
iFood

MARKETING

PROBLEM

You're a marketing analyst and you've been told by the Chief Marketing Officer that recent marketing campaigns have not been as effective as they were expected to be. You need to analyze the data set to understand this problem and propose data-driven solutions.

You are expected to use:

Python

SQL

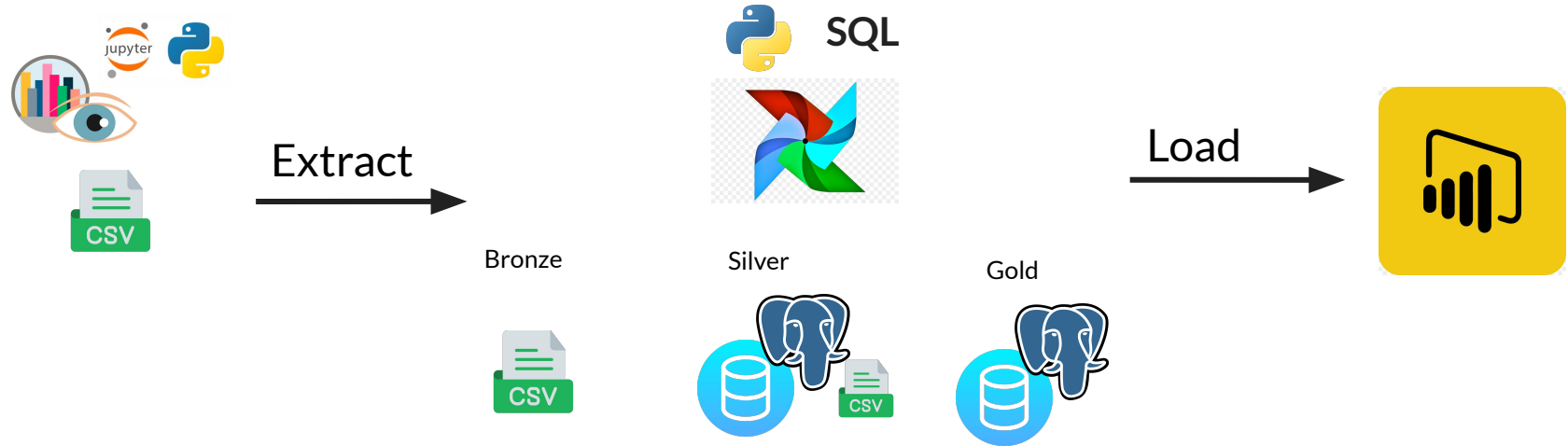
A way to visualise the data.

METHODOLOGY

1. Architecture Definition
 2. Data Source Evaluation: Identify data quality issues (e.g., missing values, duplicates, incorrect data types) and uncover patterns for field engineering.
 3. ETL
 - a. Bronze Layer:
Focus: Resolve basic data quality issues.
Tasks: Data cleaning, data type adjustments.
 - b. Silver Layer:
Focus: Integrate business logic.
Tasks: Add fields, flags, metrics relevant to business analysis.
 - c. Gold Layer:
Focus: Facilitate reporting.
Tasks: Perform data aggregations for summarized insights.
 4. Exploratory Data Analysis: Explore curated data from the Silver layer to extract insights and identify trends.
 5. Reporting: Use the Gold layer's aggregated data for detailed reports and visualizations.
 6. Conclusion
-

ARCHITECTURE

ARCHITECTURE



SOURCE DATA ANALYSIS

COLUMN NAMES

```
# Check for spaces in column names
for i in input_df.columns:
    print(i) if ' ' in i else ""
```

education_2n Cycle

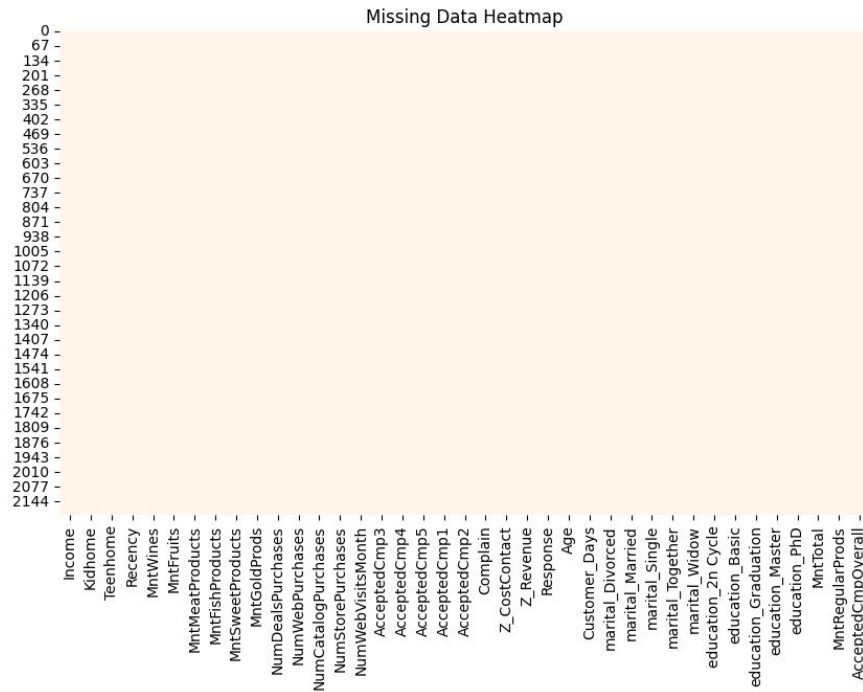
```
input_df.columns
```

```
Index(['Income', 'Kidhome', 'Teenhome', 'Recency', 'MntWines', 'MntFruits',
      'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
      'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases',
      'NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth',
      'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1',
      'AcceptedCmp2', 'Complain', 'Z_CostContact', 'Z_Revenue', 'Response',
      'Age', 'Customer_Days', 'marital_Divorced', 'marital_Married',
      'marital_Single', 'marital_Together', 'marital_Widow',
      'education_2n Cycle', 'education_Basic', 'education_Graduation',
      'education_Master', 'education_PhD', 'MntTotal', 'MntRegularProds',
      'AcceptedCmpOverall'],
      dtype='object')
```

