

DECEMBER 2021 - PROJ. 01
FACUNDO CASTELLÁ

Marketing Campaign Analysis

About the Project

A detailed analysis of the customers for a Grocery Store will be conducted. This will help the company better understand its customers and facilitate the process of modifying and updating its products based on the specific needs, behaviors, and concerns of each customer.

Goal: Conduct a clustering analysis on the Marketing Campaign to simplify and summarize customer segments.

When: December 2021

Data Analytics

Data Science

Machine Learning

Marketing

Business Analytics

Python

EDA

PCA

t-SNE

K-Means

Agglomerative Clustering

Field

Tags

Language

Skills

About the Database

Name: Marketing Campaign

The dataset used for this analysis comes from a marketing campaign dataset available on Kaggle, which focuses on customer behavior and preferences. It provides rich demographic, behavioral, and transactional data that can be used to perform customer segmentation and target specific customer groups effectively.

29x2,240

Fields x
Records

Synthetic

Type

CSV

Format

Kaggle

Source

About the Presentation

Presentation will have technical  and non-technical slides . Technical slides will be noted. If you consider yourself technical you can read all the slides, if not you should be able to understand the problem, analysis and conclusions by only reading the non technical slides

Background & Motivation

Background

In the past month, only 4.53% of the 6 campaigns executed were accepted by the target audience, with a wide variance in acceptance rates across individual campaigns (e.g., Campaign 2 reached 1.34%, while Campaign 6 performed best at 14.9%). This suggests inconsistent engagement and the need to reassess campaign strategies. Additionally, only 8.55% of accepted campaigns converted into a purchase, indicating a significant drop-off at the conversion stage. A deeper concern is that only 0.19% of total monthly purchases resulted in deal closures, highlighting inefficiencies in closing opportunities. Finally, of all website visits, a mere 0.77% ended in a purchase, signaling the potential for optimization in the online customer journey to improve conversion rates.

4.53%

Accepted
Campaigns

8.55%

Deals
Purchases /
Accepted
Campaigns

0.19%

Deals
Purchases /
All Purchases

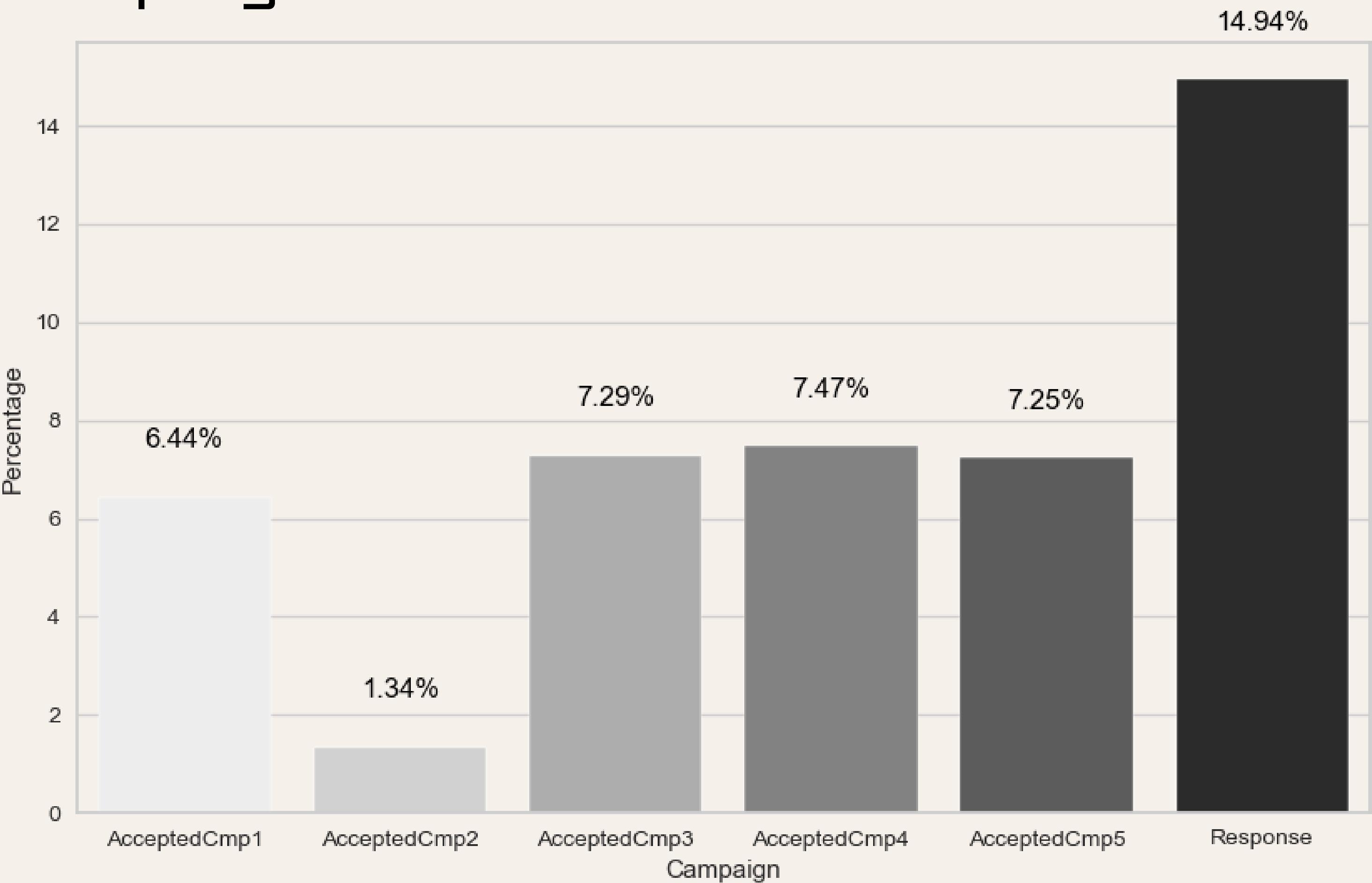
0.77%

Web
Purchases /
Web Visits

Motivation

To maximize campaign performance, it's crucial to segment the target audience, analyze their behaviors, and identify which campaign types resonate best. This approach will enable us to tailor strategies more effectively, driving higher acceptance and conversion rates across all touchpoints.

Percentage Of Acceptance Per Campaign



Methodology

Methodology

00 INITIAL FEATURES

Customer Attributes:

- **ID:** Unique identifier for each customer
- **Year_Birth:** Year of birth of the customer
- **Education:** Education level of the customer
- **Marital_Status:** Marital status of the customer
- **Income:** Annual household income of the customer
- **Kidhome:** Number of children in the customer's household
- **Teenhome:** Number of teenagers in the customer's household
- **Dt_Customer** Date when the customer enrolled with the company
- **Recency:** Number of days since the customer's last purchase
- **Complain:** 1 if the customer has complained in the last two years, 0 otherwise

Product Attributes:

- **MntWines:** Amount spent on wine in the last two years
- **MntFruits:** Amount spent on fruits in the last two years
- **MntMeatProducts:** Amount spent on meat in the last two years
- **MntFishProducts:** Amount spent on fish in the last two years
- **MntSweetProducts:** Amount spent on sweets in the last two years
- **MntGoldProds:** Amount spent on gold products in the last two years

Methodology

00 INITIAL FEATURES

Promotion Attributes:

- **NumDealsPurchases:** Number of purchases made with a discount
- **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
- **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

- **Response:** 1 if the customer accepted the offer in the latest campaign, 0 otherwise

Place Attributes:

- **NumWebPurchases:** Number of purchases made through the company's website
- **NumCatalogPurchases:** Number of purchases made via catalogue
- **NumStorePurchases:** Number of purchases made directly in stores
- **NumWebVisitsMonth:** Number of visits to the company's website in the last month

Methodology

01 DATA CLEANSING & FEATURE ENGINEERING

- “Education” was simplified into broader groups (Undergraduate, Graduate, PostGraduate) and converted the column to a categorical data type.
- “Marital_status” categories were grouped into “Living With” (Partner, Alone) and converted to a categorical data type, with unnecessary categories being dropped.
- “Customer since” is created by converting the “Dt_Customer” column into a datetime format and calculating the difference between the current date and the enrollment date, storing the result as the number of days since enrollment.
- “Age” is created by subtracting the “Year_Birth” from the current date, which calculates each individual’s age based on their year of birth.
- “TotalSpent” is created by summing the amounts spent on various product categories, including wine, fruits, meat, fish, and sweets.
- “Sons” is calculated by adding the number of children at home (Kidhome) and the number of teenagers at home (Teenhome).
- “Is_Parent” is created using a conditional statement that assigns a value of 1 if the customer has any children (if the total number of “Sons” is greater than 0) and 0 if not.

Methodology

01 DATA CLEANSING & FEATURE ENGINEERING

- “Living_With_Num” is created by replacing the “Living_With” values with numerical equivalents, where “Alone” is mapped to 1 and “Partner” is mapped to 2, and then converting the result to an integer type.
- “fam_size” is calculated by adding the number of “Sons” to the “Living_With_Num” value, which gives the total family size.
- “Generation” is calculated using the Year_Birth column. The following categories were defined: Silent, Boomers, Generation X, Millennials, and Generation Z.
- Outliers were removed in ‘Income’, and ‘Age’ columns
- Columns were dropped: [‘Year_Birth’, ‘Dt_Customer’, ‘Kidhome’, ‘Teenhome’, ‘Z_CostContact’, ‘Z_Revenue’]
- Label Encoding: Label Encoding is a popular encoding technique for handling categorical variables. In this technique, each label is assigned a unique integer based on alphabetical ordering.

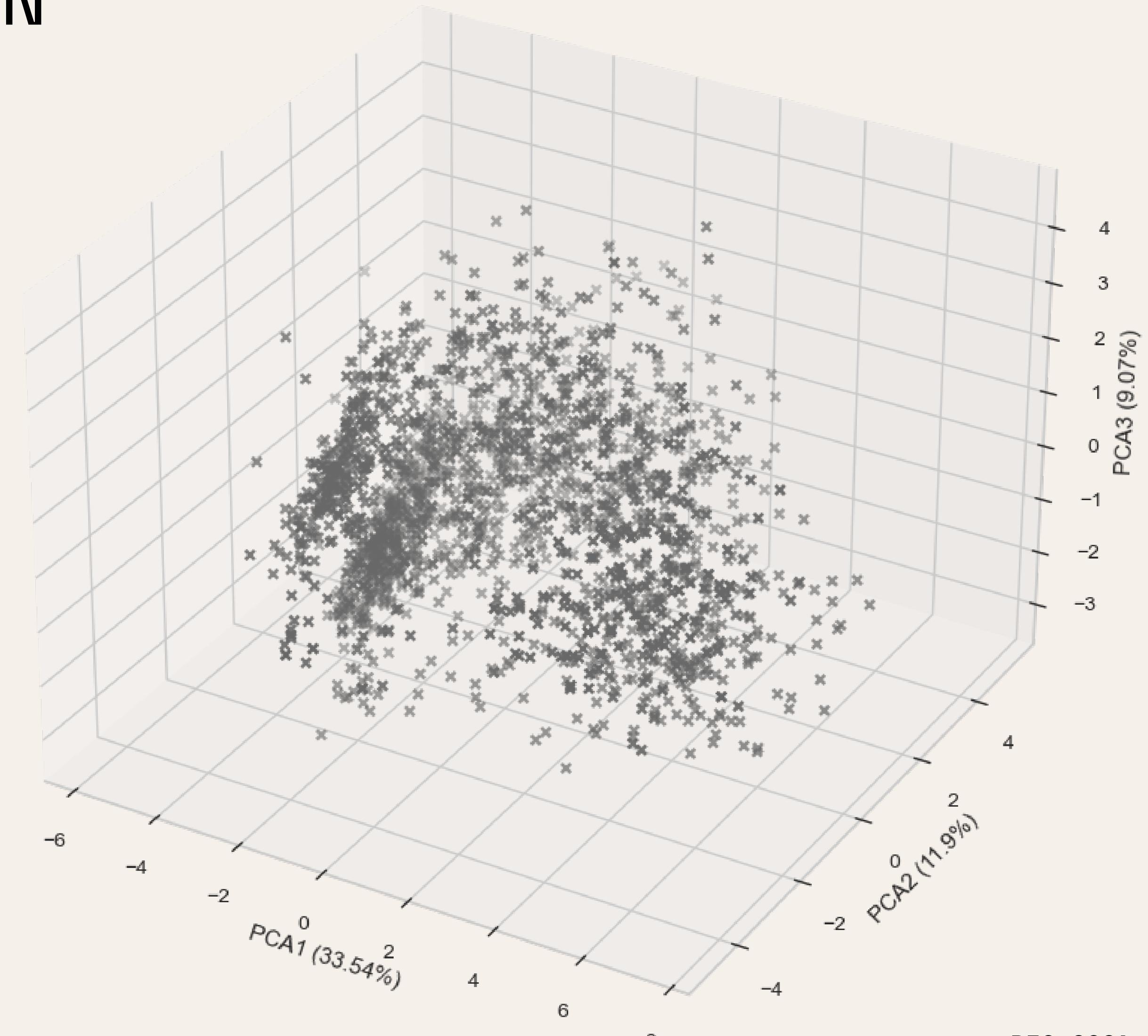
Methodology

02 DIMENSIONALITY REDUCTION

Two key approaches were applied for dimensionality reduction: PCA and t-SNE. For visualization purposes, we focused on reducing the dimensions to 3 main components.

On the right, the PCA results are displayed, where these 3 components capture 54.51% of the total variance. Additionally, we can already observe some clear cluster formations emerging from the data.

