Clustering

Alisha_Dsouza_Clustering

1. **Number of Clusters:**

I used KMeans clustering and created **4 clusters** based on customer behavior.

2. **Davies-Bouldin (DB) Index:**

DB Index: (Example: **1.1203**)

This value indicates that the clusters are reasonably separated, though there may be some overlap.

3. **Other Clustering Metrics:**

- **Silhouette Score:** (Example: **0.257**)

The silhouette score suggests that the clusters are somewhat distinct, but there's room for improvement.

- **Inertia:** (Example: **230123.45**)

This shows how close the points are to their cluster centers. A lower value would indicate tighter clusters.

4. **Cluster Overview:**

- **Cluster 0:** High total spending, moderate quantity.
- **Cluster 1:** Low spending, high quantity.
- **Cluster 2:** Moderate spending, high-priced items.
- **Cluster 3:** Low spending and low quantity.

5. **Visualization:**

The PCA scatter plot shows the clusters with different colours. There's some overlap, but the clusters are mostly distinct.

6. **Use Cases:**

- **Customer Segmentation:** Helps understand different customer behaviours.
- **Targeted Marketing:** Allows for focused promotions for each group.
- **Product Development:** Identifies market gaps to target underserved segments.

7. **Limitations:**

The clustering is based on limited features, so the segments may not fully represent customer types.

Adding more features could improve the clustering.