

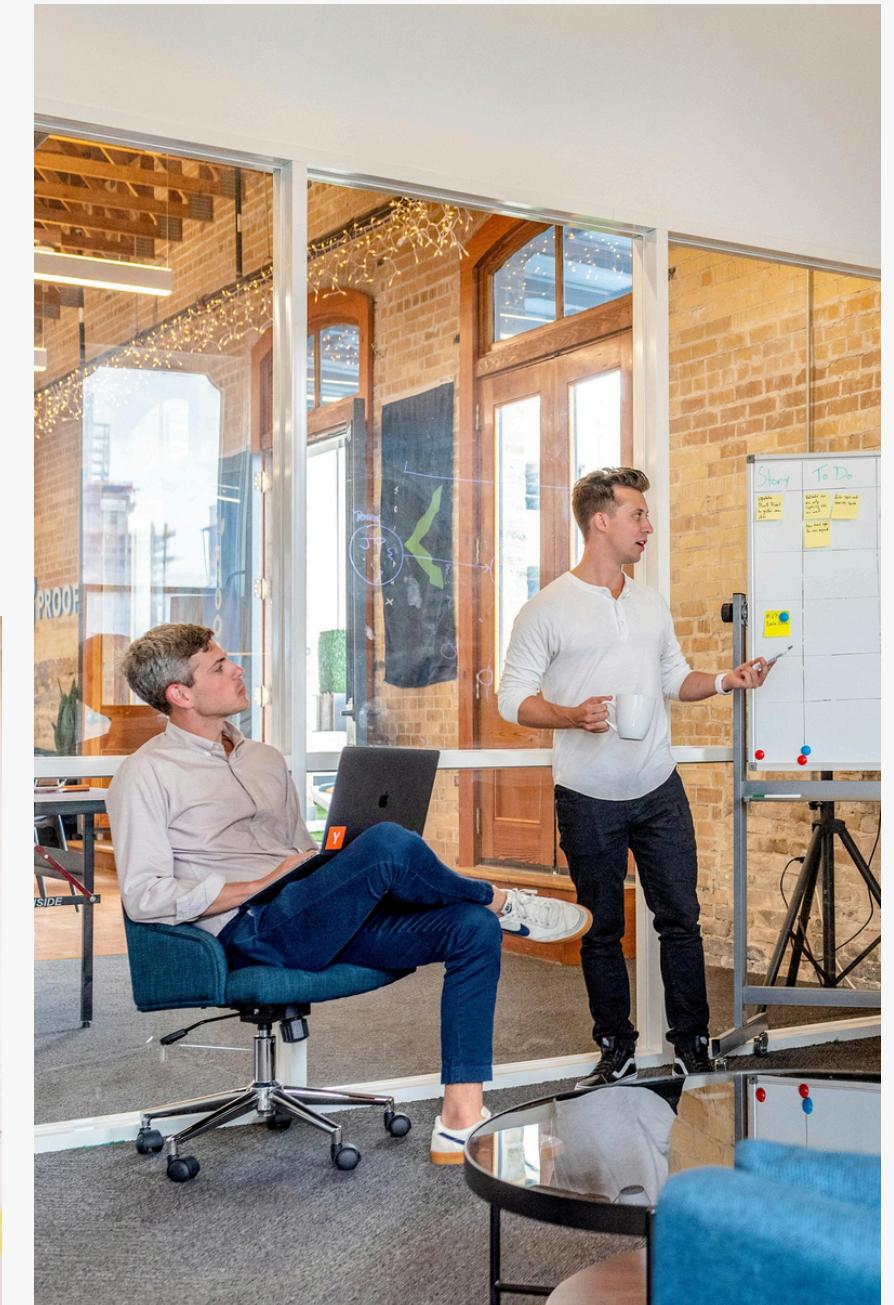
# Predict Customer Personality To Boost Marketing Campaign

Az-Zukhrufu Fi Silmi Suwondo

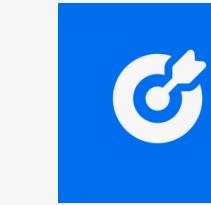


# Problem Statement

The company is experiencing **inefficiencies** in its **marketing strategy** due to the lack of detailed customer segmentation. A generalized, **one-size-fits-all approach** has led to suboptimal resource allocation, resulting in poorly targeted campaigns with low conversion rates and underwhelming ROI. Moreover, the organization **struggles to identify high-value customer segments** and **address their specific needs effectively**.



# Goal & Objectives



## Goal

**Increase customer engagement** and **revenue** through personalized marketing strategies driven by data insights.

### # Objective 1

- ★ Identify key customer segments based on shopping behavior and demographics

### # Objective 2

- ★ Understand product and channel preferences for each segment

### # Objective 3

- ★ Identify key customer segments based on shopping behavior and demographics

# Customer Demography



*Predominantly between 35–50 years old*



*Majority earn IDR 50–60 million per month*



*Nearly half belong to Gen X (born 1965–1980)*



*Typically 2 family members*



*Over half hold a university degree*



*Only 0.01% of income spent on shopping*



Customers are mostly **35–50 years old**, **Gen X**, earning **IDR 50–60 million** monthly, with **small households** and **higher education**. Despite strong purchasing power, shopping expenses remain minimal.

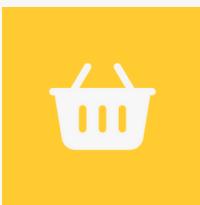
# Passive vs Active Customers

## Passive Customer



### Lifestyle & Family

- Mostly have children
- Larger family size



### Shopping Behavior

- Low total spending
- Long time since last transaction
- Rarely participates in campaigns



### Purchasing Power

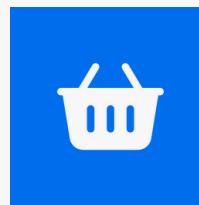
- Typically low to middle income

## Active & Responsive Customer



### Lifestyle & Family

- Mostly do not have children
- Smaller family size



### Shopping Behavior

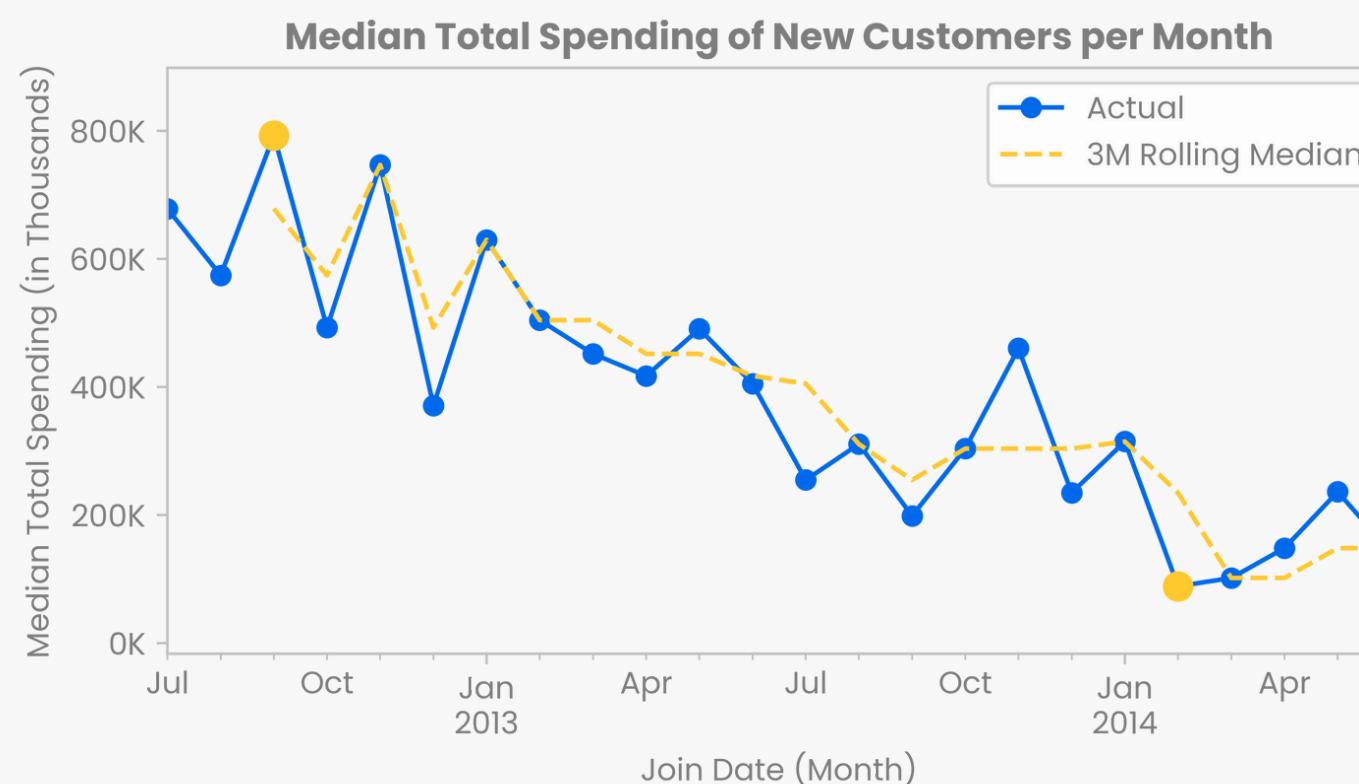
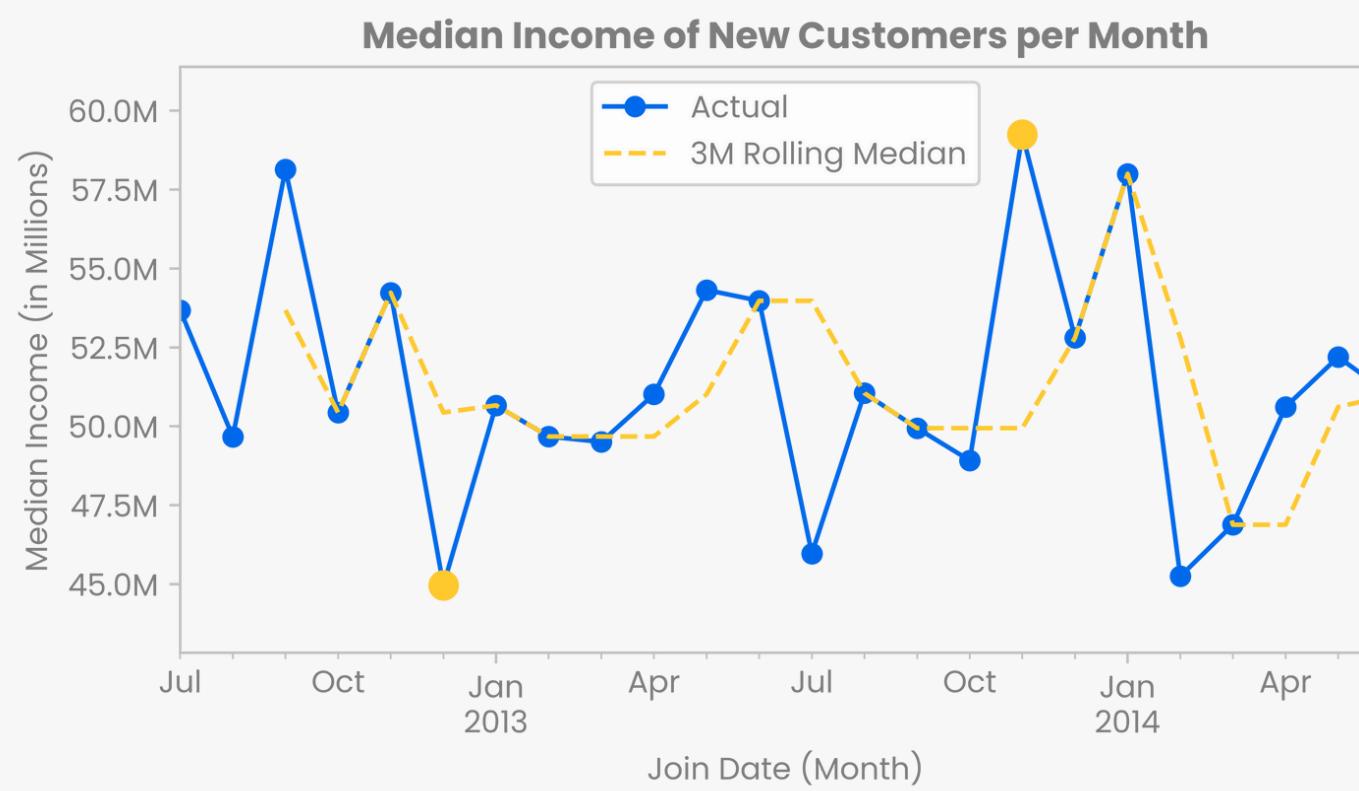
- Big spender
- Recently made a purchase
- Frequently joins campaigns



