

Market Segmentation Summary

Executive Summary

This project proposes K-Medoids with $K=5$ for market segmentation of customers frequenting our fine foods grocery store between the years of 2012-2014 (data was collected in 2016). This model boasts a Silhouette score of 0.3392 as well as reduced sensitivity to the outliers than other models explored. However, it is subject to a number of limitations, including the fact that the data may be stale, as the training data set was collected in 2016. This was before the COVID-19 pandemic, which changed the way customers shop. It is suggested that post-COVID-19 data be collected and analyzed using the K-Medoids method in order to ensure the accuracy of market segments discovered during this project. Moreover, technology advances and changes at a rapid pace and customers may or may not change with that technology. This likely has impacted the way customers shop as well. Therefore, it is recommended that stakeholders consider these variables in building improved long-term segmentation algorithms.

Problem Summary

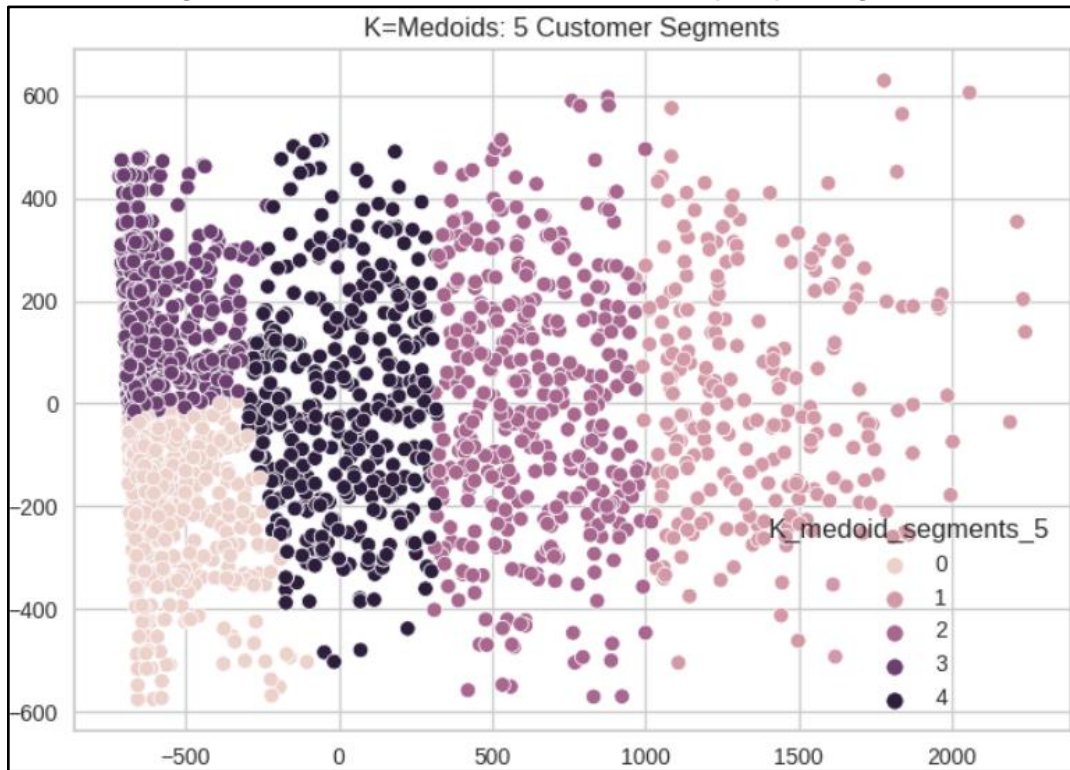
Dropping revenue is one of the most pressing challenges our company faces today. In just one year, from 2013 to 2014, revenue decreased by 61.8%. Out of the last six marketing campaigns, five campaigns had an average conversion rate of just 5.6%. As our competitors narrow in on personalized marketing strategies, it has become necessary for our company to do the same. The key objective of this project is to build a **clustering model for market segmentation of our customer base** in order to provide marketing with characteristics of various customer groups. As a result, marketing will be equipped with knowledge on how to construct personalized marketing campaigns to target specific customer segments. The clustering analysis presented here provides insight into the aforementioned customer group characteristics, such as demographics, purchase amounts, engagement activity and purchasing channels. The discussed analysis will serve as a basis for marketing recommendations for future personalized campaigns.

Solution Design

A number of clustering methods were explored as part of the solution design. These include K-Means, K-Medoids, Hierarchical Clustering, DBSCAN, and Gaussian Mixture Model. The final proposed solution is the **K-Medoids model** using the 2016 data. The 2016 dataset was the data available to the data science team.

