ADVANCED PROGRAMMING AND DATA VISUALIZATION - A.Y. 2024/25

PROJECT

MAXIMIZING MARKETING PROFITS: PREDICTIVE MODELING AND CUSTOMER SEGMENTATION FOR CAMPAIGN SUCCESS

By Sacha Hangimana, Viola Brevi, Benie Nezerwa

01 CONTEXT

In this study, our goal is simple but powerful: maximize the profitability of a direct marketing campaign for a new gadget.

The previous campaign, which targeted 2,240 randomly selected customers, ended in a loss of 3.046 MU due to a disappointing 15% conversion rate.

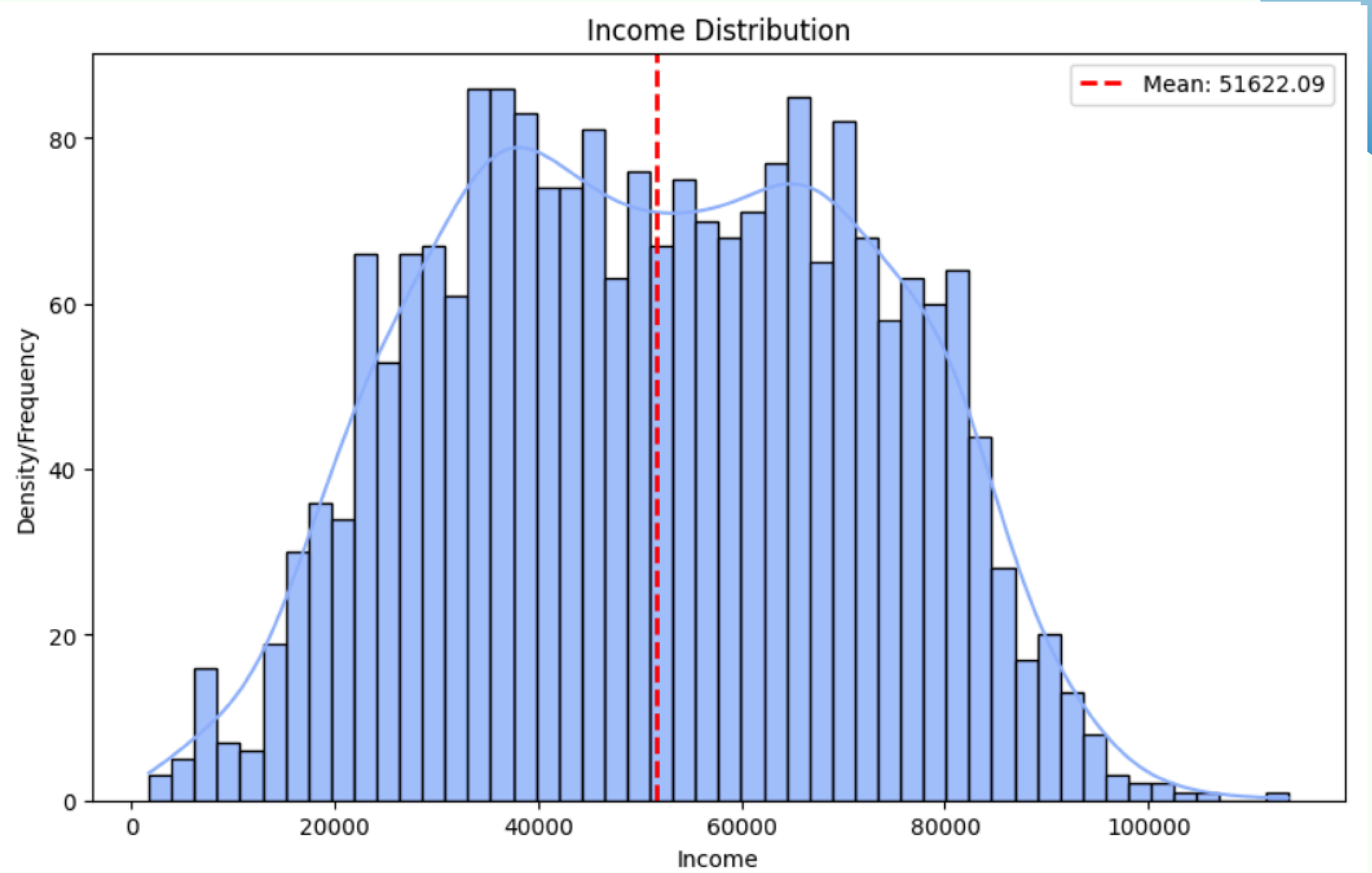
To turn this around, we delve into:

- Understanding customer behavior
- Segmenting the market intelligently
- Comparing predictive models
- Optimizing the contact strategy

The key? Targeting the most likely buyers while avoiding wasting resources on non-respondents. Let's explore how we can achieve that!

02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS

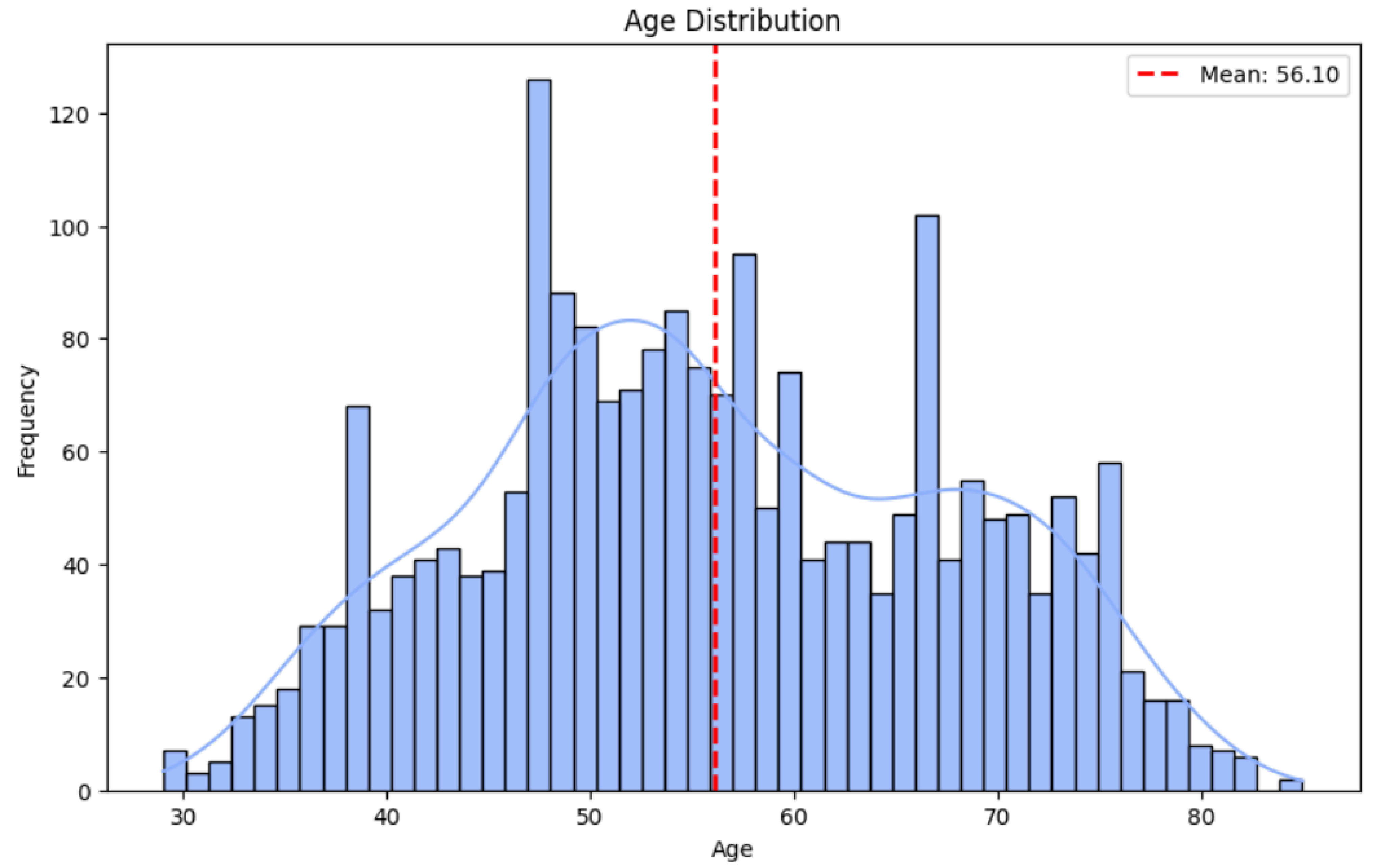
We begin by exploring the key characteristics of our customers



Most customers earn between 30,000 MU and 70,000 MU, with peaks around 40,000 MU and 60,000 MU. The income distribution is slightly right-skewed, with a small group earning above 100,000 MU.

→This suggests a **financially stable, middle-income majority**, while the high-income segment offers opportunities for premium sales.