To Clean

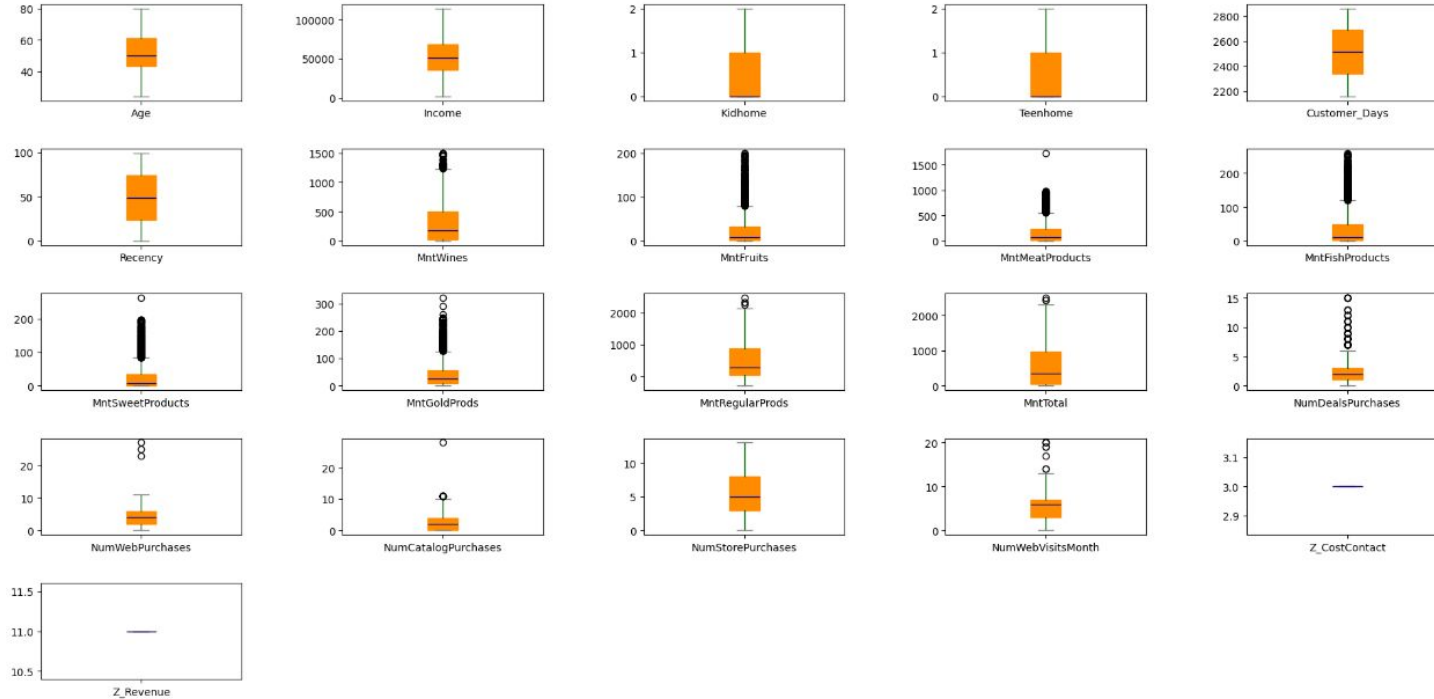
- one field with a space
- we have fields that are not normalized (camel case or containing _)

NULL VALUES



No Null
Values Found

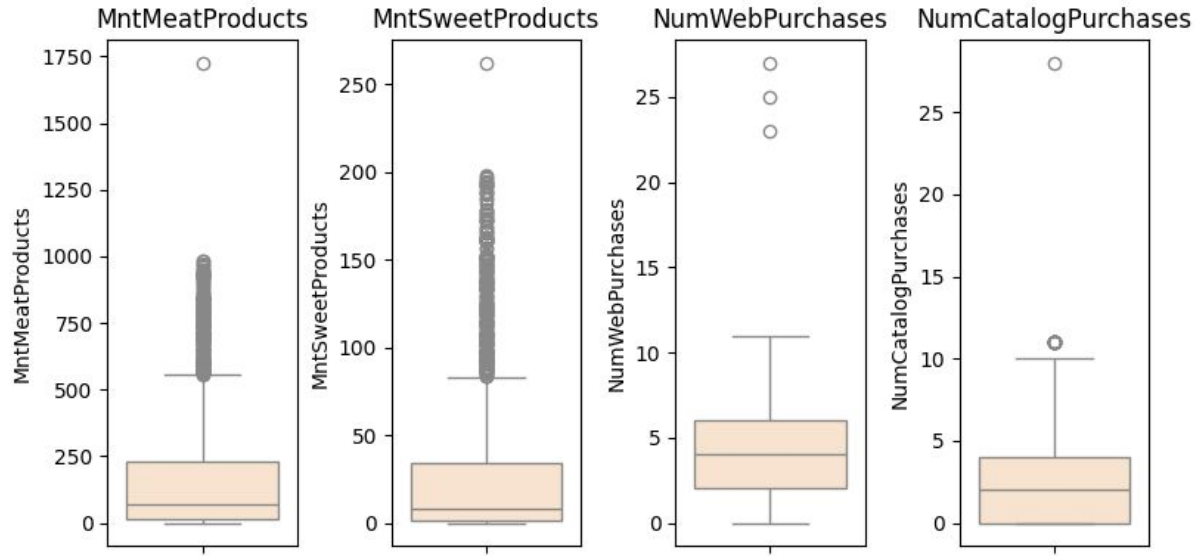
OUTLIERS - Qualitative



Outliers

- Number of Catalog Purchases
- Num Web Purchases
- Mount of Sweet Products
- Mount of Meat Products

OUTLIERS - Qualitative



Given the lack of detailed business context to determine if the identified outliers are anomalies that don't align with expected data patterns, we will not remove them from our analysis.

```
input_df.nunique()
```

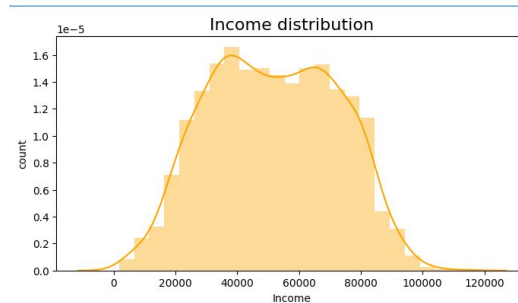
Income	1963
Kidhome	3
Teenhome	3
Recency	100
MntWines	775
MntFruits	158
MntMeatProducts	551
MntFishProducts	182
MntSweetProducts	176
MntGoldProds	212
NumDealsPurchases	15
NumWebPurchases	15
NumCatalogPurchases	13
NumStorePurchases	14
NumWebVisitsMonth	16
AcceptedCmp3	2
AcceptedCmp4	2
AcceptedCmp5	2
AcceptedCmp1	2
AcceptedCmp2	2
Complain	2
Z_CostContact	1
Z_Revenue	1
Response	2
Age	56
Customer_Days	662
marital_Divorced	2
marital_Married	2
marital_Single	2
marital_Together	2
marital_Widow	2
education_2n Cycle	2
education_Basic	2
education_Graduation	2
education_Master	2
education_PhD	2
MntTotal	897
MntRegularProds	974
AcceptedCmpOverall	5

UNIQUE VALUES

- Z fields are constants
- The flags in the dataset contain only two unique values ✓

VALUES DISTRIBUTION

	count	mean	std	min	25%	50%	75%	max
Income	2205.0	51622.094785	20713.063826	1730.0	35196.0	51287.0	68281.0	113734.0
MntRegularProds	2205.0	518.707483	553.847248	-283.0	42.0	288.0	884.0	2458.0



