MARKETING
DATA ANALYSIS

Midproject for
IRONHACK DA BOOTCAMP

INDEX PRESENTATION

01

THE CHALLENGE

What do we have here?

The purpose of this project is to **analyze and interpret a dataset** with information about the **customers of a store**. We don't know what type of business it is except that **they sell grocery items**. We only know what **types of products** they provide, the **ways in which customers have previously purchased** and some **metrics about marketing campaigns** launched.

It is up to us to **determine what type of clientele we have**, what their **relationship is with the products on sale** and to interpret how the **campaigns launched have gone**.

We will also try, with the final reading of all the data, to **give some guidelines to the marketing department** to increase sales.


Instructions
Keep refrigerated or frozen. Thaw in refrigerator or microwave.
Keep raw meats and poultry separate from other foods. Wash working surfaces (including cutting boards), utensils, and hands after touching raw meat or poultry.
Cook thoroughly.
Keep hot foods hot. Refrigerate leftovers immediately or discard.
No Antibiotics No Hormones
PREMIUM CUTS
FINEST BEEF
All Natural Angus
(Product of USA)
シャブシャブ
SHABU SHABU

02

EDA

DATAFRAME SNEAK PEAK

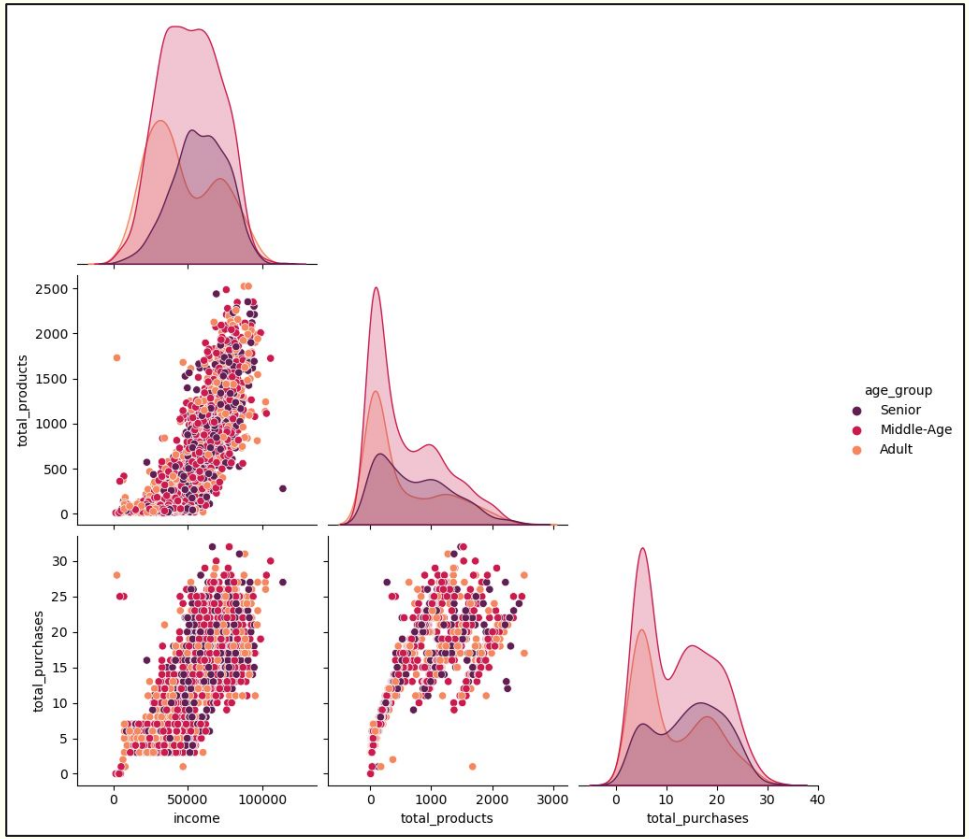
	id	age	age_group	education	marital_status	income	childrenhome	dt_customer	recency	wine_products	fruit_products	meat_products	fish_prod
0	5524	67	Senior	Graduation	Single	58138.0	0	2012-09-04	58	635	88	546	
1	2174	70	Senior	Graduation	Single	46344.0	2	2014-03-08	38	11	1	6	
2	4141	59	Middle-Age	Graduation	Together	71613.0	0	2013-08-21	26	426	49	127	
3	6182	40	Adult	Graduation	Together	26646.0	1	2014-02-10	26	11	4	20	
4	5324	43	Adult	PhD	Married	58293.0	1	2014-01-19	94	173	43	118	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2235	10870	57	Middle-Age	Graduation	Married	61223.0	1	2013-06-13	46	709	43	182	
2236	4001	78	Senior	PhD	Together	64014.0	3	2014-06-10	56	406	0	30	
2237	7270	43	Adult	Graduation	Divorced	56981.0	0	2014-01-25	91	908	48	217	
2238	8235	68	Senior	Master	Together	69245.0	1	2014-01-24	8	428	30	214	
2239	9405	70	Senior	PhD	Married	52869.0	2	2012-10-15	40	84	3	61	

