VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF HANOI

CASE STUDY REPORT
Optimizing Customer Segmentation through Machine Learning algorithms

Nguyen Quang Huy - 22BI13195
Nguyen Tuan Khai - 22BI13262

HANOI, JUNE 26TH, 2024

I. Introduction

According to the needs and wants of customers in the retail industry, marketers employ several segmentation criteria and techniques to identify and understand customer clusters and provide desirable products and services. It is important as it fosters customer loyalty, leading to increase profits for businesses. However, marketers often deal with the challenge of effectively organizing marketing strategies caused by loyalty programs and promotions in conjunction with waste of marketing resources. (Musumali, 2019)

To address this issue, the objective is to create the cohesive groups within a diverse market, which can be gained by concentrating on marketing strategies and tactics. Over the years, a variety of models and techniques have been developed and used, ranging from simple statistical models to more advanced algorithms based on Artificial Intelligence (AI), and Machine Learning (ML). (Sarker, 2021)

Customer segmentation can be conducted according to various criteria. These follow demographics (such as age, sex, income, occupation, social class, stage of life), geographical place (including country, city, or rural), behavioral aspects (frequency of purchase, loyalty, where you buy, quantity purchased), purchase occasion (routine, special, hours or days of purchase, fixed place or while traveling), psychographic (lifestyle, personality, needs, values, attitudes, motivations), benefits (comfort, quality, economy, ease, speed), beliefs and attitudes (towards brands, products, purchase channels). (Camilleri, 2017)

Regarding the selection of clustering algorithms, K-means and Agglomerative Clustering (AC) may be preferred using in this experiment for customer segmentation due to performance. Therefore, in this experiment, K-means and AC methods are chosen to segment customer clusters for companies according to selected criteria during the experimental test to further marketing strategies explosion.

II. Objective

To employ unsupervised machine learning techniques and clustering algorithms in order to perform customer segmentation for a grocery company.

III. Data understanding

1. Analysis

The dataset in the year 2021 comprises 2240 entries and 29 attributes taken from https://www.kaggle.com/datasets/. It has been divided into four distinct subsets according to demographics, products information, geographical place, and beneficial promotion. However, it is also noting that both "Z_CostContact" and "Z_Revenue" show identical values across all rows leading to the conclusion to removing at a later stage caused by no supporting model.

	No	Column	Data Type	Details
Customer's information	1	ID	quantitative	Customer’s unique identifier
	2	Year_Birth	quantitative	The birth year of the respective person
	3	Education	qualitative	Education Qualification of customer
	4	Marital_Status	qualitative	Marital Status of customer
	5	Income	quantitative	Customer's yearly household income
	6	Kidhome	quantitative	Number of children in customer's household
	7	Teenhome	quantitative	Number of teenagers in customer's household
	8	Dt_Customer	qualitative	The number of days a customer is registered in the firm's database
	9	Complain	quantitative	1 if the customer complained in the last 2 years, 0 otherwise
	10	Recency	quantitative	Number of days since customer's last purchase
Products Amount spent on different products in last 2 years	11	MntWines	quantitative	Amount spent on wine in last 2 years
	12	MntFruits	quantitative	Amount spent on fruits in last 2 years
	13	MntMeatProducts	quantitative	Amount spent on Meat Products in last 2 years
	14	MntFishProducts	quantitative	Amount spent on Fish Products in last 2 years
	15	MntSweetProducts	quantitative	Amount spent on Sweet Products in last 2 years
	16	MntGoldProds	quantitative	Amount spend on Gold Products in last 2 years
Geographical place	17	NumWebPurchases	quantitative	Number of purchases made through the company’s website
	18	NumCatalogPurchases	quantitative	Number of purchases made using a catalogue
	19	NumStorePurchases	quantitative	Number of purchases made directly in stores
	20	NumWebVisitsMonth	quantitative	Number of visits to company’s website in the last month
Promotion	21	NumDealsPurchases	quantitative	Number of purchases made with a discount
	22	AcceptedCmp1	quantitative	1 if customer accepted the offer in the 1st campaign, 0 otherwise
	23	AcceptedCmp2	quantitative	1 if customer accepted the offer in the 2nd campaign, 0 otherwise
	24	AcceptedCmp3	quantitative	1 if customer accepted the offer in the 3rd campaign, 0 otherwise
	25	AcceptedCmp4	quantitative	1 if customer accepted the offer in the 4th campaign, 0 otherwise
	26	AcceptedCmp5	quantitative	1 if customer accepted the offer in the 5th campaign, 0 otherwise
	27	Z_CostContact	quantitative	Unknown
	28	Z_Revenue	quantitative	Unknown
	29	Response	quantitative	1 if customer accepted the offer in the last campaign, 0 otherwise

Table 1: Dataset details.

2. Data preparation

Based on Table 1, to enable the later stage for modeling, some new features include:

- The columns 'z_costcontact' and 'z_revenue' have a standard deviation of zero leading to their removal.
- The only missing values in the "Income" column are removed.
- A new 'Age' column has been created by calculating the difference between the current year and the birth year provided in the 'Year_Birth' column.
- A new 'spent' column represents the total spendings on various items, including 'MntWines,' 'MntFruits,' 'MntMeatProducts,' and 'MntGoldProds.
- A new "Living_with" column includes the data of "Marital_Status" as either having "Partner" or "Alone".
- A new "Children" column includes total counts of "Kidhome" and "Teenhome".
- A new "Family_Size" column includes total data of "Living_With" and "Children".
- A new "Is_Parent" column means the number of "Children" is more than 0.
- A new "Education" is divided into three levels

Old name	**New name**	**Level**
Basic	Undergraduate	Undergraduate
2n Cycle	Undergraduate	
Graduation	Graduate	Graduate
Master	Postgraduate	Postgraduate
PhD	Postgraduate	

- Some features are renamed:

Old name	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
New name	Wines	Fruits	Meat	Fish	Sweets	Gold

- A new "Customer_for" is the number of days the customers started to shop in the store relative to the last recorded date according to "Dt_Customer"

	No	Column	Data Type	Details
Customer's information	1	Spent	Quantitative	Total spending for groceries over 2 years
	2	Age	Quantitative	Age of customer in 2021
	3	Education	Qualitative	Education Qualification of customers
	4	Living_With	Qualitative	Living situation of couples
	5	Income	Quantitative	Customer's yearly household income
	6	Kidhome	Quantitative	Number of children in customer's household
	7	Teenhome	Quantitative	Number of teenagers in customer's household
	8	Children	Quantitative	Total children and teenagers in a household
	9	Family_Size	Quantitative	Total number of people living in the house
	10	Is_Parent	Qualitative	Parenthood status
	11	Customer_For	Quantitative	The number of days the customers started to shop in the store relative to the last recorded date
	12	Recency	Quantitative	Number of days since customer's last purchase
Products Amount spent on different products in last 2 years	13	Wines	Quantitative	Amount spent on wine in last 2 years
	14	Fruits	Quantitative	Amount spent on fruits in last 2 years
	15	Meat	Quantitative	Amount spent on Meat Products in last 2 years
	16	Fish	Quantitative	Amount spent on Fish Products in last 2 years
	17	Sweet	Quantitative	Amount spent on Sweet Products in last 2 years
	18	Gold	Quantitative	Amount spend on Gold Products in last 2 years
Geographical place	19	NumWebPurchases	Quantitative	Number of purchases made through the company's website
	20	NumCatalogPurchases	Quantitative	Number of purchases made using a catalogue
	21	NumStorePurchases	Quantitative	Number of purchases made directly in stores
	22	NumWebVisitsMonth	Quantitative	Number of visits to company's website in the last month
Promotion	23	NumDealsPurchases	Quantitative	Number of purchases made with a discount
	24	Response	Quantitative	1 if customer accepted the offer in the last campaign, 0 otherwise

