

Customer Profile

Group Name: Customer Profile

Group Members:

First name	Last Name
Aiswarya	Prabhalan
Bini	Abraham
Darling	Oscanoa
Shreyash	Banawala
Suruchi	Pokhrel

Submission date: *2024-04-15*

Table of Contents

List of Figures	3
Abstract.....	4
1. Introduction	5
1.1. Overview	5
1.2. Goals	5
2. Methods.....	6
2.1. Dataset Used	6
2.2. Libraries Used.....	8
2.3. Data Preprocessing	8
Missing values	8
Unique values.....	9
EDA (Exploratory Data Analysis).....	9
2.4. Feature Engineering.....	12
Transform Year_Birth to Age	12
Other transforms:.....	12
3. Results.....	13
Feature Engineering II:	17
Outliers:.....	17
DBSCAN Clustering:	17
4. Conclusion and Future Work	20
5. References	21

List of Figures

Figure 1: Marketing Mix, 4Ps	6
Figure 2: Dataset provided.....	7
Figure 3: Raw DataSet just after imported	7
Figure 4: Quantity of records per subgroup in numerical columns	9
Figure 5: Histogram for all columns	10
Figure 6: Boxplot for all columns	11
Figure 7: Transform Year_Birth to Age and Encoding by ranges	12
Figure 8: Question a, first approach combining the results	13
Figure 9, Question a, first approach intersection of the results	14
Figure 10, Question a, second approach RAFT Marketing Technique	14
Figure 11, Question b, RFM values and intersection	15
Figure 12, Question b, Histogram of components.....	15
Figure 13, Question c, low-value customers.....	16
Figure 14, Question e, DBSCAN Clustering for the 4 groups, found it.....	18
Figure 15, Question e, DBSCAN Clustering Income vs Frequency	18
Figure 16, Question f, Characteristics of highest meat buyers.....	19
Figure 17, Question f, Meat Product by Age Range and Marital Status	19

Abstract

The retail internet sector has experienced major transformations as it navigates the changing terrain of marketing tactics. The industry has shifted over the past few decades from product-focused Marketing 1.0 (4 Ps: Product, Place, Price, and Promotion) to customer-focused Marketing 5.0, which now strongly emphasizes providing outstanding customer experiences. According to Digital Commerce 360, with more than 46% of the companies spending more than 50% of their total marketing budget on digital marketing by 2023, companies are still investing to gain bigger profits and stay competitive in this evolving retail landscape.

This report shows the company's targeting and retention efforts of consumer data presented in this report. Data on consumer demographics, purchase patterns, and campaign engagement are all included in the dataset.

The analysis reveals several important insights. First, the business should focus on attracting new clients who have high salaries, have completed college, and have a track record of buying pricey goods like beef, gold, and wine. It's conceivable that these "Premium Buyers" will accept premium offerings. Customers who typically buy inexpensive, basic products and have low incomes and levels of education, on the other hand, are likely to gain little from targeted initiatives.

Additionally, the business can always count on its current "Loyal Diversified Buyers" base or clients with frequent recent purchases spanning multiple product categories. The focus of retention initiatives must be these dependable sources of income.

Furthermore, the research identifies a subset of "Emerging Buyers"—new clients with inconsistent purchasing patterns. Before making a significant acquisition investment, the business should thoroughly assess the potential lifetime worth of these clients.

All things considered, the analysis's conclusions can assist the business in refining its methods for attracting new clients and keeping existing ones so that it can concentrate on the most lucrative and promising clientele. The company can drive sustainable growth and maximize its return on investment by aligning its marketing and product development efforts with the identified customer characteristics and behaviors.

1. Introduction

1.1. Overview

The company's target and retention efforts are informed by the thorough analysis of consumer data presented in this report. Data on consumer demographics, purchase patterns, and campaign engagement are all included in the dataset.

Numerous important insights are revealed by the investigation, which can aid the business in maximizing its efforts to attract and retain customers. The business can promote long-term growth and optimize return on investment by matching its marketing and product development tactics with the traits and actions of its target market.

1.2. Goals

This analysis's main objectives are:

- ✓ Determine which new customer segments the business should focus on selling its expensive products to.
- ✓ Ascertain which clients the business can depend on to provide continuous revenue.
- ✓ Determine which current clients are not likely to gain much from targeted efforts so that the business can concentrate its resources on other promising markets.
- ✓ Identify new clients who need to have their long-term potential carefully assessed to avoid them becoming unproductive or with low value.
- ✓ Divide the consumer base into discrete groups according to how they spend their money and learn about the traits of the most ardent consumers of particular product categories, like meat.

By focusing on these objectives, the business may create more successful client acquisition and retention plans, resulting in long-term profitability and growth.

2. Methods

In this project, our "Online retail company" provided us with a Dataset where they recorded all their customer information stored under traditional Marketing from 1.0 to 3.0 format (The 4Ps of Marketing by E. Jerome McCarthy in his book "Basic Marketing: A Managerial Approach"), which is Product Oriented. Now, our online retail company wants to move forward by adding some customer features when defining new target products called "Marketing Mix."

