

Customer Segmentation Using K-Means Clustering

Executive Summary

This report presents the results of a **K-Means clustering analysis** performed to segment customers based on their purchasing behavior and geographical location. The analysis aimed to identify distinct customer groups to inform targeted marketing strategies.

Key highlights:

- **Five clusters** were identified.
 - Clustering performance was evaluated using the **Silhouette Score** and the **Davies-Bouldin Index**.
 - Results indicate **moderately well-defined clusters**, providing valuable insights into customer profiles across different regions.
-

Table of Contents

1. **Introduction**
 2. **Methodology**
 - 2.1. Data Description
 - 2.2. Data Preprocessing
 - 2.3. Clustering Algorithm
 - 2.4. Evaluation Metrics
 3. **Results**
 - 3.1. Evaluation Metrics Results
 - 3.2. Cluster Profiles
 4. **Discussion**
 - 4.1. Interpretation of Clusters
 - 4.2. Insights
 5. **Conclusion**
 6. **Recommendations**
-

Introduction

Understanding customer behavior is crucial for businesses aiming to enhance **customer satisfaction** and increase **profitability**. Clustering analysis helps segment customers into groups with similar characteristics.

This report details the application of **K-Means clustering** to a dataset containing customer purchasing behavior and geographical information. The objective is to identify **natural groupings of customers** to facilitate targeted marketing and personalized services.

Methodology

2.1 Data Description

The dataset includes the following features for each customer:

- **Recency:** Number of days since the customer's last purchase.
- **Frequency:** Total number of purchases made by the customer.
- **Monetary:** Total monetary value of the customer's purchases.
- **Geographical Indicators:**
 - **Asia**
 - **Europe**
 - **North America**
 - **South America**Each geographical indicator is a binary variable indicating the customer's region.

2.2 Data Preprocessing

The following preprocessing steps were performed:

1. **Handling Missing Values:** Checked for and addressed missing/null values in the dataset.
2. **Scaling Features:** Standardized the features using **z-score normalization** to ensure equal contribution to distance calculations.
3. **Encoding Categorical Variables:** Ensured geographical indicators were properly encoded as binary variables.

2.3 Clustering Algorithm

- **Algorithm Used:** K-Means Clustering
- **Number of Clusters (k):** 5 (chosen based on domain knowledge and preliminary analysis)
- **Initialization Method:** K-Means++ to improve convergence.
- **Number of Initializations (n_init):** 10 (to avoid local minima convergence).
- **Random State:** Set for reproducibility.

2.4 Evaluation Metrics

- **Silhouette Score:**
 - Measures the cohesion and separation of clusters.
 - Range: [-1, 1], with higher scores indicating better-defined clusters.

- **Davies-Bouldin Index (DB Index):**
 - Evaluates intra-cluster similarity and inter-cluster differences.
 - Range: $[0, \infty)$, with lower scores indicating better clustering.
-

Results

3.1 Evaluation Metrics Results

- **Silhouette Score: 0.4846** (moderately high, indicating reasonably well-defined clusters).
 - **Davies-Bouldin Index: 0.8381** (relatively low, indicating good separation between clusters).
-

3.2 Cluster Profiles

K-Means Cluster Profiles:

	Recency	Frequency	Monetary	Asia	Europe	North America	South America
Cluster_KMeans							
0	275.916667	1.333333	764.795000	0.333333	0.166667	0.333333	0.166667
1	58.491228	5.298246	3827.558947	0.000000	0.000000	0.000000	1.000000
2	68.000000	5.690476	3560.888810	0.000000	0.000000	1.000000	0.000000
3	66.354167	4.791667	3414.071458	0.000000	1.000000	0.000000	0.000000
4	55.425000	5.325000	3730.360000	1.000000	0.000000	0.000000	0.000000

The table below summarizes the average values of key features for each cluster:

Discussion

4.1 Interpretation of Clusters

- **Cluster 0:**
 - **Profile:** High Recency, Low Frequency, and Monetary Value; diverse geographical distribution.
 - **Interpretation:** Inactive or churned customers.
- **Cluster 1:**
 - **Profile:** Low Recency, High Frequency, and Monetary Value; exclusively South America.
 - **Interpretation:** Loyal and valuable customers from South America.
- **Cluster 2:**
 - **Profile:** Similar to Cluster 1; exclusively North America.
 - **Interpretation:** Key customers from North America.
- **Cluster 3:**
 - **Profile:** Moderate engagement; exclusively Europe.

- **Interpretation:** Moderate-value customers with growth potential.
 - **Cluster 4:**
 - **Profile:** Low Recency, High Frequency, and Monetary Value; exclusively Asia.
 - **Interpretation:** Highly engaged and valuable customers from Asia.
-

4.2 Insights

1. **Regional Segmentation:**
 - Clusters 1 to 4 show strong regional purchasing behaviors.
 - Tailored regional marketing strategies can enhance engagement.
 2. **Customer Engagement Levels:**
 - Cluster 0 requires re-engagement strategies.
 - Clusters 1, 2, and 4 are high-value segments that merit loyalty programs.
 3. **Potential for Growth:**
 - Cluster 3 represents an opportunity to convert moderately engaged customers into high-value ones.
-

Recommendations

1. **Re-Engagement Campaigns for Cluster 0:**
 - Use discounts and personalized communications to reactivate inactive customers.
 2. **Loyalty Programs for High-Value Clusters:**
 - Strengthen loyalty programs tailored to regional preferences for Clusters 1, 2, and 4.
 3. **Upselling and Cross-Selling Opportunities:**
 - Analyze purchasing patterns in high-value clusters to identify products for upselling.
 4. **Increase Engagement in Cluster 3:**
 - Target European customers with promotional offers and cultural customization.
 5. **Regional Marketing Strategies:**
 - Design campaigns that consider **language, cultural nuances, and regional holidays**.
-

Conclusion

The K-Means clustering analysis successfully segmented customers into **five distinct clusters** based on purchasing behavior and geographical location. Evaluation metrics indicate **reasonably well-defined clusters** that provide actionable insights for marketing and customer retention.