- The income field displays a broad distribution. Without additional data to assess the significance of this observation, it will remain unaltered for now.
- **Rows with negative values in the MntRegularProds column must be removed to avoid data inconsistencies.**

FINDINGS

- The fourth quartile for recency is 74 days or more, indicating potential churn.
 - The minimum age of participants is 24, while the maximum is 80
 - Among all product categories, MntWines has the highest average spending.
 - Purchases are predominantly made in-store and web.
 - The second campaign has the lowest acceptance rate, whereas the fourth campaign is the most successful.
 - The target (Response) shows higher average acceptance compared to previous campaigns.
 - The campaigns were aimed at customers who have been with the company for between 5 to 8 years.
 - Marital status: The campaigns were primarily targeted at couples (married and living together)
 - Education: The focus was more on individuals with a graduation level of education.
 - Spending: People tend to spend more on regular products than on gold products.
-

ISSUES SUMMARY (Bronze)

- The income field displays a broad distribution. Without additional data to assess the significance of this observation, it will remain unaltered for now.
 - Rows with negative values in the MntRegularProds column must be removed to avoid data inconsistencies.
 - One column name (*education2nd Cycle*) contains a space and we have fields that are not normalized (camel case while others contains _)
-

FIELDS/FEATURE ENGINEER (Silver)

General Fields

- **Dependents:** The total number of dependents in the home can be engineered from the sum of 'Kidhome' and 'Teenhome'
- **TotalPurchases:** The total purchases can be engineered from the sum of all features containing the keyword 'Purchases'

Flags

- **isGraduated:** 1 if has graduation, master or PhD. 0 otherwise
 - **isCouple:** 1 if together or married. 0 otherwise
 - **hasChild:** 1 if has child or teen at home. 0 otherwise
 - **isSingleParenting:** 1 if is_couple is false and has_child is true. 0 otherwise
 - **isChurn:** 1 if customer present in the 4th quartil of recency (recency > 74 days), 0 otherwise, 0 otherwise
-

FIELDS/FEATURE ENGINEER (Silver)

Demographic and behavioral segmentation

- **RFMSegment:** based on the Recency(R), TotalPurchases(F) and MntTotal(M)
 - **AgeGroup:**
 - 1: 0-30 #Emerging Adults
 - 2: 31-45 #Established Adults
 - 3: 46-60 #Experienced Adults
 - 4: 61-74 #Senior Adults
 - 5: 75+ #Wisdom Elders
 - **SpenderGroup**
 - **Low:** < 500
 - **Medium:** Between 25% and 50% percentiles 500 to 1000
 - **High:** Between 50% and 75% percentiles 1000 to 2000
 - **VeryHigh:** Above the 75% percentile > 964
 - **IncomeGroup**
 - **Low:** Below the 25% percentile (< 35196)
 - **Middle:** Between the 25% and 75% percentiles (35196 to 68281)
 - **High:** Above the 75% percentile (> 68,281)
 - **Education:** based on the existing flags
 - **MaritalStatus:** based on the existing flags
 - **PreferredChannel:** channel with the greater amount of purchases (web,deal,catalog,store)
-

ETL

DAG

☐ DAG: ifood_pipeline A simple DAG to process iFood data

queued Schedule: 1 day, 0:00:00 Next Run: 2024-04-04, 00:00:00

[Tree](#) [Graph](#) [Calendar](#) [Task Duration](#) [Task Tries](#) [Landing Times](#) [Gantt](#) [Details](#) [Code](#)



2024-04-03T00:00:01Z [Runs](#) 25 [Run](#) scheduled__2024-04-03T00:00:00+00:00

[Layout](#) Left > Right [Update](#)

Find Task...

PythonOperator

[queued](#) [running](#) [success](#) [failed](#) [up_for_retry](#) [up_for_reschedule](#) [upstream_failed](#) [skipped](#) [scheduled](#) [deferred](#) [no_status](#)

☐ Auto-refresh



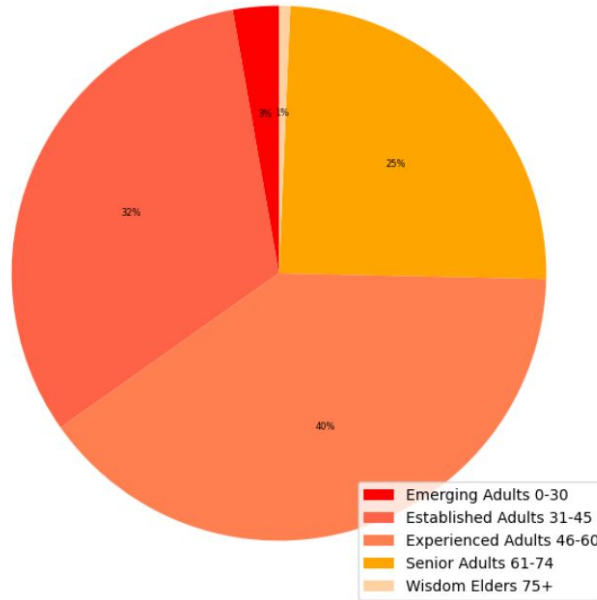
EXPLORATORY DATA ANALYSIS (EDA)

AUDIENCE SEGMENTATION

- how emerging adults and established adults reacted to the different campaigns?
- how the others reacted to the different campaigns?

Distribution by Age Group

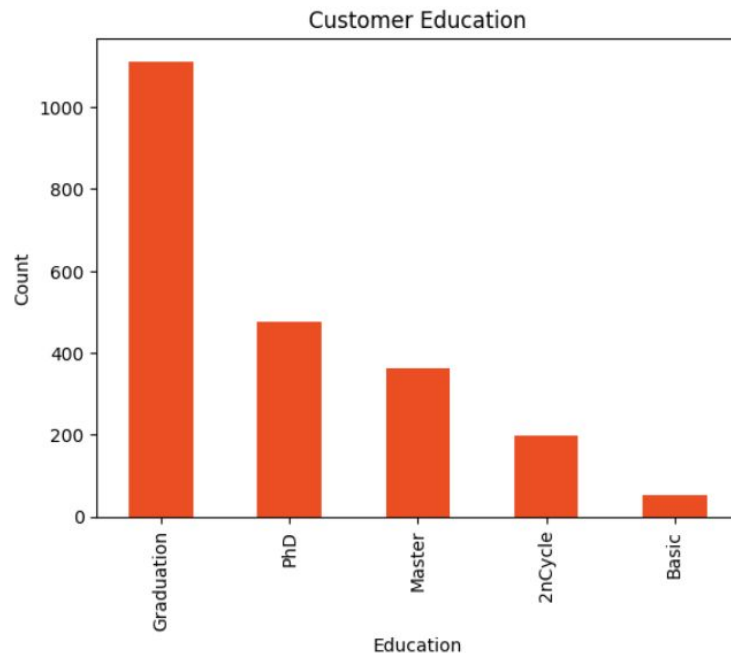
Customer Distribution by Age Group



- 40% of the targeted customer are between 46 and 60 age range
- 60% of the customers are aged 46 and above

-
- how graduated reacted to the different campaigns?
 - how the others reacted to the different campaigns?

