EXECUTIVE SUMMARY

This analysis focuses on segmenting Marketplace Supermarket's customers to understand their distinct characteristics and develop personalized strategies that enhance their shopping experiences.

The primary objective of this analysis was to identify meaningful customer segments based on five key variables: age, income, spending on products, online shopping activity, and household composition. The aim was to create targeted strategies that cater to the unique needs of each group and are designed to enhance customer satisfaction, increase loyalty, and drive revenue growth for Marketplace.

Analysis

A Two-Step Cluster Analysis was conducted on the customer data to identify distinct segments. Five key variables were analyzed: age, income, product spending, online purchases, and household with teenagers. ANOVA and Chi-Square tests were used to confirm the statistical significance of the differences between the segments.

Key findings
The analysis revealed three distinct customer segments:

1. Moderate Spenders (47.9%): Middle-aged with moderate income and spending, active online shoppers.
2. High Spenders (26.1%): Wealthy, middle-aged with the highest product spending and frequent online purchases.
3. Low Spenders (26.0%): Younger, low-income, minimal spending and online activity.

The statistical tests (ANOVA and Chi-Square) confirmed that these segments significantly differ in behavior and characteristics.

Strategies

- Moderate Spenders: Offer loyalty programs and targeted discounts and emphasize online shopping convenience to maintain engagement.

- High Spenders: Introduce exclusive deals, premium product lines, and personalized shopping experiences to enhance satisfaction.
- Low Spenders: Promote affordable bundles and targeted advertising and provide incentives for online shopping to increase spending.

INTRODUCTION

This analysis aimed to segment Marketplace Supermarket's customer base into distinct groups based on essential characteristics such as age, income, spending habits, online shopping activity, and household composition. By identifying these segments, the goal was to develop tailored strategies that address each group's unique needs and preferences. This analysis is crucial for improving customer targeting, enhancing the shopping experience, and driving sales through personalized marketing initiatives. Understanding these segments allows Marketplace to make data-driven decisions that align with customer expectations, ultimately leading to increased satisfaction and loyalty.

METHODOLOGY

Variables Used:

1. **Age**: Represents the demographic distribution of the customers.
2. **Income**: Indicates customers' purchasing power and financial stability. Income helps differentiate between high- and low-spending segments.
3. **Spend on Products**: Reflects the actual spending behavior of customers in-store. This variable helps identify price-sensitive customers versus those with high disposable income.
4. **NWebPurchases**: Measures the frequency of online shopping. This variable helps identify internet-friendly customers who prefer convenience versus in-store shopping.
5. **TeenHouseholds**: Indicates the presence of teenagers in the household, which influences spending patterns on family-oriented products.

Data Cleansing: The age variable had extreme values while running the frequency analysis. There were three outliers in the age variable, more than 120 years (*Figure 1.1)*. So, I identified the outliers using the "explore" function under the descriptive statistics. Then, I permanently removed the outliers through the "select cases" function in the Data tab of SPSS and saved the file (*Figure 1.2*). This step removed the outliers from my dataset, which will provide more accurate test results.

Recoding into different variable
Since the "NWebPurchases" variable was categorical, its values ranged from 0 to 27. So, I recoded the same variable into a continuous variable(scale) because its high values (27) would dominate the cluster analysis if it were calculated as a categorial variable. (*Figure 2*)

Clustering Analysis

Clustering is ideal for this analysis because it allows us to segment customers into meaningful groups based on their similarities across multiple variables (age, income, spending, etc.). By identifying natural clusters, we can categorize customers in a way that reflects their purchasing behavior, enabling tailored marketing strategies.

The objective is to uncover patterns in the data without predefining groups. This method gives a data-driven approach to identifying how customer characteristics combine to form distinct groups.

Bivariate Analysis

1. ANOVA: ANOVA was used to test if there were significant differences in **age**, **income**, **spending on products**, and **NWebPurchases** between the identified clusters.
   Our analysis helped confirm that the differences between the customer segments are statistically significant, providing insight into how these key variables differentiate the groups.

2. Chi-Square Test: The Chi-Square test was applied to explore the relationship between **TeenHouseholds** and the clusters, determining if the presence of

teens in the household was significantly associated with the customer segments.
This test helped confirm if the family structure significantly impacts customer segmentation, essential for targeting family-oriented products.