### Purchasing Power

- Typically middle to high income

# Stable Income, But Why Is Spending Down?



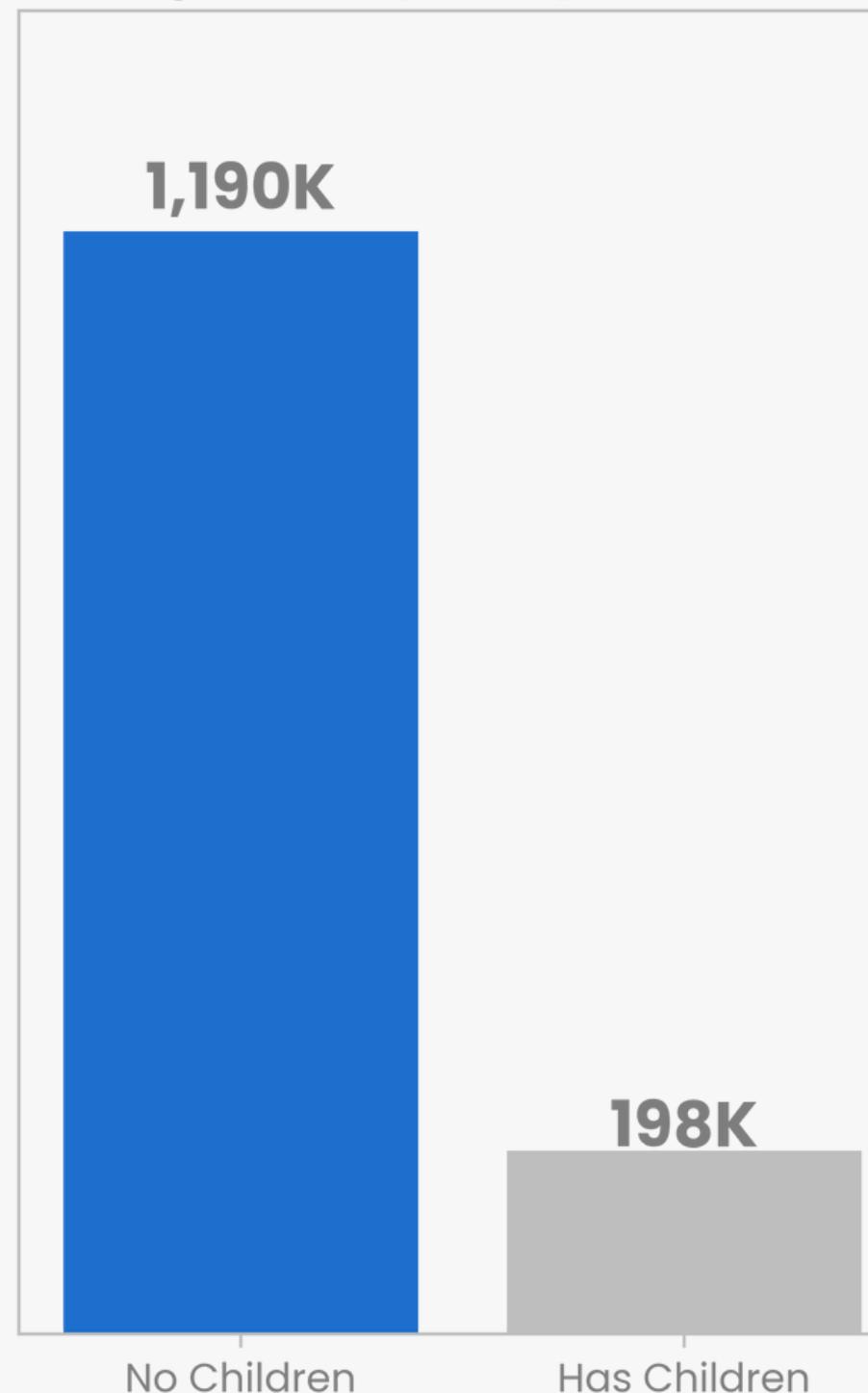
**Median spending of new customers dropped significantly** from around 800K to 100K between September 2013 and January 2014, while **median income remained stable** in the range of 45–58 million. Outliers and inconsistent seasonal patterns suggest potential data quality issues or external factors that require further investigation.



## Recommendation

Check for any campaign or event that may have influenced the data, and ensure monthly sample sizes are sufficient to avoid biased analysis.

Average Total Spent by Has Children

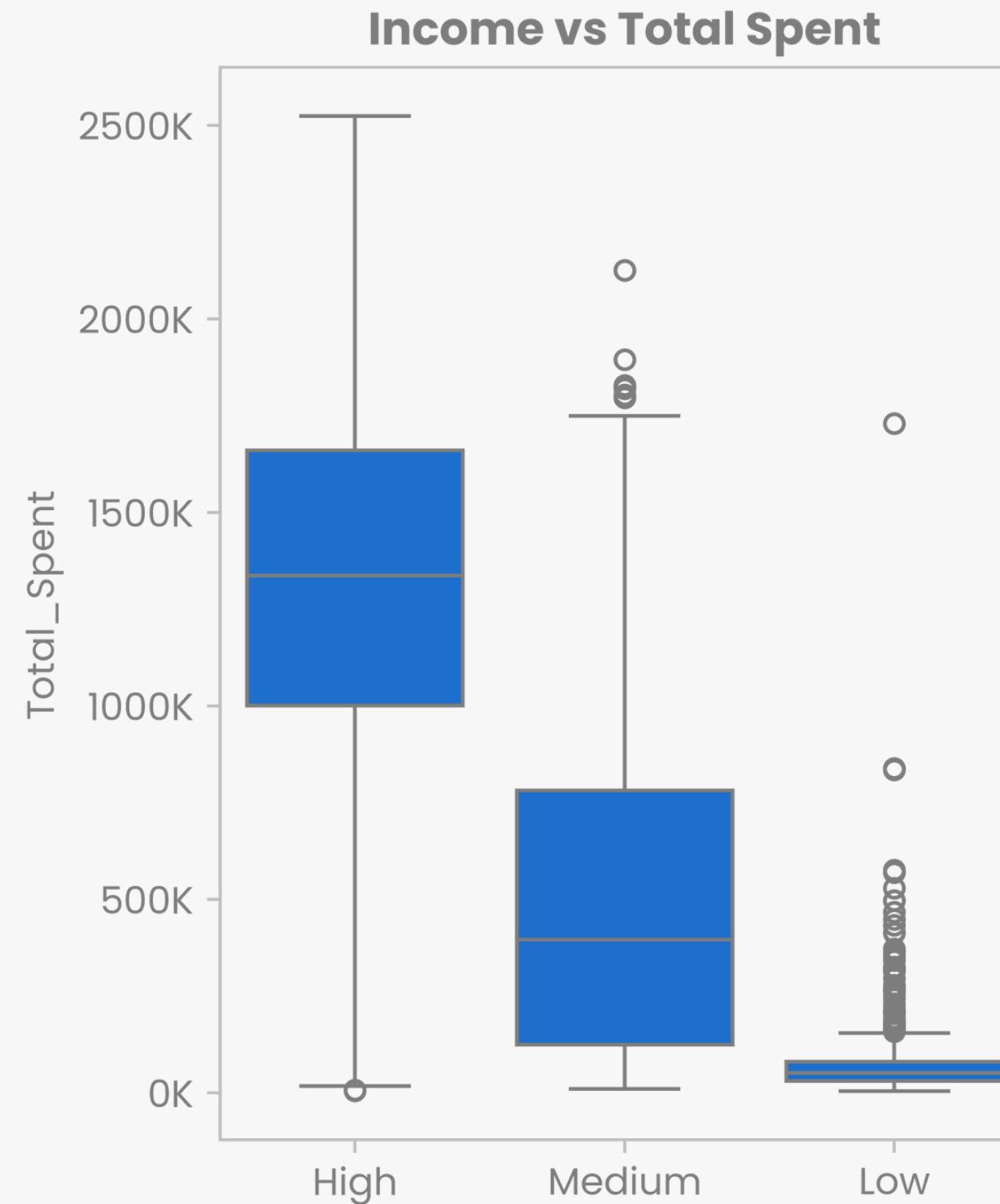


# Child-Free Customers Spend 6x More

Annual spending analysis shows two clear customer types. **Childless customers spend a steady median** of \$1.2M yearly, while **parents vary widely**, with a \$200K median and a range from \$5K to \$2.2M. The 6x gap highlights fundamentally different buying behaviors.