Distribution by Education



The vast majority of the targeted audience possesses academic qualifications, with graduation being the most common level of education.

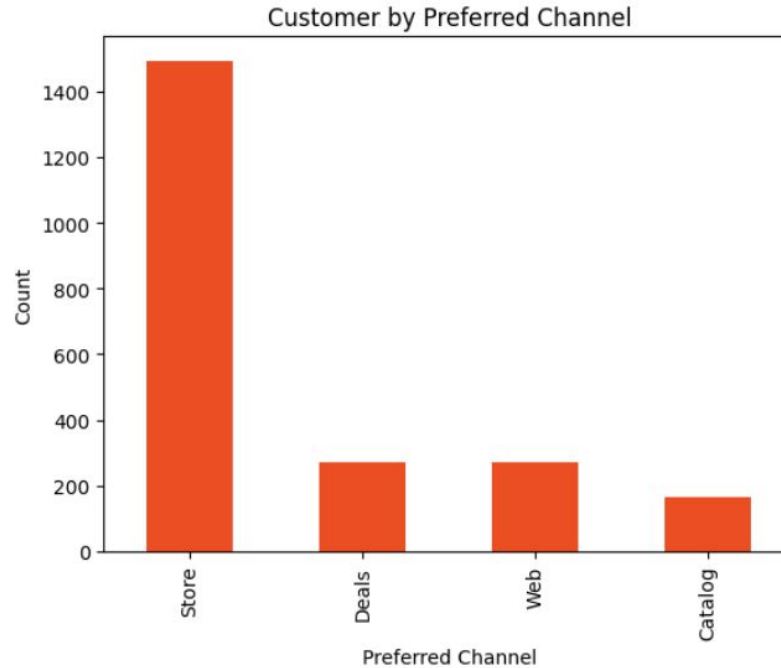
-
- how couples reacted to the different campaigns?
 - how the others reacted to the different campaigns?

Distribution by Marital Status



The vast majority of the targeted audience is couple (married or together)

Distribution by Preferred Channel

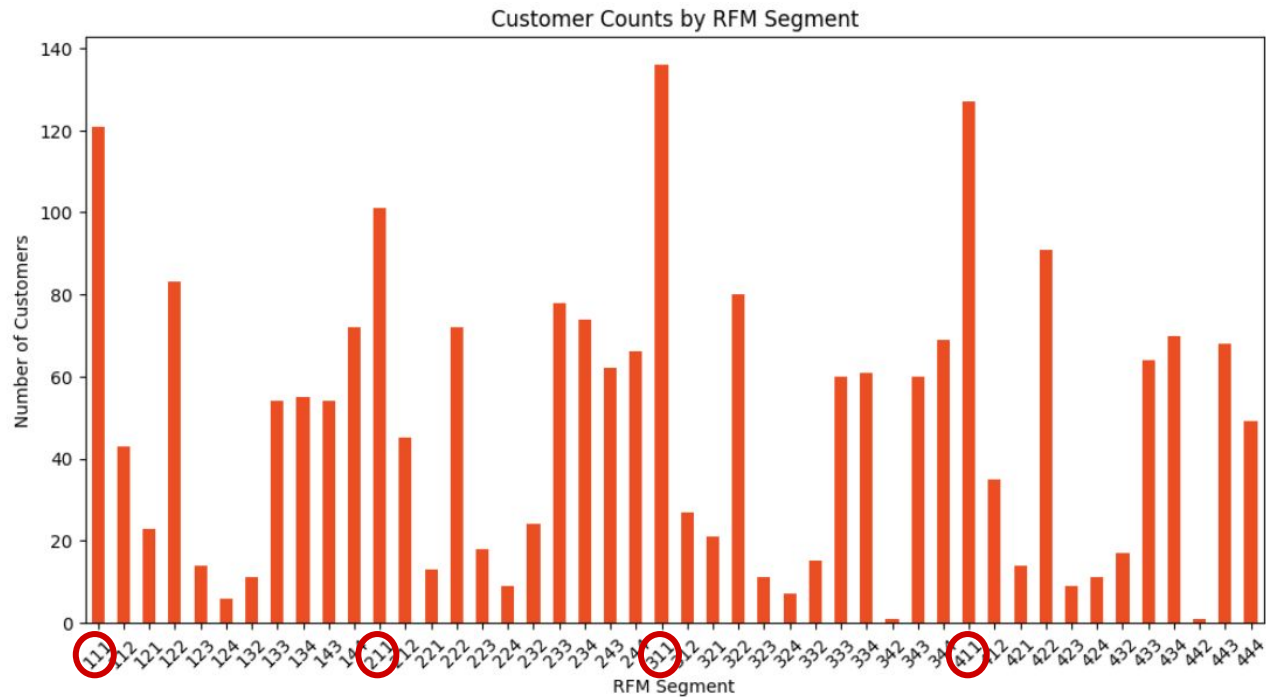


The vast majority of the targeted audience prefers to purchase in Store

RFM Segmentation

- **Recency:** A lower recency value (more recent interaction) results in a higher score, indicating recent engagement.
 - **Frequency:** Greater purchase frequency leads to a higher score, reflecting more frequent transactions.
 - **Monetary:** Higher spending amounts earn a higher score, signifying greater customer value through monetary contribution.
-

RFM Segmentation



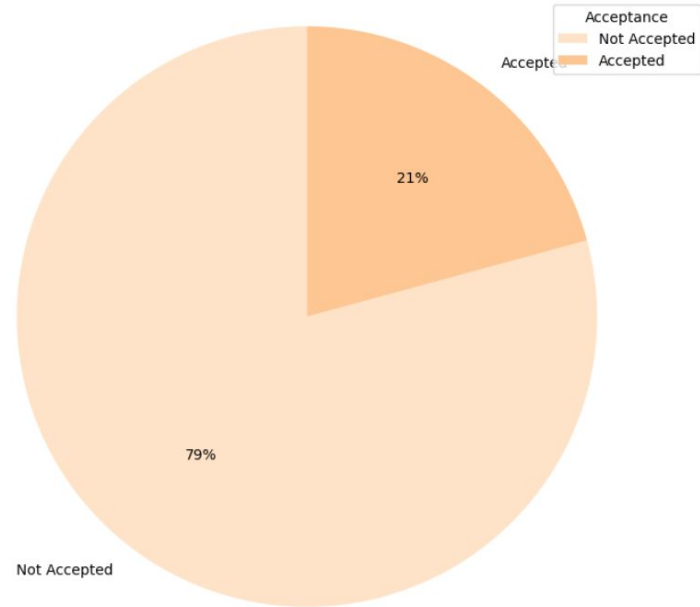
RFM Segmentation - Score

It appears that these campaigns were aimed at customers who are infrequent purchasers and low spenders.