2225 rows × 31 columns

1. This dataset contains a total of **2240 entries** and **29 columns** corresponding to clients information and their related activities with our business.
2. Almost all data types are numeric (**25 integers** and **1 float**). Then, we have **3 objects**.
3. Our first impression is that in the **categorical variables we will need to group and transform the values for optimization**. And we will see for the numeric variables, because some of them are one hot encoding, others are continual and some discrete.
4. After this first look, our main objective will be to demographically sectorize the type of customer that participates in our campaigns in order to **detect which is our best type of customer**.
5. Other objectives will be to interpret the **feasibility of campaigns**, **preferences** and **expenditures** of our clients.
6. We will **design specific buyer personas** thanks to the data reading we'll have at the end.

CLIENTS STATISTICS

Income vs Total Continuous Clusters



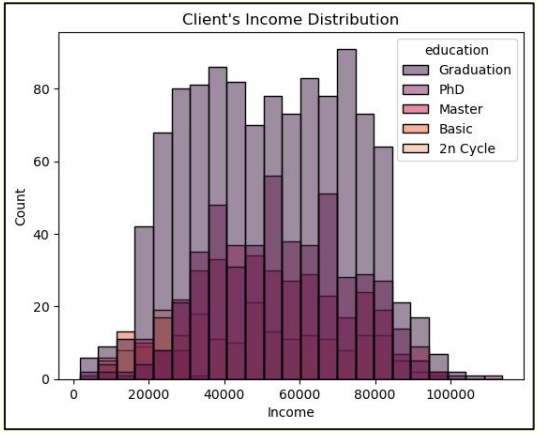
- In all **Income** ranges, the **Middle-Age** sector predominates. Followed in the first half by the **Adult** group and in the second half by the **Senior** group

- For **Amount Spent**, **Middle-Age** again dominates, followed by the **Adult** and **Senior** groups. During the increase in spending, we noted a drop in the **Adult** group and an increase in the **Senior** group

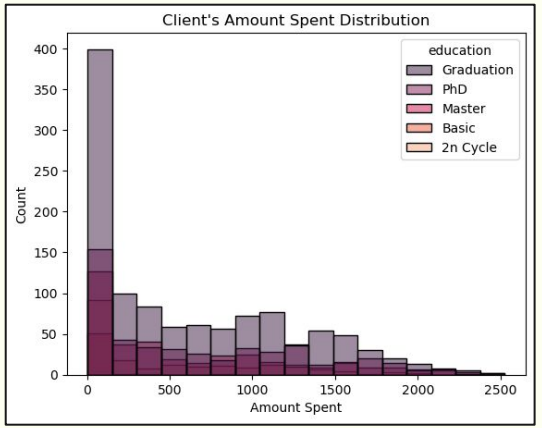
- For **Total Purchases**, the evolution is quite similar to that observed for other variables. **Middle-Age** predominates, followed by **Adult** and **Senior**. As in Expenditure, **Senior** overpowers **Adult** as the number of purchases increases

CLIENTS STATISTICS

Income & Amount Spent Histograms



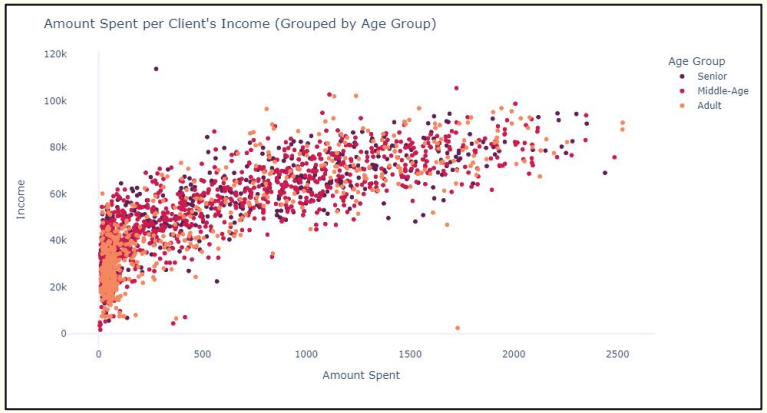
- For the distribution of **Income** grouped by **Level of Education**, we can see that **Graduation** dominates all ranges. We can understand that the **purchasing power** is correlated with the **grade of education** (the **Post-Graduation** groups are below, being less frequent)