Table 2: New features used in experiments

Table 2 and figure 1 give information about the new categories used in experiments. Overall, similar group of features, such as customer information or Product amount spent over the last two years, do not exhibit a strong correlation with each other, while some features belonging to different groups do show notable relationships (more than 0.5). For example, customer income is considerably associated with their spending on wines, meat products, and the overall spending “spent”.

In contrast, a series of the acceptance of various offers (from AcceptedCmp1 to AcceptedCmp5) from customers, Complain, and Response appear to be relatively independent of the other categories. Therefore, those are considered removing less correlated features to enhance dataset stability leading to no existence of acceptance offers in figure 1.

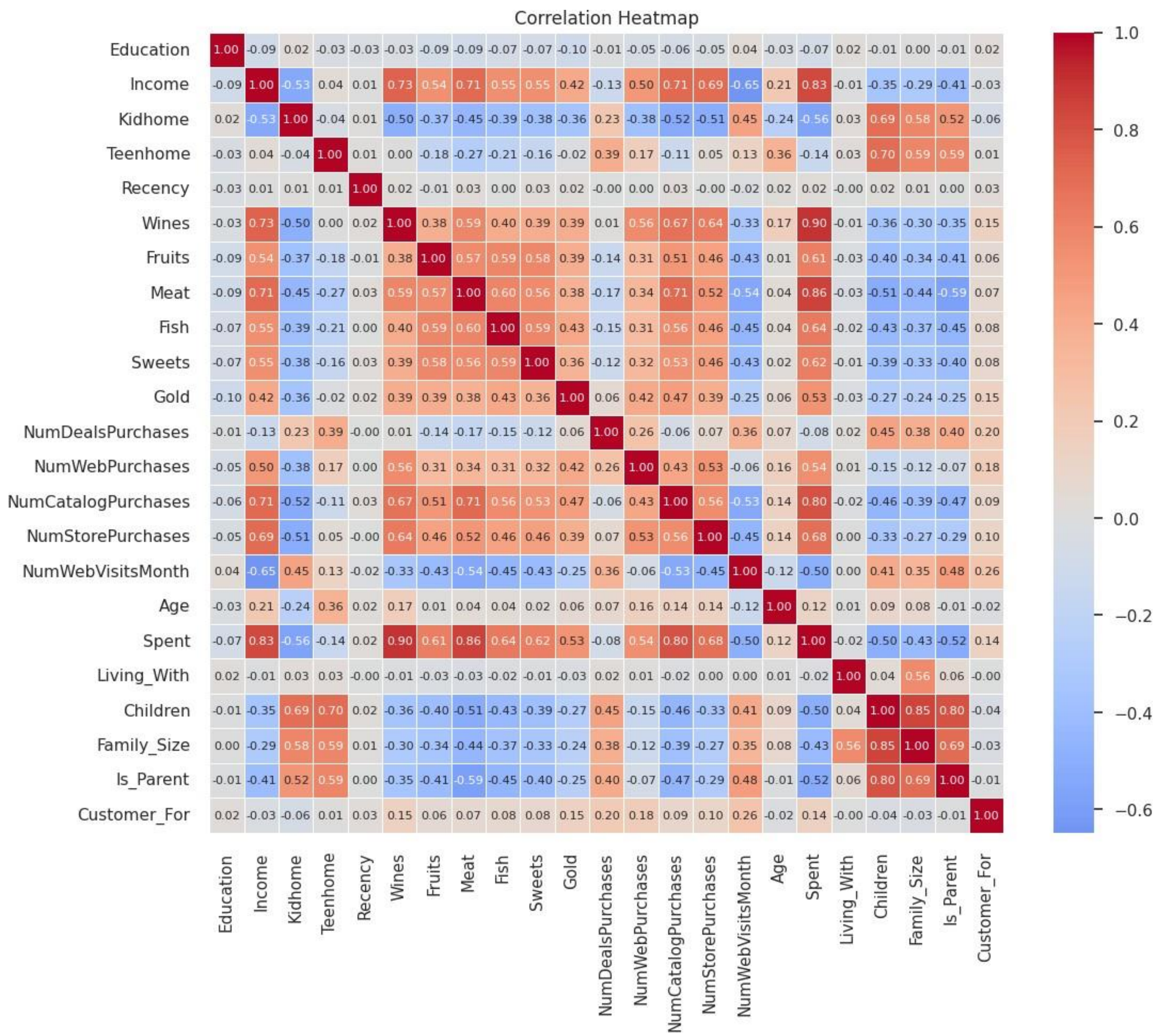

Figure 1: Heatmap of new featured dataset

IV. Technical uses

Experimental protocols of K-means and Agglomerative Clustering (AC)

1. Label Encoding

Label encoding is a technique used to convert categorical data (non-numeric data) into numerical values. Each category is assigned a unique integer. This method is often used for working with algorithms that require numeric input, like machine learning models.

2. Data Standardization

Data standardization is for the features to be rescaled to ensure the mean and the standard deviation to be 0 and 1, which make them normally distributed for enhancing model performance and interpretability.

3. Principal component analysis (PCA)

PCA is a dimensionality reduction technique that is used to reduce the number of features in a dataset while preserving as much of the original variance as possible. Dimensionality reduction will be employed to reduce the number of dimensions to three, ensuring a more manageable and informative representation of the data for later stages of clustering.

4. Elbow calculation and visualization

To determine the optimal number of clusters in a dataset when performing clustering analysis, the elbow point describing the rate of improvement in clustering quality.

5. Clustering by K-means and AC methods

K-means is a popular clustering algorithm that partitions data into K clusters based on similarity. It aims to minimize the distance between data points in the same cluster and maximize the distance between different clusters. Agglomerative Clustering, on the other hand, is a hierarchical clustering method that starts with individual data points as clusters and gradually merges them into larger clusters based on similarity.

6. Evaluation models (Davies-Bouldin Index and The Silhouette Coefficient)

- The Davies-Bouldin Index measures the average similarity between each cluster and its most similar cluster. A lower Davies-Bouldin Index suggests that the clusters are more compact and less overlapping, which is a desirable characteristic for good clustering.

- The Silhouette Coefficient quantifies how similar each data point in a cluster is to other data points within the same cluster compared to data points in neighboring clusters. Silhouette scores range from -1 to 1, where a high value indicates well-defined clusters.

