# **Marketing Analytics Final Report**

## **1. Introduction**

In today’s fast-moving consumer goods or FMCG industry, companies generate vast amounts of customer data through daily transactions, online interactions, and marketing campaigns. Analysis shows that leveraging this data effectively is essential for businesses aiming to drive growth, retain customers, and expand their market presence. Marketing analytics, powered by statistical techniques and machine learning methods such as K-means clustering, enables companies to uncover meaningful patterns in customer behavior and make data-driven decisions. By identifying distinct customer segments based on demographic, behavioral, and transactional characteristics, businesses can tailor their marketing strategies to meet the needs of each group better, ultimately improving engagement, loyalty, and conversion rates. This analysis provides actionable insights derived from a real-world FMCG customer dataset, offering a foundation for strategic marketing planning and personalized customer outreach.

Analysis of a retail FMCG customer dataset using R programming reveals key behavioral trends and allows for the segmentation of customers into distinct groups. These segments are based on age, income, spending behavior, and purchase frequency. Insights derived from this analysis can guide the development of targeted marketing strategies that improve customer engagement, strengthen brand loyalty, and increase conversion rates.

## **2. Dataset Overview**

The dataset contains detailed records of 2,240 customers, with features covering demographics, purchasing habits, marketing campaign responses, and product spending. The core data fields include:

* **Demographic variables**: Unique ID, year of birth, education level, marital status, income, and household composition..
* **Purchase and product variables**: Amount spent on various product categories over the last year, and counts of purchases made via different channels (web, catalog, store).
* **Promotion-related variables**: Number of deals used, participation in various marketing campaigns, and response to the latest campaign.
* **Customer tenure and activity**: Date of first purchase and recency of last purchase.

Additional variables engineered for analysis included **Age, Family\_Size, Total\_Spent, Total\_Purchases,** and **Customer\_Since\_Years**.

## **3. Data Preprocessing and Feature Engineering**

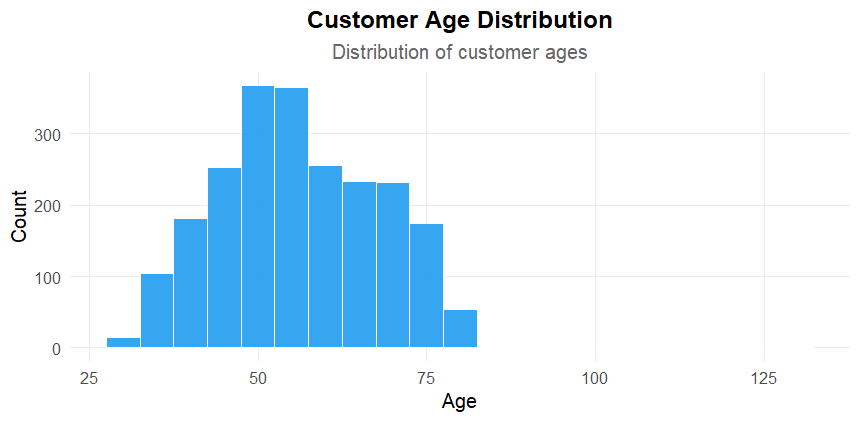
To prepare the data for analysis, the following steps were conducted:

* **Missing values**: Detected and imputed, particularly in the **Income** column, using the median to ensure a realistic and unbiased estimate.
* **Data cleaning**: Removed inconsistencies and converted relevant categorical fields like **Education, Marital\_Status, Complain, Response** to factor variables for better handling in analysis.
* **Feature engineering**: Created new variables to capture key aspects of customer behavior better:  
  + **Age** (from year of birth)
  + **Family\_Size** (sum of children and teenagers at home)
  + **Total\_Spent** (total spending across all categories)
  + **Total\_Purchases** (aggregate of purchases across channels)
  + **Customer\_Since\_Years** (years since first recorded transaction)
* **Scaling**: Standardized continuous features to ensure all variables contributed equally to clustering.