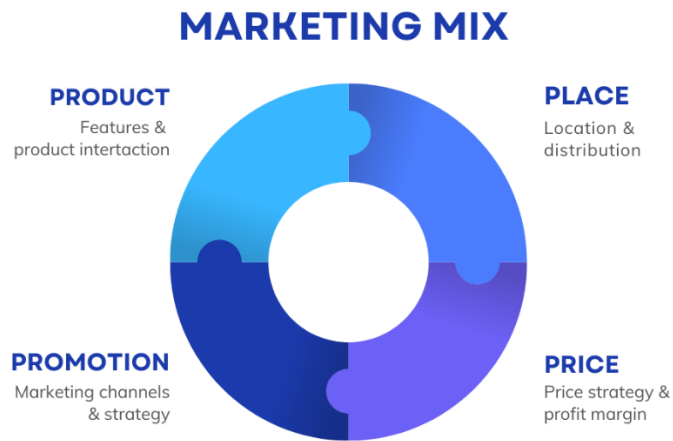


Figure 1: Marketing Mix, 4Ps

2.1. Dataset Used

The dataset provided for this analysis includes a comprehensive set of customer information covering demographics, purchasing behavior, and engagement with marketing campaigns. The key variables in the dataset are described in Figure 1. As follow:

Field in 4Ps	Features	Description
Customers	ID	Customer ID
	Year_Birth	Year of birth of the customer
	Education	Education level of the customer
	Marital_Status	Marital status of the customer
	Income	Income level of the customer
	Kidhome	Number of children in the customer's household
	Teenhome	Number of teenagers in the customer's household
	Dt_Customer	Date the customer became a customer
Product	Recency	Number of days since the customer's last purchase
	MntWines	Amount spent on wine in the last 2 years
	MntFruits	Amount spent on fruits in the last 2 years
	MntMeatProducts	Amount spent on meat products in the last 2 years
	MntFishProducts	Amount spent on fish products in the last 2 years
	MntSweetProducts	Amount spent on sweet products in the last 2 years
Place	MntGoldProds	Amount spent on gold products in the last 2 years
	NumDealsPurchases	Number of purchases made with discount
	NumWebPurchases	Number of purchases made through the company's website
	NumCatalogPurchases	Number of purchases made using a catalog
Promotion	NumStorePurchases	Number of purchases made directly in stores
	NumWebVisitsMont	Average number of visits to the company's website in the last month
Customers	AcceptedCmp3	1 if the customer accepted the offer in the 3rd campaign, 0 otherwise
	AcceptedCmp4	1 if the customer accepted the offer in the 4th campaign, 0 otherwise
	AcceptedCmp5	1 if the customer accepted the offer in the 5th campaign, 0 otherwise
	AcceptedCmp1	1 if the customer accepted the offer in the 1st campaign, 0 otherwise
	AcceptedCmp2	1 if the customer accepted the offer in the 2nd campaign, 0 otherwise
Customers	Complain	1 if the customer complained, 0 otherwise
	Z_CostContact	Cost of contacting the customer
	Z_Revenue	Revenue generated by the customer
Promotion	Response	1 if the customer responded to the last campaign, 0 otherwise

Figure 2: Dataset provided

This comprehensive dataset allows for a detailed analysis of customer demographics, purchasing behavior, and engagement, enabling the company to develop targeted customer acquisition and retention strategies.

Before starting any data preprocessing or EDA, Figure 2 shows a quick view of the data type for all columns.

```
[12]: df.head(10)
```

	Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Complain	Z_CostContact	Z_Revenue	Response
1957	Graduation	Single	58138.0	0	0	2012-09-04	58	635	...	7	0	0	0	0	0	0	0	3	11	1
1954	Graduation	Single	46344.0	1	1	2014-03-08	38	11	...	5	0	0	0	0	0	0	0	3	11	0
1965	Graduation	Together	71613.0	0	0	2013-08-21	26	426	...	4	0	0	0	0	0	0	0	3	11	0
1984	Graduation	Together	26646.0	1	0	2014-02-10	26	11	...	6	0	0	0	0	0	0	0	3	11	0
1981	PhD	Married	58293.0	1	0	2014-01-19	94	173	...	5	0	0	0	0	0	0	0	3	11	0
1967	Master	Together	62513.0	0	1	2013-09-09	16	520	...	6	0	0	0	0	0	0	0	3	11	0
1971	Graduation	Divorced	55635.0	0	1	2012-11-13	34	235	...	6	0	0	0	0	0	0	0	3	11	0
1985	PhD	Married	33454.0	1	0	2013-05-08	32	76	...	8	0	0	0	0	0	0	0	3	11	0
1974	PhD	Together	30351.0	1	0	2013-06-06	19	14	...	9	0	0	0	0	0	0	0	3	11	1
1950	PhD	Together	5648.0	1	1	2014-03-13	68	28	...	20	1	0	0	0	0	0	0	3	11	0

Figure 3: Raw DataSet just after imported

2.2. Libraries Used

For this project we used the basic python libraries:

Description:

- ✓ pandas: Allows to make all Dataframe treatment
- ✓ matplotlib.pyplot: Provides support for build graphs during the lifecycle project.
- ✓ seaborn: Offers sub-modules to create specific graphs such as histogram, boxplot or violin.
- ✓ sklearn.impute: Helps with KNNImputer to estimate and fill up missing values.
- ✓ sklearn.preprocessing: Helps with LabelEncoder to encode the categorical variables that can then be used by machine learning models.
- ✓ numpy: Used for working with arrays and functions in linear algebra.

2.3. Data Preprocessing

Before the data can go to the machine learning algorithms, it needs to be prepared in advance. This step helps to clean and this project is divided into 2 groups of questions for the online retail company:

- ✓ Analytical questions (Questions 1 to 4)
- ✓ Cluster segmentation for new target products (Question 5)

We are going to do the Data Processing for the part I first

Missing values

The dataset has 2240 records, and the only column with missing values is the Income, with 24 missing values, as that is less than 1% of the records we could quickly delete. However, we used KNNImputer and recorded in a list those 24 Customer IDs that we imputed their values to use later. Online retail does not want to send a campaign for a customer "ID" with an Imputed value

Unique values

After reviewing the unique values, some columns have the same value for all records, not providing any information such as: “Z_CostContact”, “Z_Revenue”

From other values, we also recorded the quantity per any subgroup inside those unique values, as seen in Figure 4.

