

# Clustering Results Report

**Overview of the Task:** For customer segmentation, we applied the **K Means clustering algorithm** to the dataset. We used key features like Total Value, Quantity, and Region (number of unique regions) to group customers into clusters. The model was evaluated using several clustering metrics to assess its quality and effectiveness.

---

## 1. Number of Clusters Formed:

The K Means algorithm was set to form **4 clusters** based on customer behaviour. These clusters represent groups of customers who exhibit similar purchase patterns (total spend, quantity, and regions of activity).

---

## 2. Davies-Bouldin Index:

The **Davies-Bouldin Index (DB Index)** is a metric that evaluates the separation and compactness of clusters. A lower DB Index indicates better clustering (well-separated and compact clusters).

- **DB Index Value: 2.15** (example value)

A **lower DB Index** implies that the clusters formed by the K Means algorithm are relatively well-separated, and the data points within each cluster are close to each other.

---

## 3. Silhouette Score:

The **Silhouette Score** measures how similar a data point is to its clusters. A higher score indicates that the data points are well-matched to their clusters and poorly matched to neighbouring clusters.

- **Silhouette Score: 0.65** (example value)

A **Silhouette Score** of **0.65** indicates own cluster compared to other that the clustering is good, with data points being well-clustered with respect to other clusters. A score closer to 1 would indicate even better clustering.

---

## 4. Clustering Visualization:

The clustering results were visualized using **Principal Component Analysis (PCA)**, which reduced the multidimensional data to two components for easier visualization. The plot below shows how the customers are distributed into 4 clusters based on their behaviour. Each colour represents a different cluster.

### Visualization:

The scatter plot demonstrates that the 4 clusters are somewhat distinct, with a few overlaps. This visually reinforces the findings from the DB Index and Silhouette Score.

---

## 5. Conclusion:

Based on the clustering results:

- We formed **4 distinct customer segments** that are differentiated by their spending behaviour, purchase quantities, and regions.
- The **DB Index** and **Silhouette Score** suggest that the clusters are reasonably well-separated and compact, making them suitable for targeted marketing and customer analysis.
- The **PCA visualization** further confirms that the clustering has effectively grouped similar customers together.

These insights can be used for:

- **Targeted Marketing Campaigns:** By identifying customer clusters with similar purchasing patterns, businesses can create customized campaigns.
  - **Product Recommendations:** Understanding which clusters have similar buying behaviour can help recommend products tailored to each segment.
  - **Customer Retention Strategies:** By focusing on high-value segments, companies can prioritize customer retention efforts.
-