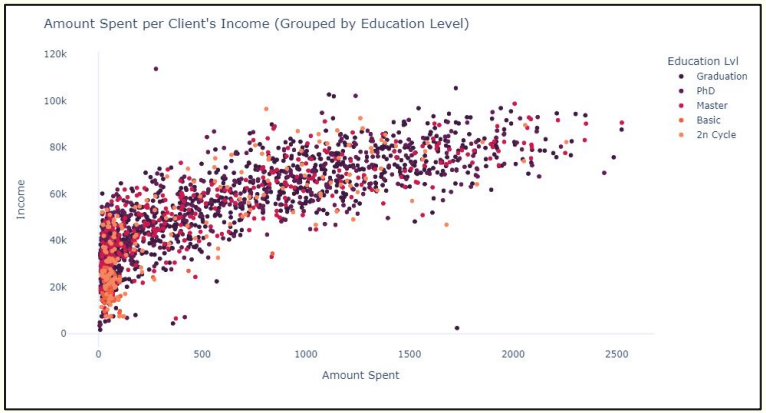
- In the distribution of Amount Spent, we see that the **Graduation** level also predominates. From the right-skewed distribution we understand that the **highest amount spent** is in the **first ranges**, diluting towards higher amounts spent

CLIENTS STATISTICS

Amount Spent vs Income



- In this plot, we can see how the mass of **Adults** is agglomerated in the first ranks and **Middle-Age** is gaining relevance in a linear ascent

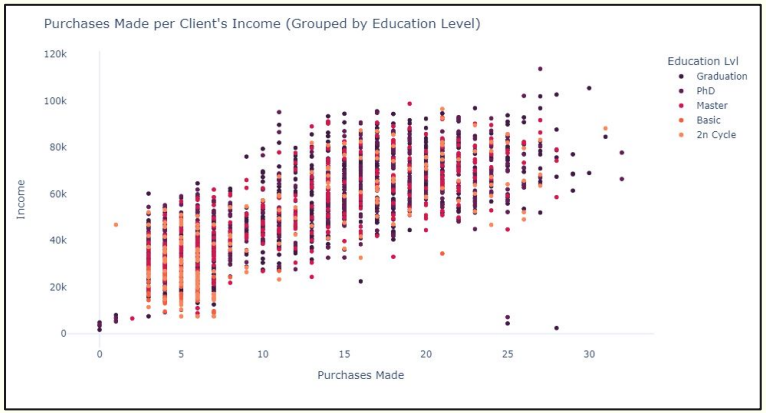


- In the same plot but grouped by **Education**, we see that the lowest levels of education are grouped in the first ranks. The progression is linear and the higher levels of **Education** predominate

CLIENTS STATISTICS

Purchases vs Income

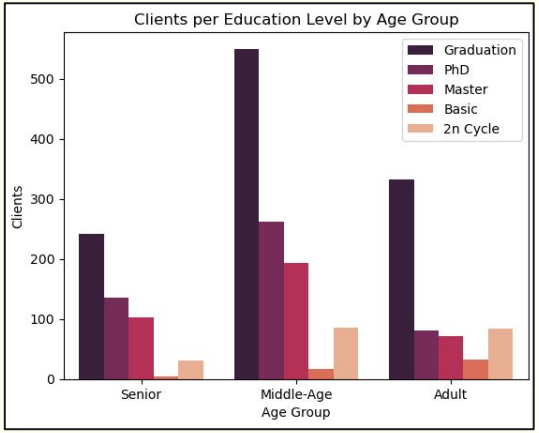
- Once again, we can see an upward trend. **Adults** in the lower levels giving way to **Middle-Age**



- For the grouping by **Education**, patterns similar to the previous graphs occur. The most basic levels in the first ranks and the highest levels gain prominence in the largest number of purchases

SPECIFIC STATISTICS

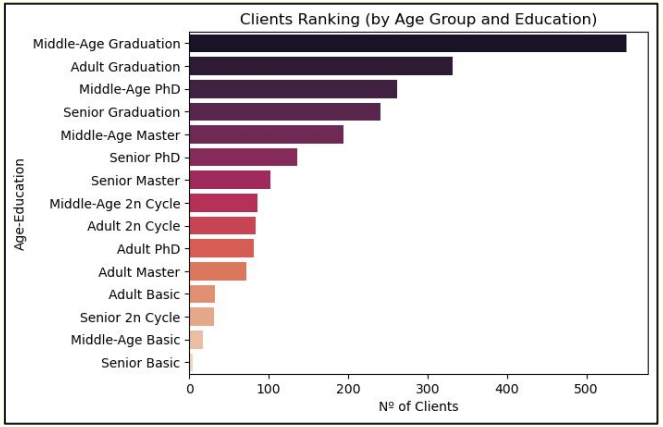
Group Age / Education



- The correlation observed in the previous plots is confirmed here. Both age (**Middle-Age**) and education (**Graduation**) predominate as frequent customers