Frequency of values in 'Education':		Frequency of values in 'Marital_Status':	
Education		Marital_Status	
Graduation	1127	Married	864
PhD	486	Together	580
Master	370	Single	480
2n Cycle	203	Divorced	232
Basic	54	Widow	77
Name: count, dtype: int64		Alone	3
		Absurd	2
		YOLO	2
		Name: count, dtype: int64	

Figure 4: Quantity of records per subgroup in numerical columns

EDA (Exploratory Data Analysis)

To perform EDA in all columns with almost raw data, the project plots using matplotlib and seaborn libraries to release Histogram, Violin, and Boxplot. The results are shown in Figures 5 and 6

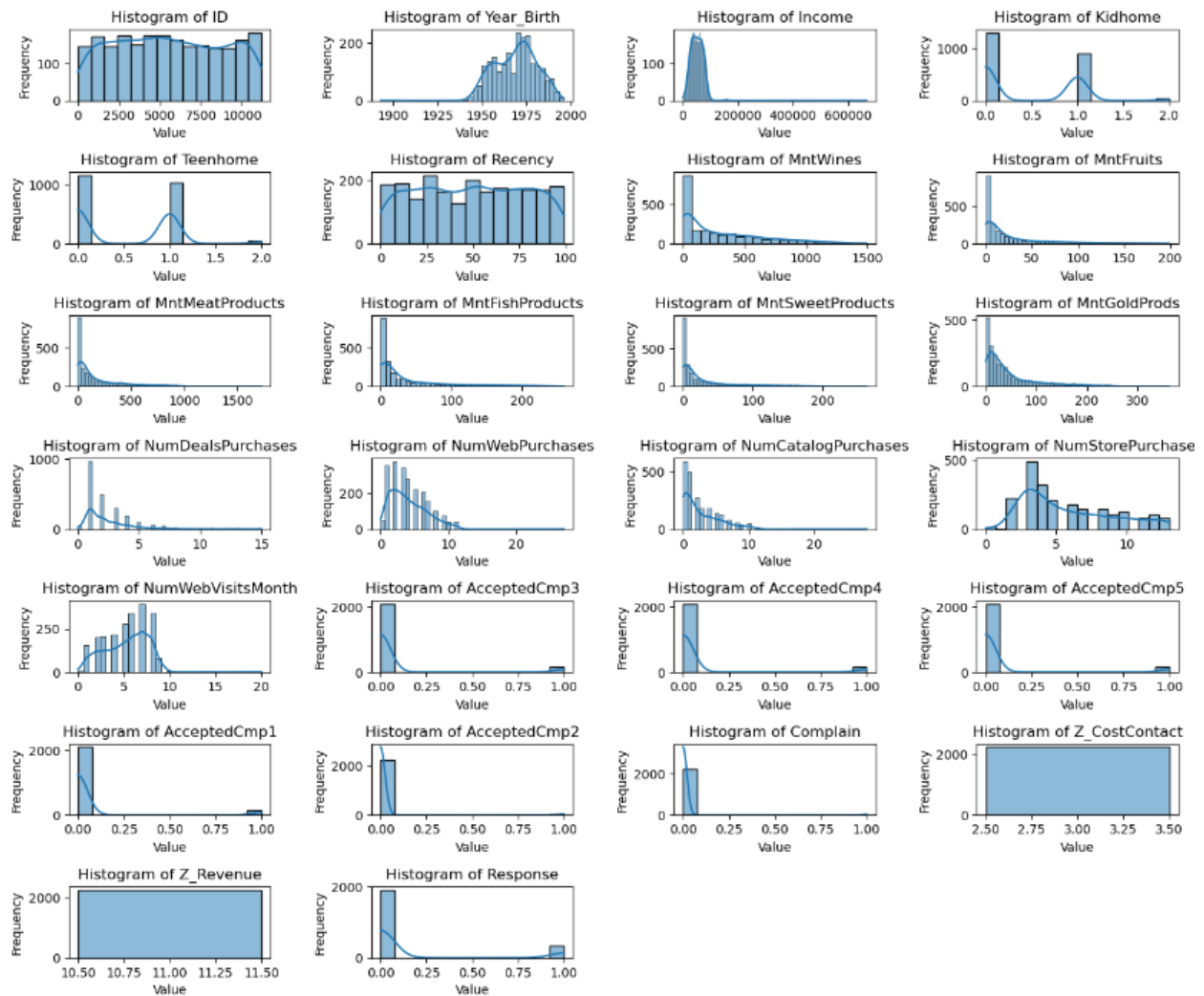


Figure 5: Histogram for all columns

From the Histogram, it is seen:

- The most amount spent is on Wine and GoldProducts, not surprisingly also MeatProducts
- Recency is almost flat, so is almost flat.
- The ones who Accepted any campaign are very few in each category, so is better to combine them to be more significant