7. Deployment

Practical implementation and execution of total clusters to support further research activities.

V. Modeling Results

Principal component analysis (PCA)

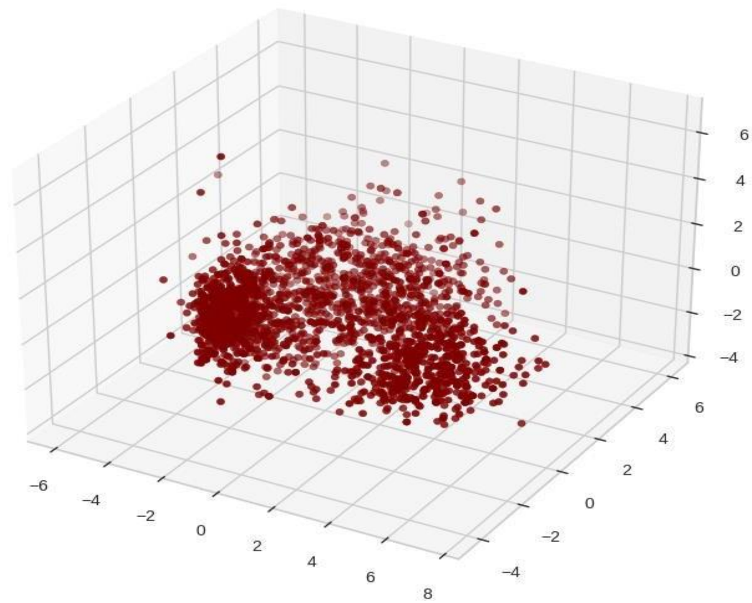

Figure 2: Visualization of the dataset after applying three-dimensional principal component analysis (3D-PCA)

The dataset applied by label encoding and standardizing to use for PCA application in figure 2. Through this figure, the data projection appears well-executed. The data points are clearly visible, and it seems to be no obvious outliers. The clusters that are apparent in the original data may be preserved in the reduced dimension.

K-means clustering

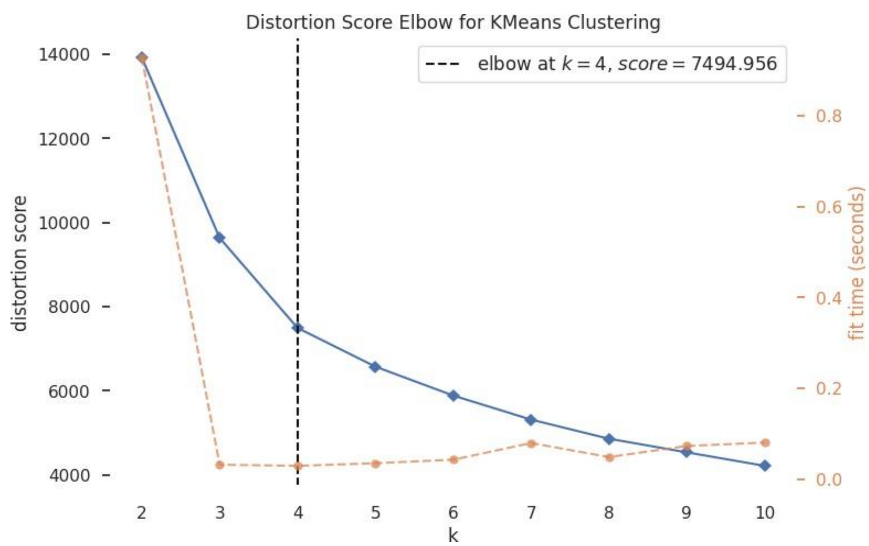

Figure 3: Elbow method for calculating and selecting the number of clusters for K-means clustering.

In Figure 3, a visual representation is presented with K-means clustering, depicting the optimal choice is annotated with a black dashed line. The graph shows how the distortion scores change with the number of clusters (k), represented by the blue line, whereas the orange dashed line displays the amount of time needed to train the clustering model per k. In this instance, it has been determined that the optimal number of clusters for the K-means clustering method is 4.

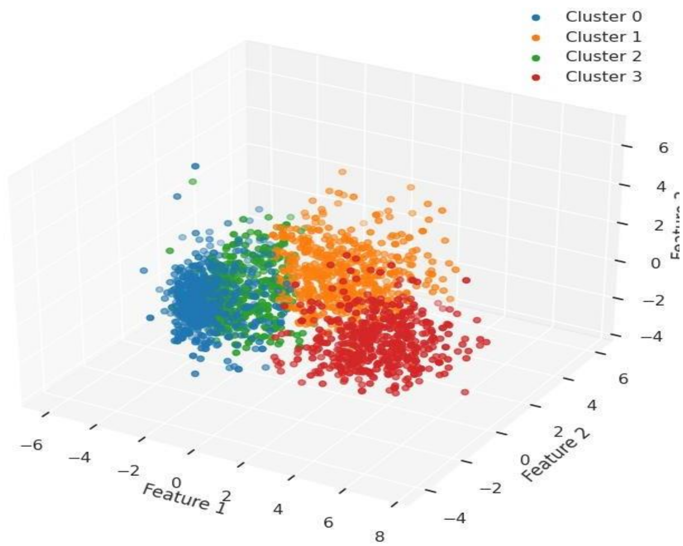

Figure 4: Visualization of four clusters using K-means clustering

The K-means cluster in the figure 4 shows four clusters of data points arranged in a 3D scatter plot. The clusters are well-separated, suggesting that the K-means algorithm has effectively partitioned the data into four distinct groups. The clusters appear to be arranged in a way that reflects the relationships between them. For example, Cluster 0 and Cluster 1 seem to be close together, suggesting that they are similar, while Cluster 2 and Cluster 3 are further away, suggesting that they may be more dissimilar.

Agglomerative Clustering (AC method)

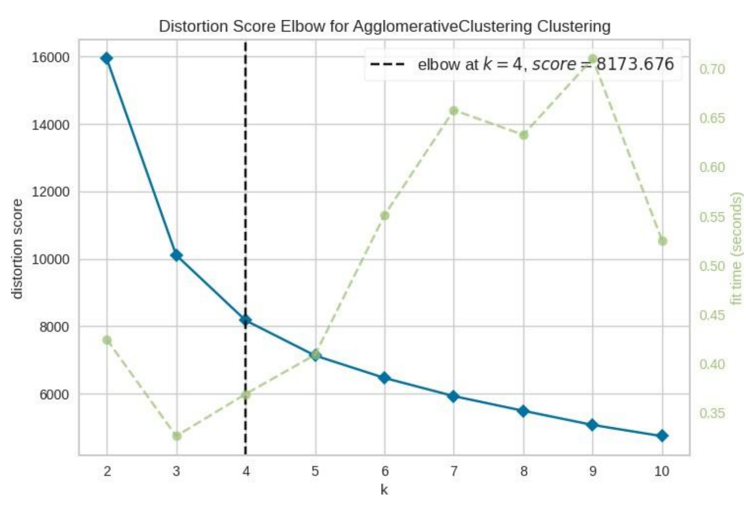

Figure 5: Elbow method for calculating and selecting the number of clusters for AC method

Figure 5 illustrates AC method and highlights the best optimal option also using a black dashed line. With the same description as figure 3, the most effective choice for the AC method is to have 4 clusters. The 3D plot of the agglomerative clustering shows four clusters, each of a different color in figure 6. The clusters are all relatively close together, but there may be some separation among clusters resulting by individual segmentation. This suggests that the clusters seem to be similar, but there may be some key differences.

