Customer Segmentation Using K-Means Clustering

Objective

The objective of this analysis is to segment customers based on their purchasing behavior using K-Means clustering. This segmentation will help identify distinct customer groups and support targeted business strategies.

Methodology

1. Data Used:

The analysis was performed using three key features from the customer_profiles dataset:

• **Price**: The total price of products purchased by each customer.

• **Quantity**: The total number of products purchased by each customer.

• **TotalValue**: The total value of transactions for each customer.

2. Preprocessing:

• The features were normalized using StandardScaler to ensure uniform scaling across all variables.

3. Dimensionality Reduction:

• Principal Component Analysis (PCA) was applied to reduce the dimensions to 2 components for visualization while preserving most of the variance in the data.

4. Clustering Algorithm:

• K-Means clustering was applied to segment customers into distinct groups. The number of clusters (k) was tested in the range of 2 to 10.

5. Evaluation Metric:

• The Davies-Bouldin Index (DB Index) was used to evaluate cluster quality. A lower DB Index indicates better-defined clusters. The optimal number of clusters was selected based on the lowest DB Index.

Results

1. Optimal Number of Clusters:

The optimal number of clusters was found to be **2**, with a Davies-Bouldin Index value of **0.730**.

2. Cluster Characteristics:

The two clusters exhibit the following characteristics:

• **Cluster 0**:

• Customers with high transaction values.

• Representing high spenders or frequent buyers.

• **Cluster 1**:

• Customers with moderate transaction values.

• Representing occasional or budget-conscious buyers.

3. Visualization:

The scatter plot below illustrates the clusters in PCA-reduced 2D space. Each point represents a customer, and the colors indicate the assigned cluster.

Customer Segments (K-Means Clustering)
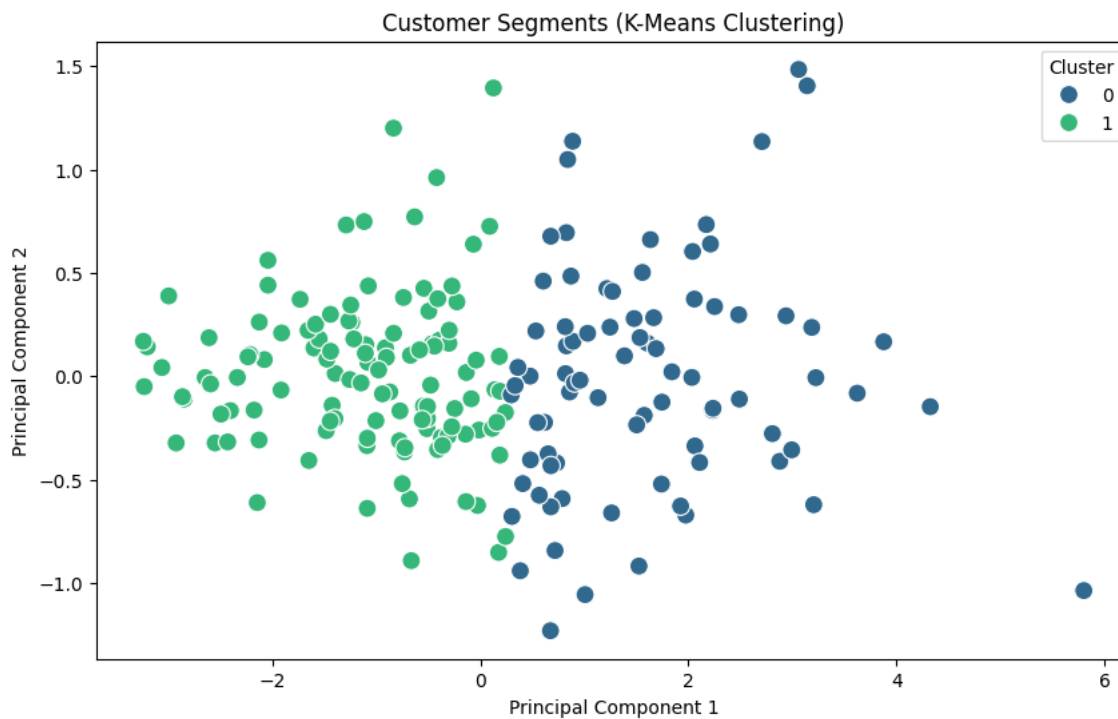
Conclusion

This analysis grouped customers into two distinct clusters based on their purchasing behavior:

• **Cluster 0**: High spenders.

• **Cluster 1**: Moderate spenders.

These insights can guide the business in implementing targeted marketing strategies, loyalty programs, and personalized recommendations for each customer segment.