3D Projection of the Reduced Data (PCA)



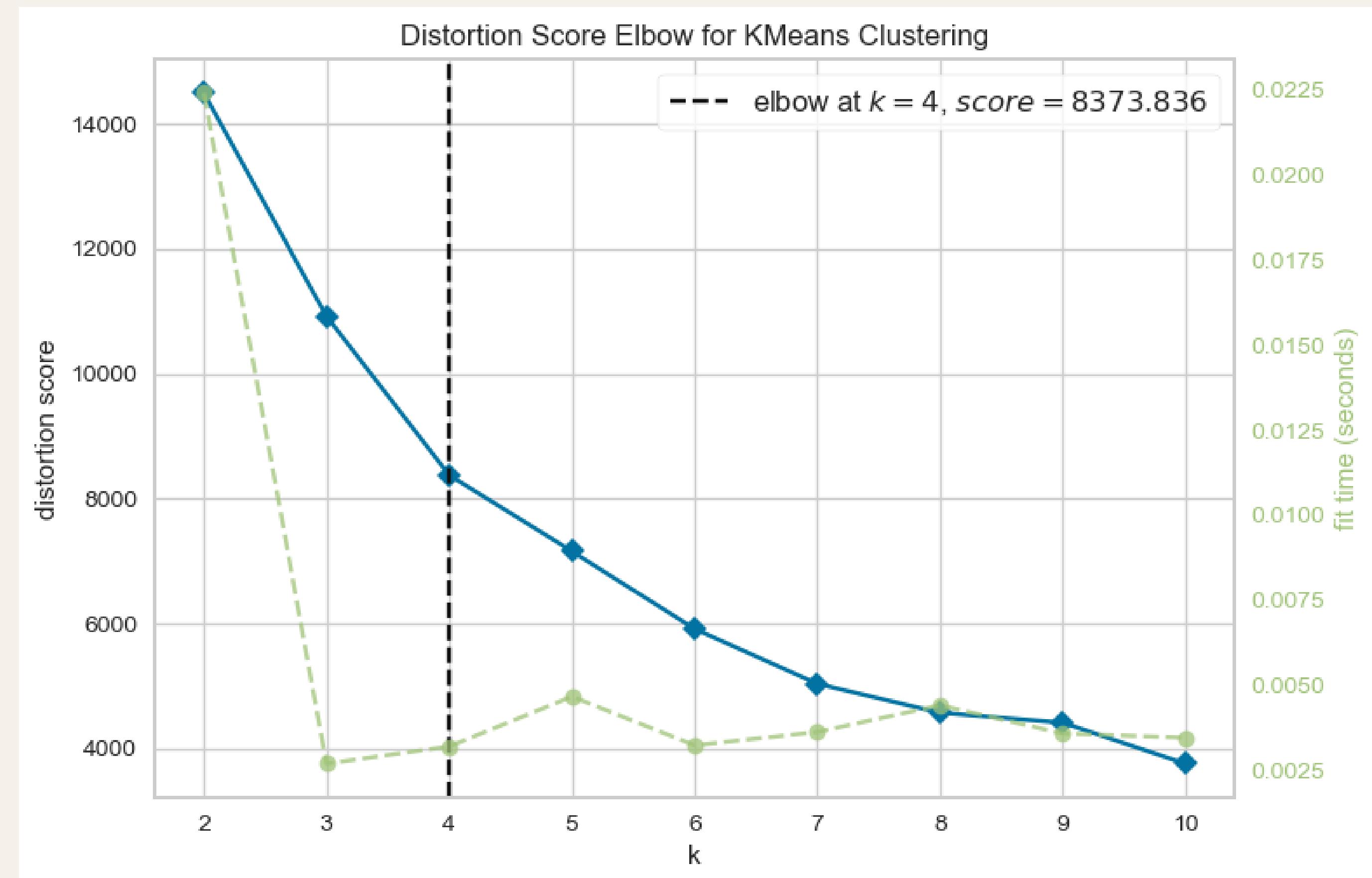
Methodology

03 CLUSTERING

We began by using the elbow method to identify the optimal number of clusters.

The method involves plotting the distortion score (within-cluster sum of squares) against the number of clusters, and the point where the distortion starts to decrease more slowly (forming an “elbow”) suggests the ideal number of clusters.

In the image, we can see that the elbow occurs at $k = 4$, where the score significantly drops before leveling off, indicating that 4 clusters is the most suitable choice.



Methodology

03 CLUSTERING

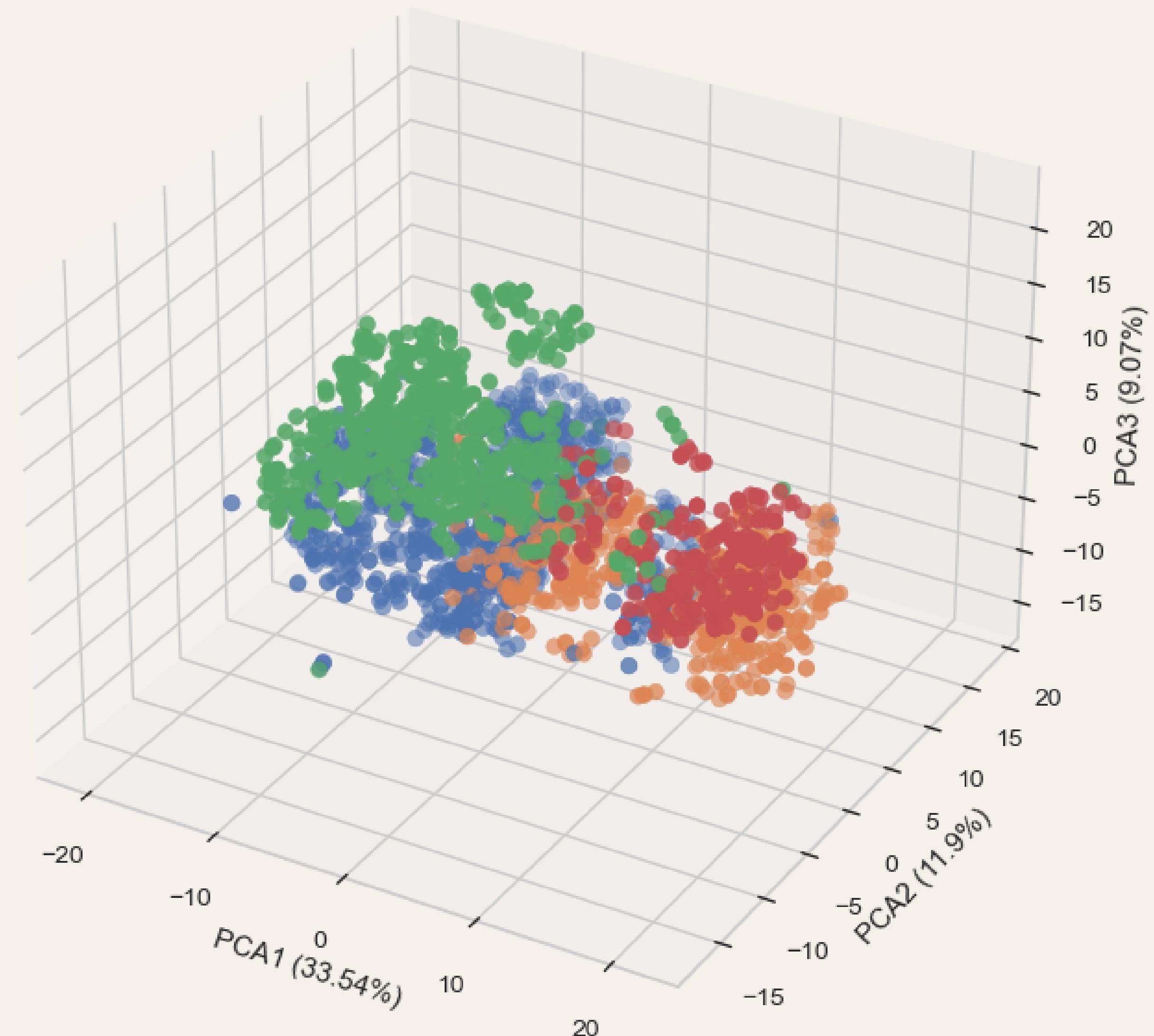
For clustering, we explored two methods:
Agglomerative Clustering and K-Means.

On the right, the results of Agglomerative Clustering are displayed, which differs from K-Means by its hierarchical approach, merging clusters iteratively based on similarity.

The four distinct clusters are clearly visible in the PCA-reduced dimensional space, highlighting the natural groupings within the data.

Now we can proceed to analyze each cluster...

Cluster Plot (Agglomerative Clustering)



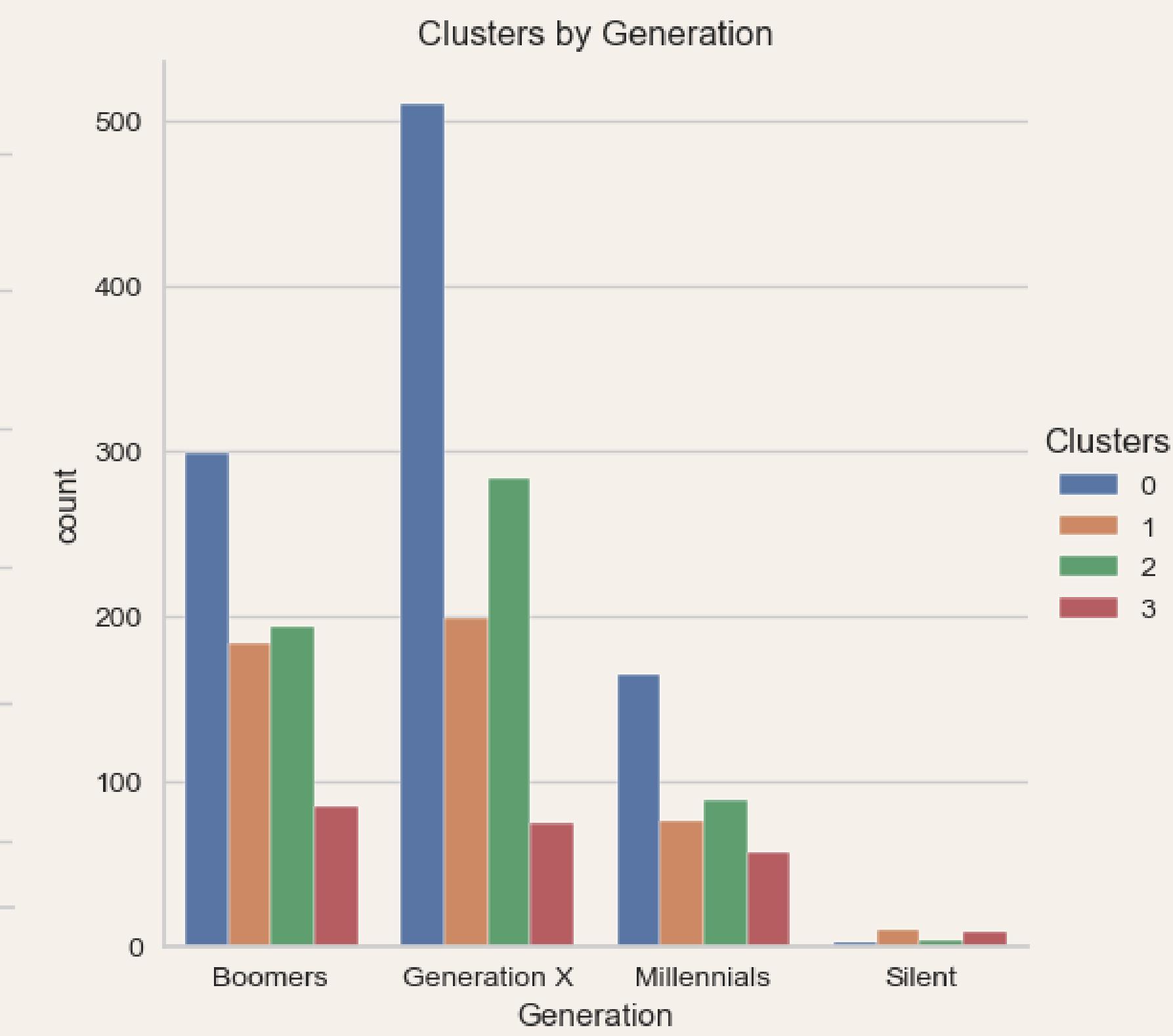
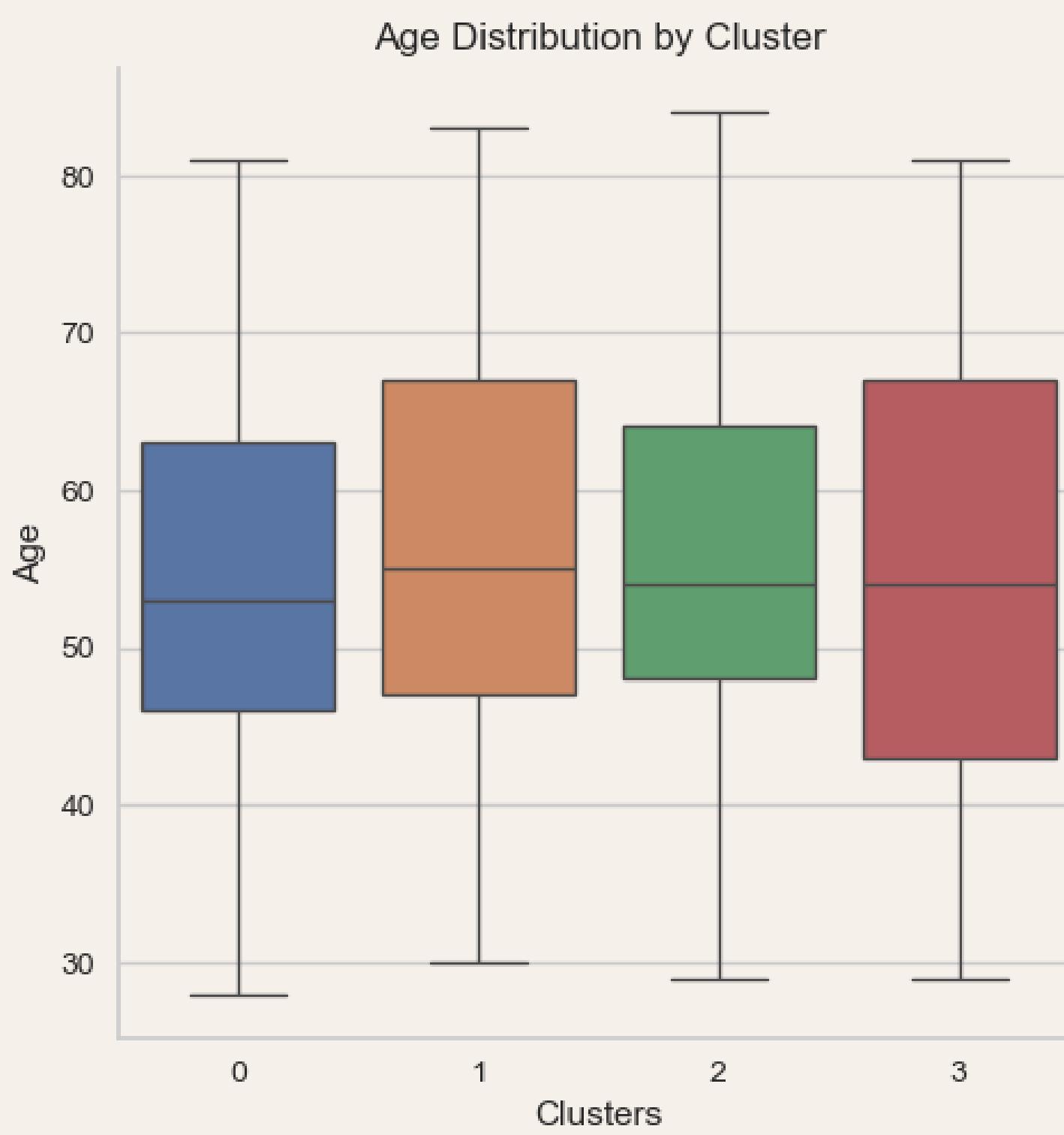
Analysis & Results

Analysis & Results

01 PERSONA MAIN ATTRIBUTES

All clusters have similar overall ranges, spanning from approximately 30 to 80 years of age. And, the median age across clusters seems to hover around 50 to 60 years.