## 💡 Recommendation

Marketing budgets should reflect segment value: spend more to acquire childless customers due to consistent high spending; prioritize scale for parents, given broader but variable patterns.



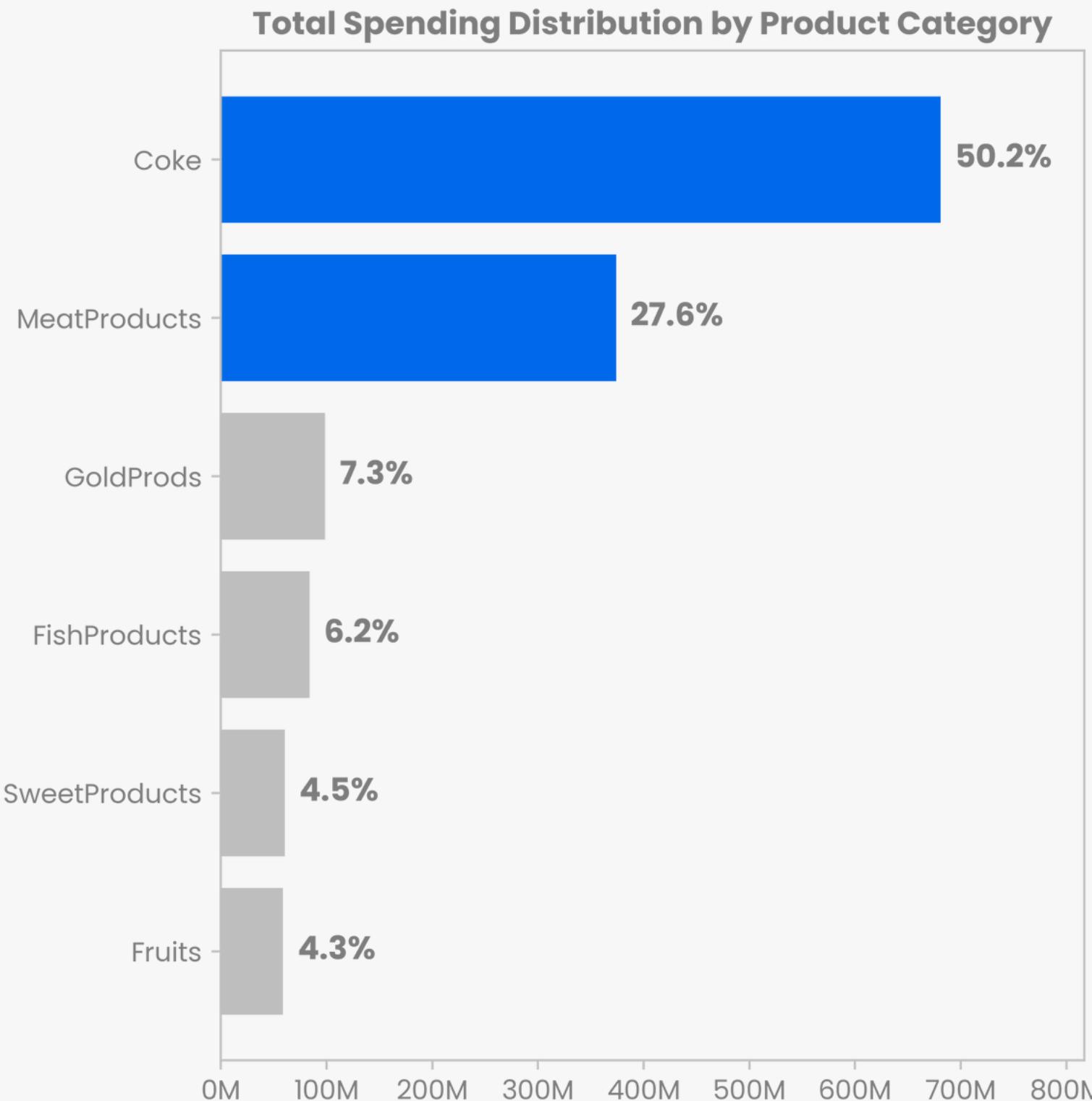
# Different Income, Different Spending Habits

**High-income** customers are **stable premium spenders** (median Rp 1.34M). **Medium-income** customers are diverse, with **power spenders lifting the average** (median Rp 396K). **Low-income** customers are **mostly frugal** (median Rp 52K) but have the most big-spending outliers.

## 💡 Recommendation

Tailor offers by segment: premium perks for high-income, tiered promos for medium, and value deals plus installments for low-income outliers.

# Coke & Meat Drive 78% of Sales



Customer spending is **highly concentrated**: **Coke** (50.2%) and **MeatProducts** (27.6%) together drive almost **78% of total sales**, a clear example of the 80/20 rule. Spending behavior is also skewed, average spend is much higher than the median, indicating a small group of heavy spenders while the majority spend at more moderate levels.

## 💡 Recommendation

Focus on star categories through strong promotion and availability, while boosting regular customer spend with cross-selling and bundled offers like Coke, Meat, and Sweets.

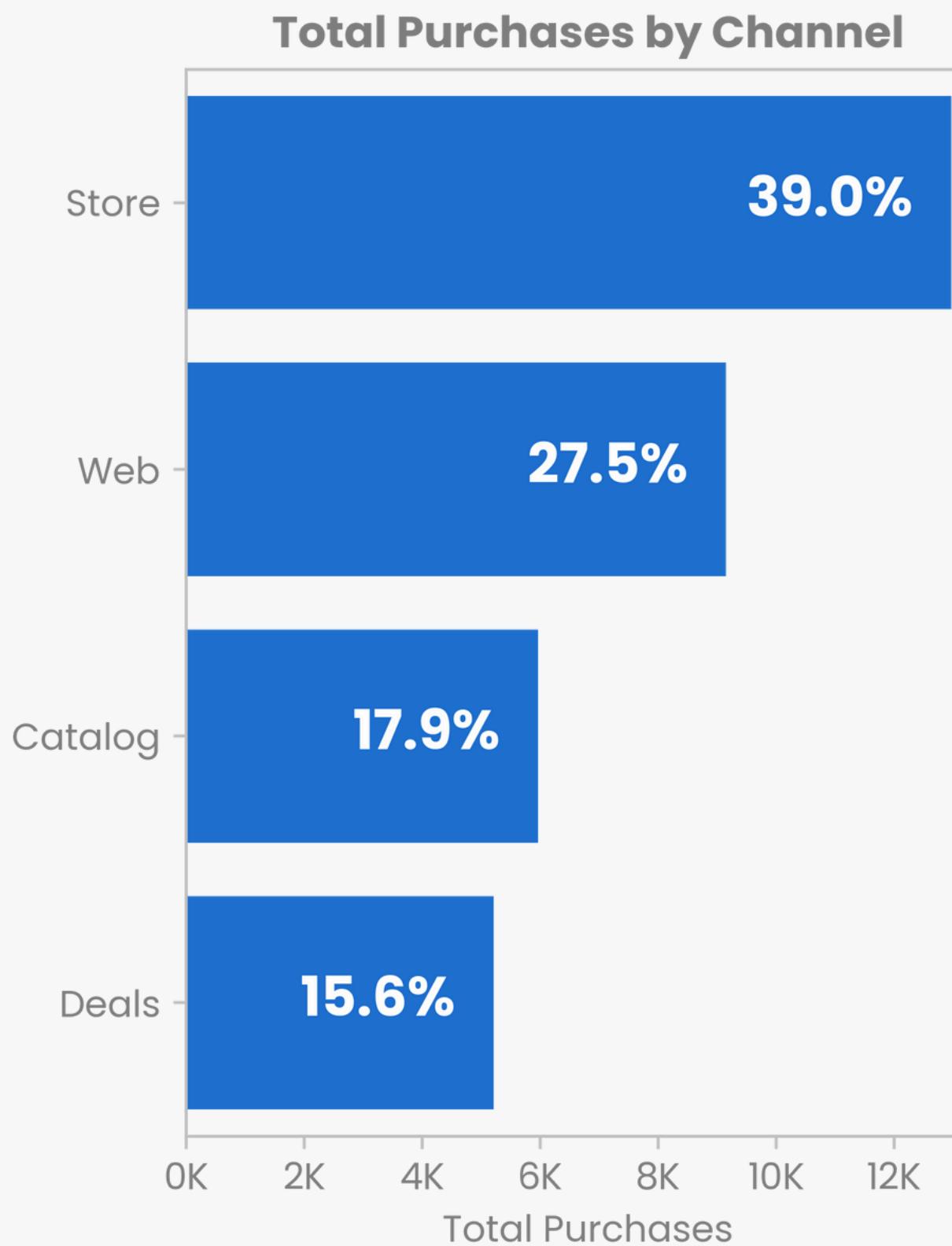


# Cross-Buying Strategy

There are **7 product category pairs** showing **strong correlations** above 0.5, indicating consistent cross-buying patterns across certain categories.

## 💡 Recommendation

Implement a cross-merchandising strategy by grouping the four core categories in a single area, while positioning Coke as a complementary add-on across each category.