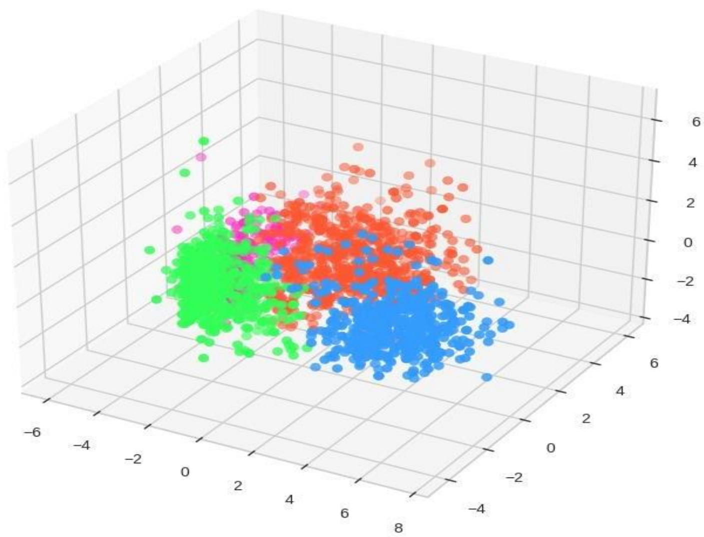

Figure 6: Visualization of four clusters using AC method

VI. Evaluation models

1. K-means

The Silhouette Score of 0.44566 indicates that the clusters are fairly well-defined and distinguishable, with some potential for enhancement. Even though the Davies-Bouldin Index is greater than 0.5 (0.85723), it implies a reasonable separation between clusters in table 3.

Evaluation methods	K-means clustering results	AC Results
Silhouette Score	0.44566	0.43735
Davies-Bouldin Index	0.85723	0.8272

Table 3: Evaluation model results of K-means clustering and AC methods

2. Agglomerative Clustering

The Silhouette Score of 0.43735 suggests that the clusters formed by Agglomerative Clustering exhibit a moderate level of cohesion and separation. With a Davies-Bouldin Index of 0.8272, the Agglomerative Clustering demonstrates moderate separation between clusters. Although there is some overlap between clusters, the overall separation is reasonable in table 3.

VII. Deployment - Cluster Analysis

1. The cluster analysis of K-means

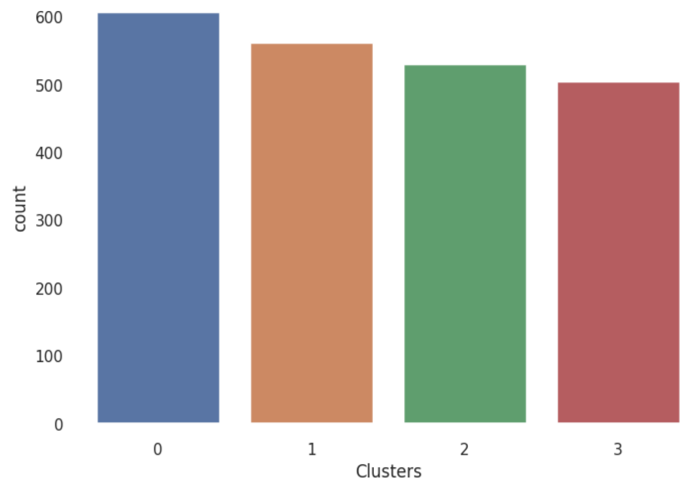

Figure 7: Distribution of K-means clustering results

Figure 7 provides the outcome of the K-means clustering analysis indicates that the clusters are fairly distributed. According to the results, there are 12 different categories to provide information of 4 clusters contributing to the discussion. Despite having the small difference from cluster 0 to cluster 3. While table 4 provides a comprehensive overview of various characteristics within four distinct clusters. Particularly, it offers insights into how these clusters are defined across different 3categories, including age, family size, spending behavior, annual income, and customer engagement.

No.	Category	Cluster 0	Cluster 1	Cluster 2	Cluster 3
1	Age (years old)	37 - 50	47 - 64	49 - 64	41 - 65
2	Number of Children (people)	1	1	1 - 2	0
3	Family size (people)	2 - 3	2 - 3	3 - 4	1 - 2
4	The total spendings for groceries ($)	37 - 122	605 - 1,105	45 - 249	1049 - 1,672
5	Customer_For (days)	2.95e+16 - 5.83e+16	3.79e+16 - 6.43e+16	2.25e+16 - 5.25e+16	2.75e+16 - 5.94e+16
6	Income ($)	22,609 - 37,398	55,218 - 69,078	35,791 - 50,773	69,520 - 82,427
7	Recency (days)	24 - 74	26 - 72	24 - 75	23 - 73
8	Discounts used (times)	1 - 2	2 - 5	1 - 4	1
9	Website visits (times)	6 - 8	4 - 7	5 - 7	1 - 3
10	Buying groceries through websites (times)	1 - 3	5 - 8	1 - 4	3 - 6
11	Purchases using a catalogue (times)	0 - 1	2 - 5	0 - 1	4 - 7
12	Direct in-store purchases (times)	3 - 4	6 - 11	3 - 5	6 - 10

