



Segmentation Strategy

Optimizing Customer Segmentation by Machine Learning algorithms - A case study

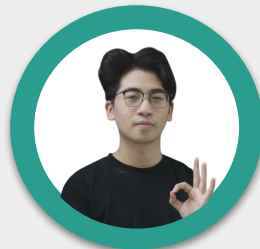
Subject: Fundamentals of Data Science

Presented by DS team

OUR TEAM



Pham Hai Nam



Pham Minh Hoang



Phung Duc Thai



Hua Hai Minh



Le Mau Minh Phuc

TABLE OF CONTENTS



**Introduction
and Objective**

A



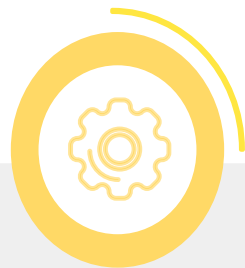
**Dataset
Understanding
& Preparation**

B



**Segmentation
Process**

C



**Model
Evaluation &
Profiling**

D



**Marketing
suggestions**

E

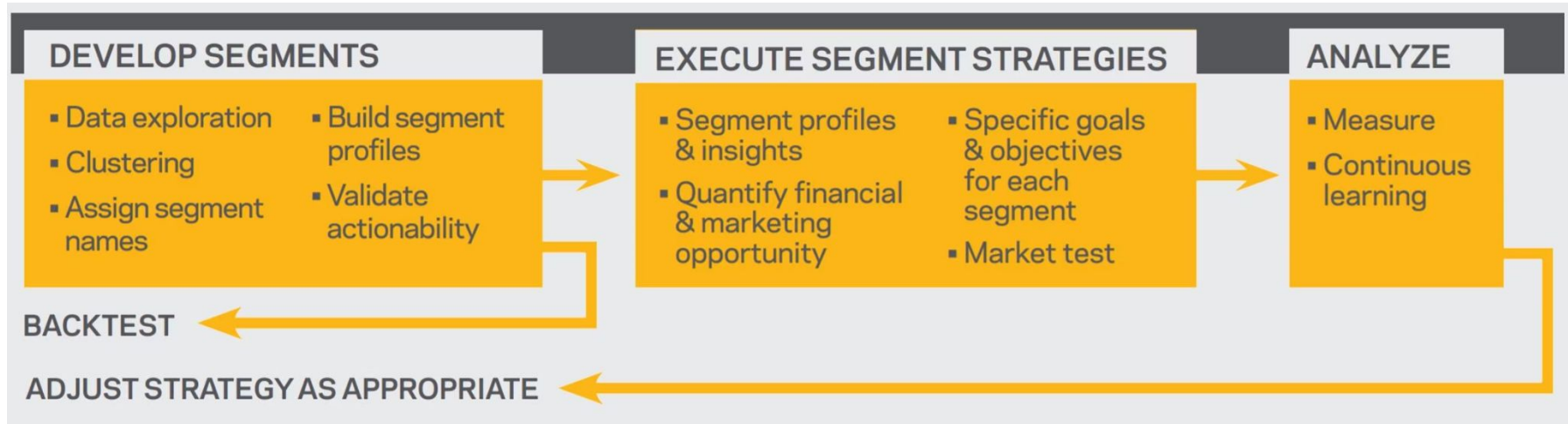


Conclusion

F

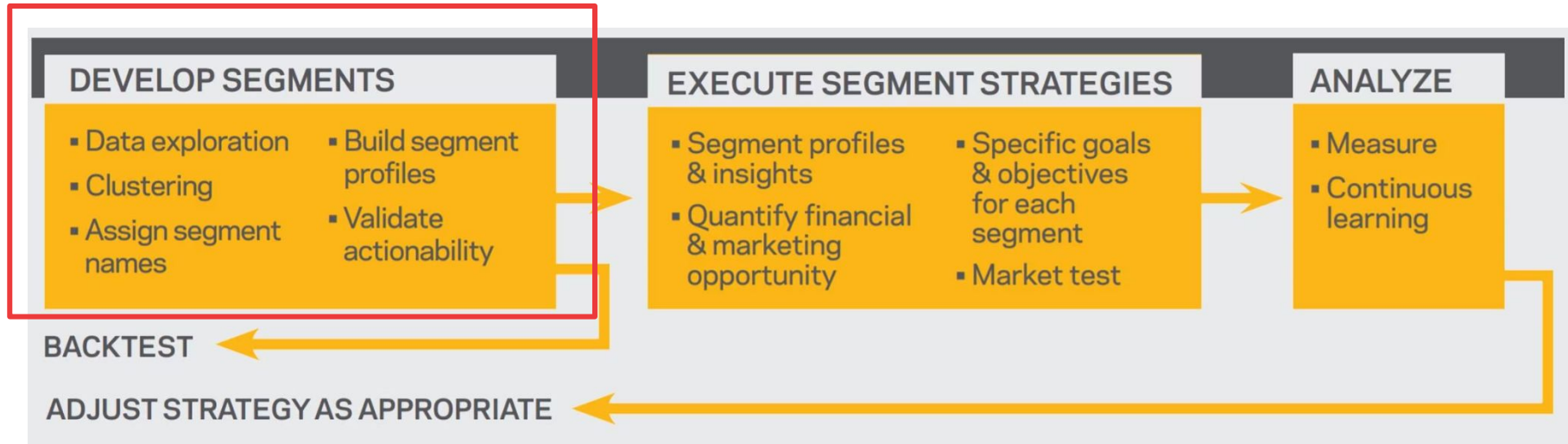
Introduction: Customer targeting strategies according to segmentation

In the retail industry, the key to segmentation strategies begins with gathering customer data. Once enough customer data is collected and analyzed, certain characteristics and behaviors rise to the top, forming a distinct and unique segment profile.



Introduction: Customer targeting strategies according to segmentation

Objective: To employ unsupervised machine learning techniques to analyze and cluster the dataset in order to perform customer segmentation for a grocery company.



Process for developing customer segmentation strategy

- + Visualize collected data to understand each category.