# Offline Stores Rank #1, Websites #2

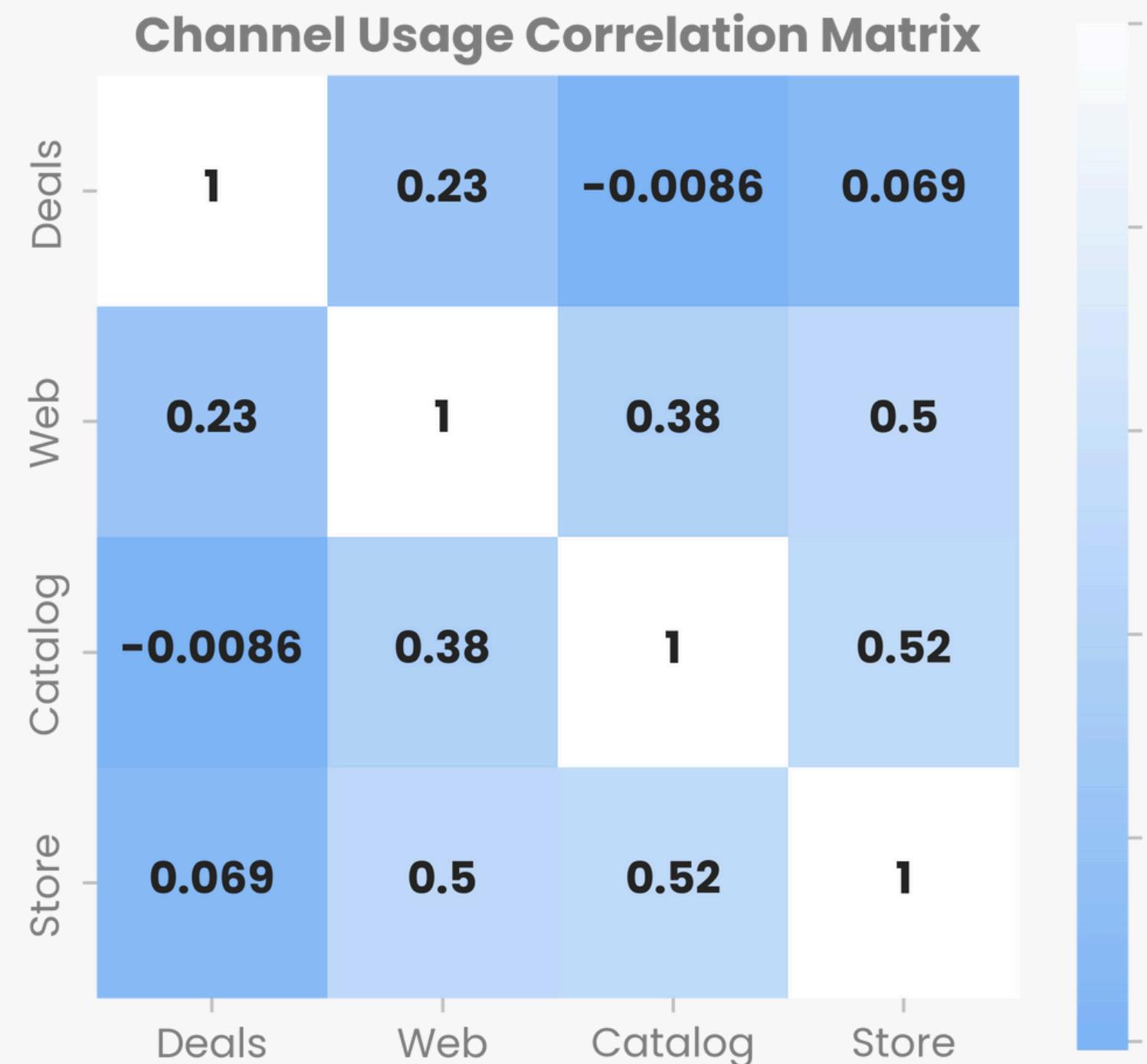
**Stores** remain the **primary choice** for customers (39.0%), **followed by websites**, which reflect the ongoing digitalization of the customer journey (27.5%). Catalogs and deals, however, play a less dominant role in total customer purchases.



## Recommendation

Maximize the potential of the 39% market share by enhancing the in-store customer experience, while simultaneously optimizing omnichannel integration to ensure a seamless and consistent journey across all touchpoints.

# Cross-Channel Behavior

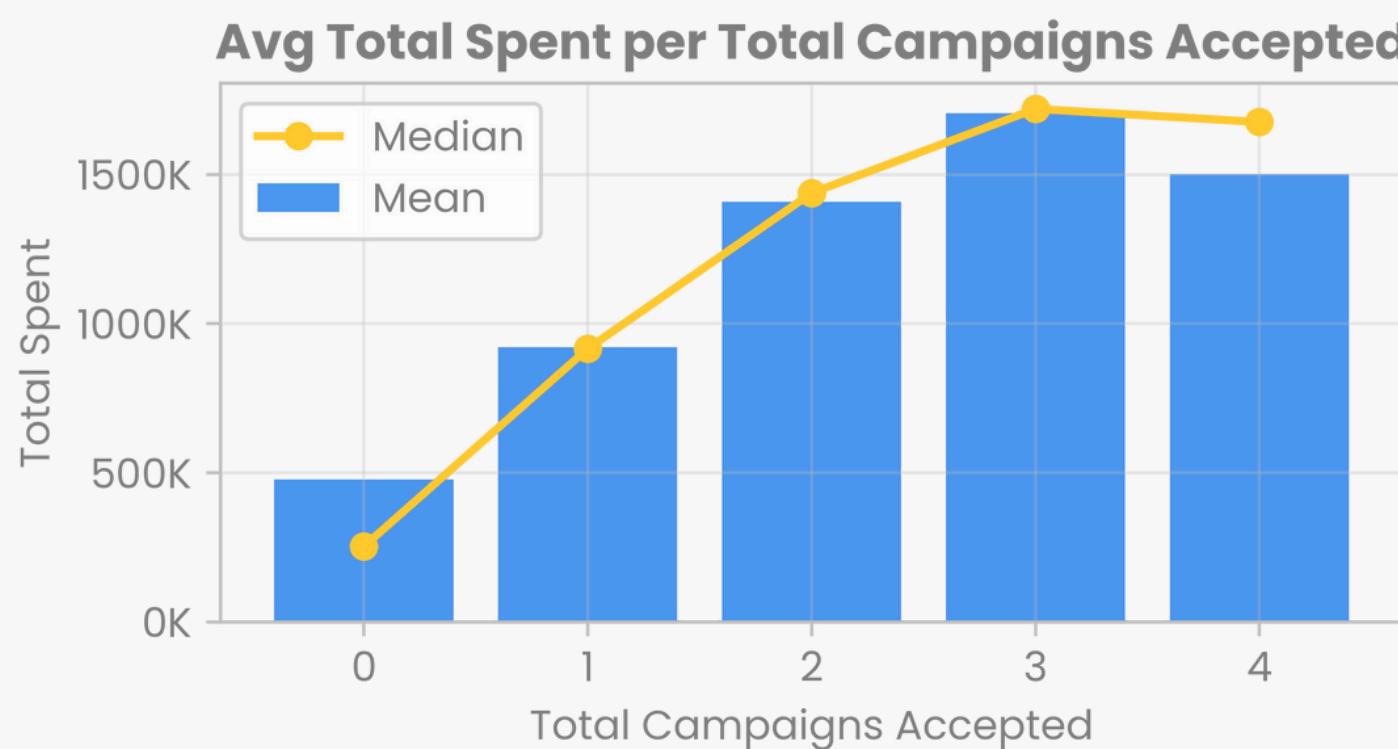
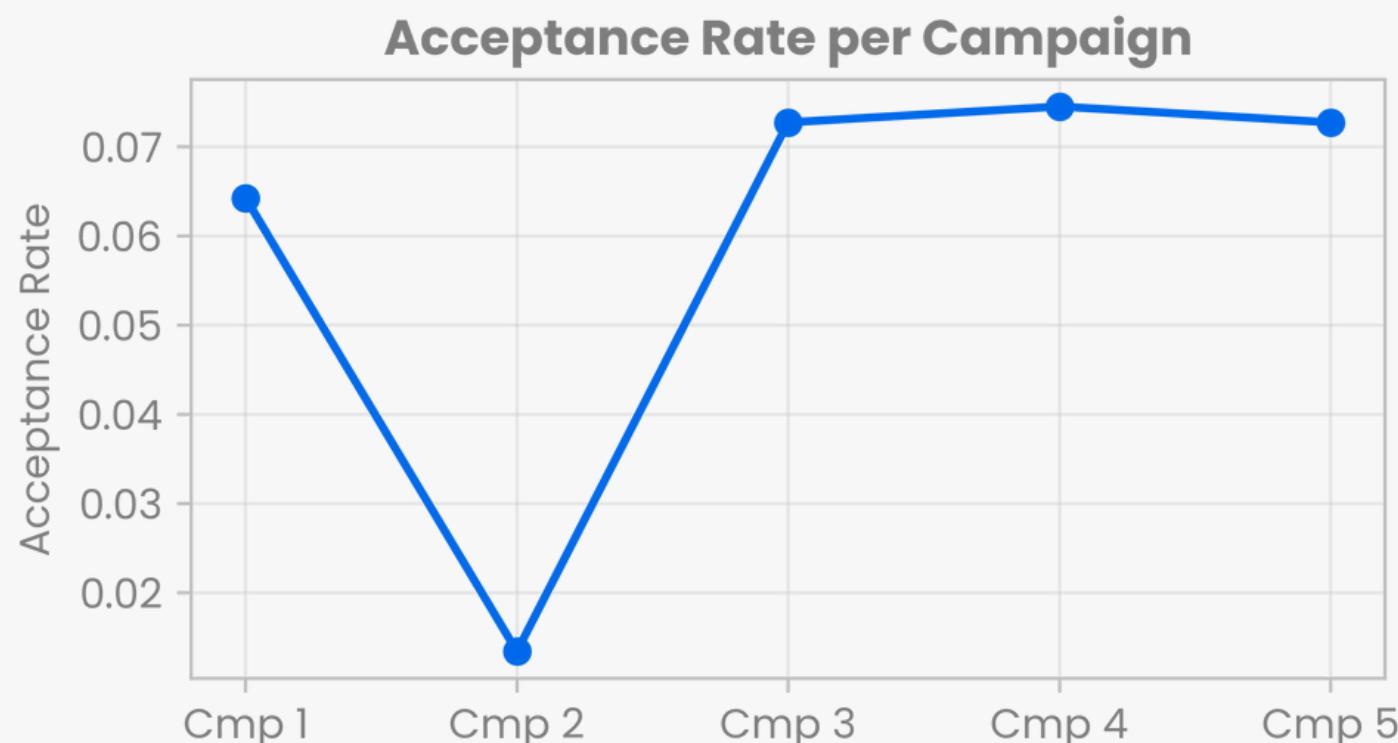


There is a **strong synergy** between **Web, Catalog, and Stores**, showing customers move across these channels seamlessly. In contrast, Deals is isolated, appealing mainly to bargain-driven customers.



## Recommendation

Enhance omnichannel with consistent promotions and click & collect. Engage Deals customers through targeted upsell and cross-sell to connect them with other channels.



# Winning Campaigns Boost Customer Spending

**Marketing campaigns significantly increase customer spending and purchase frequency**, even a single campaign drives strong results. Most customers, however, remain untouched by any campaign, representing a major growth opportunity. Campaign performance varies sharply: Campaign 2 failed, while Campaigns 3–5 succeeded.

## 💡 Recommendation

Analyze Campaigns 3–5 to replicate successful strategies, while reviewing Campaign 2 to avoid past mistakes in future efforts.

