Clustering Results and Evaluation Report

1. Number of Clusters Formed

After applying the K-Means clustering algorithm, the analysis identified **5 distinct customer clusters**. These groups were formed based on key behavioral metrics such as total spending, transaction frequency, and regional activity. Each cluster represents a unique customer segment, providing valuable insights for tailored engagement strategies.

2. Evaluation Metrics

To assess the quality of the clustering, three key metrics were used:

a. Davies-Bouldin Index (DB Index)

- **Explanation:**
  The DB Index evaluates the relationship between within-cluster distances and the separation between clusters. A lower value suggests better clustering, indicating that clusters are both compact and well-separated.

b. Silhouette Score

- **Explanation:**
  This score measures how similar a data point is to its own cluster compared to other clusters. Scores range from -1 to 1, with values closer to 1 indicating strong cohesion within clusters and clear separation between them.

c. Calinski-Harabasz Index

- **Explanation:**
  This index calculates the ratio of between-cluster dispersion to within-cluster dispersion. Higher values indicate more distinct and well-defined clusters.

3. Cluster Insights

The clustering revealed five distinct customer profiles:

- **Cluster 1:** High-spending, frequent shoppers who are likely loyal customers and contribute significantly to revenue.
- **Cluster 2:** Moderate spenders with occasional purchases, representing a segment with potential for growth.
- **Cluster 3:** Customers concentrated in specific regions, exhibiting unique preferences and behaviors.
- **Cluster 4:** Low-spending, infrequent buyers who may require targeted re-engagement strategies.
- **Cluster 5:** Seasonal or erratic spenders with irregular transaction patterns.

4. Data Visualizations

To better understand the clusters, several visualization techniques were used:

- **t-SNE Visualizations:**
  t-SNE was used to reduce the dimensionality of the data, allowing us to plot the clusters in a 2D space. The resulting scatterplots confirmed the distinctiveness of each cluster.

- **Bar Plots and Heatmaps:**
  Bar plots displayed the average feature values (e.g., total spending, transaction frequency) for each cluster. Heatmaps illustrated the regional distribution of customers, highlighting areas of high concentration.

5. Practical Implications

The findings from this analysis offer actionable insights for business decision-making:

- **Marketing Strategies:**
  - Target high-spending clusters with loyalty programs to reinforce their engagement.
  - Incentivize seasonal or erratic spenders to make purchases during off-peak periods.
- **Resource Allocation:**
  - Focus resources (e.g., inventory, staffing) in regions with high customer density.
  - Optimize supply chains to align with regional and cluster-specific demands.
- **Customer Retention:**
  - Develop tailored strategies to re-engage low-spending and infrequent buyers.

6. Recommendations

To further enhance the analysis, consider the following steps:

- Experiment with alternative clustering methods, such as DBSCAN or Gaussian Mixture Models, to compare results.
- Incorporate temporal data to track how customer behavior evolves over time.
- Integrate additional demographic data (e.g., age, income) to refine the clustering and uncover deeper insights.