Figure 1 shows the best model, which yielded the highest Silhouette score for each model that used a Silhouette score (0.3392 for training dataset). The model was effective in clustering the observations into different customer segments and providing insight into purchasing behavior.

Figure 1: Best overall model K-Medoids (K=5) using 2016 data



From the boxplots in Appendix 1, customer segments become more clear as we analyze the statistics for each segment. Characteristics of the 5 clusters identified by K-Medoids include:

Table 1: K-Medoids Key Characteristics of Customer Segments

Customer Segments		Demographic Averages			Purchase Averages			Purchase Channel Preferences	
Segment #	Segment Name	Age	Income	Family Size	Total # of Purchases	Purchase Amount	Product Most Purchased	First Preferred Channel	Second Preferred Channel
0	Budget-Conscious Shoppers	45	\$33,631	2.9	9	\$12.44	Gold	Store	Web
1	Wine Connoisseurs	48	\$77,802	1.8	1778	\$92.91	Wine	Store	Catalog
2	Cooking Enthusiasts	50	\$69,836	2.2	1145	\$54.84	Most Goods	Store	Web
3	Convenience Seekers	46	\$37,434	2.9	97	\$10.82	Gold	Store	Web
4	Foodie Adventurers	50	\$58,281	2.7	611	\$30.62	Fish	Store	Web

The Gaussian Mixture Model was also tested (see Appendix 2). While the GMM produced very similar customer segment results as the K-Medoids model, it yielded a lower Silhouette score for the test data (0.1023 for training dataset). That is because GMMs are more sensitive to outliers than K-Medoids models. Additionally, GMMs can be more computationally expensive than K-Medoids as K-Medoids is a partitioning algorithm that is less computationally demanding than GMMs. By selecting the K-Medoids model, we don't lose any important information that may be gleaned from the GMM and there is more computational efficiency with this model. Moreover, K-Medoids can be more scalable than GMMs. As the customer base grows, we will want the clustering algorithm to grow with it.

The Hierarchical Clustering model (Agglomerative Clustering) produced a Silhouette score of approximately 0.339 when $K = 3$ (see Appendix 3). While it appears to be a robust clustering model based on Silhouette score, the clusters are grouped by high, medium, and low incomes. This does not provide deep insight into customer purchasing behavior like the K-Medoids model provides. Therefore, the K-Medoids model with $K=5$ was selected.

Analysis & Key Insights

From the selected model, we observe the following characteristics of each segment:

Budget-Conscious Shoppers: These consumers have the lowest average income out of the five segments. The average family size for this segment is about 3 people. Only one other segment (Convenience Seekers) have an average family size of 3 people; the rest have fewer than 3 people. They typically purchase items when there is a discount or deal happening, and this is reflected in the average amount of expenses for this segment. These customers visit the store website more often than 80% of the other segments. This may indicate they search the web for deals before going to the store to make a purchase. This segment spends the most on wine, followed by meat, gold, fruit, fish, and sweets.

In summary, customers in this segment appear to be young and middle-aged parents with children or teens. They are on tight budgets due to lower incomes and larger family sizes, and because of this they are most interested in scoring good deals on fine foods. This segment is not as likely to accept marketing campaigns.

Wine Connoisseurs: These consumers have the highest average income of the five segments. The average family size for this segment is around 2 people. The majority of these customers do not have minors (children or teens) in the home. Perhaps it is because of this that the item they spend the most on is wine. Of the five segments, this segment spends the most on wine. This segment also spends much on meat, fish, fruit, sweets, and gold. This may be because they have the most money to spend and the fewest number of family members to take care of. This segment uses the fewest deals, indicating they aren't especially concerned with pinching pennies or saving money. As for purchase channels, it appears these customers prefer in-store purchases, but they also purchase through catalogs and the web. In summary, the customers in this segment appear to be middle-aged individuals in a relationship but with few or no children.

In summary, these customers make decent money and do not have large families to take care of. Because of this, they are more interested in expensive, fine foods and wine irrespective of price. This segment is most likely to accept marketing campaigns.

Cooking Enthusiasts: This segment spends the second to highest amount on all goods except gold products. For gold they are tied with the Wine Connoisseurs segment. These customers have similar demographics to the Wine Connoisseurs, with some variation in precise numbers such as income, expenses, etc. This segment has a slightly larger family size (consisting mostly of teens) than the Wine Connoisseur segment. They also appear to be slightly older, on average, than the Wine Connoisseurs. Because this segment does not have young children, because they earn high incomes, and because they spend the second to most on food products, they likely have more time to cook and enjoy fine foods.

