MARKETING CAMPAIGN ANALYSIS
CUSTOMER SEGMENTATION

KAITLIN BENNETT

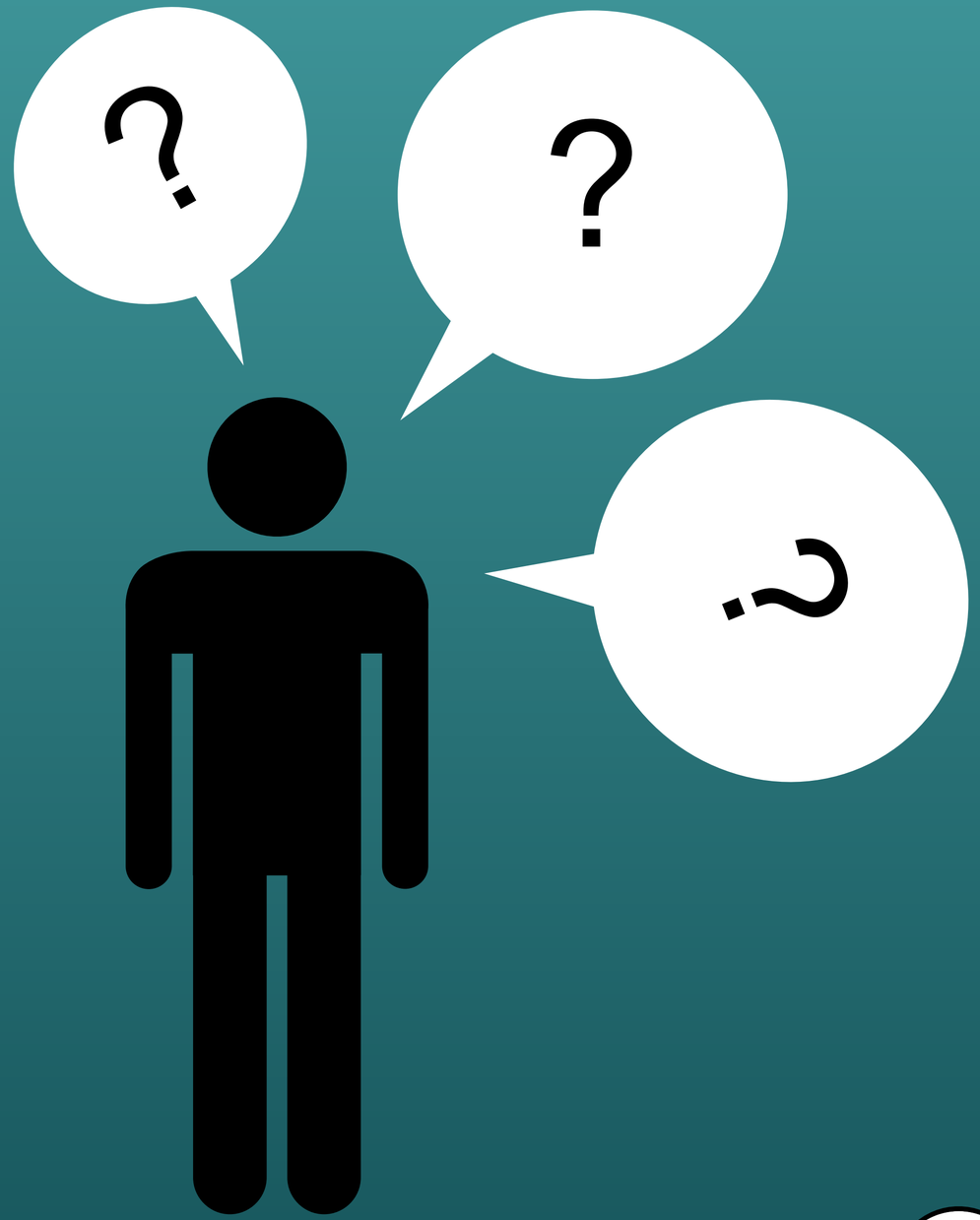
4/15/23

PROBLEM DEFINITION

We need to segment the data that we have on our customers into groups that show their commonality.

A good understanding of customer behavior and defining characteristics is key in order for marketing strategies to be successful.

Customer segmentation is an effective tool to maximize the effect of marketing campaigns and make better use of business funds

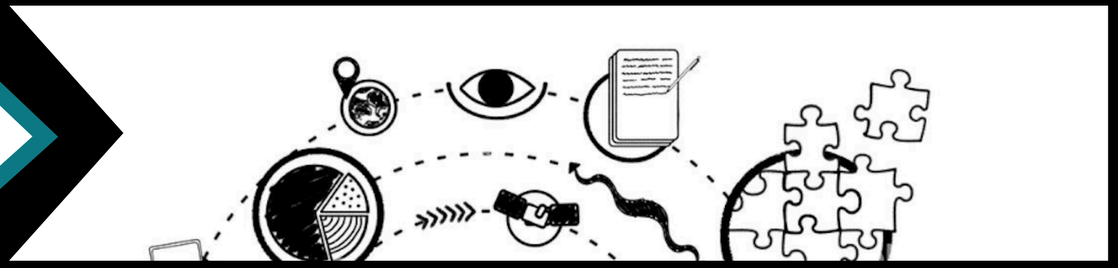


OBJECTIVES

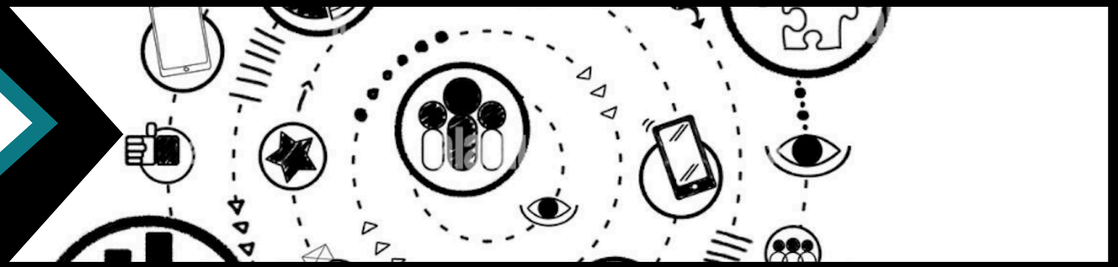
- Identify defining characteristics and trends among customers
- Determine the variables that should be used for segmentation
- Choose the best possible customer segments to increase ROI

SOLUTION APPROACH

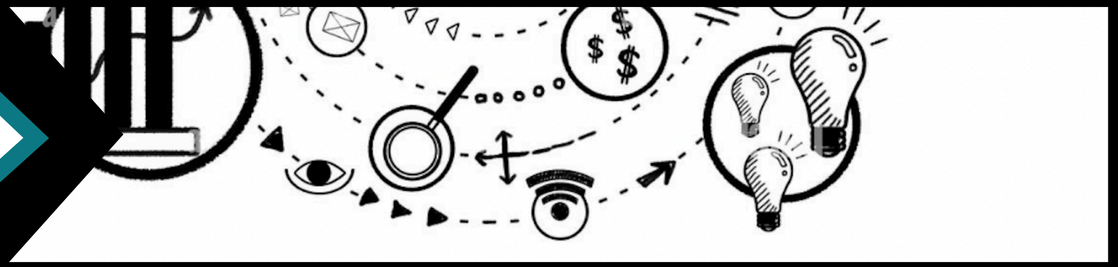
Exploratory analysis – Bivariate analysis



