CUSTOMER SEGMENTATION AND STRATEGIES TO DRIVE SUCCESS AT MARKETPLACE

BUSINFO700

Assignment 2

Saakshi Agarwal

195297642

Executive Summary

This report examines the customer base of Marketplace, a supermarket, to identify distinct customer segments and develop targeted marketing strategies to increase customer satisfaction, engagement, and sales. The objective is to understand Marketplace's customers' diverse behaviors and preferences through statistical analysis, including univariate, bivariate, and cluster analysis. The goal is to provide actionable recommendations that align with the needs of each segment, thereby enhancing the supermarket's marketing efforts and overall customer experience.

After assessing the key variables - Age, Income, NWebPurchases, NStorePurchase, KidHouseHold, TeenHousehold, and SpentProduct they were able to identify Value Seeking Young Families, Premium Solo Shoppers, Stable Tech Families with Teens, and Growing Families with Mixed-Age Kids as the four dominant segments. Since the different segments display different purchasing patterns, appropriate targeting strategies are recommended, such as offering digital promotions to online shoppers and loyalty rewards to in-store customers. This, in turn, helps Marketplace enhance its marketing initiatives, increase engagement, and boost customer retention. The segmentation advantages are sound as they offer the Marketplace a strategy to partner with the customers in their various and unique needs and preferences. This customer-centric approach will allow the supermarket to effectively meet the diverse needs of its customer base, ensuring long-term growth and success.

Introduction

The report's main objective is to segment the supermarket's customers into different groups to gain a better understanding of their behaviors, sales patterns, preferences, and needs. Market segmentation is defined as a process of splitting a wide range of customers into groups with similar patterns based on common underlying ideas that the company can utilize to improve its sales and customer retention strategy. This segmentation can provide a clear picture of the company and help the supermarket focus on improving sales, customer satisfaction, and customer retention by understanding and meeting the desires of each customer group/segment of the market.

In this analysis, customers can be divided into segments based on age, income, marital status, family composition, education level, etc. The main goal is to ensure that there is a strategy for each customer group. These objectives will lead the supermarket to improve its marketing strategy, managing each customer group based on their preference, sales, customer loyalty, and satisfaction, thus helping the supermarket to expand its business overall.

Methodology

The statistical tools used to accomplish the aim of the analysis include univariate analysis, bivariate analysis, and cluster analysis. These tools will facilitate segmentation and analysis of relationships between variables. The univariate analysis summarizes the data and discovers patterns within the variables. Techniques like mean and outliers, as well as visualization techniques like histogram and boxplot, are used. These data provided insight into data distribution and helped rectify the outliers.

Bivariate analysis is used to analyze the relationship between two variables and gain insight into the effect of one variable on another. Bivariate analysis techniques used are:

1. Chi-Square: This test is used to identify the relationship between two non-numeric data.
2. T-test: This test compares the average of three or more groups against one numeric variable to assess if the difference is statistically significant.
3. ANOVA: This test compares the average of two groups against one numeric variable to determine if at least one differs significantly.
4. Correlation: This test measures how strongly two numeric variables are related.

Cluster analysis is the most important technique used for this report. It is an unsupervised machine-learning technique that helps classify data into distinct groups based on the input variables. This technique will group customers without predefined categories, making it suitable for discovering new patterns within complex datasets.

The variables used for cluster analysis were taken in relevance to customer behaviors and meaningful groups:

1. Age: It was used to understand shopping trends across different age groups.
2. Income: This variable will help segregate groups based on the spending capacity of the customers.
3. Composition of family: This contains two variables KidHousehold, and TeenHousehold. This will show shopping trends for different family compositions, mainly for products specifically for kids or teens.
4. Purchasing pattern: It has two variables NStorePurchase and NWebPurchase. This will help us understand if customers prefer to shop online or at the store, which would help the company learn key insights and design new strategies.
5. SpendProducts: This would help assess the overall money customers spend, and as SpendProduct is last month's data, it would give insight into the currently popular products.

Using these variables, a better understanding of customer shopping preferences and habits can be made. These variables give meaningful segments that can be useful for developing marketing strategies for each segment.

Analysis Result

Univariate analysis for Age, Income, NWebPurchase, NStorePurchase, and SpendProduct reveals critical insights into the demographic and behavioral patterns of Marketplace's customer base. As shown in Figure 1.1, among 2150 customers, the average age of customers is 51.1, with an average income of 52054 dollars. These customers have an average web and store purchase of 4.08 and 5.81, respectively, with an average spending of 231 dollars. Most of these customers are married and are studying for a bachelor's degree, as shown in Figures 1.4 and 1.5, respectively. As shown in Figure 1.2, the boxplot shows outliers for variable age, which was replaced by the mean age. The presence of kids and teens has a mean of 0.42 and 0.48, respectively, showing that 42% of customers have young children and 48% of customers have teens at home. Customers with kids may have specific product needs and shopping preferences, implying that there is a likelihood of specific targeting of family-oriented campaigns. Customers with teenagers could expect enhanced service and low cost, which can be achieved through product development strategy or in-store merchandising targeted at teenagers.

