

Business insights inferred from the graphs produced in the EDA file. Here are the following insights:

1. Customer Distribution by Region: South America has the highest proportion of customers (30%), indicating strong market penetration. Focused strategies in this region could further solidify market dominance.
2. Balanced Customer Base: Europe and North America follow closely, with similar percentages of customers. These regions present stable growth opportunities for targeted campaigns.
3. Asia's Potential: Asia has the lowest customer share among the regions. Expansion efforts and localized marketing could tap into this underrepresented market.
4. High Engagement in South America: The high customer count in South America suggests it could be a primary revenue driver. Prioritizing customer retention here is essential.
5. Growth Opportunities in North America: Despite a decent customer base, North America's percentage indicates room for more aggressive marketing efforts.

Clustering Results Report

1. Number of Clusters Formed

The clustering process resulted in **4 clusters**. The breakdown of customers in each cluster is as follows:

- **Cluster 0:** 58 customers
- **Cluster 1:** 60 customers
- **Cluster 2:** 28 customers
- **Cluster 3:** 53 customers

2. Davies-Bouldin Index (DB Index)

The Davies-Bouldin Index (DB Index) is a measure of clustering quality. Lower values indicate better-defined clusters.

- **DB Index:** 1.06
 - This value suggests moderately well-separated clusters.

3. Silhouette Score

The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. It ranges from -1 to 1.

- **Silhouette Score:** 0.31
 - This indicates that the clusters are moderately compact but may have some overlap.

4. Cluster Features

The clustering was performed using three key customer features:

- **TotalTransactions:** Total number of transactions per customer.
- **TotalAmountSpent:** Total monetary value of transactions per customer.
- **AvgTransactionValue:** Average value of each transaction.

5. Summary Statistics

- **Dataset Size:** 199 customers
- **Feature Details:**
 - TotalTransactions ranges from 1 to 11.
 - TotalAmountSpent ranges from \$335.52 to \$10,673.87.
 - AvgTransactionValue ranges from \$335.52 to \$1,323.13.

6. Observations

- **Cluster Differences:** Clusters appear to group customers based on spending habits and transaction frequency.
 - High spenders with frequent transactions form one cluster.
 - Low spenders with fewer transactions form another cluster.
- **Cluster Range Length:** The clustering was tested across 9 possible cluster configurations to find the optimal number of clusters.
- **Inertia:** The inertia values were analyzed to assess the optimal cluster count.

7. Recommendations

- **For Marketing:** Focus on Cluster 3 for high-value customers with frequent transactions.
- **For Engagement:** Identify Cluster 2 customers (smallest group) for personalized offers to improve engagement.
- **Further Analysis:** Consider additional features, such as customer region or product categories, to refine clustering.