ANALYSIS RESULTS

Based on the **Two-Step Cluster Analysis**, I identified three distinct customer segments. (*Figure 3.1 and Figure 3.2)*

Cluster 1: Family-focused spenders (47.9%)

- **Age**: Middle-aged (~55 years)
- **Income**: Moderate (~$52,606)
- **Spending**: Moderate spenders on products (~$145.97)
- **Online Purchases**: Average online shopper (4.53)
- **Household**: Includes teenagers (TeenHouseholds = 1)

Family-focused spenders: The group comprises middle-aged customers with moderate income and spending habits. They are family-oriented, as indicated by the presence of teenagers in the household. They balance their spending between in-store purchases and online shopping.

Cluster 2: Wealthy professionals (26.0%)

- **Age**: Younger (~42 years)
- **Income**: Low (~$29,386)
- **Spending**: Minimal spenders on products (~$42.53)
- **Online Purchases**: Rare online shoppers (~2.18)
- **Household**: No teenagers (TeenHouseholds = 0)

Wealthy professionals: This segment comprises high-income, middle-aged high spenders in-store and online. They have no teenagers in the household, which suggests focusing on personal or professional needs.

Cluster 3: Budget Conscious Youth

- **Age**: Younger (~42 years)
- **Income**: Low (~$29,386)
- **Spending**: Minimal spenders on products (~$42.53)
- **Online Purchases**: Rare online shoppers (~2.18)
- **Household**: No teenagers (TeenHouseholds = 0)

Budget Conscious Youth: This group comprises younger, low-income customers with minimal spending and online shopping activity. They are highly price-sensitive and unlikely to engage in high-value purchases.

Bivariate Analysis (Figure 3.3):

1. ANOVA:

- **Age:** $p < 0.001$ (significant)
- **Income:** $p < 0.001$ (significant)
- **Spend on Products:** $p < 0.001$ (significant)
- **Online Purchases**: $p < 0.001$ (significant)

The objective of the ANOVA test was to check for significant differences in **age, income, spending on products**, and **online purchases** across the segments. These results confirm that the customer segments differ significantly across these variables.

2. Chi-square Test:
   The objective of the chi-square test was to establish the relationship between **TeenHouseholds** and customer segments. The result indicates that teenagers in the household are a significant factor in differentiating the customer segments, particularly in **Cluster 1**(Family-focused spenders). (Figure 3.4)

SEGMENT-SPECIFIC STRATEGIES

1 **Family-focused Spenders (47.9%) (majority):**

- **Strategy:**
  - **Bundle Discounts** on Family-Oriented Products: This segment is family-oriented, with many customers having teenagers in their households. By offering discounts or bundle deals on family

essentials (e.g., groceries, household products, and family meal packs), Marketplace can encourage higher sales per transaction.

ii **Loyalty points**: As family-focused customers with moderate spending habits, they are likely to value rewards. A loyalty program offering discounts for repeat purchases or points for spending could encourage them to shop more frequently.

iii **Family advertising**: Use targeted advertising to promote family products and services, highlighting convenience and savings for family households. For example, ads on social media platforms and email campaigns can focus on parenting and family needs.

- **Key Justifications:**
  i **Family Orientation:** With a significant presence of teenagers in the household, the Chi-Square test highlighted the relationship between household type and spending behavior, confirming the importance of targeting family-focused products for the **Family-Focused Spenders** group. This group will likely be responsive to promotions catering to family needs (Holbrook & Hirschman, 1982).
  ii **Moderate Spending:** They tend to spend moderately, so bundle offers or loyalty incentives will appeal to their price-conscious behavior while encouraging increased shopping frequency (Chandon, Wansink, & Laurent, 2000).