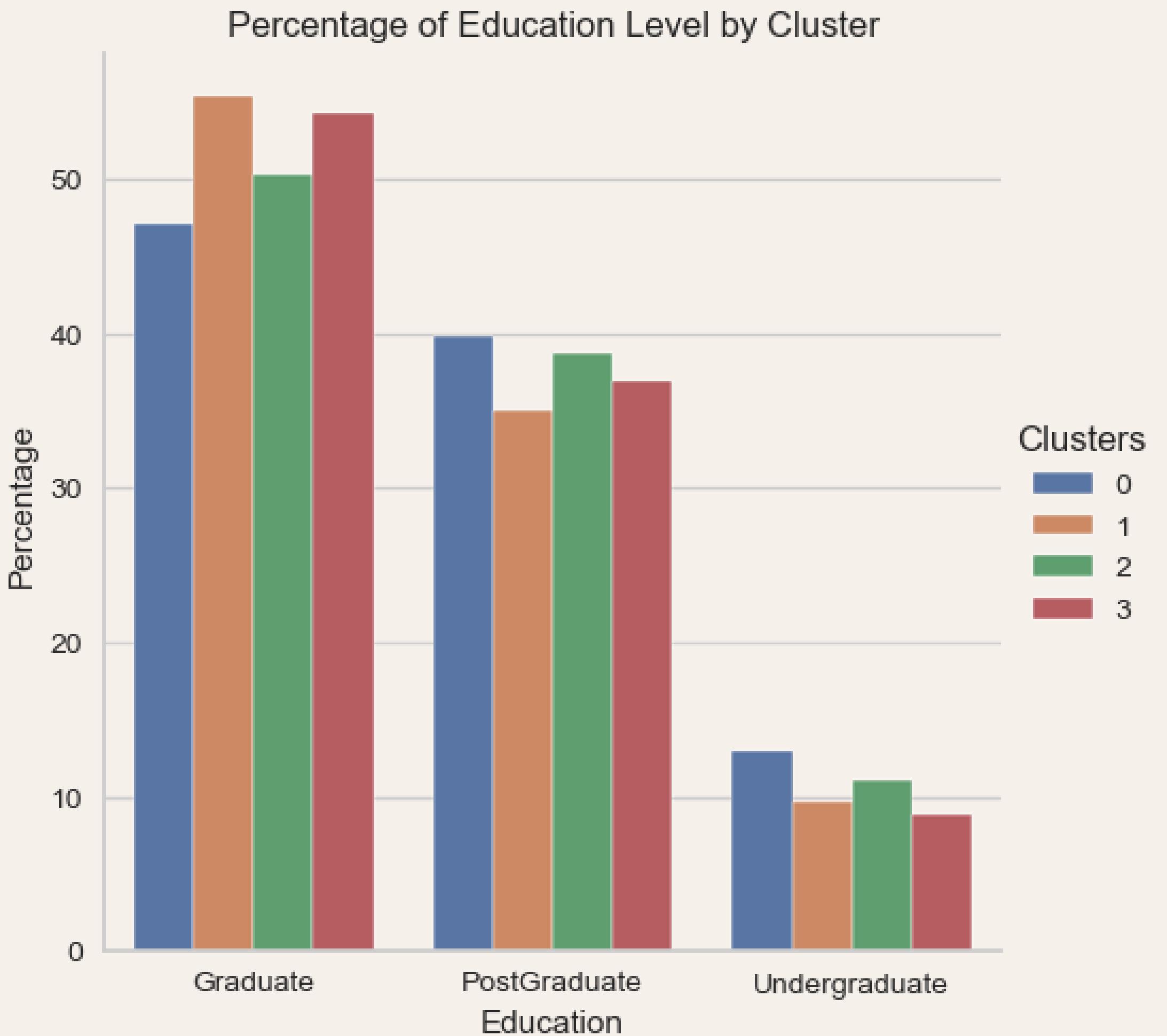
Clusters seem to capture distinct generational and age-based groupings. Cluster 0 and 2 appears to contain more middle-aged (Generation X and Boomers), while Cluster 1 and 3 have a more evenly distributed generational composition.



Analysis & Results

01 PERSONA MAIN ATTRIBUTES

All clusters show a similar distribution in terms of education, with approximately 50% of individuals being Graduates, 40% being Postgraduates, and around 10% being Undergraduates.

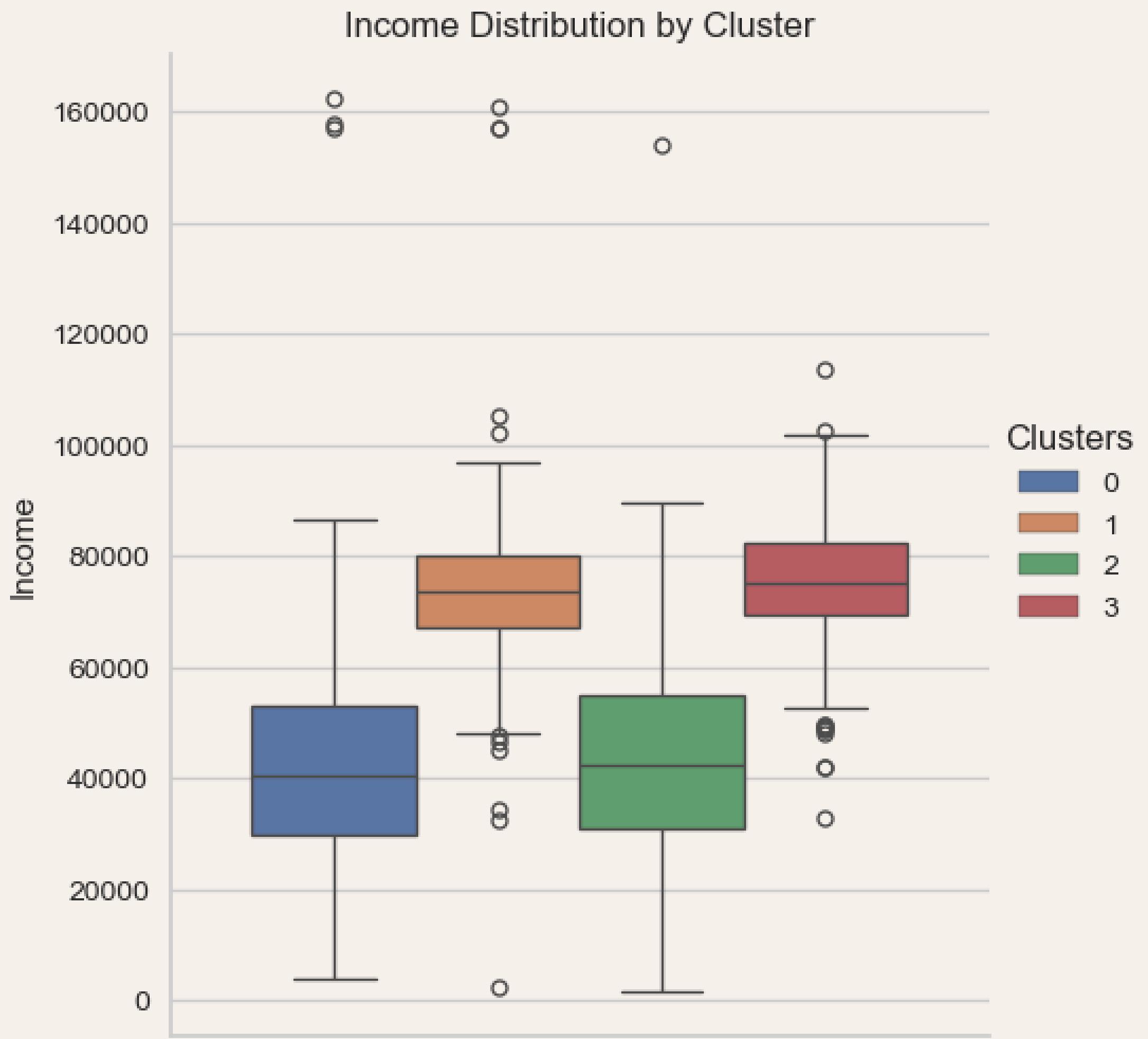


Analysis & Results

01 PERSONA MAIN ATTRIBUTES

The overall pattern shows that while Clusters 1 and 3 represent higher income groups, Clusters 0 and 2 tend to have lower incomes.

Having into account that this is a US dataset, we could classify as Cluster 0 and 2 as Lower-middle class, Cluster 1 as Middle class; and Cluster 3 as Middle class and as Upper-middle class.

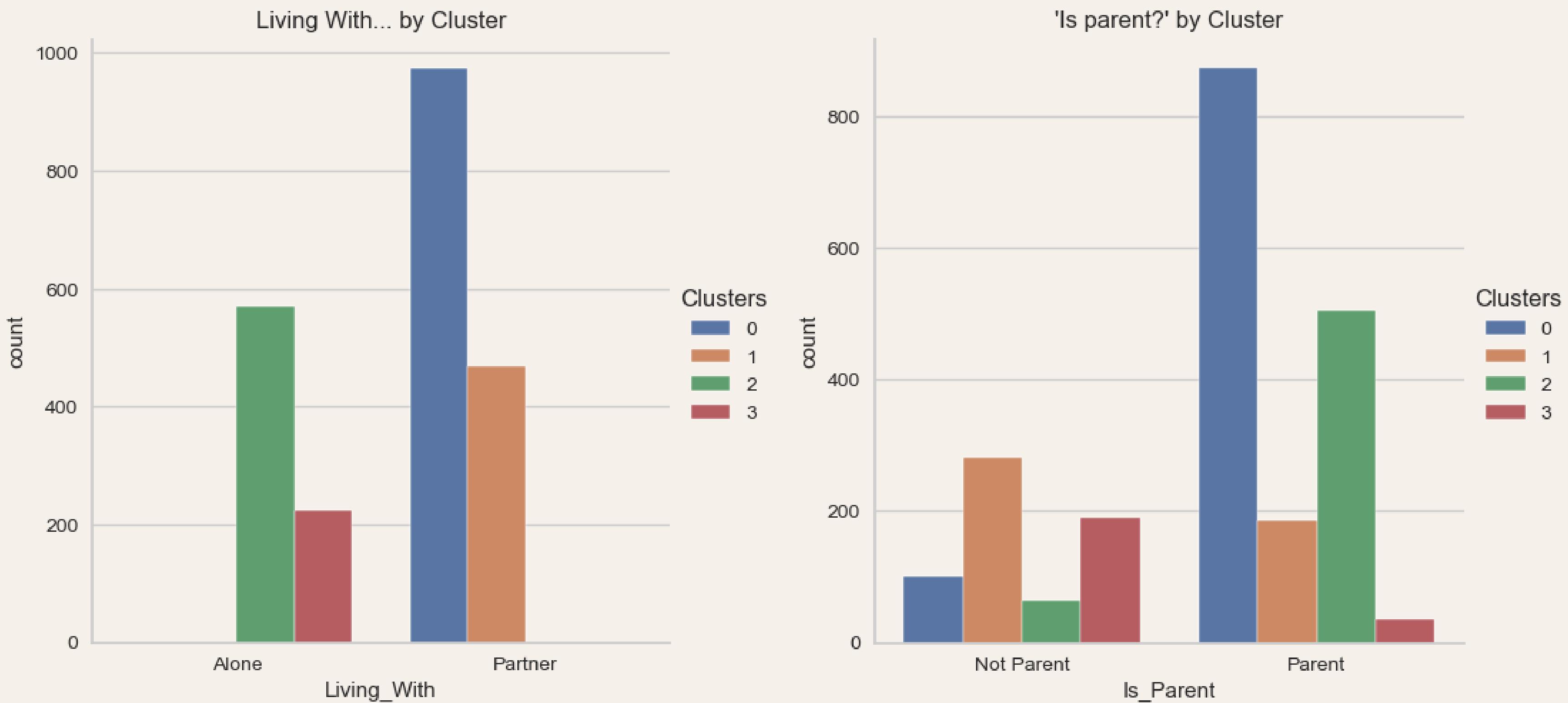


Analysis & Results

02 FAMILY SHAPE

Clients that belong to cluster 2 and 3 mainly live alone; and clients that belong to cluster 0 and 1 live with a Partner.