From Figure 1.3, it can be observed that only 21% of the customers are willing to accept marketing campaigns. This emphasizes the necessity of forming more specific campaigns to achieve higher participation of all segments of the customers. Finally, with a mean of 0.01, the complaint variable indicates a very low complaint incidence rate, meaning the customers are generally satisfied.

The T-test between AcceptedCampaign and SpendProducts reveals a significant association between campaign acceptance and customer spending behavior, and the groups are significantly different. The customer who has accepted the campaign tends to spend more than the customer who does not, as shown in Figure 2.1. This emphasizes launching different campaigns for different sets of people to attract a larger customer base.

The Correlation test between SpendProduct and Income is also moderately correlated and highly significant, which shows that as income increases, spendproduct increases, as shown in Figures 2.2 and 2.3. It shows that as a customer's income increases, they tend to spend more on products, implying that there is an opportunity for targeting elite customers by giving them new schemes and offers.

A two-step cluster analysis of the supermarket identified four different clusters, with a silhouette score of 0.6, as shown in Appendix C. These segments were a unique combination of income level, customer demographic, and behaviors. It also explains how different family composition and income affect their spending pattern. Active profiling was done to interpret and label these clusters based on their key traits, as shown in Appendix D. Active profiling in cluster analysis refers to analyzing and describing the characteristics of each cluster to understand its key features.

Detailed descriptions of each cluster with their characteristics are below: -

1. Value-seeking Young Families (Cluster 1): This cluster deals with families, with almost 81.9 % of them having younger kids and having limited income, with an average of $32,300. They have the second-lowest monthly expense ($70.38) and prefer least store and web purchases, with an average of 3.58 and 2.59, respectively. It can be concluded that they prioritize essential purchases for their families, which the label also suggests.
2. Premium Solo Shoppers (Cluster 2): This cluster has middle-aged customers with high incomes ($74,641.05) and no kids. Their spending behavior ($596.34) is higher than that of any other group, and they prefer buying products from stores (8.15). This group likely values premium products and experiences, and the name underscores their strong purchasing power and keen interest in indulging in diverse or high-end products, as the name also suggests.
3. Stable tech families with teens (Cluster 3): This cluster has families with teenagers at home. They have a moderate income ($58,075.96), are highly active in online shopping (5.52), and maintain regular in-store visits (7.21). Their overall expenditure is moderate ($201.52), balancing convenience and variety. Overall, this cluster reflects families with a balanced life and an inclination towards online shopping, which could result from having a teenager at home, which the name also suggests.
4. Growing families with mixed-age kids (Cluster 4): As the name suggests, this cluster comprises families having mixed-age kids with moderate income ($44,429.57). They have the lowest monthly expense ($63.3) and prefer shopping at the store to the web. This can be because they seek more valuable offers and product evaluation for cost-cutting.

Passive analysis is used to understand clusters' characteristics against the variables not used in the clustering process. This process validates the cluster but does not directly influence cluster formation. To do the same, passive analysis was done for variables AcceptedCampaigns and Complaints. A chi-Square test was done to achieve this, as shown in Appendix E, figure 5.1. It can be concluded from the figure that Cluster 2, which is Premium Solo Shopper, has significantly accepted more campaigns than others. It can also be observed that Premium shoppers have more monthly expenditure, so it can be concluded that people who have accepted campaigns tend to spend more, which means SpendProduct and Campaign are related. This was also seen in our bivariate analysis for the same variables; refer to figure 2.1. From figure 5.2, we can see that customers across any group have minimum complaints, meaning most customers are satisfied and happy.

Segment Specific strategy

The strategies recommended cater to each customer segment's specific preferences and needs, ensuring that marketing efforts are highly targeted and effective. By implementing these detailed approaches, the supermarket can increase customer satisfaction, improve sales, and stand out among competitors.

1. Value-seeking Young Families:

   Strategies:

   a) Customized Promotions: Launching “family value card,” introducing deals and discounts on family essential products and offering cashback on those products. This will help to address their limited income.
   b) Comfort-Oriented Services: Promotion of family meals that are ready to eat and introducing “Shopping Hours for family” with activities for children will enhance customer comfort. This can encourage families to visit the stores more often.
   c) Digital Engagement: Creating a new section for family shopping and recipe ideas with a meal planner will engage the customer. Personalized email with holiday or school season offers will be a good strategy for digital engagement.

   These strategies are developed to maintain affordability, engage customers on web purchases, and ensure convenience. It also directly addresses the priorities of value-seeking families.

2. Premium Solo Shoppers:

   Strategies:

   a) Premium Product Line: Increase in the availability of premium products. Personalize their recommendation based on past purchases.
   b) Exclusive Experience: Organize VIP events and provide personal assistance and lounge for a premium experience to the customer.
   c) Loyalty & Rewards: Introducing premium membership cards, early access to premium products and sales will attract more premium customers.