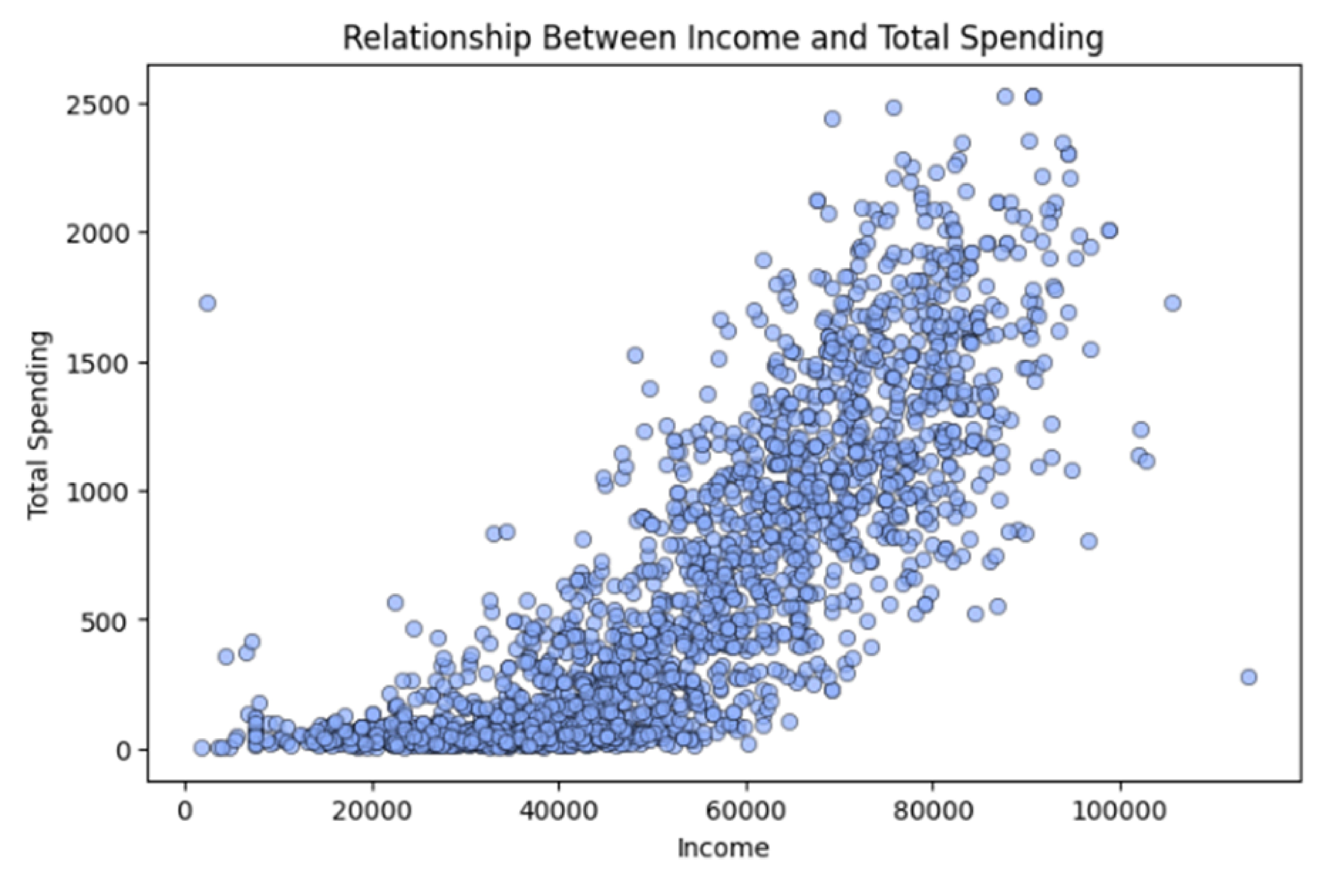
02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS



The majority of customers are aged between 40 and 70, with a peak at 50 years.

→This implies that marketing efforts should **focus on the preferences of this age group**, while exploring **strategies to attract younger** customers for long-term growth.

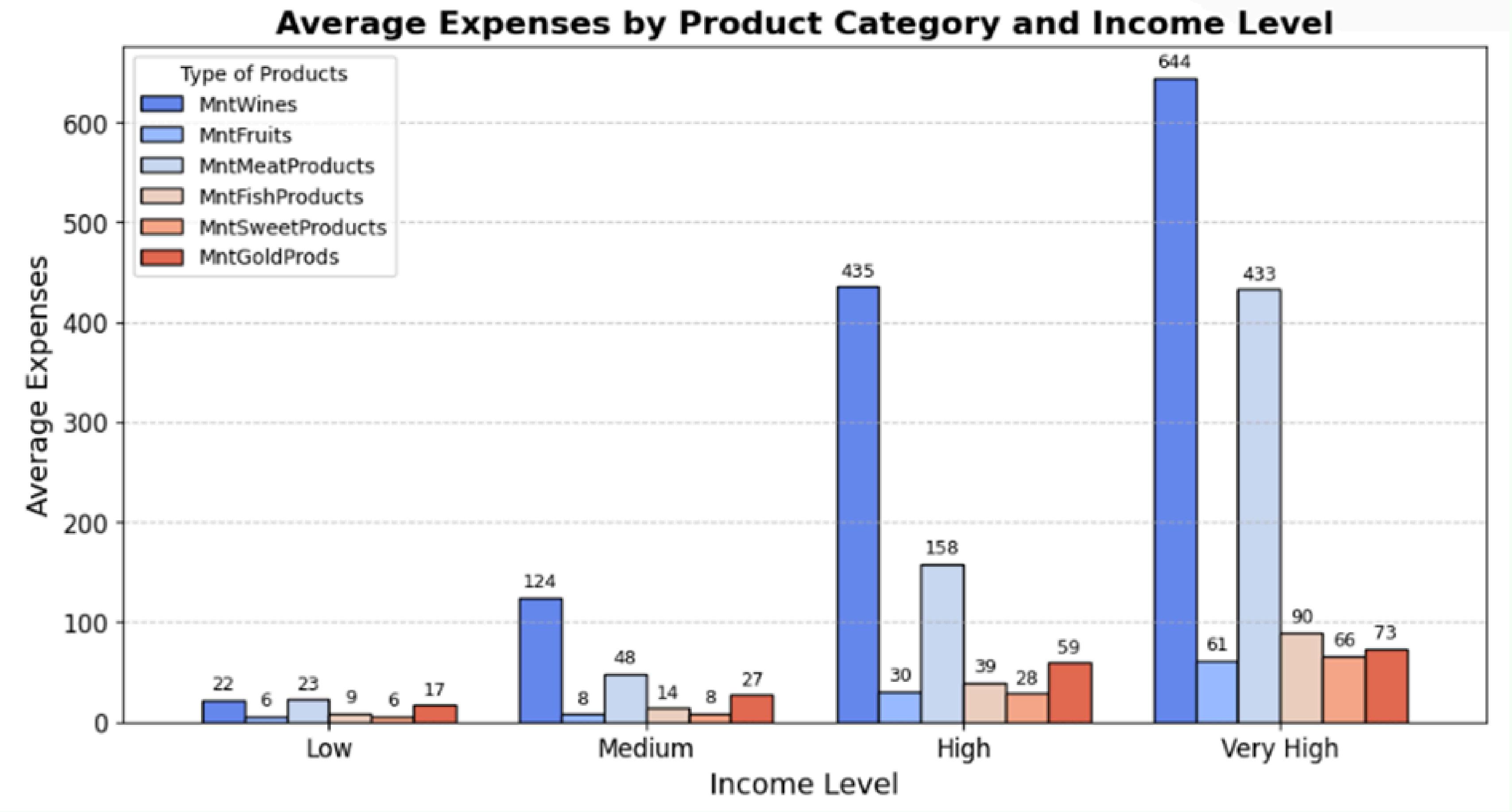
02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS

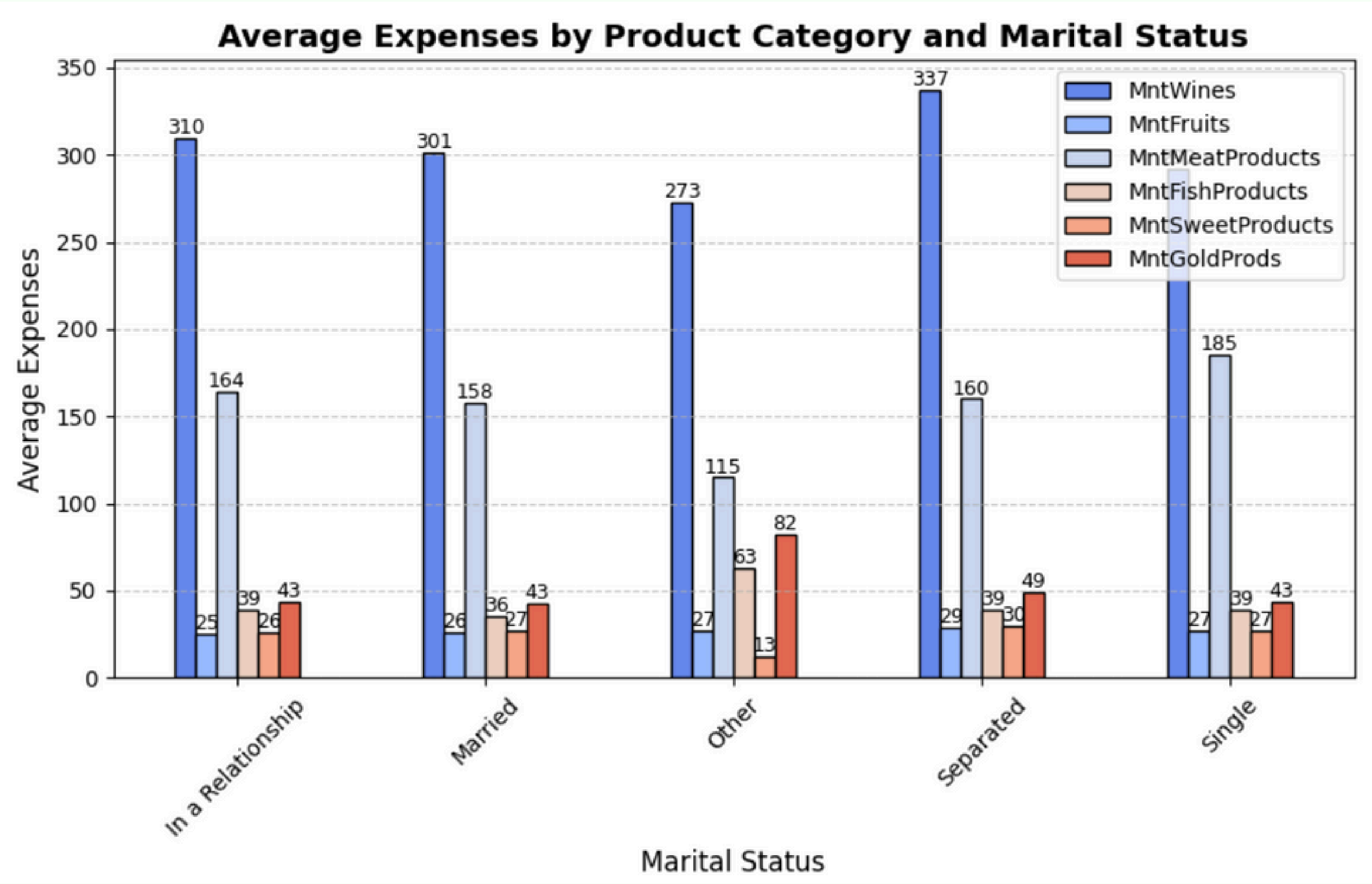


A **positive correlation** exists between income and Spending.

→Higher-income customers tend to spend more, but spending varies within income brackets.

