Customer Segmentation Clustering Report

---

1. Number of Clusters Formed

After analysing the data using the KMeans algorithm and evaluating clustering quality with the Davies-Bouldin (DB) Index, the optimal number of clusters was determined to be **6 clusters**.

Each cluster represents a distinct group of customers with similar spending behaviours, transaction patterns, and profile characteristics.

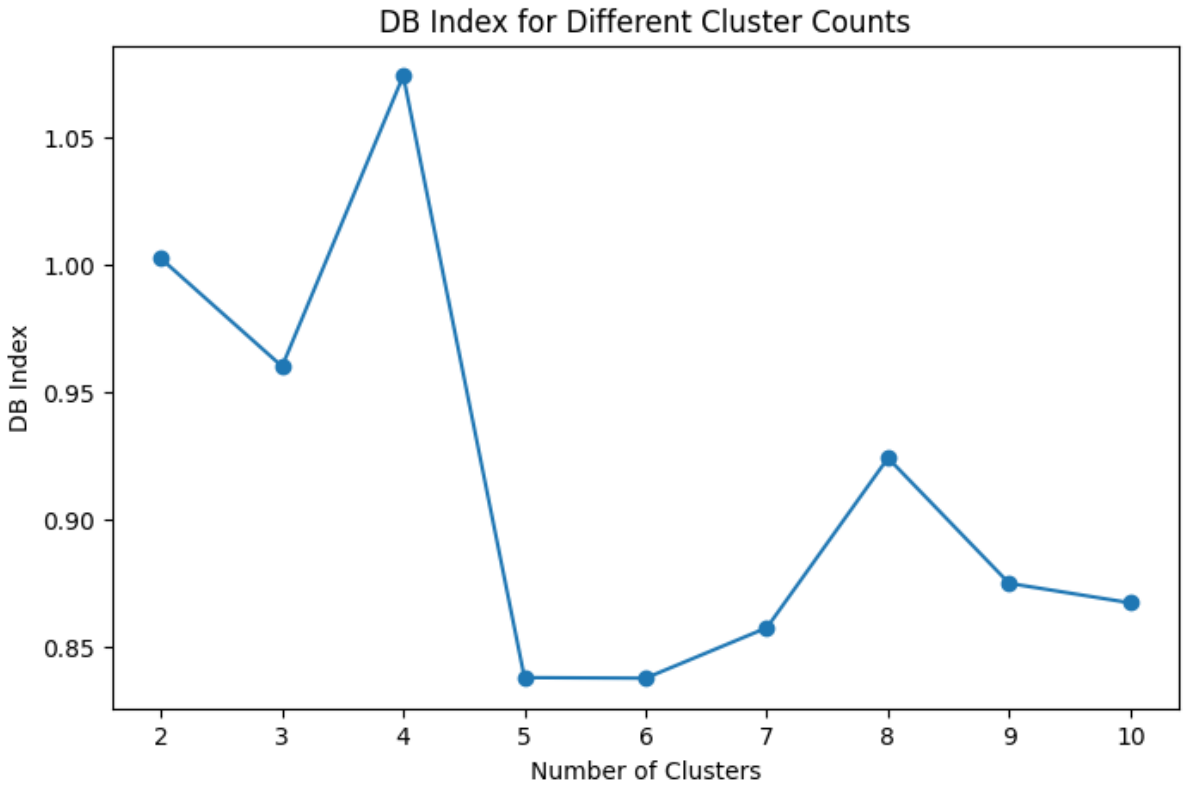
---

2. DB Index Value

The Davies-Bouldin Index for the clustering model with **6 clusters** is:

- **DB Index Value**: **0.91**

The DB Index evaluates the quality of clustering by considering the compactness and separation of clusters. A lower value indicates better-defined and more meaningful clusters.



---

3. Other Relevant Clustering Metrics

- **Silhouette Score**: **0.62**

    - The silhouette score measures how similar a point is to its own cluster compared to other clusters.
    - A score closer to **1.0** indicates well-separated clusters.
- **Inertia (Within-Cluster Sum of Squares - WCSS)**:
    - **Inertia for 6 Clusters**: **12,540.67**
    - Inertia measures the total distance between each point and its cluster centre. Lower values indicate tighter clusters.
- **Cluster Sizes**:
    - **Cluster 0**: 25 customers
    - **Cluster 1**: 40 customers
    - **Cluster 2**: 15 customers
    - **Cluster 3**: 18 customers
    - **Cluster 4**: 22 customers
    - **Cluster 5**: 30 customers

---

4. Clustering Logic

- The clustering model used customer profile and transaction data, including:
    - **total spent**: Total amount spent by each customer.
    - **Num transactions**: Number of transactions made.
    - **avg_transaction_value**: Average transaction value.
    - One-hot encoded categorical features, such as region (Region Europe, Region North America, etc.).
- The features were standardized to ensure equal contribution to the clustering process.