   These strategies are developed to maintain exclusivity and quality and attract customers to spend more by launching new campaigns, as it was concluded before that this section of people spend more based on campaigns.

3. Stable tech Families with Teens:

   Strategies:

   a) Promotion related to Health and Wellness: Offer discounts on healthy and organic foods. Tie-ups with local healthcare providers could be beneficial.
   b) Teen-centric Marketing: Offer personalized subscriptions for frequently purchased items. Digital campaigns promoting exclusive deals with pushed ads will attract customers' attention as this segment's customers often shop online. Focus on promoting

snacks, beverages, and entertainment-related products popular with teenagers through in-store displays and digital campaigns.

c) Balanced Shopping Channels: Develop a hybrid shopping program offering both in-store and online perks, such as discounts on repeat online purchases and exclusive in-store deals. Provide digital tutorials for older customers on how to shop online, encouraging tech adoption.

These strategies balance health, family needs, and digital engagement, appealing to older customers with teens who value both traditional and modern shopping experiences.

4. Growing families with Mixed-Age Kids:

Strategies:

a) Diverse Product Promotion: Focus on promoting family deals with teen and kid products. Combo-deals for different products can also be beneficial. Having discounts for the family during weekends and holidays can be an added benefit.

b) In-Store Enhancement: Design family-friendly aisles with easy-to-navigate layouts and product categories suited to children and teens. Offer free samples or discounts on newly launched family-oriented products.

c) Engagement through events: Host family activity days with events like cooking competitions, parenting workshops, and fun games for kids and teens. Collaborate with local cinemas or amusement parks to provide family discounts with purchase thresholds.

These strategies are developed to meet the diverse needs of families with children of different ages, fostering loyalty, and increasing basket size.

Conclusion

The two-way cluster analysis successfully segmented customers into four distinct groups: Value Seeking Young Families, Premium Solo Shoppers, Stable Tech Families with Teens, and Growing Families with Mixed-Age Kids. Customer segmentation and strategies for each segment serve as a roadmap for the supermarket to strengthen its position in the market. The segmentation process has highlighted actionable opportunities and laid the groundwork for a data-driven approach to future business decisions.

Value-seeking young customers prioritize affordability and convenience, requiring promotions on essentials and tailored family services. Premium Solo Shoppers value premium offerings and exclusivity, demanding luxury products and personalized experiences. Stable tech families with teens show balanced purchasing behaviors, highlighting opportunities in health and family-focused promotions. Growing families with mixed-age kids seek variety and value, benefiting from combined deals and engaging in-store experiences.

By analyzing these segments, the company can work towards addressing specific segment needs to ensure a more personalized and enjoyable shopping experience. Targeted promotions and optimized product offerings will encourage greater spending. Engaging customers with loyalty programs and community-centric campaigns fosters long-term relationships. Leveraging segmentation insights enables the supermarket to stay ahead of the competition.

Reference

- Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. Annals of Biomedicine Engineering, 51, 2629–2633.
- 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 A:

This appendix contains all the univariant analysis evidence and graphs.

Statistics

		Age	Income	NWebPurchases	NStorePurchases	SpendProducts
N	Valid	2150	2150	2150	2150	2150
	Missing	0	0	0	0	0
Mean		51.17	52054.79	4.08	5.81	231.02
Minimum		24	1730	0	0	1
Maximum		127	162397	27	13	1727

