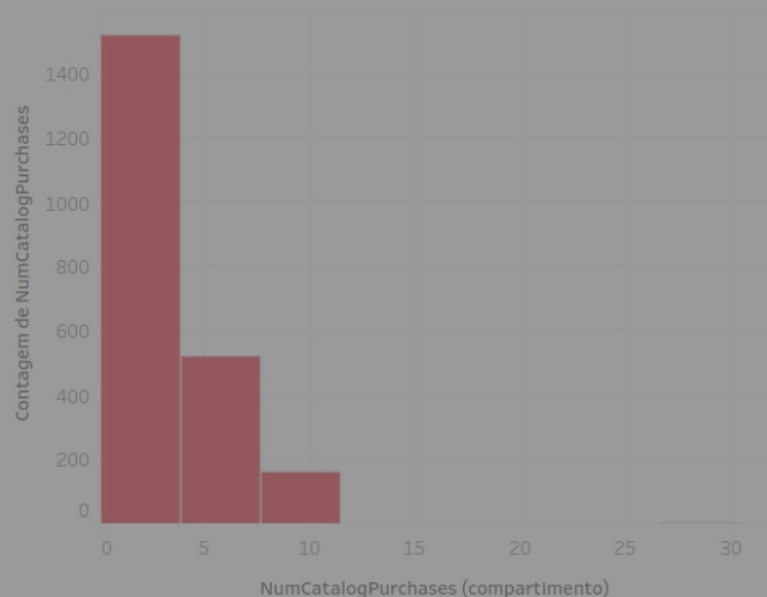


Amount Catalog Purchase

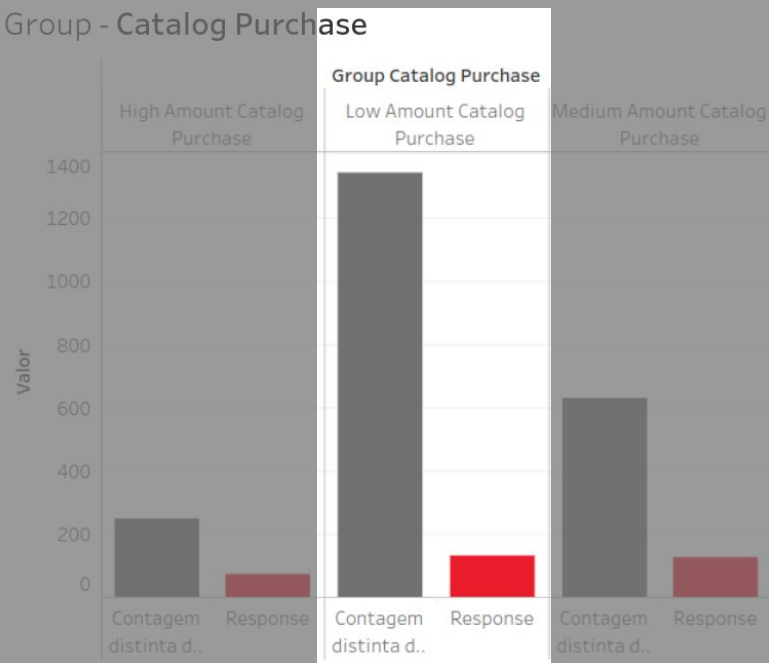
Analysis

Low Catalog Purchases (0 to 2):
concentrates **60.55% of the users**,
but concentrates **39.63% of the**
responses