Feature engineering – Feature scaling

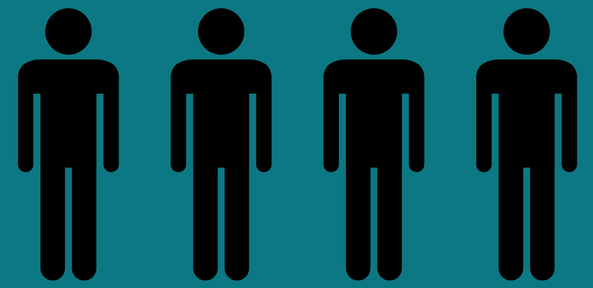


Test clustering methods – Profile clusters



PROPOSED SOLUTION METHOD

PCA for dimensionality reduction



Elbow method and Silhouette score to determine number of clusters

K-Means clustering method k=5

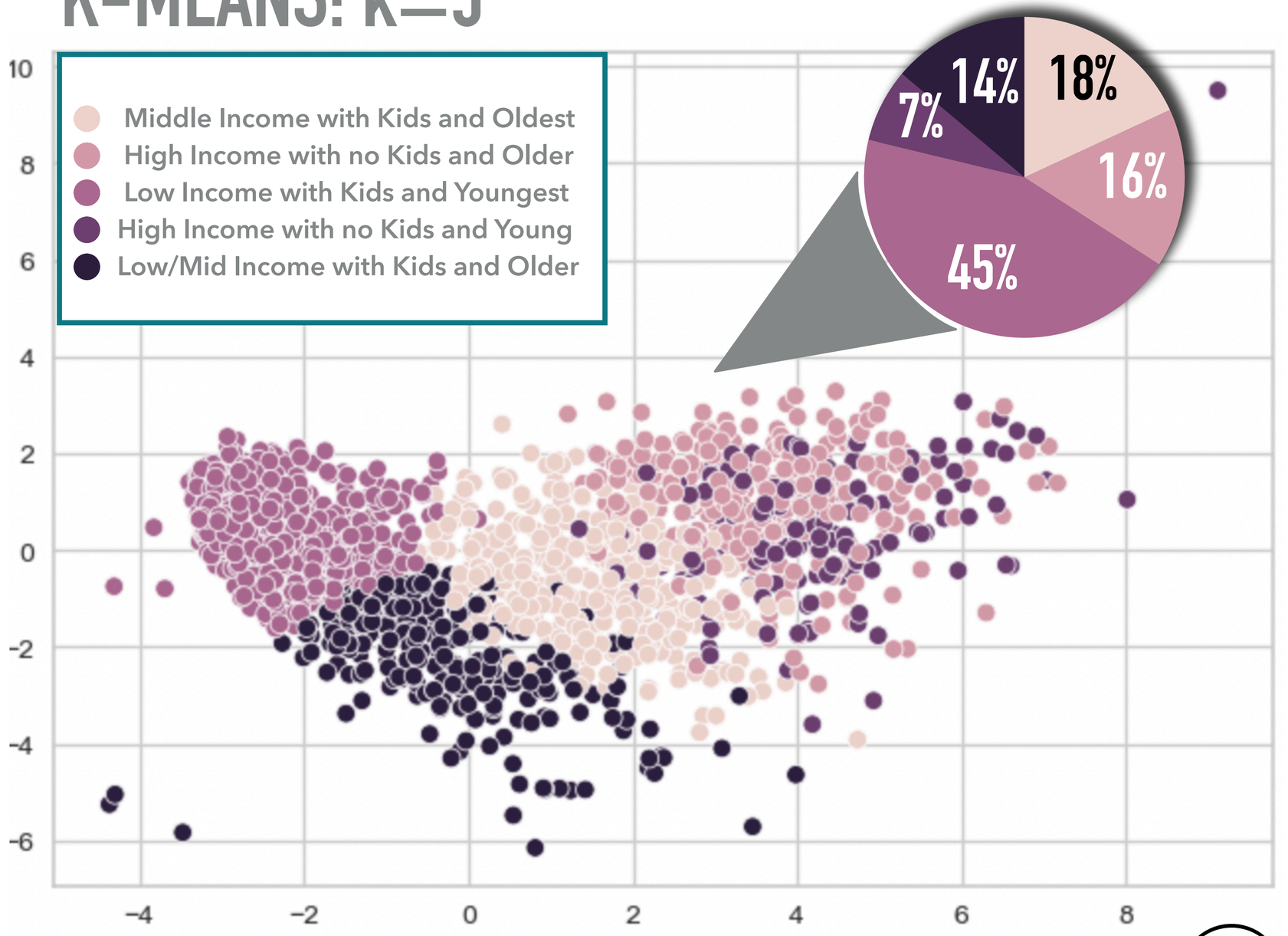
FINAL MODEL SOLUTION

Our best results came from using:

K-Means
k=5

This gave use a silhouette score of .24

K-MEANS: k=5



CLUSTER PROFILES

CLUSTER 0: MIDDLE INCOME WITH KIDS-OLDEST

- 403 customers aged 50 with teens
- Shops in store and through the website and makes lots of purchases
- Responded best to campaign 4

CLUSTER 1: HIGH INCOME WITHOUT KIDS- OLDER

- 357 customers aged 48 without children
- Shop in store and through the catalog but not the website. Not using deals
- Responded best to campaigns 1 and 5

CLUSTER 2: LOW INCOME WITH KIDS- YOUNGEST

- 994 customers aged 44 with most children
- Lowest purchase amounts but look at the website 6 times a month
- Responded best to campaign 3

CLUSTER 3: HIGH INCOME WITHOUT KIDS- YOUNG

- 165 customers aged 46 without children
- Highest catalog purchases, lowest for deals
- Responded best to campaign 5 and best overall. **High Profile**

CLUSTER 4: LOW/MIDDLE INCOME WITH KIDS- OLDER

- 308 customers aged 49 with most teens
- Shop very little but visit the website 7 times a month, and are highest for deals
- Responded best to campaign 4

PROPOSED BUSINESS SOLUTION

FOR K-MEANS CLUSTERING METHOD K=5

Cluster 0: Recommendations systems for website purchases. Target items for teens through email marketing.

Cluster 1: Use the catalog and ads through the mail that are time sensitive to decrease recency.

Cluster 2: Email marketing and recommendations systems for website purchases targeting items for families with young kids.

Cluster 3: **HIGH VALUE CUSTOMERS**. Provide incentives for them to promote through a platform of choice. Advertise new high priced items on the catalog. Anticipate this groups needs.

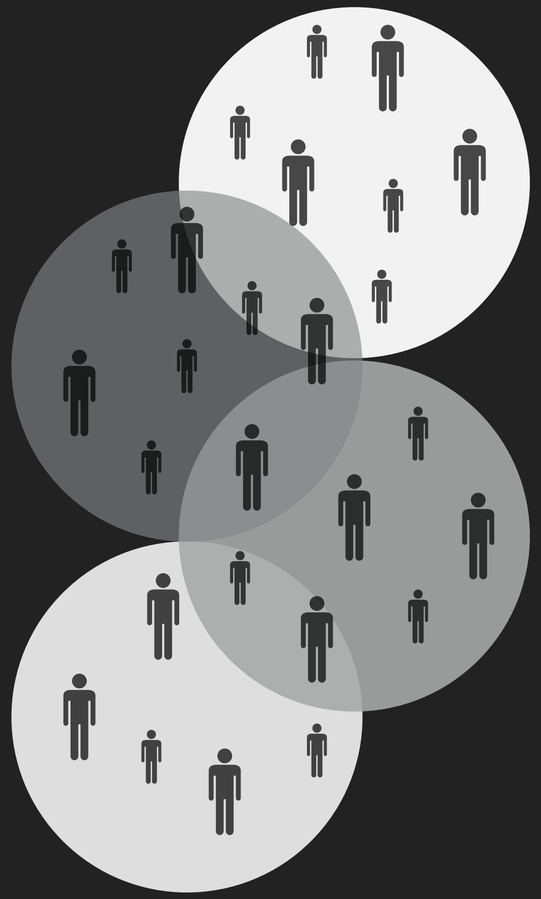
Cluster 4: Recommendation systems similar to cluster 0. Ads for less specialty items and more for larger families.

EXECUTING BUSINESS SOLUTION

Begin with tailoring marketing strategies for each cluster according to their behaviors and preferences.

New customer acquisition to update our dataset. Use marketing research and incentives.

Explore cloud services to run clustering algorithms and store our data.



EXECUTIVE SUMMARY

Unsupervised learning methods help to identify patterns in our clients dataset of customers, increasing ROI and maximizing marketing potential

Our most successful clustering method that provided insight into patterns and diverse profiling was K-Means with k=5 which had a silhouette score of 0.24

We are able to form conclusions from our clusters that can lead to more successful marketing campaigns and increase in sales and ROI.

RISKS AND CHALLENGES

- Production cost of marketing campaigns
- Changes in the market
- Customer characteristics can change


THANK
YOU
QUESTIONS?

APPENDIX

CLUSTERING METHOD SUMMARY

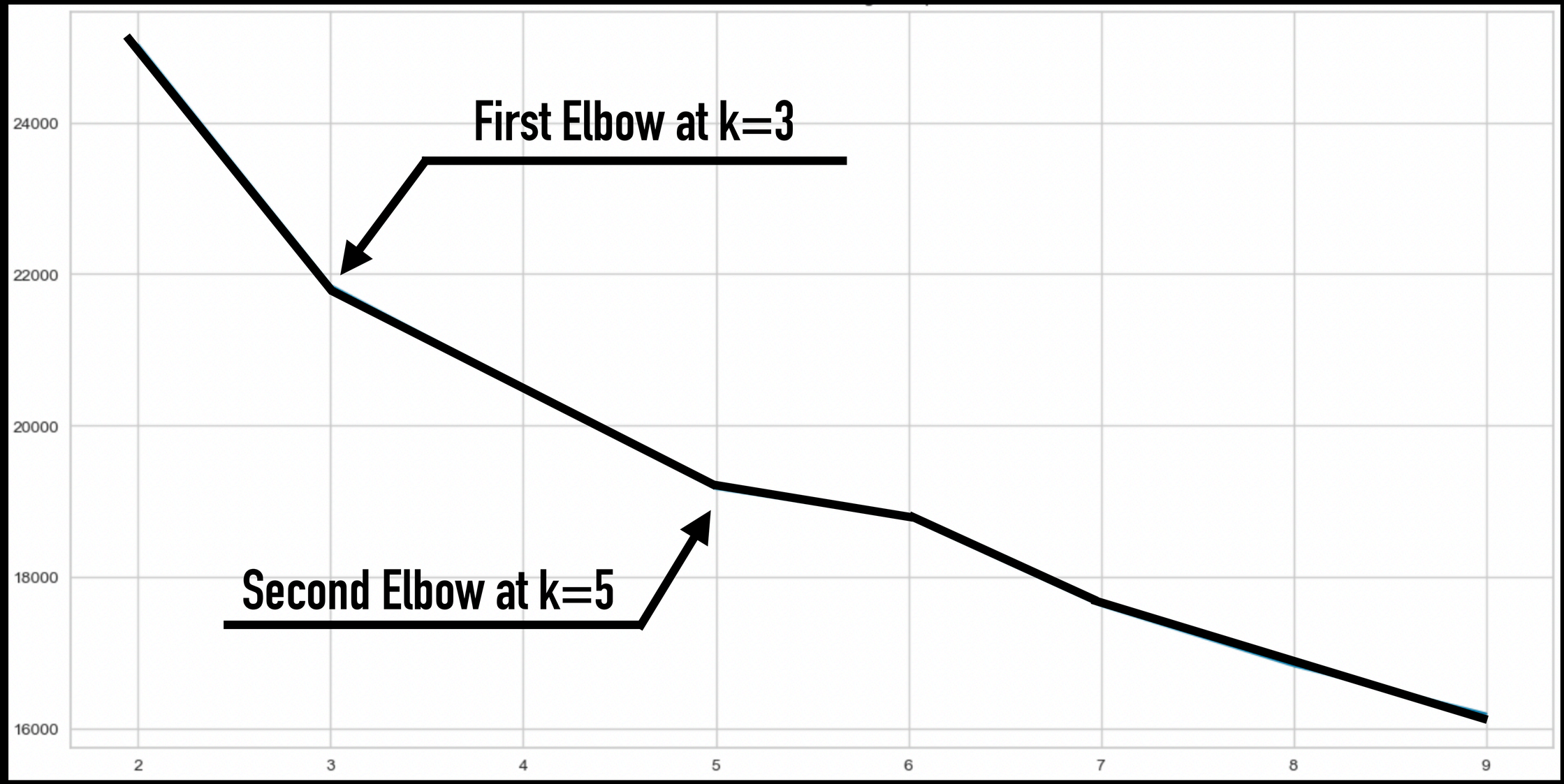
When choosing which clustering method to use we had to consider the nature of our data and our intended purpose.

Two methods stood out, K-Means with k=3 gave us the best silhouette score.

K-means with k=5 gave us more insight into our different groups.