- + Complete the objectives in experiment.

Data Understanding, Preparation and EDA



Business case

- + Customer Segmentation in the Retail Industry.
- + Determine the objectives.



Segmentation

- + Use Elbow method to optimize segments.
- + Perform segmentation to cluster customers with K-means clustering and Agglomerative Clustering.



Model Evaluation

- + Internal metrics: Silhouette Score and Davies-Bouldin Index.
- + External Metrics : Visual Evaluation



Profiling

Summarize and characterize the properties, behaviors, characteristics of the data points.



Marketing Suggestions

Suggest the potential marketing strategy according to clusters.



Data required to create Customer Segmentation



kaggle

Customer information

- ID
- Year_Birth
- Education
- Martial_Status
- Income
- Kidhome
- Teenhome
- Dt_Customer
- Recency
- Complain

01

Promotion

- NumDealsPurchases
- AcceptedCm1
- AcceptedCm2
- AcceptedCm3
- AcceptedCm4
- AcceptedCm5
- Response

02

Products

Amount spent on different products in last 2 years

- MntWines
- MntFruits
- MntMeatProducts
- MntFishProducts
- MntSweetProducts
- MntGoldProducts

03

Place

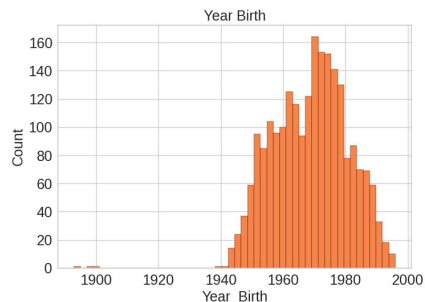
- NumWebPurchases
- NumCatalogPurchases
- NumStorePurchases
- NumWebVisistsMonth

04

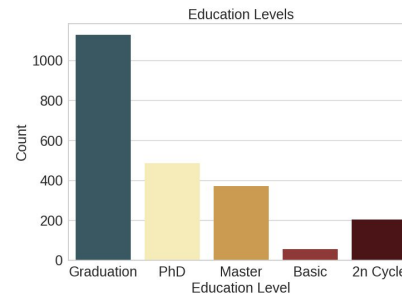
2240 entries and 29 attributes (2021)