In summary, customers in this segment have higher incomes, are willing to spend money on fine foods but also look for deals and discounts, have the highest average in-store purchases, the highest web purchases, and are likely to accept campaigns.

Convenience Seekers: This segment consists of middle-aged customers who have both children and teens in the home. They earn a little more income than Budget-Conscious Shoppers and they also have a slightly larger average family size. They have the lowest number of purchases and the least amount of purchases. This may be because they are less interested in the fine wine and foods this particular grocery store specializes in. This customer segment may be extra busy with children and teenagers, and they may find it more convenient to shop at a supermarket (such as Walmart) where they can purchase items for children and teens rather than only for adults. This fine foods grocery store does not carry baby food, diapers, and kid-friendly foods like macaroni and cheese. Rather, this store retails high-end wines, meats, and gold products. However, we do see this segment has the second highest average web visits per month, indicating the parents in the family do appreciate going to this store, perhaps when there is a good deal happening.

In summary, this segment consists of busy parents who may value the convenience of supermarkets over the luxury of fine foods, but they can be persuaded to purchase if there is a good deal going on.

Foodie Adventurers: The customers in this segment average \$60,000 in income per annum. This is the third highest income in the segments. They have one of the highest average ages (around 50 years) and the third largest family size. They have the third most children in the home and the most teens in the home (on average). These folks purchase the most deals. They enjoy fine wines, meat, fish, fruit, gold, and sweets (listed in order of most dollars spent on the product category). Their preferred purchase channel is in-store, but they also don't seem to mind purchasing on the web.

In summary, this segment appears to consist of older parents with many of their children grown to teenagehood. Because of this, they may feel more free to purchase fine foods as their teens are less likely to waste the food or turn up their noses at it and opt for Frosted Flakes or PB&J instead.

Limitations & Recommendations for Further Analysis

Limitations: The potential risk of the proposed solution design was the quantity of data available for the data science team. Only 2,240 observations were provided, and of those 8 observations were dropped due to outliers. There may need to be additional data needed to test the selected solution in an effort to see if similar results are attained. Another risk is stale data. The data was collected in 2016, pre-COVID-19. COVID-19 changed the way consumers shop. According to the National Institute for Health, "Prior to COVID-19, more consumers shopped in person, even if it was not preferred...During the pandemic, more consumers adopted online grocery shopping," (Bridges and Fowler). In order to be up to date with consumer trends and preferences, it is necessary to collect and analyze post-COVID-19 data. While each segment analyzed above prefers in-store shopping above other methods, this may have changed since COVID-19.

Recommendations: After thoroughly analyzing the data at hand and creating customer segments based on the K-Medoids model, the data science team recommends the following actions to the marketing team.

Budget-Conscious Shoppers: To capture the attention of this segment, we recommend constructing campaigns surrounding deals and discounts. This segment spends the most on wine, meat, and gold products. The campaign could emphasize deals on these products. The advertising channel that may work best for this segment is via the internet, as this segment visits the store website most often. Being of a younger age, the customers in this segment may feel more comfortable with technology and may look to technological sources for advertisements over printed advertisements. Therefore, we recommend electronic means of advertisements, such as emails, advertisements on the store website, and social media ads.

Wine Connoisseurs: To capture the attention of the classy folks in this segment, we recommend pushing the fine wines and high-end meats. The customers in this segment are approaching or have surpassed age 50, and many of them do not have children. In conjunction with this, we observe that this segment makes around the same number of purchases in-store as they do via catalog. Therefore, printed advertisements, such as mailers or ads in newspapers and magazines, may catch the eye of customers in this segment.

Cooking Enthusiasts: These folks spend more on food products than on gold products. They seem to appreciate deals and discounts. This segment also has the highest average number of web purchases and second highest number of catalog purchases. There are some overlapping qualities between Cooking Enthusiasts and the two aforementioned segments, so utilizing both the technological, deal-based advertisements and the printed, product-based advertisements may be beneficial for this segment.

Convenience Seekers: This customer segment shares characteristics of Budget-Conscious Shoppers. This segment has the most children and teens at home. It is also the segment that spends the least at the store. Constructing advertisements targeting parents of children and teens may help to drive up sales. Additionally, offering services such as curbside pickup may appeal to busy, working fathers and mothers. Further research and analysis on offering curbside pickup would need to be conducted; we recommend further research on this potential offering.

