

Clustering report

1. Davies-Bouldin Index (DB Index)

- This evaluates the compactness of clusters (how close data points within a cluster are to each other) and their separation (how distinct clusters are from one another).
- Value (1.35):
 - This value is relatively low, which indicates that the clusters are reasonably well-formed.
 - However, a smaller DB Index (closer to 0) would suggest tighter, more distinct clusters.
 - The result suggests that the clustering algorithm effectively differentiated the customers but may struggle with certain overlapping segments.

2. Silhouette Score:

- This measures how well each data point fits within its cluster compared to other clusters.
- A score of 1 indicates perfect clustering, 0 indicates overlap between clusters and negative values indicate incorrect clustering (data points assigned to the wrong cluster).
- Value (0.216):
 - The score is positive, showing that most points fit better within their assigned cluster than in neighbouring clusters.
 - the score is low, indicating moderate overlap, which could result from suboptimal cluster shapes, highly correlated features and a high number of dimensions relative to the number of clusters.

3. Cluster Characteristics:

- Each cluster represents a distinct group of customers based on their behavioural and profile features. These clusters could be interpreted as follows:

High-Value Customers (e.g., Cluster A):

- Customers with the highest TotalValue and average price_x.
- Likely frequent buyers of high-priced products.
- Strategies: Offer loyalty rewards, personalized discounts, or early access to sales.

Budget Shoppers (e.g., Cluster B):

- Customers with moderate TotalValue but lower average price_x.
- Likely focus on affordable product categories or discounted items.
- Strategies: Promote affordable product bundles or discounts to encourage more purchases.

Infrequent Buyers (e.g., Cluster C):

- Customers with low TotalValue and fewer transactions.
- May represent new or inactive customers.
- Strategies: Reactivation campaigns, such as email marketing with limited-time offers.

Region-Specific Buyers (e.g., Cluster D):

- Clusters are strongly tied to specific geographic regions.
- Buying patterns could reflect regional preferences or seasonality.
- Strategies: Tailored marketing campaigns for regional trends.

Category-Specific Customers (e.g., Cluster E):

- Customers who predominantly buy from specific product categories (Category mode).
- Likely focus on niche products or specific interests.
- Strategies: Cross-sell related products or suggest new arrivals in their preferred category.

Insights from Centroids:

- Analyzing the centroid values of each cluster helps identify key patterns:
- Spending Habits: Average TotalValue and price_x per cluster reveal customer value tiers.
- Product Preferences: The most common Category gives insights into preferences within each group.