02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS

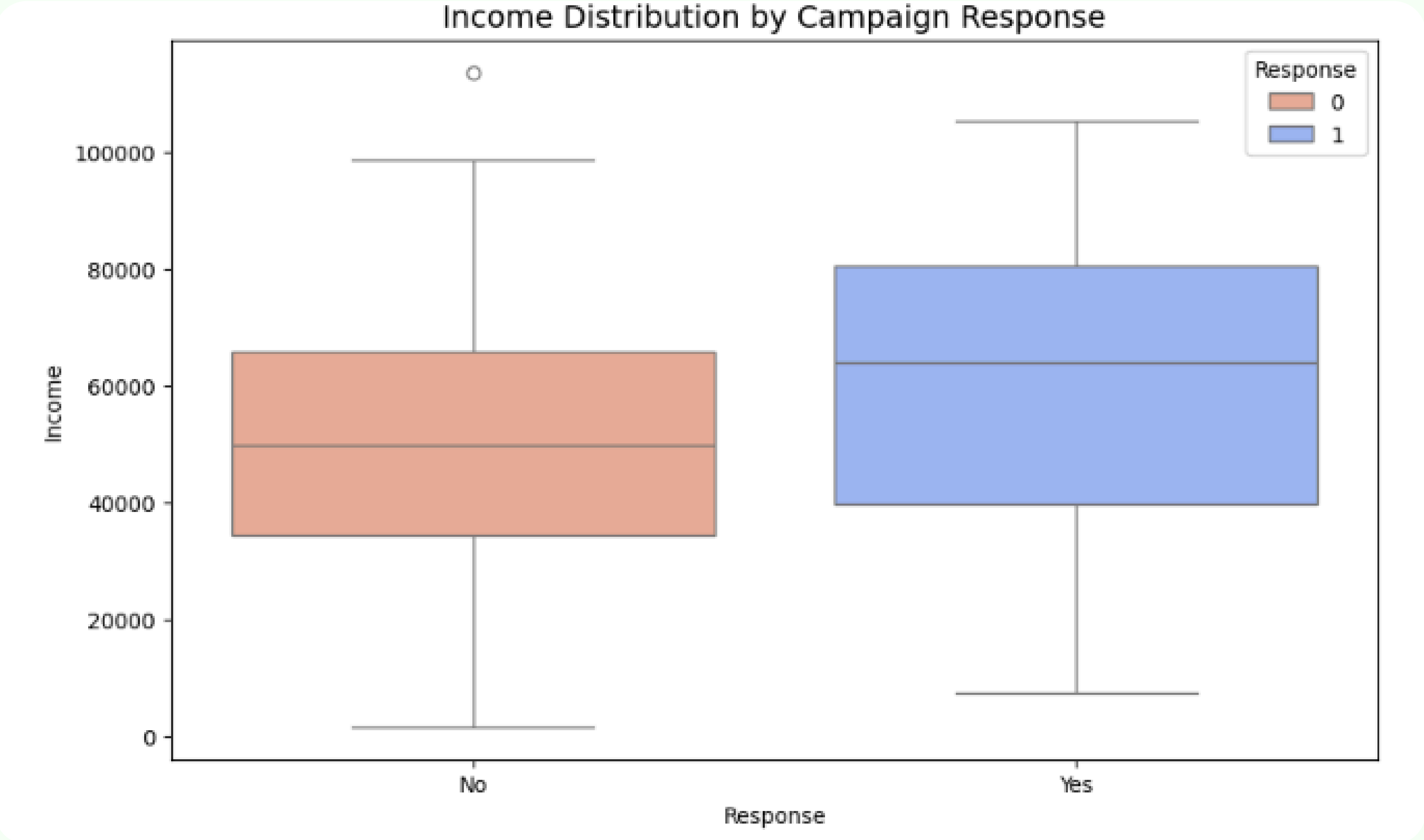




Wine is the top seller, especially among high-income earners, followed by meat. Gold products attract wealthier customers, while fish, sweets, and fruit have stable demand, rising with income.

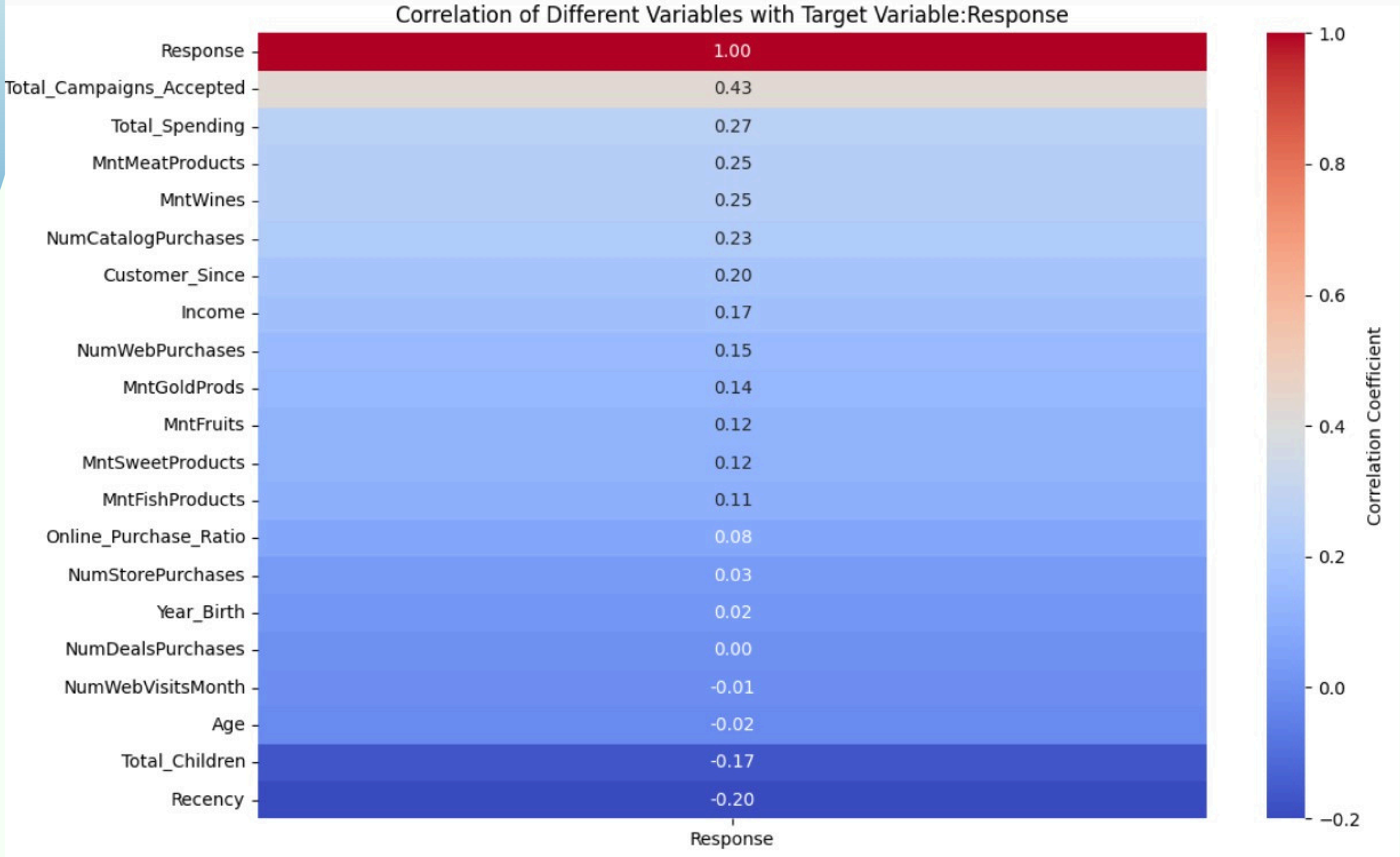
Married couples spend more on essentials like meat and fish, while singles focus more on luxury items like gold.

02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS



The box plot indicates that **higher-income customers are more likely to respond.**
However, since **both groups share similar income distributions**, income alone does not determine response. While targeting wealthier individuals may improve success, **spending habits and engagement should also be considered.**