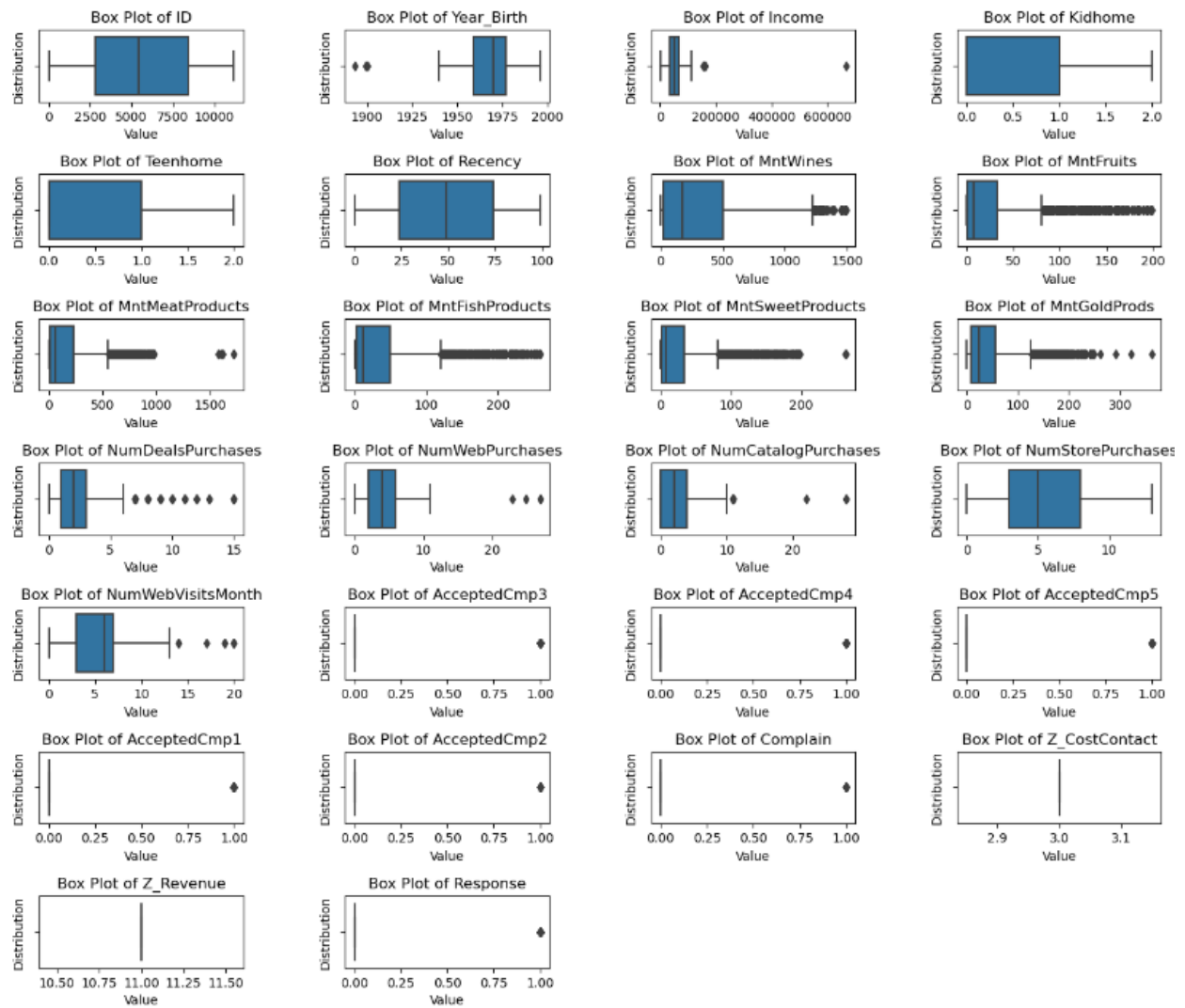


Figure 6: Boxplot for all columns

From Violin:

-There are many outliers for Year_Birth, Income, and all amounts in products and purchases.

From Boxplot:

-Z_CostContact and Z_Revenue are constants, which is better to delete as those don't provide information.

2.4. Feature Engineering

The online retail company still needs to provide the current date. However, we have information on dates for the feature "Dt_Customer" from 2012 until 2014, and this project, to be productive and concise, needs updated data.

So, we can define for this study that today is 2014-12-31, and we started planning for the following year.

Transform Year_Birth to Age

Considering the current date assumption, we transformed the column "Year_Birth" to "Age" and encoded it to have it for ranges. Figure 7 shows the ranges defined.

```
import datetime
today = datetime.date(2014, 12, 31)
df['Age'] = today.year - df['Year_Birth']

df['Age_Range'] = df['Age'].apply(lambda x:
                                0 if x <= 18 else # Child
                                1 if x <= 24 else # Youth
                                2 if x <= 34 else # Young Adult
                                3 if x <= 44 else # Adult
                                4 if x <= 54 else # Middle-Aged
                                5 if x <= 64 else # Pre-Senior
                                6)                # Senior

df = df.drop(['Year_Birth', 'Age'], axis=1)
df.head(10)
```

Figure 7: Transform Year_Birth to Age and Encoding by ranges

Other transforms:

For "Education": As there are 5 levels of education and are categorical, we are going to encode those

For "Marital_Status": There are 7 values, then we used LabelEncoder from the scikit-learn library.

For "Dt_Customer": As Dt_Customer is the date when a customer becomes a customer, the number of days being a customer rather than the date is most important.

Recreate the column Z_Revenue

As in steps before, the "Z_Revenue" was deleted because of their flat value; in this step, we calculate taking into consideration all amount spent on it

3. Results

In this Stage, we are going to cover the main objective of this analysis by answering the questions that business has:

a) Which new customers can the company target for high-cost products?

The first approach to answer this question started by looking for people with high income (more than the q3), then analyzing purchase behavior by filtering only beyond the q3 for the most "luxury" products we have as "Meat", "GoldProduts" and "Wine", looking for the top spenders; then by also taking the top visitor to the website. Finally, the results are combined to remove the intersection, as shown in Figure 8.

```
high_web_visits_customers = high_web_visits_customers['ID'].tolist()
high_income_customers = high_income_customers['ID'].tolist()
high_spending_customers = high_spending_customers['ID'].tolist()

high_value_customers = list(set(high_web_visits_customers) & set(high_income_customers) & set(high_spending_customers))

#We will remove any high-value customers that coincide with the "list_imputed"
high_value_customers = [cid for cid in high_value_customers if cid not in list_imputed]

print("Customer IDs of High-Value Customers:")
print(high_value_customers)

Customer IDs of High-Value Customers:
[7441, 6566, 4910, 8755, 5299, 10678, 1079, 5831, 10057, 2379, 4299, 6606, 3667, 7899, 6749, 5341, 3426, 5989, 4070, 10727, 2535, 10473, 9451, 4974, 7919, 6384, 10992, 9212]

len(high_value_customers)

28
```

Figure 8: Question a, first approach combining the results

We found 28 customers with their customer "IDs" that, after comparing with the Imputed values for Income, don't have the same value, so we can conclude that this is a good list to show to the business.

