Customer Segmentation / Clustering - Deliverables Report

1. **Number of Clusters Formed:**
   The KMeans clustering algorithm was applied to the dataset with the number of clusters set to 3. This configuration captures distinct customer segments and helps identify meaningful patterns in customer behavior based on their profile and transaction history. **Number of clusters: 3**

2. **DB Index Value (Davies-Bouldin Index):**
   The Davies-Bouldin Index (DBI) quantifies the quality of the clustering by evaluating both the compactness and separation of the clusters. A lower DBI indicates better clustering. The DBI for the clustering result with 3 clusters is:

   DB Index: 1.03

   A DBI of 1.03 suggests that the clustering is relatively good, with better separation between clusters compared to previous results. However, there is still some room for improvement as lower DBI values (closer to 0) are generally ideal, indicating even better-defined and more distinct clusters.

3. **Silhouette Score:**
   The Silhouette Score indicates how well each customer fits into its assigned cluster. A higher score indicates better cohesion within the cluster and more separation between clusters. The Silhouette Score for the clustering result with 3 clusters is:

   Silhouette Score: 0.38

   A Silhouette Score of 0.38 suggests moderate cluster separation. While this is an improvement over previous results, it still indicates that the

clusters could benefit from further refinement. A score closer to 1 would indicate better-defined clusters.

4. **Cluster Centers:**

The cluster centers provide insight into the average characteristics of the customers within each segment. Here are the cluster centers for the 3 clusters:

- Cluster 1:
  - Total Value: 0.52
  - Quantity: 0.60
  - Price: 0.54
  - Region: 0.22
- Cluster 2:
  - Total Value: 0.31
  - Quantity: 0.37
  - Price: 0.50
  - Region: 0.86
- Cluster 3:
  - Total Value: 0.20
  - Quantity: 0.24
  - Price: 0.49
  - Region: 0.18

These values provide insight into the typical characteristics of customers within each cluster, helping to identify distinct behaviors and preferences.

5. **Visual Representation of Clusters:**

The clusters were visualized using scatter plots of the scaled features, Total Value and Quantity, to better understand the distribution of customers across clusters.

- Scatter Plot (with Cluster Labels):
  The plot shows how customers are distributed across the 3 clusters. Each cluster is represented by a different color, visually helping to identify the separation between them.

- Scatter Plot (with Cluster Centers):
  The cluster centers are marked with red 'X' symbols to show the
  typical characteristics of the customers in each cluster.

Customer Segments (KMeans with 3 clusters)

Cluster Centers with 3 Clusters