Dashboard for Customer Segmentation in the Dataset

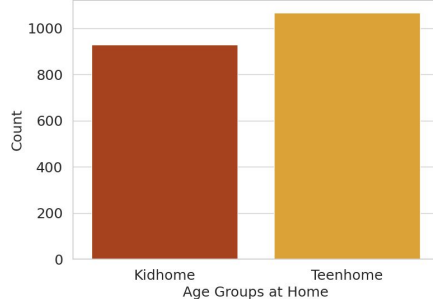
Year birth



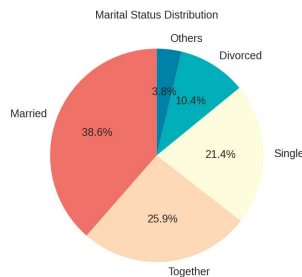
Education



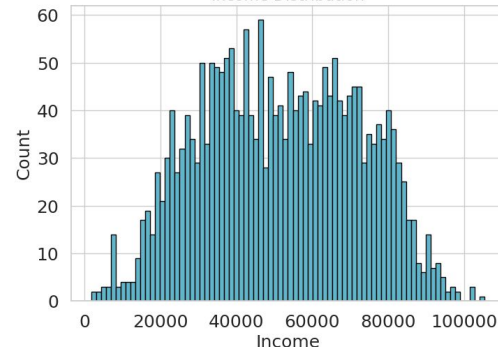
Number of Records with Nonzero Kidhome and Teenhome Values



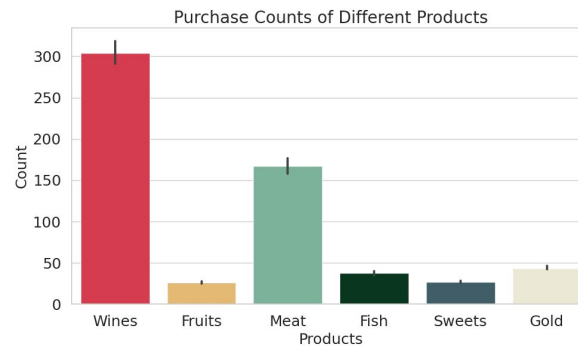
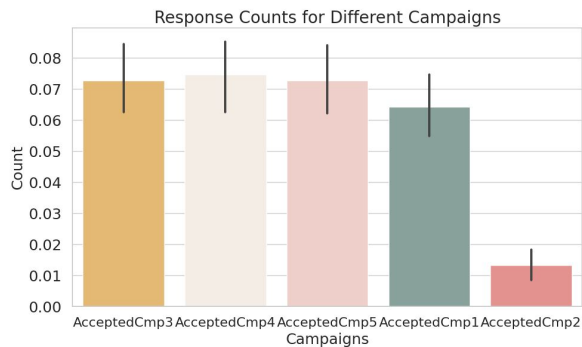
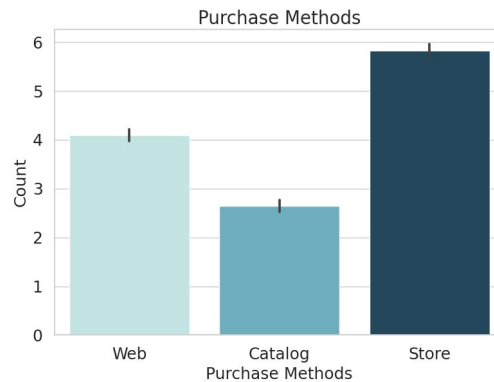
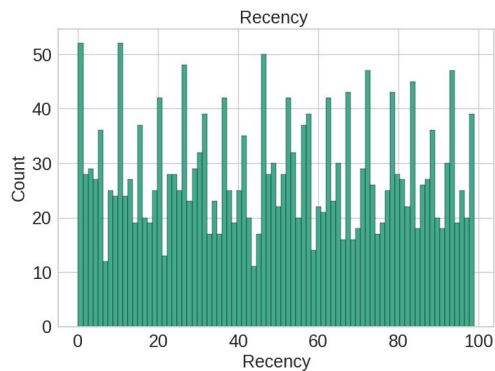
Household



Income Distribution



Dashboard for Customer Behavior Purchases



Data Preparation

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

Name and Criteria Change

Basic, 2n Cycle $\Rightarrow$ Undergraduate
Graduation $\Rightarrow$ Graduate
Master, PhD $\Rightarrow$ Postgraduate