To add, by reviewing the details of the intersections as shown in Figure 9, we can tell the business by taking the customers with more than 75% in each field, we have those 28 customer IDs, so if the business wants a bigger list, we could reduce until 60% and re-run the model to find more customer IDs as a 2nd round of the campaign depending of the result from the previous one.

For the 2nd round, we can take those 163 customers with High Web Visits and High Spending; then we could take those 491 customers with High Income and High Spending, as seen in Figure 9

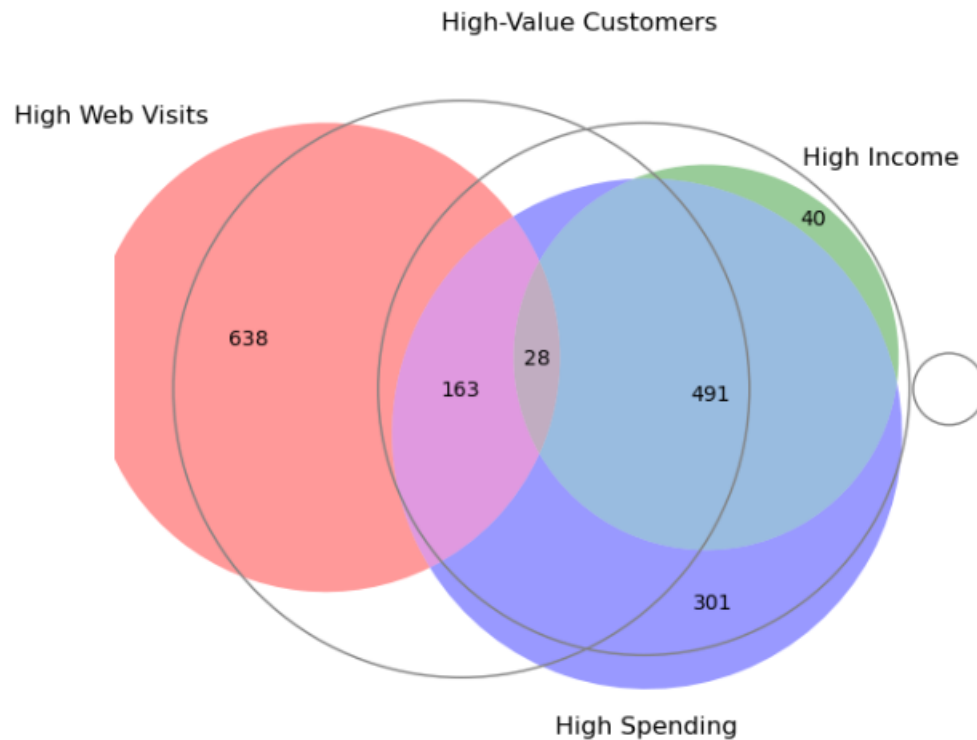


Figure 9, Question a, first approach intersection of the results

A 2nd approach or technique is well-known in the marketing field called RAFT (Recency, Amount, Frequency, Tenure) to find highly valuable customers. This technique assesses clients' monetary value (how much money they spend), frequency (how often they make purchases), and recency (how long ago they made a purchase). Combining all together, as shown in Figure 10.

```
income_75 = df['Income'].quantile(0.75)
web_visits_75 = df['NumWebVisitsMonth'].quantile(0.75)
accepted_cmp_75 = df[['AcceptedCmp1', 'AcceptedCmp2', 'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5']].sum(axis=1).quantile(0.75) #Now we are considering the previous Accepted marketing campaign
recency_25 = df['Recency'].quantile(0.25) #Now we are taking only the most recent buyers (the lower quartile)

#Now combining all of them:
high_value_new_customersRAFT = df[(df['Income'] >= income_75) &
    (df['NumWebVisitsMonth'] >= web_visits_75) &
    (df[['AcceptedCmp1', 'AcceptedCmp2', 'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5']].sum(axis=1) >= 2) &
    (df['Recency'] <= recency_25)]

# Filtering the imputed
target_customer_idsRAFT = high_value_new_customersRAFT['ID'].tolist()
target_customer_idsRAFT = [cid for cid in target_customer_idsRAFT if cid not in list_imputed]
target_customer_idsRAFT

[3667, 2535]
```

Figure 10, Question a, second approach RAFT Marketing Technique

We can see the list reduced because now “Recency” was taken into consideration, and with that, the output we got was only 2 customer IDs (included in the first approach, too).

b) Which old customers have the company consistently relied on?

According to the Clientbook marketing study, one key aspect of effective retail customer segmentation is identifying the company's "old customers" - those with a history of repeat purchases and high total spending. The study suggests that companies should focus on these loyal, high-value customers as they are the ones the business has consistently relied upon over time.