02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS

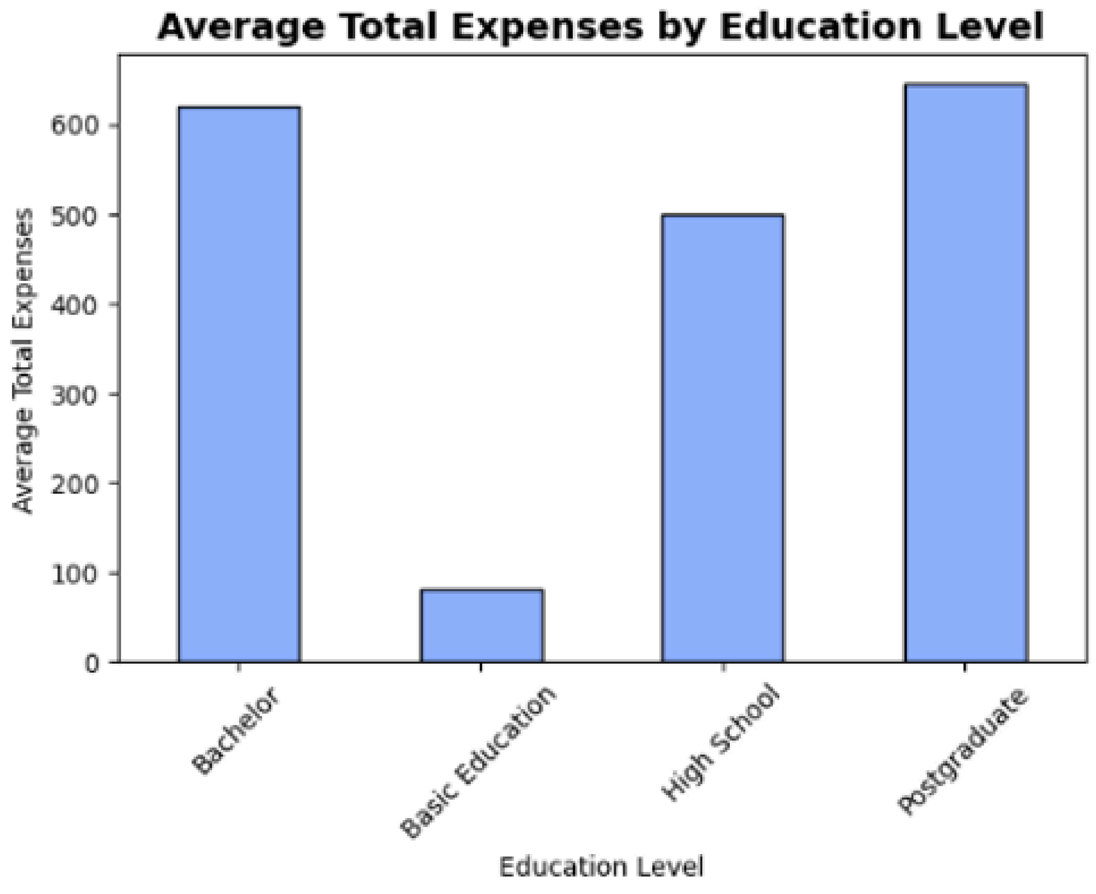


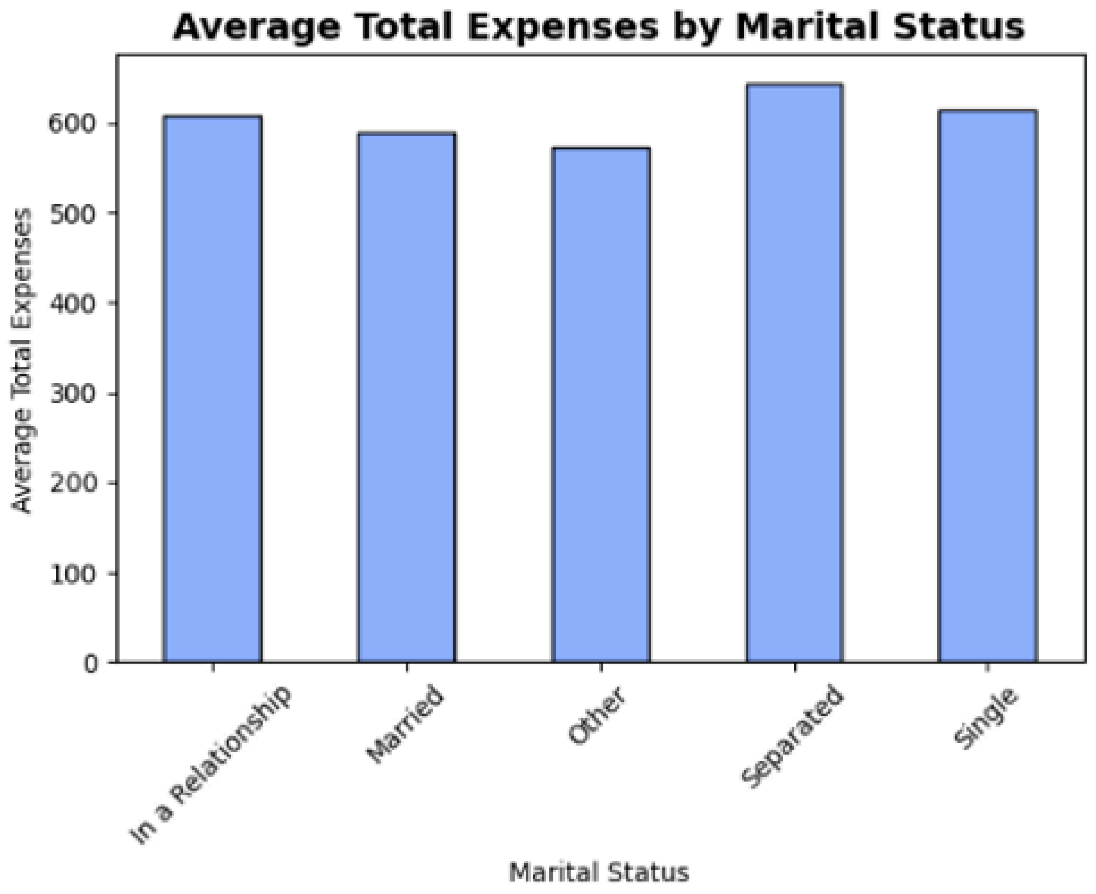
The chart shows that the most correlated factor with response is the total **number of accepted campaigns**, followed by total spending and purchases of specific products such as meat and wine.

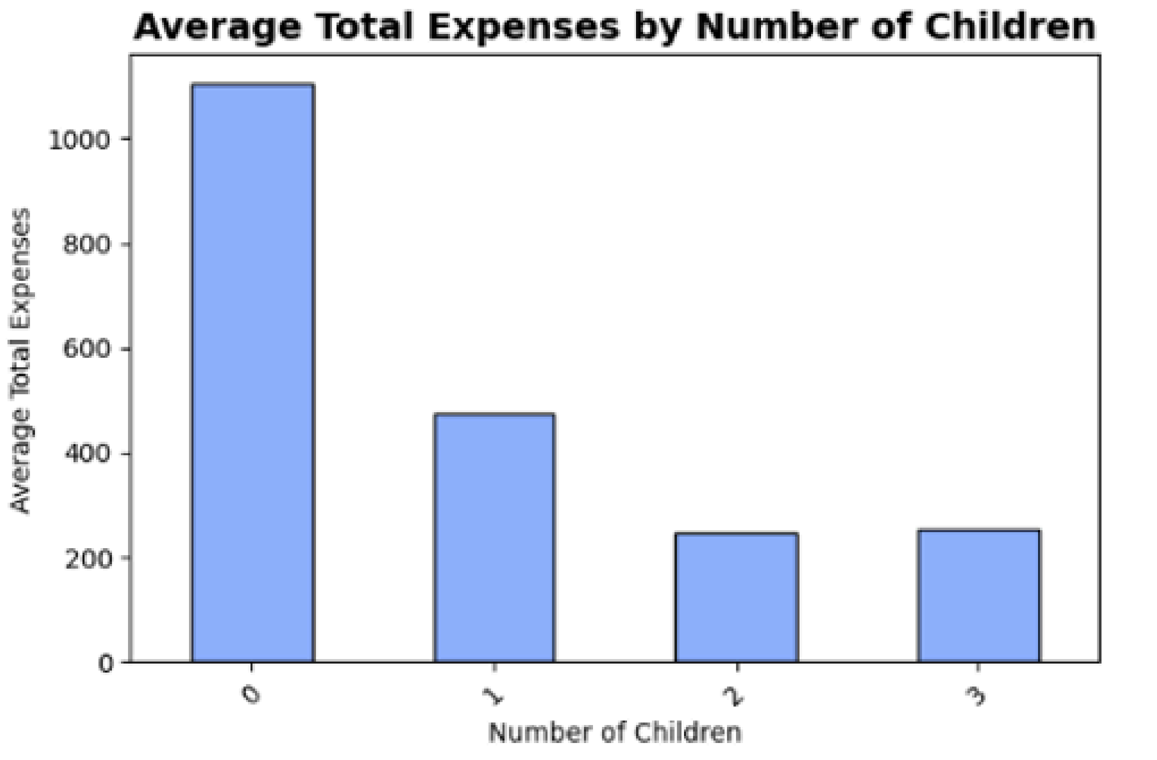
On the other hand, variables like recency* and the total number of children have a **negative correlation**, suggesting that a longer gap since the last purchase and a higher number of children may reduce the likelihood of response.

*Recency: Time since the last purchase

02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS





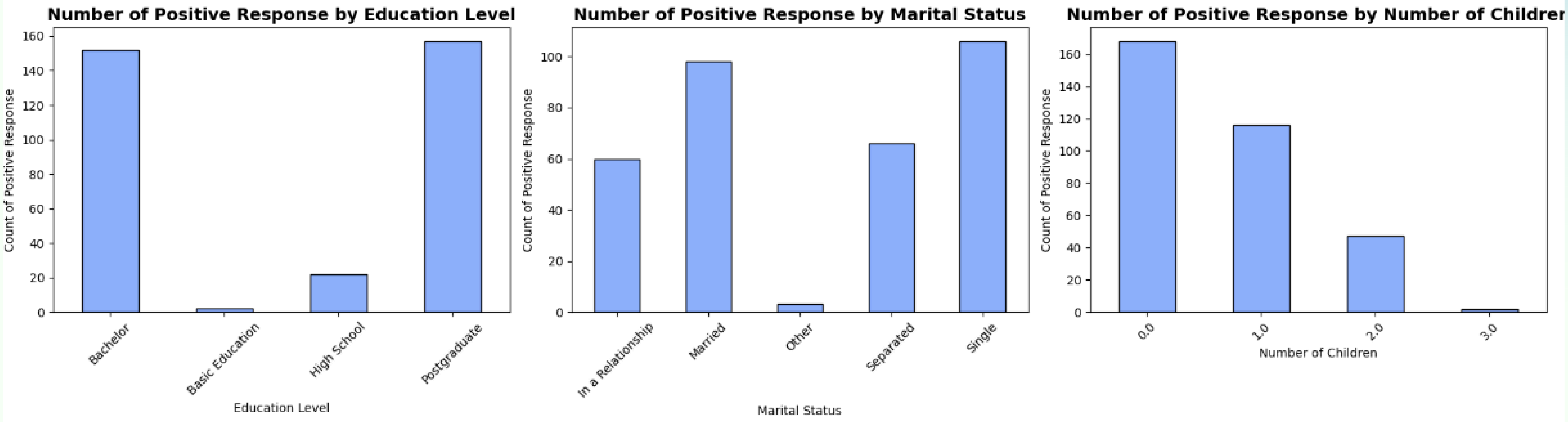


Customers with **higher education levels** tend to have **higher total expenses**, likely due to greater financial stability and higher earning potential.

Marital status appears to have **minimal impact on spending**, as customers across different relationship statuses exhibit similar average expenses.

Families with **more children** tend to **spend significantly less**, likely due to budget constraints and the need to allocate resources more carefully.

02 DATA ANALYSIS & KEY DESCRIPTIVE STATISTICS



Customers with **higher education levels** tend to be **more receptive** to marketing, likely due to a greater interest in premium products and higher purchasing power.