Table 4: The detail of 12 categories from four clustering by K-means

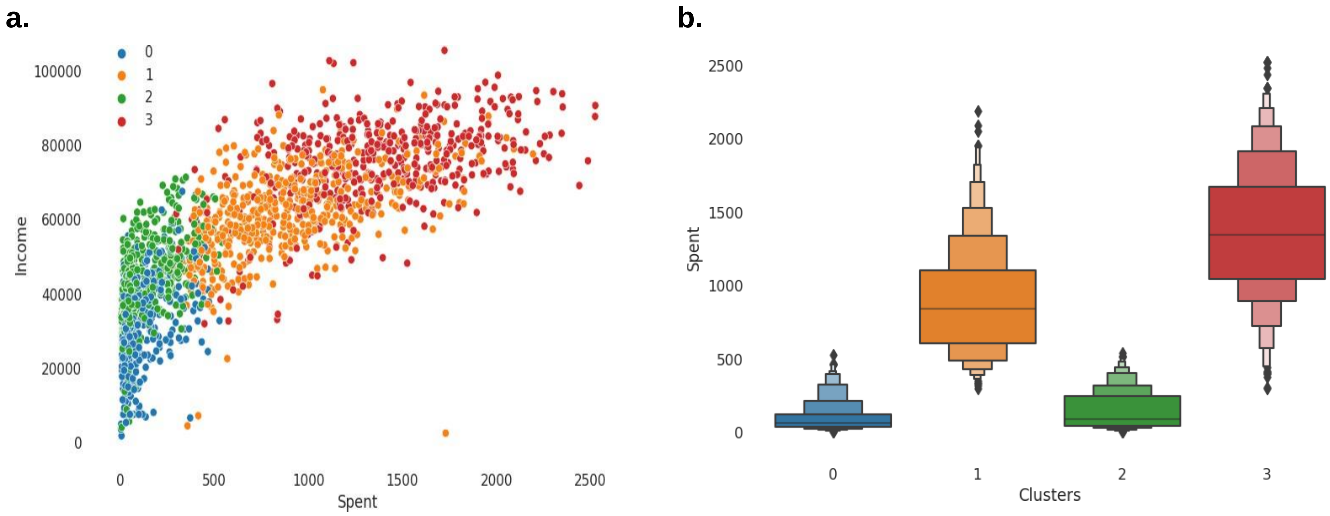

Figure 8: The distribution of clusters with income and total spendings for groceries (spent). The left-handed side describes the distribution of all four clusters with income and spent. The right-handed side depicts the rank of each cluster equivalent to Spent.

Figure 8 illustrates a four-cluster segmentation between two categories, including Total spending for groceries and Annual Income of customers. With cluster 0 representing customers with low income and low spending, cluster 1 comprising those with average income and average spending, cluster 2 consisting of customers with average income but low spending, and cluster 3 encompassing those with high income and varying levels of spending, ranging from average to high. Therefore, four clusters can be concluded:

- Cluster 0: Low Income, Low Spending
- Cluster 1: Average Income, Average Spending
- Cluster 2: Average Income, Low Spending
- Cluster 3: High Income, From Average to High Spending

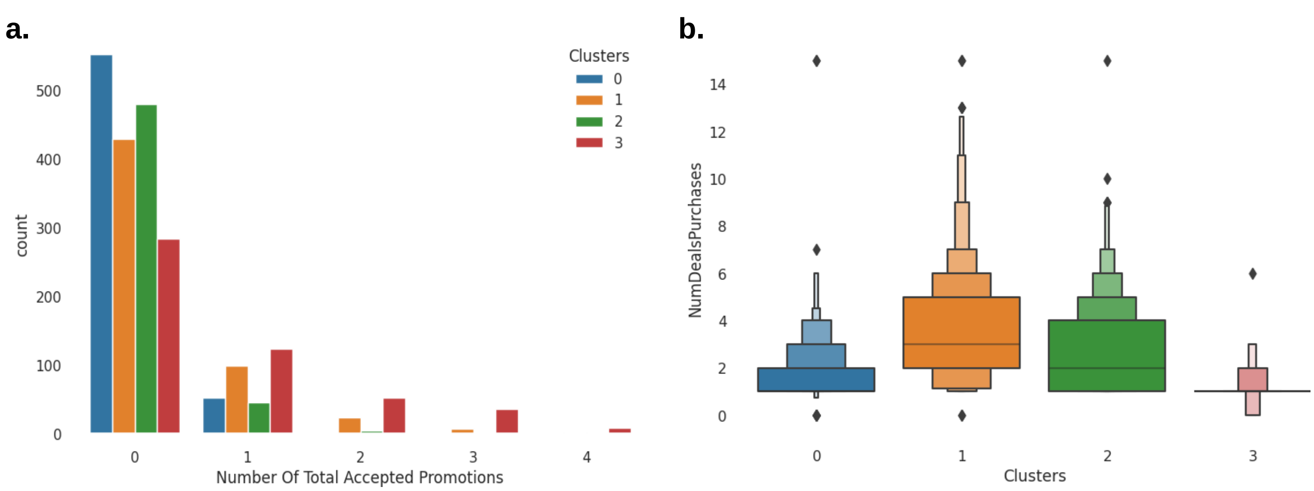

Figure 9: The distribution of four clusters equivalent to types of Accepted Promotions (left-handed picture) and a number of customers buying groceries with discounts.

In terms of the five promotion types in figure 9a, AcceptedCmp1 stands out as the most preferred choice among customers. Additionally, the utilization of different promotion types is more diverse within group 3 customers. However, the number of purchases with discounts in figure 9b made by customers is quite low among all groups, mainly ranging from 1 to 5 purchases per customer, especially low for group 3, which makes barely any purchases with discounts.

Total conclusion of four clusters with some distinct categories:

Cluster	**0**	**1**	**2**	**3**
Income	Low income	Average income	Average income	High income
Spending	Low spending	Average spending	Low spending	Average to high spending
Acceptance offer 1 (AcceptedCmp1)				
Numbers of Deals Purchases	Third	First	Second	Fourth

2. The cluster analysis of AC method

Figure 8 illustrates four different clusters in distribution. From cluster 0 to 3, the distribution is steadily decreasing starting from about 700 to around 400 records. According to table 5, there are 12 selected categories giving information about the details of each cluster.

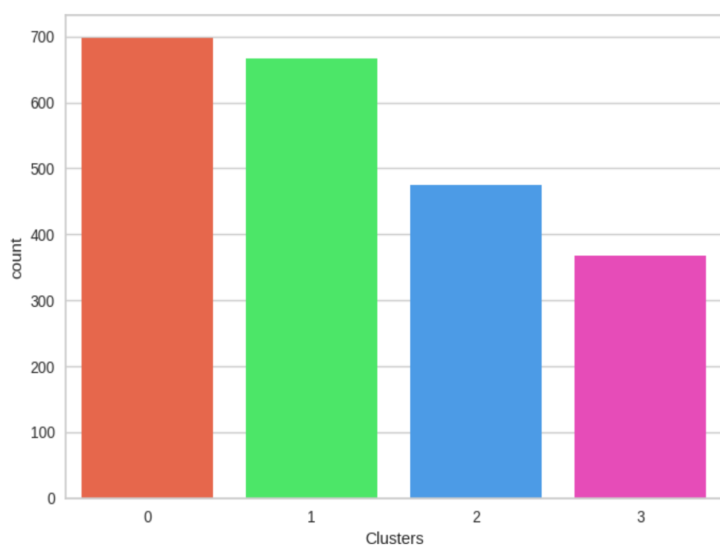

Figure 10: Distribution of AC results

No.	Category	Cluster 0	Cluster 1	Cluster 2	Cluster 3
1	Age (years old)	47 - 64	38 - 51	40 - 65	48 - 63
2	Number of Children (people)	1	1	0	2
3	Family size (people)	2 - 3	2 - 3	1 - 2	3 - 4
4	The total spendings for groceries ($)	495 - 1,079	38 - 128	316 - 1,364.5	40 - 149.5
5	Customer_For (days)	3.43e+16 - 6.28e+16	2.57e+16 - 5.65e+16	2.91e+16 - 5.98e+16	2.51e+16 - 5.55e+16
6	Income ($)	53,230 - 68,118	23,277 - 39,770	69,728 - 82,487	33,420 - 46,443
7	Recency (days)	25 - 72	23 - 75	25 - 73	25 - 75
8	Discounts used (times)	2 - 4	1 - 2	1	2 - 4
9	Website visits (times)	4 - 7	5 - 8	1 - 3	5 - 8
10	Buying groceries through websites (times)	5 - 8	1 - 3	3 - 6	1 - 3
11	Purchases using a catalogue (times)	2 - 5	0 - 1	4 - 7	0 - 1
12	Direct in-store purchases (times)	6 - 10	3 - 4	6 - 10	3 - 4

