

Retail Customer Segmentation

The Problem

Customer segmentation is the process of dividing customers into groups based on common characteristics or patterns so that businesses can market their products effectively and significantly to each group.

In business-to-consumer marketing, most companies categorize their customers based on their age, gender, marital status, location (urban, rural), life stage (single, married, divorced, etc.).

Segmentation helps marketers to gain a better understanding of their product and identify ways to improve existing products or identify new product or service opportunities, establish better customer relationships, focus on the most profitable customers, and so on.

The Dataset

Customers of a retail shopping store are represented in the dataset. There are 10 variables and 1000 records in the dataset. (Before data clean up).

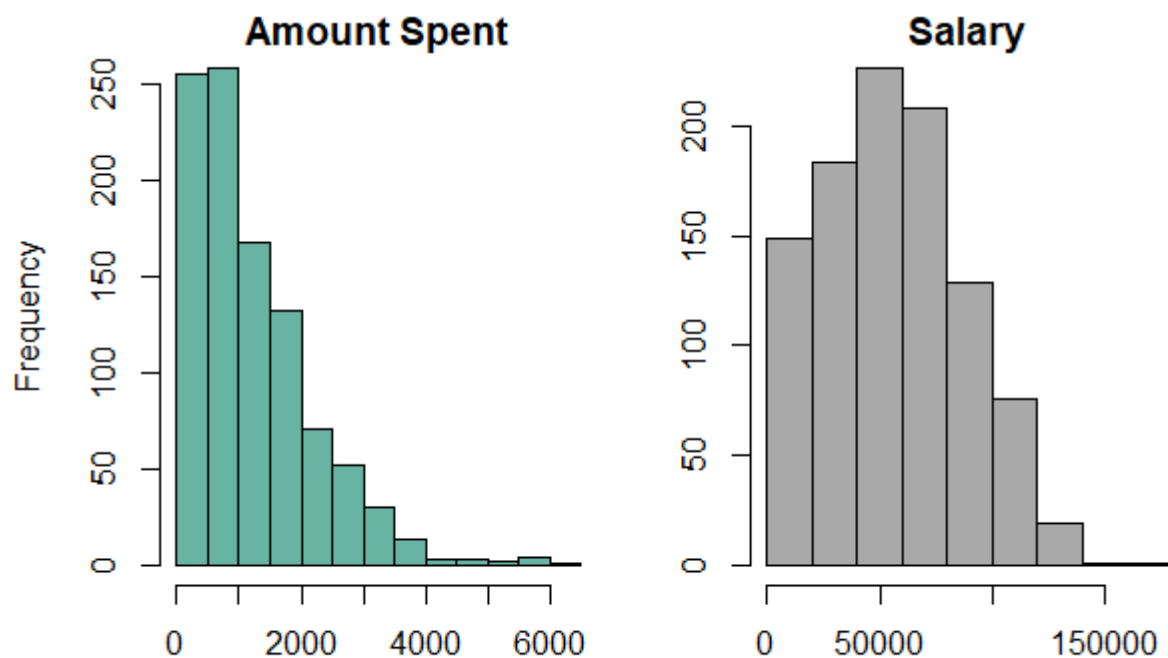
After data clean-up, I was left with 9 variables and 997 records to examine. I continued forward after the initial stage of cleaning and organizing the data to learn more about the dataset. The variables consist of 7 factors and 2 integers in the dataset. Discrete variables: Age, Gender, Own Home, Married, Location, Children, and Catalogs.

A glimpse of the dataset.

	Age	Gender	OwnHome	Married	Location	Salary	Children	History	Catalogs	AmountSpent
1	Old	Female	Own	Single	Far	47500	0	High	6	755
2	Middle	Male	Rent	Single	Close	63600	0	High	6	1318
3	Young	Female	Rent	Single	Close	13500	0	Low	18	296
4	Middle	Male	Own	Married	Close	85600	1	High	18	2436
5	Middle	Female	Own	Single	Close	68400	0	High	12	1304
6	Young	Male	Own	Married	Close	30400	0	Low	6	495
7	Middle	Female	Rent	Single	Close	48100	0	Medium	12	782
8	Middle	Male	Own	Single	Close	68400	0	High	18	1155
9	Middle	Female	Own	Married	Close	51900	3	Low	6	158
10	Old	Male	Own	Married	Far	80700	0	NA	18	3034
11	Young	Male	Rent	Married	Close	43700	1	NA	12	927
12	Middle	Male	Own	Married	Far	111800	3	High	18	2065

Data Source: <https://www.kaggle.com/kerneler/starter-retail-marketing-af201342-b>

Continuous variables: Salary and Amount spent.



