MIT Applied Data Science Program

# Marketing Campaign Customer Segmentation

October 7, 2022

Kristi Yamashita

# Problem and Solution Summary

## PROBLEM SUMMARY

- Customer segmentation is a crucial aspect of marketing operations for businesses and organizations
  - Effectively use their money, time, and other resources to target the correct customer groups and provide customized communications and offerings
- We must divide the dataset of customers into groups based on similar characteristics of the customers
- Are there any issues with the data that we need to address before analyzing?
- What characteristics are common among the customers in the dataset to create customer groups?
- What variables can we create to further analyze the data?
- Which technique should we adopt to create clusters?
- How should we position our marketing campaigns and strategies to effectively target the given customer segments?

## FINAL PROPOSED SOLUTION DESIGN

- Data exploration
- Utilize Hierarchical Agglomerative Clustering to create customer segments
- Analyze 3 clusters created
  - High-Income Customers
  - Medium-Income Customers
  - Low-Income Customers
- Make appropriate campaigns and offerings

# Executive Summary

## KEY TAKEAWAYS

- At least one customer accepted the offer of all the campaigns. The offers from campaign 3 and 5 were accepted the most, and the offer from campaign 2 was accepted the least. More customers have accepted the offer in the last campaign compared to any of the offers from campaign 1, 2, 3, 4, and 5.

| Algorithm | Best Solution | Silhouette Score | Remarks |
|---|---|---|---|
| K-Means | K = 5 | 0.13021374284739756 | In the elbow plot, the elbow is seen for K=3 and K=5. We used both of these K values to see which provides better results. We get deeper insights into different types of customers using K=5. |
| K-Medoids | K = 5 | 0.107528069592116 | Gives us similar clusters to K-Means algorithm. |

| Algorithm | Best Solution | Silhouette Score | Remarks |
|---|---|---|---|
| Hierarchical Clustering | • Euclidean distance and ward linkage<br>• K = 3 | 0.25378332321341474 | The cophenetic coefficient has an average of 0.736922421650532 and highest coefficient of 0.8671371105092277, which means it is a pretty good measure of how faithful the dendrogram preserves the pairwise distance between the original unmodeled data points. |
| DBSCAN | Eps = 3 and min sample = 20 | 0.3398851568849134 | Highest silhouette average is 0.3398851568849134 for eps = 3 and min sample = 20. Only yields 2 clusters. |
| Gaussian Mixture Model | K = 5 | 0.14344403792681099 | Gives us similar clusters to K-Means and K-Medoids algorithm. |

## Cluster 0: High-Income Customers

- Consists of 583 customers
- Spend a lot and don't tend to make purchases through the company's website or with a discount
- Average income of 74435.61
- Spends significantly more compared to the other clusters with an average total amount spent of 1362.68 and average of 71.11 spent per purchase
- Spend the most on wine, fruit, meat, fish, sweet, and gold products in the last 2 years
- Highest average for purchases made using a catalog or directly in store
- Average age of 48
- Don't have kids
- Highest average for accepting the offers in campaigns 1, 2, 3, 4, 5, and the last campaign

## Cluster 1: Low-Income Customers

- Twice as big as the other two clusters with a total of 1072 customers
- Don't spend a lot and makes more purchases through the company's website using a discount
- Spend the least on average for all the products in this dataset
- Average income for customers in this cluster is 35708.22
- Spend an average total amount of 112.64 and only 11.81 per purchase
- On the younger side compared to the other two clusters with an average age of about 45
- Have a small child in their household

## Cluster 2: Medium-Income Customers

- Consists of 572 customers
- Spend a medium amount and seem to make the most purchases through the company's website using a discount compared to the customers in the other two clusters
- Average income for this group is 58347.45
- Generally older as the average age is about 50
- In a relationship
- Have a teenager in the household
- Average amount these customers spend is 759.70 with an average of 34.94 per purchase
- Make the most purchases
- Been with the company the longest

## KEY NEXT STEPS

- Create campaigns and other offerings that are personalized to the customers based on the clusters created
- Understand the successes and failures of prior campaigns and mimic what worked well

# Recommendations for Implementation

## KEY RECOMMENDATIONS AND ACTIONABLES

- High-Income Customers
  - Mimic offerings in campaigns 1, 2, 3, 4, 5, and the last campaign
  - Promote expensive products in catalogs and in stores

