Clustering Results Report

The analysis focuses on clustering customer data using the k-means algorithm to segment customers based on their purchasing behaviour. The insights obtained from this clustering process include the number of clusters formed, the Davies-Bouldin (DB) Index value, and other metrics like the silhouette score.

---

1. Number of Clusters Formed

The clustering process involved determining the optimal number of clusters by analysing the **Elbow Method**, the **Davies-Bouldin Index**, and the **Silhouette Score**. Based on these metrics:

- The **Elbow Method** suggested **5 clusters** as a good initial choice, based on the point where the rate of inertia reduction slows down.
- Further analysis of clustering metrics identified **7 clusters** as the most optimal for segmenting customers based on their purchasing behaviour.

Thus, the final number of clusters chosen for segmentation was **7**.

---

2. Davies-Bouldin Index (DB Index)

The **Davies-Bouldin Index** measures the average similarity between clusters, with a lower value indicating better clustering quality.

- The DB Index for different cluster sizes ranged between **0.62 and 0.8**, depending on the number of clusters.
- For **7 clusters**, the DB Index value was found to be **0.88**, which indicates reasonably well-separated clusters. Lower DB Index values suggest that the clusters are compact and distinct.

---

3. Silhouette Score

The **Silhouette Score** measures how well-separated clusters are, with values ranging between -1 and 1:

- A value closer to **1** indicates well-defined clusters.
- The Silhouette Score for **7 clusters** was approximately **0.37**, which suggests that while most data points were reasonably assigned to their clusters, there might be some overlaps or less distinct separations between certain clusters. This score indicates moderate clustering quality.

---

4. Cluster Characteristics

Visualizations of the clusters provided insights into customer segments:

- The segmentation revealed clear groupings of customers based on their **Total Quantity Purchased** and **Total Spend**.

- Customers in certain clusters displayed high spending with lower quantities, while others made frequent purchases with moderate spending.
- Clusters also highlighted potential groups of high-value customers (high spenders with moderate frequency) and low-value customers (low spending and low frequency).

---

5. Insights and Recommendations

- **High-Value Segments**: Specific clusters with high spenders can be targeted for loyalty programs or premium offers to retain and further engage these customers.
- **Low-Value Segments**: Clusters with low spending and low transaction frequency might benefit from personalized marketing campaigns or discounts to encourage higher engagement.
- **Diverse Patterns**: Some clusters indicate high-frequency, low-spend customers who might be purchasing smaller-ticket items. Promotions aimed at upselling or increasing basket size could be effective for these groups.

---

6. Final Observations

The clustering process successfully identified **7 distinct customer segments**, providing valuable insights into purchasing behavior. The combination of metrics (DB Index and Silhouette Score) confirms that the clusters are reasonably well-defined and actionable. These insights can be leveraged for strategic decision-making, including personalized marketing, inventory planning, and enhancing customer experience.