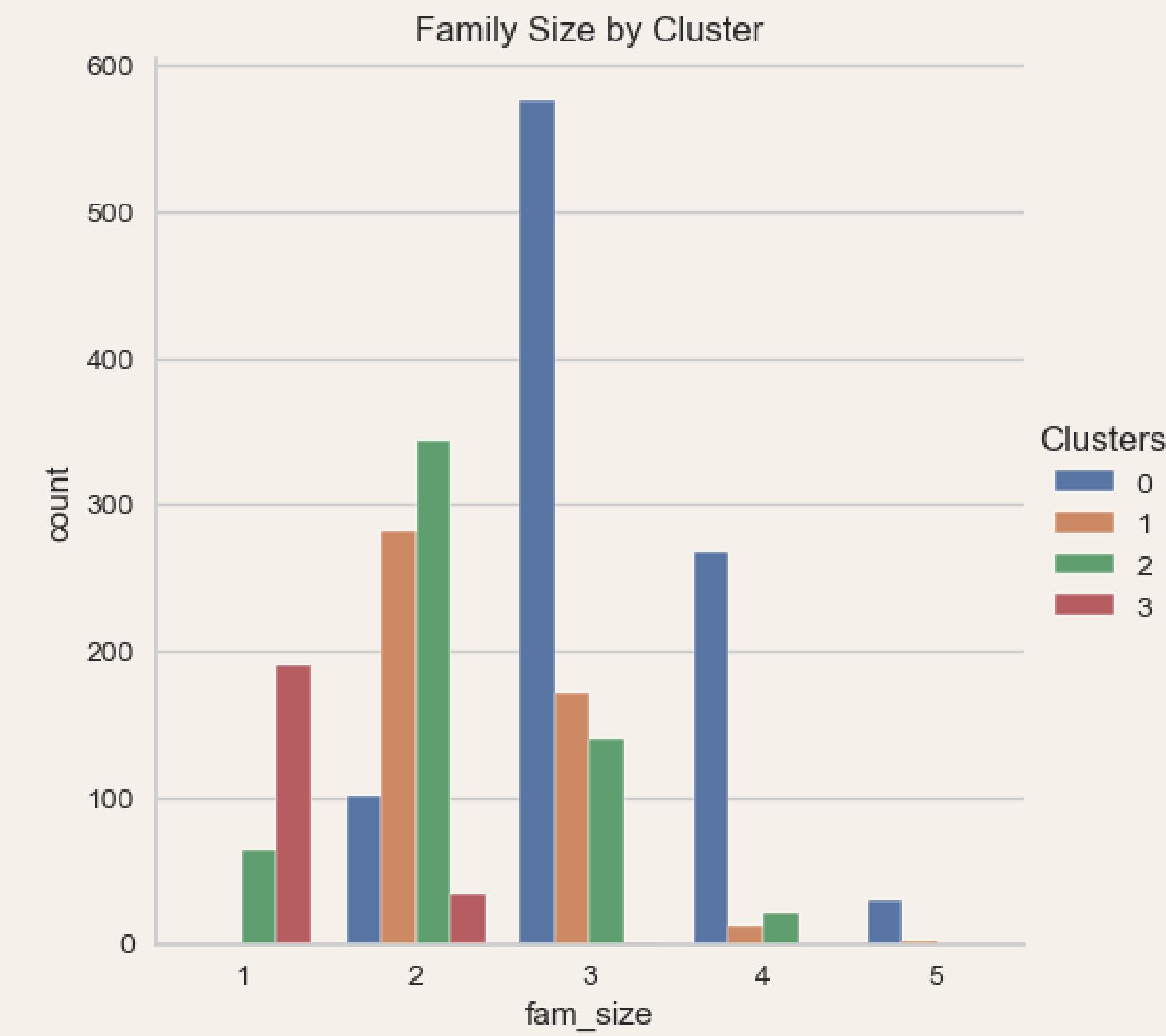
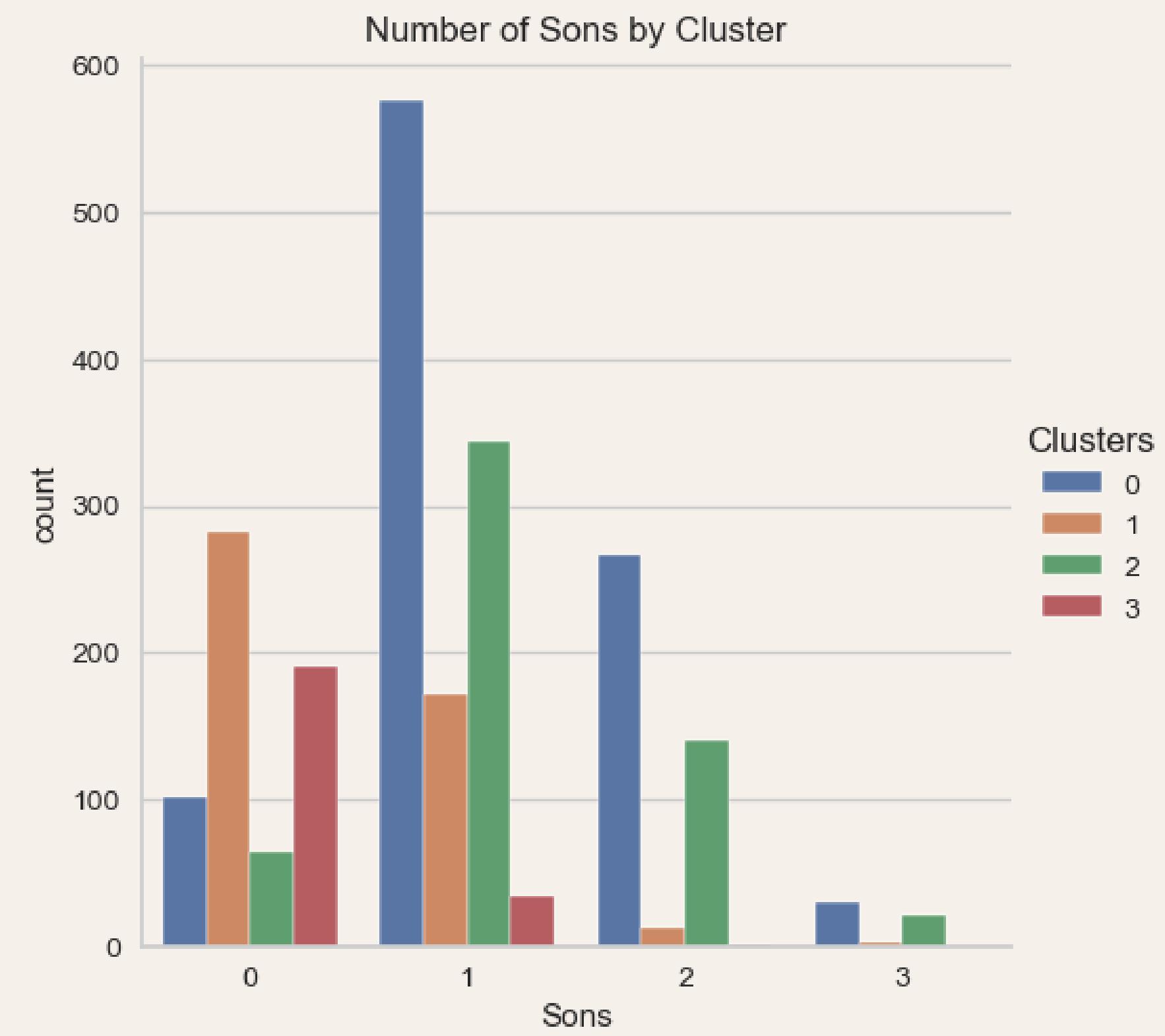
Additionally, Clusters 0 and 2 have a strong representation of parents, whereas Cluster 3 has more non-parents, suggesting differences in family structure across clusters.



Analysis & Results

02 FAMILY SHAPE

In summary, Cluster 0 represents larger family structures, while Clusters 1 and 2 show more diversity in family sizes and number of sons. Cluster 3 generally represents smaller families with fewer children and single people.

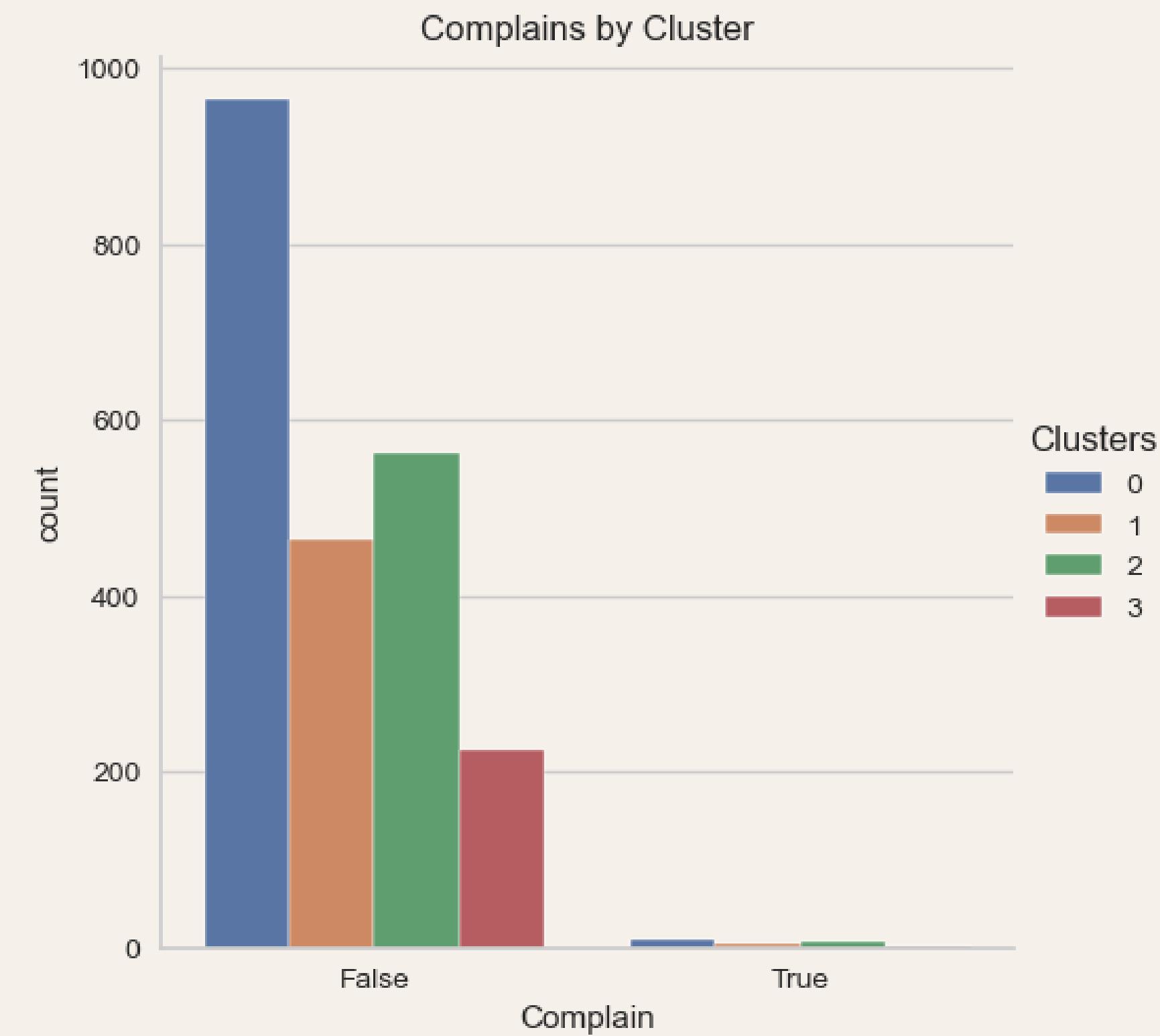
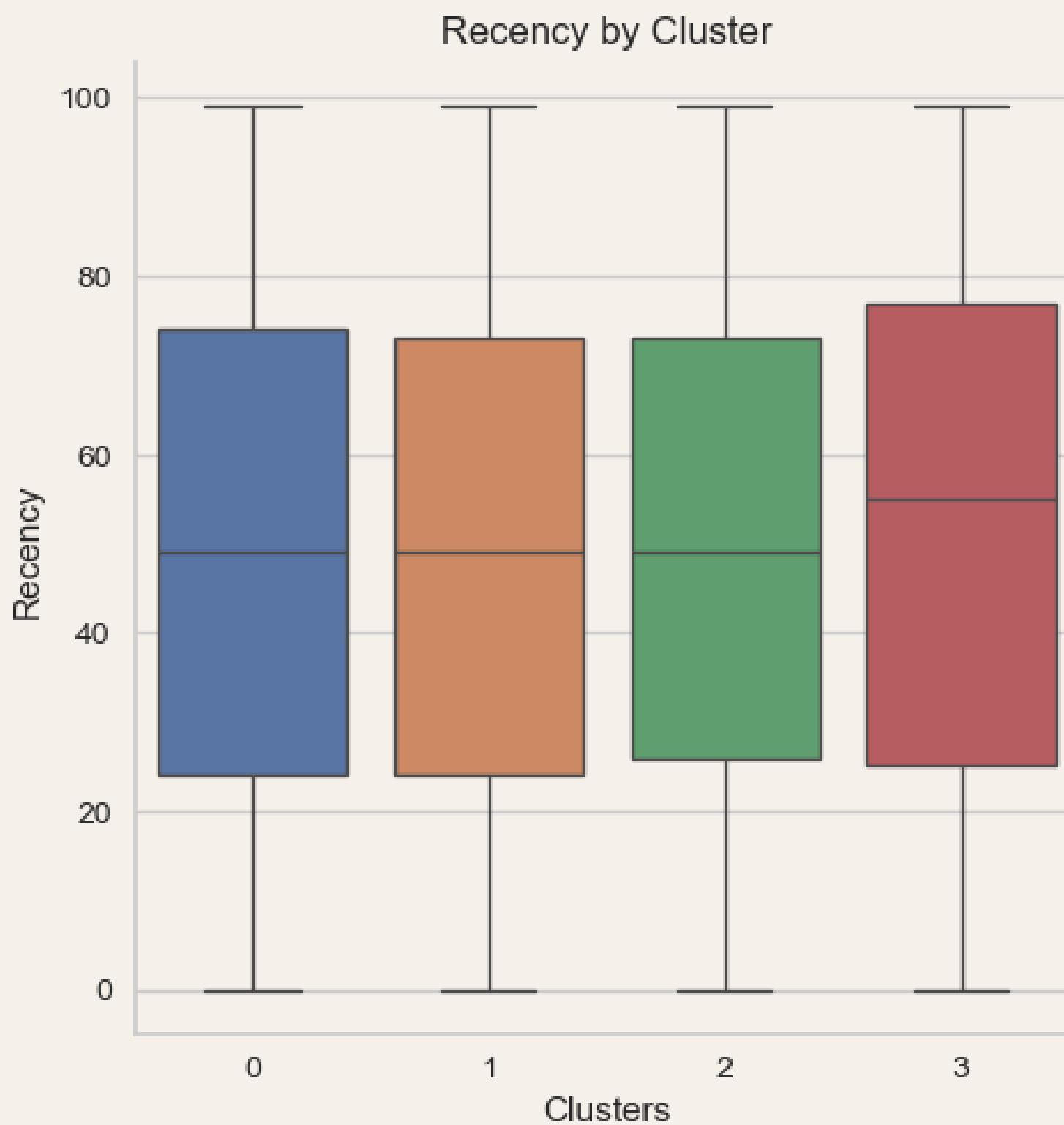


Analysis & Results

03 FIDELITY

There doesn't seem to be a distinct pattern in terms of one cluster having significantly more recent customers than another, suggesting that recency isn't a strong differentiating factor between clusters.

