IFOOD CRM ANALYSIS CASE

Emilia López Reviriego
Ironhack, Data Analytics PT 2023

Description

The dataset contains information regarding a pilot campaing involving **2.240 customers**, customers who bought the offer were properly labeled. The total cost of the sample campaign was **6.720$** and the revenue generated by customers was **3.674$**. Globally the campaing had a profit of -3.046$ and the success of the campaign was **15%**.

The Chief Marketing Officer notifies me as markeyting analyst that recent marketing campaings have not been as effective as they were expected to be. The company wants to produce the highest profit for the next marketing campaign, scheduled for next month.

6 categories of products

Objectives

The objective of the team is to develop a model that predicts the amount that customers are willing to spend and learn more about the customer profile to implement the necessary product and marketing adjustments.

Moreover, other than maximizing the profit of the campaign, the CMO is interested in understanding to study the characteristic features of those customers. The steps are:

- **Data Exploration**
- **Segmentation**
- **Predictive Model of spending amount**

DATA PROCESSING

Cleaning Data

- Data type conversión
- Feature engineering: age, spending, dependents, marital status, total purchases, enrollment customer year...
- Remove missing data and irrelevanrt columns
- Outliers (years, income), EDA

Remove higly correlated columns
- Normalize the data
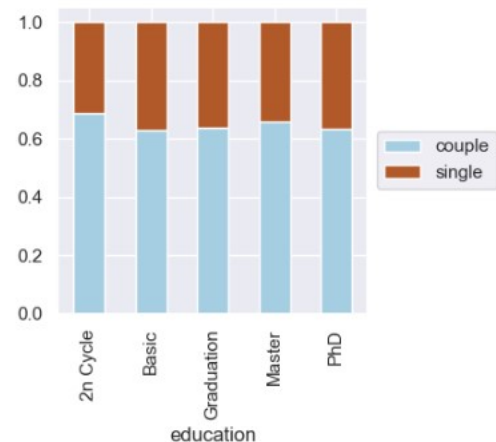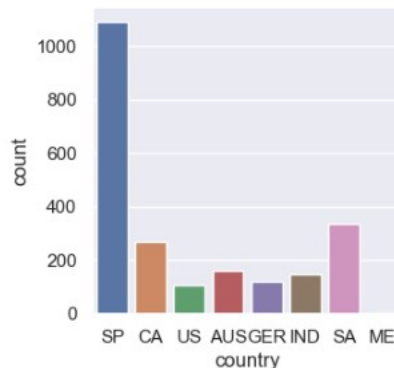- Aplication of label encoding method to categorical variables

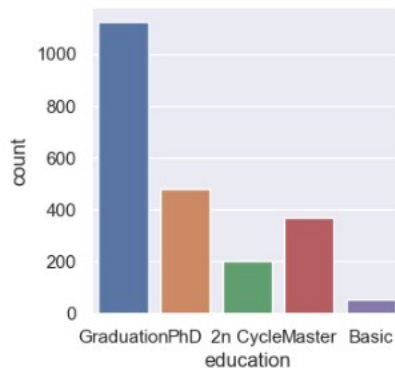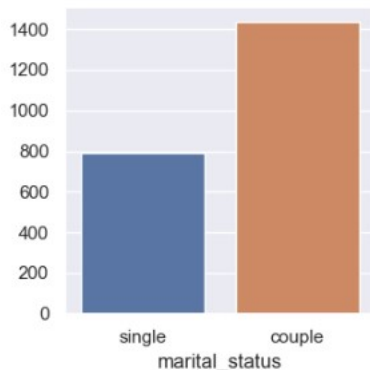

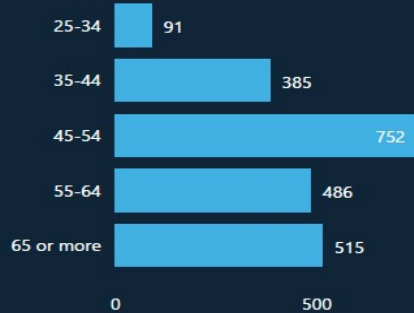CORRELATIONS HEATMAP FOR NUMERICAL DATA
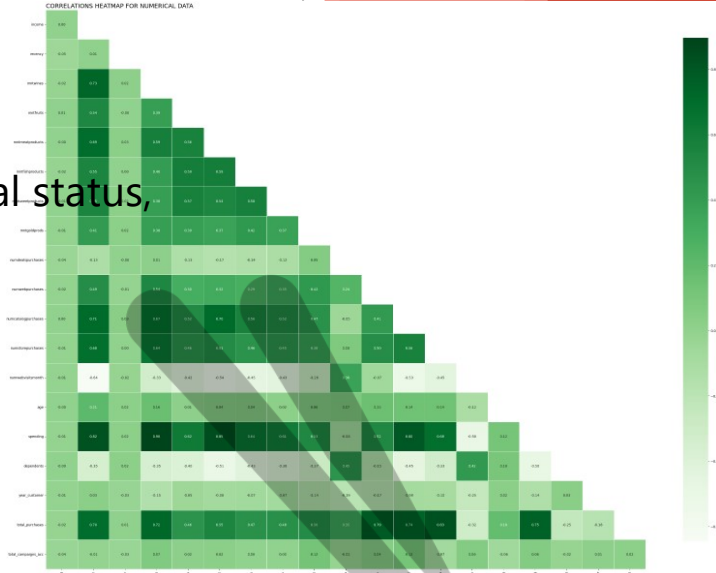
DATA FROM