# Data Preparation



## Handling Missing & Duplicate Data

- Checked for missing values using `df.isnull().sum()`
- Imputed missing numeric values with column **medians**
- Identified duplicate rows using `df.duplicated().sum()`

### Feature Creation

#### Online Engagement

- Avg\_Web\_Visits
- Online\_Purchase\_Ratio



## Feature Engineering

- Converted **Dt\_Customer** to **datetime**
- Mapped Education to **Graduate/Undergraduate**
- Mapped Marital\_Status to **Menikah, Cerai, Lajang**
- Created **new features**

### Feature Creation

#### Demographic Features

- Age
- Is\_Parent



## Feature Selection & Encoding

- Dropped **irrelevant columns**
- Applied **one-hot encoding** to Education and Marital\_Status

### Feature Creation

#### Behavioral Features

- Total\_Campaigns\_Accepted
- Average\_Spend
- Total\_Spent



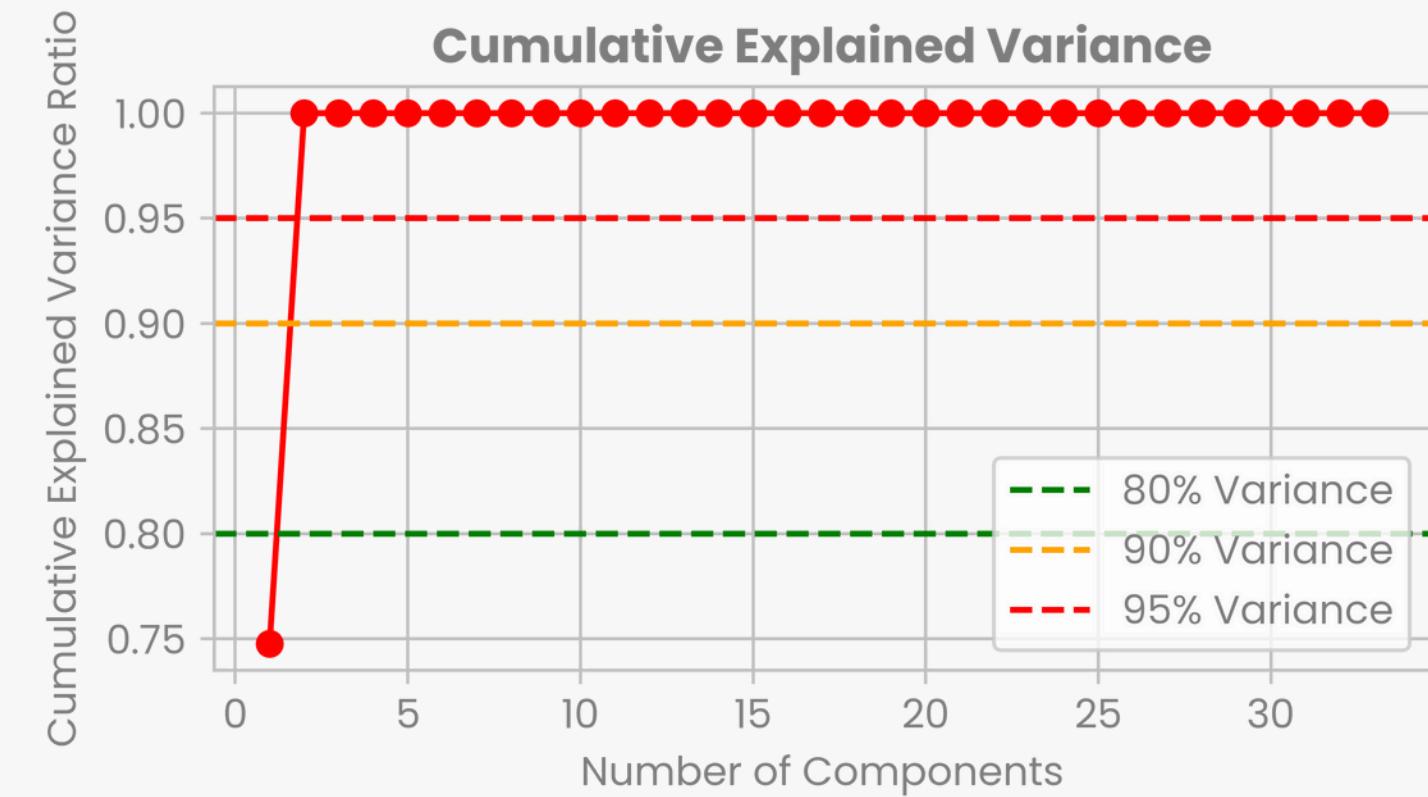
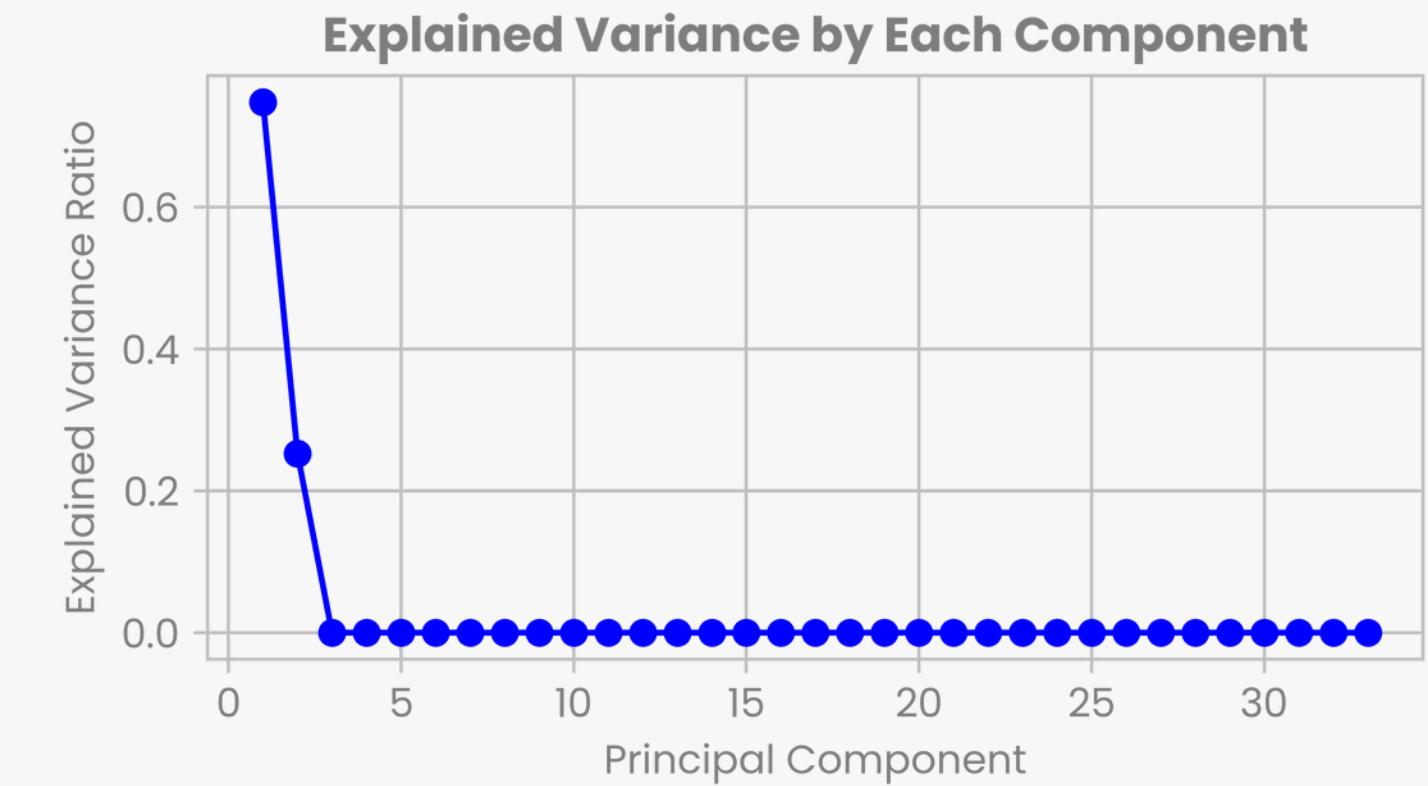
## Feature Scaling

- Scaled numerical features using **RobustScaler**

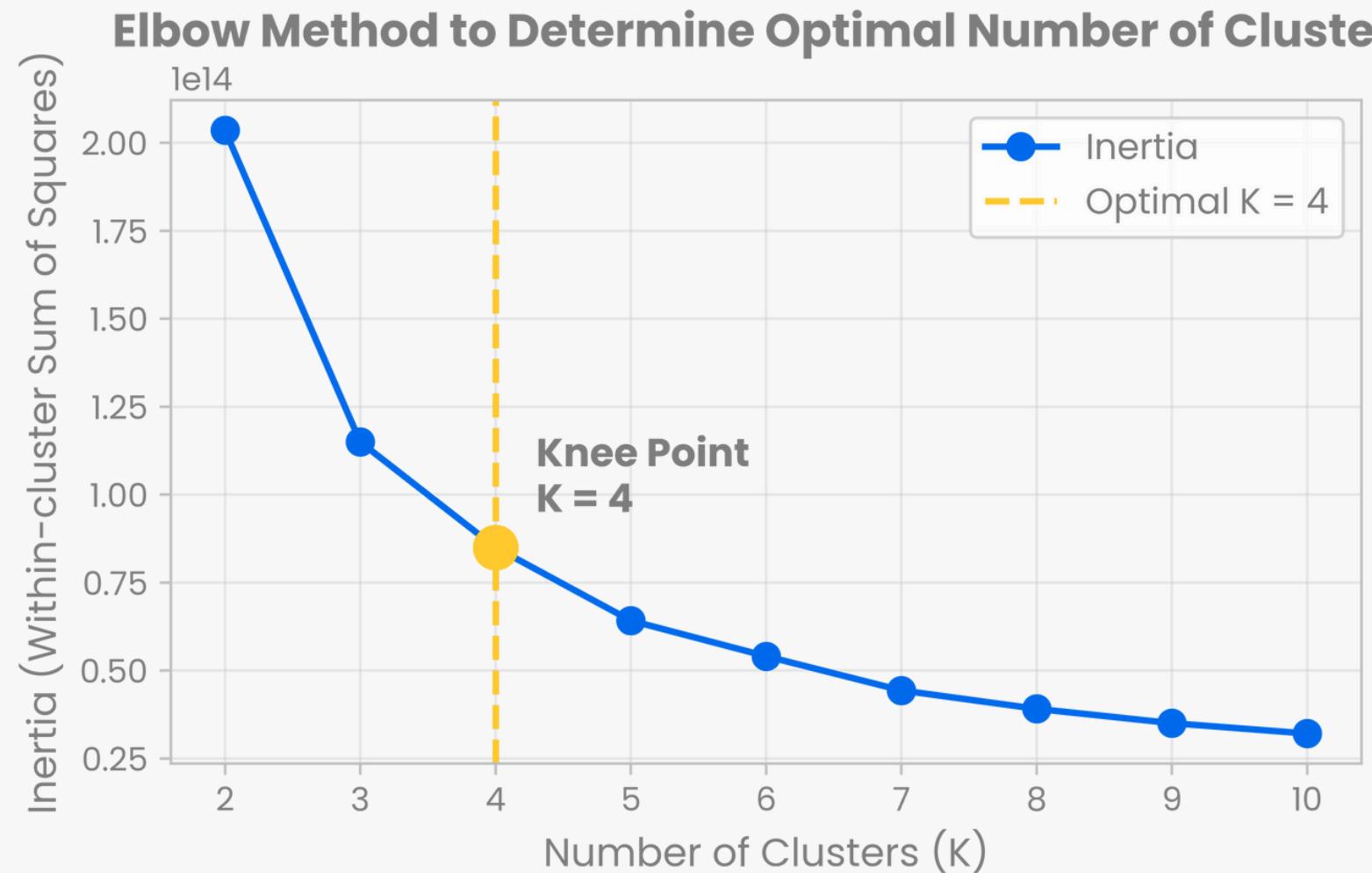
# Dimensionality Reduction

The PCA analysis shows that just `{n_components_90}` components are enough to capture **90% of the dataset's variance**, making it **efficient to reduce dimensionality without losing essential information**. Using **2 components** allows for simplified visualization while preserving key patterns.