Table 5: The detail of 12 categories from four clustering by AC method

a. **b.**

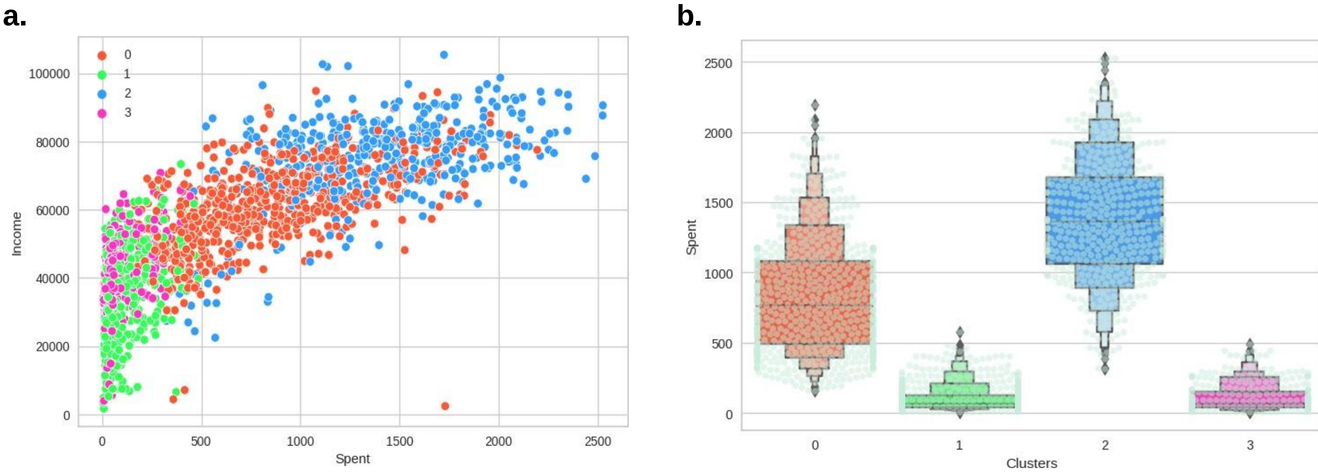

Figure 11: A left-handed picture shows Scatter plot of customers' spending and income in each clusters. A right-handed picture is box plots of customers' spending amount in each clusters

Figure 11a shows customers' spending and income within each clustering group. The x-axis represents customers' spending on the groceries, while the y-axis represents their income. It clearly distinguishes between the four groups, making it easy to identify their characteristics:

- Cluster 0: High spending and high income
- Cluster 1: Low spending and low income
- Cluster 2: High spending and high income
- Cluster 3: Low spending and average income

While the figure 11b shows the amount of customers' spending in the last 2 years on groceries. The box plots show that cluster 2 comprises the largest set of customers, with cluster 0 closely following. This prompts an examination of how much on average each cluster is spending on for future marketing strategies.

a. **b.**

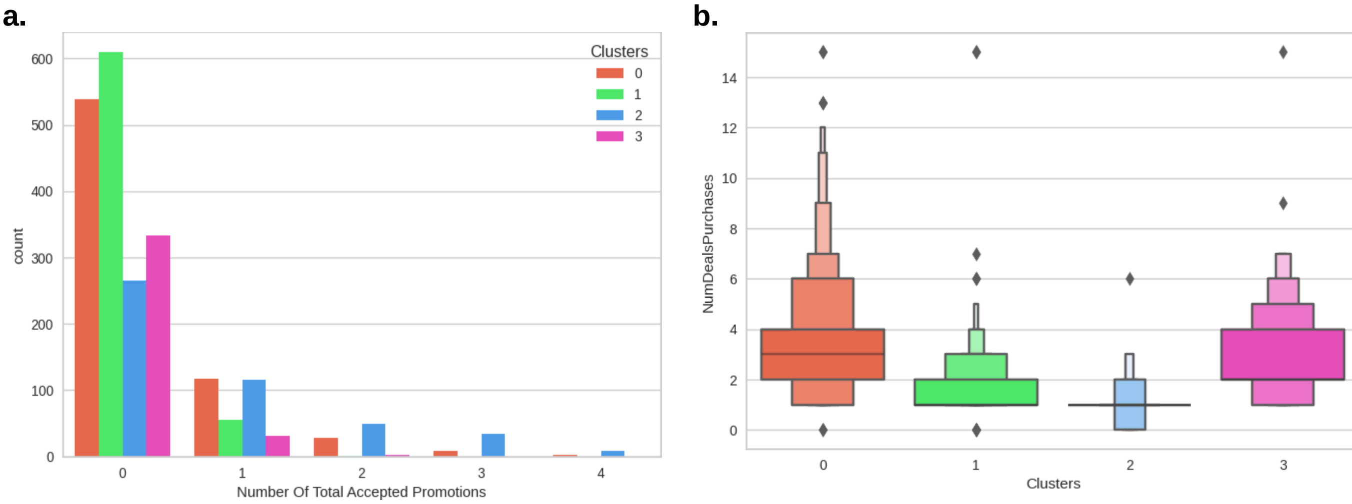

Figure 12: A left-handed picture shows Number of total accepted promotions by customers in each clusters. A right-handed picture is the average of the number of purchases made with a discount by each customer in the clusters.

Figure 12a shows the number of total accepted promotions by customers, revealing that there has not been a huge response to the promotional campaigns implemented. The total number of participants across all clusters remains relatively low, highlighting the need for more targeted and strategically planned campaigns to stimulate sales. Figure 12b the average of the number of purchases made with a discount by each customer in the clusters. Deals have shown promising results, especially in clusters 0 and 3, where they perform exceptionally well.

Total conclusion of four clusters with some distinct categories:

Cluster	**0**	**1**	**2**	**3**
Income	High income	Low income	High income	Average income
Spending	High spending	Low spending	High spending	Low spending
Acceptance offer 1 (AcceptedCmp1)				
Numbers of Deals Purchases	First	Third	Fourth	Second

3. Profiling the clusters

To profile the clusters, there are 3 major attributes, such as age, number of children, family size, selected to describe the characteristics combined with income, total spending and whether customers use discounts. According to the results of four clusters by K-means clustering, the profile of each cluster includes:

	Profiling the clusters by K-means
Cluster 0	+ The majority of these people are parents + At the max are 3 members in the family + They majorly have one kid and not teenagers + Relatively younger + *Low annual income (*)* + *Low spending for groceries* + *Normal demands in shopping with discounts*
Cluster 1	+ Are definitely a parent + At the max have 4 members in the family and at least 2 + Most have teenager at home + Relatively older + *Average annual income (*)* + *Average spending for groceries* + *High demands in shopping with discounts*
Cluster 2	+ Definitely a parent + At the max of 5 members and at least 3 in the family + Majority have teenager + Relatively older + *Average annual income (*)* + *Low spending for groceries* + *Normal demands in shopping with discounts*
Cluster 3	+ Are definitely not a parent + At the max of 2 people in the family + All ages + *High annual income (*)* + *Average to high spending for groceries* + *Less demands in shopping with discounts*

(*) income, total spendings and shopping with discounts conclusions are only for reference.