Histogram - Catalog Purchase



Group - Catalog Purchase

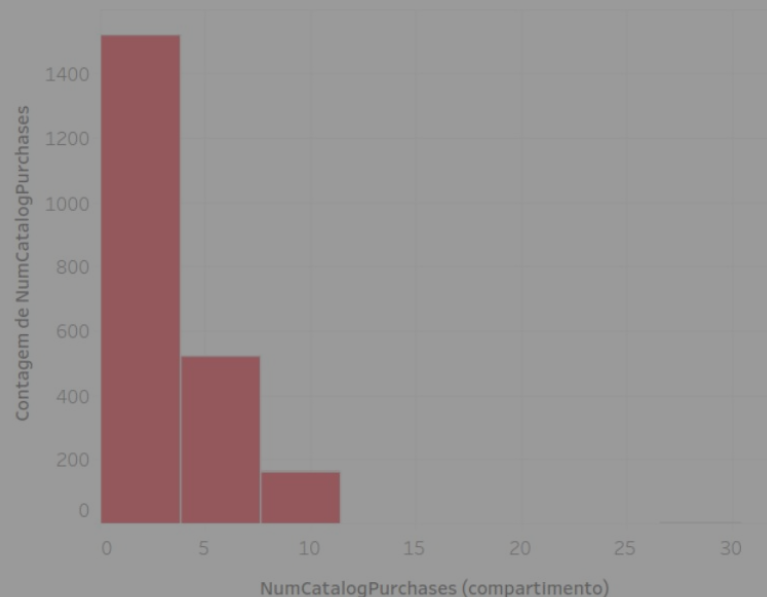


Amount Catalog Purchases

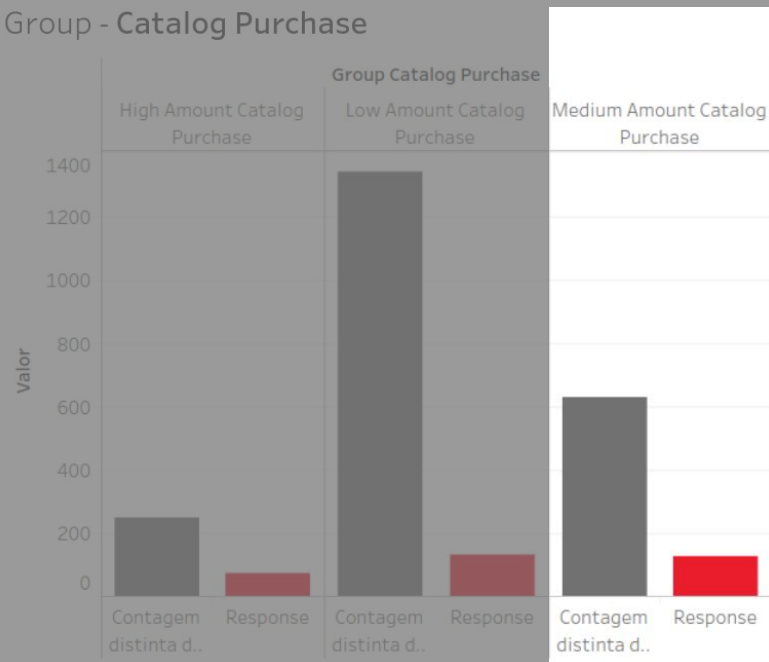
Analysis

Medium Catalog Purchases (2 to 6):
concentrates **28.33% of the users**,
but concentrates **37.83% of the responses**