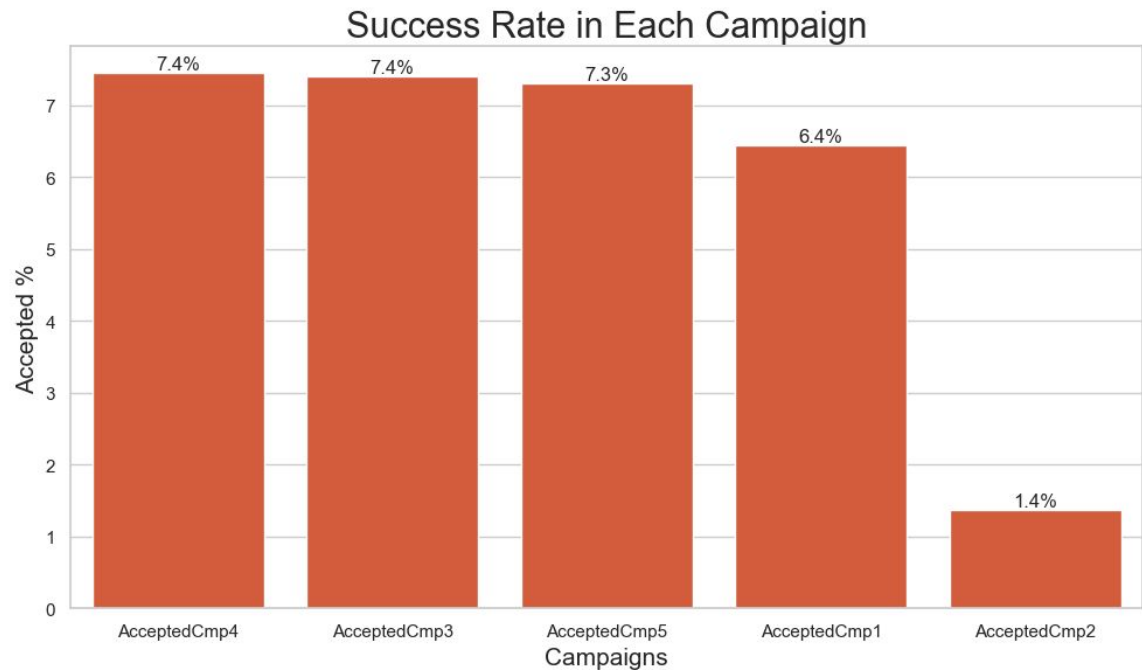
CAMPAIGN PERFORMANCE

General Acceptance

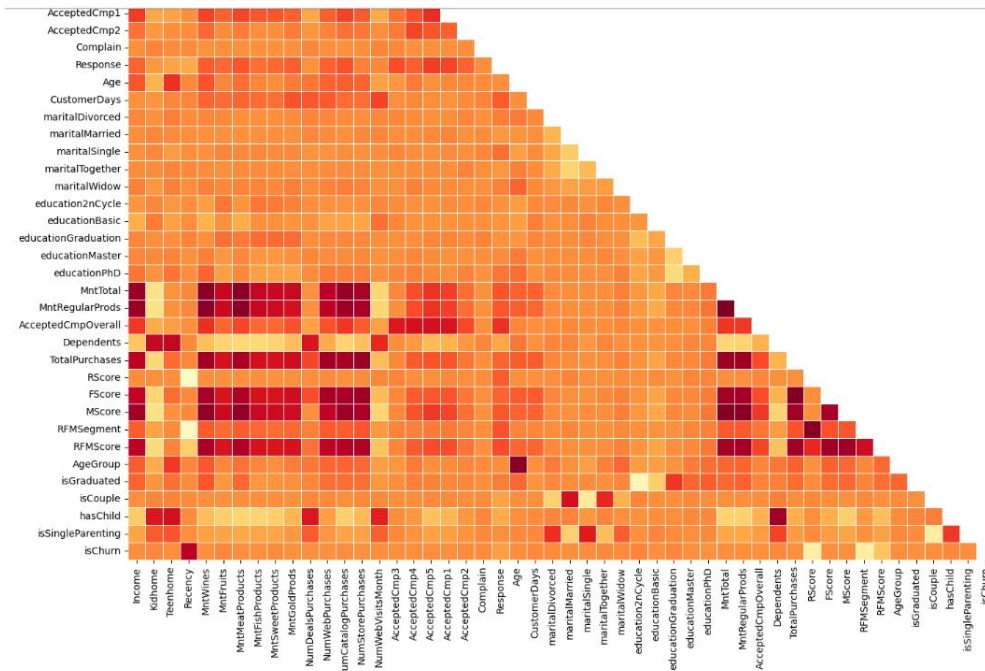
How many customers accepted the campaign?



Success Rate by Campaign



CORRELATION

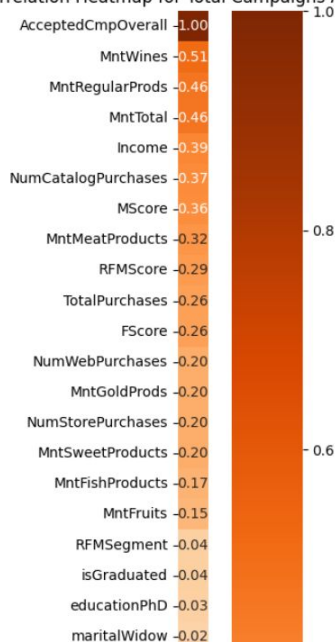


- Income and the total purchases have a strong correlation
- Income and the total amount spent have a strong correlation
- Income and overall campaign acceptance have a strong correlation

Unfortunately, we can observe lack of substantial and meaningful correlations beyond these

Correlation Total Campaigns Accepted - Positive

Correlation Heatmap for Total Campaigns Accepted



Customer

- Positive Correlation with **Income**

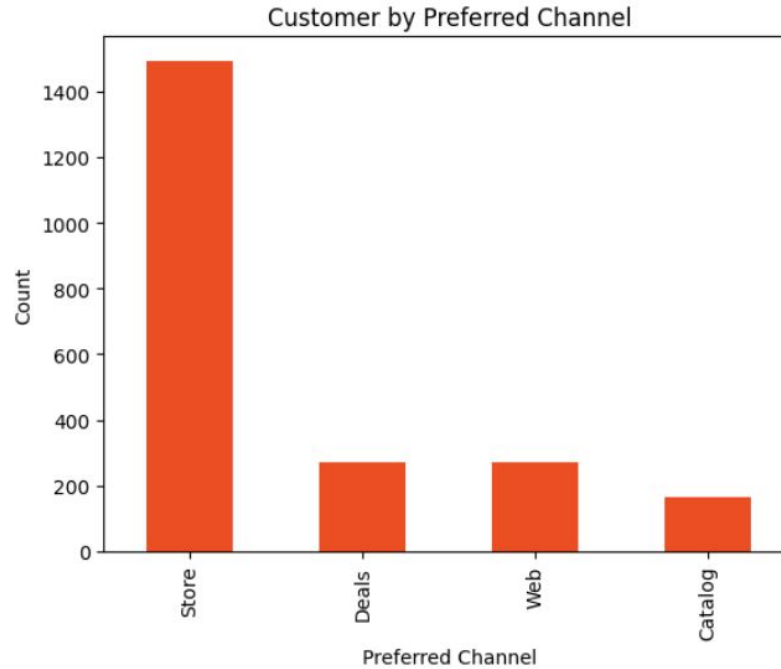
Orders

- Strong correlation with the **amount spent in wines and regular products**
- Positive correlation with the **total amount spent**