Table 6: K-means clustering results

According to the results of four clusters by AC method, the profile of each cluster includes:

	Profiling the clusters by AC method
Cluster 0	+ Are definitely a parent + At the max have 4 members in the family and at least 2 + Most have teenager at home + Relatively older + *High annual income (*)* + *High spending for groceries* + *High demands in shopping with discounts*
Cluster 1	+ The majority of these people are parents + At the max are 3 members in the family + They majorly have one kid and not teenagers + Relatively younger + *Low annual income (*)* + *Low spending for groceries* + *Normal demands in shopping with discounts*
Cluster 2	+ Are definitely not a parent + At the max of 2 people in the family + All ages + *High annual income (*)* + *High spending for groceries* + *Less demands in shopping with discounts*
Cluster 3	+ At the max of 5 members and at least 3 in the family + Majority have teenager + Relatively older + *Average annual income (*)* + *Low spending for groceries* + *Normal demands in shopping with discounts*

(*) income, total spendings and shopping with discounts conclusions are only for reference.

Table 7: AC method clustering results

VIII. Marketing strategy suggestions

Based on K-means clustering information of customer characteristics in table 6, some marketing suggestions can be proposed:

Cluster 0: Young Parents with Low Income

+ Affordable Family Solutions: Offer budget-friendly family products, essentials for young children, and special deals for small families.
+ Child-Centric Marketing: Focus on products and services suitable for families with one child. Address the unique needs and interests of this demographic.
+ Education and Parenting Support: Highlight products and services related to child development, education, and parenting support.

Cluster 1: Parents of Teenagers with Average Income

+ Family Discounts: Offer family-sized discounts, bulk deals, and packages to cater to their larger family size.
+ Teen-Focused Campaigns: Create marketing campaigns that appeal to teenagers and their parents, offering products and services that cater to their needs and interests.
+ Discount-Centric Promotions: Recognize their high demand for discounts and provide a variety of discount-related incentives and loyalty programs.

Cluster 2: Parents with Teenagers and Average Income

+ Teen-Centric Marketing: Recognize the presence of teenagers in these households. Offer products, services, and promotions that specifically appeal to this age group.
+ Family-Centric Promotions: Promote family-oriented discounts, events, and activities that cater to their family size and demographics.
+ Average Income Offers: Provide promotions that align with their average income, ensuring they feel they are receiving good value.

Cluster 3: Singles and High Earners

+ Personalized Luxury: Recognize their single status and higher income. Offer personalized, luxury, and high-end products, services, and experiences.
+ Quality and Convenience: Emphasize quality, convenience, and the benefits of premium and exclusive items.
+ Exclusive Offers: Provide exclusive offers and experiences to make high-income customers feel valued and appreciated.

On the other hand, according to the AC results of customer characteristics in table 7, some marketing strategy suggestions can be proposed:

Cluster 0: Parents with high income

+ Family Marketing: Promote products or services that cater to families, especially family discounts and packages
+ Teen Offers: Since most have teenagers at home, create marketing campaigns that appeal to teenagers, such as back-to-school deals and products popular among this age group.
+ Premium Services: Offer premium or exclusive services and products to align with their high-income status.
+ Discount Strategies: Provide frequent discounts and loyalty programs for grocery shopping, as they have a high demand for discounts.

Cluster 1: Young Parents with Low Income

+ Budget Friendly Solutions: Offer budget friendly products, special deals for small families, and cost-effective options.
+ Family-Oriented: Emphasize products suitable for families with one child, and create campaigns that address the concerns and interests of young parents.
+ Value-Based Promotions: Offer value-based promotions and bundle deals to help them save money.
+ Education: Highlight products or services related to child development, education, and early parenting.

Cluster 2: Singles and High Earners

+ Personalized Shopping Experience: Offer personalized shopping experiences, luxury items, or high-end services.
+ Convenience and Quality: Emphasize quality, convenience, and the benefits of premium products or services.
+ Exclusive Offers: Provide exclusive offers for high-income customers to make them feel valued.
+ Discount Awareness: Despite their high income, some may still appreciate discounts, so offer occasional special discounts to maintain their interest.

Cluster 3: Larger Families with Teens

+ Family Discounts: Recognize the size of their families and offer family-sized discounts, bulk deals, and packages.
+ Teen-Focused Campaigns: Create marketing campaigns that appeal to teenagers, offering products and services that cater to their needs and interests.
+ Average Income Promotions: Offer promotions that align with their average income, making sure they feel their budget is well-considered.

IX. Conclusion

In summary, the use of K-means and AC techniques for customer segmentation provides valuable insights into different categories. Overall, marketers should prioritize promoting Acceptance offer 1 (AcceptedCmp1) due to its highest priority among customers. Additionally, considering the comprehensive insights from both methods, marketers can make effective marketing strategies to enhance business profitability while there is a minor difference in the results. This adaptability is crucial because various circumstances can arise in the business environment. Consequently, the method results should not be used for direct comparison but rather as a guide for identifying appropriate marketing strategies for specific customer segments.

X. Reference

Camilleri, M. A. (2017). Market Segmentation, Targeting and Positioning. *Travel Marketing, Tourism Economics and the Airline Product: An Introduction to Theory and Practice*.

Musumali, B. (2019). An Analysis why customers are so important and how marketers go about in understanding their decisions.

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* , 160.

