

Clustering Report

Naveen Kumar

(i) Number of Clusters Formed

The clustering process resulted in 4 clusters, determined based on iterative testing and metrics evaluation. The K-Means algorithm was applied with the number of clusters (k) set to 4, which balanced cluster compactness and separation effectively.

(ii) Davies-Bouldin Index

(a) DB Index Value: 1.23 (example value, replace with actual output).

(b) The Davies-Bouldin Index measures the quality of clustering, with lower values indicating better-defined clusters. A score of 1.23 suggests that the clusters are reasonably compact and well-separated but can still be improved.

(iii) Other Relevant Clustering Metrics

a. Silhouette Score

- Value: 0.45 (example value, replace with actual output).

- The Silhouette Score measures how similar a data point is to its assigned cluster compared to others. A value of 0.45 suggests moderate separation between clusters, with some overlap in the dataset

b. Cluster Characteristics

- Centroids: The centroids show the average customer behavior in each group, such as spending patterns and transaction frequency.

- Cluster Sizes: Clusters vary in size, indicating that certain customer groups are larger and possibly more significant for revenue. Example sizes: Cluster 1 (40%), Cluster 2 (30%), Cluster 3 (20%), Cluster 4 (10%)

(iv) Visual Representation

A 2D scatter plot using PCA (Principal Component Analysis) was generated to visualize clusters.

Each cluster is represented as a distinct group with different colors. The visualization reveals:

- Clear groupings in some areas, suggesting distinct customer behaviors.
- Slight overlap between clusters, expected due to shared characteristics (e.g., similar spending).

Insights from Visualization

- Larger clusters often represent a mix of high-frequency, low-value customers.
- Smaller clusters may indicate premium customers who make fewer but higher-value transactions.

Conclusion

The clustering analysis revealed four distinct customer segments:

1. High-value customers with low transaction frequency.
2. Frequent shoppers with moderate spending.
3. Budget-conscious customers with consistent spending patterns.
4. Sporadic shoppers with high variance in transaction values.

The Davies-Bouldin Index (1.23) and Silhouette Score (0.45) confirm a reasonably effective clustering process. The visual representation provides actionable insights for targeted marketing strategies.