All groups seem to have customers who do not present complaints, with a majority in each cluster showing no complaints at all.

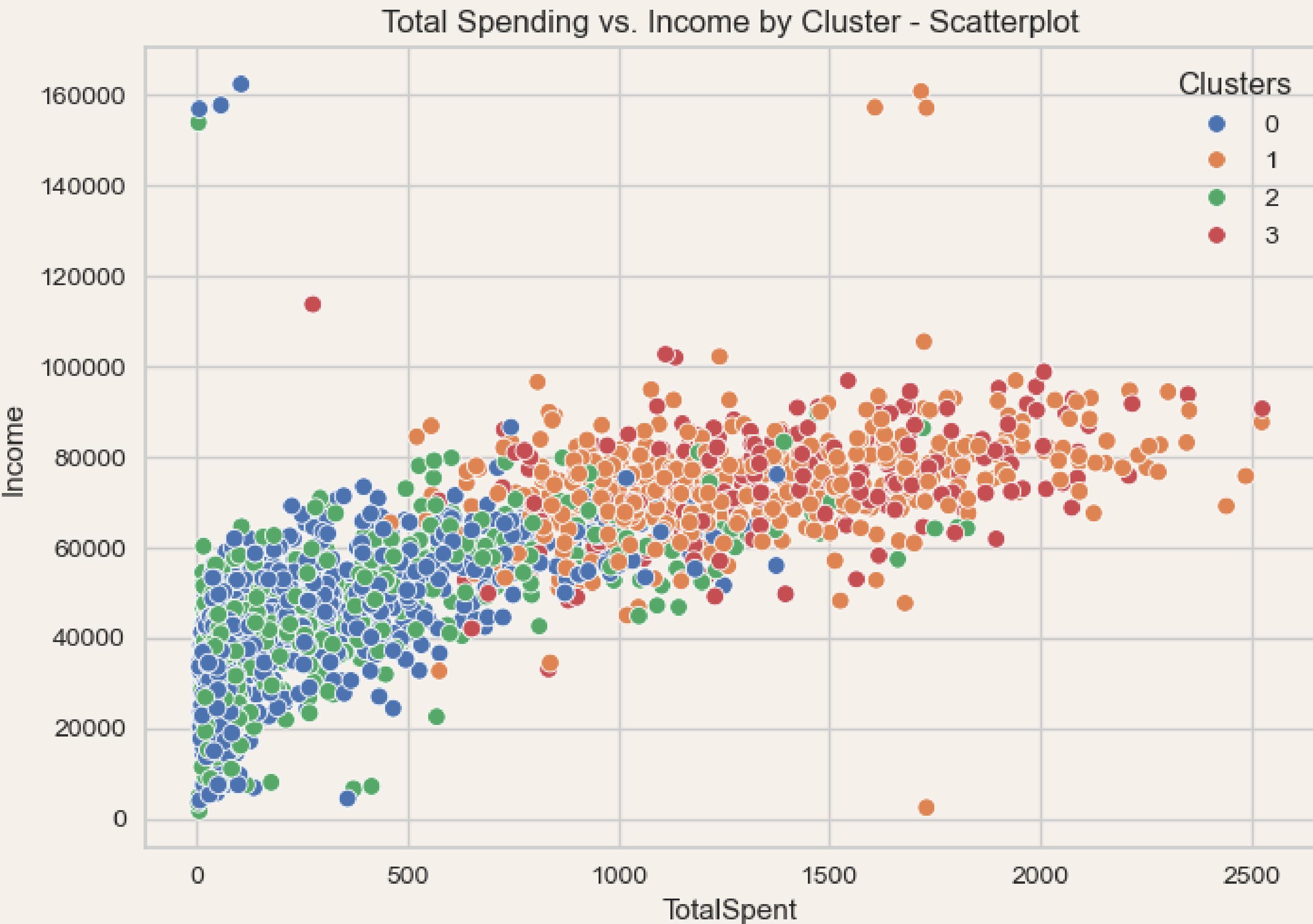


Analysis & Results

04 PRODUCT

Across all clusters, there is a clear positive correlation between total spending and income. As income increases, total spending tends to increase as well, with higher-income individuals generally spending more.

Clusters 0 (**Blue**) and 2 (**Green**) are concentrated in the lower income and spending range, with most individuals earning below 60K and spending under 1K. In contrast, Clusters 1 (**Orange**) and 3 (**Red**) show more variation, with higher earners and spenders, including individuals with incomes over 100K and spending between 1K and 2K, especially in Cluster 1. Cluster 3, though smaller, also includes higher income and spending individuals.



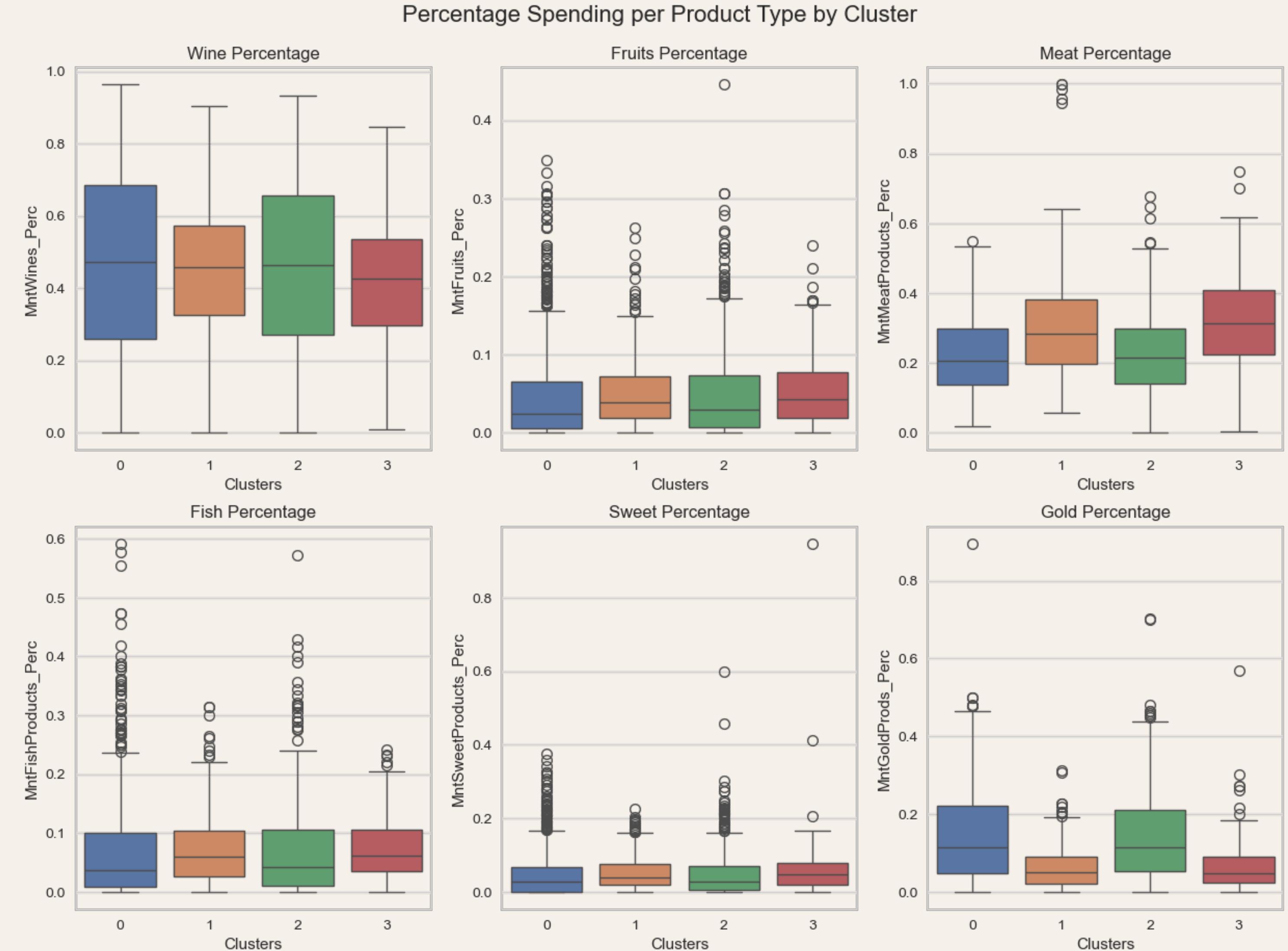
Analysis & Results

04 PRODUCT

Wine and Meat Preferences: All Clusters have similar median wine spending, with Cluster 0 and 2 showing more variability.

On the other hand, Cluster 3 spends the most on meat, suggesting a stronger preference for meat products in this group, while Clusters 0 and 2 place less emphasis on meat.

Consistent Spending: Cluster 1 exhibits the most consistent spending behavior across categories like sweet products and wine, showing less variability in its consumption patterns compared to other clusters.

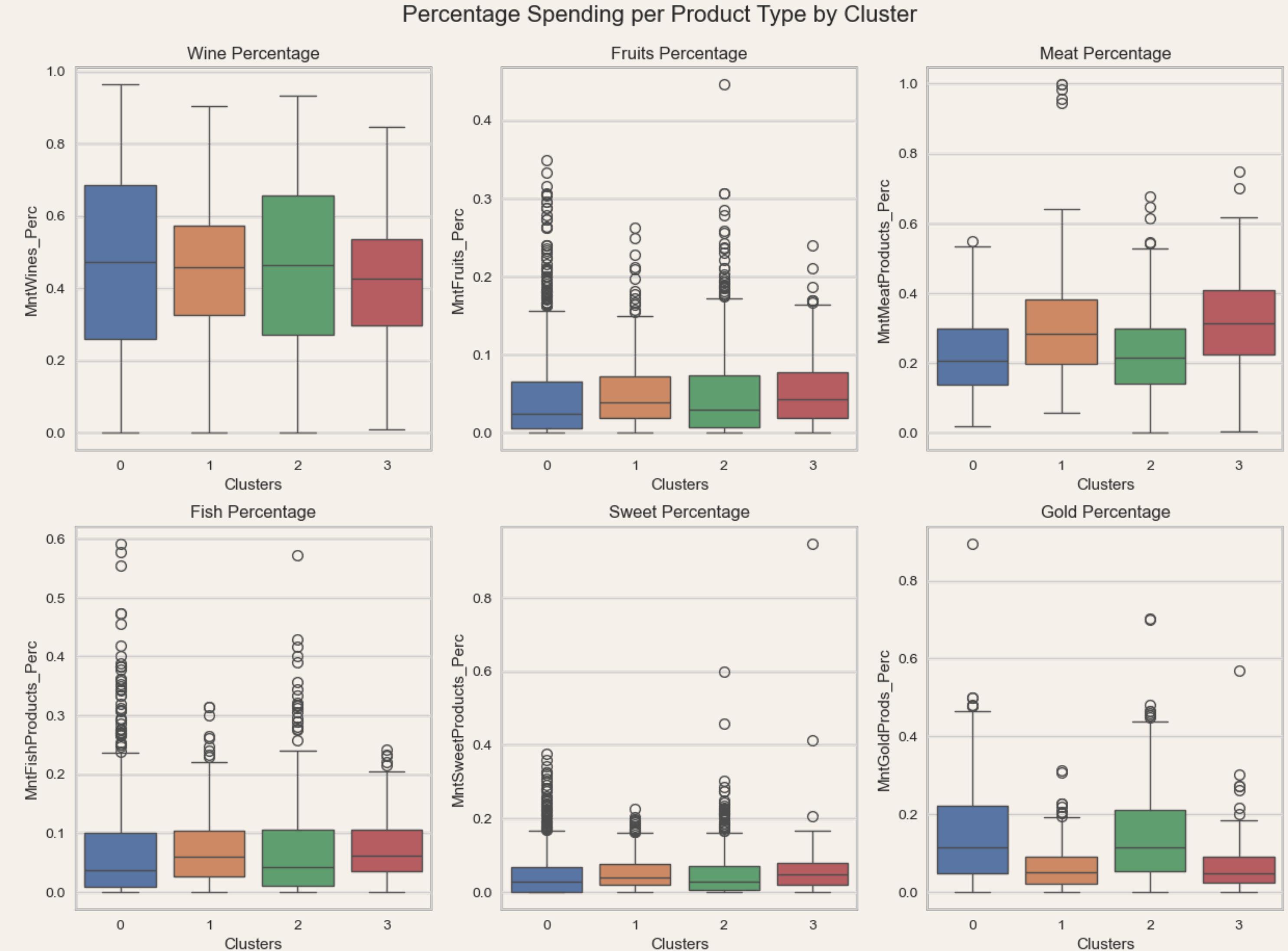


Analysis & Results

04 PRODUCT

Outliers in Spending: Clusters 0 and 2 consistently show outliers across several product categories, including fish, sweets, and gold products, indicating that some individuals in these clusters have spending habits that significantly deviate from the norm.