Foodie Adventurers: Like the Budget-Conscious and Cooking Enthusiasts segments, Foodie Adventurers seek deals. But unlike those two segments, this segment spends more at our store. This may be due to fewer children in the home and more disposable income to spend on fine foods. This segment seems to be comfortable with website navigation and also feels confident with purchasing via catalog. Therefore, pushing deals electronically and via catalog may be effective for this segment.

In addition to the above recommendations, we also recommend continuing to collect data on these segments. Continued collection of data will be useful in determining the effectiveness of these segmentation and marketing efforts.

Work Cited

Bridges, Elieen, and Kendra Fowler. "Grocery Shopping before, during and after the Pandemic: A Qualitative Study." *NCBI*, 3 August 2022, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9537917/>. Accessed 9 December 2023.

("Online Advertising Costs In 2021 | Top Draw")

("How Much Does Email Marketing Cost? Pricing Guide 2024")

Works Cited

Bridges, Elieen, and Kendra Fowler. "Grocery Shopping before, during and after the Pandemic: A Qualitative Study." *NCBI*, 3 August 2022, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9537917/>. Accessed 9 December 2023.

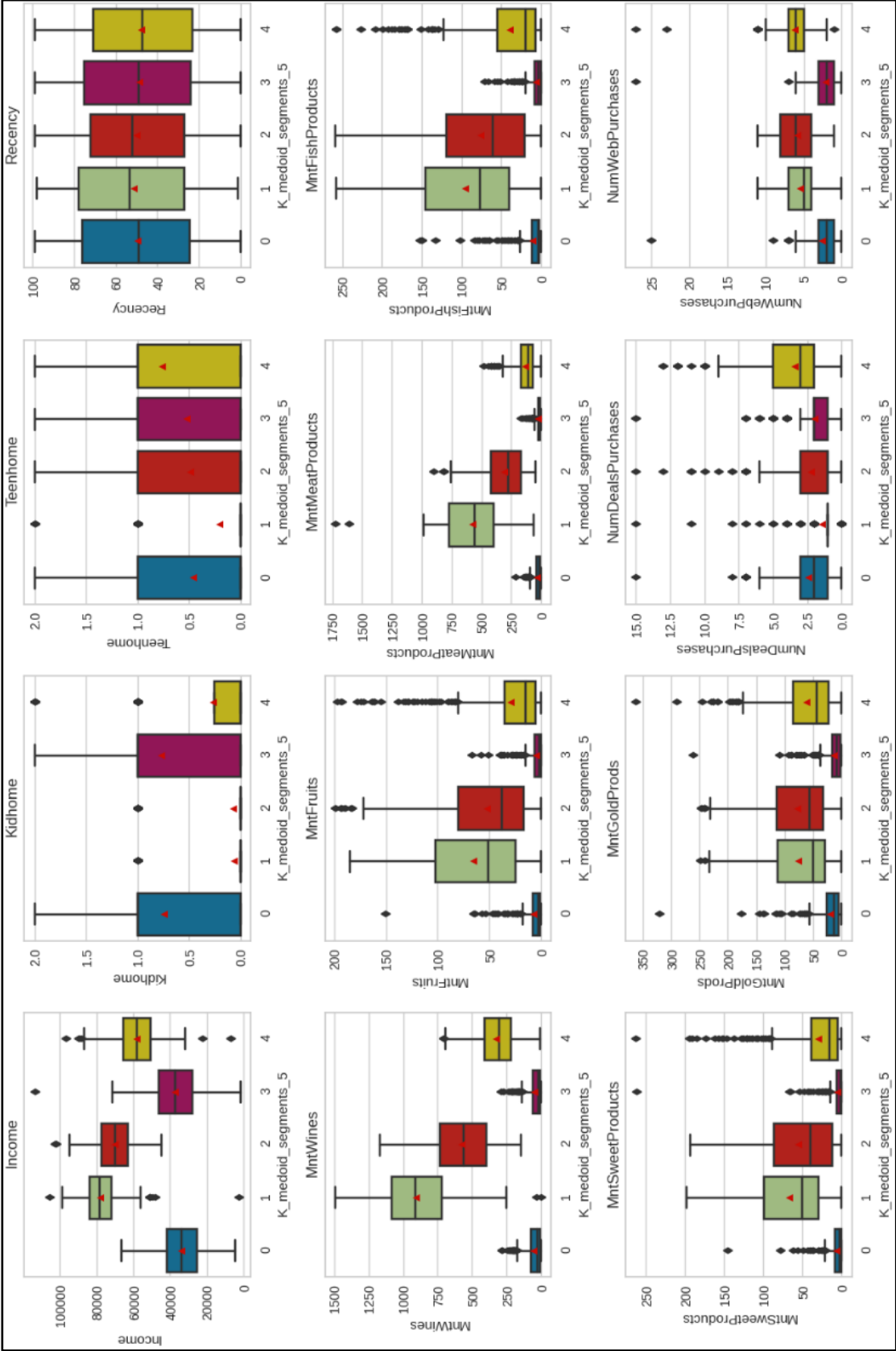
"Catalog Printing & Mailing (How Much It Costs & More)." *Conquest Graphics*, 3 May 2023, <https://www.conquestgraphics.com/blog/conquest-graphics/2018/10/10/how-much-does-it-cost-to-print-catalogs>. Accessed 15 December 2023.