Channel

- Positive correlation with **Catalog Purchases**

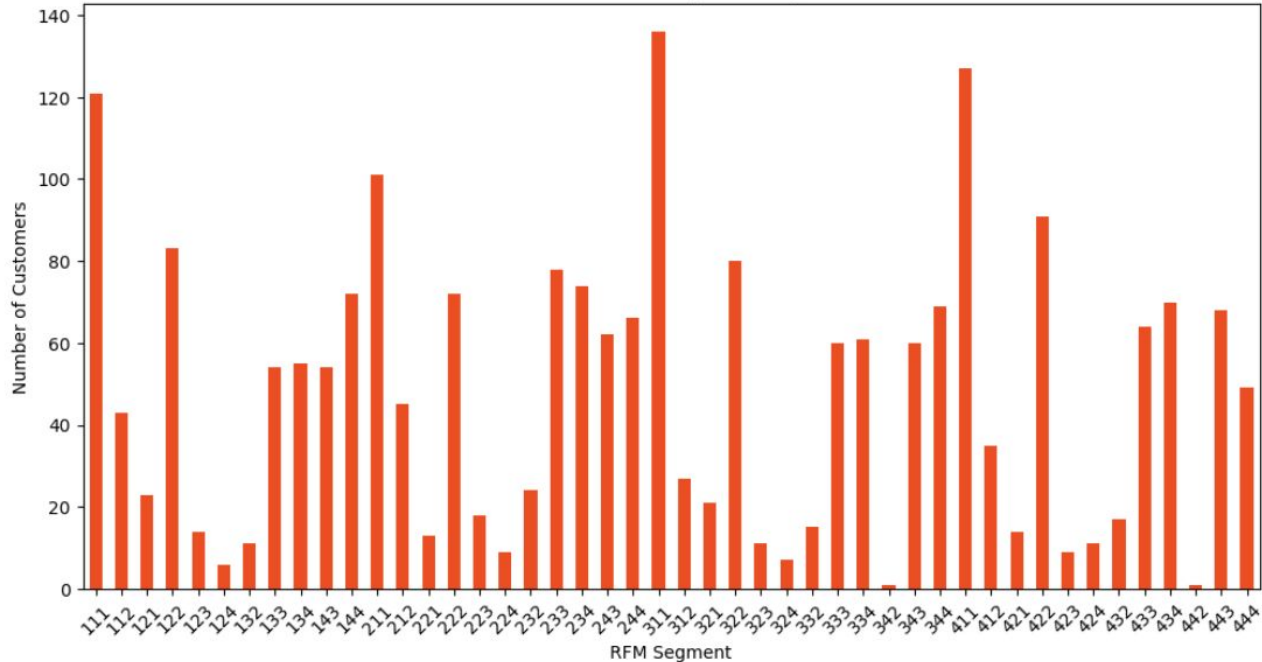
ANALYSIS



The campaigns appeared to be targeted towards customers who prefer the Store and Deals channels, although they were more effective with customers favoring the Catalog channel.

ANALYSIS

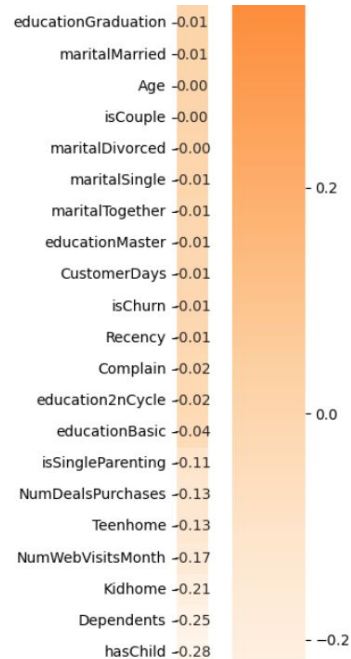
Customer Counts by RFM Segment



The campaigns seemed to be directed at customers with **low frequency** of purchases and **minimal spending**.

However, there is a **positive correlation** between campaign success and the total amount spent by customers.

Correlation Total Campaigns Accepted - Negative



Customer

- Negative correlation with **Dependents**
- Negative correlation with **Basic and 2ndCycle Education**

Orders

- Negative correlation with **recency**

Channel

- Negative correlation with **Deals Purchases**
- Negative correlation with **Number of Web Visits**

INSIGHTS

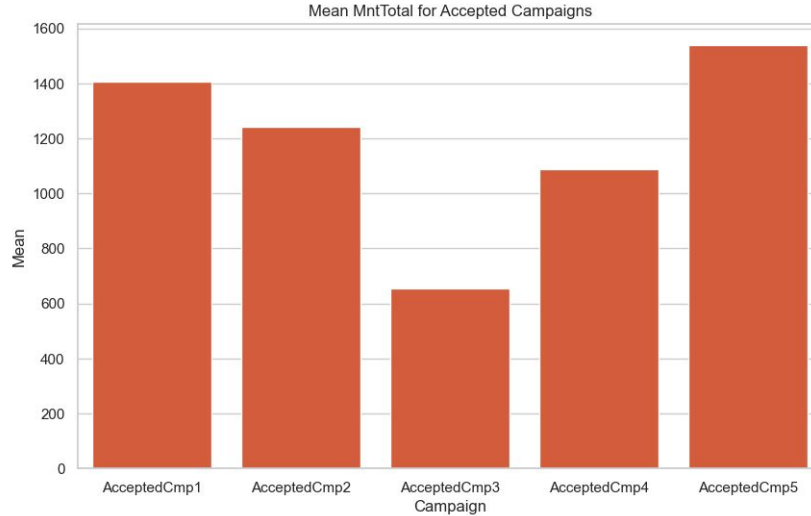
Correlation Observations

- **Education and Marital Status:** No significant relationship between a customer's education or marital status and their acceptance of the campaigns.
- **Income and Spending on Wines:** A noticeable correlation exists between a customer's income and their expenditure on wines, which aligns with the overall campaign performance.