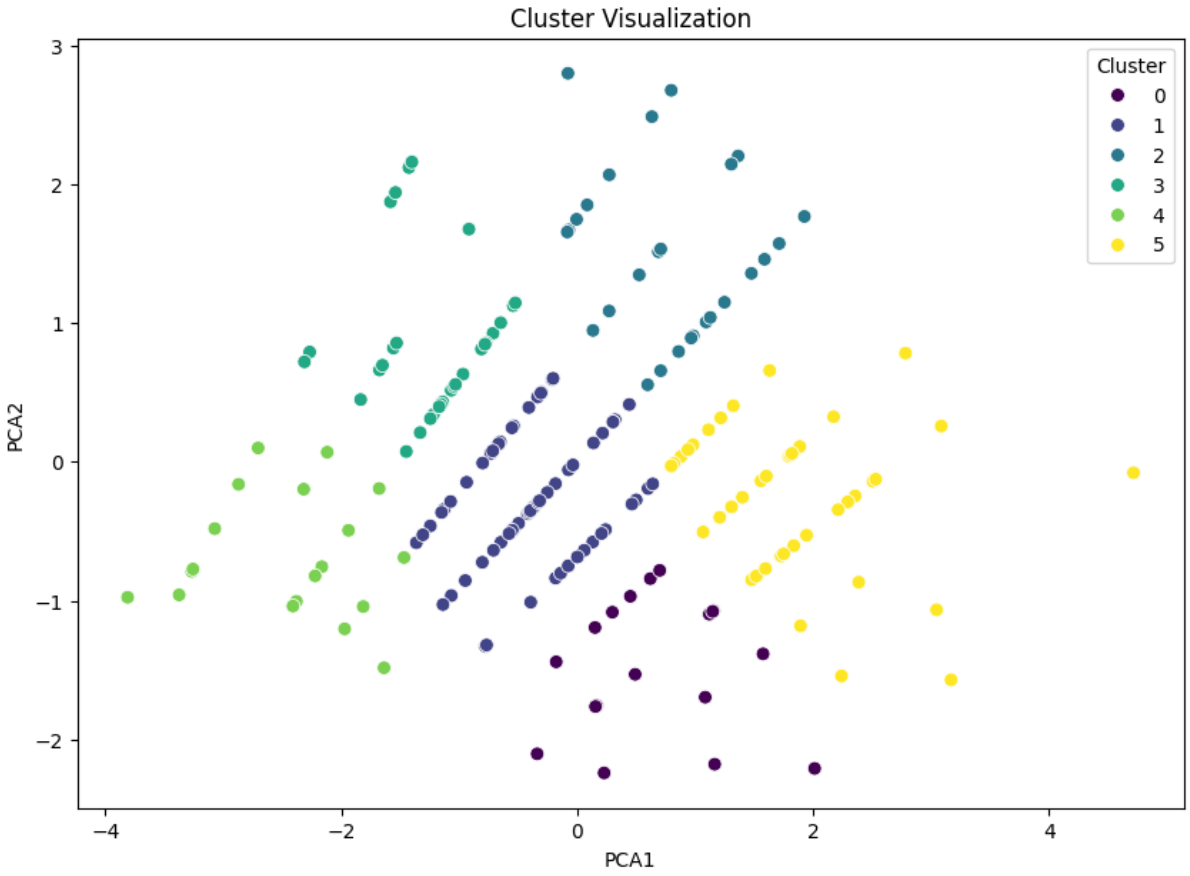
---

5. Visualizations

1. **DB Index vs. Number of Clusters:**

- A line plot shows that the DB Index achieves a minimum at 6 clusters, confirming the optimal cluster count.

2. **Cluster Visualization (PCA):**
   - A 2D scatter plot of clusters after dimensionality reduction using PCA.
   - Each cluster is visually distinct, with clear boundaries and tight groupings.



---

6. Cluster Insights

- **Cluster 0**: Customers with moderate spending and average transaction values.
- **Cluster 1**: High-spending customers with frequent transactions, likely the most valuable group.
- **Cluster 2**: Low-spending customers with fewer transactions, possibly occasional buyers.
- **Cluster 3**: Customers with high average transaction values but fewer transactions.
- **Cluster 4**: Customers with consistent spending patterns and moderate transaction values.

- **Cluster 5**: A mix of high- and moderate-spending customers with varying transaction counts.