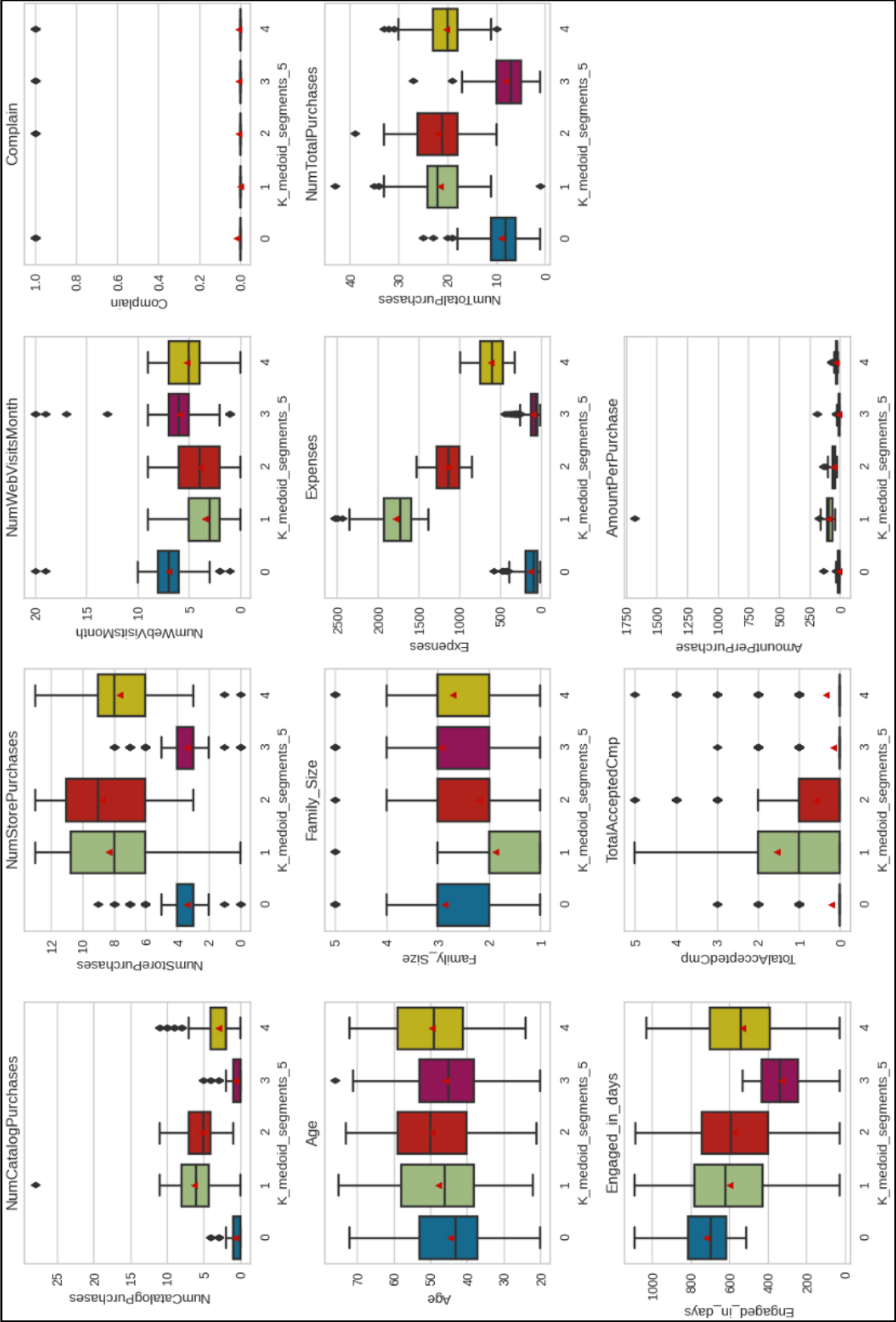
"How Much Does Email Marketing Cost? Pricing Guide 2024." *Email Vendor Selection*, 10 December 2023, <https://www.emailvendorselection.com/email-marketing-cost-pricing-guide/>. Accessed 15 December 2023.