Some key characteristics of the "old customers" that companies should target include:

- High Recency, Frequency, and Monetary (RFM) values
- Customers who are part of the company's loyalty program
- Customers with a history of repeat purchases and high total spending

```
# Recency (R)
high_recency_customers = df[df['Recency'] <= df['Recency'].quantile(0.25)]
print("R:", len(high_recency_customers))

# Frequency (F)
high_frequency_customers = df[(df['NumDealsPurchases'] >= df['NumDealsPurchases'].quantile(0.75)) &
(df['NumWebPurchases'] >= df['NumWebPurchases'].quantile(0.75)) &
(df['NumCatalogPurchases'] >= df['NumCatalogPurchases'].quantile(0.75)) &
(df['NumStorePurchases'] >= df['NumStorePurchases'].quantile(0.75))]
print("F:", len(high_frequency_customers))

# Monetary Value (M)
high_monetary_customers = df[df['Z_Revenue'] >= df['Z_Revenue'].quantile(0.75)]
print("M:", len(high_monetary_customers))

# Tenure (T) #We ran first with 75% percentile but the result was only 3 records, so then iterate until convergence got that with 60% percentile
high_tenure_customers = df[df['Dt_Customer_Days'] >= df['Dt_Customer_Days'].quantile(0.6)]
print("T:", len(high_tenure_customers))

# The Intersection
target_customers_b = high_recency_customers.merge(high_frequency_customers, on='ID', how='inner', suffixes=('_', '_f'))
target_customers_b = target_customers_b.merge(high_monetary_customers, on='ID', how='inner', suffixes=('_', '_m'))
target_customers_b = target_customers_b.merge(high_tenure_customers, on='ID', how='inner', suffixes=('_', '_t'))
print("Intersection (R, F, M, T):", len(target_customers_b))
target_customers_b

#Loyalty program
loyalty_customers = df[df['AcceptedCmp1'] == 1]

R: 567
F: 58
M: 560
T: 899
Intersection (R, F, M, T): 7
```

Figure 11, Question b, RFM values and intersection

As shown in the Figure 11 we found 7 loyal customers to share with the business, we can also see in the Figure 12 how those values interact

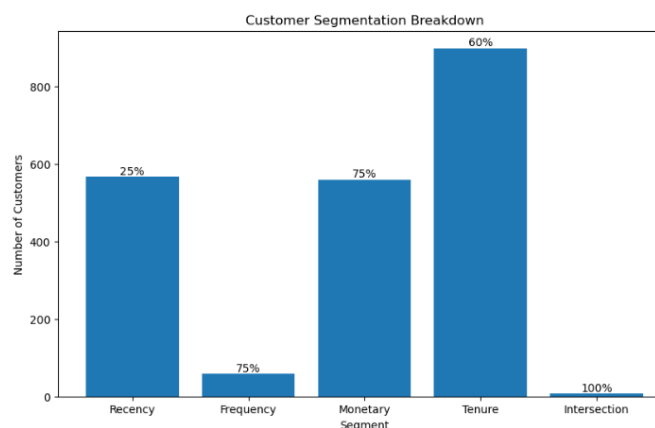


Figure 12, Question b, Histogram of components

c) For which old customers are targeting efforts, not going to yield much benefit?

Another concern for the business is determining which clients are likely to gain little from targeted efforts so that the business can concentrate its resources on other promising markets.

With the same concept from Clientbook but opposite values to the previous question, we need to look for the following:

- Customers with Low Recency, Frequency, and Monetary (RFM) values
- Customers who have not purchased in a long time (low Recency)
- Customers with low total spending and purchase frequency

In this case, we found 74 customer IDs with low value to report immediately to the business as shown in Figure 13

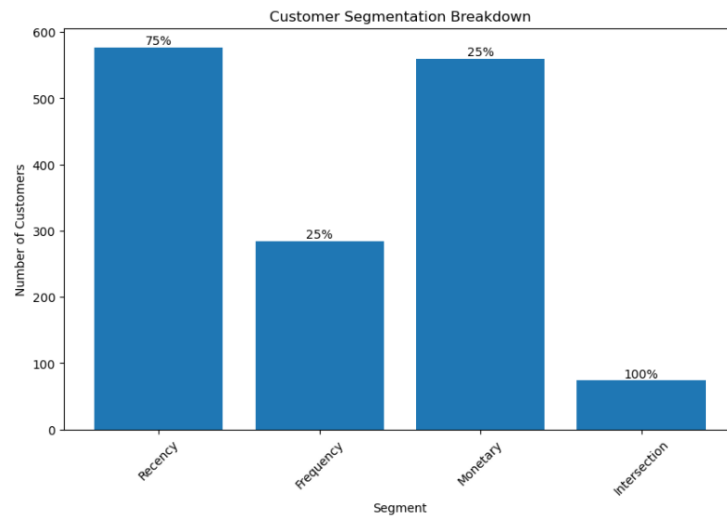


Figure 13, Question c, low-value customers

d) Which new customers are in danger of going into territory-c?

Another concern from the business is detecting new clients who must carefully assess their long-term potential to avoid them becoming unproductive or with low value and take action immediately. In this case, If we would like to identify new customers at risk of becoming low-value:

- Customers with high Recency (recent first purchase) but low Frequency and Monetary values
- Customers who have not responded to previous marketing campaigns (low AcceptedCmp values)
- Customers with low website engagement (low NumWebVisitsMonth)

By running our model, we found 17 new customers with the risk of going to territory-c

e) Using 'spending' behavior, classify the customers into at least 4 categories

Finally, the business wants to know how they can divide the consumer base into discrete groups according to how they spend their money and learn about the traits of the most ardent consumers of particular product categories, like meat.

We need to continue with feature engineering and deal with outliers to do that.

Feature Engineering II:

We calculate the "MntTotal" by adding all spent per customer, then calculate the "TicketAverage" by dividing the previous one against "NumAllPurchases." We also calculate the "TotalAcceptedCmp" by adding all accepted previous campaigns; we also remove unnecessary columns like "ID" in this case

Outliers:

As discussed before, we saw many outliers in "Income" and "TicketAverage." We apply the flooring and capping techniques with 10% and 90% percentiles, respectively.

DBSCAN Clustering:

Finally, we used DBSCAN to make our clustering and after some interaction, we see the best names for our clusters are:

-Outliers (-1): This cluster represents customers who are significantly different from the rest of the customer base and may need to be analyzed separately.

