Clustering Analysis Report

Optimal Number of Clusters:

The optimal number of clusters was determined using the Silhouette coefficient.

Silhouette Coefficient for k=3: 0.279179
Silhouette Coefficient for k=4: 0.227721
Silhouette Coefficient for k=5: 0.188433
Silhouette Coefficient for k=6: 0.193583

Optimal value of k: 3
Highest Silhouette Coefficient: 0.279179

After performing K-means clustering for different values of k (ranging from 3 to 6), the optimal value of k was found to be 3. This value of k yielded the highest Silhouette coefficient, indicating the best clustering result.

Analysis of Similarity Coefficients:

Step 2: K-means Clustering
The Silhouette coefficient was used to evaluate the quality of the K-means clustering results. The Silhouette coefficient measures how well-separated clusters are, with values closer to 1 indicating better separation

Optimal value of k: 3
Highest Silhouette Coefficient: 0.279179

Step 4: Comparison with Hierarchical Clustering
To compare the K-means clustering results with hierarchical clustering, Jaccard similarity scores were computed between corresponding clusters from both methods. The average Jaccard similarity score among K-means clusters and hierarchical clusters was 0.38517290500049123. This indicates the degree of similarity between the two clustering methods.
Jaccard Similarity score for K-means cluster 1 and best matching Hierarchical cluster 1: 0.620690
Jaccard Similarity score for K-means cluster 2 and best matching Hierarchical cluster 3: 0.490385
Jaccard Similarity score for K-means cluster 3 and best matching Hierarchical cluster 2: 0.044444
Overall average similarity among K-means cluster and Hierarchical cluster: 0.385173

Runtime Analysis : 10 seconds approx.