Clustering Algorithm	Best Method	Silhouette Score	Key Points
K-Means	k=3	0.27	Most obvious elbow visible at k=3 and best silhouette score of all the models.
K-Means	k=5	0.24	Elbow also visible at k=5. These clusters gave better insight.
K-Medoids	k=5	0.11	Simliar clusters to K-Means
Hierarchical Clustering	Euclidean distance Ward linkage k=3	0.25	High silhouette score with k=3 but computationally expensive.
DBSCAN	eps=3 and min=20	0.34	Overlapping of clusters caused excessive grouping and high silhouette score.
GMM	k=5	0.14	Similar to K-Means but this method is better able to handle overlapping.

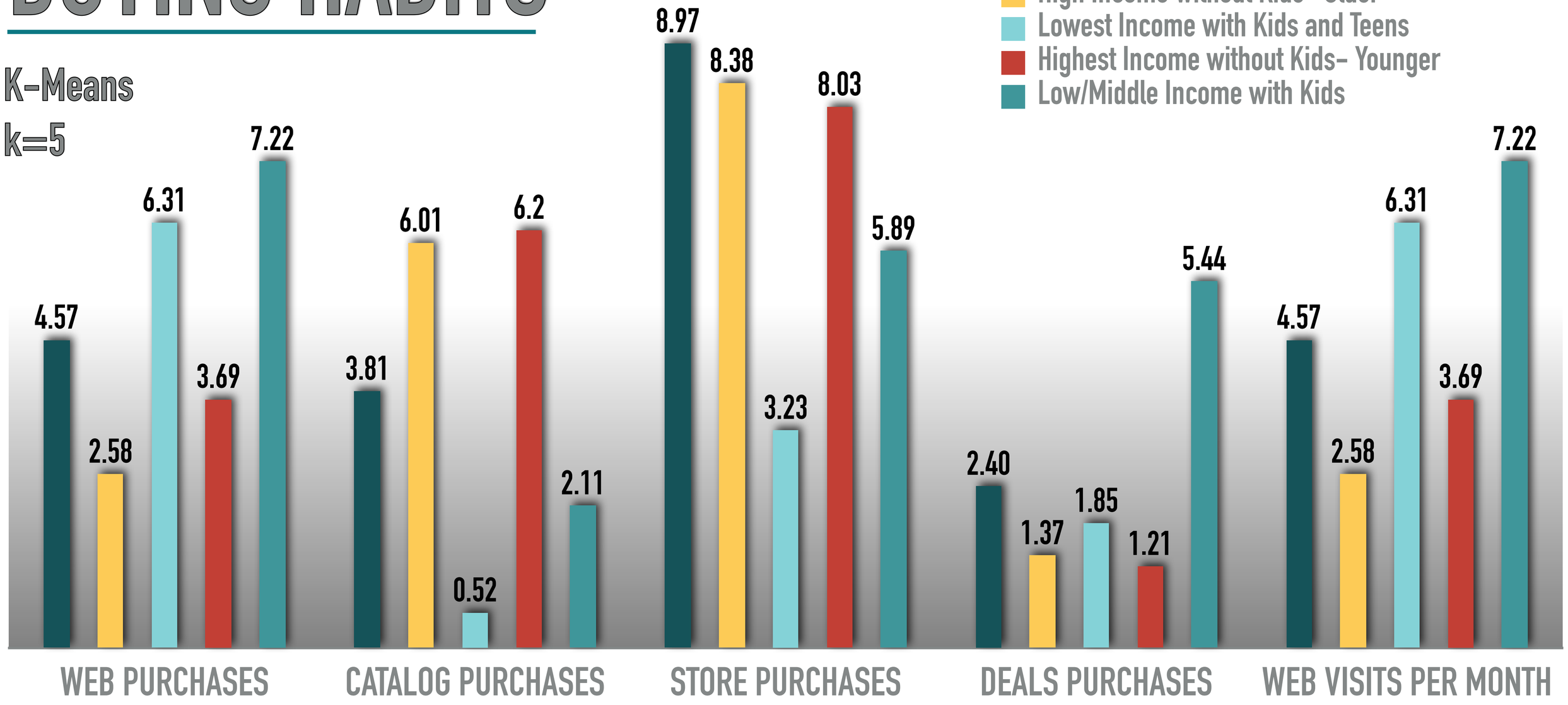
THE ELBOW METHOD SHOWING THE OPTIMAL K



BUYING HABITS

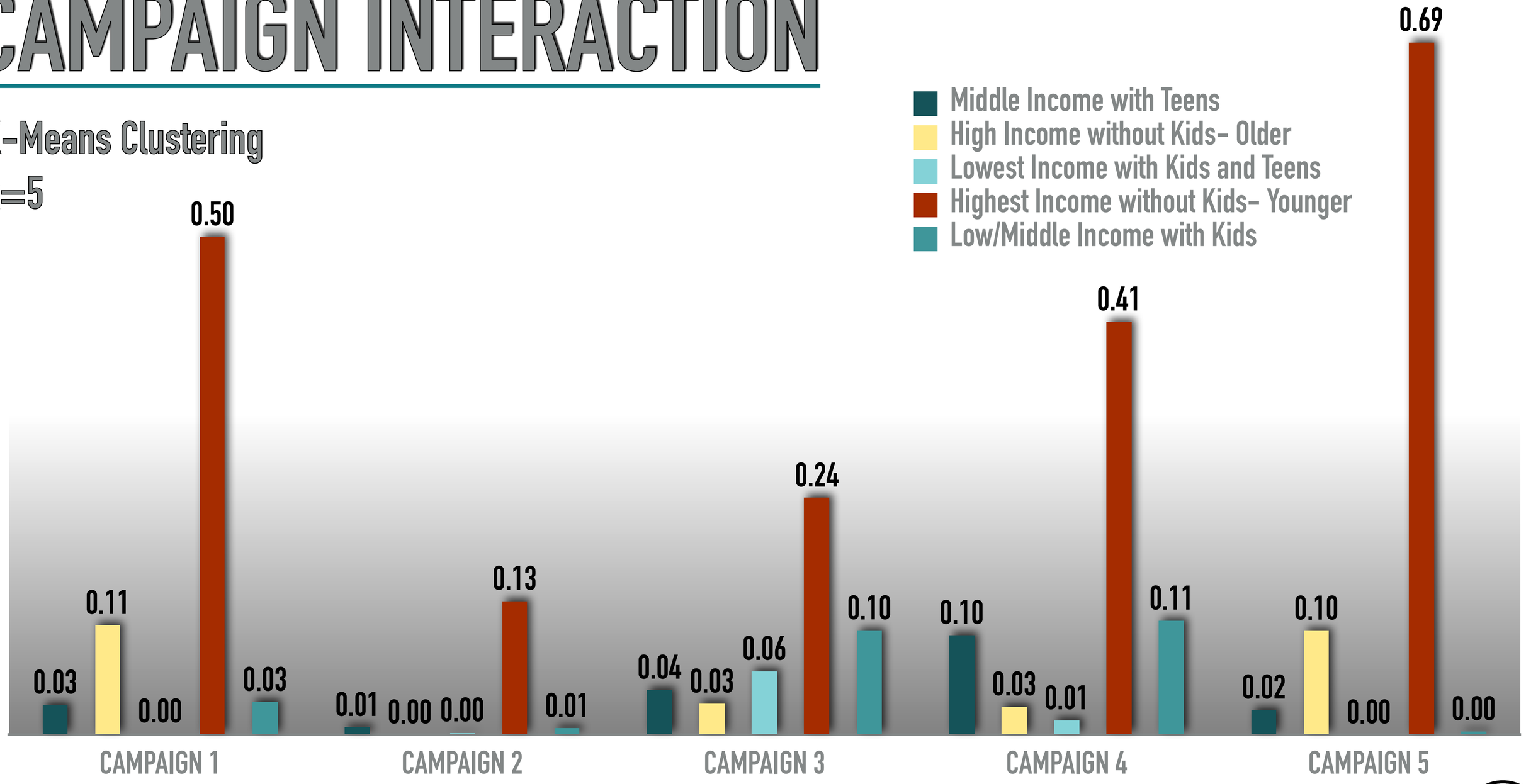
K-Means

k=5



CAMPAIGN INTERACTION

K-Means Clustering
k=5