-High-Value Customers (0): These customers are likely to have a high income, a large number of household members (kids and teens), and high spending across various product categories (wines, fruits, meat products, fish products, sweet products, and gold products). They are the most valuable customers for the business.

-Moderate-Value Customers (1): These customers have moderate income, household size, and spending patterns compared to the high-value customers.

-Low-Value Customers (2): These customers have lower income, smaller household size, and lower spending across the various product categories than high-value and moderate-value customers.

As shown in Figure 14, the outlier's group is dispersed in all data points. However, for other groups, there are some areas of the scatter plots that they move between.

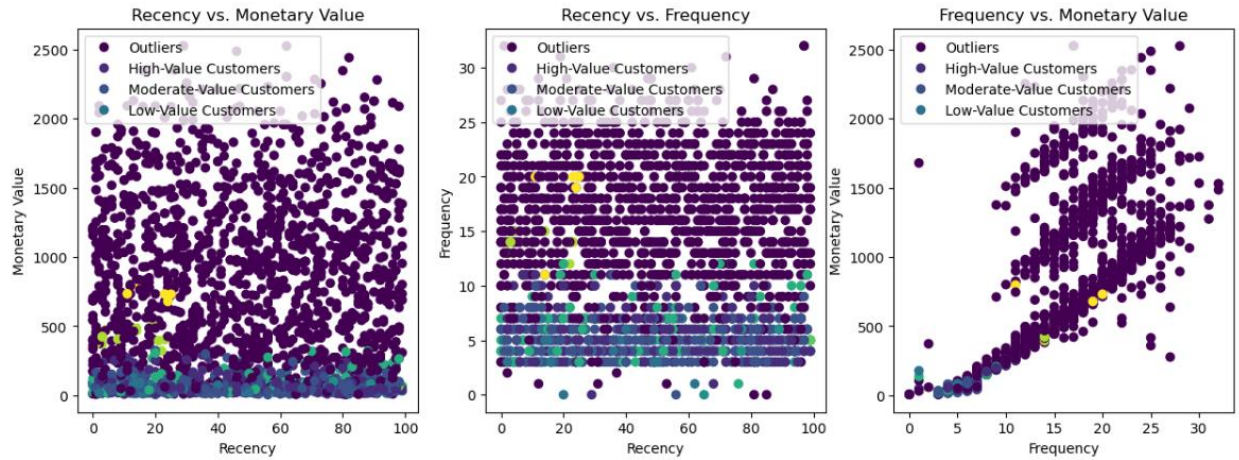


Figure 14, Question e, DBSCAN Clustering for the 4 groups, found it

As expected, low-value customers are located on low incomes. However, there are some high-value customers with low incomes, as shown in Figure 15.

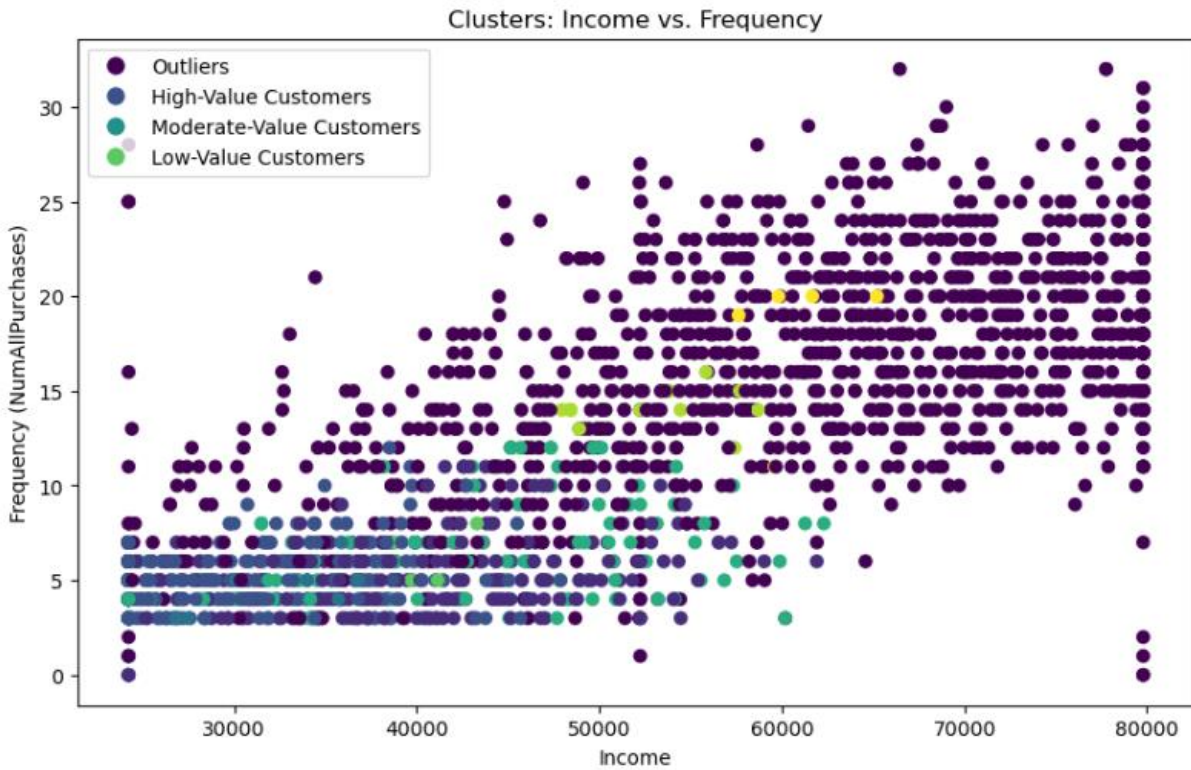


Figure 15, Question e, DBSCAN Clustering Income vs Frequency

f) What are the characteristics of customers who are the highest buyers of Meat products?

