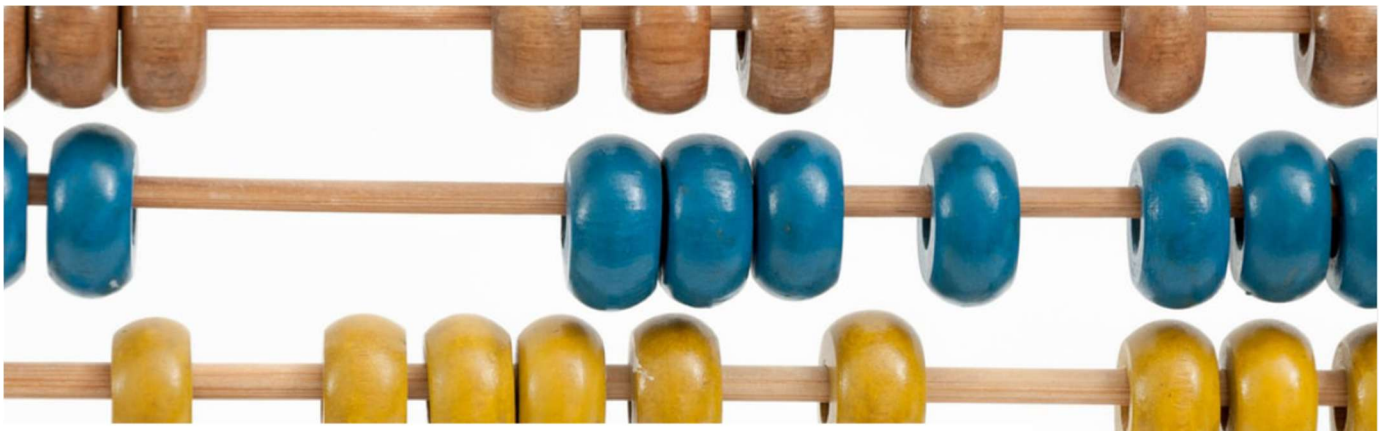


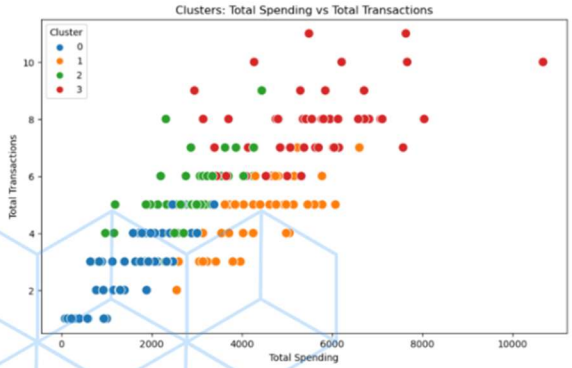
Clustering Analysis Using K-Means

Exploring the results of K-Means clustering, including cluster formation and performance evaluation metrics.

Keertipriya Bellary

Clustering Results Report

An In-Depth Look at the Clustering Analysis and Findings



Chosen Clustering Algorithm

The K-Means algorithm was selected for this clustering analysis due to its simplicity and efficiency in handling large datasets. K-Means works by partitioning the data into K distinct clusters, where each data point belongs to the cluster with the nearest mean. This method is particularly effective for our dataset, allowing us to derive meaningful groupings.

Number of Clusters Formed

Four clusters were determined through a combination of manual selection and the elbow method. The elbow method is a technique used to identify the optimal number of clusters by plotting the sum of squared distances from each point to its assigned cluster center. Observing the 'elbow' point on the graph indicates the ideal number of clusters, which in this case is four.

Davies-Bouldin Index (DB Index)

Evaluating Clustering Performance

1.

DB Index Value

The current DB Index value is recorded at 1.29, which is a crucial metric in evaluating clustering quality.

2.

Interpretation of DB Index

A DB Index value of 1.29 is considered relatively low, indicating effective clustering performance. This suggests that the clusters formed are both compact and well-separated, which is desirable in clustering analysis.

3.

Importance of Low DB Index

A low DB Index signifies that the data points within each cluster are close to each other, while the clusters themselves are distant from one another. This quality is essential for the reliability of the clustering results.

4.

Application of DB Index

The DB Index can be utilized in various applications of clustering, such as market segmentation, image processing, and social network analysis. Its ability to quantify cluster quality aids in selecting the most appropriate clustering algorithm.