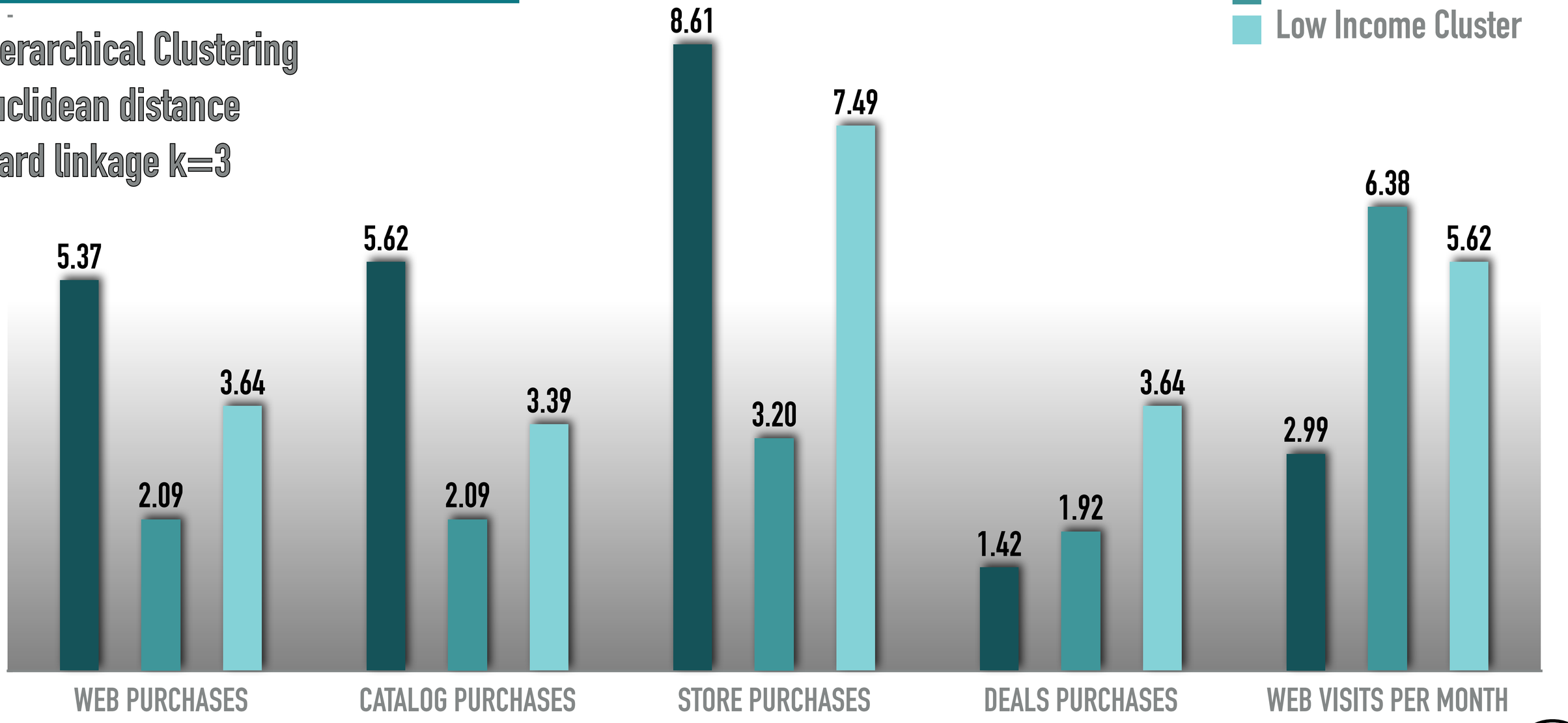


Problems to solve

- Are there problems with the data that we need to address before we begin our analysis?
- Which variables will be key factors in our clustering?
- Are all the variables necessary or can we streamline some?
- Are there any variables that we can create to make our analysis more efficient?
- Which clustering methods should we use and how should we determine their success?
- How can we best position our findings to improve our sales and marketing strategies.

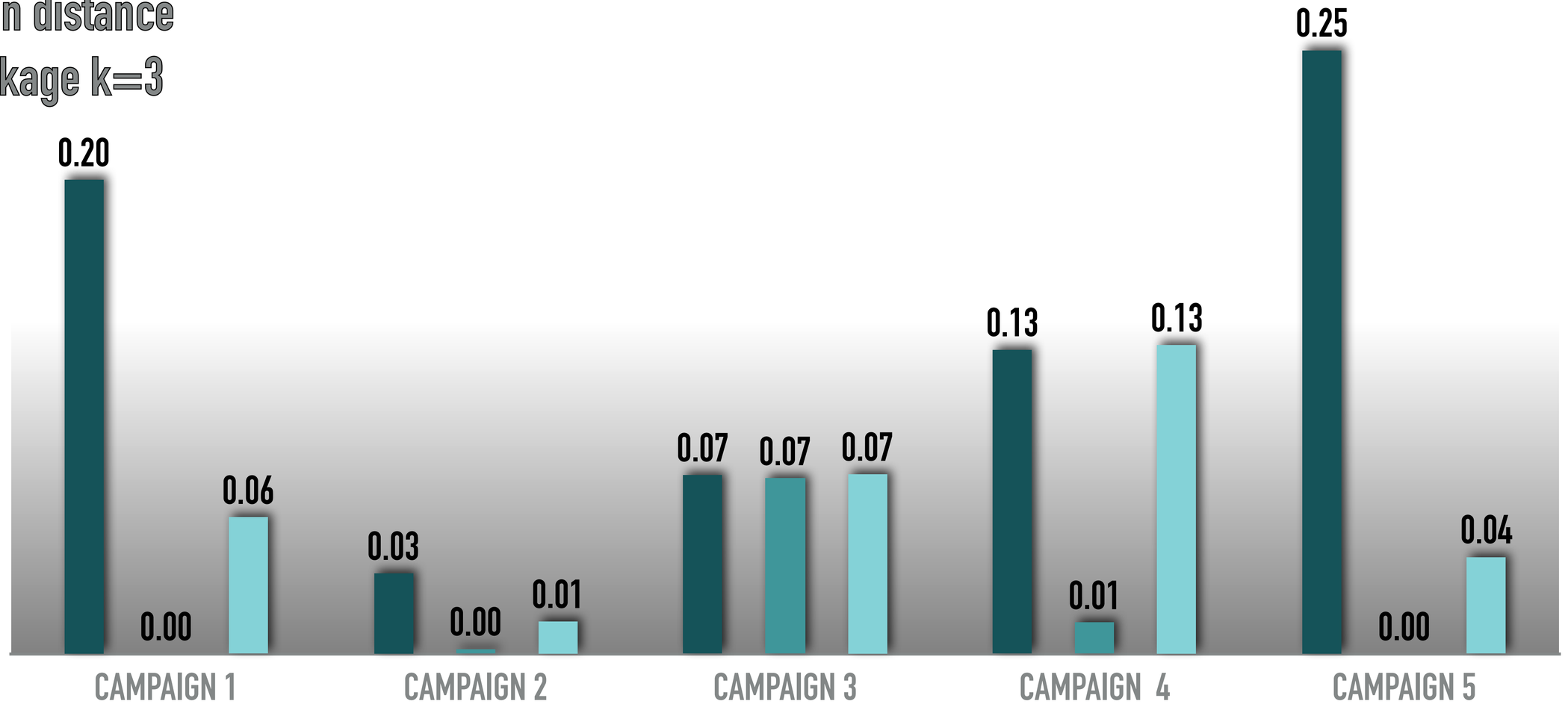
BUYING HABITS

Hierarchical Clustering
Euclidean distance
Ward linkage k=3



CAMPAIGN INTERACTION

Hierarchical Clustering
Euclidean distance
Ward linkage k=3



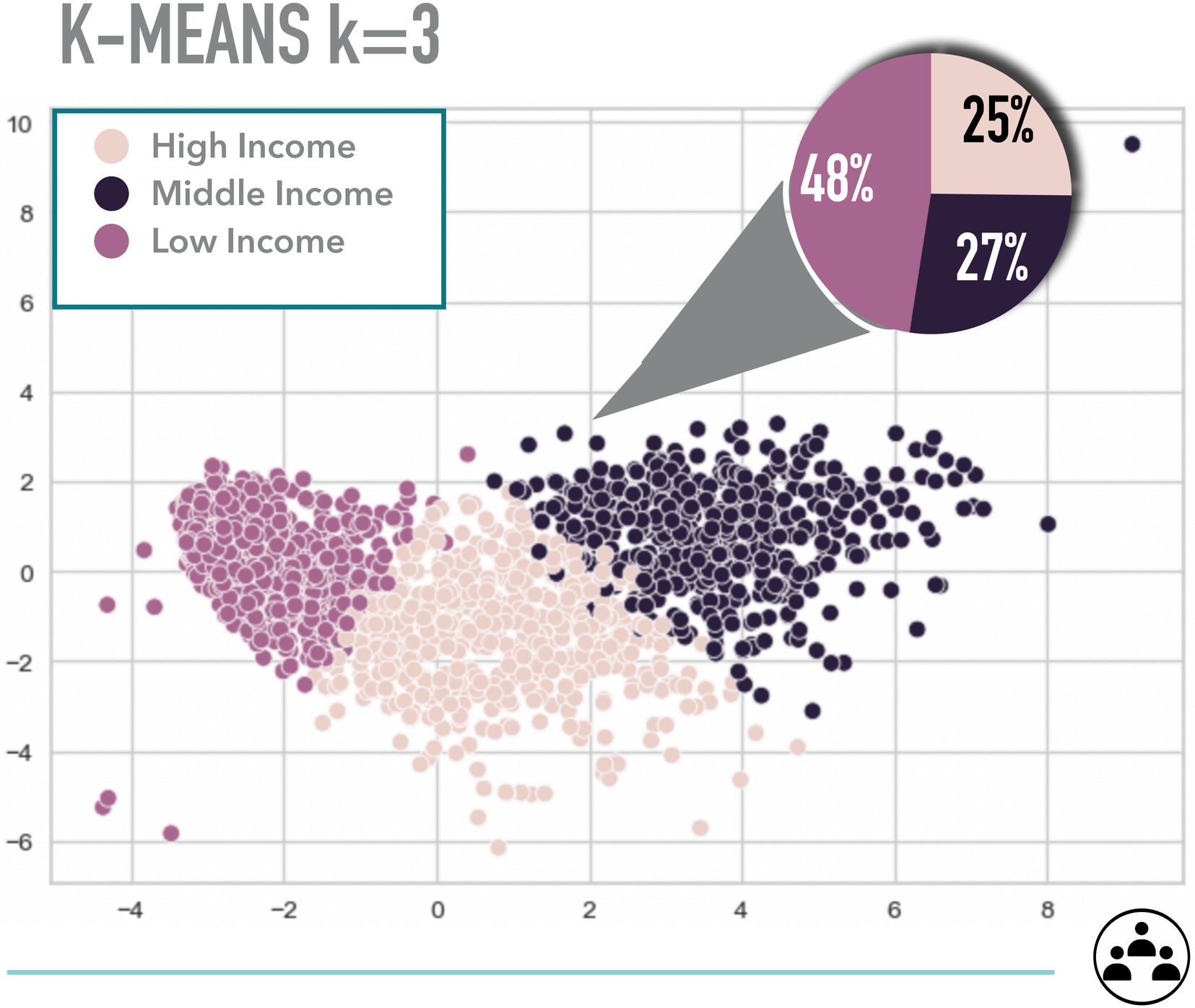
ALTERNATIVE MODEL SOLUTION

Our best results came from using:

K-Means

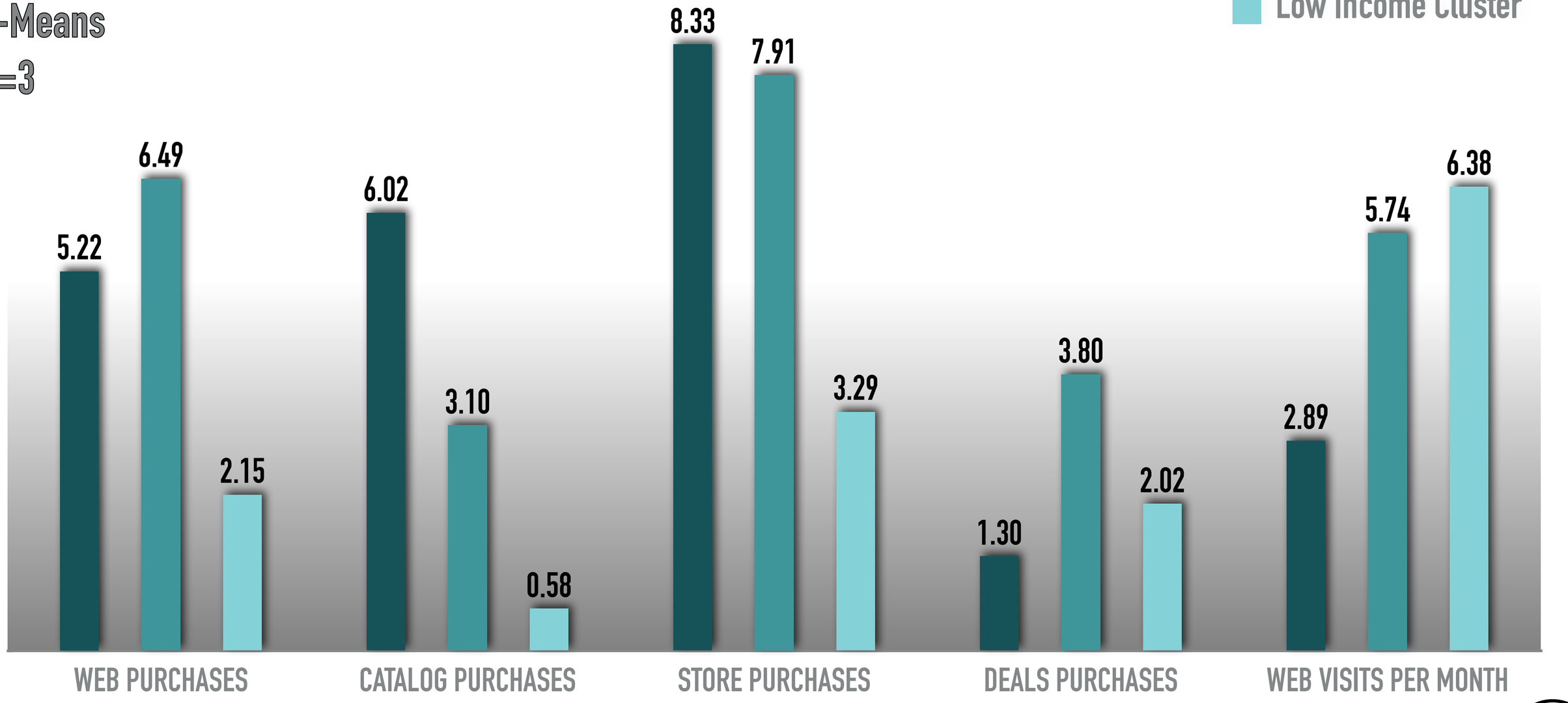
k=3

This gave use a silhouette score of .27



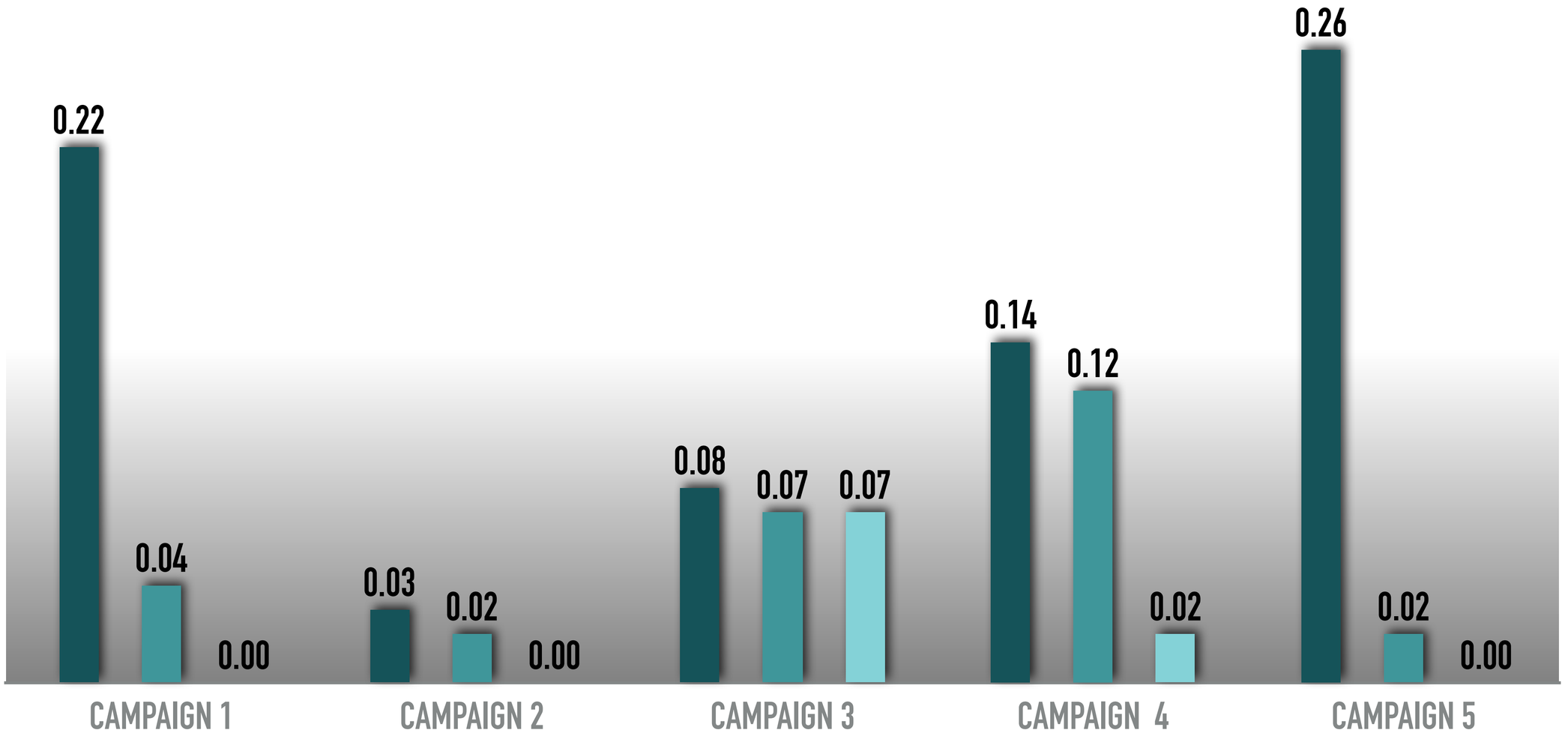
BUYING HABITS

K-Means
k=3



CAMPAIGN INTERACTION

K-Means
k=3



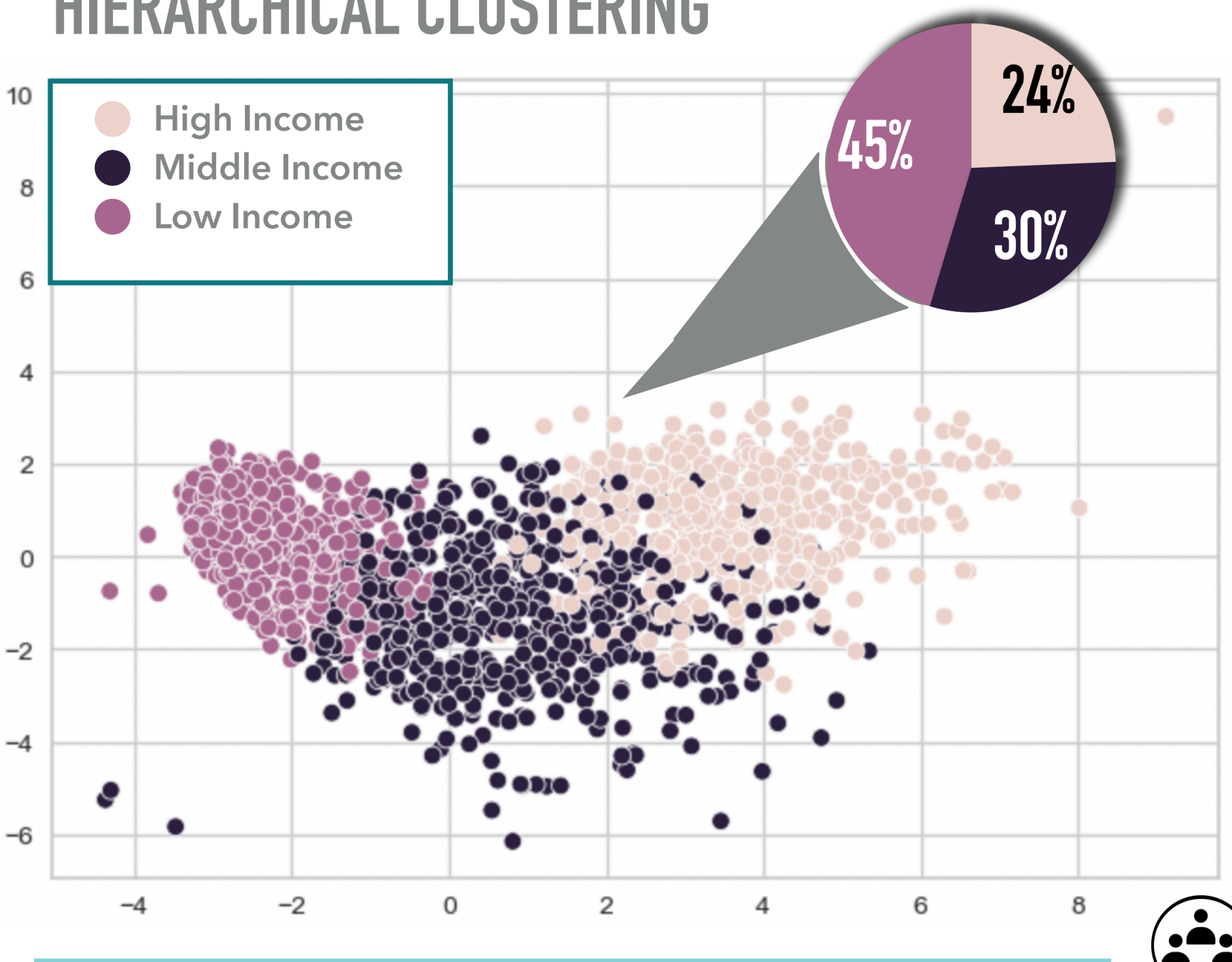
ALTERNATIVE MODEL SOLUTION

Our best results came from using:

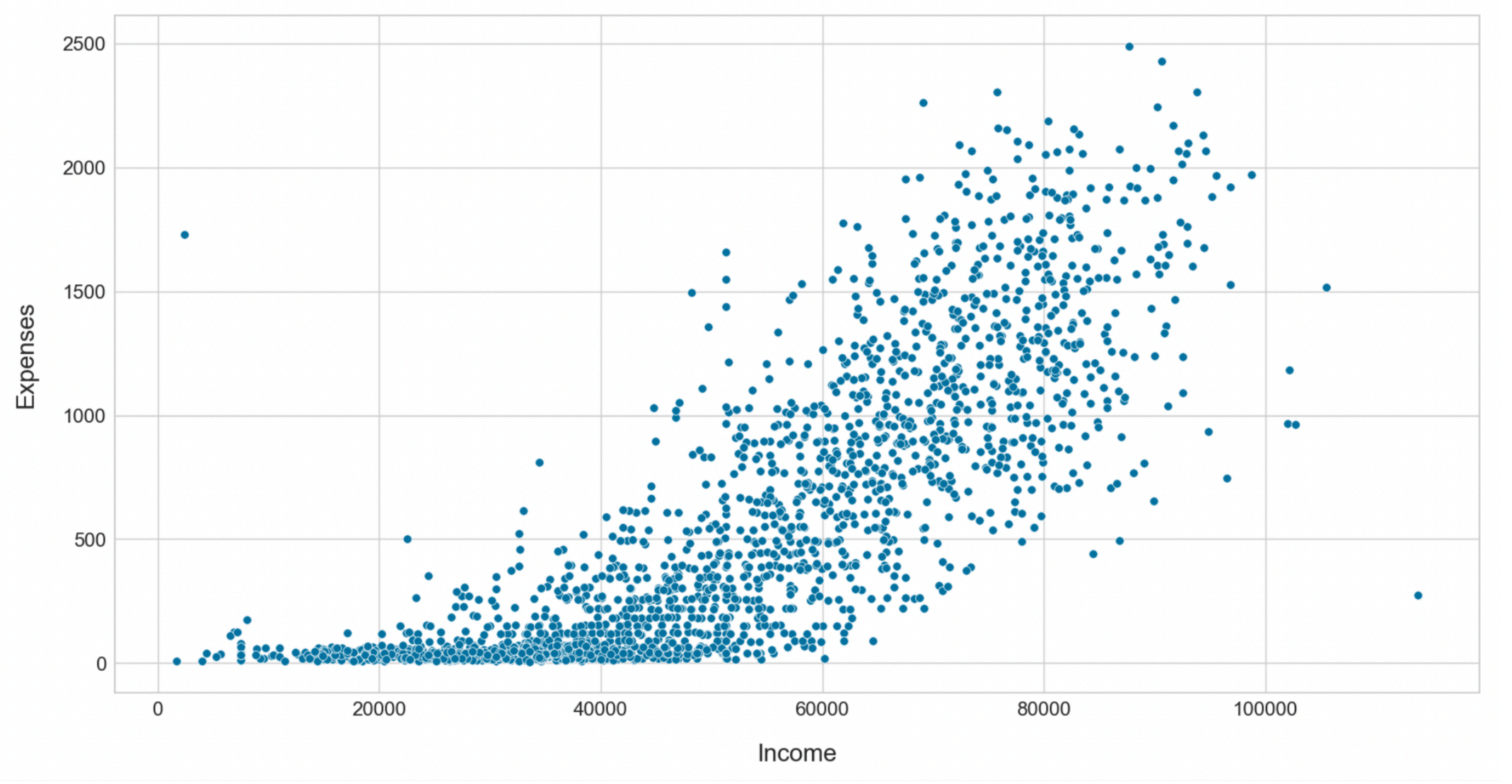
Euclidean distance
Ward linkage
k=3.