Histogram - Catalog Purchase



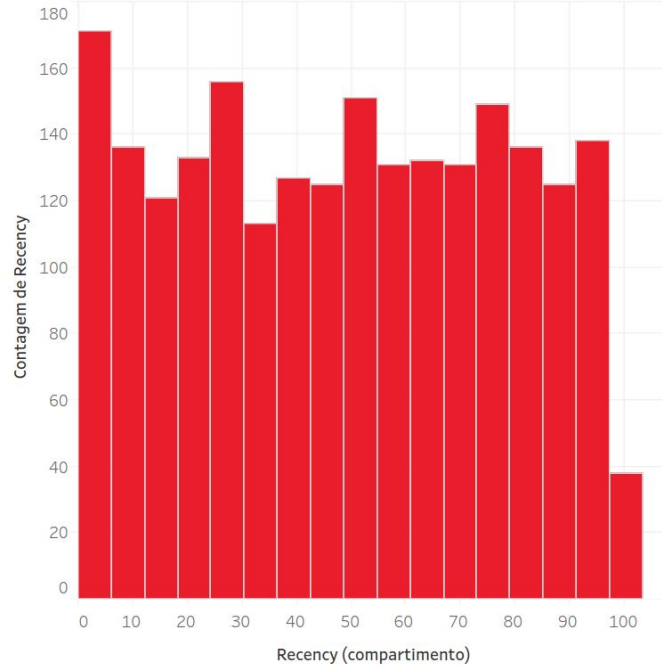
Group - Catalog Purchase



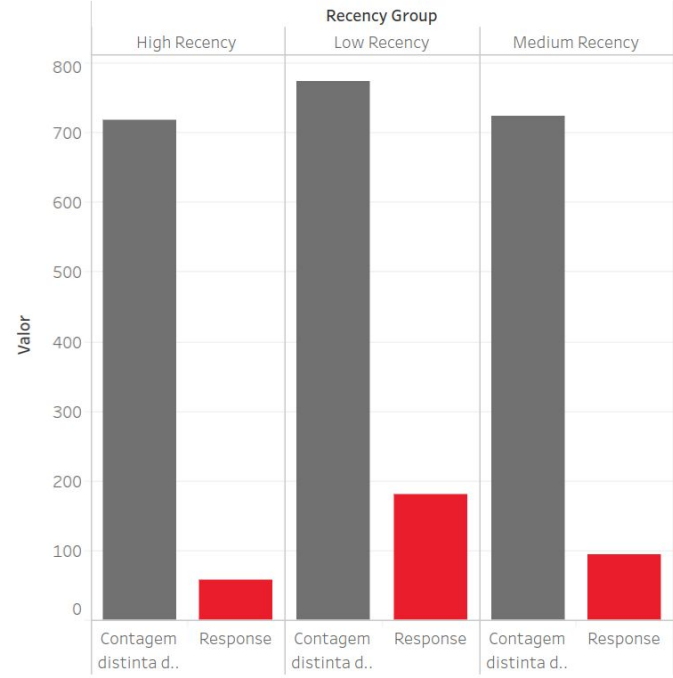
Recency

Analysis

Histogram - Recency



Group - Recency

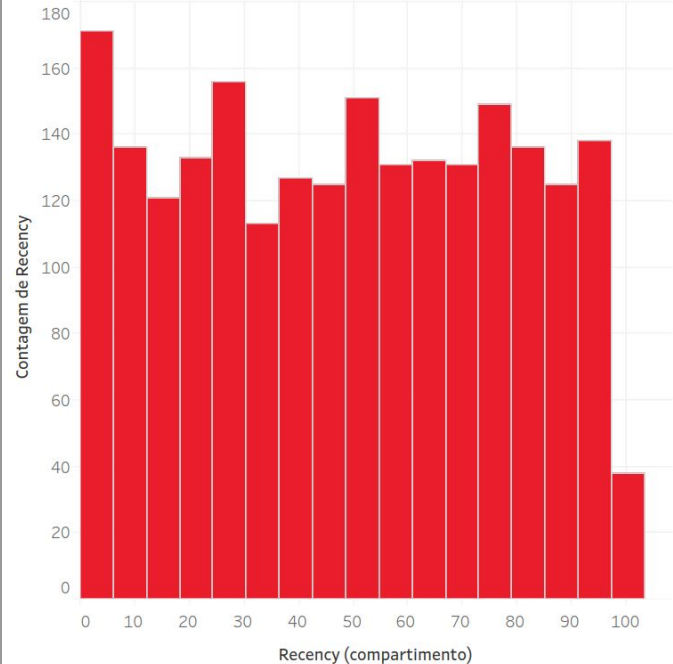


Rec

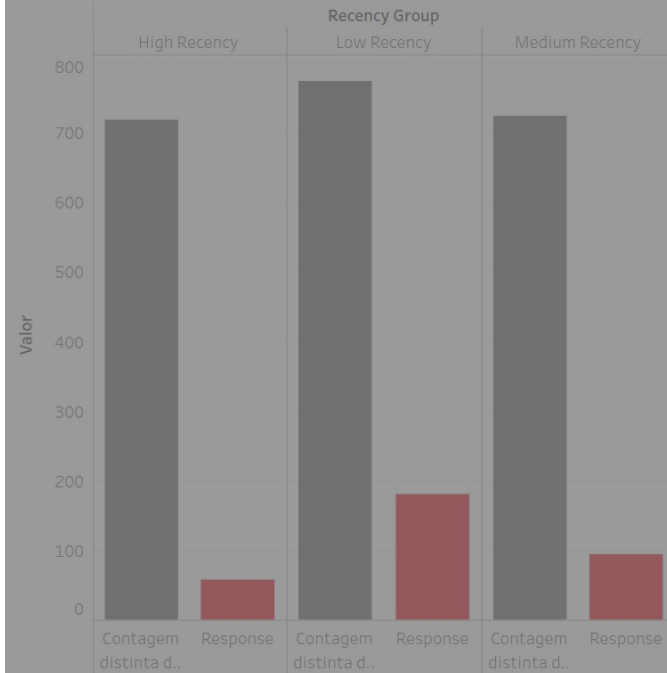
Analysis

This variable doesn't have a concentration on any specific amount of days. So we decided to divide equally into the Recency Groups.