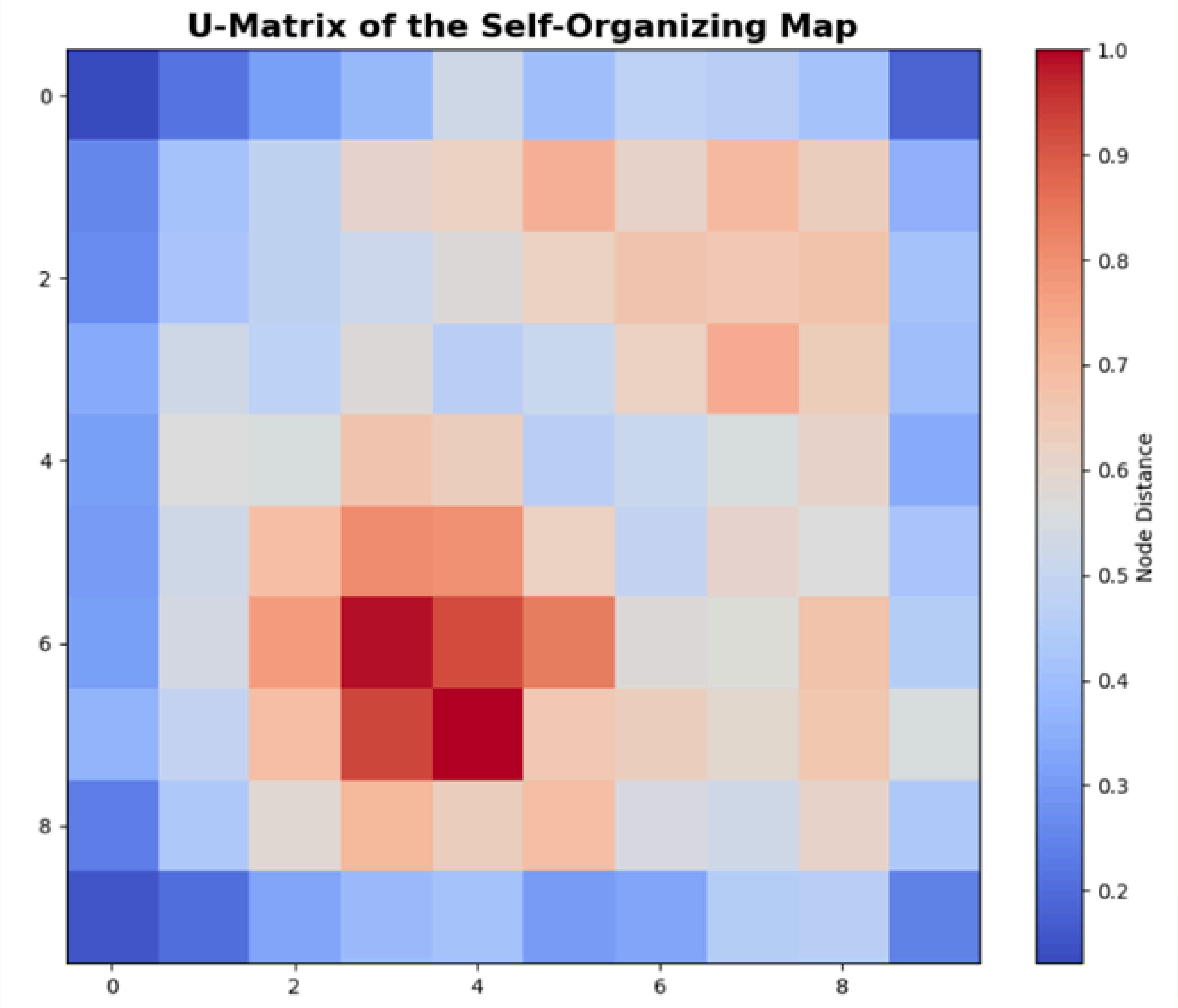
Marital status influences response rates, with **single and married** individuals showing the **highest engagement**, while other groups have lower acceptance rates.

Households with **fewer children** are **more engaged**, with response rates declining as the number of children increases. This suggests that financial flexibility plays a significant role in purchasing decisions.

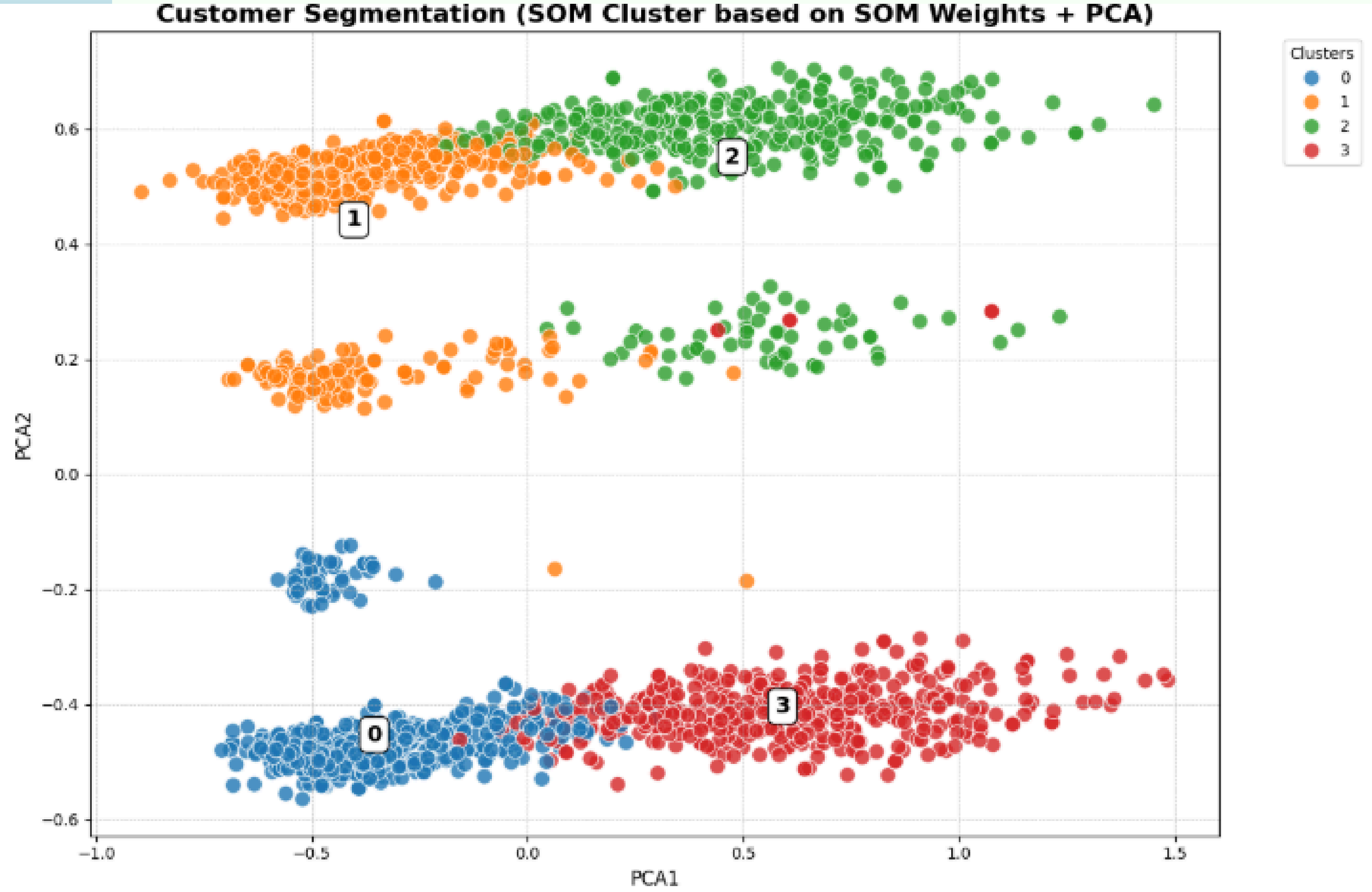
03 MARKET SEGMENTATION USING SELF-ORGANIZING MAPS (SOM) AND K-MEANS

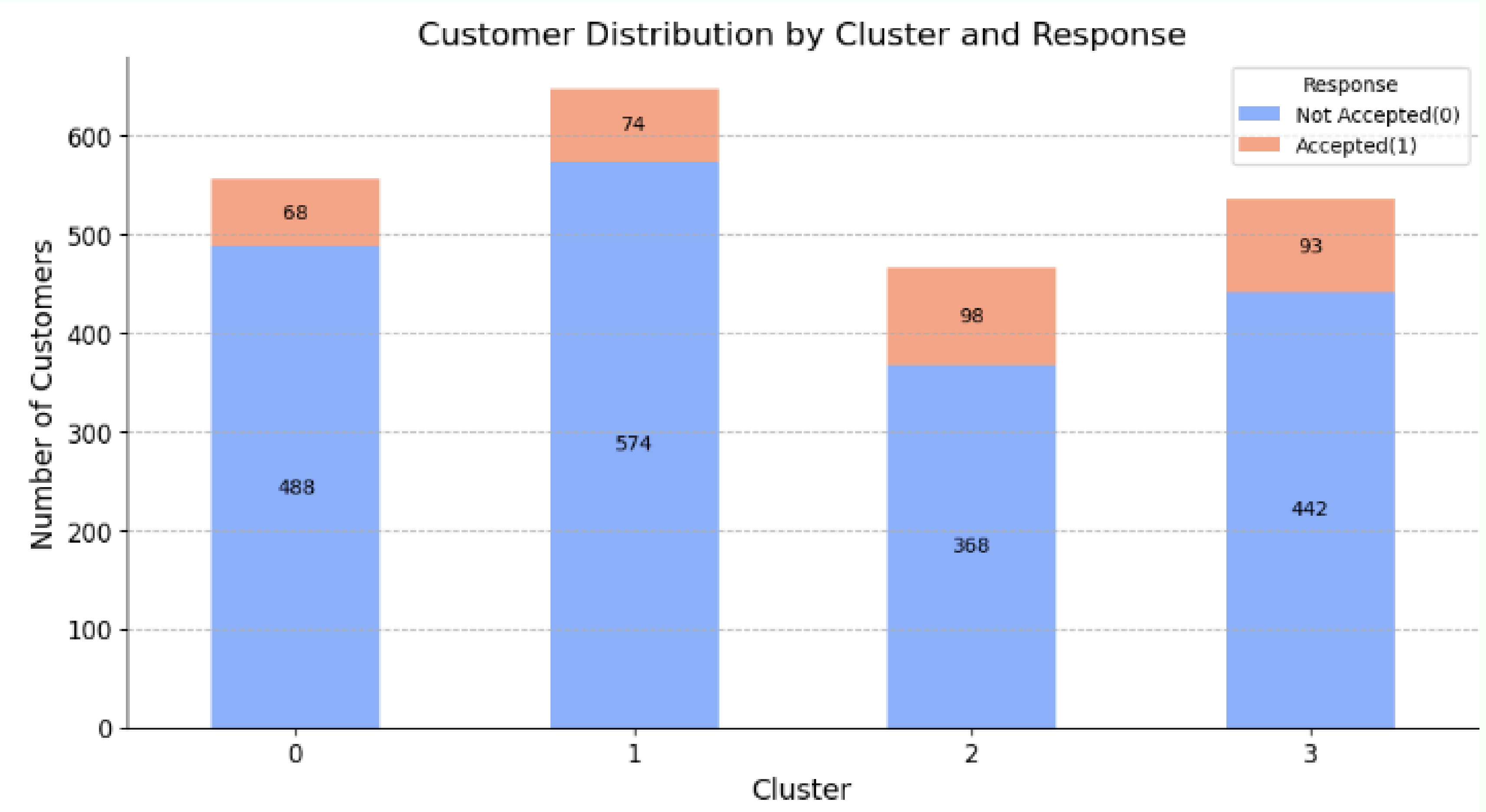
Now, we take a deep dive into market segmentation by applying **Self-Organizing Maps (SOM)**, a powerful machine-learning technique that clusters customers based on **demographics and purchasing behavior**.

By leveraging this approach, we move beyond surface-level insights to uncover **distinct customer segments**, each with unique characteristics, preferences, and responses to marketing strategies. This enables us to tailor our offerings, optimize engagement, and drive more effective decision-making.