"How Much It (Truly) Costs to Mail 1000 Flyers + How To Get Discounts." *Conquest Graphics*, 8 December 2023, [https://www.conquestgraphics.com/blog/conquest-graphics/2022/11/14/how-much-it-\(truly\)-costs-to-mail-1000-flyers-how-to-get-discounts](https://www.conquestgraphics.com/blog/conquest-graphics/2022/11/14/how-much-it-(truly)-costs-to-mail-1000-flyers-how-to-get-discounts). Accessed 15 December 2023.

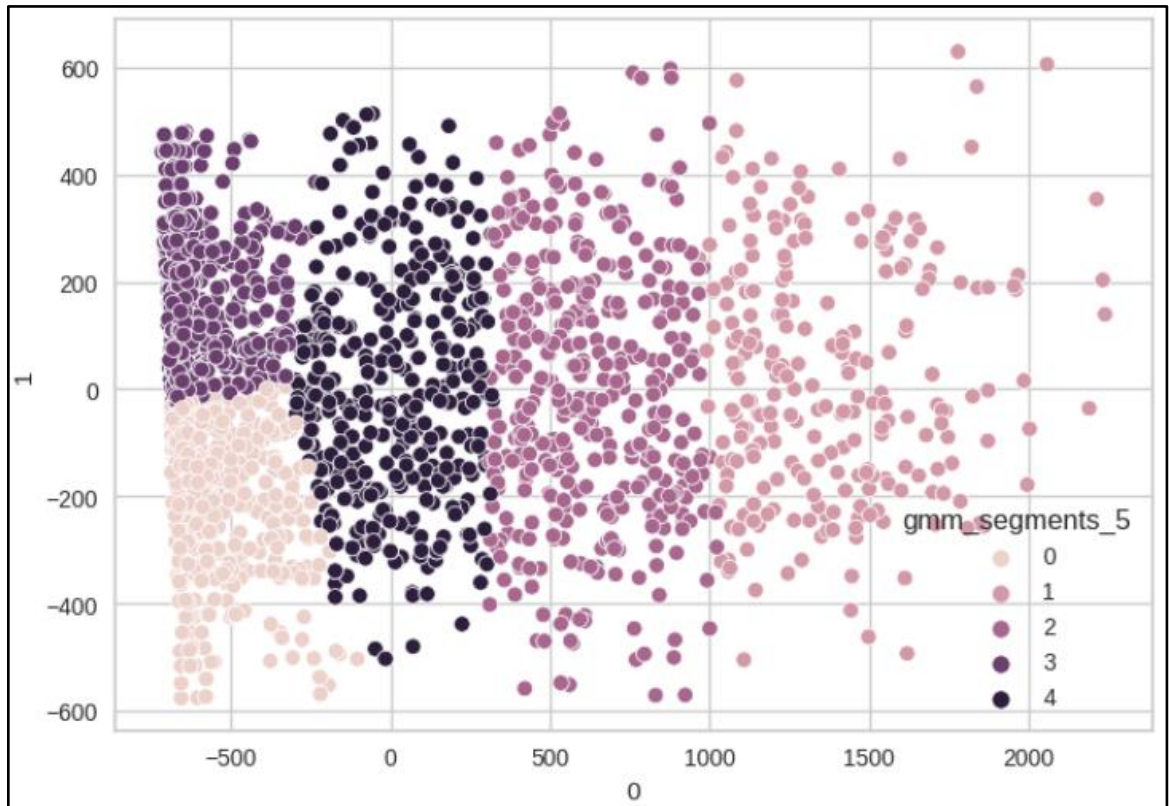
"Online Advertising Costs In 2021 | Top Draw." *Top Draw Inc.*, 26 March 2021, <https://www.topdraw.com/insights/is-online-advertising-expensive/>. Accessed 15 December 2023.

Appendix 1: K-Medoids Segment Boxplots, Where K=5





Appendix 2: Gaussian Mixture Model



Appendix 3: Hierarchical Clustering Model Silhouette Scores