Histogram - Recency



Group - Recency

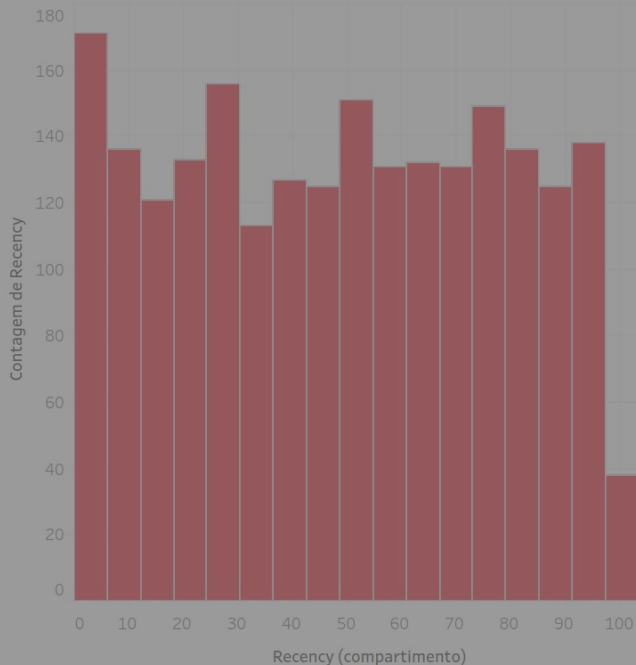


Recency

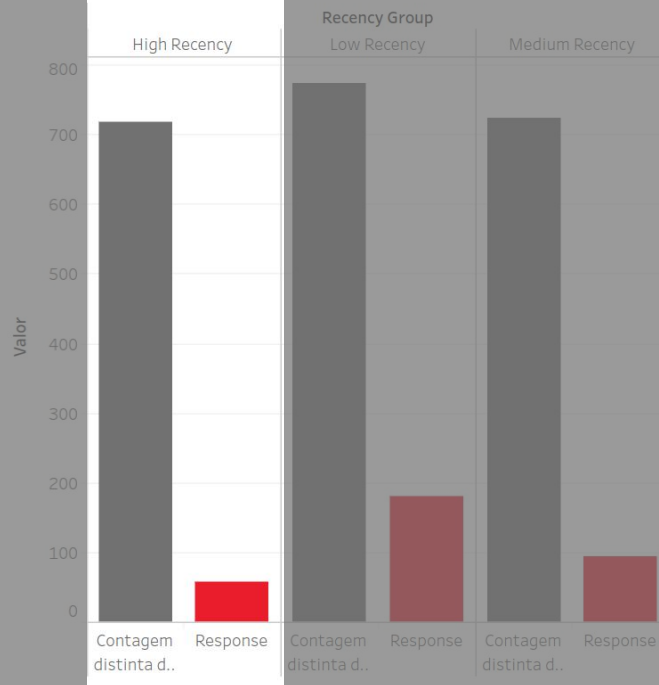
Analysis

High Recency (+66): concentrates 33.12% of the users, but concentrates 17.71% of the responses