Distribution of continuous variables

Approach

After cleaning up and rearranging the data, we were left with 996 records and 9 variables. Using K-mean we defined 4 clusters, 4 different types of customer segments, and each one with its unique characteristics of customers. The decision of choosing 4 clusters was backed by different validity measures: The elbow method and the Silhouette score.

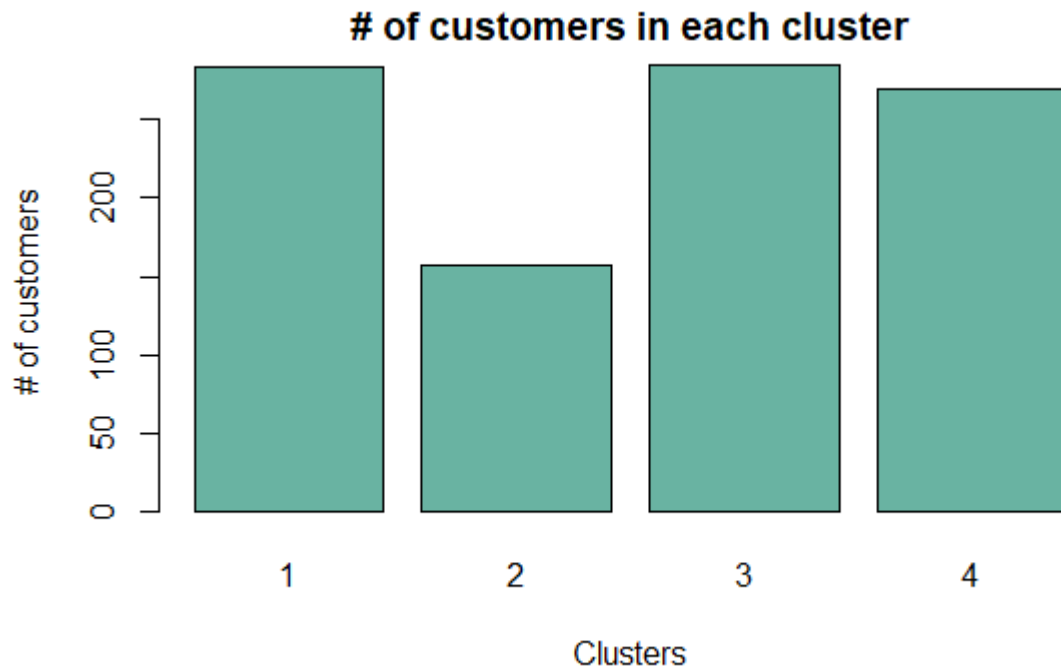
On the dataset we have, we need to apply customer segmentation to this historical data. Customer segmentation is the practice of partitioning a customer base into groups of individuals that have similar characteristics. It is a significant strategy as a business can target these specific groups of customers and effectively allocate marketing resources. Case in point, one group might contain customers who are high-profit and low-risk that are more likely to purchase products or subscribe for a service; another group might include customers who are low-profit.

A business task is to retain those who are high-profit and low-risk customers.

Analysis

Clusters 1,3,4 are distributed almost evenly with 283, 285, and 269 clients respectively, while cluster number 2 has 157 clients. The results of the clustering are as follows:

- Cluster number-1 This cluster has customers with the lowest average amount spent.
- Cluster number-2 This cluster has customers with the second-lowest amount spent.
- Cluster number-3 This cluster has customers who spent the most.
- Cluster number-4 This cluster has customers who spent the second-highest amount.



Conclusion

K-means has proven to be a simple yet effective classification method. The results are clear, and each clustered segment has different distinguishing characteristics, which is why these methods are widely used in the marketing industry not only for the purpose we demonstrated in our case study—client segmentation—but also for matching the best products similar to what the clients may have purchased.

[Link to my GitHub](#)