03 MARKET SEGMENTATION USING SELF-ORGANIZING MAPS (SOM) AND K-MEANS

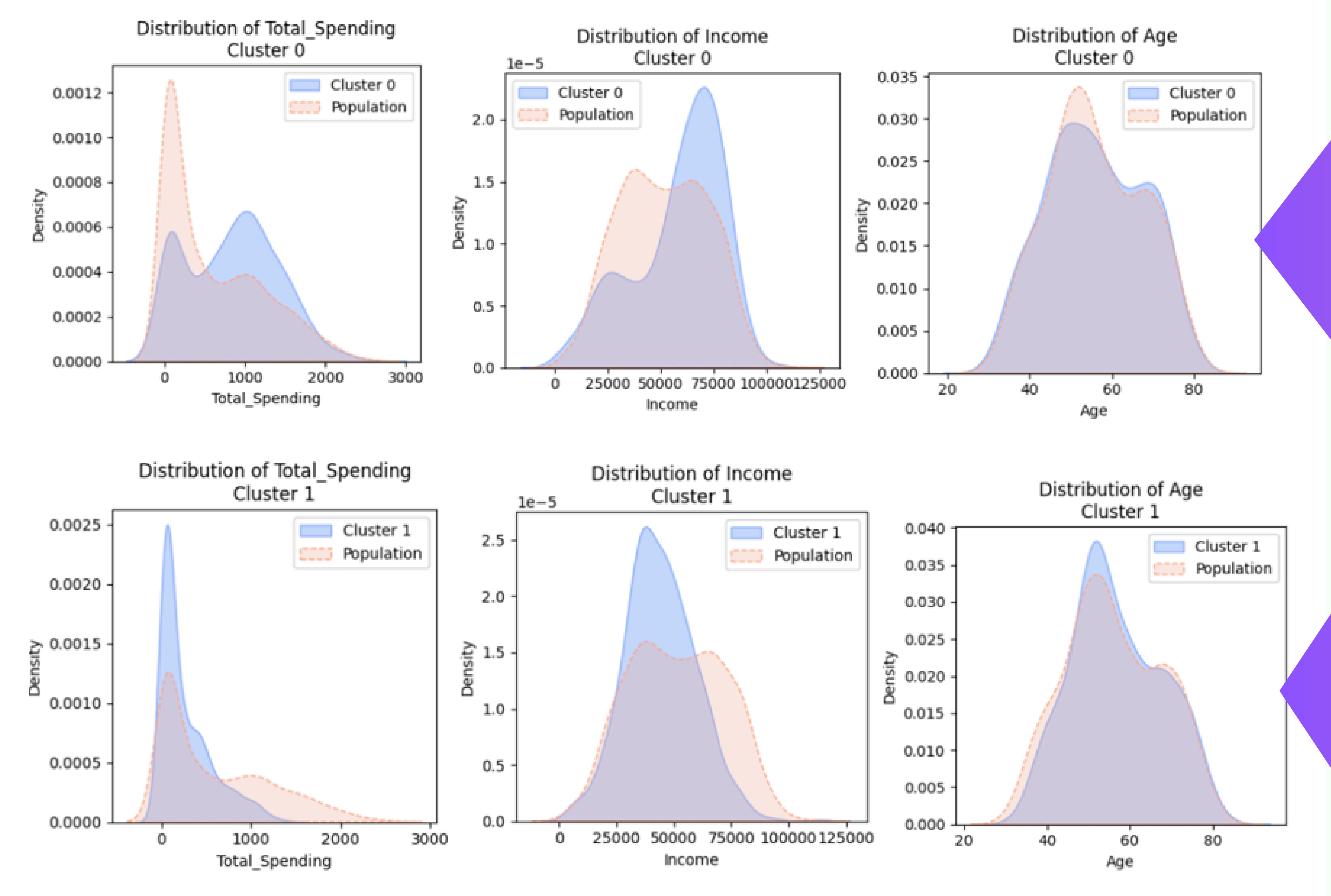




SOM and K-Means segmentation reveals **four distinct customer clusters**, each with unique purchasing patterns.

The left graph shows well-defined clusters, while the right highlights groups like **Cluster 2** with **higher acceptance rates**, making them prime targets for promotions. This segmentation approach enhances **marketing efficiency and profitability** by focusing on **highly engaged customers**.

03 MARKET SEGMENTATION USING SELF-ORGANIZING MAPS (SOM) AND K-MEANS



Cluster 0: Moderately Engaged High-Income

Profile: Middle-aged (56 years old), high income and moderate spending

Engagement: Moderate offer acceptance, selective in choices

Key Insight: Financially stable but require more tailored or compelling offers to increase engagement.

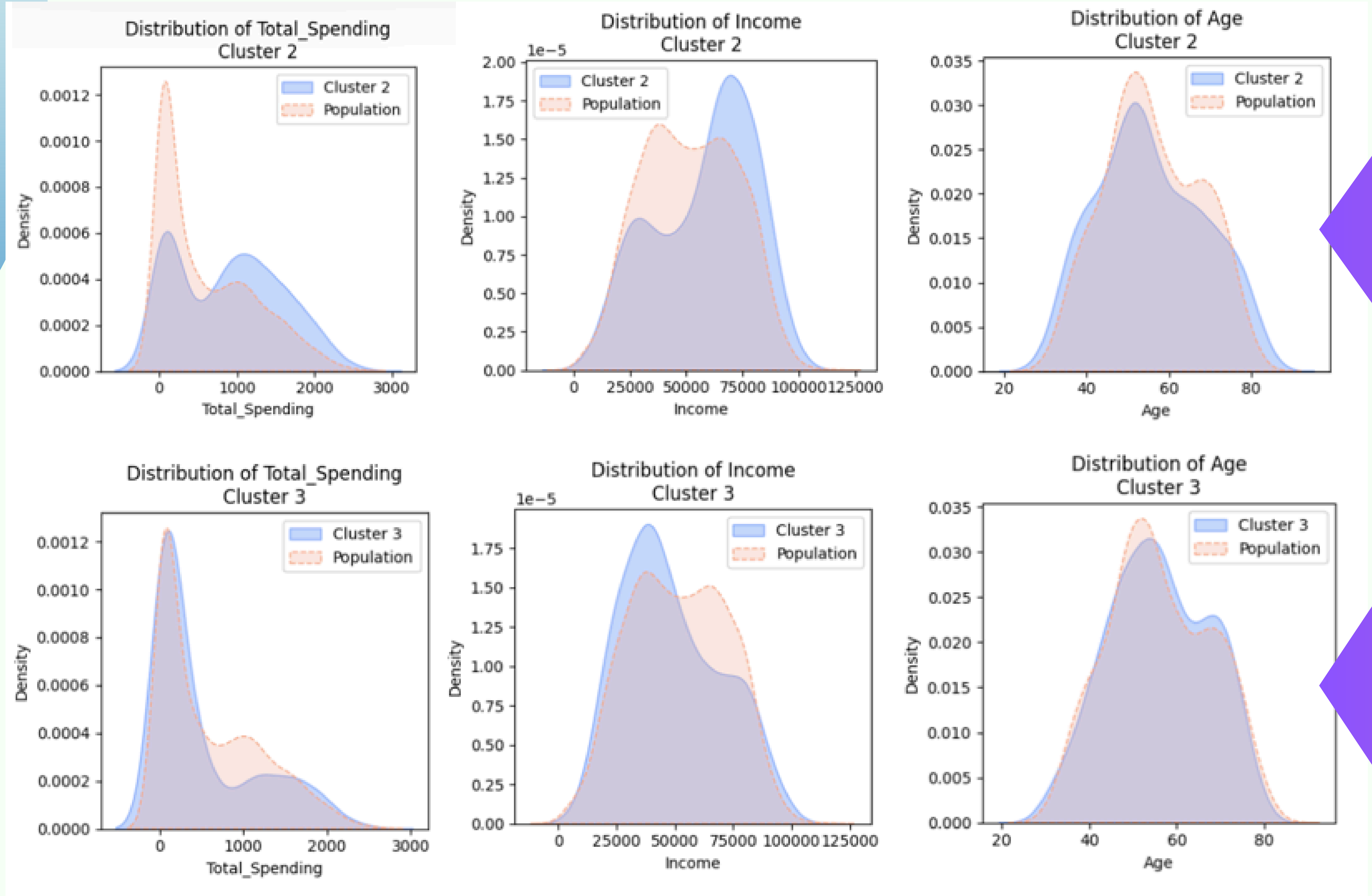
Cluster 1: Budget-Conscious Rejecters

Profile: Middle-aged (56.8 years old), low income, very low spending

Engagement: Low offer acceptance despite recent engagement

Key Insight: More likely to reject offers, possibly due to affordability concerns or lack of interest in the offers provided.

03 MARKET SEGMENTATION USING SELF-ORGANIZING MAPS (SOM) AND K-MEANS



Cluster 2: Affluent Selective Shoppers

Profile: Middle-aged (55.5 years), high income, high spending

Engagement: Moderate to high offer acceptance, but selective

Key Insight: They have strong purchasing power but prefer premium, personalized offers. Exclusive or high-end products could further increase their engagement.

Cluster 3 : Middle-Income Low Spenders

Profile: Middle-aged (55.8 years), moderate income, low spending

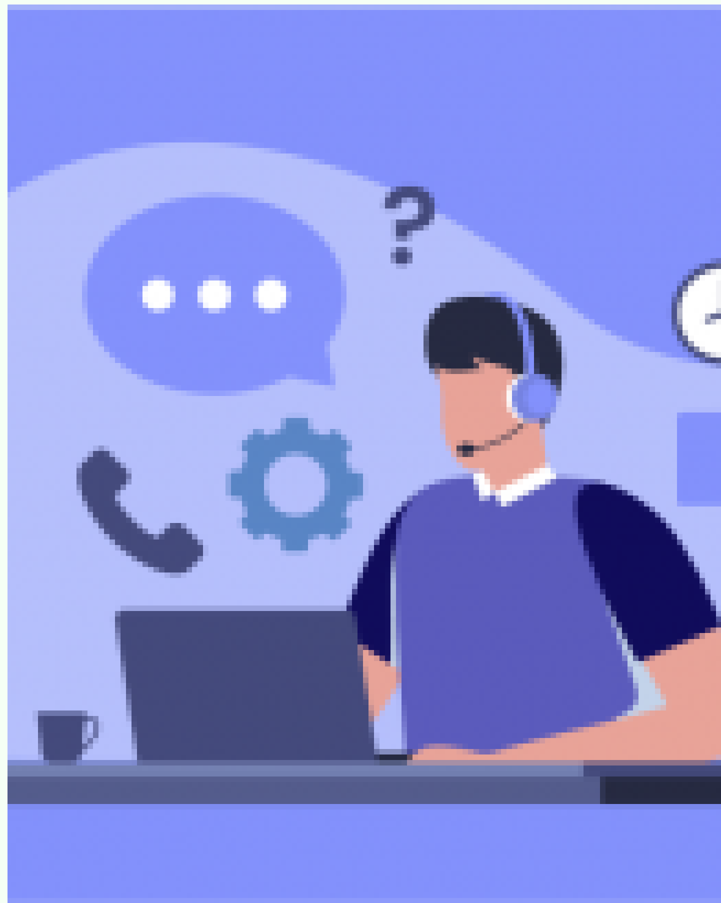
Engagement: Low offer acceptance despite engagement

Key Insight: While financially stable, they are cautious spenders. Budget-friendly offers or targeted discounts.value-driven offers may encourage higher engagement.