Histogram - Recency



Group - Recency

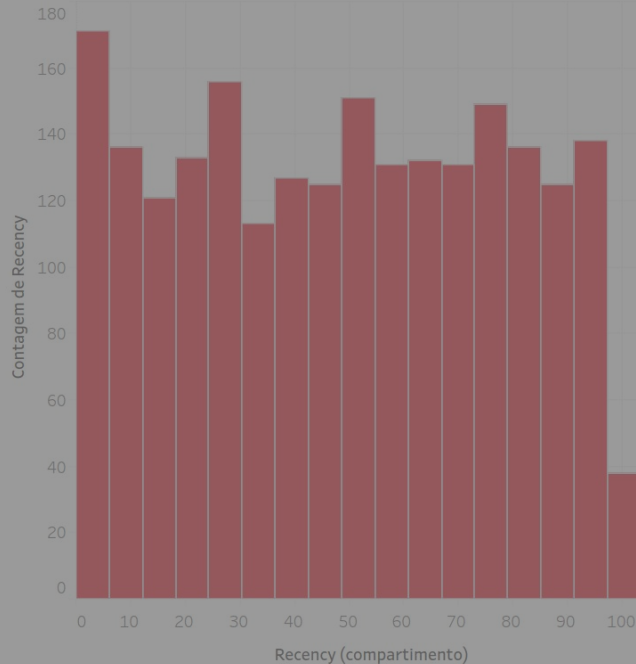


Recency

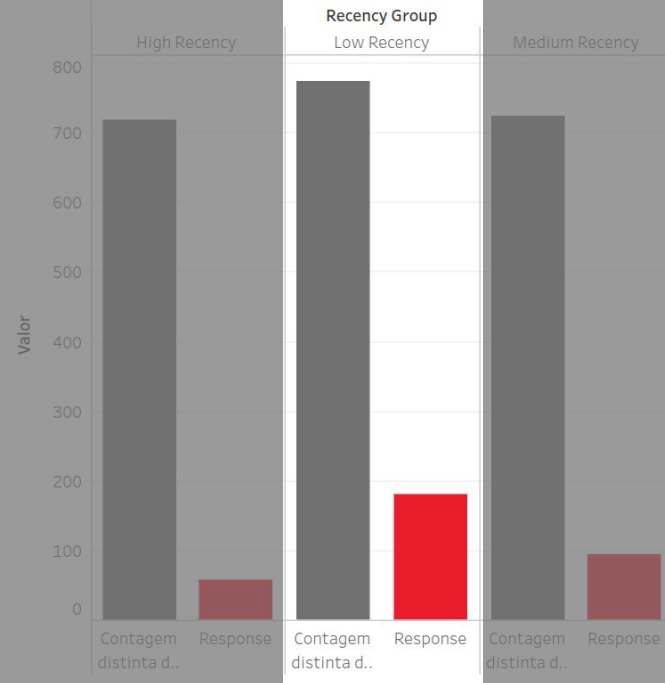
Analysis

Low Recency (0 to 33):
concentrates **34.92% of the users**,
but concentrates **54.35% of the responses**