`n_components_90 = 2`



# Optimal Number of Clusters



Based on the **Elbow Method**, the **optimal number of clusters** was identified at  **$K = 4$** , where the curve shows a clear inflection point. Therefore, K-Means clustering with four clusters was selected as the most suitable configuration for this dataset.

- **Silhouette Score: 0.62**
- **Calinski-Harabasz Index: 3192.6**
- **Davies-Bouldin Index: 0.82**

# Exploring Clustering Methods

## ❖ Algorithms Applied

- **K-Means**: 4 clusters
- **Agglomerative Clustering**: 3 clusters
- **DBSCAN**: density-based clustering
- **Gaussian Mixture Model (GMM)**: 2 components

## ⚙ Optimization Techniques

- **K-Means**: Elbow Method
- **Agglomerative**: Dendrogram analysis
- **DBSCAN**: k-distance plot
- **GMM**: AIC and BIC criteria

## ↗ Performance Metrics



Silhouette Score

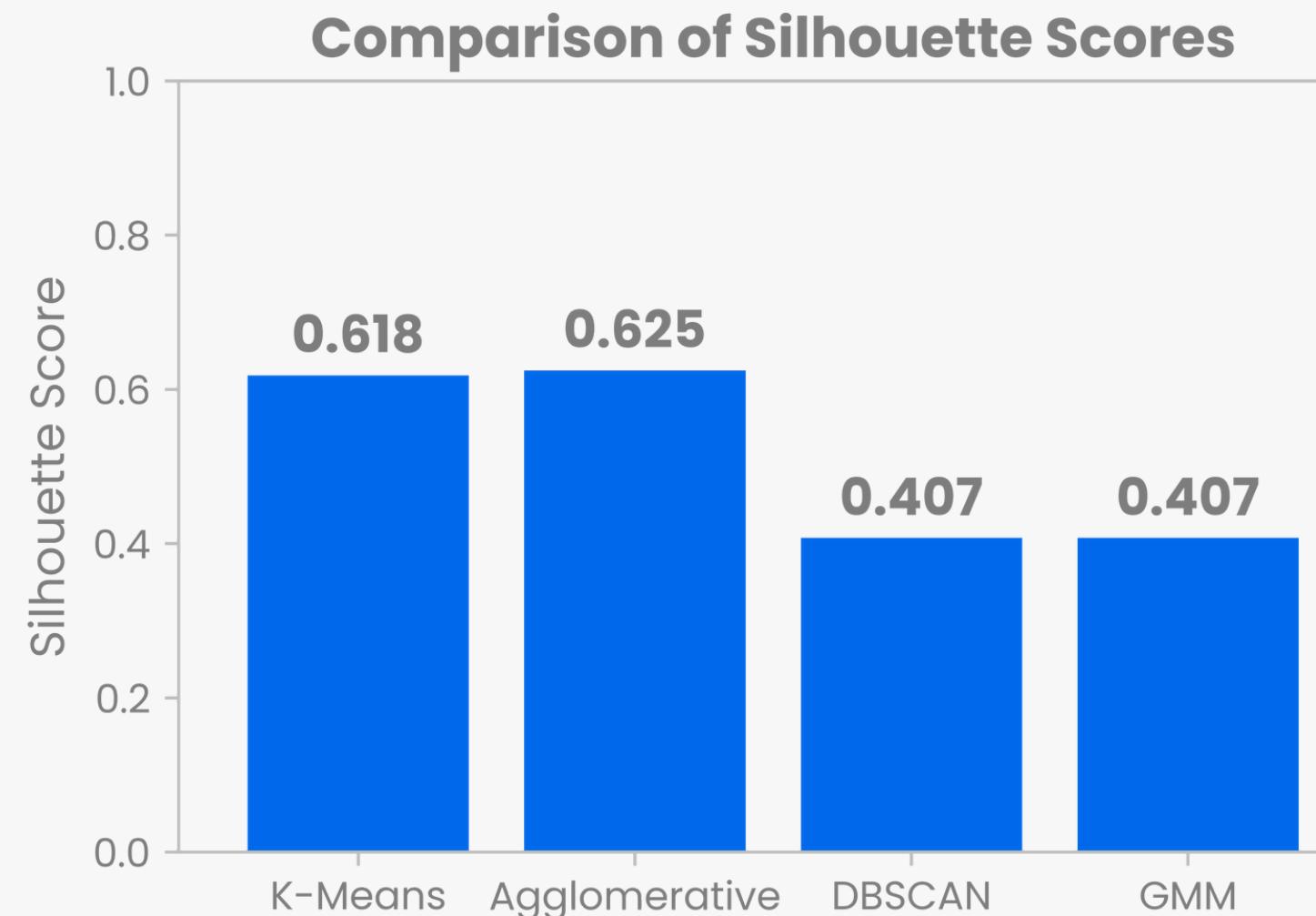


Calinski-Harabasz Index



Davies-Bouldin Index

# Model Selection: K-Means

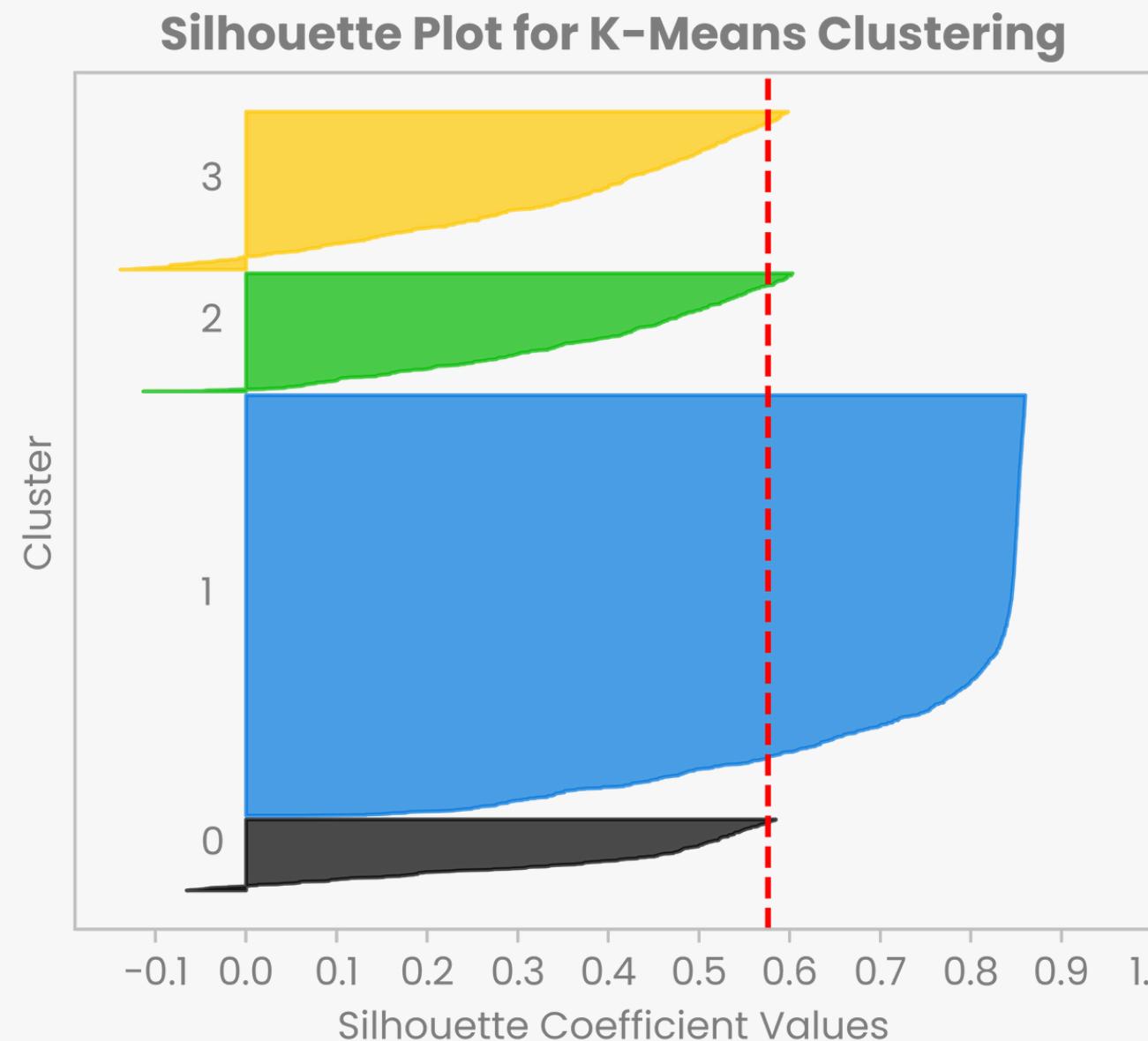


Although Agglomerative Clustering achieved slightly better scores (Silhouette: 0.625, CH: 3347.925, DB: 0.766), K-Means was chosen for practical considerations:

- $O(n)$  vs  $O(n^3)$  **complexity**
- **Better** performance on **large** datasets
- **Clear centroid-based** cluster representation
- More **suitable** for **production** environments
- **Reliable** results across **different data distributions**

Final Model: K-Means (4 clusters)