Campaign Targeting and Performance

- **Preferred Channels:** Although the campaigns targeted customers favoring the Store and Deals channels, they were notably more successful among those who preferred the Catalog channel.
 - **Customer Spending Habits:** The focus of the campaigns was on customers characterized by infrequent purchases and lower spending levels. However, there was a positive correlation between campaign success and higher customer spending.
-

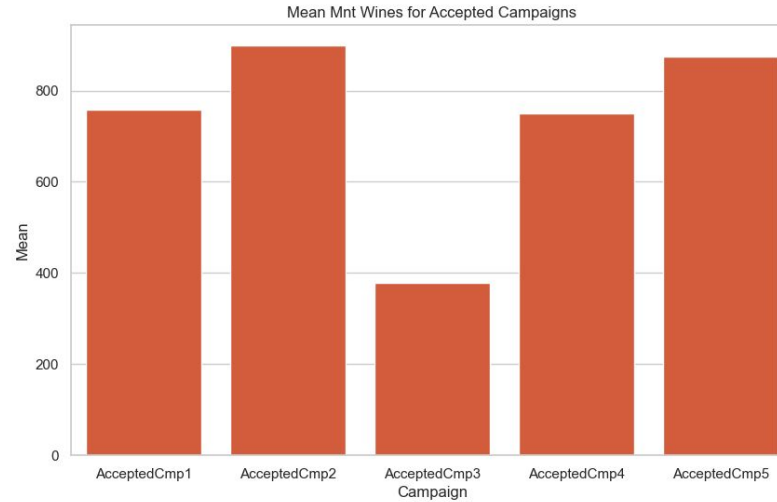
Strong Correlations - Total Spent



The customer who accepted Campaign 3 have lower total amount spent in the last 2 years. While the ones who accepted Campaign 5 have higher amount spent.

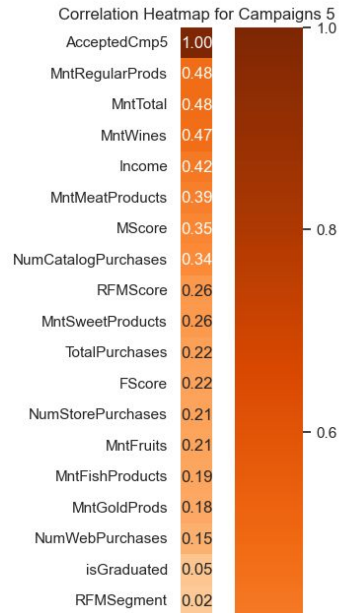
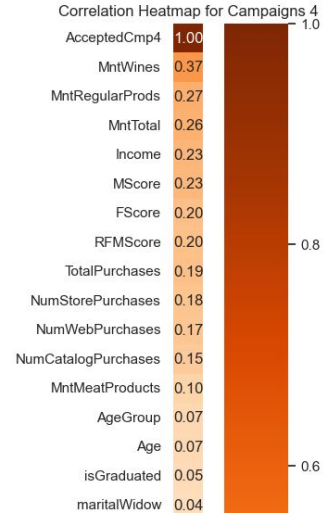
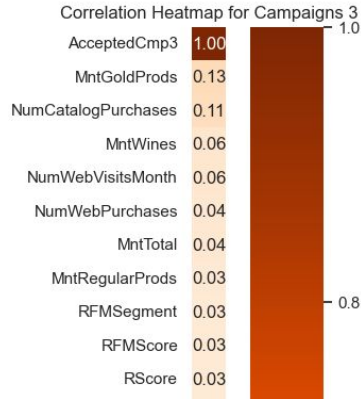
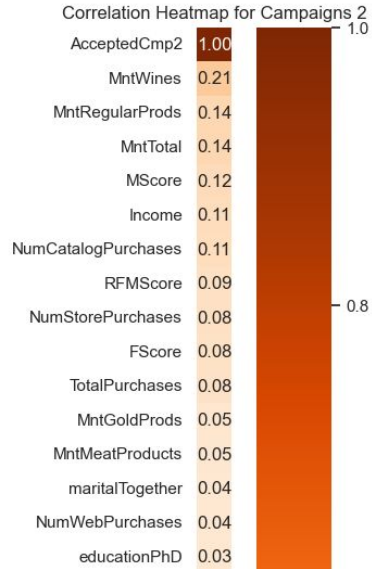
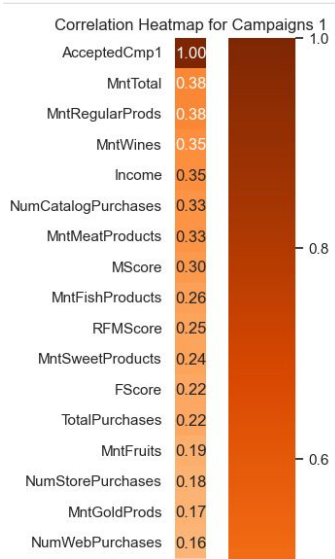
Campaign 2 which was the less successful one has the lower max value (2056) compared to the other campaigns

Strong Correlations - Wine



The customer who accepted Campaign 2 have higher total amount spent in wine in the last 2 years. Since this is the campaign with lower acceptance the target audience should be redirected

Correlation by Campaign Accepted



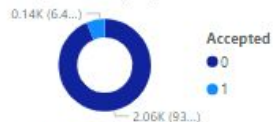
INSIGHTS

- Campaign 1,4 and 5 are the ones with strongest correlations and have the same strong correlations observed before in the overall campaign correlation assessment
 - Campaign 2 has a positive correlation with amount spent in wines as we observed before -> **should target customers who spend a significant percentage in wine**
 - Campaign 3 has weak correlations overall. The strongest is related to Gold Product purchases. This focus on higher-cost gold items may explain the lower overall spending -> **can be used to lower the stock of gold products** if needed, targeting the audience to gold product frequent buyers
-

REPORTING

PERFORMANCE REPORT

Accepted Campaign 1



Accepted Campaign 2



Accepted Campaign 3



Accepted Campaign 4



Accepted Campaign 5



142
AcceptedCmp1
30
AcceptedCmp2
163
AcceptedCmp3
164
AcceptedCmp4
161
AcceptedCmp5

MaritalStatus	IncomeGr...	Education	SpenderGro...
<input type="checkbox"/> Divorced	<input type="checkbox"/> High	<input type="checkbox"/> 2nCycle	<input type="checkbox"/> High
<input type="checkbox"/> Married	<input type="checkbox"/> Low	<input type="checkbox"/> Basic	<input type="checkbox"/> Low
<input type="checkbox"/> Single	<input type="checkbox"/> Middle	<input type="checkbox"/> Graduati...	<input type="checkbox"/> Medium
<input type="checkbox"/> Together		<input type="checkbox"/> Master	<input type="checkbox"/> VeryHigh

Sum of total_customers by acceptance and campaign



CUSTOMER REPORT

AGE (average)



INCOME (average)