Histogram - Recency



Group - Recency

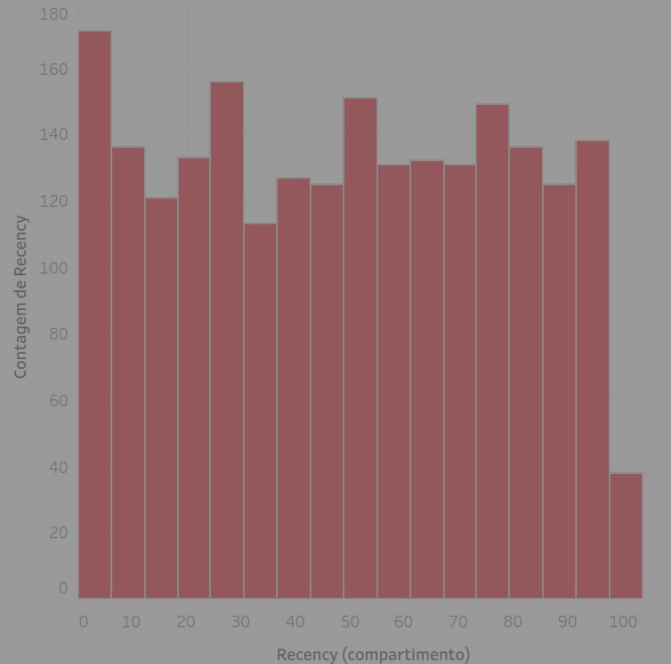


Recency

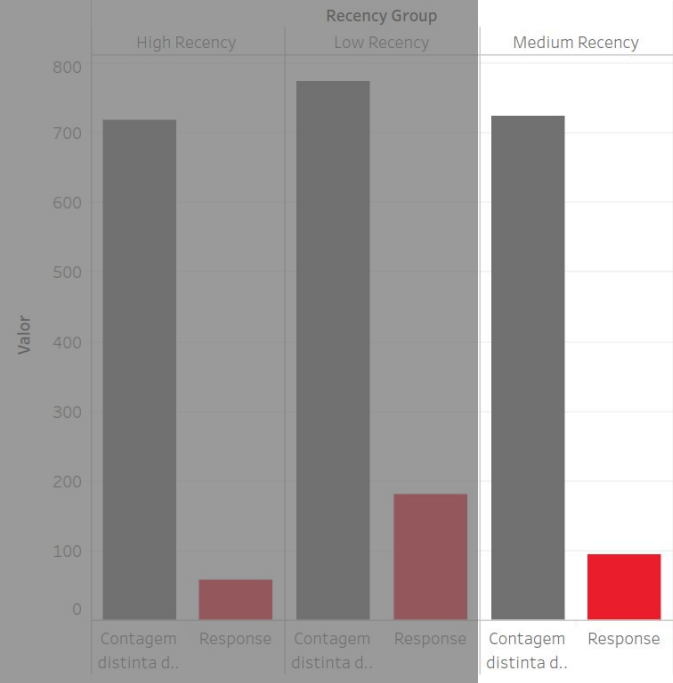
Analysis

Medium Recency (34 to 66):
concentrates **31.94% of the users**,
but concentrates **27.92% of the**
responses