03 MARKET SEGMENTATION USING SELF-ORGANIZING MAPS (SOM) AND K-MEANS

Customer Segmentation Overview

CLUSTER	DESCRIPTION
0	**Moderately Engaged High-Income:** Frequent website visits but low overall spending, mainly on wine. Limited response to marketing.
1	**Budget-Conscious Rejecters:** Low-income, low-spending individuals with a low acceptance rate for offers.
2	**Affluent Selective Shoppers:** High-income, high-spending, discerning customers with selective offer acceptance.
3	**Middle-Income Low Spenders:** Middle-income, low-spending individuals who need more compelling offers to increase spending.



03 MARKET SEGMENTATION USING SELF-ORGANIZING MAPS (SOM) AND K-MEANS

Customer Segmentation Results

SUMMARY OF RESULTS

- **Cluster Identification:** The use of SOM and K-Means allowed us to identify homogeneous customer groups, distinguishable by demographic characteristics and purchasing behaviors.
- **Key Differences:** Each cluster shows unique needs and sensitivities towards products, promotions, and marketing messages.
- **Personalization Opportunities:** Marketing strategies and offers can be tailored for each segment, maximizing the effectiveness of promotional initiatives.

OPERATIONAL IMPLICATIONS

- **Targeted Approach:** Prioritize high-potential segments with specific actions.
- **Resource Optimization:** Reduce marketing efforts on clusters with lower purchase propensity or profit margins to increase overall efficiency.
- **Continuous Monitoring:** Incorporate new data to verify the stability of the clusters over time and update the segmentation as needed.

04 PREDICTIVE MODEL: WHY?

To make the next marketing campaign more profitable, we need to focus our efforts on the customers most likely to buy the new gadget.
Predictive models help us do this by analyzing past campaign data and identifying patterns in customer behavior.
By leveraging these insights, we can contact only the most promising customers, reducing costs and increasing revenue. This data-driven approach allows us to make smarter decisions, improving the overall success of the campaign.

SUPPORT VECTOR MACHINE

This model helps us find clear distinctions between customers who are likely to buy and those who are not, even when the differences are subtle.

RANDOM FOREST

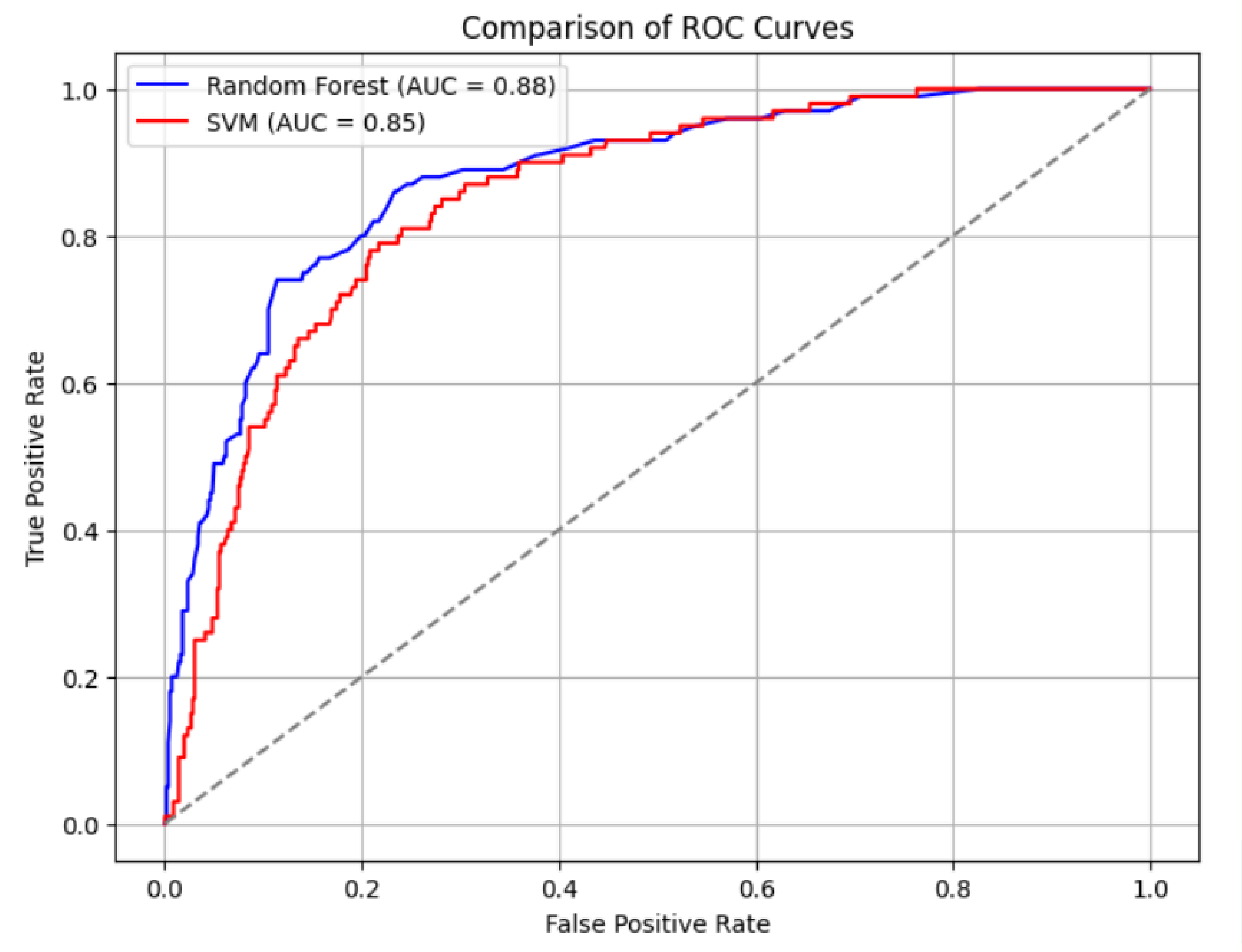
This model makes predictions by combining multiple decision-making processes, making it more reliable and adaptable to different types of customers.

04 PREDICTIVE MODEL: COMPARISON

The **ROC curve (Receiver Operating Characteristic curve)** helps us evaluate the performance of our predictive models by showing how well they differentiate between customers who are likely to buy and those who are not.
A higher **area under the curve (AUC)** means that the model is more accurate in predicting customer behavior.

Random Forest slightly outperformed SVM, achieving a higher AUC of 0.88 compared to 0.85.

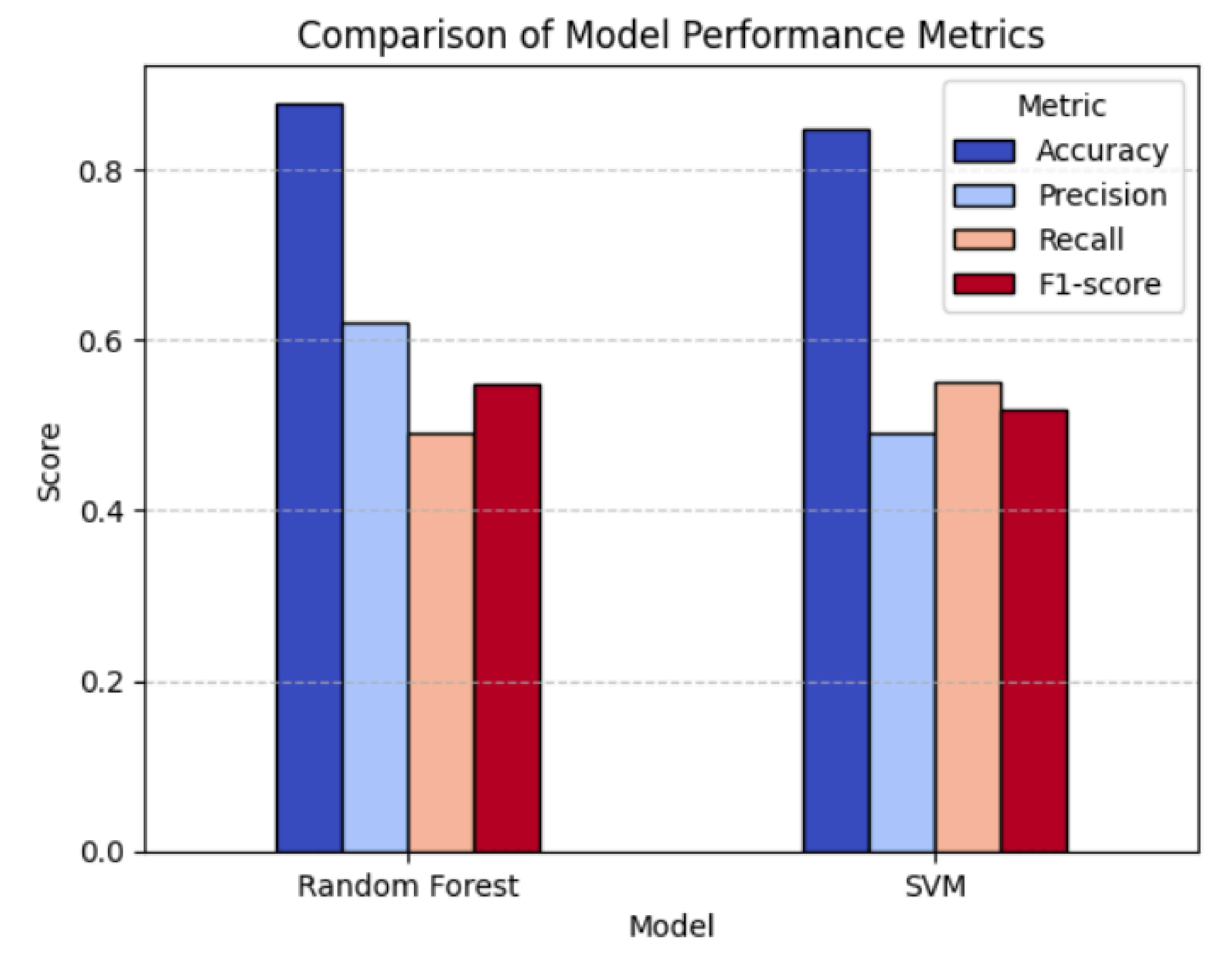
This indicates that **Random Forest was better** at distinguishing between customers likely to buy and those who were not.