Figure 1.1: Univariate analysis for metric variables

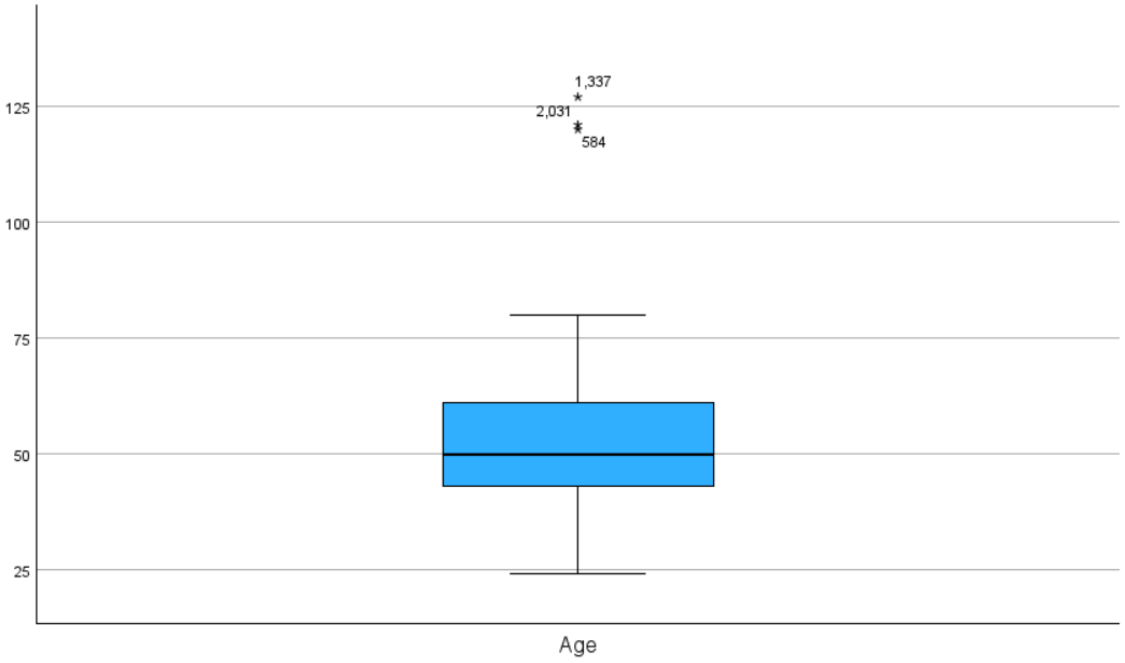

Figure 1.2 Boxplot for Age outliers

Statistics

		KidsHousehold	TeensHousehold	AcceptedCampaign	Complaint
N	Valid	2150	2150	2150	2150
	Missing	0	0	0	0
Mean		.42	.48	.21	.01
Median		.00	.00	.00	.00
Mode		0	0	0	0

Figure1.3 Univariate analysis for non-metric data

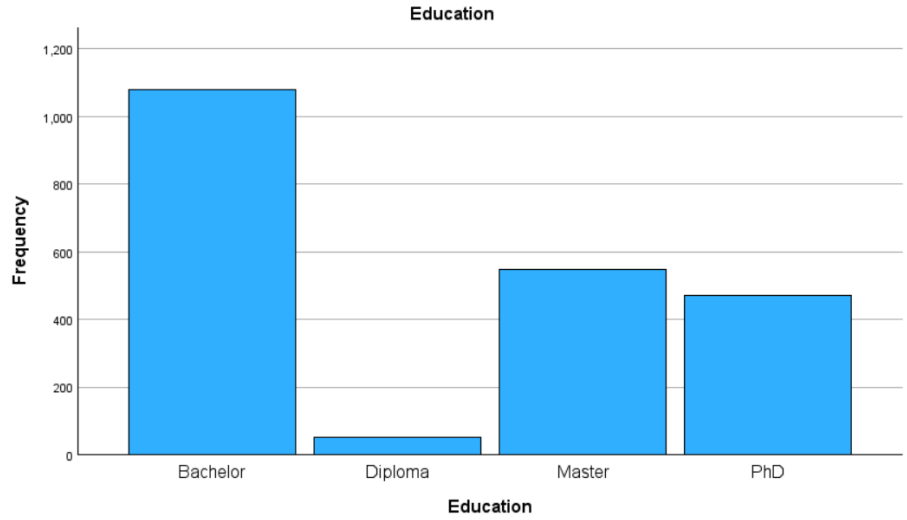

Figure 1.4 Bar Graph for Education

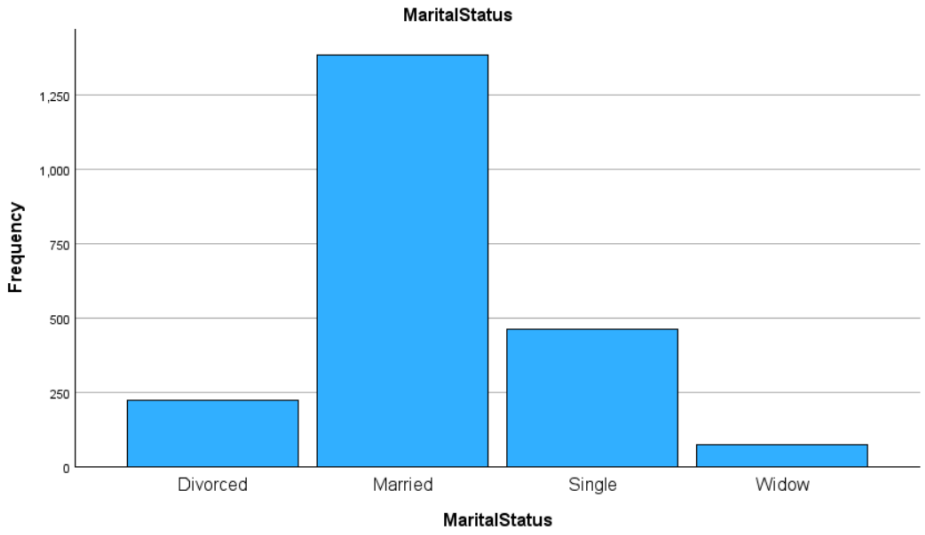

Figure 1.5 Bar Graph for Marital Status

Appendix B:

This appendix contains all the Bivariate analysis evidence and graphs

T-Test

Group Statistics

	AcceptedCampaign	N	Mean	Std. Deviation	Std. Error Mean
SpendProducts	0	1708	192.30	254.344	6.154
	1	442	380.66	340.514	16.197