These steps ensured that the dataset was reliable and suitable for both statistical analysis and machine learning.

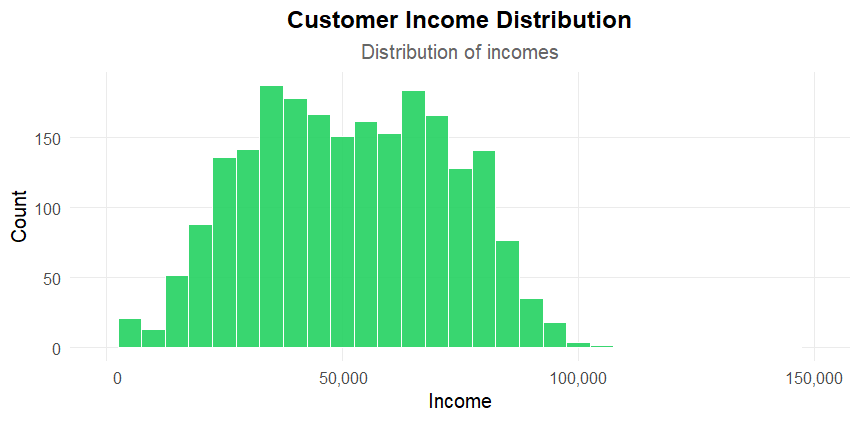
## **4. Exploratory Data Analysis (EDA)**

An Exploratory Data Analysis was conducted to understand the underlying patterns and distributions within the dataset before proceeding with advanced modeling techniques such as customer segmentation. This step provided foundational insights into customer demographics, income levels, purchasing behavior, and campaign response rates. The visualizations generated during EDA are described below.

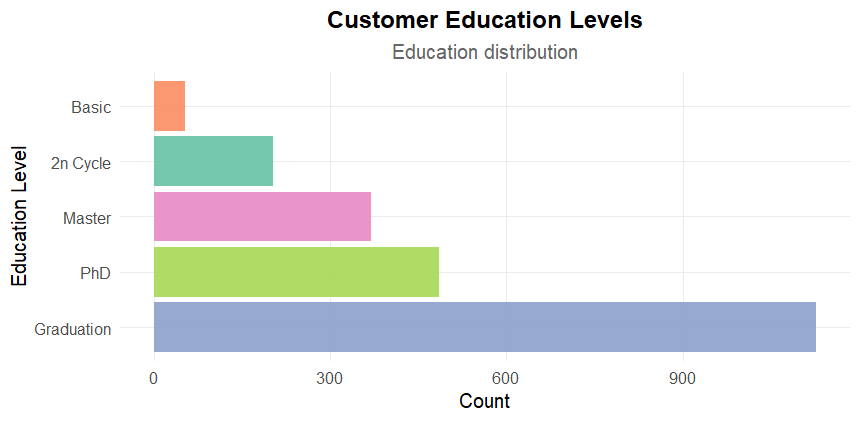
**Age**: The histogram of customer age revealed that the majority of customers fall within the 45–75 years range, with a peak around 60 years old, as shown in Figure 1. Younger customers were less represented, suggesting potential opportunities for market expansion among younger adults. This indicates that the retail company primarily serves an older demographic, which may influence marketing strategies.

**Figure 1:** Showing Customer Age Distribution

**Income**: Most customers reported annual household incomes between about $30,000 and $80,000, with a smaller group earning significantly more, as shown in Figure 2. This distribution suggests a core customer base with moderate financial capacity, but also highlights the presence of high-income individuals who could be targeted with premium products or exclusive offers.