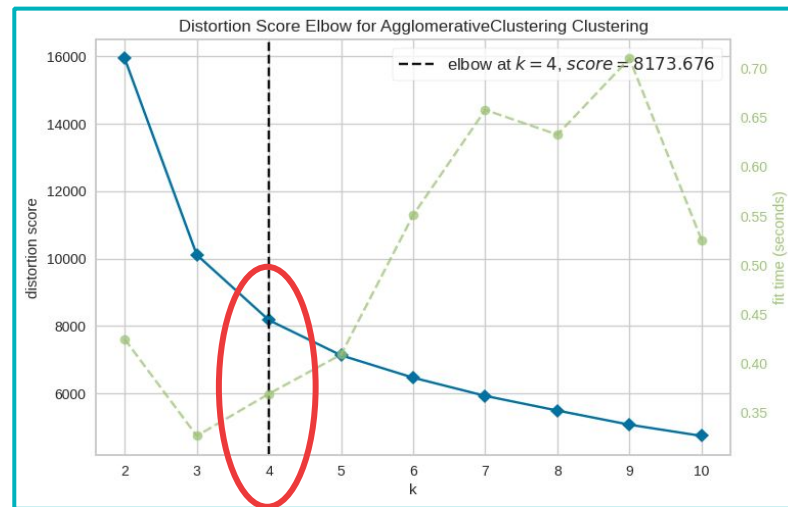
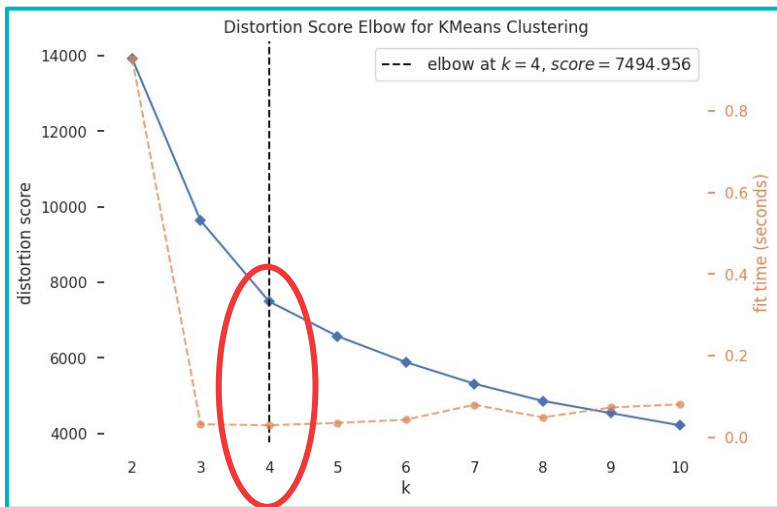
Removal

'Z_costcontact'
'Z_revenue'

Creation

Age	Current year - 'Year_Birth'
spent	Total of 'MntWines' 'MntFruits' 'MntMeatProducts' 'MntGoldProds'
Living_with	"Marital_Status" $\rightarrow$ "Partner" or "Alone"
Children	Total of "Kidhome" and "Teenhome"
Family_Size	Total of "Living_With" and "Children"
Is_Parent	If "Children" > 0
Education	Divided into three levels