Histogram - Recency



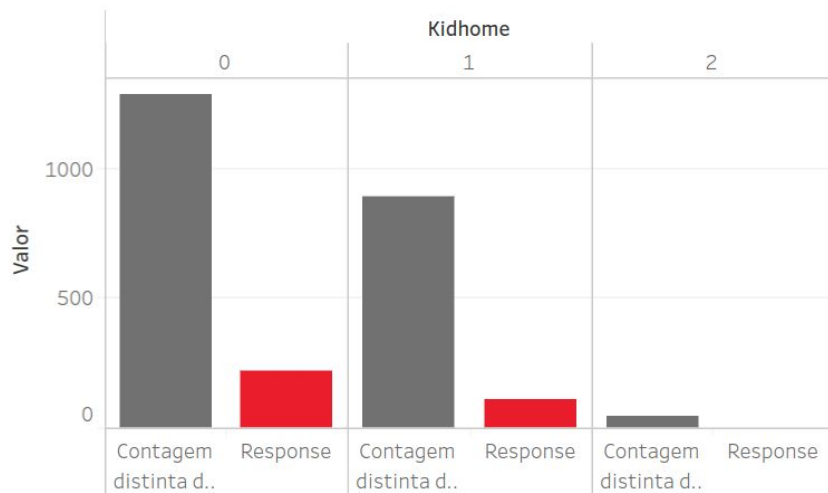
Group - Recency



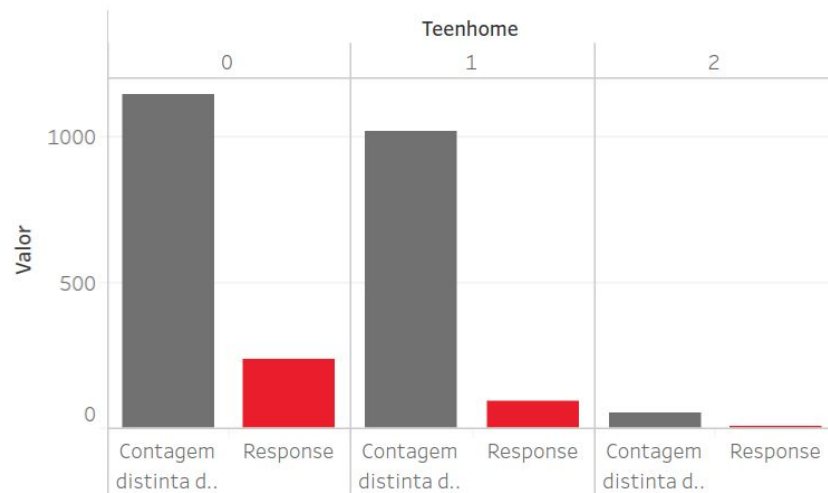
Kid or Teen at Home

Analysis

Kids at home

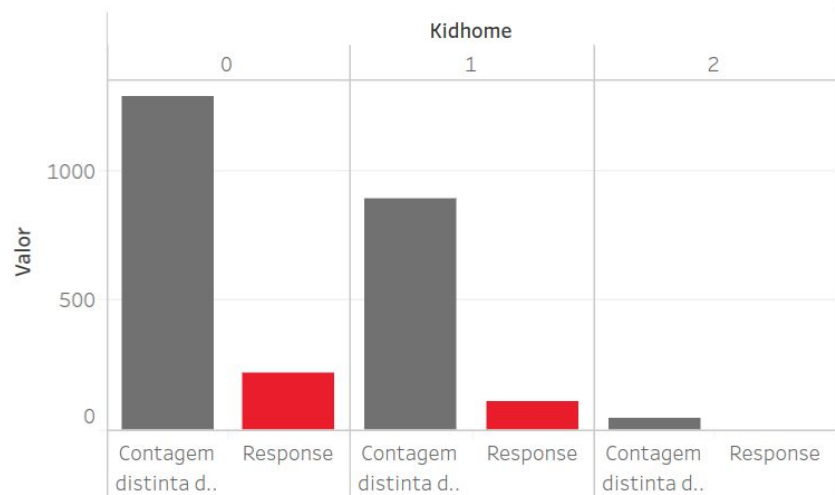


Teens at home

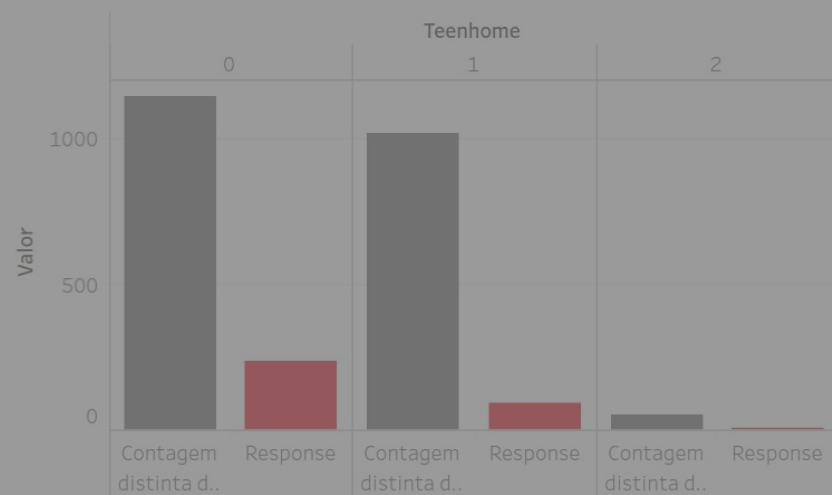


o **kids** represent **57.88%** of the customer base and concentrate **66.34%** of the responses

Kids at home



Teens at home

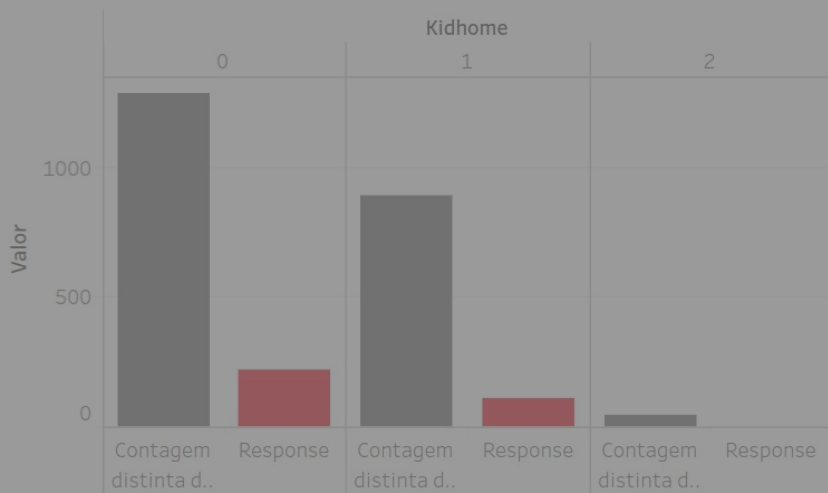


Kid or Teen at Home

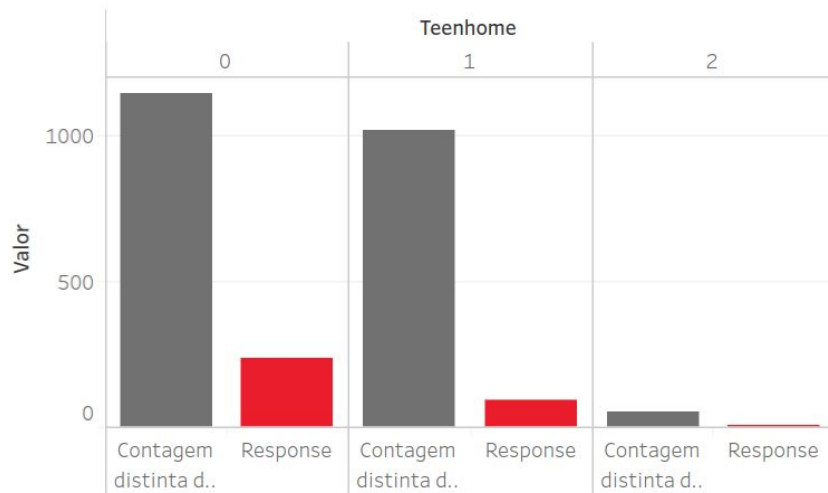
Analysis

0 teens represent **51.73%** of the customer base and concentrate **70.87%** of the responses

