**NAME: Kriyanshi Ghodasara**

**Email_id: kriyanshighodasara@gmail.com**

**Topic: Report on Clustering**

### 1. Number of Clusters Formed

The K-means clustering algorithm was applied to the dataset with the goal of segmenting customers into meaningful groups. After evaluating various potential cluster counts, the analysis resulted in the formation of **2 distinct clusters**. These clusters represent customer segments characterized by their purchasing behaviour, specifically based on features like **Total Value**, **Quantity**, and **ProductID**.

- **Cluster 0**: Customers with higher total values and greater quantities of products purchased.
- **Cluster 1**: Customers with moderate total values and smaller quantities of products purchased.

This segmentation provides valuable insights into customer types, allowing for targeted marketing strategies and product recommendations.

### 2. Davies-Bouldin Index (DBI)

The **Davies-Bouldin Index (DBI)** is a metric used to evaluate the quality of clustering by assessing both the separation and compactness of clusters. A lower DBI indicates better separation between clusters and tighter, more compact clusters.

- **DBI Value**: **0.6098**

**Interpretation**:
The DBI value of **0.6098** indicates that the clusters formed are relatively well-separated and compact, but there is room for improvement. Ideally, a DBI closer to 0 would suggest better-defined clusters with minimal overlap. The current value suggests that the clusters are moderately distinct, but further refinement (e.g., adjusting the number of clusters or choosing a different clustering algorithm) could enhance the quality of the segmentation.

### 3. Silhouette Score

The **Silhouette Score** is another crucial metric that assesses the quality of the clusters by evaluating how similar each point is to its own cluster compared to other clusters. A higher score indicates well-separated and well-formed clusters.

- **Silhouette Score**: **0.5416**

**Interpretation**:
The Silhouette Score of **0.5416** suggests that the clusters have a moderate level of cohesion and separation. A score above 0.5 generally indicates a decent clustering result, where the points are fairly well clustered within their respective groups. Although the clusters are reasonably well-formed, there may be potential for further refinement to achieve a higher level of distinctness and compactness.

## Conclusion

The clustering analysis has successfully identified **2 distinct customer segments** based on their purchasing behaviour. The **Davies-Bouldin Index (0.6098)** and **Silhouette Score (0.5416)** indicate moderate cluster quality, with room for potential improvement. While the current clustering solution offers valuable insights into customer segmentation, further optimization, such as adjusting the number of clusters or experimenting with alternative clustering algorithms, could enhance the results.