

Ecommerce Customer Segmentation with Machine Learning

Avery Willets

University of Advancing Technology

September 29, 2024

Ecommerce Customer Segmentation with Machine Learning

Clustering and segmentation are vital aspects of data science. They allow analysts to break an overall population of data into groups with similar traits. In this way, data scientists can use clustering and segmentation to find connections in the data they may otherwise not see. K-Means Clustering is a clustering machine learning algorithm that sorts data into an amount of clusters, K, defined by the user. In this project, I created a dummy dataset of customer data, fitted a K-Means Clustering algorithm to it, and analyzed the results.

Clusters

Cluster 1

	Age	Income	Total purchases in the last 100 days	Average order value	Time spent on website	cluster
count	31.000000	31.000000	31.000000	31.000000	31.000000	31.0
mean	40.935484	74761.258065	18.419355	52.032258	56.903226	0.0
std	12.055249	7759.305132	9.932352	22.498124	28.937121	0.0
min	20.000000	60300.000000	3.000000	16.000000	8.000000	0.0
25%	31.000000	68666.000000	10.500000	35.500000	31.500000	0.0
50%	41.000000	74835.000000	17.000000	49.000000	56.000000	0.0
75%	52.000000	81517.000000	28.000000	69.500000	84.500000	0.0
max	58.000000	87105.000000	38.000000	96.000000	98.000000	0.0

Cluster 1 is characterized by varying age groups, higher income levels, and the longest amount of time spent on the website.

These customers may be wealthier middle class and may be attracted to deals or promotions on the website. It may be beneficial to further examine this cluster by implementing

targeted ads on the website and identifying any correlations between customers in this cluster and customers who clicked on the ad.

Cluster 2

	Age	Income	Total purchases in the last 100 days	Average order value	Time spent on website	cluster
count	31.000000	31.000000	31.000000	31.000000	31.000000	31.0
mean	38.935484	102576.645161	20.451613	48.483871	55.935484	1.0
std	11.030066	9454.948555	10.689056	23.517187	32.255165	0.0
min	20.000000	89730.000000	0.000000	11.000000	3.000000	1.0
25%	30.000000	93981.000000	14.000000	28.000000	29.000000	1.0
50%	39.000000	102316.000000	20.000000	43.000000	50.000000	1.0
75%	48.500000	110547.500000	27.000000	68.500000	90.000000	1.0
max	59.000000	119194.000000	38.000000	88.000000	98.000000	1.0

The second cluster is characterized by a slightly younger age range, very high income levels, and the highest number of purchases in the last 100 days.

This cluster could represent younger wealthy individuals who may purchase items more frequently. They may be influenced by ads or trends. It may also be beneficial to research methods and channels for consumer trends of younger individuals to know where and how to place ads.

Cluster 3

	Age	Income	Total purchases in the last 100 days	Average order value	Time spent on website	cluster
count	38.000000	38.000000	38.000000	38.000000	38.000000	38.0
mean	41.210526	45811.289474	18.684211	59.000000	48.526316	2.0
std	9.989611	8782.836921	10.984277	25.954222	25.961236	0.0
min	20.000000	30957.000000	0.000000	10.000000	1.000000	2.0
25%	34.000000	38433.500000	10.000000	37.250000	26.250000	2.0
50%	43.500000	46258.000000	20.000000	67.000000	49.000000	2.0
75%	48.000000	53576.750000	28.750000	79.000000	69.500000	2.0
max	59.000000	59683.000000	35.000000	97.000000	93.000000	2.0

The third cluster is characterized by an older age range, lower income levels, and the least amount of time spent on the website.

This cluster may represent older individuals with lower incomes who are less familiar with technology. They may be influenced by in-person sales, budget-friendly items, and promotional emails.

Next Steps

To further analyze the data, I recommend the following:

1. Identify important features in each cluster and compare them with known audience metrics.
2. Create more detailed personas for each cluster based on market research. Identify target markets and their purchase incentives.
3. Plan and implement a marketing campaign around one or more of these clusters and record results over time.
4. Report success metrics of the campaign and understand audience growth patterns.