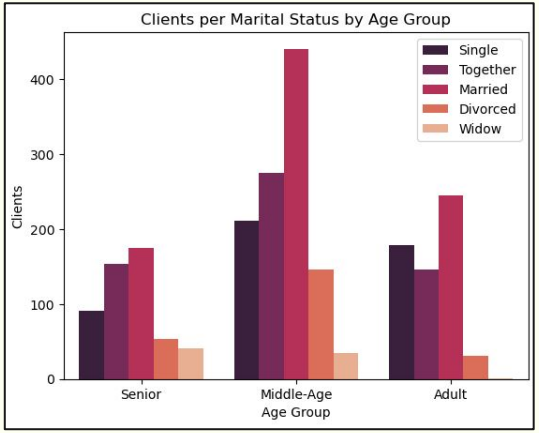
- **Combining the tags** of the variables by input we will be able to see more accurately what our customers are like

```
Middle-Age Graduation: 550 clients of 2225; 24.72% of total clients
Adult Graduation: 332 clients of 2225; 14.92% of total clients
Middle-Age PhD: 262 clients of 2225; 11.78% of total clients
Senior Graduation: 241 clients of 2225; 10.83% of total clients
Middle-Age Master: 194 clients of 2225; 8.72% of total clients
```



SPECIFIC STATISTICS

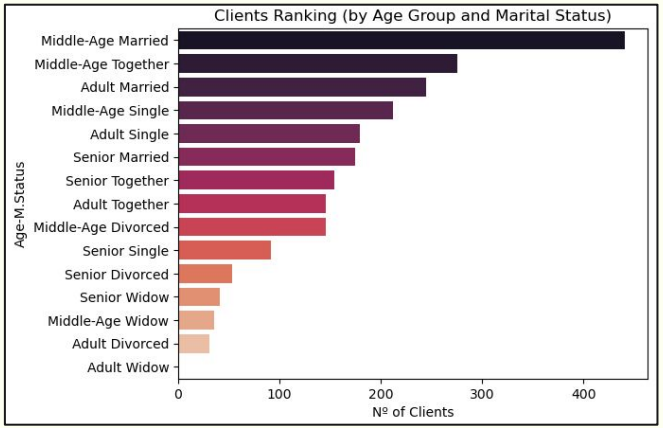
Group Age / Marital Status



- This plot indicates the predominance of the **Married** and **Together** groups among our clients. Except in the case of **Adults** cluster, where **Single** takes the lead behind **Married**

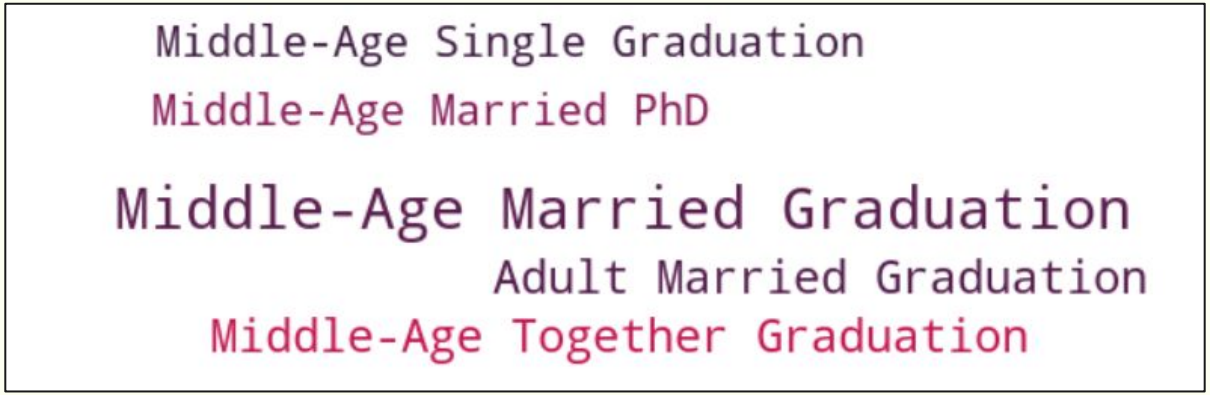
- As we did before, we can do the **customer type ranking** by these new classification labels

```
Middle-Age Married: 441 clients of 2225; 19.82% of total clients
Middle-Age Together: 275 clients of 2225; 12.36% of total clients
Adult Married: 245 clients of 2225; 11.01% of total clients
Middle-Age Single: 212 clients of 2225; 9.53% of total clients
Adult Single: 179 clients of 2225; 8.04% of total clients
```



SPECIFIC STATISTICS Group Age / Marital Status / Marital Status

How about considering three factors together?