2 **Wealthy Professionals (26.1%) (Highly profitable):**

- **Strategy:**
  i **Premium products:** This segment, with a high-income level and frequent online shopping habits, will likely be attracted to premium offerings. Marketplace should introduce exclusive, high-end products, such as organic food, gourmet goods, or luxury home appliances.
  ii **Personalized shopping experience:** Premium customers expect convenience and a luxury shopping experience, so features like customized recommendations on the website or a VIP customer service hotline can improve satisfaction.
- **Key Justifications:**

i. **High income and spending:** With a higher income and more significant spending potential, this group is more likely to respond positively to premium products and personalized offers. Research suggests that higher-income customers are willing to pay a premium for convenience and exclusivity (Keller, 2003).
ii. **Online shopping:** Frequent online shopping habits make this group more responsive to digital engagement, such as personalized emails or exclusive online deals, aligning with trends in e-commerce personalization (Grewal, Iyer, & Sharma, 2003).

3. **Budget-conscious youth (26%) (Price sensitive):**

- **Strategy:**
  i. **Time-sensitive discounts and flash sales:** The Budget-Conscious Youth segment, with low income and minimal spending habits, will likely respond to limited-time offers and flash sales. Marketplaces should create special promotions, such as “one-day flash sales” or "weekend discount codes," to entice budget-conscious customers.
  ii. **Affordable product lines:** Focus on low-cost, high-value items such as basic groceries, household essentials, or budget-friendly meal kits.
  iii. **Social media and influencer marketing:** Use social media platforms like Instagram and TikTok to engage with this segment. Collaborate with influencers or student ambassadors to promote affordable products, using catchy hashtags and student discounts.
- **Key justification:**
  i. **Price sensitive:** Research indicates that younger, budget-conscious consumers respond exceptionally to immediate savings promotions (Zhou, Liao, & Palacios, 2015). Price sensitivity is critical for this segment, and value-based promotions will resonate with their spending habits.
  ii. **Low online purchases:** Since this group tends to make fewer online purchases, it is essential to use digital engagement methods that increase visibility and ease of access, such as offering promo codes or discounts directly through social media platforms they frequent (Mangold & Faulds, 2009).

Conclusion

The segmentation analysis has provided valuable insights into the distinct needs and behaviors of **Family-Focused Spenders**, **Wealthy Professionals**, and **Budget-Conscious Youth**. By leveraging cluster analysis, along with Chi-Square and ANOVA tests, we developed tailored strategies for each segment. **Family-focused spenders** will respond to bundle discounts and loyalty programs, while **Wealthy Professionals** will be attracted to premium, personalized offerings. **Budget-conscious youth** require time-sensitive discounts and affordable products. These strategies are grounded in data-driven insights, ensuring effective customer engagement and enhanced sales. For Marketplace, implementing these targeted approaches will strengthen customer loyalty, optimize marketing efforts, and drive business growth.

References

Chandon, P., Wansink, B., & Laurent, G. (2000). A benefit congruency framework of sales promotion effectiveness. Journal of Marketing, 64(4), 65–81. https://doi.org/10.1509/jmkg.64.4.65.18077

Grewal, D., Iyer, G., & Sharma, A. (2003). The Internet and the consumer decision process. Journal of Interactive Marketing, 17(3), 8–18. https://doi.org/10.1002/dir.10049

Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. Journal of Consumer Research, 9(2), 132-140. https://doi.org/10.1086/208906

Keller, K. L. (2003). Strategic brand management: Building, measuring, and managing brand equity (3rd ed.). Pearson Education.

Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. Business Horizons, 52(4), 357–365. https://doi.org/10.1016/j.bushor.2009.03.002

Zhou, L., Liao, Z., & Palacios, M. (2015). Customer satisfaction and customer loyalty in e-commerce: A comparative study of online shopping platforms. Journal of Electronic Commerce Research, 16(4), 232-250.

Appendix

Figure 1.1

The mean of the original data (51.17) is that the maximum age of the variable is 127 years, meaning there are outliers in the data.

Frequencies

Statistics

		Age	Income	SpendProducts
N	Valid	2150	2150	2150
	Missing	0	0	0
Mean		51.17	52054.79	231.02
Median		50.00	51381.50	91.00
Mode		44	7500	10
Std. Deviation		11.994	21506.956	284.564
Minimum		24	1730	1
Maximum		127	162397	1727

Figure 1.2

The mean after cleaning the data (51.07) is that the maximum age of the variable is 80 years, meaning there are no outliers in the dataset.

