Clustering Analysis Report

1. Number of Clusters Formed:

In this analysis, we applied KMeans clustering and specified the number of clusters as 5. This resulted in the segmentation of customers into five distinct groups based on their spending patterns, transaction frequency, and average product price.

2. Davies-Bouldin Index (DB Index):

- DB Index Value: 0.94 The Davies-Bouldin Index (DB Index) measures the average similarity ratio of each cluster with its most similar cluster. The lower the value, the better the separation between clusters. A value of 0.94 suggests that the clusters are reasonably distinct, but there is still some overlap that could be improved. This value indicates that the clustering solution is fairly effective, but fine-tuning the number of clusters or applying additional clustering methods might help further separate the groups.

3. Other Clustering Metrics:

- Silhouette Score: 0.31 The Silhouette Score quantifies how similar an object is to its own cluster compared to other clusters. A score closer to 1 indicates well-separated, distinct clusters, while a score near 0 suggests overlapping clusters. A value of 0.31 indicates that while some clusters are reasonably well-separated, there is significant overlap,

4. Visual Representation of Clusters:

We used a scatter plot to visually represent the clusters based on two key features: Total Spent and Transaction Count.

Each cluster is colored differently to show the distribution of customers across the identified segments.

This visual prόvides a clear overview of the segmentation: .

- Cluster 0: High spenders with more transactions, positioned at the top right of the plot.

- Cluster 2: Lower spenders with fewer transactions, located at the bottom left.

- Clusters 1, 3, and 4: Represent intermediate customer segments, with varying degrees of spending and transaction counts

5. Cluster Characteristics:

The following table summarizes the average characteristics of each cluster:

Cluster	Avg TotalSpent	Avg TransactionCount	Avg AvgPrice
0	6075.81	8.10	281.76
1	3469.36	5.56	261.75
2	1393.02	3.21	166.59
3	1671.11	2.63	283.18
4	3872.78	3.90	372.66

Cluster 0: The highest-spending and most frequent customers, with an average spending of 6075.81 and 8.1 transactions. Cluster 1: Mid-range spenders with an average of 3469.36 spent and 5.56 transactions.

Cluster 2: Customers with the lowest total spending of 1393.02, averaging 3.21 transactions.

Cluster 3: Moderate spenders with 1671.11 spent, and relatively fewer transactions at 2.63.

Cluster 4: Customers with a higher-than-average product price (372.66), spending 3872.78 across 3.9 transactions.

6. Conclusion and Recommendations:

The KMeans clustering successfully identified five distinct customer segments, with some clusters showing clear differentiation in terms of spending and transaction behavior. The Davies-Bouldin Index and Silhouette Score suggest that the clustering model is relatively effective but could benefit from further refinement. A potential next step would be to experiment with a different number of clusters or use other clustering techniques (such as hierarchical clustering) to better separate the customer groups. The visual representation confirms that most of the clusters are reasonably well-separated, with a few areas showing some overlap. The cluster summary provides valuable insights for targeted marketing and customer profiling, such as focusing more on high-spending clusters (Cluster 0) or addressing the needs of low-spending customers (Cluster 2).