

ZEOTAP DATA SCIENCE ASSIGNMENT

- Karan Kamath

1. This analysis applies clustering techniques to segment customers based on key behavioral and demographic metrics. The dataset includes customer signup details, transaction history, and regional information. Key metrics computed include the number of days since signup, transaction frequency, total spending, average transaction value, total quantity purchased, and the average time between transactions. Missing regional data is handled by imputing unknown values, and one-hot encoding is applied to regional attributes to enhance segmentation accuracy.
2. KMeans clustering is used to identify distinct customer segments. All features are standardized to ensure equal contribution, and the optimal number of clusters is determined using the Davies-Bouldin index. The analysis identifies seven clusters, with a Davies-Bouldin Index of 1.167, indicating a well-defined segmentation structure. Cluster evaluations are performed across a range of 2 to 10 clusters, with the final selection balancing compactness and separation effectiveness.
3. Cluster Profiles: Each cluster represents a unique customer segment with distinct purchasing behaviors and regional distributions:

Cluster 0: Customers with the longest tenure (691.92 days), high transaction frequency (7.73), and the highest total spend (\$6,235.18). Predominantly from South America (69.2%).

Cluster 1: Moderate tenure (583.65 days) and transaction frequency (4.57), with total spending of \$2,955.63. 100% from South America.

Cluster 2: The least active group, with only one transaction on average, spending \$340.50. A diverse regional mix, with 20% from Asia and 30% from Europe.

Cluster 3: Moderate engagement (3.28 transactions) and total spend of \$2,204.36, with the highest proportion of Asian customers (100%).

Cluster 4: Similar transaction frequency (4.63) and spending patterns to Cluster 1 but entirely composed of European customers (100%).

Cluster 5: Balanced engagement (4.97 transactions) and spending (\$3,025.00), with all customers located in North America (100%).

Cluster 6: High engagement (7.5 transactions) and total spending of \$5,379.35, with an average transaction value of \$736.28. Entirely from Asia (100%).

These clusters highlight clear differences in customer activity, spending behavior, and regional distribution. The segmentation provides valuable insights for targeted marketing, personalized recommendations, and optimized resource allocation. High-spending and frequent buyers (Clusters 0 and 6) present opportunities for loyalty programs, while low-activity customers (Cluster 2) may benefit from re-engagement strategies. The

strong regional influence in certain clusters suggests region-specific marketing approaches could be highly effective.