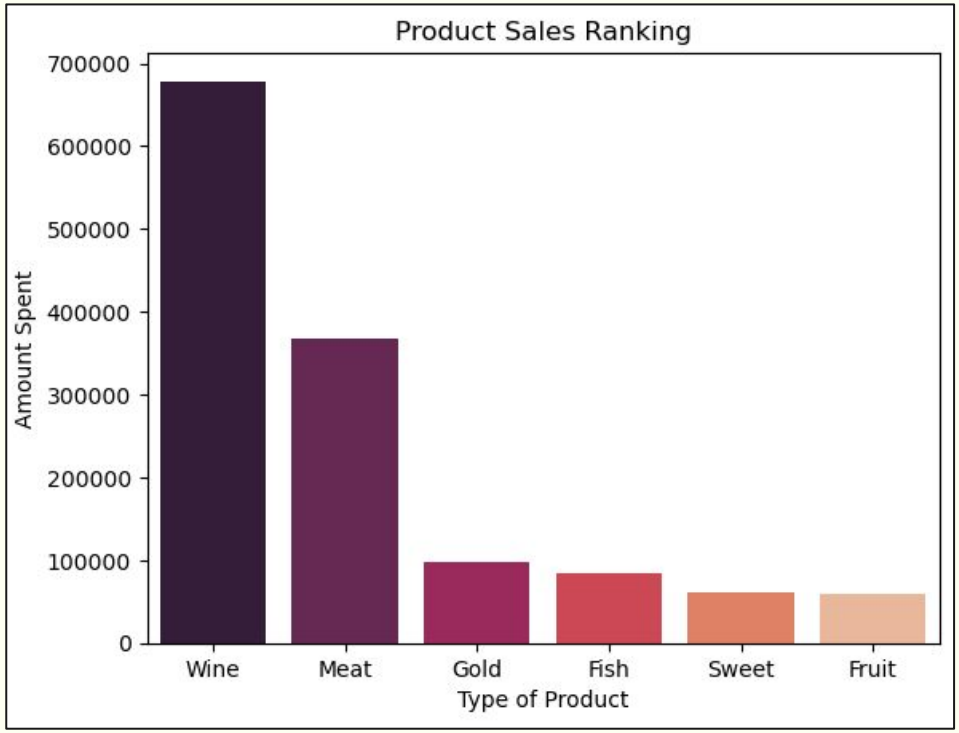


- This tag cloud reveals the **top 5 customers**, among all the combinations of characteristics they can have

```
Middle-Age Married Graduation: 213 clients of 2225; 9.57% of total clients
Adult Married Graduation: 138 clients of 2225; 6.2% of total clients
Middle-Age Together Graduation: 135 clients of 2225; 6.07% of total clients
Middle-Age Single Graduation: 113 clients of 2225; 5.08% of total clients
Middle-Age Married PhD: 109 clients of 2225; 4.9% of total clients
```

