CUSTOMER SEGMENTATION REPORT

Aim:

The primary aim of this analysis is to segment customers based on their transactional behavior to identify distinct groups with similar characteristics. By doing so, we can help the business tailor its marketing strategies, product offerings, and customer retention efforts to different customer segments, ultimately improving customer satisfaction and driving business growth.

1. Number of Clusters Selection:

Based on the **Elbow Method** applied to the Within-Cluster Sum of Squares (WCSS), I have selected **6 clusters** for this customer segmentation task. The Elbow Method is a technique used to identify the optimal number of clusters by plotting the WCSS against the number of clusters.

- **Observation:** When plotting the WCSS for cluster sizes ranging from 2 to 10, we observe a noticeable "elbow" point at **6 clusters**, where the WCSS starts to decrease at a slower rate. This point indicates a diminishing return in terms of reducing WCSS with increasing clusters. Choosing 6 clusters ensures that we capture sufficient variation among the customer groups while avoiding overfitting by using too many clusters.

Thus, **6 clusters** provide a good balance between the model's simplicity and its ability to capture meaningful customer segmentation.

2. Clustering Metrics - Davies-Bouldin Index:

The **Davies-Bouldin Index (DB Index)** for the clustering results is **1.2213**. The DB Index is a measure of the average similarity between each cluster and its most similar one, where a lower value indicates better clustering. It takes both the distance between clusters and their dispersion into account.

- **Interpretation:** A DB Index value of **1.2213** suggests a reasonably good clustering result. Generally, a DB Index closer to 0 indicates well-separated and compact clusters. A value around **1.0 to 2.0** is considered acceptable, indicating that the clusters are fairly distinct and separated. While this value is not extremely low, it still indicates that the clusters are relatively well-separated, with a moderate level of compactness.

The **DB Index** confirms that the 6 clusters chosen are sufficiently distinct, and the separation between them is reasonable.

3. Cluster Characteristics:

The customer data has been segmented into **6 clusters**, each representing distinct groups with different customer behaviors. The table below shows the **average values** of key features per cluster:

Cluster	TotalValue	Quantity	Price	TransactionID	Recency
0	6127.848378	22.027027	280.611149	8.162162	74.891892
1	1995.199787	9.659574	210.105943	4.000000	81.255319
2	1070.562353	4.823529	227.525392	1.764706	279.058824
3	2466.809259	7.074074	348.250383	3.000000	113.703704
4	3927.642069	16.068966	243.980290	6.448276	119.034483
5	4066.327381	12.595238	320.840880	5.047619	92.595238

From the table above, we can interpret the following about each cluster:

- **Cluster 0**: This cluster represents **high-value customers** who purchase in larger quantities and have relatively frequent transactions (lower recency). They are likely to be loyal and valuable customers.
- **Cluster 1**: These are **mid-value customers** with moderate purchase quantities and slightly less frequent transactions. They may be occasional buyers, potentially influenced by promotions or seasonal sales.
- **Cluster 2**: These are **low-value customers**, purchasing in small quantities and infrequently (high recency). They might be one-time buyers or customers who only buy during special promotions.
- **Cluster 3**: These customers are in the **mid-to-high value** range with moderate purchasing frequency. They might be interested in higher-priced products, indicating a preference for quality or premium items.
- **Cluster 4**: These are **valuable customers** who buy moderately priced items but in larger quantities. They are engaged fairly frequently, indicating a stable and loyal customer base.
- **Cluster 5**: These customers have a good total value, purchase medium-to-high priced items, and have relatively recent transactions. They could be newly acquired customers with potential for long-term value.