- Medium-Income Customers
  - Offer discounts that can be used with purchases through the company's website or in stores
  - Promote offers that are suitable for a families and households with primarily teenagers

- Low-Income Customers
  - Offer discounts that can be used with purchases through the company's website
  - Promote offers that are suitable for a families and households with primarily small children

## EXPECTED BENEFITS

- Deeper understanding of customers
  - Needs, motivations, interests, spending habits, demographics and lifestyles
- More efficient marketing efforts and targeted campaigns
- Better return on investment
  - Allows businesses and organizations to know how to effectively use their money, time, and other resources to target the correct customer groups

## KEY RISKS AND CHALLENGES

- Limited and/or expensive production
  - It can be hard to produce a variety of products for each segment on a mass scale
- Costly and time consuming marketing
  - Marketers have to consider all the segments in regards to the different needs, interests, habits, preferences and attitudes. Formulating and implementing several marketing strategies for different segments can take a lot of time.
- Wrong selection or change in the market
  - Chance of selecting an irrelevant or very small segments and therefore, cannot sell its products or services properly
  - Characteristics of the market may change due to a change in the customers' behavior, buying habits, income, etc.

## FURTHER ANALYSIS / ASSOCIATED PROBLEMS

- Conduct further analysis for each campaign and why certain ones performed better with customers than others
- Determine which campaigns would be most appropriate for each cluster
- Create new campaigns that specifically target each cluster based on this knowledge

# THANKS!

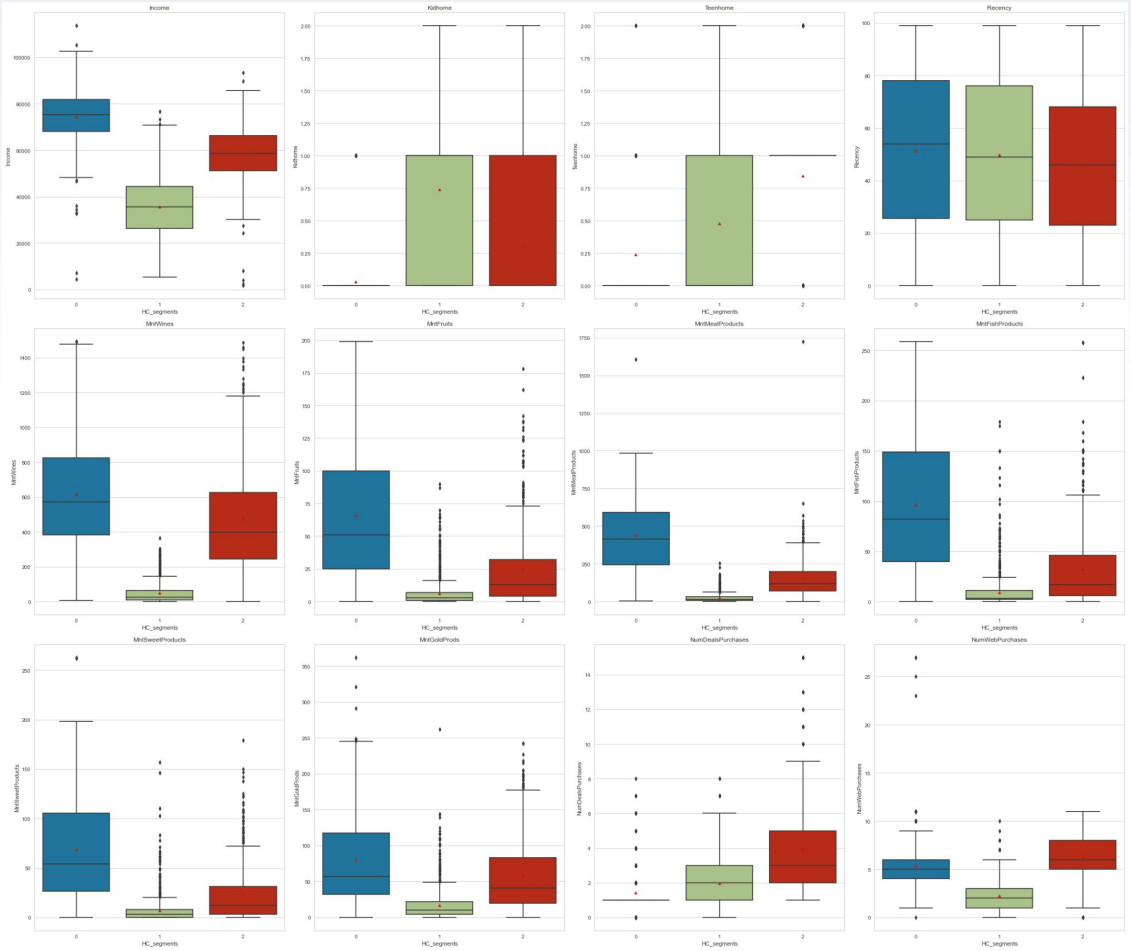
## Any questions?