SPECIFIC STATISTICS

Best Selling Products

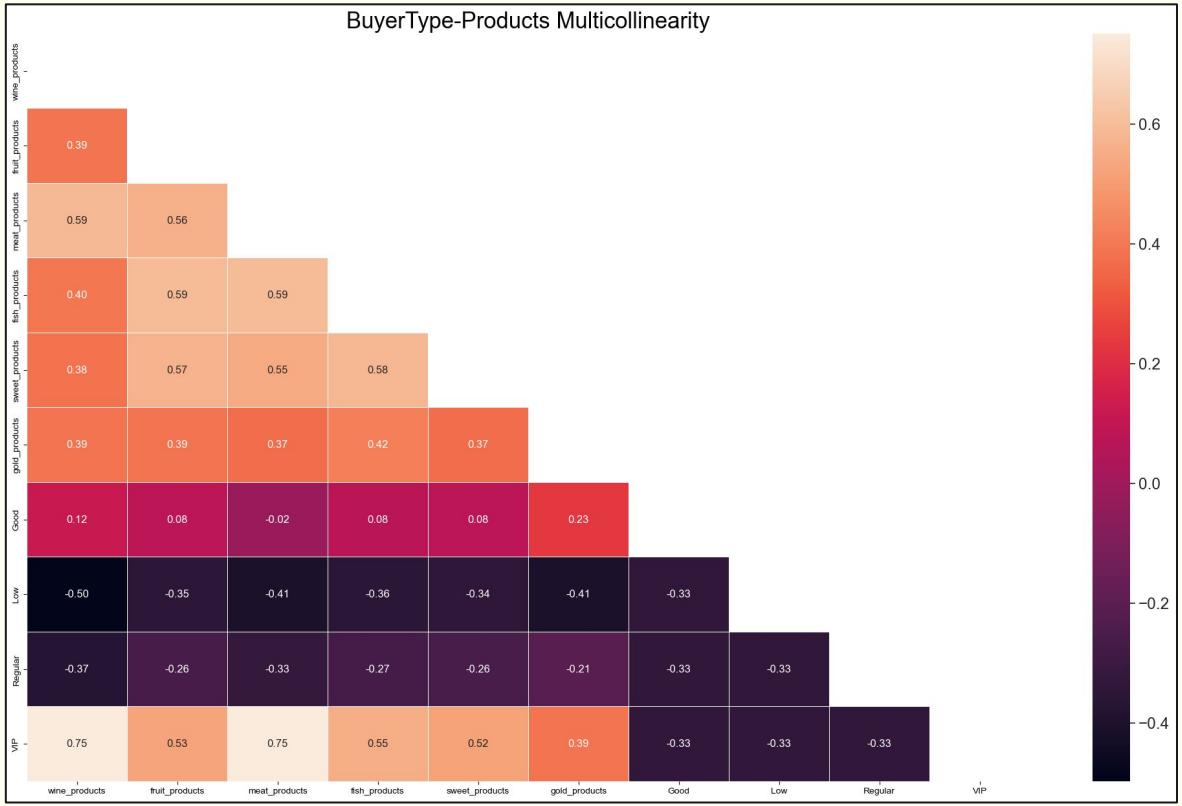


- The overwhelming difference between **Wine** and **Meat** with the other products, added to the usual profile of our customers, reveals the nature of our business: the sale of **gourmet items**

- At this stage of the presentation, with the data observed on the characteristics of our customers and now knowing what their buying preferences are, we can define the type of business: **frequented by high, medium-high well educated class** and focused on the **sale of some main categories** (**Wine** and **Meat**) and other complementary ones

SPECIFIC STATISTICS

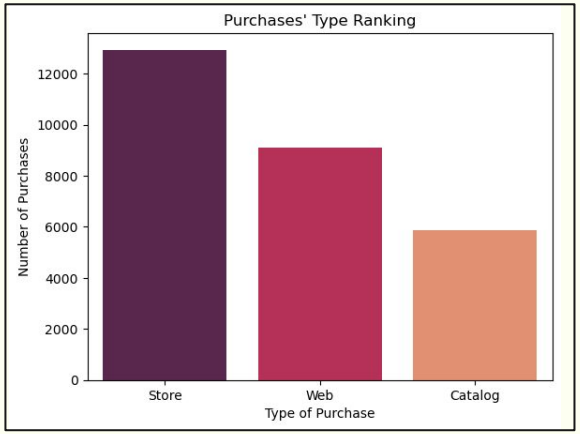
Buyer Type Correlation Products



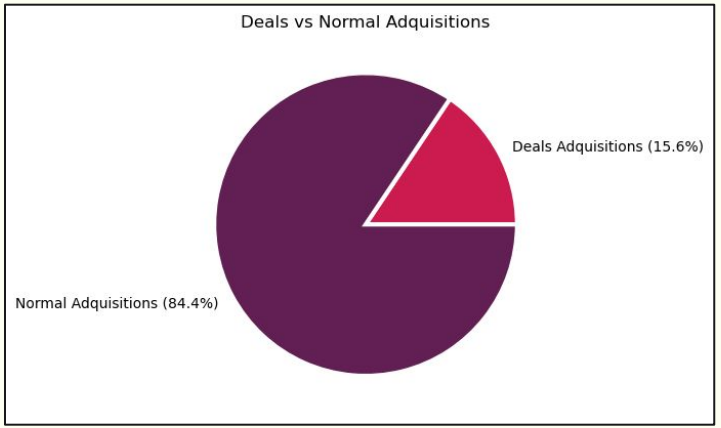
- Our **VIP** clients have a strong correlation with **Wine** and **Meat** Products. Also they have good correlation for other products as well
- For **Good** and **Regular** clients, there's no significant correlations. Good tends to be positive and Regular viceversa, but in both cases there's not enough force.
- **Low** clients have some considerable negative correlation with **Wine**, **Meat** and **Gold**. Evidence of the choice to purchase other products instead
- We observed a significant **positive correlation** (0.59) between **Meat** and **Wine**; **Meat** and **Fish** (0.59) & **Fish** and **Fruit** products (0.59). Good indicator for making **related offers**
- From here, we can suggest Marketing to focus its efforts on promoting **excellence/more expensive Wine** and **Meat** products to **VIP** clients, given their tendency to purchase these products