Customer Profile Analysis

The majority of the customers are couples and highly educated, there is no significant correlation between those two categorical information.
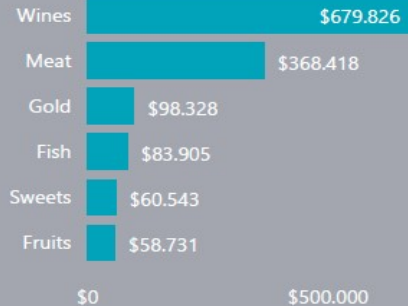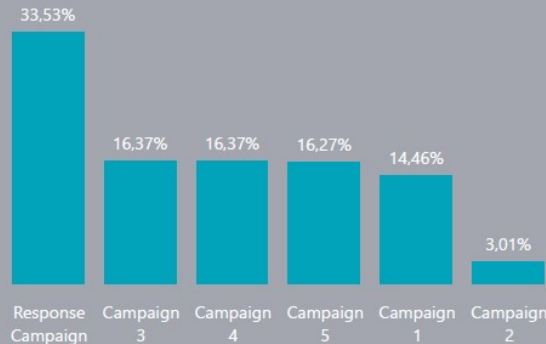
Business Insights

AGE GROUPS

Age Group	Value
25-34	91
35-44	385
45-54	752
55-64	486
65 or more	515

education	Total Customers	Amount Spent	Total Purchases	Amount per purchase
2n Cycle	201	$100.708	2.792	$36,07
Basic	54	$4.417	379	$11,65
Graduation	1.124	$696.828	16.855	$41,34
Master	369	$226.300	5.521	$40,99
PhD	481	$321.498	7.596	$42,32
Total	**2.229**	**$1.349.751**	**33.143**	**$40,73**

SPENT BY PRODUCTS

Product	Amount
Wines	$679.826
Meat	$368.418
Gold	$98.328
Fish	$83.905
Sweets	$60.543
Fruits	$58.731

CAMPAIGN SUCCESS RATE

Campaign	Rate
Response Campaign	33,53%
Campaign 3	16,37%
Campaign 4	16,37%
Campaign 5	16,27%
Campaign 1	14,46%
Campaign 2	3,01%