To answer this question, we will take a rank, let's say for 100th top highest buyers (we are not considering the outliers). Then, we will take the mean as a single value to represent how those high meat buyers are, as shown in Figure 16.

```

Characteristics of the highest buyers of meat products:
Average Income: 76614.59577964456
Average Kidhome: 0.015384615384615385
Average Teenhome: 0.03076923076923077
Average Recency: 52.207692307692305
Average MntWines: 674.1769230769231
Average MntFruits: 72.13846153846154
Average MntFishProducts: 95.04615384615384
Average MntSweetProducts: 70.4
Average MntGoldProds: 71.46923076923076
Average Response: 0.4153846153846154
Age Range Distribution:
Age_Range_Mapped
Adult      40
Young Adult 26
Middle-Aged 22
Pre-Senior 22
Senior     12
Youth      8
Name: count, dtype: int64
Education Level Distribution:
Education_Level
3    68
5    27
4    27
2     8
Name: count, dtype: int64
Marital Status Distribution:
Marital_Status_Encoded_Mapped
Married    49
Single     38
Together   33
Divorced    7
Widow       3
Name: count, dtype: int64

```

Figure 16, Question f, Characteristics of highest meat buyers

Also, as we can see in Figure 17, the biggest consumers are Adult, Married or youth in a relationship

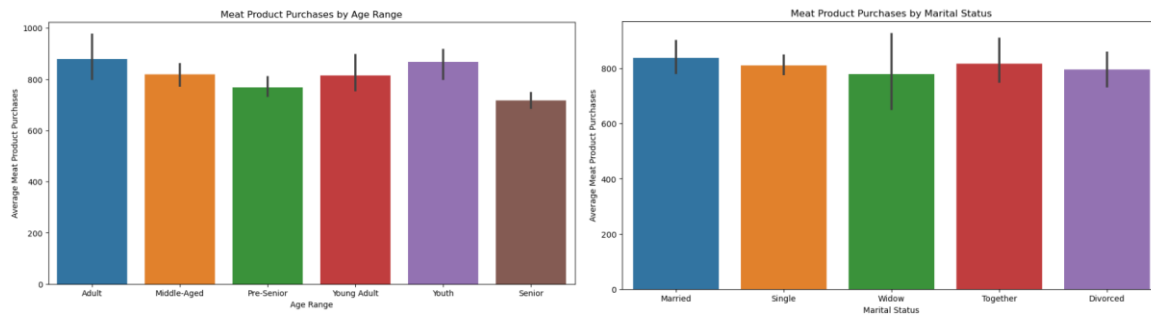


Figure 17, Question f, Meat Product by Age Range and Marital Status

4. Conclusion and Future Work

Conclusions:

- a) The business ought to go after new "Premium Buyer" clients who have a history of buying expensive goods like meat, gold, and wines, as well as high salaries and advanced degrees. Premium offerings are probably going to be accepted by these clients.
- b) The company can consistently rely on its existing "Loyal Diversified Buyers" - customers with a long history of purchases across product categories and high recency. These reliable revenue generators should be a focus of retention efforts.
- c) Targeting efforts are unlikely to yield significant benefits for customers with low incomes and inadequate education who predominantly purchase low-cost, basic products. The business ought to concentrate its efforts on more prospective clientele.
- d) The company should carefully evaluate the lifetime value potential of "Emerging Buyer" customers - new customers with mixed purchasing behavior - before investing heavily in acquisition, as they pose a risk of falling into the unproductive category described

Future Work:

The business can align its efforts with current marketing tools like Marketing 5.0, which was developed by the marketing guru Phillip Kotler in 2021, this marketing 5.0 focuses on consumer experience via technology; now the challenge is how to use all information to create a closer and more agile relationship with customers. We can focus on 2 from the 5 main components of Marketing 5.0 that we can cover in our future work:

Predictive marketing aims to end that famous phrase "half, "Which means that half of the marketing investment goes down the drain. The problem is that we don't know which half". From now on, with the use of predictive marketing, we should be a little closer to knowing which marketing investment is effective and which is not.

Contextual marketing is about using various interfaces to analyze the physical environment as experienced by the user. Once more, this data will be essential to provide the user with accurate information. Businesses may increase engagement and conversion rates by delivering more relevant and personalized content by knowing the user's context.

Finally, we can also develop data-driven capabilities by investing in machine learning for personal brand communications using customer data.

5. References

- Business.org. (2024, March 30). What is the 4P marketing matrix?
<https://www.business.org/marketing/sales/marketing-101-4p-matrix/>
- Clientbook. (2024, March 29). How retail customer segmentation works.
<https://www.clientbook.com/blog/how-retail-customer-segmentation-works>
- Digital Commerce 30. (2023, June 8). The retailer speaks: Digital marketing 2023.
<https://www.digitalcommerce30.com/2023/06/08/the-retailer-speaks-digital-marketing-2023/>
- Hurra, T. (n.d.). DBSCAN — Make density-based clusters by hand. Towards Data Science.
<https://towardsdatascience.com/dbscan-make-density-based-clusters-by-hand-2689dc335120>
- Perez-Font, R. (n.d.). What is marketing 5.0? LinkedIn. <https://www.linkedin.com/pulse/what-marketing-50-ricardo-perez-font/>
- Sthda. (n.d.). DBSCAN: Density-based clustering for discovering clusters in large datasets with noise.
http://www.sthda.com/english/wiki/wiki.php?id_contents=7940
- Taneja, A., & Gupta, S. (2021). Marketing 5.0: The era of technology and the challenges faced by it. International Journal of Advanced Engineering and Management, 6(1), 1-6.
https://ijaem.net/issue_dcp/Marketing%205.0%20The%20Era%20of%20Technology%20and%20the%20Challenges%20Faced%20By%20It.pdf