04 PREDICTIVE MODEL: COMPARISON

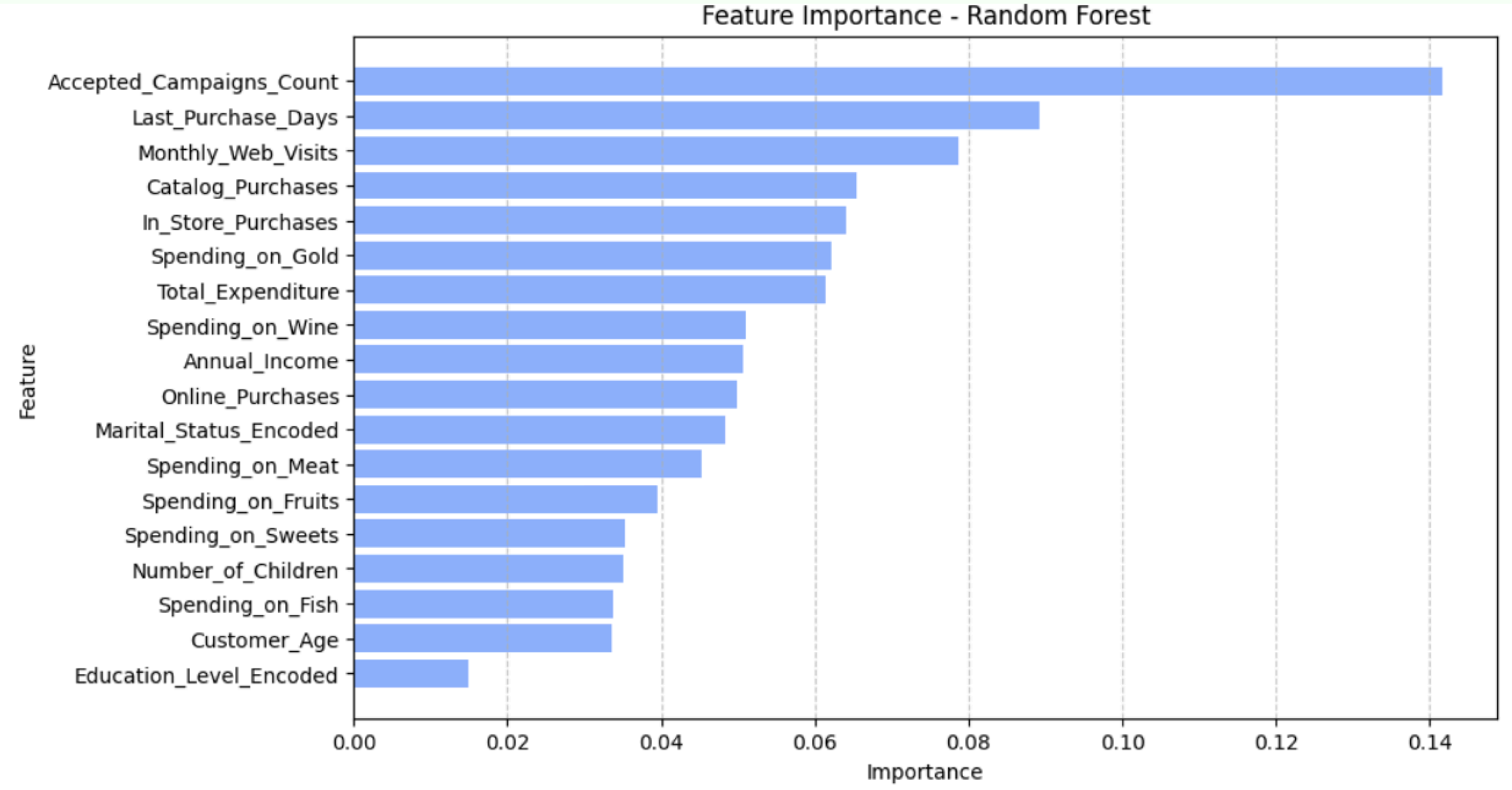
Accuracy shows how many predictions were correct overall.
Precision tells us how many of the predicted "yes" responses were actually correct.
Recall indicates how many of the actual "yes" responses were correctly identified.
F1-score balances precision and recall to give us an overall measure of the model's performance.

Metric	Random Forest	SVM
Accuracy	89%	85%
Precision (0)	93%	93%
Precision (1)	67%	50%
Recall (0)	95%	89%
Recall (1)	59%	62S%
F1-Score (0)	94%	91%
F1-Score (1)	63%	56%



Random Forest maintains a more balanced performance across all metrics, making it the more reliable choice for maximizing campaign effectiveness.

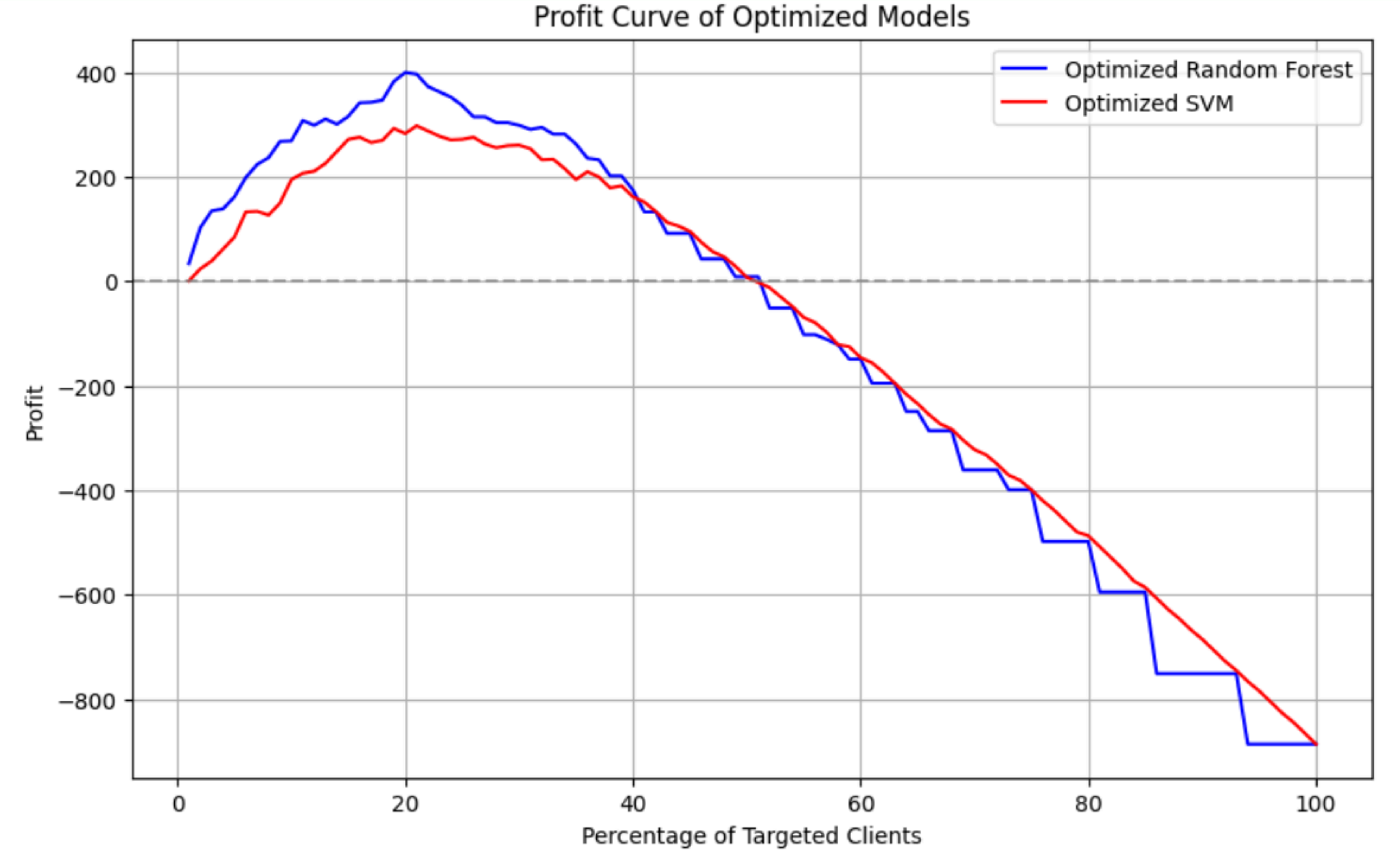
04 PREDICTIVE MODEL: COMPARISON



The Random Forest model highlights that the **most influential variables** in purchasing behavior are the **number of accepted campaigns**, **recency**, **and monthly web visits**, followed by catalog and store purchases.

Variables like i**ncome**, **age**, **and education** have a **lower impact**.

04 PROFIT CURVE ANALYSIS: OPTIMAL CUSTOMER TARGETING



The best contact strategy involves **targeting 20% of customers**, as recommended by the Random Forest model, which maximizes campaign profit at 400 MU.

While the SVM model suggests targeting 21% of customers for higher profits, returns start to diminish beyond this threshold due to increased costs from non-buyers.

Therefore, focusing on the 20% with Random Forest offers the most profitable approach.

05 CONCLUSIONS

At the end, our analysis indicates that the **Random Forest model** outperforms the alternative in predicting customer responses. The **optimal strategy** emerging from the profit curve analysis suggests targeting **the top 20% of customers** based on the **most influential predictive variables**.
This marks a **significant improvement** compared to the previous random targeting approach, which led to a **negative profit (-3.046MU)** and a **low success rate (15%)** in the pilot campaign.

By integrating these results with our **clustering approach**, we gain deeper insights into which customer segments are most valuable for the company. This enables us to design **tailored marketing campaigns** that align with the specific needs and preferences of each cluster.

For instance, **affluent selective shoppers (Cluster 2)** may respond best to premium, exclusive offers, while **moderately engaged high-income customers (Cluster 0)** might require stronger incentives to convert. On the other hand, **budget-conscious rejecters (Cluster 1)** and **middle-income low spenders (Cluster 3)** may benefit from value-driven promotions.

Ultimately, this **data-driven approach** allows us to **allocate resources efficiently**, ensuring that the next campaign is both **profitable for the company** and **valuable for the customer**, maximizing engagement and return on investment.