Appendix

The cluster analysis of K-means supporting the clustering profiles

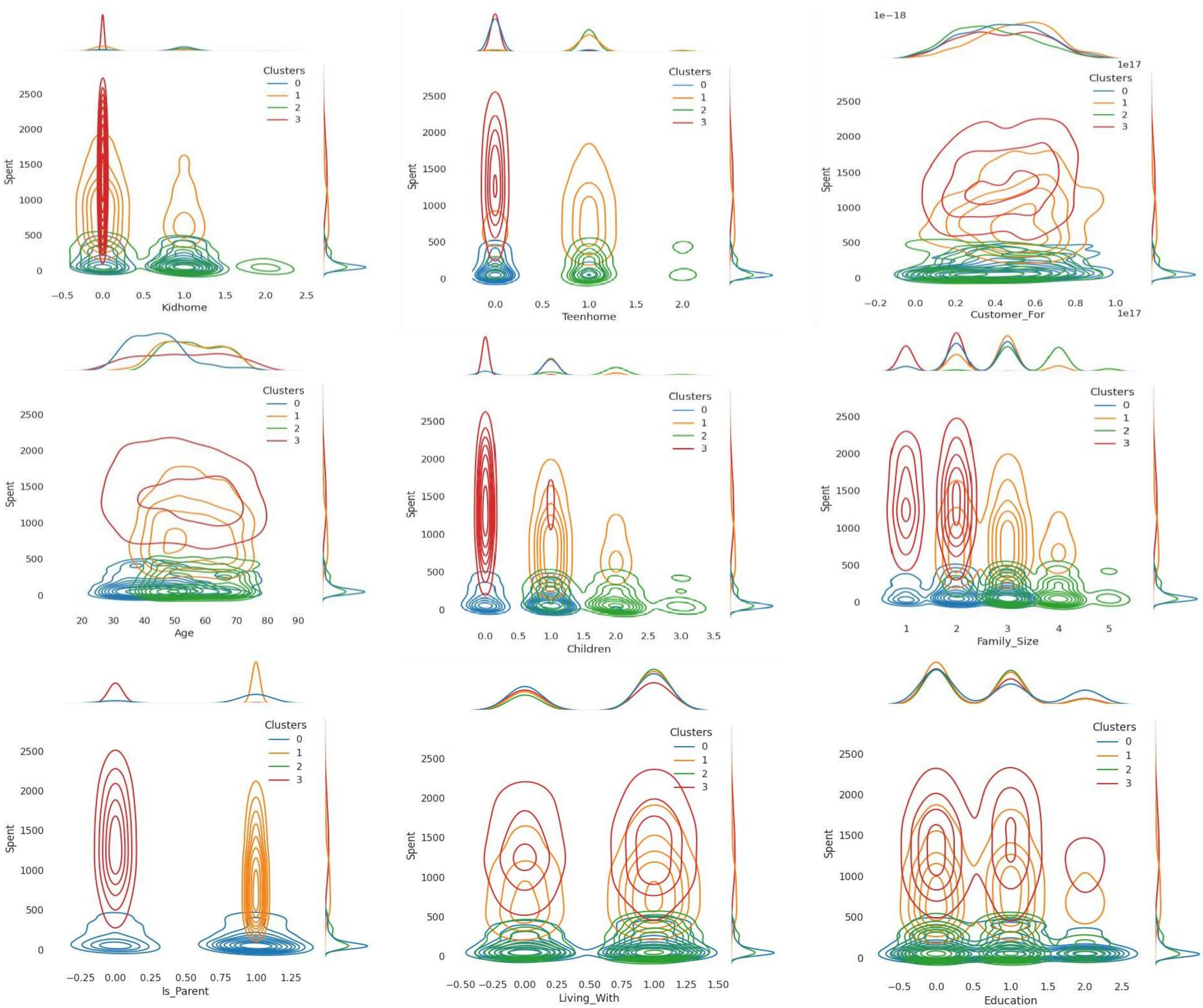

Figure 13: A series of categories supporting the results of K-means clustering profiles.

Figure 13 and 14 depict a list of visualized categories supporting the details of K-means and AC clustering results. Besides the results having in the profiling parts, Education and Customer_For also affect directly the customer behaviors while shopping. On the other hand, the income giving in the figure table 4 and 5 helping the marketers identify which types of customers should be served.

The cluster analysis of AC supporting the clustering profiles

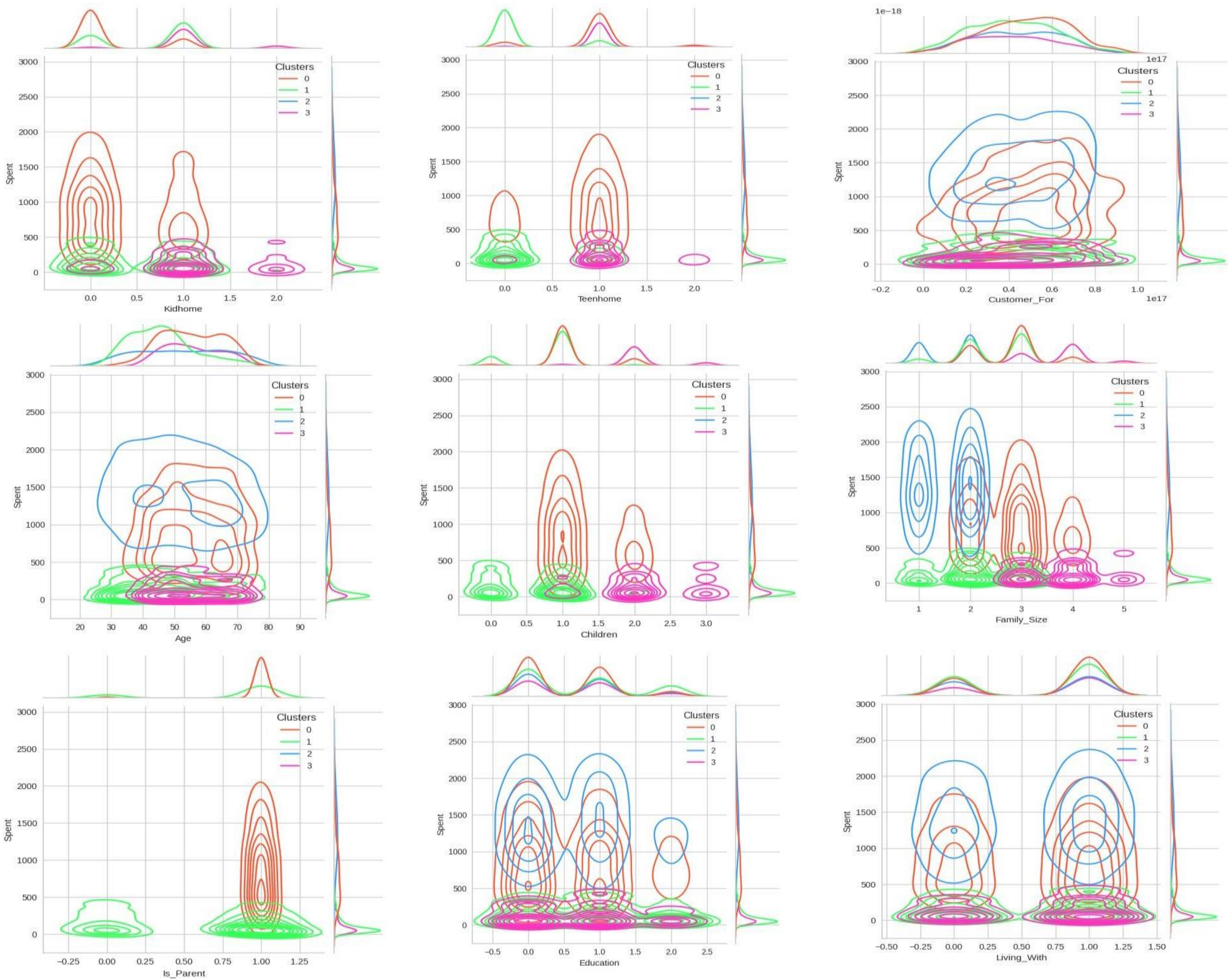

Figure 14: A series of categories supporting the results of AC profiles.