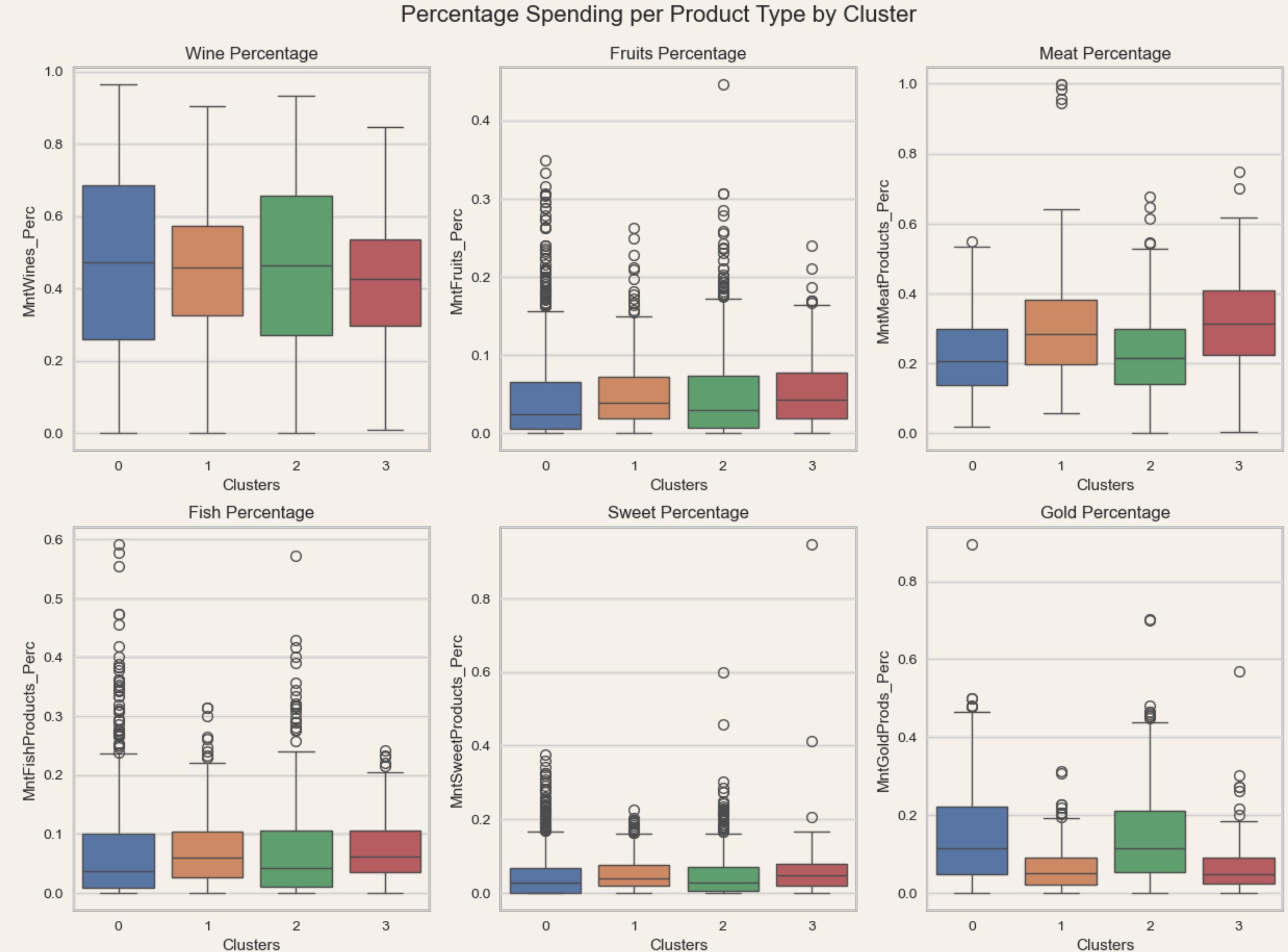
Low Spending Categories: All clusters tend to spend relatively little on fruit and fish products, with minor differences in medians. However, there are notable outliers in Clusters 0 and 1 for fruit, and in Clusters 0 and 2 for fish, reflecting more diverse spending patterns within these groups.



Analysis & Results

04 PRODUCT

Gold Product Preferences: Clusters 0 and 2 demonstrate higher spending on gold products, with outliers indicating significant purchases by certain individuals, whereas Clusters 1 and 3 spend much less on gold, showing more consistent and lower spending across the board.



Analysis & Results

05 PLACE

Clusters 0 and 2 are more concentrated in lower total purchase ranges, indicating that customers in these groups tend to make fewer purchases.

Clusters 1 and 3 dominate in higher purchase ranges, suggesting that these customers are making more frequent purchases overall, with Cluster 1 being the standout in the 15 to 25 purchase range.



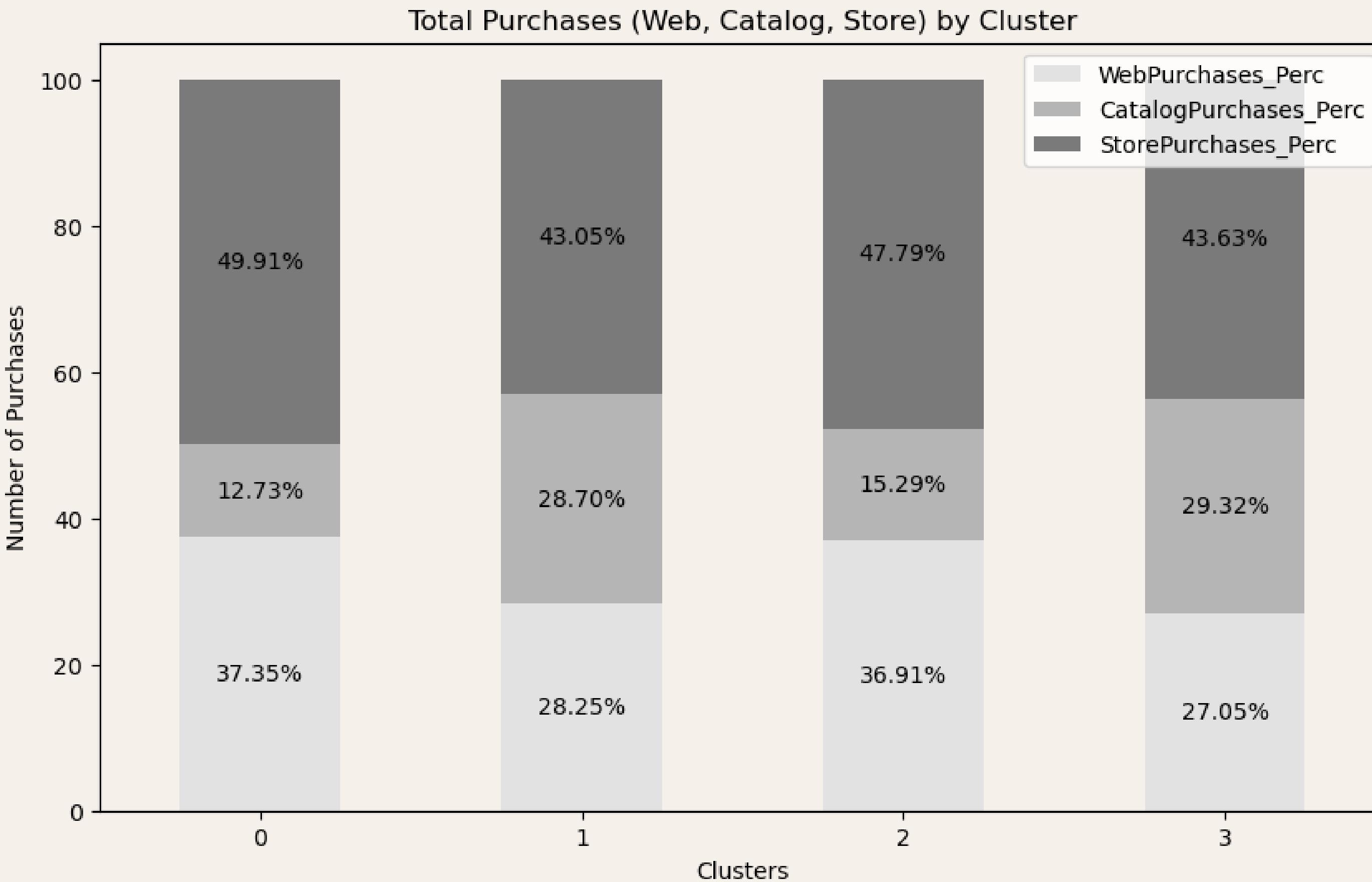
Analysis & Results

05 PLACE

Store purchases are the most dominant channel across all clusters, especially in Cluster 0 and Cluster 2.

Cluster 1 and Cluster 3 show higher engagement with catalog shopping, with Cluster 3 having the highest catalog purchase percentage at 29.32%.

Web purchases are most significant in Cluster 0 and Cluster 2, contributing over 36% to the total purchases in these clusters, but are less important in Cluster 3.

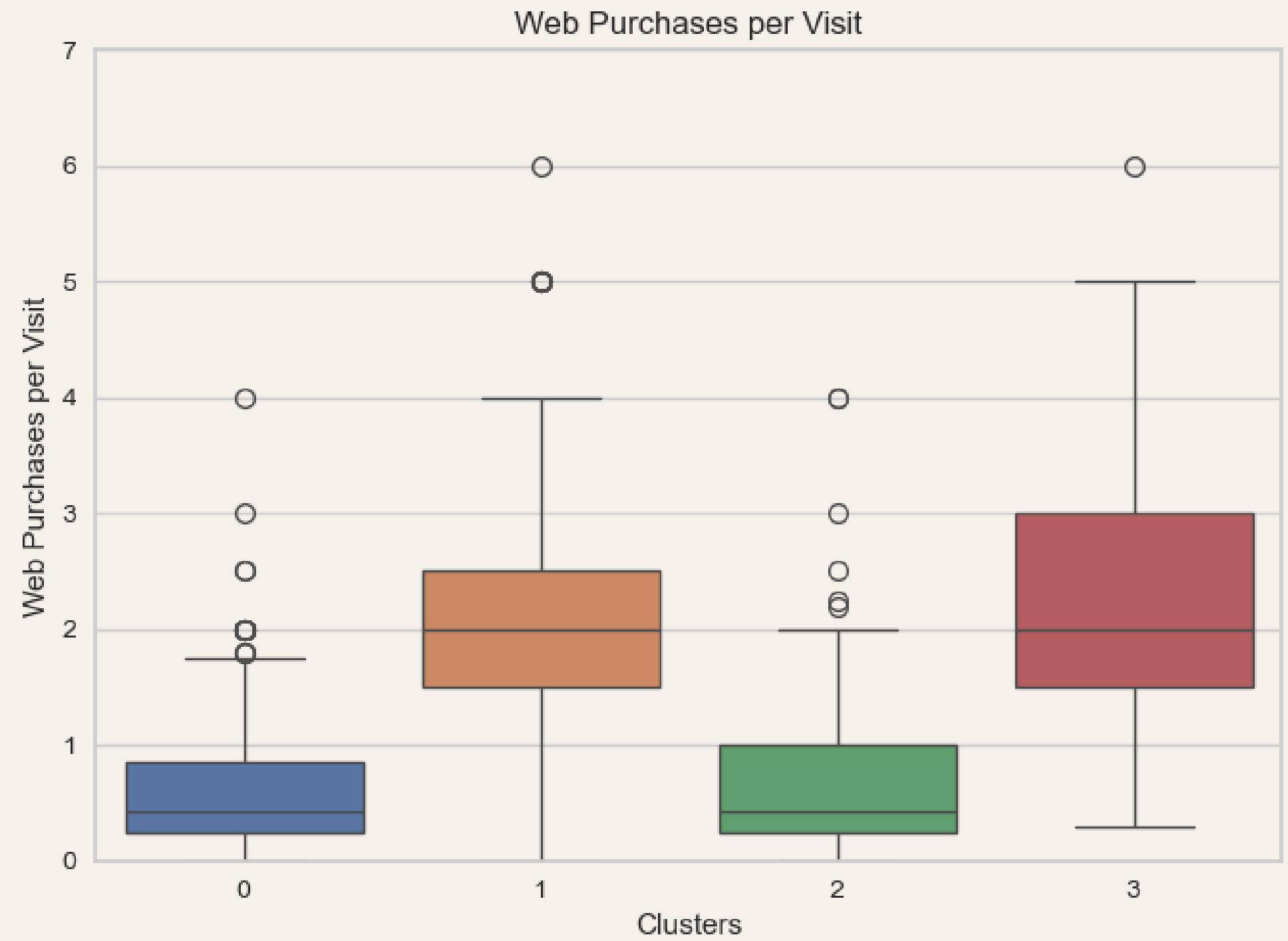


Cluster 3 stands out for its relatively balanced distribution of purchases across all channels, while Cluster 0 is more skewed toward store and web purchases, with catalog purchases playing a minor role.

Analysis & Results

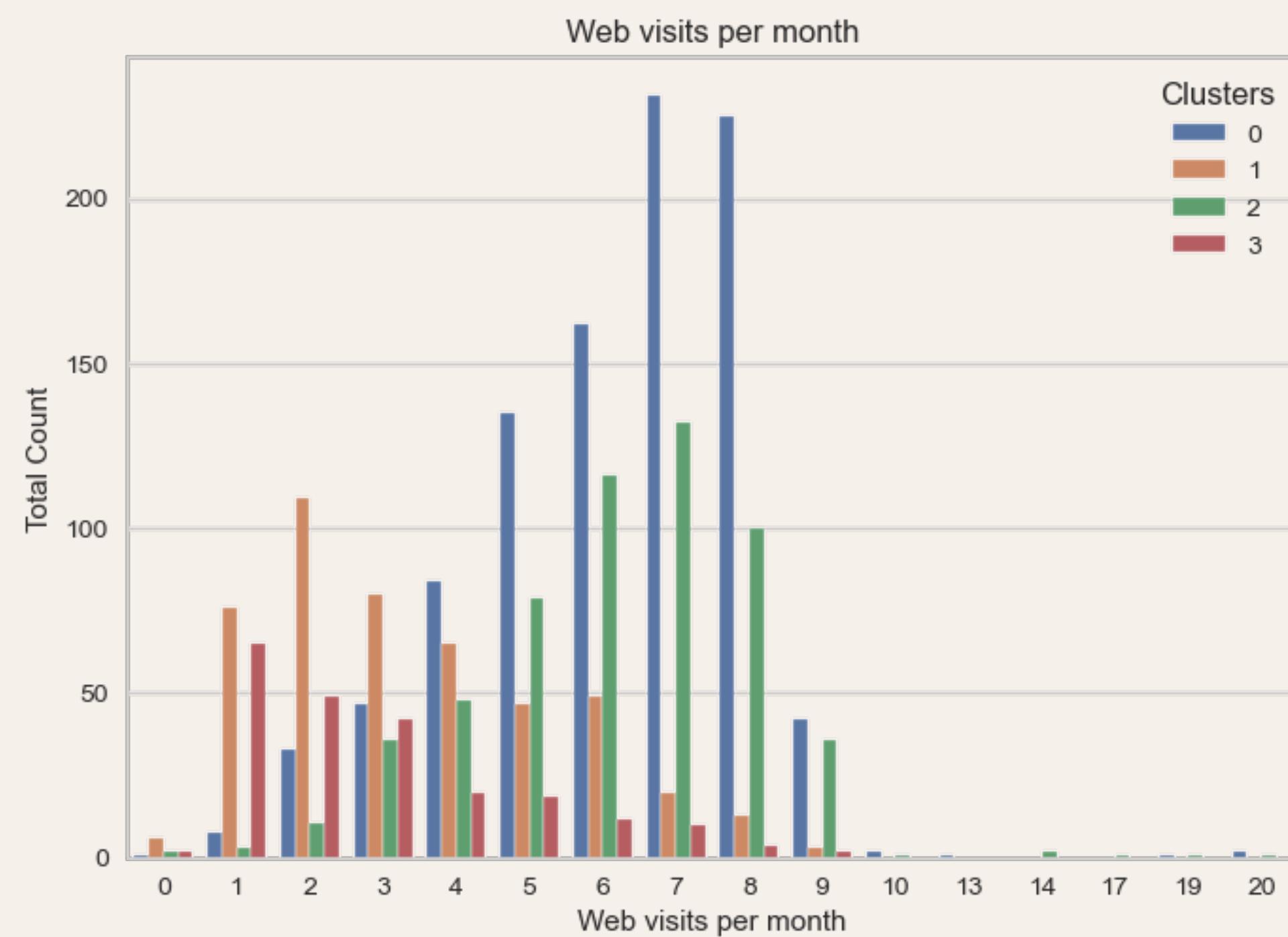
05 PLACE

Cluster 0 represents the least active web shoppers, making fewer purchases per visit. Though, they are the group that makes more visits in the Web.