You can find me at:

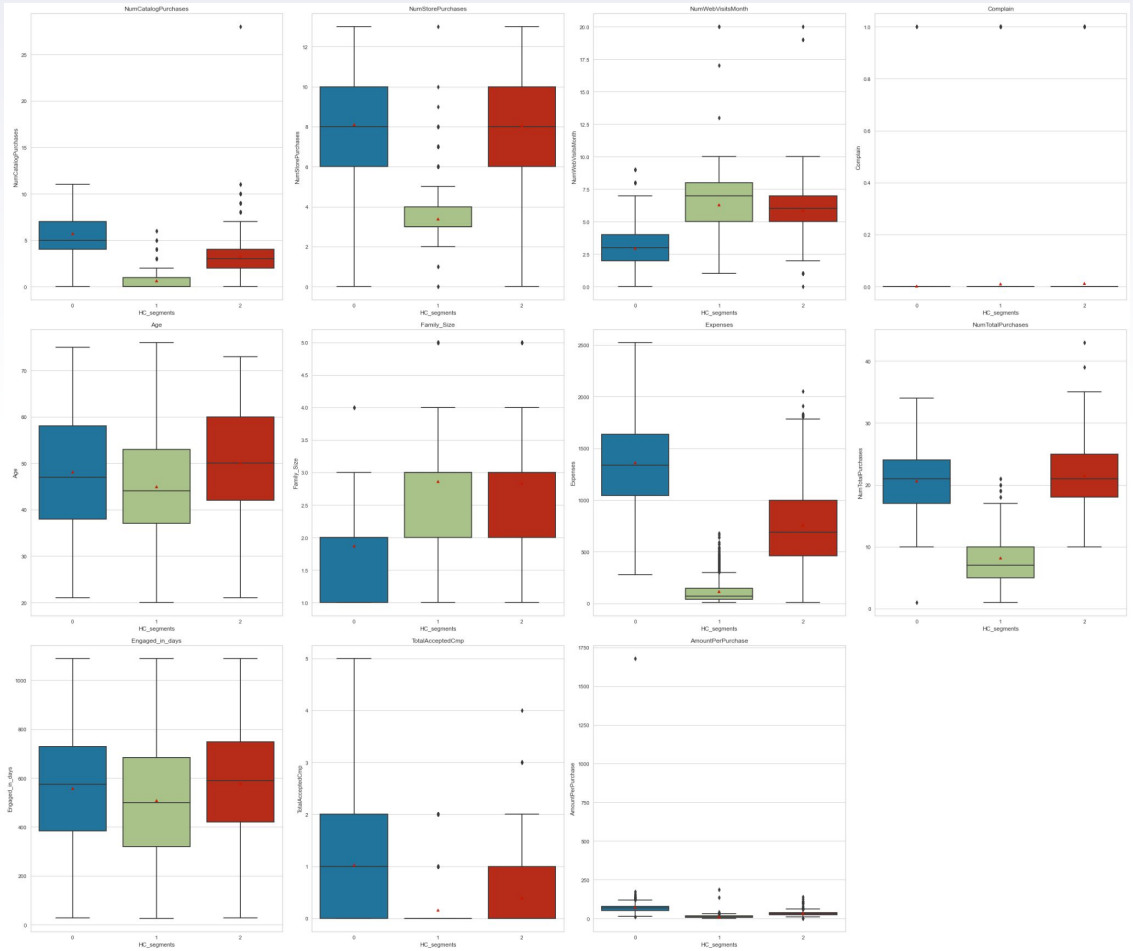
- kristi.yamashita@yum.com
- www.linkedin.com/in/kristiyamashita

# Appendix

# HIERARCHICAL AGGLOMERATIVE CLUSTERING BOX PLOT



# HIERARCHICAL AGGLOMERATIVE CLUSTERING BOX PLOT (continued...)



# HIERARCHICAL AGGLOMERATIVE CLUSTERING SCATTER PLOT