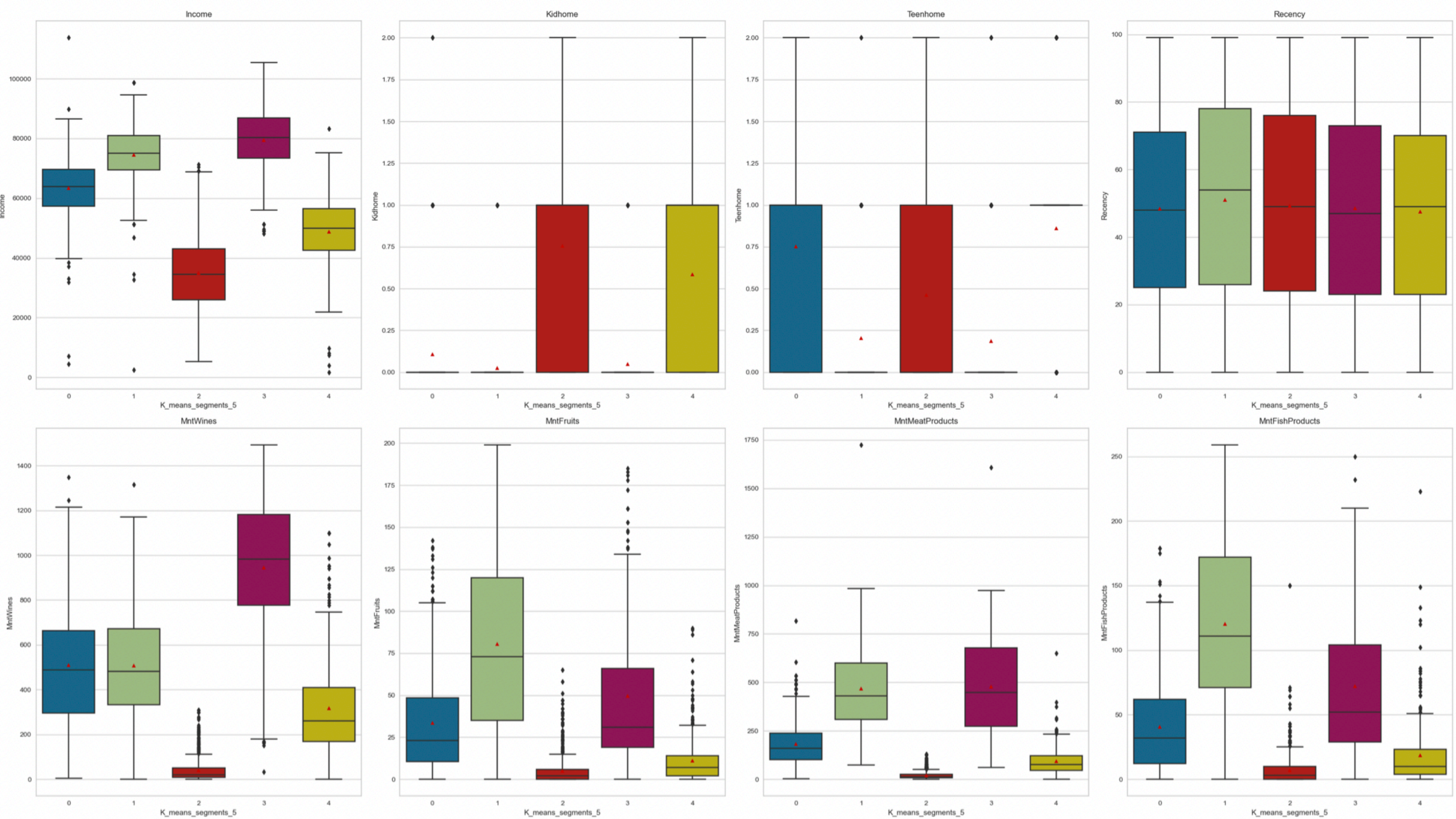
This gave use a silhouette score of .25

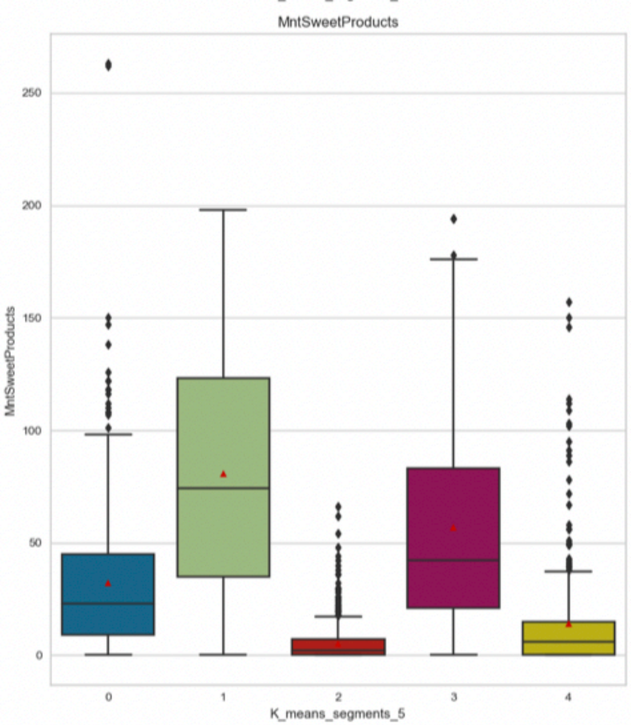
HIERARCHICAL CLUSTERING

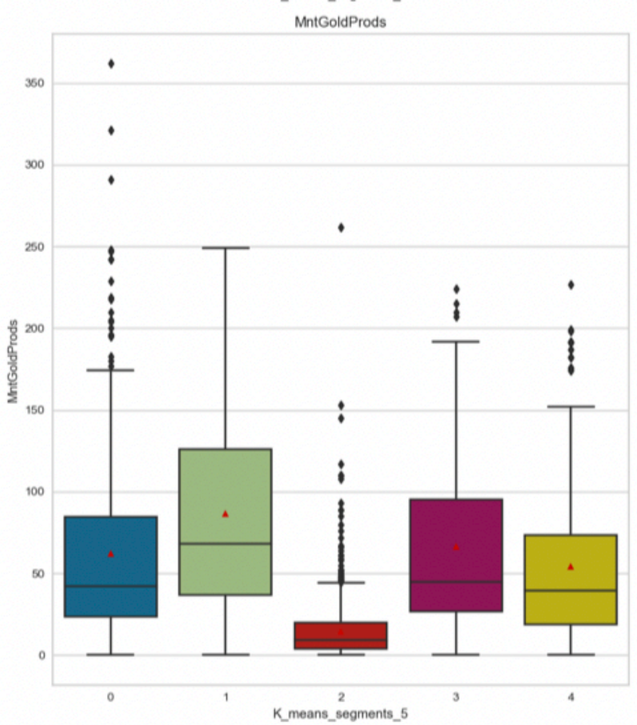


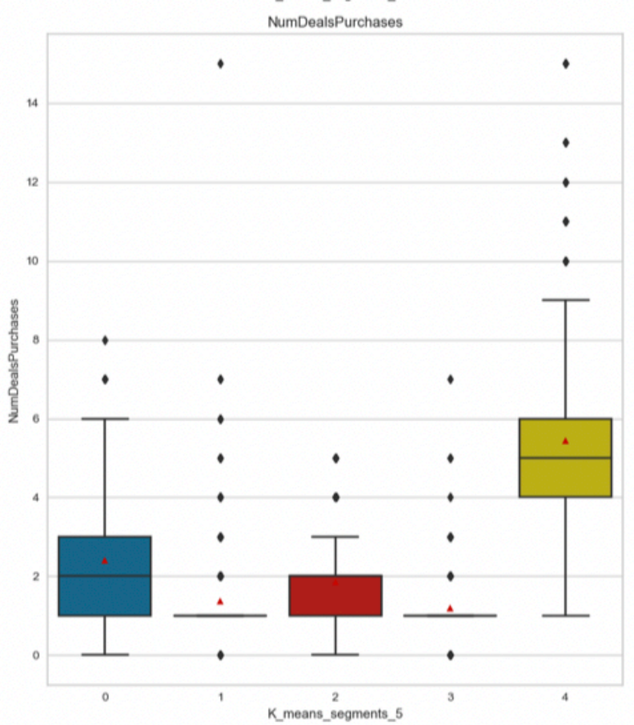
RELATIONSHIP BETWEEN EXPENSES AND INCOME

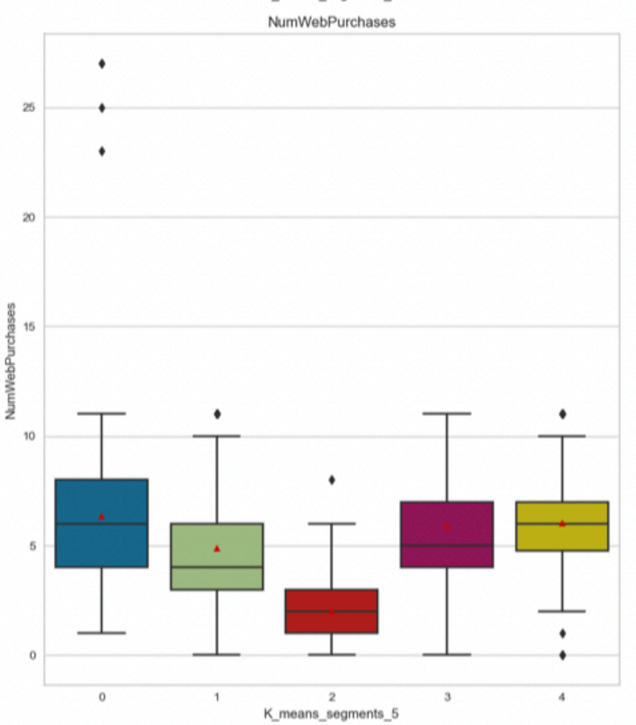


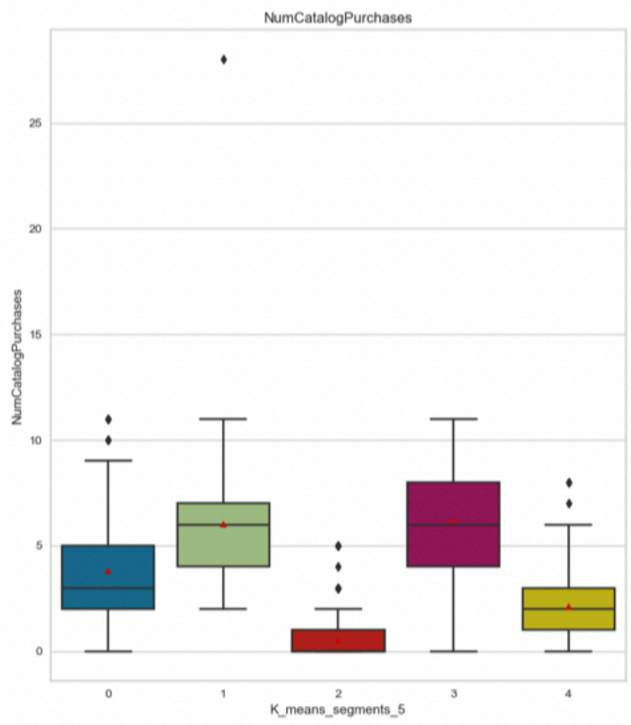

Income
Kidhome
Teenhome
Recency
MntWines
MntFruits
MntMeatProducts
MntFishProducts
K_means_segments_5