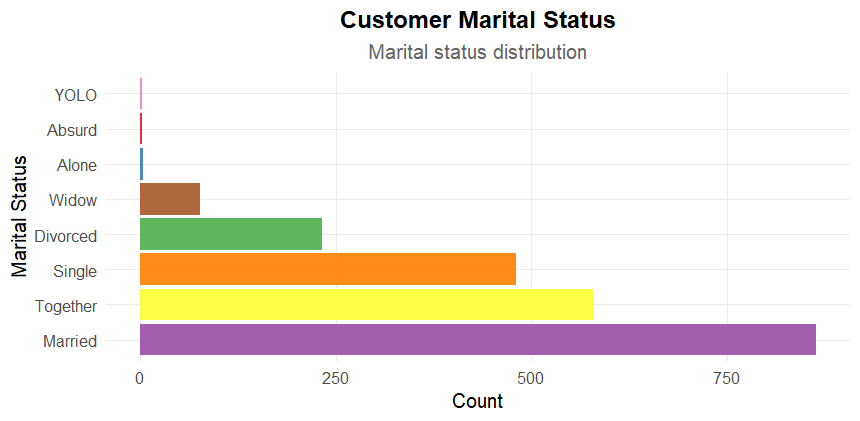


**Figure 2:** Showing Customer Income Distribution

**Education and Marital Status**: Among all educational qualifications, "Graduation" was the most common category, as shown in Figure 3. This implies that a significant portion of the customer base has completed tertiary education. This finding supports the development of marketing messages that are informative, benefit-oriented, and possibly more complex than basic advertising, appealing to a relatively educated audience.

**Figure 3:** Showing Customer Education level

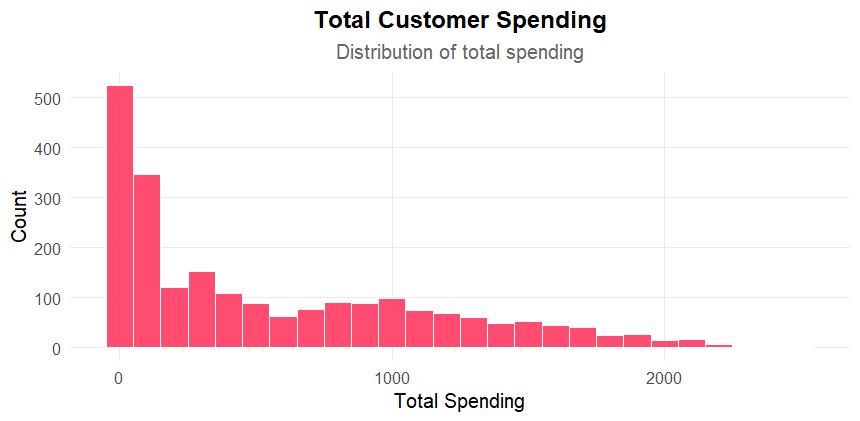
Similarly, The marital status breakdown showed that most customers were either married or Together, with fewer customers identifying as divorced or widowed, as shown in Figure 4. Married customers may represent households with shared purchasing decisions, while single individuals may prioritize convenience and personalization.



**Figure 4:** Showing Customer Marital Status

These insights can guide segmentation efforts for instance, married couples may respond well to family-sized packages, whereas singles might prefer individualized offers.

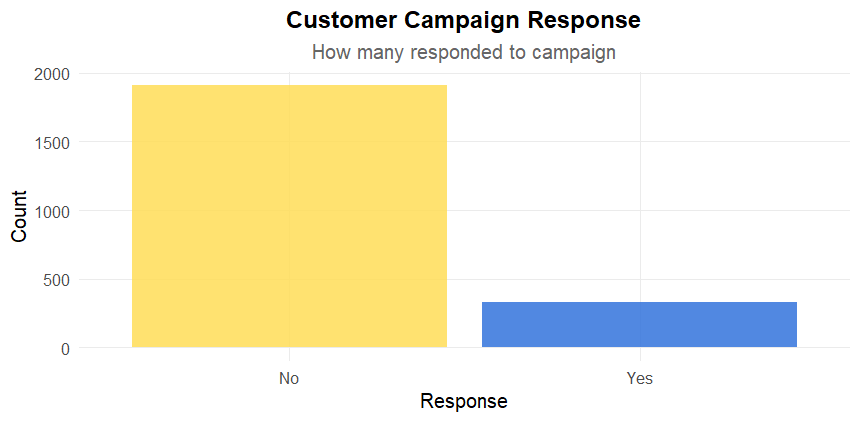
**Total Spending**: A wide range of annual spending was observed, with a few customers spending significantly more than average, as shown in Figure 5.