Technical results and Model Evaluation

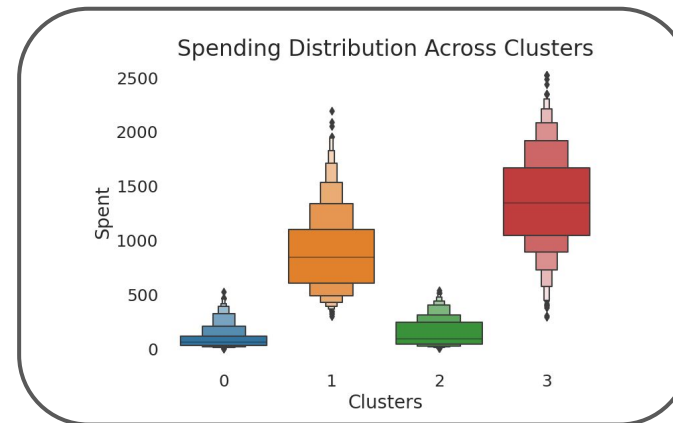
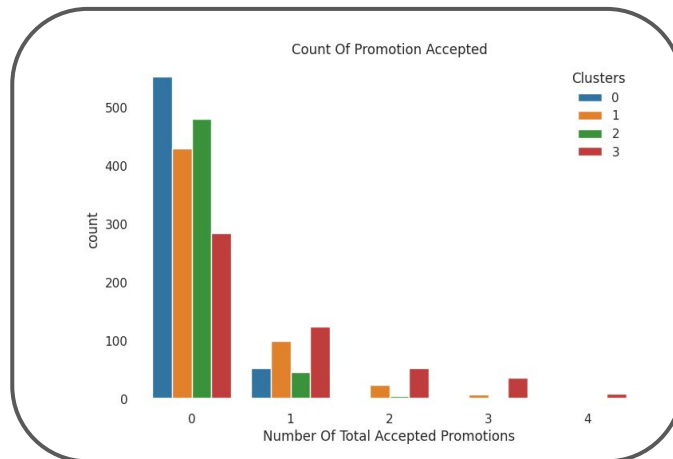
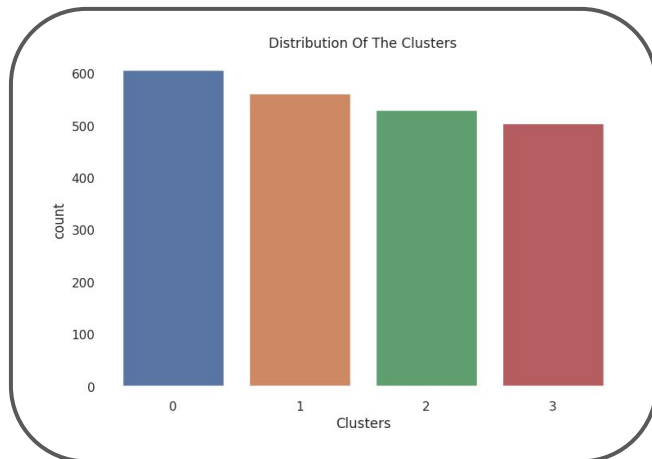


Based on the Elbow method, optimized number of clusters for both K-means and AC is 4

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

Clustering Visualization Results - K-means

Example:
Cluster 3



Profiling the clusters by K-means method

Cluster 0

Majority are parents
Majority have one kid
Family up to 3 members
Relatively younger
Low annual income
Low spending for groceries
Normal demands in discounts



Cluster 2

Definitely are parents
Majority have teenagers
Family of 3 to 5 members
Relatively older
Average annual income
Low spending for groceries
Normal demands in discounts



Cluster 1

Definitely are parents
Mostly have teenager
Family of 2 to 4 members
Relatively older
Average annual income
Average spending for groceries
High demands in discounts



Cluster 3

Definitely are not parents
Family up to 2 members
All ages
High annual income
Average to high spending for groceries
Less demands in discounts



Marketing Strategy Suggestions - K-means Clustering



Cluster 0: Young Parents with Low Income: Affordable Family Solutions, Child-Centric Marketing, Education and Parenting Support