# Cluster Quality Assessment

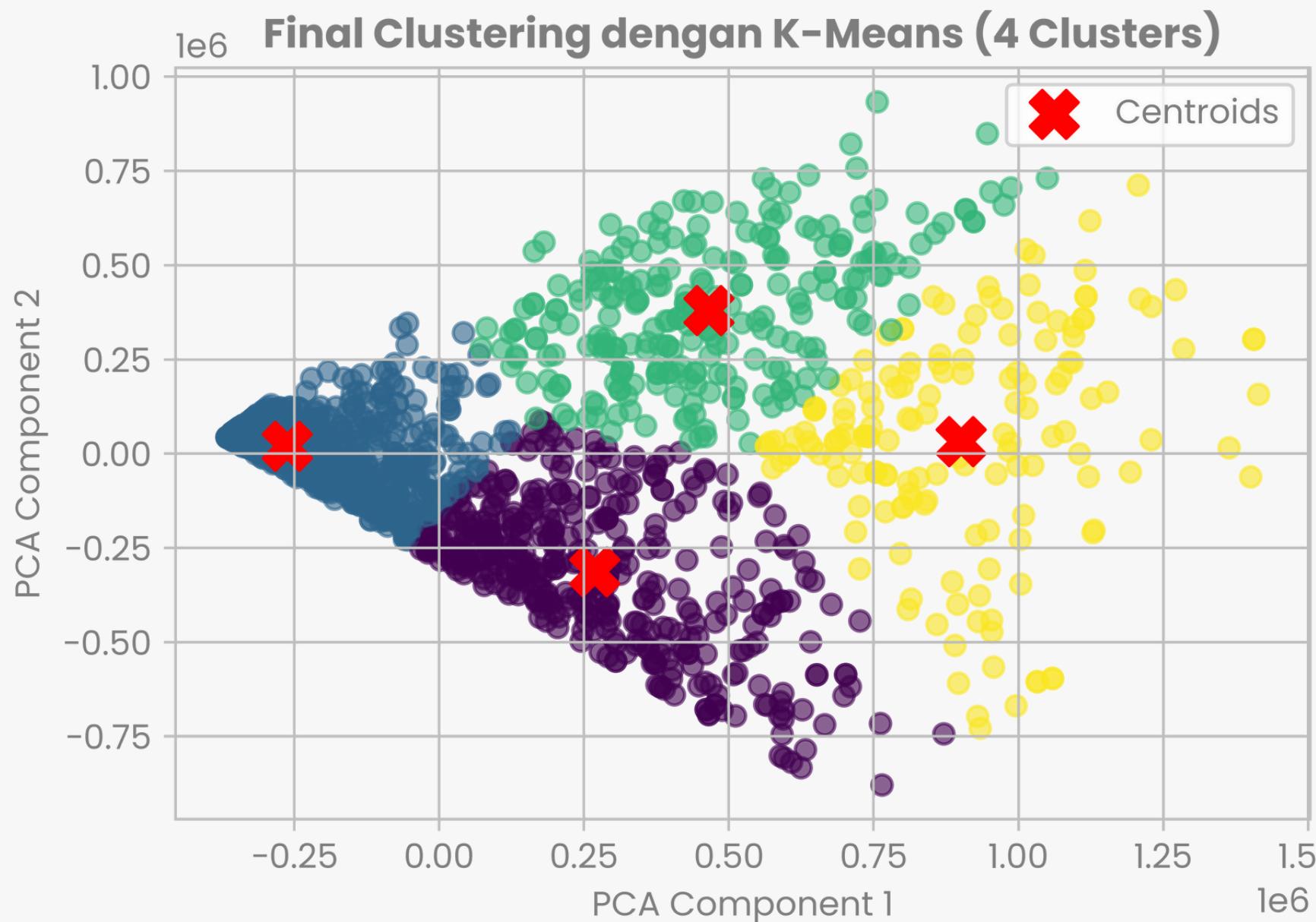


The average silhouette coefficient is 0.576, indicating a generally **well-separated** and **valid 4-cluster structure**.

- Clusters 1, 2, and 3 show **good cohesion**, especially Cluster 1.
- Cluster 0 has the **lowest quality**, with all members scoring below average.

Silhouette Score: 0.576

# Cluster Distribution and Sizes



## Cluster Sizes

- Cluster 1 = **1416**
- Cluster 0 = **413**
- Cluster 2 = **252**
- Cluster 3 = **159**

**Cluster 1 dominates** with 1416 members, while **Cluster 3 is the smallest** (159), indicating uneven segment distribution, some are concentrated, others niche. Cluster 0, with 413 members, shows weaker cohesion and warrants closer analysis based on silhouette scores.

# Customer Profile by Cluster

Feature	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Age	49.343	46.04	43.958	43.394
Income	5469.159	5843.271	5625.082	5244.831
Kidhome	0.143	0.02	0.663	0.019
Teenhome	0.77	0.116	0.541	0.1

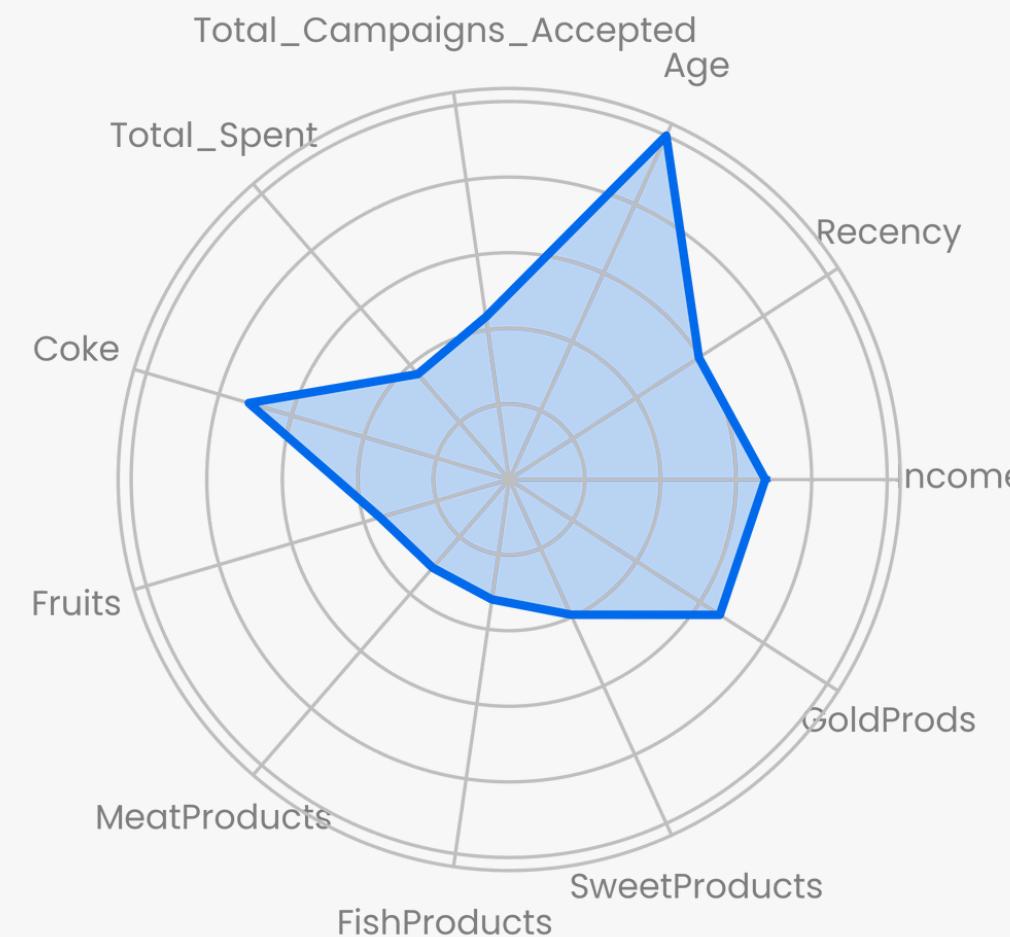


Cluster 2 has the largest households, while Cluster 1 has the highest income.

# Customer Behavior by Cluster

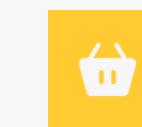
Feature	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Total_Campaigns_Accepted	0.562	0.331	0.101	1.244
Response	0.335	0.252	0.348	0.266
Total_Spent	387,691.24	822,195.22	120,462.37	834,362.50
Online_Purchase_Ratio	0.431	0.19	0.514	0.256
Avg_Web_Visits	0.335	0.252	0.348	0.266

# Cluster 0: Promo Seekers



## Behavior & Demographics

- Middle-aged with majority having children
- Low income, minimal spending
- Poor campaign response
- High web browsing



## Product Preferences

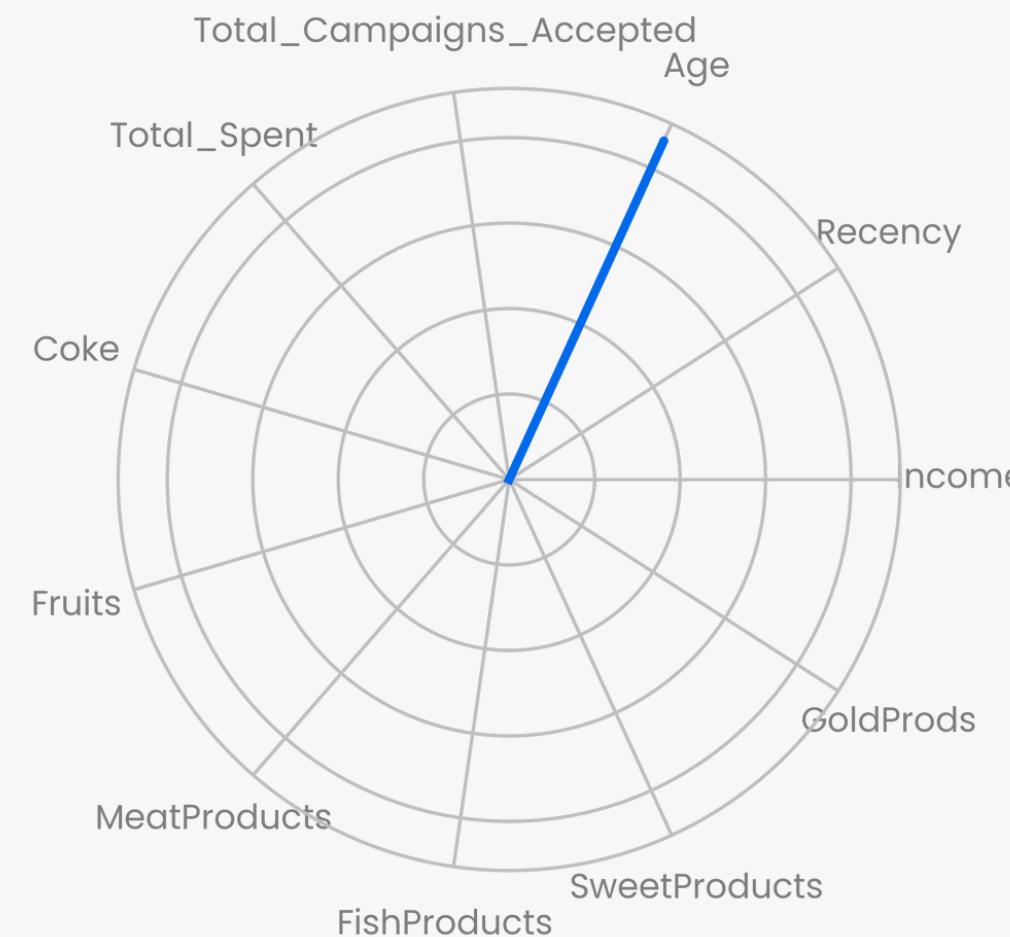
Low engagement across all product categories



## Recommendation

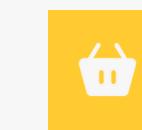
Engage them with frequent discounts, "Buy 1 Get 1" offers, and bundled packages.  
A loyalty program with point accumulation would resonate well.

