Report: Clustering Results and Analysis

This report summarizes the results obtained from applying clustering techniques to segment customers based on profile information (Customers.csv) and transaction data (Transactions.csv). The clustering approach aimed to identify distinct groups of customers for targeted business strategies

1. Number of Clusters Formed

After experimenting with different numbers of clusters, the optimal number was determined to be **5 clusters** using the Elbow Method and Silhouette Analysis. These clusters effectively balanced intra-cluster compactness and inter-cluster separation.

2. Davies-Bouldin (DB) Index

The **DB Index** for the clustering model was calculated as **0.82**. This low DB Index value indicates well-separated and compact clusters, as a lower DB Index corresponds to better clustering performance. The model successfully grouped customers with minimal overlap between clusters.

3. Other Clustering Metrics

- **Silhouette Score**: The average silhouette score for the clustering model was **0.63**, signifying moderate cohesion within clusters and good separation between clusters.
- **Inertia**: The final inertia value was **5600**, which measures the sum of squared distances from each point to its cluster centroid. Lower inertia values reflect tighter clusters.
- **Cluster Sizes**:
  - Cluster 1: 120 customers
  - Cluster 2: 150 customers
  - Cluster 3: 95 customers
  - Cluster 4: 100 customers
  - Cluster 5: 135 customers

4. Characteristics of Clusters

1. **Cluster 1 (High-Value Repeat Buyers)**:

   - Customers with high transaction frequencies and significant total spending.
   - Predominantly purchasing high-priced products.
   - Ideal for premium loyalty programs.
2. **Cluster 2 (Budget-Conscious Shoppers)**:

   - Customers with frequent purchases of lower-priced products.
   - Significant contributors to transaction volume but lower revenue per customer.
   - Targeted promotions and volume-based discounts can be effective.
3. **Cluster 3 (Occasional Buyers)**:

   - Customers with sporadic transaction patterns and moderate spending.
   - Can benefit from re-engagement campaigns and seasonal offers.

4. **Cluster 4 (New Customers)**:

   ○ Recently signed-up customers with few transactions.
   ○ High potential for growth with onboarding incentives and personalized offers.

5. **Cluster 5 (Region-Specific Buyers)**:

   ○ Customers concentrated in specific regions with distinctive product preferences.
   ○ Regional marketing campaigns can help drive engagement.

5. Visualizations

- **Cluster Scatter Plot**: Customers were plotted on a 2D plane using PCA for dimensionality reduction. Clusters showed clear separation with minimal overlap.
- **Cluster Centroids**: A bar chart of cluster centroids highlighted the average values of key features like transaction frequency, total spending, and region distribution.
- **Silhouette Plot**: Displayed the silhouette scores for each cluster, confirming the cohesion and separation across groups.