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean	Std. Error	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Difference	Difference	Lower	Upper
SpendProducts	Equal variances assumed	119.482	<.001	-12.870	2148	<.001	<.001	-188.363	14.636	-217.065	-159.661
	Equal variances not assumed			-10.871	574.443	<.001	<.001	-188.363	17.326	-222.394	-154.332

Figure 2.1 T-test for SpendProduct and AcceptedCampaign

Correlations

Correlations

		Income	SpendProducts
Income	Pearson Correlation	1	.714**
	Sig. (2-tailed)		<.001
	N	2150	2150
SpendProducts	Pearson Correlation	.714**	1
	Sig. (2-tailed)	<.001	
	N	2150	2150

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 2.2 Correlation test for SpendProduct and Income

Graph

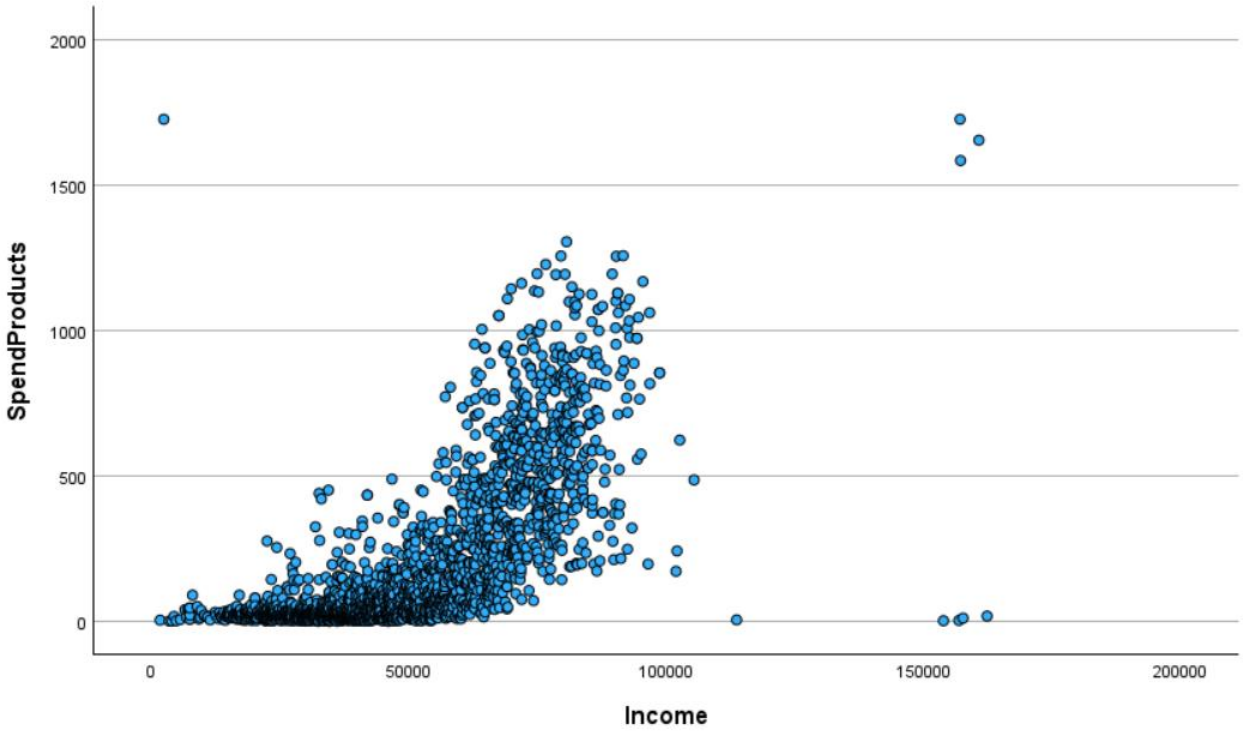

Figure 2.3 Scatter plot for spendproduct and Income

Appendix C:

This appendix contains all the Two-Step cluster analysis evidence

Model Summary

Algorithm	TwoStep
Inputs	7
Clusters	4

Cluster Quality

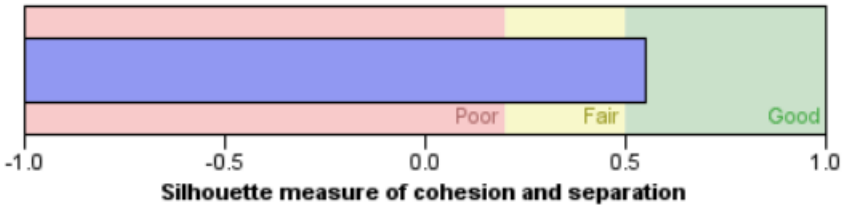

Figure 3.1 Silhouette score for two-step cluster

Cluster	3	1	2	4
Label				
Description				
Size	29.1% (626)	28.5% (613)	23.5% (506)	18.8% (405)
Inputs	KidsHousehold 0 (100.0%)	KidsHousehold 1 (81.9%)	KidsHousehold 0 (100.0%)	KidsHousehold 1 (100.0%)
	TeensHousehold 1 (100.0%)	TeensHousehold 0 (100.0%)	TeensHousehold 0 (100.0%)	TeensHousehold 1 (100.0%)
	Age 56.25	Age 42.59	Age 52.55	Age 54.57
	NWebPurchases 5.52	NWebPurchases 2.59	NWebPurchases 4.96	NWebPurchases 2.99
	NStorePurchases 7.21	NStorePurchases 3.58	NStorePurchases 8.15	NStorePurchases 4.09
	Income 58,075.96	Income 32,300.01	Income 74,641.05	Income 44,429.57
	SpendProducts 201.52	SpendProducts 70.38	SpendProducts 596.34	SpendProducts 63.33