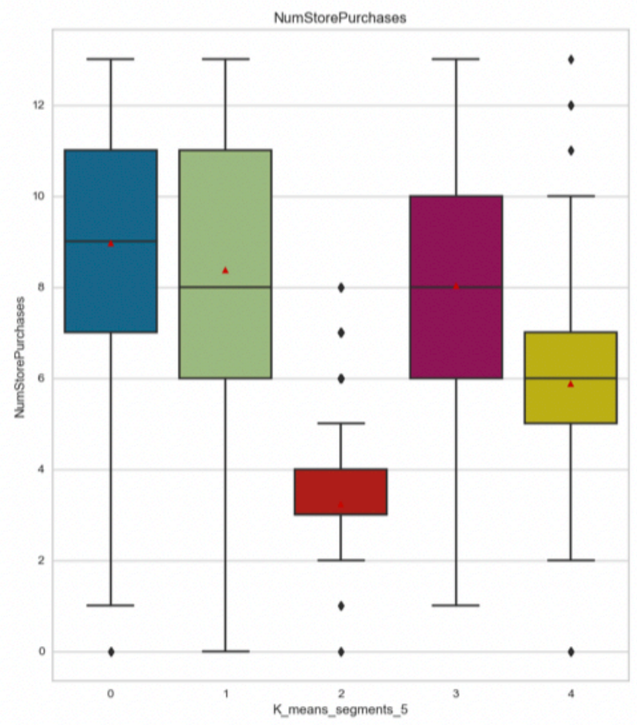

MntSweetProducts
MntSweetProducts
K_means_segments_5

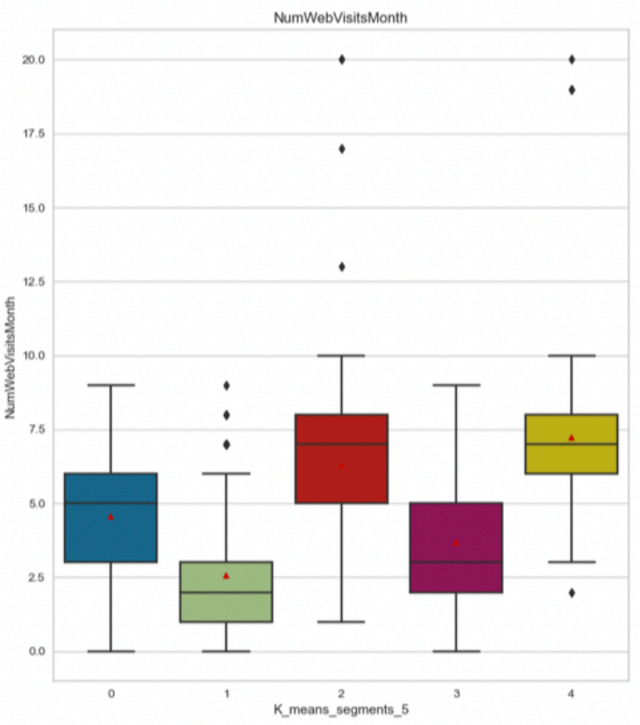

MntGoldProds
MntGoldProds
K_means_segments_5

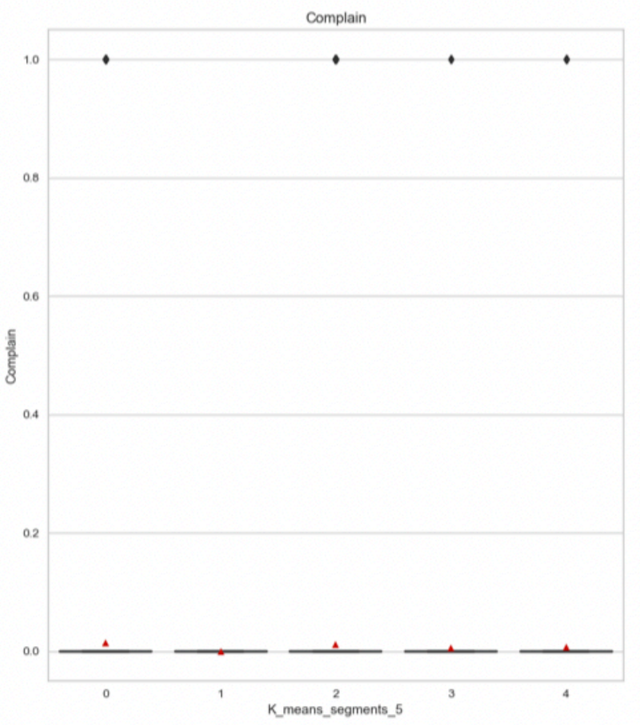

NumDealsPurchases
NumDealsPurchases
K_means_segments_5

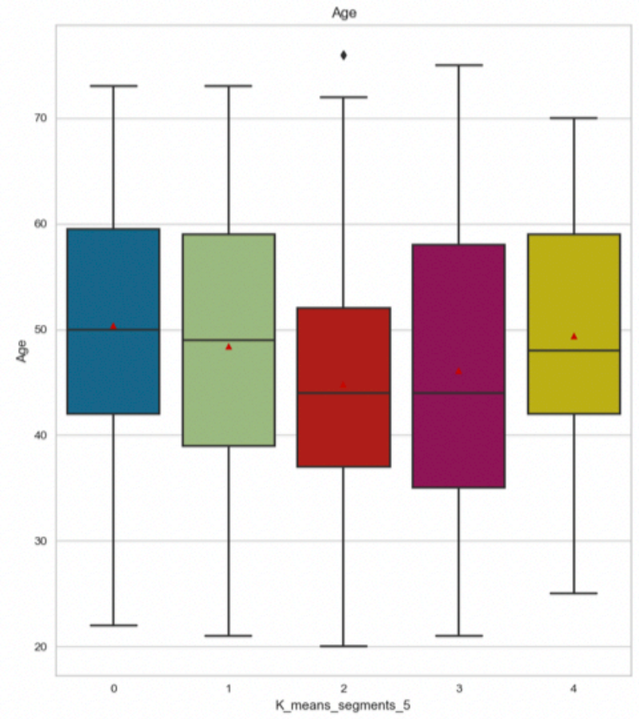

NumWebPurchases
NumWebPurchases
K_means_segments_5

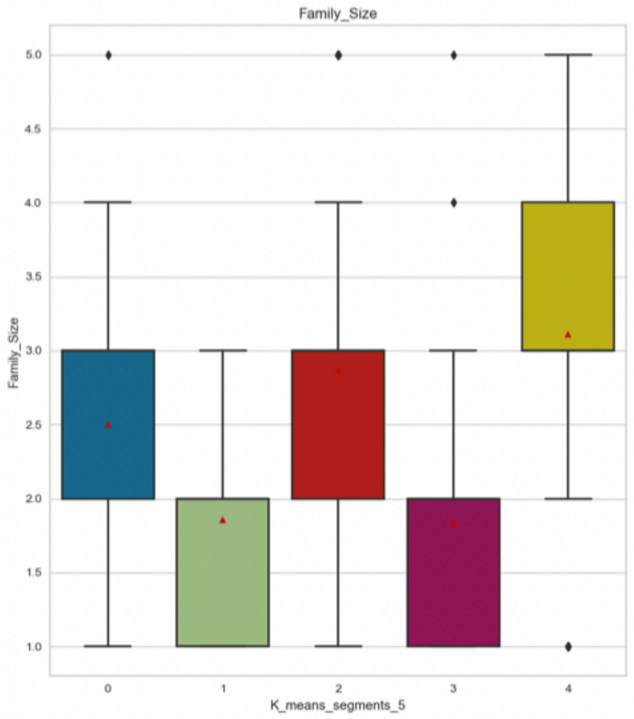

NumCatalogPurchases
NumCatalogPurchases
K_means_segments_5

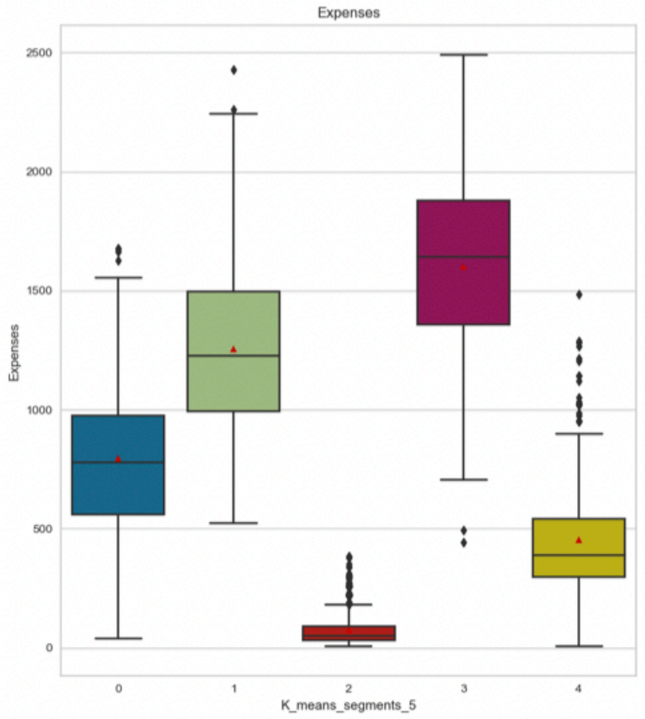

NumStorePurchases
NumStorePurchases
K_means_segments_5

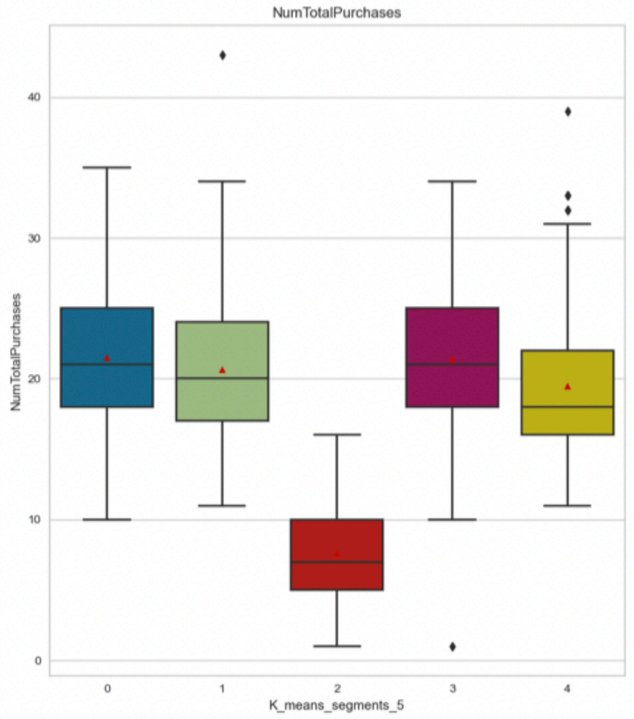

NumWebVisitsMonth
NumWebVisitsMonth
K_means_segments_5

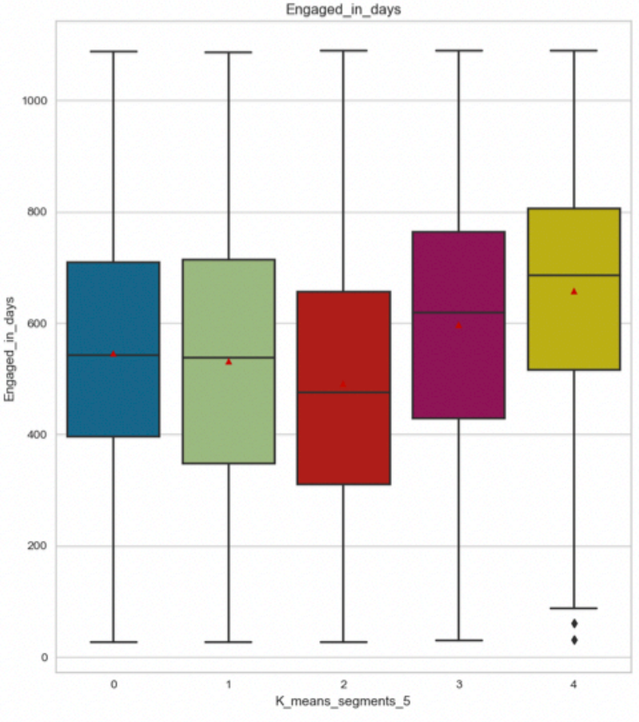

Complain
Complain
K_means_segments_5

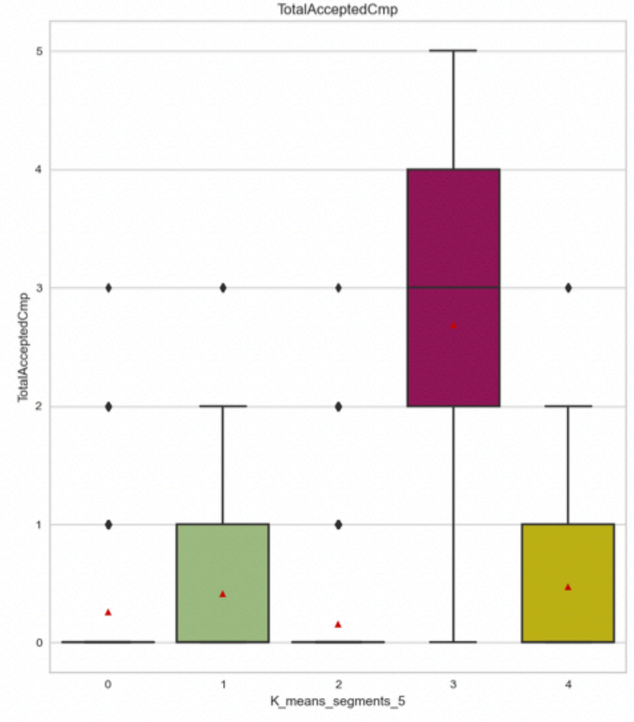

Age
Age
K_means_segments_5

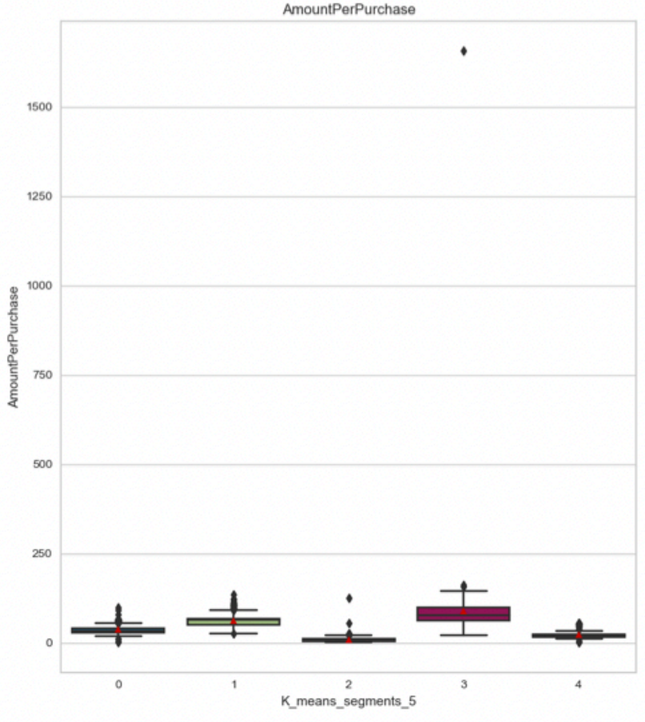

Family_Size
Family_Size
K_means_segments_5


Expenses
Expenses
K_means_segments_5


NumTotalPurchases
NumTotalPurchases
K_means_segments_5


Engaged_in_days
Engaged_in_days
K_means_segments_5


TotalAcceptedCmp
TotalAcceptedCmp
K_means_segments_5


AmountPerPurchase
AmountPerPurchase
K_means_segments_5