Frequencies

Statistics

		Age	Income	SpendProducts
N	Valid	2147	2147	2147
	Missing	0	0	0
Mean		51.07	52043.53	230.95
Median		50.00	51373.00	91.00
Mode		44	7500	10
Std. Deviation		11.700	21507.963	284.413
Minimum		24	1730	1
Maximum		80	162397	1727

Figure 2

Transforming the nominal measure of the variable "NWebPurchases" into a scale measure is necessary because it represents categorical data. The high values of "NWebPurchases" can significantly dominate the analysis. Consequently, we

recoded it into a different measure to mitigate this effect.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ID	Numeric	8	0		None	None	8	Right	Scale	Input
2	Age	Numeric	8	0		None	None	8	Right	Scale	Input
3	Education	String	8	0		None	None	10	Left	Nominal	Input
4	MaritalStatus	String	8	0		None	None	13	Left	Nominal	Input
5	Income	Numeric	8	0		None	None	8	Right	Scale	Input
6	KidsHouse...	Numeric	8	0		None	None	14	Right	Nominal	Input
7	TeensHou...	Numeric	8	0		None	None	19	Right	Nominal	Input
8	NWebPurc...	Numeric	8	0		None	None	14	Right	Scale	Input
9	NStorePurc...	Numeric	8	0		None	None	20	Right	Ordinal	Input
10	AcceptedC...	Numeric	8	0		None	None	18	Right	Nominal	Input
11	Complaint	Numeric	8	0		None	None	10	Right	Nominal	Input
12	SpendProd...	Numeric	8	0		None	None	16	Right	Scale	Input
13	TSC_4618	Numeric	10	0	TwoStep Clust...	{-1, Outlie...	None	11	Right	Nominal	Input
14											

Figure 3.1

Cluster view of the cluster analysis which shows the mean of each variable. Suitable for comparison

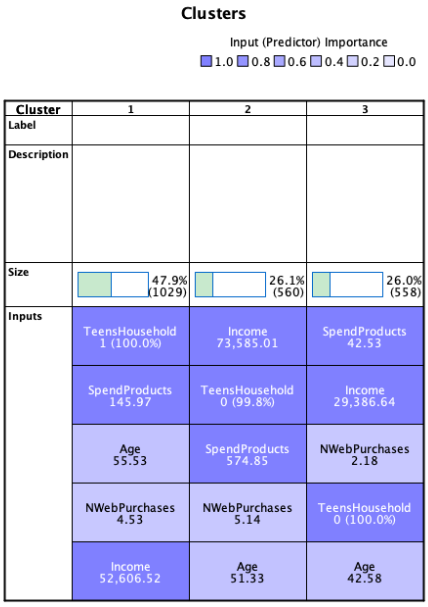

Figure 3.2

Model Summary of the Cluster Analysis used to interpret the Cluster quality, inputs, and the number of clusters formed.

Model Summary

Algorithm	TwoStep
Inputs	5
Clusters	3

Cluster Quality

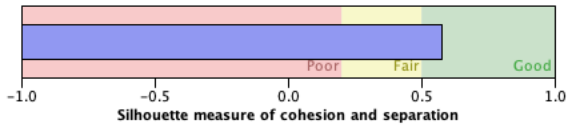

Figure 3.3

Anova test shows that the results are significant for age, income, spendproducts, and NWebPurchases

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Age	Between Groups	60795.416	2	30397.708	279.773	<.001
	Within Groups	232948.382	2144	108.651		
	Total	293743.798	2146			
Income	Between Groups	5.466E+11	2	2.733E+11	1313.581	<.001
	Within Groups	4.461E+11	2144	208067413		
	Total	9.927E+11	2146			
NWebPurchases	Between Groups	2861.560	2	1430.780	231.949	<.001
	Within Groups	13225.294	2144	6.169		
	Total	16086.854	2146			
SpendProducts	Between Groups	93472133.9	2	46736067.0	1250.652	<.001
	Within Groups	80119925.0	2144	37369.368		
	Total	173592059	2146			

Figure 3.4

Chi-Square Test shows that the results are significant as the p-value is less than 0.05 for TeenHouseholds

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2142.001[a]	2	<.001
Likelihood Ratio	2956.886	2	<.001
Linear-by-Linear Association	1739.791	1	<.001
N of Valid Cases	2146		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 267.34.