# Capstone Project: Marketing Campaign Analysis based on the domain - Data Analysis & Visualization/Unsupervised Learning

## **Executive Summary**

This project’s purpose is to understand customer behavior better and segment those customers to be better able to market to those customers based on the behaviors uncovered in the data. Based on the data, there were two distinct segments that were uncovered.

* **Segment 1:** This group has a slightly higher income and is likely to have a teen in the home. They are relatively even in age with the second segment. When they shop, they are definitely likely to spend more money. They use all mediums to shop. Mediums like the web, catalog, and store but are more in favor of shopping in person than using the other two.
* **Segment 2:** This group has a slightly lower income but is more likely to have teens and smaller children in the home. This group makes more web visits, they don’t make as many purchases on the web.

## **Problem Summary**

In an effort to drive revenue and lower marketing spending, customer behavior needs to be better understood. A better understanding of customer behavior would allow for more targeted marketing. More targeted would allow lower marketing costs and less effort because of the better understanding of customers. This would also drive revenue because of the lowered marketing cost but also get customers to shop more often and spend more per visit.

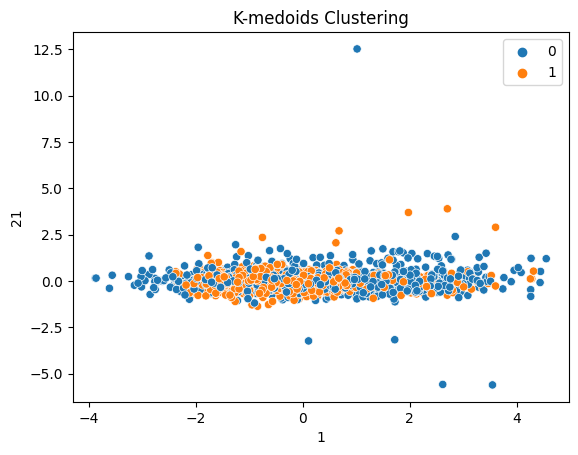
## **Solution Design**

Using the customer data, there were four clustering models that were explored. Those models were:

* KMeans
* Kmedoids
* Agglomerative Clustering
* DBSCAN

PCA and T \_SNE were both explored for each model. The solution that was chosen was KMedoids using PCA to reduce dimensionality. This was done with the marketing campaign segmentation dataset. This was data that was provided for this project. KMedoids were chosen because they are robust to outliers. This dataset had many outliers, especially for income. It is easier to interpret. Because of low dimensionality, the computation cost is low. The elbow method was used to determine the best number of clusters.

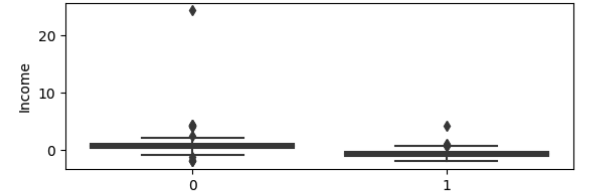
**Figure 1** shows the KMedoids cluster produced from the PCA data. In this model, you can see with the overlap that the data is close together. There is not a lot of difference between the two clusters, as shown in the plot. The appendix will show the plots of the other models. Overall the models came up with very similar plots where the clusters are close together.



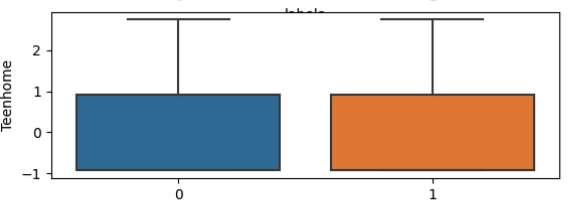
**Figure 1:** Plot of the KMedoids clustering

## **Analysis and Key Insights**

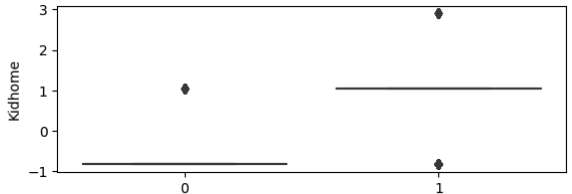
The key driver of customer behavior is the number of kids in the home. We can see from the data in **Figure 2** that Segment 2 income is slightly less than Segment 1. There does not look to be enough of a difference that income would be a key driver in behavior. In **Figure 2.1,** teens in the home are the same, but where they differ the most is having younger children in the home. Looking at **Figures 2.1** and **Figure 2.2**, this suggests that Segment 2 has both teens and younger children. Segment 2 is spending less and has more children but with a slightly lower income. Neither group make a lot of pf purchases with discounts. The number of discounts used for purchases is relatively the same for both segments. You can see an example of this in **Figure 2.3.**



**Figure 2:** The incomes of segments 1 and 2. 0 being 1 and 1 being 2.

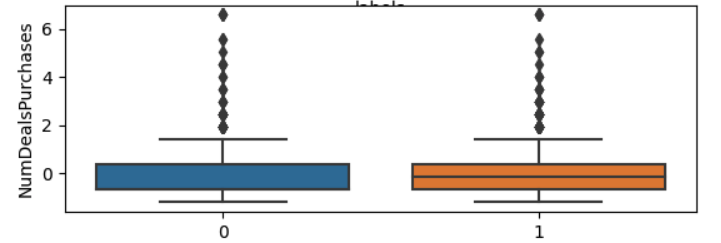


**Figure 2.1:** The number of teens in the home of each segment.



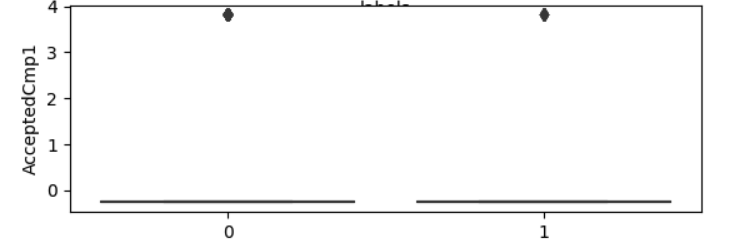
**Figure 2.2:** The number of younger kids in the home for each segment

Although incomes are not far apart, the amount Segment 1 spends vs. Segment 2 is significant. Also, segment 1 is more likely to have fewer people in the home. Neither segment use a lot of discounts when making purchases.



**Figure 2.3:** Discounts used when shopping

There is little to no interaction with any of the six marketing campaigns that have been offered. This is shown in **Figure 3.** The plots from the other five campaigns look easily the same.



**Figure 3:** Acceptance of Campaign 1 for each segment

## **Recommendations**

Segment 2 should be the focus. They are spending less and seem to have more reason to spend with likely hood of having more people in the home. Here are some possible suggestions for Segment 2:

* Make sure that customers always know about discounts.
* What are some things that both teens and smaller children like? Because Segment 2 is more likely to have both in the home, it has discounts, and marketing campaigns centered around those things.
* Make purchasing easy. Segment 2 is more likely to have both younger kids and teens in the home, which means they might have busier lives. Making purchasing easier and more efficient would help them save time.

## **Recommendations for Next Steps**

* Get an understanding of customers' media habits to know where to market.
* Partner with UX to understand the why behind customer behavior.
* Get a regular cadence of running these projects to have a continual understanding of customer behavior and how it changes over time.