Clustering Analysis Results Report

1. **Overview:**
   This document outlines the findings from a clustering analysis conducted on customer transaction data. The objective was to categorize customers based on two characteristics: overall transaction value and transaction count. These characteristics were normalized prior to employing the KMeans clustering method.
2. **Clustering Method Utilized:**
   Method: KMeans
   Number of Groups: 5 (Selected for this study)
3. **Quantity of Groups Established:**
   The KMeans algorithm created 5 groups. This quantity was selected for the preliminary analysis, and further adjustments may be made using techniques such as the Elbow Method or Silhouette Score.
4. **Davies-Bouldin Index (DBI):**
   The Davies-Bouldin Index (DBI) was utilized to assess the clustering quality. A lower DBI value denotes superior clustering, indicating that the clusters are distinct and cohesive.

   DBI Value: 1.23

5. **Interpretation:**
   A DBI value of 1.23 implies that the clusters are fairly well-defined and cohesive. An even lower value would suggest enhanced clustering quality.
6. **Additional Clustering Metrics:**
   Inertia (Sum of Squared Distances): This metric was not explicitly calculated in this report, but it can be obtained from the KMeans model using kmeans.inertia_.
7. **Silhouette Score:**
   To gain a more comprehensive understanding of clustering quality, the Silhouette Score could be calculated, providing insight into how effectively the data points are grouped into clusters.
8. **Visualization of Clusters:**
   The scatter plot displayed below illustrates the clustering outcomes, with each point symbolizing a customer. The color assigned to each point

indicates the cluster it belongs to. The axes represent the normalized total transaction value and the normalized transaction quantity for each customer.

9. **Summary: Cluster Overview:**
   The KMeans algorithm effectively divided customers into 5 segments based on their transaction patterns.
10. **The Davies-Bouldin Index of 1.23 indicates that the clusters are relatively well-defined.**
    Suggestions for Enhancement: Implementing the Elbow Method or Silhouette Score could aid in refining the selection of the number of clusters. Incorporating additional customer features may enhance the clustering model.
11. **Future Actions:**
    The clustering findings are stored in the file Customer_Clusters.csv, which contains customer IDs, total transaction value, quantity, and their assigned cluster.