SPECIFIC STATISTICS

Best Way to Purchase



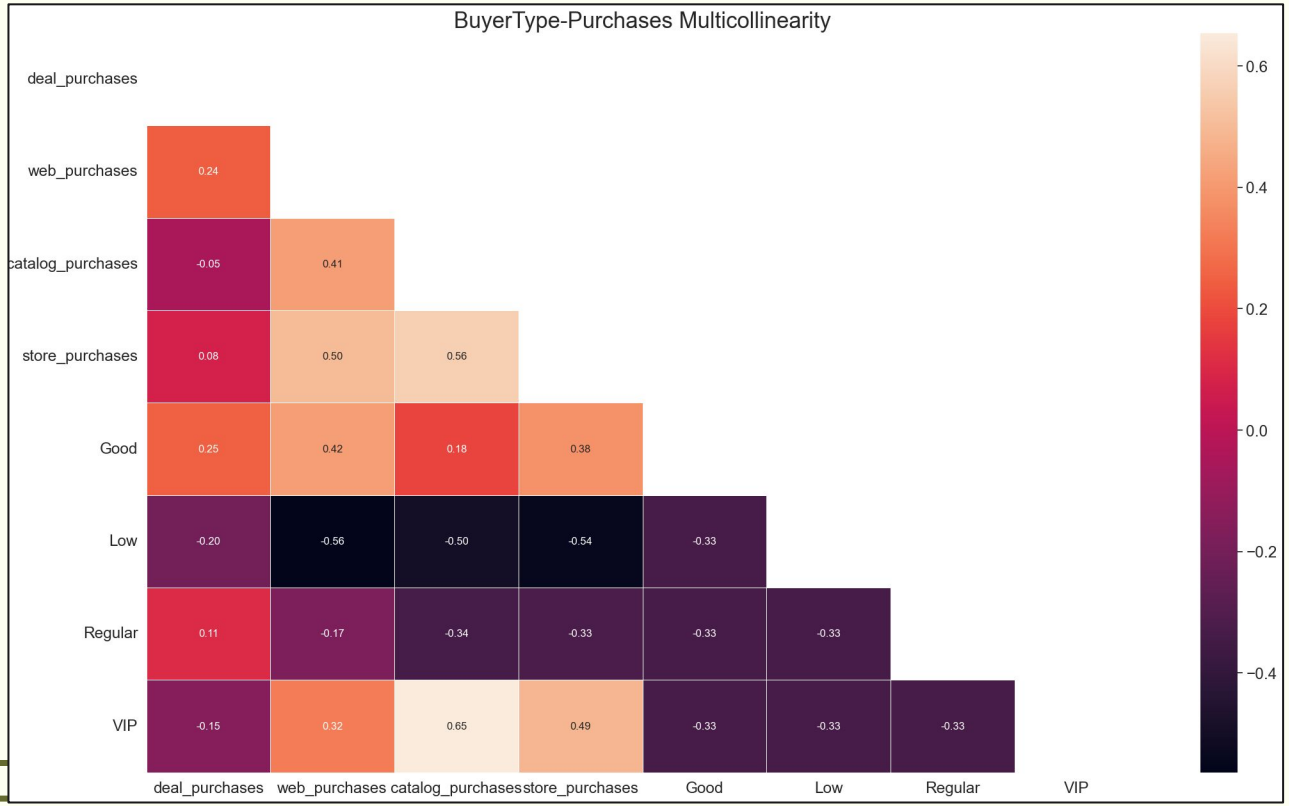
- The graph is clear. The **favourite way** our customers buy **is in person** at the store, although we also have **good numbers via the web**
- However, Catalog also has considerable numbers. Buying quite a bit through **commercial catalog** may be i**ndicative of some interesting habit**. We may see some correlation later that explains this



- **15.6%** of purchases through deals is a very good percentage. The fact that our offers are accepted is a good indication that we should **focus on offering more deals**