**Figure 5:** Showing Customer Total Spending

Identifying such groups early helps in designing VIP programs, personalized discounts, and tailored recommendations to maximize their lifetime value.

**Response Rate**: Only about 15% of customers accepted the most recent marketing campaign, as shown in Figure 6.



**Figure 6:** Showing Customer Response Rate

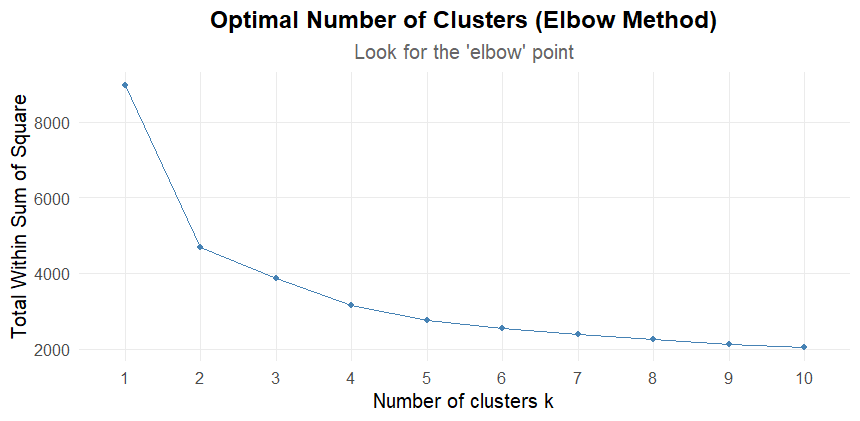
This low conversion rate suggests that the current campaign messaging or targeting may not be resonating with the broader customer base.

Each visualization served as a crucial tool for interpreting the data and informing business decisions. By understanding the composition and behavior of the customer base, the company can better align its marketing initiatives with the needs and preferences of distinct customer groups.

## **5. Customer Segmentation Using K-Means Clustering**

To discover actionable groups within the customer base, K-means clustering was applied using the following standardized features: age, income, total spending, and total purchases.

* **Choosing the number of clusters**: The "elbow method" was used to determine the optimal number of clusters by plotting the within-cluster sum of squares for different k values and locating the point adding more clusters does not significantly improve the model, as shown in Figure 7.



**Figure 7:** Showing Customer Response Rate

* **Cluster assignment**: Each customer was assigned to a segment based on similarities in the selected features.

**Cluster summary statistics**:

| **Segment** | **Count** | **% of Total** | **Avg Age** | **Avg Income** | **Avg Spent** | **Avg Purchases** | **Avg Family Size** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 508 | 22.7% | 48.0 | $72,726 | $1,137 | 19.3 | 0.6 |
| 4 | 480 | 21.4% | 68.5 | $69,817 | $1,080 | 19.4 | 0.6 |
| 3 | 480 | 21.4% | 66.6 | $43,669 | $177 | 8.4 | 1.4 |
| 1 | 772 | 34.5% | 47.5 | $33,154 | $100 | 6.4 | 1.1 |

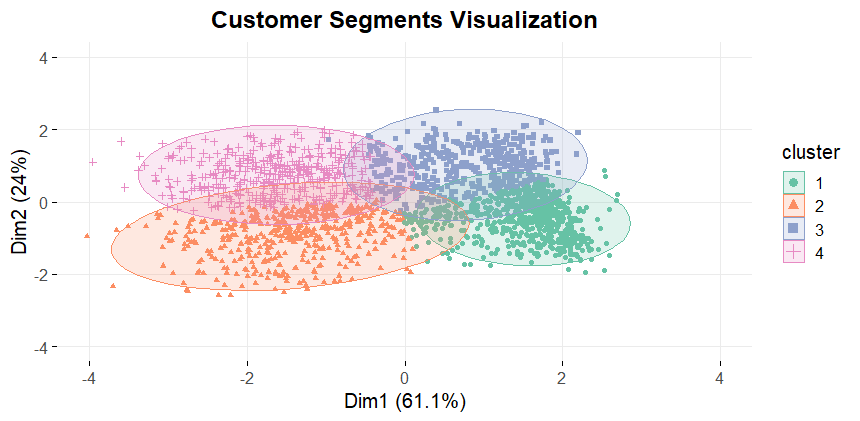
## **6. Interpretation of Customer Segments**