Cluster 3 is the most active group, frequently making multiple web purchases in a single visit, followed closely by **Cluster 1**. Also this groups are the ones that makes less visits.

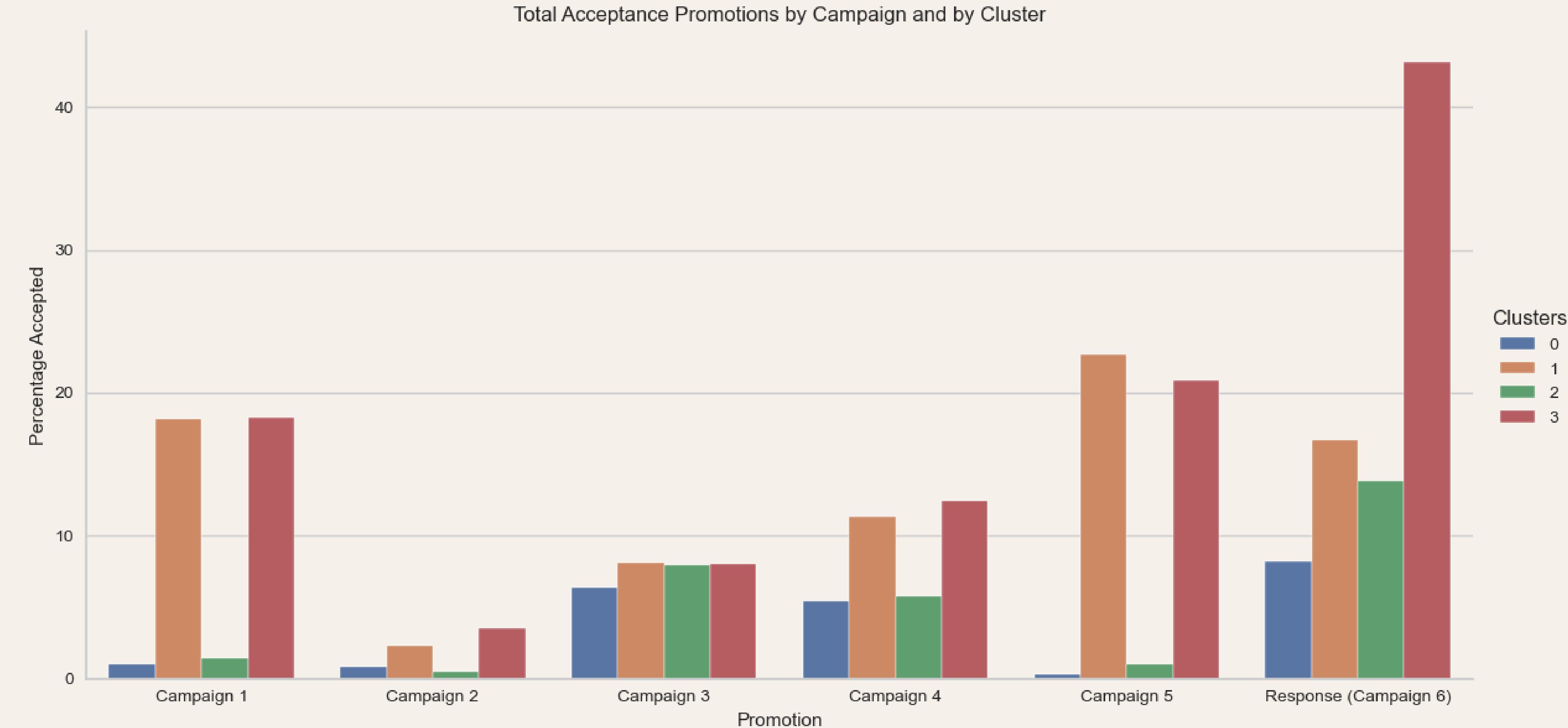
Cluster 2 displays a more mixed behavior, with some variability but generally fewer purchases per visit.



Analysis & Results

06 PROMOTION

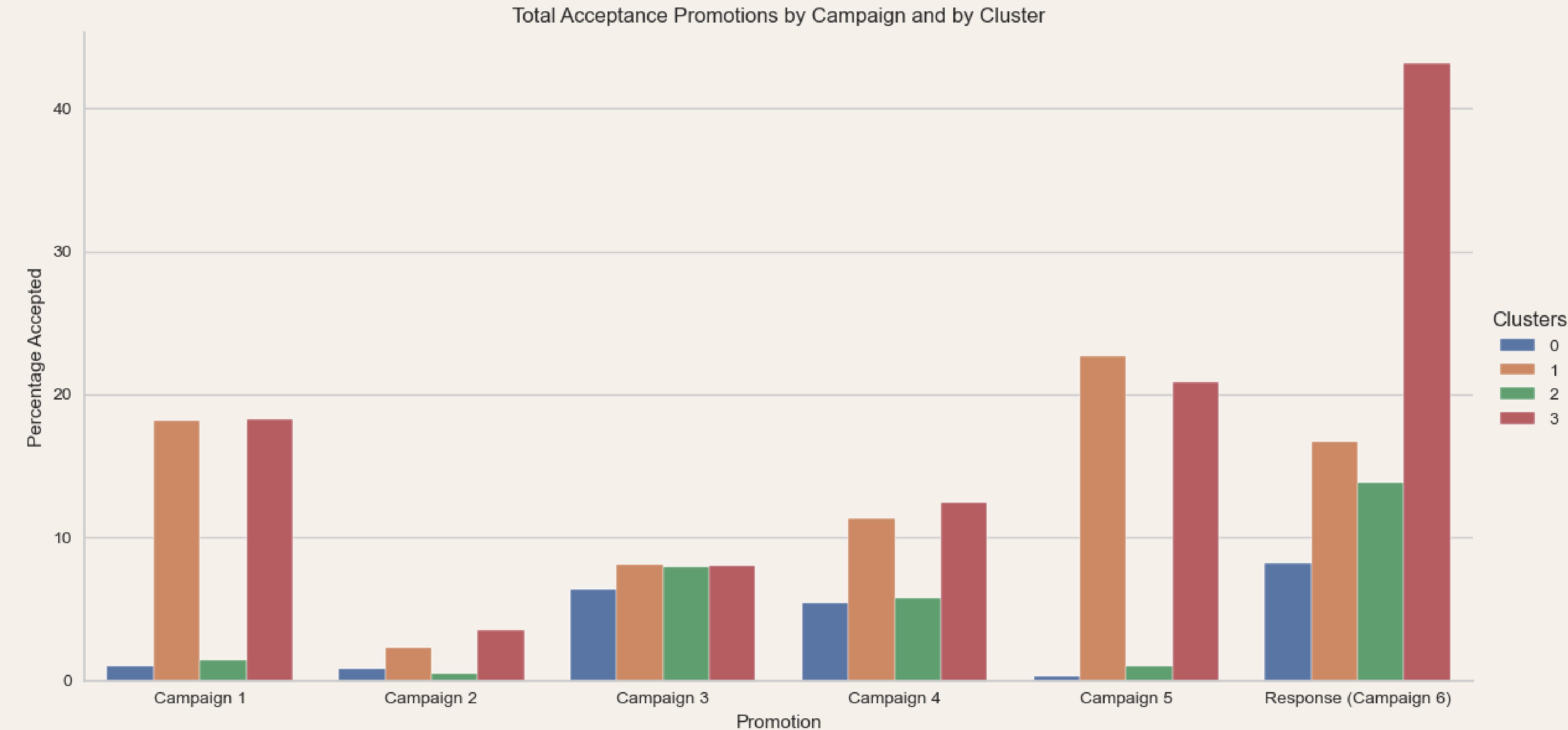
Clusters 1 and 3 are the most engaged with the promotional campaigns, with Cluster 3 consistently showing the highest acceptance rates and overall response.



Analysis & Results

06 PROMOTION

Clusters 0 and 2 display very low engagement across all campaigns, particularly in Cluster 0, which has the lowest promotion acceptance and response.

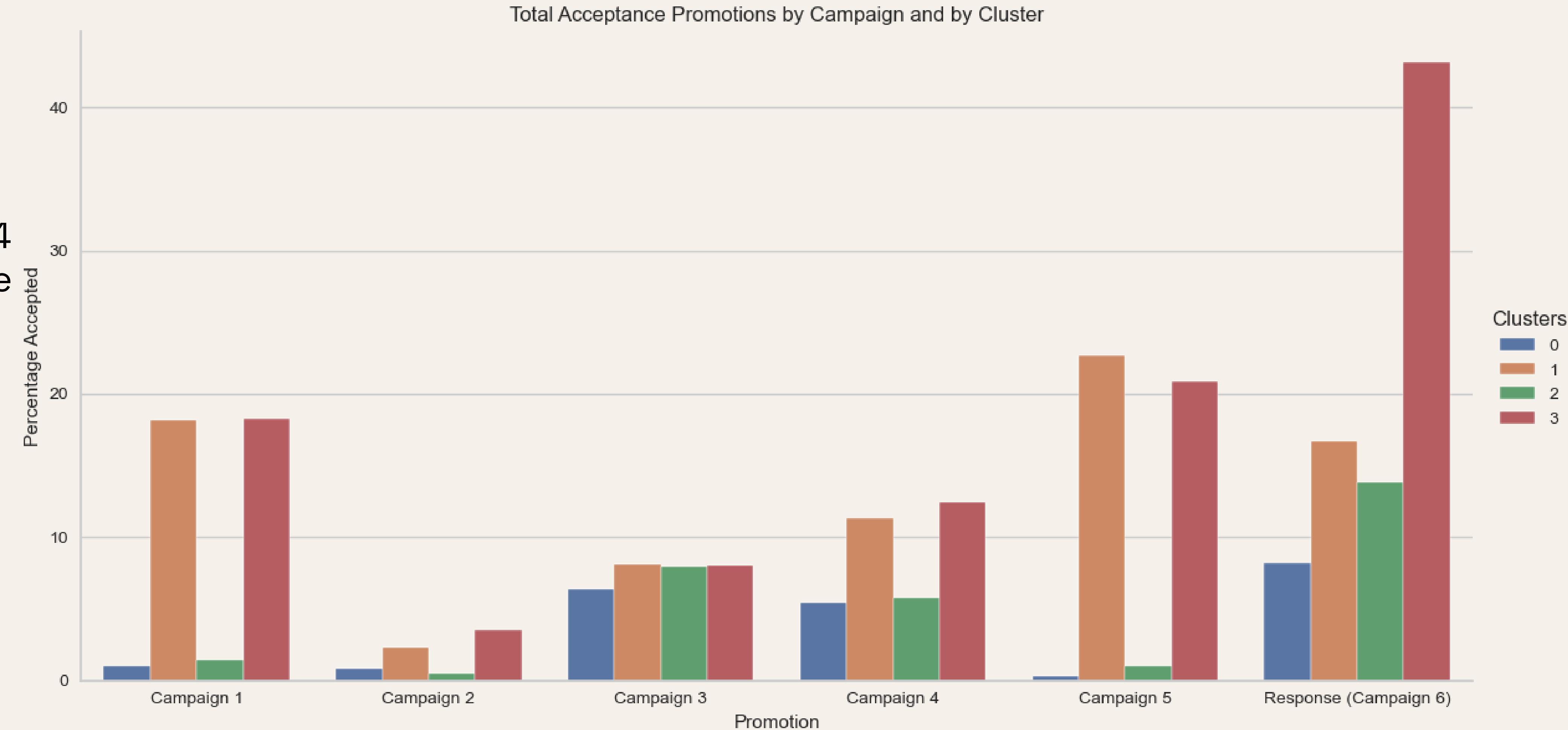


Analysis & Results

06 PROMOTION

Campaign 2 showed poor performance across all clusters.

Campaigns 3 and 4 had similar average performance for all clusters, with moderate acceptance rates.