SPECIFIC STATISTICS

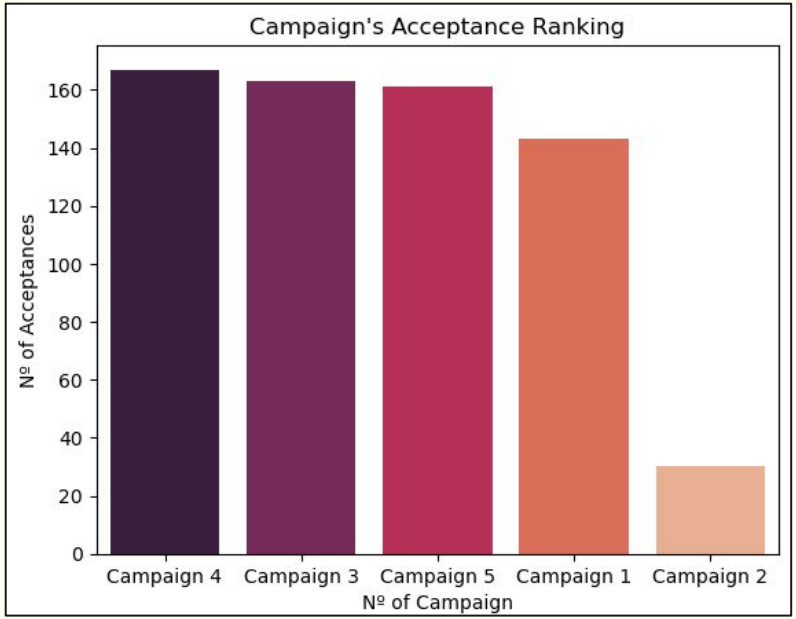
Buyer Type Correlation Purchases



- For our **VIP** customers, the **most common** form of purchase is through **Catalog**, followed by **Store**
- For **Good** customers, there's a tendency to buy through **Web** and **Store**
- For **Regular** customers there's no significant correlation, so we can deduce that they have no predilection for their purchase mode
- For **Low** customers there's a negative tendency to buy trough **Web**, **Catalog** and **Store**.
- As a recommendation, we propose to **redirect retargeting actions** from **Low** customers to the other groups

SPECIFIC STATISTICS

Acceptance of Campaigns




```
Campaign 4: 167 clients of 2225; 7.51% of total clients
Campaign 3: 163 clients of 2225; 7.33% of total clients
Campaign 5: 161 clients of 2225; 7.24% of total clients
Campaign 1: 143 clients of 2225; 6.43% of total clients
Campaign 2: 30 clients of 2225; 1.35% of total clients
```

- **Campaign 4** had the **highest acceptance**, followed closely by 3, 5 and 1
- **Campaign 2** had the **least impact** among subscribers
- Actually, these **conversion rates are quite low**. With **7.51%** as the maximum acceptance rate, we should propose to Marketing run **better segmented campaigns** to increase these metrics

03

BUYER PERSONAS

BUYER PERSONAS

Middle-Age / In a relation / University Grade

- **348** Number of Clients
- **55** Years Old
- **4370.59 $/Month** Salary
- **1** Children at home
- **5** Web visits per month
- **51** Days since last purchase
- **Good** Type of Client
- **623 $** Total spent amount
- **13** Purchases made
- **2** Deals purchased

*This archetype stands out for its high predilection for **Wine** (100000 approx) and **Meat** (60000 approx)*

BUYER PERSONAS

Adult / In a relation / University Grade

- **217** Number of Clients
- **48** Days since last purchase
- **42** Years Old
- **Good** Type of Client
- **3938.33 $/Month** Salary
- **541 $** Total spent amount
- **1** Children at home
- **11** Purchases made
- **6** Web visits per month
- **2** Deals purchased

*The only "Adult" archetype in the top. Its decrease in **Wine** is considerable (less than 50000) but its decrease in **Meat** is less pronounced (20000 less than archetype 1)*

BUYER PERSONAS

Middle-Age / Single / Post-Graduation

- **162** Number of Clients
- **50** Days since last purchase
- **56** Years Old
- **Regular** Type of Client
- **4371.08 $/Month** Salary
- **642 $** Total spent amount
- **1** Children at home
- **13** Purchases made
- **5** Web visits per month
- **2** Deals purchased

*First "Single" to appear. Quite similar to archetype 1. Compared to its counterpart, being only one person, it reduces his consumptions to half. Except for **Wine**: this category is experiencing a lower decline of 40%*

04

MARKETING IMPROVEMENTS

How can we improve?

What we learn thanks to EDA

- We have learned that **Middle-Age**, with respect to age, is the predominant group among all customers with respect to **purchasing power**.
- On **purchasing power**, we also know that **Graduate** is the predominant group with respect to educational level, followed by the **Post-Graduate** groups.
- **Purchasing power** is proportional to the **products purchased** and the **number of purchases**.
- We know exactly the **frequency ranking of our customers** for each combination of their characteristics.
- We know which are the **most popular products**
- We classify customers (**VIP**, **Good**, **Regular**, **Low**) by their spending in our business and know which products are correlated to their tag.
- We know which are the **favorite ways of purchasing**, both in general and specifically by groups.
- We **analyze campaigns** to see what campaigns are **having little or no impact on customers.**
- We build **buyers personas** to know specific habits of the **most representative customers**

How can we improve?

What can we do to improve conversions?

Through the **Zoho suite** we will **update our CRM** with the new variables (**total purchases, total products spent, buyer type**) so that the sales team has more information about them

We know the **consumption habits** of our customers and their **interests**, in addition to having segmented them, so our **email-marketing** campaigns should improve their impact

The **segmented audience** and their **habits** will allow us to better define our campaigns on Meta (**Instragram** and **Facebook**)

We know which **profile to target** more accurately, so our ads on **Google Ads** will be more effective

For the S**EO positioning work** we must work a little more to find the **keywords** related to the **profiles** we have