CLUSTERING REPORT

1. KMeans Clustering: Silhouette Scores and WCSS Metrics

Silhouette Scores for KMeans (k=2 to k=10):

- **For k=2**: Silhouette Score = **0.3802**
- **For k=3**: Silhouette Score = **0.3603**
- **For k=4**: Silhouette Score = **0.3455**
- **For k=5**: Silhouette Score = **0.3588**
- **For k=6**: Silhouette Score = **0.3697**
- **For k=7**: Silhouette Score = **0.3613**
- **For k=8**: Silhouette Score = **0.3427**
- **For k=9**: Silhouette Score = **0.3429**
- **For k=10**: Silhouette Score = **0.3552**
-

Analysis:
The silhouette score measures how well each point fits within its cluster. Based on the scores, **k=6** seems to have the highest silhouette score (0.3697), suggesting it might be the most optimal number of clusters. This is followed by **k=7** and **k=2**, which also perform relatively well.

2. KMeans Clustering: WCSS (Within-Cluster Sum of Squares) and Number of Clusters

- **For k=2**: WCSS = **340.9690**, Number of clusters = **2**

- **For k=3**: WCSS = **247.1201**, Number of clusters = **3**
- **For k=4**: WCSS = **188.9892**, Number of clusters = **4**
- **For k=5**: WCSS = **149.4134**, Number of clusters = **5**
- **For k=6**: WCSS = **125.9978**, Number of clusters = **6**
- **For k=7**: WCSS = **106.0223**, Number of clusters = **7**
- **For k=8**: WCSS = **96.1925**, Number of clusters = **8**
- **For k=9**: WCSS = **87.3271**, Number of clusters = **9**
- **For k=10**: WCSS = **78.9327**, Number of clusters = **10**

Analysis:
The **Within-Cluster Sum of Squares (WCSS)** decreases as the number of clusters increases, which is expected. A lower WCSS means that the points within a cluster are more compact. The **elbow method** can help determine the optimal number of clusters, and typically, the "elbow" in the WCSS curve occurs around **k=6** or **k=7**, which aligns with the silhouette score observations.

3. DBSCAN Clustering Metrics

- **Davies-Bouldin Index (DBI)**: **1.2752**
- **Number of clusters formed (excluding noise)**: **6 clusters**

Cluster Centers:

1. Cluster 0: [-1.4316, -1.6222, -1.0831]
2. Cluster 1: [1.3910, 0.4997, 1.1231]
3. Cluster 2: [-0.9531, 0.1955, -1.1972]
4. Cluster 3: [-0.2757, -0.3501, -0.0404]
5. Cluster 4: [0.4705, 1.5946, -0.4168]

6. Cluster 5: [0.2699, -0.8284, 1.4073]

Number of Points in Each Cluster:

- Cluster 0: **20 points**
- Cluster 1: **41 points**
- Cluster 2: **30 points**
- Cluster 3: **63 points**
- Cluster 4: **27 points**
- Cluster 5: **18 points**

Analysis:

- The **Davies-Bouldin Index (DBI)** value of **1.2752** indicates a relatively moderate clustering performance. A lower DBI is better, and this value suggests that while the clusters are reasonably distinct, there might still be some overlap or variance.
- The **number of clusters formed** by DBSCAN is **6**, excluding noise points (which are marked as -1). DBSCAN has identified a mix of small and larger clusters, with **Cluster 3** having the highest number of points (63) and **Cluster 5** the smallest (18).

4. Visualizations

- **KMeans Clustering (k=6)**: The clusters identified by KMeans can be visualized using a scatter plot, where each point is colored based on its cluster label. The visual results align with the silhouette scores, showing compact and separated clusters.
- **DBSCAN Clustering**: In the scatter plot, clusters are shown in different colors, and noise points are marked distinctly in **red**. This helps to visualize how DBSCAN handles noise and clusters based on density.

Conclusion:

- **Optimal K for KMeans**: Based on both the silhouette score and WCSS, **k=6** seems to be the most suitable choice, with the highest silhouette score and a balanced WCSS curve.
- **DBSCAN**: While DBSCAN provides useful clustering by identifying noise points, the DBI suggests that the clusters may not be as well-separated as those formed by KMeans. However, DBSCAN's advantage lies in its ability to detect arbitrarily shaped clusters and noise, which is beneficial in certain use cases.