CHANNEL PERFORMANCE

Channel	Value
Store	12.956
Web	9.139
Catalog	5.877
Deals	5.171

Customers Segmentation: K- means

Customers Groups
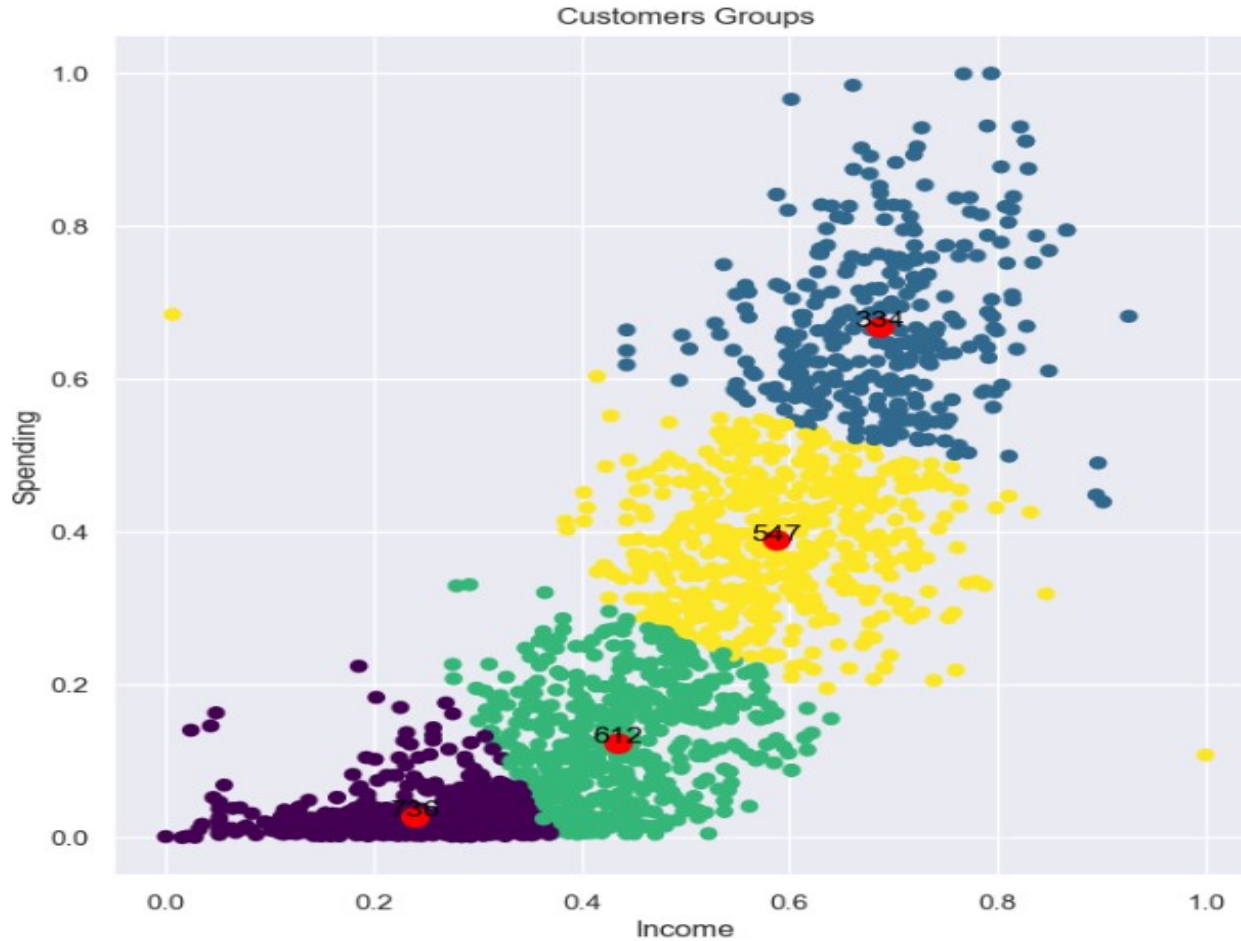
Customers Segmentation Statistics:

	Cluster 0 - Low Income	Cluster 1 - Average Income	Cluster 2 - High income	Cluster 3 - Top Income
Number of clients	736	612	547	334
Average Income	28,541$	50,399$	67,507$	78,690$
Average Expenses	72$	315$	985$	1,688$
Range Income	1,730 to 43,269$	32,632 to 73,395$	2,447$ to 113,734$	51,381$ to 105,471$
Range Expenses	5$ to 570$	16$ to 839$	277$ to 1,730$	1.112$ to 2,525$
Distribution Income	29,968$	49,973$	66,886$	79,414$
Distributipon Expenses	53$	307$	988$	1634$

Notes:

- Variability: The standard deviation of all clusters indicates a moderate degree of variability, suggesting that there are clients with different income and spending patterns within this cluster.

Features Correlation Analysis

- **Income** is a proxy for several other features, such as the amount spend, positively driven by meat and wine and it has a negative correlation with visits on the websites and dependents.

- The **amount spend on Wine** is, besides being related to high income, to the amount spend on Meat and it's purchased or in Catalog or in Stores.

- The **number of kids** is negative related to income(-0,35), amont spend in total and, consequently, related to wine.

Customer's Income	
Total_purchases	0,70
Spending	0,82
MntWines	0,73
MntMeatProducts	0,69
Dependents	-0,35
NumWebVisitsMonth	-,064

Amount Spend on Wines	
Income	0.73
MntMeatProducts	0.59
Spending	0.90
NumCatalogPurchases	0.67
NumStorePurchases	0.64

Dependents	
Numdealspurchases	0,45
NumWebVisitsMonth	0,42
Spending	-0.50
MntWines	-0.35
NumCatalogPurchases	-0,45
NumStorePurchases	-0.33

LINEAR REGRESSION MODELS

INCOME/SPENDING

R2: 0,67849190
MSE: 118045,35
MAE: 268,468
RMSE: 343,577

CAT. &NUMERICAL

R2: 0,8156524
MSE: 68735,41
MAE: 188,13
RMSE: 262,17

Recommendations to the CMO

Conclusion 1: The most recent advertising campaign was the most successful.
Action: Replicate the successful model used in the recent campaign for future advertising campaigns.

Conclusion 2: Spending is influenced by income and the presence of children or teenagers, with a positive correlation to income and a negative correlation to having kids/teens.
Action: Create two targeted advertising streams - one targeting high-income individuals without kids/teens and another targeting lower-income individuals with kids/teens.

Conclusion 3: Wines and meats are the most successful products, as customers spent the most on these items on average.
Action: Focus advertising campaigns on increasing sales of the less popular items.

Conclusion 4: Deals and catalog purchases are underperforming channels, as customers made the fewest purchases through these channels on average.

Conclusion 5: Web and store purchases are the best performing channels, as customers made the most purchases through these channels on average.
Action: Prioritize advertising campaigns on the more successful channels to reach a larger customer base.