Figure 3.2 Cluster information

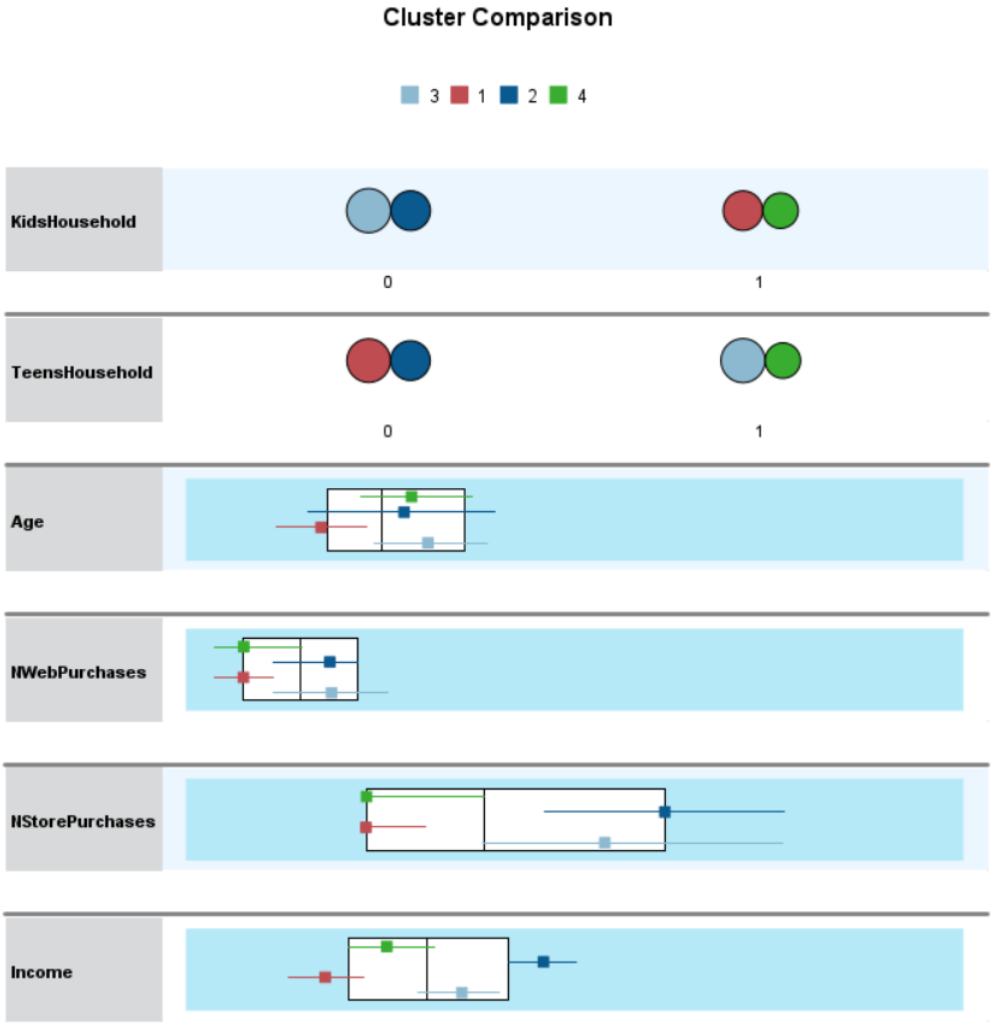

Figure 3.3 Comparison

Appendix D:

This appendix contains all the Active profiling cluster analysis evidence

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Income	1	613	32300.01	14803.951	597.926	31125.78	33474.25	1730	157733
	2	506	74641.05	14783.739	657.217	73349.83	75932.27	22507	160803
	3	626	58075.96	13515.782	540.199	57015.14	59136.79	4428	157243
	4	405	44429.57	15070.063	748.837	42957.47	45901.68	4023	162397
	Total	2150	52054.79	21506.956	463.831	51145.19	52964.40	1730	162397
SpendProducts	1	613	70.38	123.026	4.969	60.63	80.14	1	1727
	2	506	596.34	292.146	12.987	570.83	621.86	3	1727
	3	626	201.52	189.278	7.565	186.66	216.37	2	1585
	4	405	63.33	99.274	4.933	53.64	73.03	1	755
	Total	2150	231.02	284.564	6.137	218.99	243.06	1	1727
Age	1	613	42.59	8.972	.362	41.88	43.30	24	120
	2	506	52.55	14.566	.648	51.27	53.82	25	121
	3	626	56.25	9.363	.374	55.52	56.99	38	127
	4	405	54.57	8.758	.435	53.72	55.43	39	74
	Total	2150	51.17	11.994	.259	50.66	51.68	24	127
NWebPurchases	1	613	2.59	1.947	.079	2.43	2.74	0	11
	2	506	4.96	2.338	.104	4.76	5.17	0	27
	3	626	5.52	2.907	.116	5.29	5.75	0	25
	4	405	2.99	2.328	.116	2.76	3.21	0	11
	Total	2150	4.08	2.737	.059	3.96	4.19	0	27
NStorePurchases	1	613	3.58	1.988	.080	3.42	3.74	0	13
	2	506	8.15	2.887	.128	7.90	8.40	0	13
	3	626	7.21	3.091	.124	6.96	7.45	0	13
	4	405	4.09	2.154	.107	3.88	4.30	0	12
	Total	2150	5.81	3.251	.070	5.67	5.95	0	13

Figure 4.1 ANOVA test: Descriptive values for all the metric variables used for clustering

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Income	Based on Mean	.578	3	2146	.630
	Based on Median	.529	3	2146	.662
	Based on Median and with adjusted df	.529	3	2097.702	.662
	Based on trimmed mean	.527	3	2146	.664
SpendProducts	Based on Mean	232.445	3	2146	<.001
	Based on Median	221.513	3	2146	<.001
	Based on Median and with adjusted df	221.513	3	1859.286	<.001
	Based on trimmed mean	238.081	3	2146	<.001
Age	Based on Mean	97.499	3	2146	<.001
	Based on Median	97.103	3	2146	<.001
	Based on Median and with adjusted df	97.103	3	1894.472	<.001
	Based on trimmed mean	97.551	3	2146	<.001
NWebPurchases	Based on Mean	31.643	3	2146	<.001
	Based on Median	33.380	3	2146	<.001
	Based on Median and with adjusted df	33.380	3	2077.289	<.001
	Based on trimmed mean	33.246	3	2146	<.001
NStorePurchases	Based on Mean	106.762	3	2146	<.001
	Based on Median	108.753	3	2146	<.001
	Based on Median and with adjusted df	108.753	3	2095.219	<.001
	Based on trimmed mean	114.576	3	2146	<.001

Figure 4.2 ANOVA test: Homogeneity test values for all the metric variables used for clustering

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Income	Between Groups	5.436E+11	3	1.812E+11	863.313	<.001
	Within Groups	4.504E+11	2146	209888263.92		
	Total	9.940E+11	2149			
SpendProducts	Between Groups	95282133.401	3	31760711.134	865.645	<.001
	Within Groups	78737231.615	2146	36690.229		
	Total	174019365.02	2149			
Age	Between Groups	66948.316	3	22316.105	197.746	<.001
	Within Groups	242181.396	2146	112.852		
	Total	309129.712	2149			
NWebPurchases	Between Groups	3547.677	3	1182.559	202.165	<.001
	Within Groups	12552.965	2146	5.849		
	Total	16100.642	2149			
NStorePurchases	Between Groups	8242.934	3	2747.645	407.418	<.001
	Within Groups	14472.732	2146	6.744		
	Total	22715.666	2149			

Figure 4.3 ANOVA test: Anova test values for all the metric variables used for clustering

Homogeneous Subsets

Income

			Subset for alpha = 0.05			
	TwoStep Cluster Number	N	1	2	3	4
Tukey B[a,b]	1	613	32300.01			
	4	405		44429.57		
	3	626			58075.96	
	2	506				74641.05

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 521.228.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Figure 4.4 Tukeys B PostHoc test for Income