Kids at home



Teens at home



Insights

Estimated Results, based on the sample data

Insight	Priority	Total Impacted	Total Positive Responses	Campaign Rate	Total Investment	Total Return	ROI
Accepted campaigns 1, 3 and 5	1	366	172	46.99%	1,109	1,897	171.07%
High Mount Meat	1	145	59	40.69%	439	651	148.12%
High Mount Wine	1	236	96	40.68%	715	1,059	148.08%
High Catalog Purchases	1	246	75	30.49%	745	827	110.98%
Customers from 2012	1	490	132	26.94%	1,485	1,456	98.06%
Low Recency (<= 33)	1	773	181	23.42%	2,342	1,996	85.24%
Single	1	470	106	22.55%	1,424	1,169	82.10%
PhD	1	480	101	21.04%	1,454	1,114	76.60%
Divorced	2	231	48	20.78%	700	529	75.64%
Medium Mount Meat	2	786	152	19.34%	2,382	1,677	70.40%
Medium Catalog Purchases	2	901	165	18.31%	2,730	1,820	66.66%
Kidhome or Teenhome = 0	1	1,794	289	16.11%	5,436	3,188	58.64%
Year Birth - 1951 to 1960 or 1981 to 1990	2	810	126	15.56%	2,454	1,390	56.63%
Graduation	2	1,116	152	13.62%	3,381	1,677	49.58%
Customers from 2013	2	1,171	153	13.07%	3,548	1,688	47.56%
Medium Income (33k to 66k)	1	1,119	130	11.62%	3,391	1,434	42.29%
Low Mount Wine	2	937	87	9.28%	2,839	960	33.80%

Insights

Results, based on the sample data

Those are the best insights, with campaign rate beyond the premise and the ROI positive.

Insight	Priority	Total Impacted	Total Positive Responses	Campaign Rate	Total Investment	Total Return	ROI
Accepted campaigns 1, 3 and 5	1	366	172	46.99%	1,109	1,897	171.07%
High Mount Meat	1	145	59	40.69%	439	651	148.12%
High Mount Wine	1	236	96	40.68%	715	1,059	148.08%
High Catalog Purchases	1	246	75	30.49%	745	827	110.98%
Customers from 2012	1	490	132	26.94%	1,485	1,456	98.06%
Low Recency (<= 33)	1	773	181	23.42%	2,342	1,996	85.24%
Single	1	470	106	22.55%	1,424	1,169	82.10%
PhD	1	480	101	21.04%	1,454	1,114	76.60%
Divorced	2	231	48	20.78%	700	529	75.64%
Medium Mount Meat	2	786	152	19.34%	2,382	1,677	70.40%
Medium Catalog Purchases	2	901	165	18.31%	2,730	1,820	66.66%
Kidhome or Teenhome = 0	1	1,794	289	16.11%	5,436	3,188	58.64%
Year Birth - 1951 to 1960 or 1981 to 1990	2	810	126	15.56%	2,454	1,390	56.63%
Graduation	2	1,116	152	13.62%	3,381	1,677	49.58%
Customers from 2013	2	1,171	153	13.07%	3,548	1,688	47.56%
Medium Income (33k to 66k)	1	1,119	130	11.62%	3,391	1,434	42.29%
Low Mount Wine	2	937	87	9.28%	2,839	960	33.80%

Summary

1. Process
2. Main Results
3. Explanation of the Model
4. Data Analysis in Resume
- 5. Conclusion**



Conclusion

If the CRM team follow the segmentations or model recommendations, is possible to achieve the following results:

- With the **model**, is possible to get close to 100% on Campaign Rate. Also, the marketing team would bring **3.64MU for each 1MU** spend;
- With the **best insight and audience** (accepting campaigns 1, 3, and 5), the campaign rate will be close to 46%. Also, the marketing team would bring **1.71MU for each 1MU** spend.

—

Thank you

Questions?

By Israel Mendes