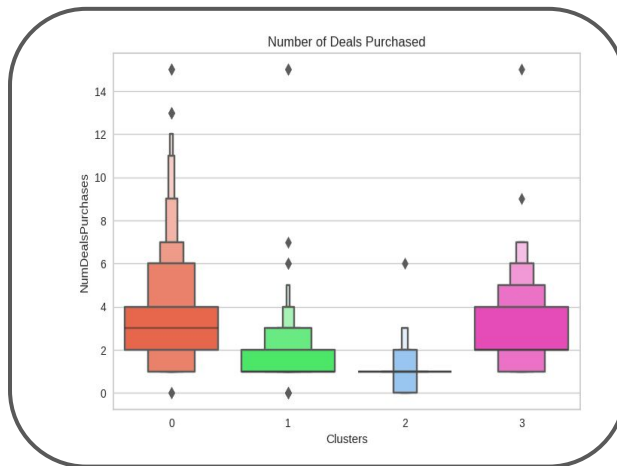
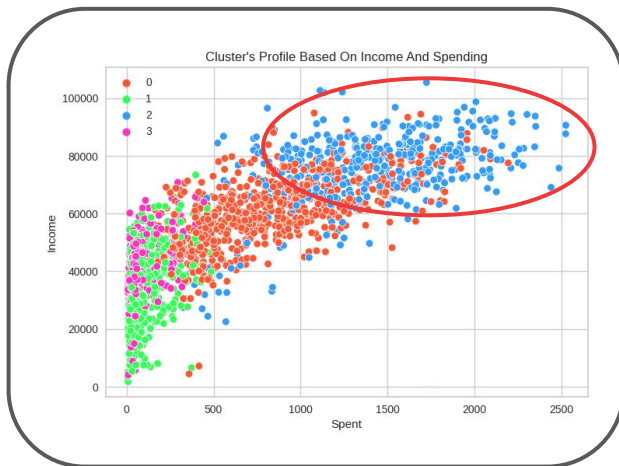
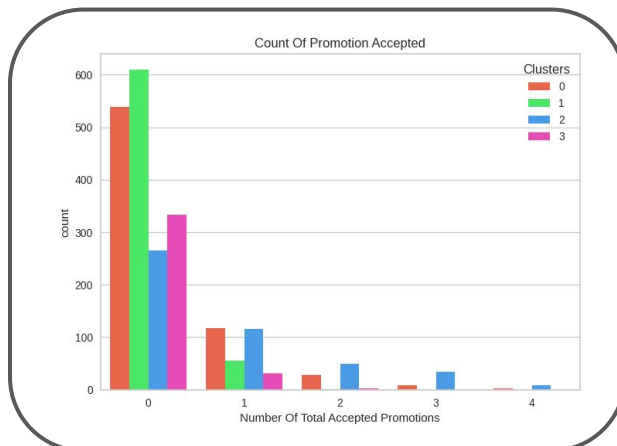
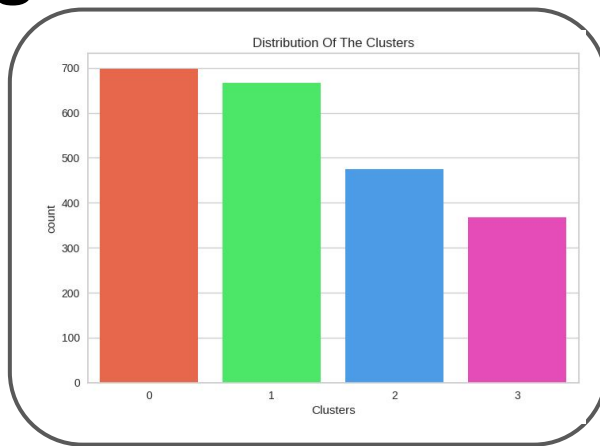
Cluster 1: Parents of Teenagers with Average Income: Family Discounts, Teen-Focused Campaigns, Discount-Centric Promotions

Cluster 2: Parents with Teenagers and Average Income: Teen-Centric Marketing, Family-Centric Promotions, Average Income Offers

Cluster 3: Singles and High Earners: Personalized Luxury, Quality and Convenience, Exclusive Offers

Clustering Visualization Results - AC method

Example:
Cluster 2



Profiling the clusters by AC method

Cluster 0

Definitely are parents
Mostly have teenagers
Family of 2 to 4 members
Relatively older
High annual income
High spending for groceries
High demands in discounts



Cluster 2

Definitely are not parents
Family up to 2 members
All ages
High annual income
High spending for groceries
Less demands in discounts



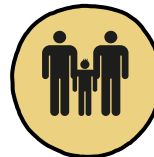
Cluster 1

Majority are parent
Mostly have 1 kid
Family of 3 members maximum
Relatively younger
Low annual income
Low spending for groceries
Normal demands in discounts



Cluster 3

Mostly have teenagers
Family of 3 to 5 members
Relatively older
Average annual income
Low spending for groceries
Normal demands in discounts



Marketing Strategy Suggestions - AC method



Cluster 0: Parents with high income: Family Marketing, Teen Offers, Premium Services, Discount Strategies

Cluster 1: Young Parents with Low Income: Budget Friendly Solutions, Family-Oriented, Value-Based Promotions, Education

Cluster 2: Singles and High Earners: Personalized Shopping Experience, Convenience and Quality, Exclusive Offers, Discount Awareness

Cluster 3: Larger Families with Teens: Family Discounts, Teen-Focused Campaigns, Average Income Promotions



CONCLUSION

K-means and Agglomerative Clustering techniques are valuable for customer segmentation with unsupervised datasets.

Utilize insights from both methods to create effective marketing strategies.

Do not use the method results for direct comparison; use them as a guide for tailored marketing strategies on specific customer segments.

“We see our customers as invited guests to a party, and we are hosts. It’s our job every day to make every important aspect of the customer experience a little bit better.”

~ Jeff Bezos