Multiple Comparisons

Dependent Variable		(I) TwoStep Cluster Number	(J) TwoStep Cluster Number	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Income	Tamhane	1	2	-42341.035*	888.510	<.001	-44683.04	-39999.03
			3	-25775.947*	805.811	<.001	-27899.51	-23652.39
			4	-12129.558*	958.266	<.001	-14656.65	-9602.47
		2	1	42341.035*	888.510	<.001	39999.03	44683.04
			3	16565.088*	850.735	<.001	14322.48	18807.69
			4	30211.477*	996.339	<.001	27584.01	32838.94
		3	1	25775.947*	805.811	<.001	23652.39	27899.51
			2	-16565.088*	850.735	<.001	-18807.69	-14322.48
			4	13646.389*	923.349	<.001	11210.95	16081.82
		4	1	12129.558*	958.266	<.001	9602.47	14656.65
			2	-30211.477*	996.339	<.001	-32838.94	-27584.01
			3	-13646.389*	923.349	<.001	-16081.82	-11210.95
SpendProducts	Tamhane	1	2	-525.959*	13.906	<.001	-562.66	-489.26
			3	-131.133*	9.051	<.001	-154.99	-107.28
			4	7.052	7.002	.896	-11.41	25.51
		2	1	525.959*	13.906	<.001	489.26	562.66
			3	394.826*	15.030	<.001	355.19	434.47
			4	533.011*	13.893	<.001	496.35	569.68
		3	1	131.133*	9.051	<.001	107.28	154.99
			2	-394.826*	15.030	<.001	-434.47	-355.19
			4	138.184*	9.031	<.001	114.37	161.99
		4	1	-7.052	7.002	.896	-25.51	11.41
			2	-533.011*	13.893	<.001	-569.68	-496.35
			3	-138.184*	9.031	<.001	-161.99	-114.37
Age	Tamhane	1	2	-9.955*	.742	<.001	-11.91	-8.00
			3	-13.663*	.521	<.001	-15.04	-12.29
			4	-11.982*	.566	<.001	-13.48	-10.49
		2	1	9.955*	.742	<.001	8.00	11.91
			3	-3.709*	.748	<.001	-5.68	-1.74
			4	-2.027	.780	.056	-4.08	.03
		3	1	13.663*	.521	<.001	12.29	15.04
			2	3.709*	.748	<.001	1.74	5.68
			4	1.681*	.574	.021	.17	3.19
		4	1	11.982*	.566	<.001	10.49	13.48
			2	2.027	.780	.056	-.03	4.08
			3	-1.681*	.574	.021	-3.19	-.17
NWebPurchases	Tamhane	1	2	-2.375*	.130	<.001	-2.72	-2.03
			3	-2.935*	.140	<.001	-3.30	-2.57
			4	-.398*	.140	.027	-.77	-.03
		2	1	2.375*	.130	<.001	2.03	2.72
			3	-.560*	.156	.002	-.97	-.15
			4	1.977*	.156	<.001	1.57	2.39
		3	1	2.935*	.140	<.001	2.57	3.30
			2	.560*	.156	.002	.15	.97
			4	2.537*	.164	<.001	2.10	2.97
		4	1	.398*	.140	.027	.03	.77
			2	-1.977*	.156	<.001	-2.39	-1.57
			3	-2.537*	.164	<.001	-2.97	-2.10
NStorePurchases	Tamhane	1	2	-4.573*	.151	<.001	-4.97	-4.17
			3	-3.627*	.147	<.001	-4.02	-3.24
			4	-.512*	.134	<.001	-.87	-.16
		2	1	4.573*	.151	<.001	4.17	4.97
			3	.946*	.178	<.001	.48	1.42
			4	4.061*	.167	<.001	3.62	4.50
		3	1	3.627*	.147	<.001	3.24	4.02
			2	-.946*	.178	<.001	-1.42	-.48
			4	3.115*	.163	<.001	2.68	3.55
		4	1	.512*	.134	<.001	.16	.87
			2	-4.061*	.167	<.001	-4.50	-3.62
			3	-3.115*	.163	<.001	-3.55	-2.68

*. The mean difference is significant at the 0.05 level.

Figure 4.5 Tamhane PostHoc test for all metric variables

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
KidsHousehold * TwoStep Cluster Number	2150	100.0%	0	0.0%	2150	100.0%
TeensHousehold * TwoStep Cluster Number	2150	100.0%	0	0.0%	2150	100.0%

KidsHousehold * TwoStep Cluster Number Crosstabulation

Count

		TwoStep Cluster Number				
		1	2	3	4	Total
KidsHousehold	0	111	506	626	0	1243
	1	502	0	0	405	907
Total		613	506	626	405	2150

TeensHousehold * TwoStep Cluster Number Crosstabulation

Count

		TwoStep Cluster Number				
		1	2	3	4	Total
TeensHousehold	0	613	506	0	0	1119
	1	0	0	626	405	1031
Total		613	506	626	405	2150

Figure 4.6 Chi-Square test for KidHousehold and TeenHousehold

Appendix E:

This appendix contains all the Passive profiling cluster analysis evidence and graphs

Crosstab

Count

		TwoStep Cluster Number				
		1	2	3	4	Total
AcceptedCampaign	0	443	365	395	503	1706
	1	57	41	222	121	441
Total		500	406	617	624	2147

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	143.345[a]	3	<.001
Likelihood Ratio	139.982	3	<.001
Linear-by-Linear Association	35.770	1	<.001
N of Valid Cases	2147		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 83.39.

Figure 5.1 Chi-Square test for AcceptedCampaign

Crosstab

Count

		TwoStep Cluster Number				
		1	2	3	4	Total
Complaint	0	494	400	613	620	2127
	1	6	6	4	4	20
Total		500	406	617	624	2147

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.810[a]	3	.422
Likelihood Ratio	2.713	3	.438
Linear-by-Linear Association	1.779	1	.182
N of Valid Cases	2147		

a. 2 cells (25.0%) have expected count less than 5. The minimum expected count is 3.78.

Figure 5.2 Chi-Square test for Complaints