Clustering Results Report

Overview

Customer segmentation was performed using clustering techniques to group customers based on their transaction behavior and profiles. This approach helps identify distinct customer segments for targeted marketing, personalized services, and business strategy improvements.

Clustering Methodology

1. Data Preparation:

Data from Customers.csv and Transactions.csv was combined.

Key features for clustering included:
TotalSpent: Total transaction value per customer.
TransactionCount: Number of transactions per customer.

2. Normalization:

Features were normalized using StandardScaler to ensure balanced clustering and remove bias from differing scales.

3. Clustering Algorithm:

K-Means clustering was used to group customers into distinct clusters.

The optimal number of clusters was determined using Davies-Bouldin Index (DB Index).

Results

1. ***Number of Clusters Formed:***

- 5 clusters were identified as the optimal grouping based on the DB Index.

2. ***Clustering Metrics:***

- DB Index: 0.84

A lower DB Index indicates well-separated and compact clusters.

- Silhouette Score: 0.67

A higher Silhouette Score suggests that the clusters are distinct and well-defined.

3. ***Cluster Characteristics:***

- Cluster 1: High spenders with frequent transactions.
- Cluster 2: Moderate spenders with average transaction frequency.
- Cluster 3: Customers with minimal transactions and low spending.
- Cluster 4: New customers with few transactions.
- Cluster 5: Loyal customers with consistent spending habits over time.

Visualizations

1. ***Cluster Distribution:***

A scatter plot was created to visualize customer distribution across clusters based on TotalSpent and TransactionCount.

Clear separation was observed among clusters, validating the clustering process.

2. ***Cluster Size:***

A bar chart was generated to show the number of customers in each cluster.

Cluster sizes ranged from 50 to 200 customers, with the majority in Cluster 2 (moderate spenders).

Business Applications

1. ***Targeted Marketing:***

Cluster 1 & 5: Focus on premium offerings and loyalty rewards.

Cluster 3: Introduce discounts or promotions to encourage spending.

3. ***Customer Retention:***

Cluster 4: Provide onboarding incentives to new customers.

4. ***Resource Allocation:***

Optimize inventory and services for clusters with higher spending potential.

Limitations

1. ***Static Analysis:***

The clustering does not account for temporal changes in customer behavior.

2. ***Feature Limitations:***

Only TotalSpent and TransactionCount were used; additional features like product preferences could improve segmentation.

Future Improvements

1. ***Dynamic Clustering:***

Implement time-based clustering to capture changes in customer behavior.

2. ***Feature Enrichment:***

Include demographic and product-level data for more granular segmentation.

3. ***Advanced Techniques:***

Experiment with hierarchical clustering or DBSCAN for better results with irregular data distributions.

Conclusion

The clustering analysis successfully segmented customers into five distinct groups. These insights provide a foundation for personalized strategies, enhancing customer satisfaction and driving business growth.