Clustering Results :

Overview

In this analysis, we performed customer segmentation using K-Means clustering based on transaction data. The goal was to identify distinct customer groups based on their spending behaviour and transaction frequency.

Clustering Details

1. Number of Clusters Formed:
    - The optimal number of clusters determined through the Elbow method was **4**.
2. Davies-Bouldin Index (DB Index):
    - The calculated Davies-Bouldin Index for the chosen number of clusters (4) is **0.9881**.
    - DB Index: 0.9880667545355788
    - A lower DB Index indicates better clustering quality, suggesting that the clusters are well-separated and distinct.
3. Silhouette Score:
    - The Silhouette Score for the optimal number of clusters is **0.3964**.
    - Silhouette Score: 0.39637746297313897
    - This score ranges from -1 to 1, where a value closer to 1 indicates that the points are well-clustered. A score around 0.4 suggests that the clusters are reasonably separated but may not be perfectly distinct.

Visualizations

1. Elbow Method Visualization:
    - A plot was generated to visualize the inertia values against the number of clusters. The "elbow" point, where the rate of decrease sharply changes, indicates the optimal number of clusters.
2. Customer Segmentation Scatter Plot:
    - A scatter plot was created to visualize the segmentation of customers based on their total spending and the number of transactions. Each cluster is represented by a different colour, allowing for easy identification of distinct customer groups.

Conclusion

The clustering analysis successfully identified 4 distinct customer segments based on their transaction behaviour. The metrics used, including the Davies-Bouldin Index and Silhouette Score, indicate that the clusters are reasonably well-defined, although there is room for improvement in separation. Further analysis could involve exploring the characteristics of each cluster to tailor marketing strategies effectively.