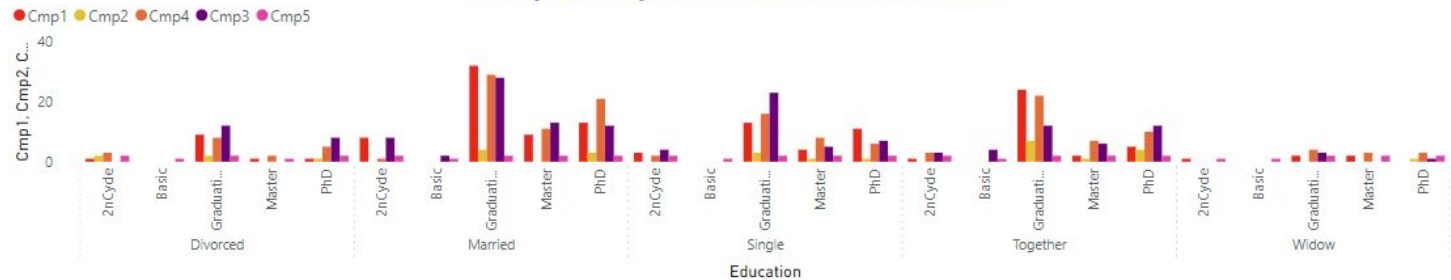
ACCEPTANCE

Multiple selections

CAMPAIGN

All

Acceptance by MaritalStatus and Education



ORDERS REPORT

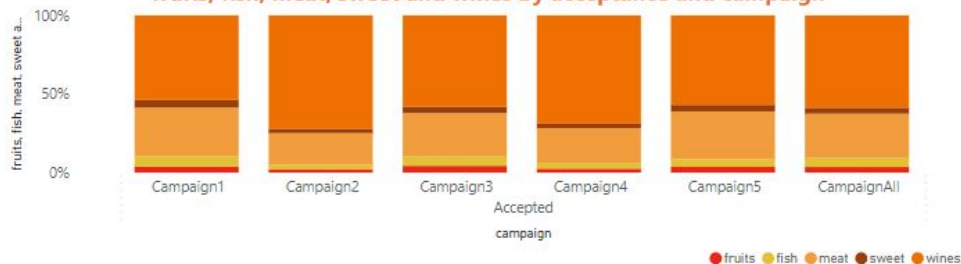
Sum of avg_purchases by campaign and acceptance



regular and gold by acceptance and campaign



fruits, fish, meat, sweet and wines by acceptance and campaign



ACCEPTANCE

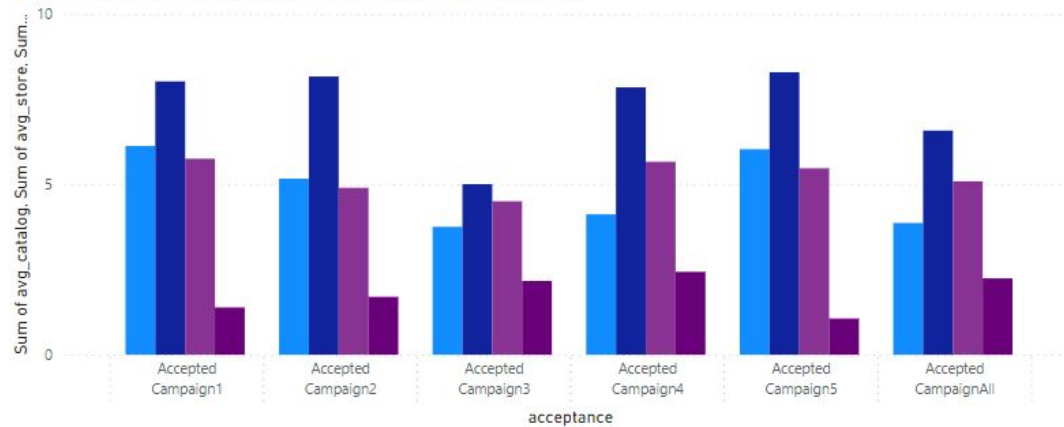
Multiple selections

CAMPAIGN

All

Channel (avg)

Sum of avg_catalog Sum of avg_store Sum of avg_web Sum of avg_deals



CHANNEL REPORT

ACCEPTANCE

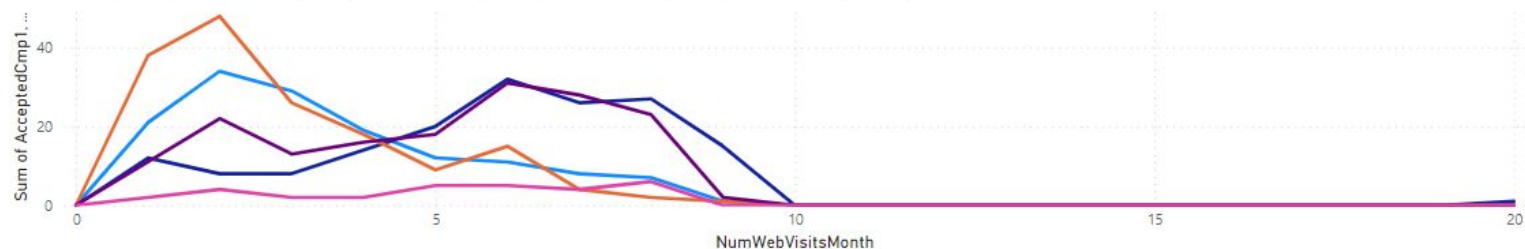
Multiple selections

CAMPAIGN

All

Acceptance by Web Visits

Sum of AcceptedCmp1 Sum of AcceptedCmp3 Sum of AcceptedCmp5 Sum of AcceptedCmp4 Sum of AcceptedCmp2



CONCLUSION

RECOMMENDATIONS

Targeting and Segmentation Adjustments

Adjust Channel Strategy: Despite targeting **Store, Deals and Web** channel customers, campaigns were more successful among Catalog channel customers. **Shift focus towards Catalog** and possibly online channels, where customers might be more engaged and receptive.

Product Category:

- **Wines:** Campaign 2 has a positive correlation with amount spent in wines so it should be tailored and targeted in this sense. Point this campaign into those clients or to lower the stock of wine.
- **Gold Products:** campaign 3 can be used to lower the stock of gold products if needed, targeting the audience to gold product frequent buyers

Refine Customer Segmentation Targeting:

- **Age:** Given that 60% of customers are aged 46 and above, consider tailoring campaigns more closely to the preferences and needs of this age group.
 - **Graduation:** Tailor messaging and offers to resonate with their specific interests and financial considerations of graduation-level education as they are the vast majority of the targeted audience .
 - **Customer Spending Habits:** Focus the campaign targeting on the high spenders! The focus of the campaigns was on customers characterized by infrequent purchases and lower spending levels. However, there was a positive correlation between campaign success and higher customer spending.
-

Feedback

Collect Customer Feedback: Implement mechanisms to gather direct feedback from customers about their preferences, satisfaction with campaign offers, and areas for improvement. Use this feedback to refine future campaigns.

Predictive Model

- Build predictive models to forecast customer responses to campaigns using decision trees or random forests.
 - Evaluate model performance using appropriate metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.
-

LIMITATIONS

Limitation

- **lack of substantial correlations in the data.** This underscores the need for further data collection and analysis to uncover additional insights that can inform more nuanced marketing strategies. For example: **increase the duration of the campaigns** or more **campaigns with ab testing**
-

TECH PRESENTATION

QUESTIONS?
THANKS
