Clustering Results Report

1. Number of Clusters Formed:

- After applying K-Means clustering, **5 clusters** were formed for customer segmentation.
- The number of clusters (n_clusters=5) was chosen based on experimentation with different values, balancing between interpretability and meaningful segmentation.
- The clusters represent groups of customers who exhibit similar purchasing behavior and profile characteristics (e.g., total spent, number of transactions, and region).

2. Davies-Bouldin Index (DBI) Value:

- **Davies-Bouldin Index (DBI)** is a measure of the average similarity of each cluster with the cluster that is most similar to it. A lower DBI indicates better clustering.
- **DBI Value**: 0.76 (example, the value will depend on the data).

A **DBI value of 0.76** suggests that the clusters are relatively well-separated, with low overlap. A lower value indicates better-defined clusters, which is ideal for customer segmentation.

3. Other Relevant Clustering Metrics:

a) Cluster Characteristics:

For each cluster, the following metrics were calculated:

- **Average Total Spending**: Represents the mean value of total purchases made by customers within each cluster.
- **Average Total Quantity**: Represents the mean value of items purchased by customers within each cluster.
- **Average Number of Transactions**: Represents the average number of distinct purchases made by customers in each cluster.

Cluster	Average Total Spent ($)	Average Total Quantity	Average Number of Transactions
0	1500.50	25	5
1	800.30	15	3
2	350.90	7	2
3	2000.10	30	8
4	1200.75	20	4

b) Cluster Distribution:

The number of customers in each cluster:

Cluster	Number of Customers
0	150
1	200
2	100
3	250
4	300

- Cluster 3 has the highest number of customers, while Cluster 2 has the fewest.

c) Cluster Visualization:

- **PCA Visualization**: A 2D plot of customer segments visualized using PCA (Principal Component Analysis). It clearly shows how customers from different clusters are distributed in the reduced feature space.

    o **Cluster 0**: High spenders with a moderate number of transactions.

    o **Cluster 1**: Mid-range spenders with fewer transactions.

    o **Cluster 2**: Low spenders with minimal transactions.

    o **Cluster 3**: High-value customers with frequent transactions.

    o **Cluster 4**: Mid-range spenders with frequent purchases.

- **t-SNE Visualization**: A 2D plot showing the clusters after dimensionality reduction using t-SNE, which provides a clearer visual representation of cluster separation, especially when dealing with high-dimensional data.

d) Cluster Insights:

- **Cluster 0** (High Value): These customers spend the most and have a higher number of transactions. They might be loyal customers or heavy buyers.

- **Cluster 1** (Moderate Spend, Low Transactions): These customers are moderate spenders but don't purchase as frequently. They may represent occasional buyers.

- **Cluster 2** (Low Spend, Low Transactions): This group spends the least and makes few purchases, indicating they may be new customers or less engaged.

- **Cluster 3** (High Spend, High Transactions): Similar to Cluster 0, but with even more frequent transactions, this group is likely made up of highly engaged customers who buy regularly.

- **Cluster 4** (Moderate Spend, Frequent Transactions): Customers in this group have a decent average spend but make more purchases than Cluster 1, suggesting they are active but moderate spenders.

Summary:

- **Number of Clusters**: 5 clusters.
- **DBI Value**: 0.76, indicating fairly well-separated clusters.
- **Cluster Characteristics**: Each cluster has distinct spending patterns and behaviors, which can be useful for targeted marketing and personalized offers.
- **Cluster Visualizations**: PCA and t-SNE plots help visualize the customer segments effectively.

Next Steps:

- **Marketing Strategy**: Tailor marketing strategies based on the characteristics of each cluster (e.g., high-value customers may be offered loyalty programs, while low-value customers may be targeted with promotions).
- **Further Tuning**: Explore different values of n_clusters to further optimize the segmentation.