# Cluster 1: Frugal Families



## Behavior & Demographics

- Middle-aged, mostly childless
- Highest income, premium spending
- Excellent campaign acceptance
- Active online shoppers



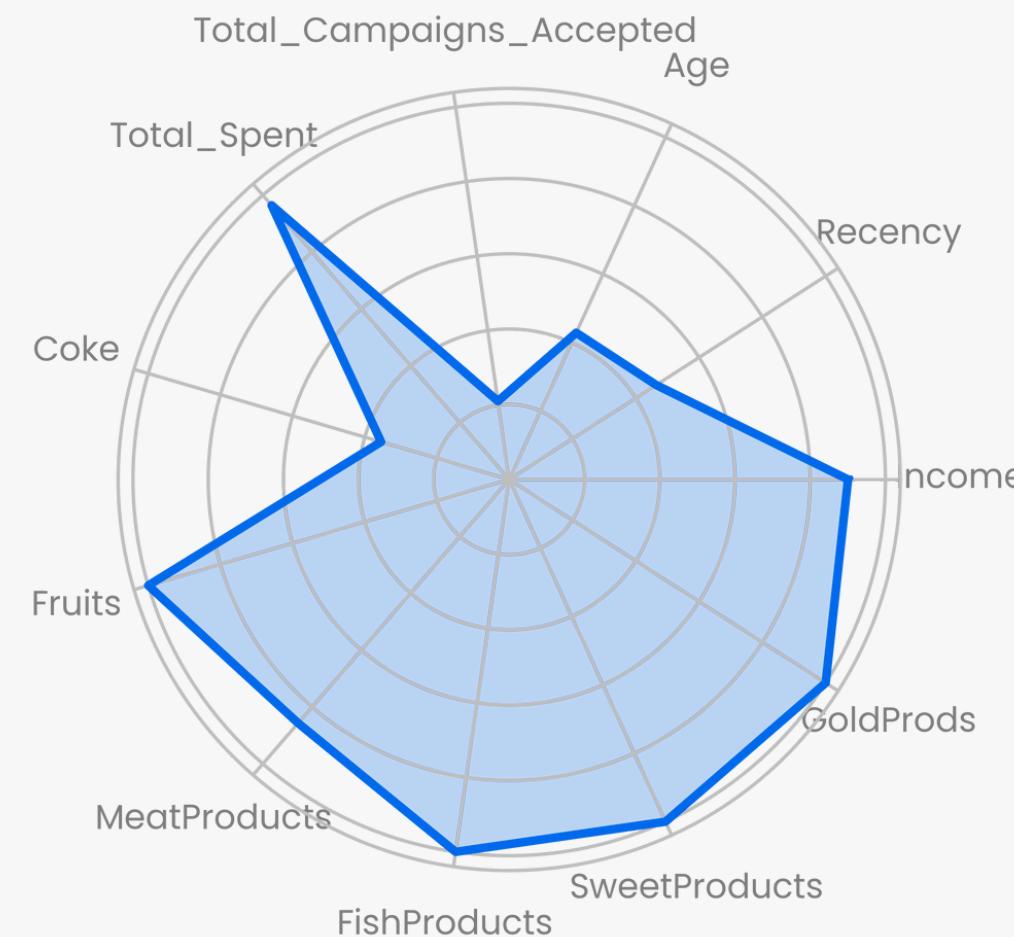
## Product Preferences

Heavy investment in Coke products, gold items, premium categories

## 💡 Recommendation

Focus on affordability and essentials. Offer bulk discounts and family-friendly pricing. Without proper engagement, they are at risk of churn.

# Cluster 2: Affluent Selectives



## Behavior & Demographics

- Oldest segment, majority with children
- Upper-middle income, moderate spending
- Good campaign response
- Active online shoppers



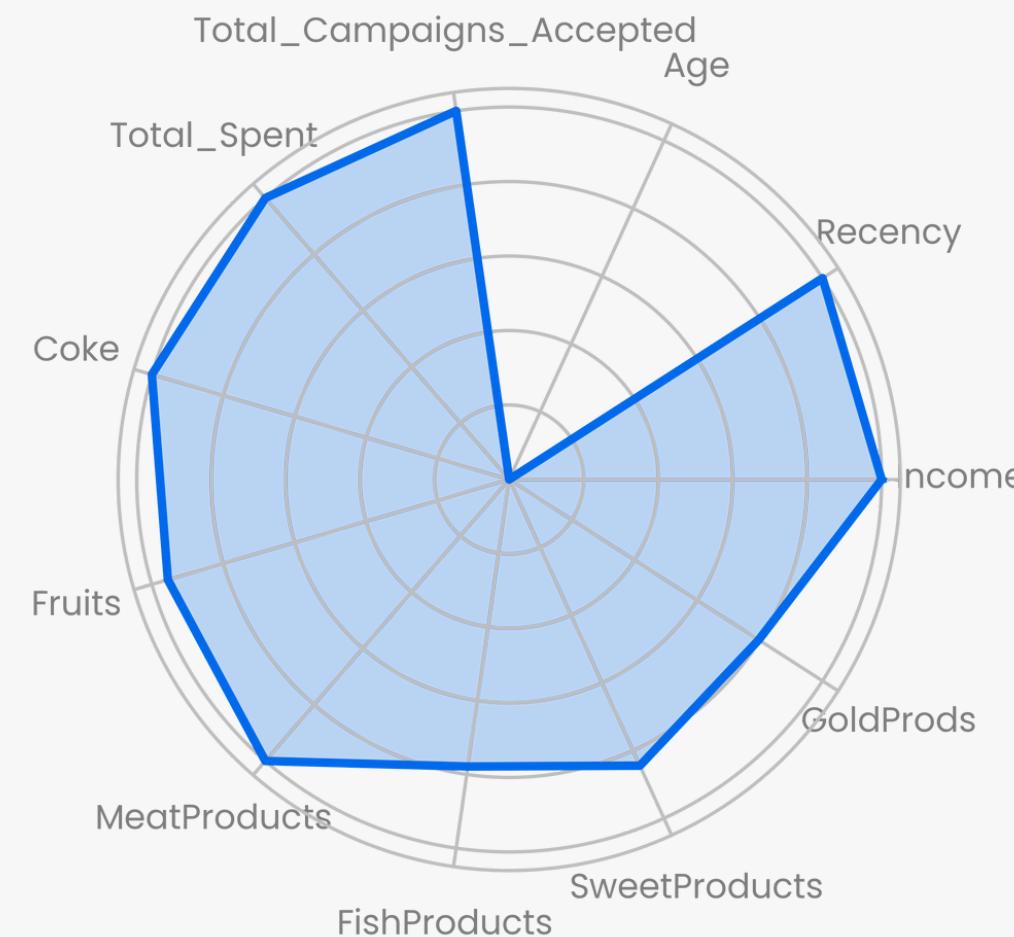
## Product Preferences

Balanced spending, preference for Coke and meat products

## 💡 Recommendation

Skip the discounts. Instead, highlight product quality, uniqueness, and benefits.  
Content marketing (e.g., healthy recipes) may be more effective.

# Cluster 3: Priority Customers



## Behavior & Demographics

- Middle-aged, few children
- High income, selective high spending
- Poor campaign response
- Strong store preference



## Product Preferences

Focus on sweet products and gold items, selective about other categories



## Recommendation

Prioritize this group. Offer exclusive loyalty programs, early access to new products, and personalized deals to make them feel valued.

# Key Strategic Recommendations



## Prioritize “Priority Customers”

Allocate the **largest portion of the budget** to this segment. Offer **exclusive services**, **early access** to products, and **personalized deals** to maintain their loyalty.



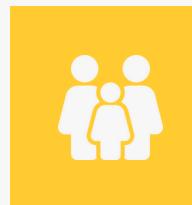
## Activate “Promo Seekers”

Tap into their love for deals with **“Buy 1 Get 1”** promos, **bundle** offers like Coke + Meat, and **point-based** loyalty programs to boost engagement.



## Retain “Affluent Selectives”

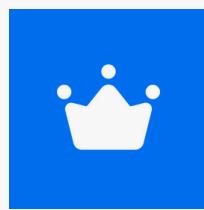
Skip discounts, **highlight product quality, uniqueness, and benefits**. Use content like premium recipes or brand stories to connect with customers.



## Stabilize “Frugal Families”

Offer **essential products** at **fair prices**, with **family packs** or **bulk deals** to boost loyalty and repeat buys.

# Potential Impact Analysis



## Priority Customers

With an **average spend of 834,362.50**, a **15% increase** due to heightened loyalty and exclusive offers would result in an additional **~20 million in revenue**.



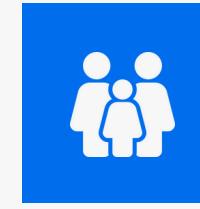
## Promo Seekers

With an **average spend of 387,691.24**, a **20% increase** driven by targeted promotions could generate an additional **~32 million in revenue**.



## Affluent Selectives

This group has the **lowest average spend** at **120,462.37**. A **10% increase** from its **252 members** could yield an additional **~3 million in revenue**.



## Frugal Families

This large segment of **1,416 customers** has an average spend of **822,195.22**. A modest **5% increase** could generate an additional **~58.2 million in revenue**.

# Get in Touch



[Source Code](#)



[LinkedIn](#)



[afsilmis@gmail.com](mailto:afsilmis@gmail.com)