**Segment 1: Budget-Conscious Families**This group is younger (avg. 47.5 years), has the lowest income and spending, and tends to have more family members. They purchase less frequently, likely making selective purchases based on necessity.

**Segment 2: High-Spending Young Professionals**Moderate in age (avg. 48), this segment has the highest income and spending, and makes purchases frequently. They are less likely to have children at home, making them ideal for targeted premium offers and exclusive digital campaigns.

**Segment 3: Middle-Income Retirees**Older customers (avg. 66.6 years) with moderate income and spending. They have a larger average family size, possibly indicating multi-generational households or active social lives.

**Segment 4: Loyal Senior Shoppers**The oldest group (avg. 68.5 years) but with high income and spending, and the highest purchase frequency. Their small family size suggests empty nesters. They are highly engaged and represent a loyal customer base.



**Figure 8:** Showing Customer Segments (K=4)

## **7. Recommended Marketing Strategies**

Based on the insights from segmentation, the following strategies are suggested:

**Segment 1: Budget-Conscious Families**Focus on discounts, bundled offers, and loyalty rewards to increase engagement and wallet share. Time promotions around school holidays and family events.

**Segment 2: High-Spending Young Professionals**Deploy targeted digital marketing, early access to premium products, and VIP benefits. Personalized online experiences and exclusive app offers are likely to be effective.

**Segment 3: Middle-Income Retirees**Promote convenience-focused services, such as home delivery or in-store assistance. Consider tailored email campaigns highlighting product value and health benefits.

**Segment 4: Loyal Senior Shoppers**Strengthen loyalty with personalized communication, appreciation programs, and in-store perks. Consider senior-specific discounts and special recognition for long-term engagement.

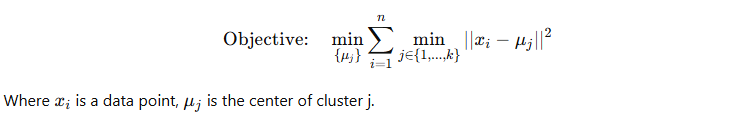
Across all segments, optimize channel mix according to preferences revealed by data (e.g., web for younger and high-spend customers, catalogs or in-person for seniors).

## **8. Explanation of the Machine Learning Algorithm**

**K-means Clustering** is an unsupervised machine learning algorithm that groups data into *k* clusters by minimizing the within-cluster variance. The algorithm works by:

1. Initializing *k* cluster centers randomly.
2. Assigning each data point to the nearest cluster center based on Euclidean distance.
3. Recalculating cluster centers as the mean of assigned points.
4. Iterating steps 2–3 until cluster assignments stabilize.

**Mathematically**, the objective is to minimize the sum of squared distances between each point and its assigned cluster center:



**Feature scaling** is essential because K-means relies on distances; unscaled features can bias the clustering.

## **9. Conclusion**

This analysis demonstrates the value of customer segmentation for targeted marketing. By combining descriptive analytics, data visualization, and machine learning, the company can better understand its customers and design strategies that cater to distinct needs, thus optimizing engagement and retention.

**Future directions** could involve testing alternative algorithms (such as hierarchical clustering or Gaussian Mixture Models) and incorporating external data sources (e.g., customer reviews, social media data) for even deeper insights.

## **Appendix**

* **R code file**: Marketing\_Analysis.R
* **Full output dataset**: Customer\_Segments\_Output.csv
* **Segment statistics table**: Segment\_Summary.csv