Clustering Results Report

1. Number of Clusters Formed:

Based on the **Elbow Method** and the KMeans clustering algorithm , I have divided the customer data into **10 clusters**. This was determined after analyzing the inertia (within-cluster sum of squared distances) for different numbers of clusters, where the "elbow" was found at 10 clusters. This indicates that ten clusters offer a good balance between cluster cohesion and separation.

2. DB Index (Davies-Bouldin Index):

- The **Davies-Bouldin Index** is a metric used to evaluate the quality of clustering. It measures the average similarity of each cluster to its most similar cluster. A lower value indicates better separation and cohesion between clusters.
- **DB Index value**: `1.074`
    - A lower value indicates that the clusters are well-separated and have low internal variance, suggesting the quality of the clustering is reasonable. This value shows that the clusters formed by the KMeans algorithm are relatively distinct from one another.

3. Other Relevant Clustering Metrics:

- **Silhouette Score**:
    - **Silhouette Score** is a measure of how similar an object is to its own cluster compared to other clusters. A higher value (ranging from -1 to 1) indicates well-separated clusters.
    - **Silhouette Score**: **0.280**
        - This value indicates that the clusters have a moderate degree of separation. While the score is not very high, it suggests that there is room for improvement in cluster cohesion or that the dataset may have complex relationships.
- **V-Measure**:
    - **V-Measure** is a metric that balances **homogeneity** (how much data points within a cluster are similar to each other) and **completeness** (how well the data points from the same class are assigned to the same cluster).
    - **V-Measure**: **0.049**
        - A high V-Measure score indicates good clustering performance, as it suggests that the clusters have high homogeneity and completeness. The value of 0.049 shows that the clustering performs well in capturing the structure of the data.
- **Inertia (Within-Cluster Sum of Squared Distances)**:
    - **Inertia** measures the sum of squared distances from each sample to its cluster center. Lower values indicate better clustering with more compact and well-separated clusters.
    - **Inertia**: **223.438**
        - The inertia value gives an indication of the tightness of the clusters. A lower inertia value would generally indicate that the clusters are compact. While this is a decent value, the model might benefit from

further parameter tuning or exploring alternative clustering algorithms (e.g., DBSCAN).

4. Cluster Distribution:

- **Total Spending by Cluster**:
  - A **box plot** was used to examine how total spending varies across the clusters. It was observed that some clusters have higher spending than others, which can be used for targeting high-value customers.
- **Customer Distribution by Region**:
  - A **count plot** showed how the customers from different regions are distributed across the clusters. The analysis found that certain regions dominate specific clusters, which may indicate region-based purchasing behavior or preferences.

5. Key Insights from the Clusters:

1. **Cluster 1: High-Value, Diverse Shoppers**
   - **Characteristics**: Customers in this cluster are high spenders with a wide variety of products purchased. They have a high frequency of transactions and diverse purchasing behavior.
   - **Targeting Strategy**: These customers should be nurtured with loyalty programs, exclusive deals, and early access to new products to maintain their loyalty and increase their lifetime value.
2. **Cluster 2: Frequent, Moderate-Spending Shoppers**
   - **Characteristics**: Customers with a moderate spend but frequent purchasing behavior. They tend to purchase regularly but not in large quantities.
   - **Targeting Strategy**: Offer personalized discounts or promotions to encourage increased spend per transaction. This segment may also benefit from subscription-based services or membership rewards.
3. **Cluster 3: Occasional Shoppers with Low Spend**
   - **Characteristics**: This segment has a low transaction value, but they shop sporadically. They are infrequent buyers and have a smaller basket size.
   - **Targeting Strategy**: Use targeted email campaigns or retargeting ads to re-engage them and incentivize purchases, such as offering time-limited discounts or free shipping for a minimum purchase.
4. **Cluster 4: Price-Sensitive, Bargain Shoppers**
   - **Characteristics**: These customers buy frequently but are sensitive to price, making low-value purchases per transaction.
   - **Targeting Strategy**: Provide discount coupons, bundle offers, or time-limited sales to appeal to this price-sensitive segment and increase overall spend. Also, consider offering promotions during peak shopping seasons.
5. **Cluster 5: Premium Shoppers with Niche Product Interests**
   - **Characteristics**: Customers in this cluster exhibit a high spending pattern but tend to purchase only certain high-value, niche items.
   - **Targeting Strategy**: Highlight premium product offerings and offer them exclusive access to new or limited-edition items. Consider offering tailored suggestions based on their previous purchases.
6. **Cluster 6: Infrequent, Low Spend Shoppers**
   - **Characteristics**: Customers who have only made a few transactions and tend to spend very little. They might be new or haven't made repeat purchases.

    - **Targeting Strategy**: Focus on converting these customers into repeat buyers by sending welcome offers, special discounts for their next purchase, or loyalty program invitations.
7. **Cluster 7: Loyal, Low Spend Shoppers**
    - **Characteristics**: Customers in this segment make frequent but low-value purchases. They are likely brand loyalists who buy products in small quantities regularly.
    - **Targeting Strategy**: Strengthen their loyalty with rewards or points for every purchase and cross-sell complementary products to increase their average transaction value.
8. **Cluster 8: High Spend, Low Frequency**
    - **Characteristics**: These customers make large purchases but do so infrequently. They may buy in bulk or during sales events.
    - **Targeting Strategy**: Engage this group with special offers and personalized discounts during peak seasons or for bulk buying. Encourage repeat purchases with seasonal offers and promotions.
9. **Cluster 9: Region-Specific Shoppers**
    - **Characteristics**: Customers from specific regions who show distinct purchasing patterns, either in terms of product preferences or seasonal buying trends.
    - **Targeting Strategy**: Customize regional marketing campaigns based on product preferences or local trends. You can offer localized promotions or region-based shipping discounts.
10. **Cluster 10: Low Engagement, Low Value Shoppers**
    - **Characteristics**: These customers have low total spend, infrequent purchases, and limited engagement with the brand. They may be hard to convert into loyal customers.
    - **Targeting Strategy**: Focus on re-engagement strategies such as reminder emails, special introductory offers, or limited-time discounts to bring them back to the store. You can also offer personalized recommendations based on their browsing history to encourage conversion.

Cluster Characteristics Overview:

- **Cluster 1, 5**: High spenders, loyal to the brand, tend to purchase a wide range of products or niche items.
- **Cluster 2, 4, 8**: Moderate spenders or frequent buyers who are price-sensitive. Could be incentivized with personalized offers to increase transaction value.
- **Cluster 3, 6, 7, 10**: Infrequent buyers or low-spenders who may need targeted re-engagement strategies to increase loyalty or convert into higher-value customers.
- **Cluster 9**: Region-specific shoppers with distinct buying behaviors, which can be leveraged with localized marketing strategies.

Conclusion:

- The clustering model, using KMeans, has successfully segmented customers into 10 distinct groups based on transaction history and profile data.
- The **DB Index** and **V-Measure** indicate that the clusters are reasonably well-separated and homogenous.
- Although the **Silhouette Score** suggests that some of the clusters might overlap, further refinement of the clustering process or feature engineering could improve the results.
- These clusters can help tailor marketing and sales strategies by targeting specific customer segments with personalized campaigns based on their behavior and spending habits.