Other Relevant Clustering Metrics

Analyzing cluster metrics to enhance business strategies

Silhouette Score: 0.23

Moderate cluster separation

The Silhouette Score of 0.23 indicates a moderate level of separation between clusters. This suggests that while the clusters are somewhat distinct, there may be overlaps that could affect the

Inertia: 474.45

Cluster compactness

The inertia value of 474.45 reflects the sum of squared distances of samples to their closest cluster center, with a lower inertia value indicating tighter clustering. This suggests that while the clusters are

Cluster 0: 62 customers

Largest cluster size

Cluster 0, with 62 customers, is the largest grouping, indicating a significant segment that may share common characteristics or behaviors. This cluster could be a primary target for marketing strategies

Cluster 1: 44 customers

Second largest cluster

Cluster 1 consists of 44 customers, representing a smaller but noteworthy segment. Understanding the dynamics of this cluster could yield insights into niche market opportunities or tailor-made

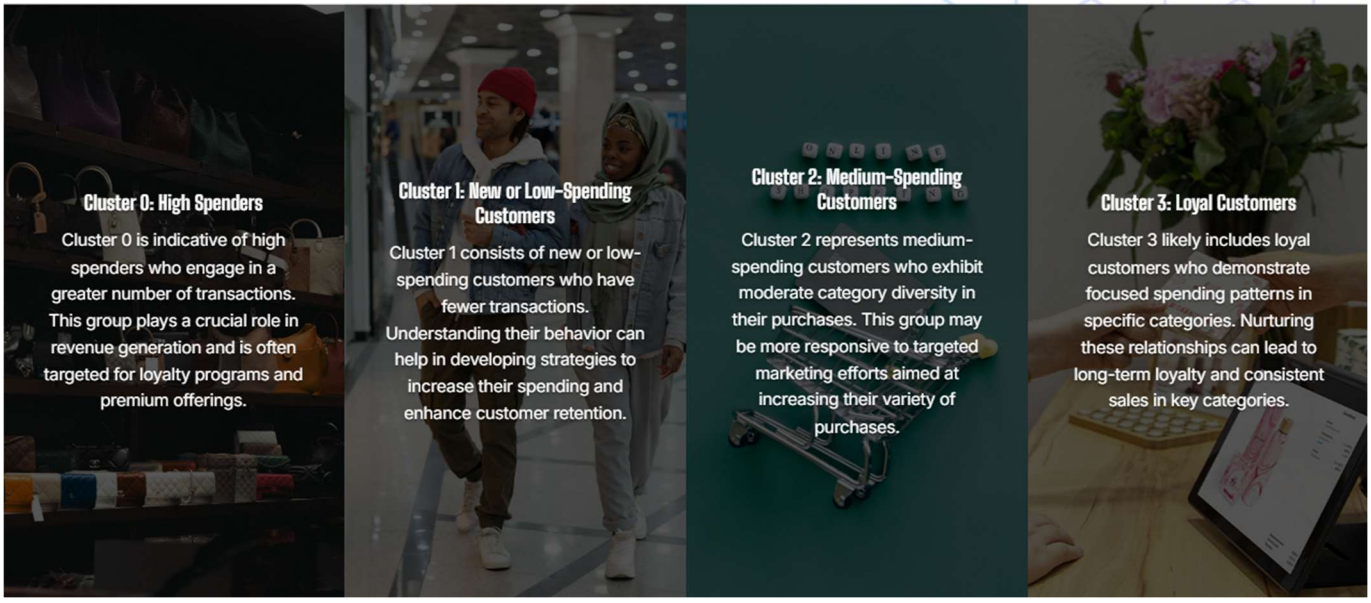
Cluster 2: 47 customers

Moderate cluster size

With 47 customers, Cluster 2 offers another moderate grouping that could present unique challenges and opportunities. Analyzing this cluster's preferences may improve customer engagement and

Insights from the Clusters

Understanding Customer Segments for Strategic Marketing



Cluster 0: High Spenders

Cluster 0 is indicative of high spenders who engage in a greater number of transactions. This group plays a crucial role in revenue generation and is often targeted for loyalty programs and premium offerings.

Cluster 1: New or Low-Spending Customers

Cluster 1 consists of new or low-spending customers who have fewer transactions. Understanding their behavior can help in developing strategies to increase their spending and enhance customer retention.

Cluster 2: Medium-Spending Customers

Cluster 2 represents medium-spending customers who exhibit moderate category diversity in their purchases. This group may be more responsive to targeted marketing efforts aimed at increasing their variety of purchases.

Cluster 3: Loyal Customers

Cluster 3 likely includes loyal customers who demonstrate focused spending patterns in specific categories. Nurturing these relationships can lead to long-term loyalty and consistent sales in key categories.

Cluster Analysis Visualizations

Insights from Cluster Scatter Plot and Distribution

Total Spending vs Total Transactions

Distinct clusters indicate varied customer behavior.

The cluster scatter plot demonstrates how customers are segmented based on their Total Spending and Total Transactions. The distinct regions occupied by each cluster suggest differing spending habits and transaction frequencies among customers, which can guide targeted marketing strategies.

62 customers in Cluster 0

Largest cluster size highlights potential focus area.

Cluster 0, containing 62 customers, is the largest group identified. This insight indicates a significant segment of customers that could be targeted for loyalty programs or special promotions, leveraging their potential to drive sales.

Silhouette score indicates overlap

Some clustering overlaps warrant further investigation.

The silhouette score suggests that there is some overlap between the clusters, indicating that certain customer segments may not be clearly defined. Further analysis could be beneficial to refine the clusters and enhance targeting efforts.