Customer Segmentation Analysis Report

Overview

Hi, this summarises my work on the customer segmentation analysis project. Using clustering techniques, I used e-commerce transaction data to group customers into different segments. The main focus of this report is to explain the number of clusters we formed, the Davies-Bouldin Index value, and some other key metrics that helped evaluate the clustering results.

Number of Clusters Formed

After analyzing the data and testing different configurations, we decided on **10 clusters** as the best number for segmenting customers. I used tools like the elbow curve and the Davies-Bouldin Index to figure this out, which helped strike a balance between having enough detail and not over-complicating the model. These clusters helped us group customers into categories like high-spending, frequent shoppers and budget-conscious, less active buyers. This way, we can better understand their behaviour and preferences.

Davies-Bouldin Index Value

The clustering model got a **Davies-Bouldin Index (DB Index) score of 1.4851**, which means the clusters are fairly distinct and compact. The DB Index measures how similar the customers in a cluster are to each other versus how different they are from other clusters. Since a lower DB Index is better, this score shows that the clustering worked well and the groups make sense.

Other Relevant Clustering Metrics

1. **Silhouette Score**: The Silhouette Score for the model was **0.3247**. While this isn't super high, it still shows that most customers were placed in the right clusters. Some overlap happened at the edges, which is expected when dealing with real-world data.

2. **Elbow Curve Analysis**: The elbow curve was really useful for figuring out the number of clusters. It showed that after 10 clusters, the improvement in the model wasn't significant, so we stopped there to keep things simple and effective.

3. **PCA Visualization**: I also used PCA to reduce the data's dimensions and visualize the clusters in 2D. This made it easier to see the groups. While some clusters were close to each other, the high-value ones, like clusters 8 and 9, stood out clearly.

Conclusion

In conclusion, the 10-cluster segmentation model seems to work well, based on the metrics I looked at. The DB Index of 1.4851, Silhouette Score of 0.3247, and the elbow curve analysis

all support this choice. These clusters can help the business make smarter decisions in areas like marketing, inventory management, and keeping customers engaged. Let me know if there are more ways I can improve this analysis!