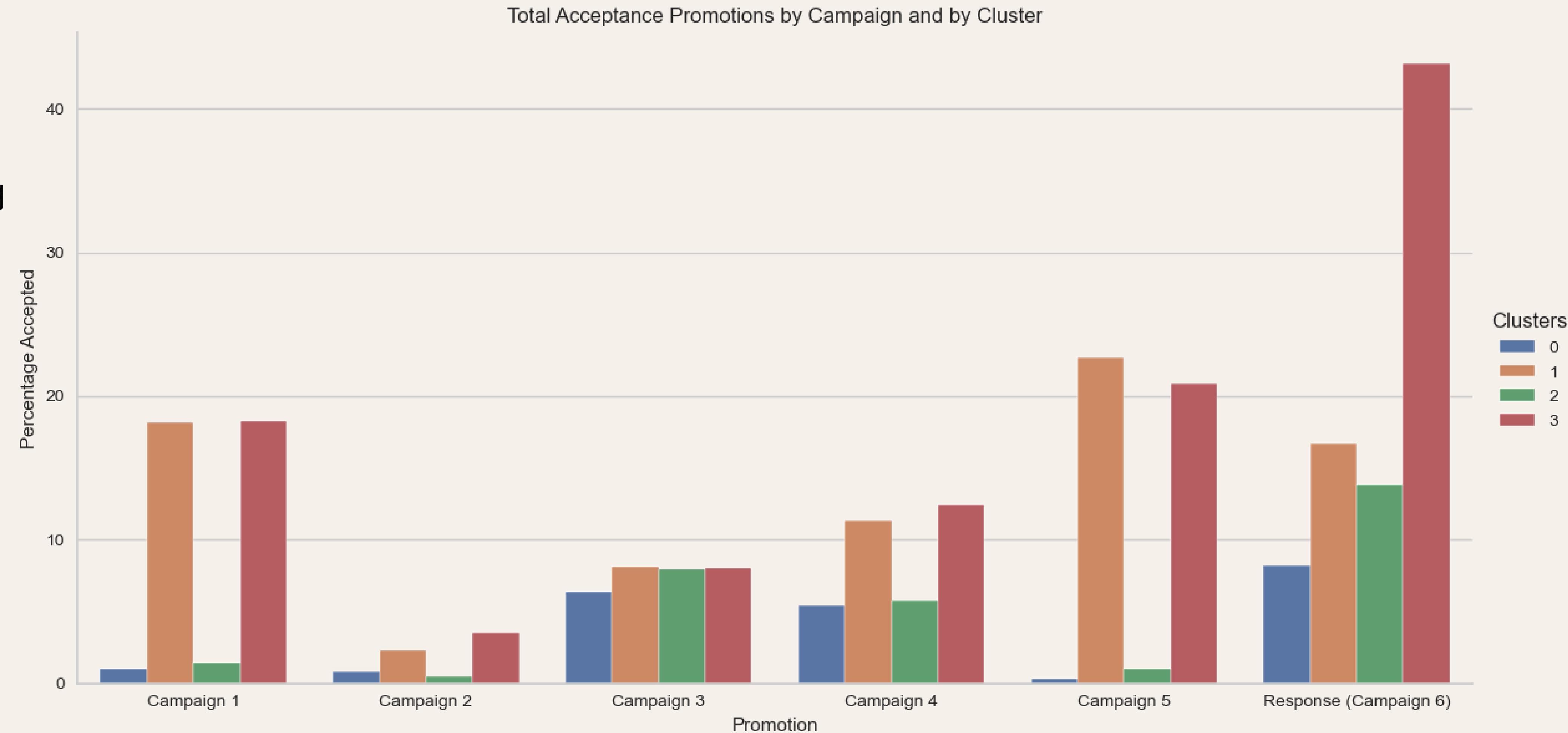


Analysis & Results

06 PROMOTION

Campaign 6 (Response) performed well across all clusters, indicating a strong overall response.

Campaigns 1 and 5 were particularly effective for Clusters 1 and 3, demonstrating high acceptance rates in these groups.



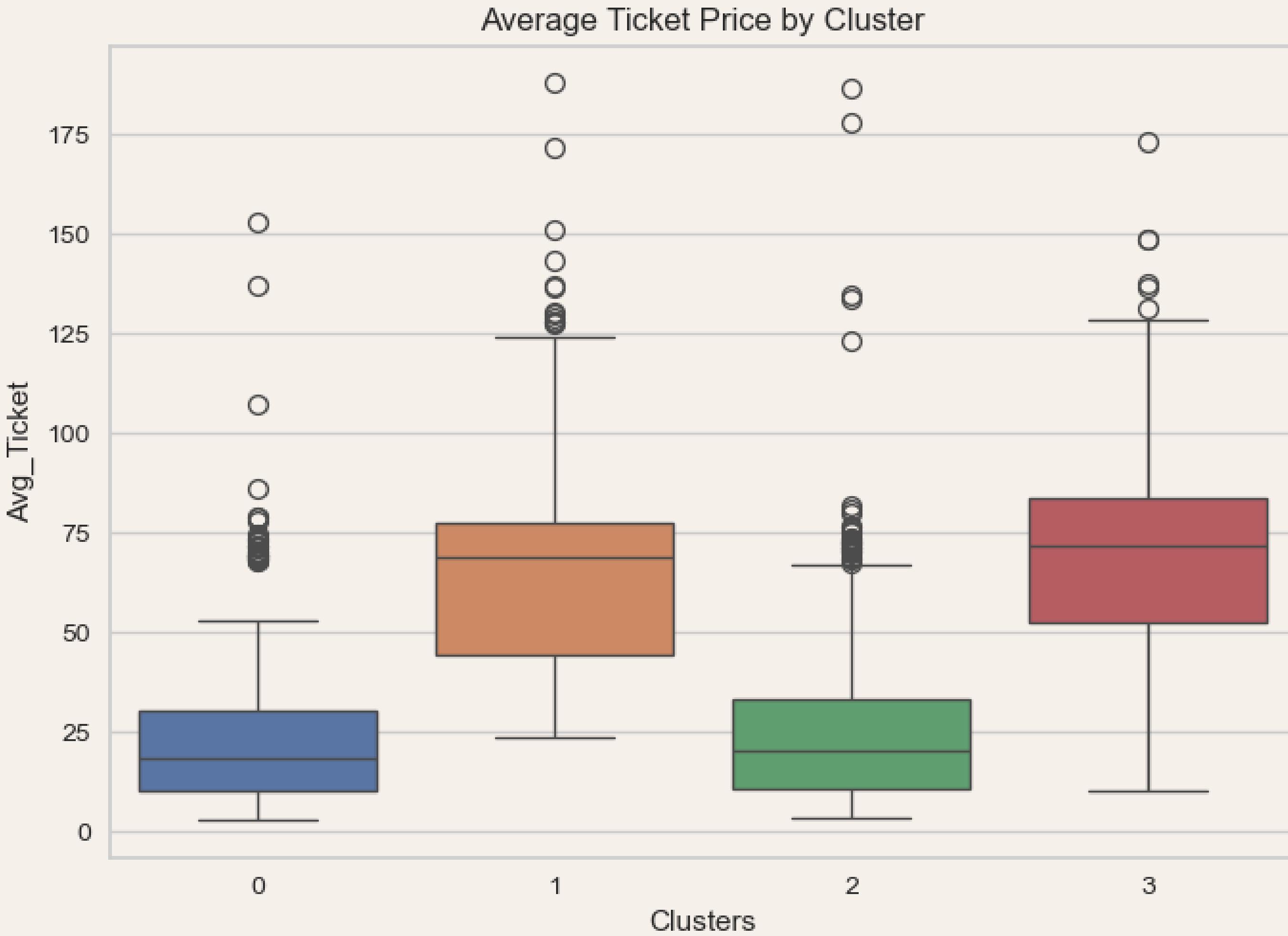
Analysis & Results

07 PRICE

Clusters 1 and 3 have the highest median ticket prices, with Cluster 1 displaying more variability and higher outliers, indicating that some individuals spend significantly more.

Clusters 0 and 2 exhibit lower spending behavior, with Cluster 0 having the lowest average ticket prices overall.

The outliers in Clusters 1 and 3 suggest that these groups contain a subset of customers who are willing to pay higher prices on average.



Avg. Ticket Price (def.): average total spend per purchase. It represents the typical amount of money a customer spends on a single purchase, calculated by dividing the total spending by the number of purchases made.

Targeting & Conclusions

“The clustering analysis, a model to group customers based on their similarities, showed four types of customers to target with the program.”

GROUP #0

PERSONA SUMMARY

01 MAIN
Dom. Gen.: X & Boomers
Education Level: Mostly Graduate & Post Graduate
Income: Lower-middle class

02 FAMILY
Status: With Partner
Kids: Mainly Yes (1-2)
Family Size: 3-5

03 PRODUCT

Low Spending (<1K), Higher Spending in Gold Products, Present Outliers (less consistency)

04 PLACE

Less Purchases (3-7 / month), High Store & Web Purchases, Web: Many Visits but low Purchase/Visit Rate

05 PROMOTION

Low Engage (Lowest), Higher Engage in Campaign Type 3, 4, and 6

06 PRICE

Lower Average Ticket Price (\$15-\$30)

GROUP #1

PERSONA SUMMARY

01 MAIN

Dom. Gen.: not conclusive

Education Level: Mostly Graduate & Post Graduate

Income: Middle Class

02 FAMILY

Status: With Partner

Kids: not conclusive (0-1)

Family Size: 2-3

03 PRODUCT

High Spending ('1K-2K),
Higher Spending in Meat,
Consistent Spending
across all cat.

04 PLACE

Higher Purchases (15-25 / month), High Store & Catalog, Web: Fewer Visits but high Purchase/Visit Rate

05 PROMOTION

Highly Engaged, Higher Engage in Campaign Type 1, 5, and 6

06 PRICE

Higher Average Ticket Price (\$50-\$90) & Preset Outliers (willingness to pay higher prices)

GROUP #2

PERSONA SUMMARY

01 MAIN

Dom. Gen.: X & Boomers

Education Level: Mostly Graduate & Post Graduate

Income: Lower-middle class

02 FAMILY

Status: Single

Kids: Mainly Yes (1-2)

Family Size: 2-3

03 PRODUCT

Low Spending (<1K), Higher Spending in Gold Products, Present Outliers (less consistency)

04 PLACE

Less Purchases (3-7 / month), High Store & Web Purchases, Web: Many Visits but low Purchase/Visit Rate

05 PROMOTION

Low Engage, Higher Engage in Campaign Type 3, 4, and 6

06 PRICE

Lower Average Ticket Price (\$15-\$30)

GROUP #3

PERSONA SUMMARY

01 MAIN

Dom. Gen.: not conclusive

Education Level: Mostly Graduate & Post Graduate

Income: Middle & Upper Middle Class

02 FAMILY

Status: Single

Kids: Mainly No (0)

Family Size: 1

03 PRODUCT

High Spending ('1K-2K),
Higher Spending in Meat

04 PLACE

Higher Purchases ('15-25 / month), Balance
Distribution of Place, Web:
Fewer Visits but high Purchase/Visit Rate

05 PROMOTION

Highly Engaged (Highest),
Higher Engage in Campaign Type 1, 5, and 6

06 PRICE

Higher Average Ticket Price (\$50-\$90) & Preset Outliers (williness to pay higher prices)

“Targeting allows us to tailor each campaign to the unique needs of our audience, maximizing engagement and conversions while ensuring every marketing dollar is spent where it delivers the most value.”

GROUP #0

NEXT STEPS

TARGETTING & MESSAGE

Focus communication towards Gen X and Boomers, emphasizing low-cost products and gold as differentiators. The messaging should highlight affordable quality and family benefits, using a familiar tone.

CHANNELS

Since they prefer physical stores and websites, prioritize an omnichannel strategy that strengthens both online and offline experiences. Improve the website's UX to increase conversion rates.

PROMOTION STRATEGY

Design campaigns that align with types 3, 4, and 6 (those they find appealing), with a focus on offers and discounts for gold products and low-cost items. Boost remarketing campaigns to capture unconverted website visits.

PRICE STRATEGY

Focus on lower average ticket prices, using messaging around saving and family value. Encourage more frequent purchases and increase the average ticket with cross-promotional offers.

GROUP #1

NEXT STEPS

TARGETTING & MESSAGE

Since this group has higher purchasing power but is still middle-class, focus on high-quality products, particularly meat and consistent spending across categories. Highlight the exclusivity and long-term value of their purchases.

CHANNELS

Drive sales both in-store and through catalogs, ensuring visually appealing catalogs. Improve the web experience with personalized recommendations to maximize conversion rates.

PROMOTION STRATEGY

Campaigns of types 1, 5, and 6 with a focus on personalized content, loyalty rewards, and exclusive promotions, as they are highly engaged.

PRICE STRATEGY

Leverage their willingness to pay more by offering premium options and bundles that increase the average ticket price (\$50-\$90), highlighting quality and added benefits.

GROUP #2

NEXT STEPS

TARGETTING & MESSAGE

While similar to Group #0, this group is more likely to shop for themselves as singles with kids. Messaging should emphasize convenience and practicality when buying gold and low-cost products.

CHANNELS

Use similar tactics as Group #0, but with additional focus on retargeting campaigns to convert their frequent website visits into actual purchases.

PROMOTION STRATEGY

Campaigns of types 3, 4, and 6, emphasizing limited-time offers and cross-promotions to build loyalty. Position gold products as small family gifts or investments.

PRICE STRATEGY

Encourage lower ticket purchases but more frequently through promotions that reinforce value in each transaction.

GROUP #3

NEXT STEPS

TARGETTING & MESSAGE

This group has higher purchasing power and no kids. Messaging should focus on quality of life, personal enjoyment, and exclusivity. Meat products should be highlighted as part of a more sophisticated lifestyle.

CHANNELS

Balance in-store and online sales. Use personalized content and a smooth experience across both channels to maintain their high web conversion rate.

PROMOTION STRATEGY

High engagement in campaigns types 1, 5, and 6, focusing on rewarding their loyalty with early access to promotions, VIP programs, or exclusive products.

PRICE STRATEGY

Since they're willing to pay higher prices, increase the average ticket by promoting premium, high-end products, focusing on the shopping experience and quality.

Key Aspects for All Groups

RETARGETING CAMPAIGNS

Particularly important for groups with high web visits but low conversions.

WEBSITE OPTIMIZATION

Improving the online experience is crucial to boost conversion rates, especially for Groups 0 and 2.

LOYALTY PROGRAMS

Implement programs to retain highly engaged consumers like Groups 1 and 3.

PERSONALIZED CONTENT

Tailoring content and promotions based on each group's preferences will help increase engagement and sales.

GitHub Project Link

Open Link

DECEMBER 2021 - PROJ. 01
FACUNDO CASTELLÁ

Marketing Campaign Analysis