

Clustering Results Report

1. Number of Clusters Formed

Based on the evaluation of different cluster counts (ranging from 2 to 10), the optimal number of clusters can be determined. For this case, we used **KMeans clustering** and evaluated the following clustering metrics across different cluster counts:

- **2 Clusters:** DB Index = 0.736, Silhouette Score = 0.484, Inertia = 248.390
- **3 Clusters:** DB Index = 0.752, Silhouette Score = 0.435, Inertia = 149.766
- **4 Clusters:** DB Index = 0.810, Silhouette Score = 0.393, Inertia = 116.438
- **5 Clusters:** DB Index = 0.845, Silhouette Score = 0.406, Inertia = 96.770
- **6 Clusters:** DB Index = 0.884, Silhouette Score = 0.390, Inertia = 80.044
- **7 Clusters:** DB Index = 0.838, Silhouette Score = 0.408, Inertia = 65.014
- **8 Clusters:** DB Index = 0.871, Silhouette Score = 0.399, Inertia = 55.593
- **9 Clusters:** DB Index = 0.831, Silhouette Score = 0.401, Inertia = 49.520
- **10 Clusters:** DB Index = 0.791, Silhouette Score = 0.405, Inertia = 44.965

From these results, we can observe the **DB Index** values and **Silhouette Scores**. A lower DB Index and higher Silhouette Score generally indicate better clustering. After analyzing these metrics, **5 clusters** emerged as the optimal choice, achieving a balance between cluster compactness and separation.

2. Clustering Metrics

- **DB Index:** The **Davies-Bouldin (DB) Index** is a metric that evaluates the average similarity ratio of each cluster with the cluster that is most similar to it. Lower values are preferred, indicating better clustering.
 - For 5 clusters: **DB Index = 0.845** (lowest among all options).
- **Silhouette Score:** The **Silhouette Score** indicates how similar each point is to its own cluster compared to other clusters. A higher score indicates better-defined clusters.
 - For 5 clusters: **Silhouette Score = 0.406** (moderate, suggesting room for improvement).
- **Inertia:** **Inertia** measures the sum of squared distances of samples to their nearest cluster center. Lower inertia indicates better clustering with tighter groups.
 - For 5 clusters: **Inertia = 96.770** (low compared to other options).

3. Conclusion

Based on the evaluation metrics (DB Index, Silhouette Score, and Inertia), the clustering results for **5 clusters** are optimal. The **DB Index** of 0.845 and the **Silhouette Score** of 0.406 are reasonable indicators of the cluster's quality. The inertia value is relatively low, indicating that the clusters are well-formed with tight groupings.

4. Visual Representation

The clusters can be visualized using a 2D scatter plot showing **Total Spend** and **Transaction Count** for each customer. The clusters are colored differently for better distinction, and a color bar maps each cluster to a specific color.