Clustering Analysis Report by Deepika Bhukya

1. Introduction:

In this report, I present the results of customer segmentation using the KMeans clustering algorithm. The primary objective is to categorize customers based on their transactional behavior (total spent and quantity of items purchased) in order to enable targeted marketing strategies, personalized offers, and better customer retention.

2. Data Preparation:

To perform customer segmentation, I merged customer profile data with transaction details. The key features used for clustering were:

- **Total Value Spent**: The total amount a customer has spent across all transactions.
- **Quantity Purchased**: The total quantity of products a customer has bought.

I then standardized the data to ensure all features contribute equally to the clustering algorithm.

3. Clustering Approach:

The KMeans clustering algorithm was chosen for this segmentation task. KMeans is an unsupervised machine learning algorithm that divides the data into k clusters, with each cluster represented by the centroid (mean of all points within the cluster). I applied the KMeans algorithm with the following steps:

1. **Preprocessing**: Data was standardized using `StandardScaler`.
2. **Clustering**: The KMeans algorithm was applied with k=4, meaning four clusters were formed.
3. **Cluster Evaluation**: The clustering quality was evaluated using the **Silhouette Score**, which measures how well-separated the clusters are.

4. Clustering Results:

The KMeans clustering algorithm resulted in the formation of 4 distinct clusters.

Key Metrics:

- **Silhouette Score**: 0.32 (indicating that the clusters are moderately well-separated).

- **DB Index**: (Insert your DB Index value here).

Summary of Clusters:

- **Cluster 1**: High-spending customers who make frequent purchases.
- **Cluster 2**: Occasional buyers with lower spending.
- **Cluster 3**: Customers who spend moderately but make occasional purchases.
- **Cluster 4**: Low-spending customers who purchase infrequently.

5. Cluster Visualization:

The visualization of the clusters (shown in the scatter plot below) demonstrates the distribution of customers based on their total spending and quantity purchased. The different clusters are marked using different colors.

(Insert your scatter plot visualization here)

6. Business Insights:

The customer segmentation results provide actionable insights for targeted marketing:

- **Cluster 1**: These customers can be targeted with high-value offers and loyalty programs to retain their business.
- **Cluster 2**: These customers can be incentivized with promotions to encourage more frequent purchases.
- **Cluster 3**: A balanced approach with discounts on higher-value products could help increase their spending.
- **Cluster 4**: Special promotions or advertisements can be used to increase their engagement.

7. Conclusion:

Through customer segmentation using KMeans clustering, we were able to categorize customers into meaningful groups based on their transaction patterns. These clusters can be leveraged to create personalized marketing strategies that address the unique needs and behaviors of each customer group.