CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING FOR A U.S. ONLINE RETAILER

1. Background and Problem Statement

A mid-sized online retail company based in the United States observed diminishing returns on blanket marketing strategies. Despite high traffic and frequent promotional campaigns, conversion rates and average order values varied significantly across customer groups. The business lacked a structured approach to understanding behavioral differences among its shoppers. To enable precision marketing, the company needed to segment its customers based on purchasing patterns and interaction history. The goal was to apply unsupervised data mining techniques in SPSS to define meaningful segments for targeted marketing, personalized offers, and product recommendation strategies.

2. Objectives

- To apply K-Means clustering in SPSS to identify distinct customer groups based on behavior and purchase history
- To understand key traits that differentiate each segment (e.g., high-value loyalists, discount-seekers, infrequent browsers)
- To assist the marketing and product teams in developing targeted campaigns and bundle offerings for each cluster
- To build a scalable segmentation framework that can be reused with new quarterly datasets

3. Methodology

3.1 Data Collection and Preparation

- Dataset: Transaction data from January to December 2023 (95,000 customers)
- Source: Internal CRM system and order history export
- Variables included:
    - Total spending (12 months)
    - Number of purchases
    - Average order value

- Number of items per order
- Return frequency
- Response to discounts (% orders purchased using promo codes)
- Days since last purchase

Data was cleaned in SPSS by handling missing values, normalizing numerical variables, and transforming skewed distributions where necessary.

3.2 K-Means Clustering Implementation

- SPSS's TwoStep Cluster and K-Means clustering procedures were tested
- Optimal number of clusters determined using Elbow Method and Silhouette scores
- Final model used K=4 clusters
- Standardized Z-scores of input variables used for improved clustering accuracy

3.3 Cluster Profiling Post-clustering, descriptive statistics were run for each cluster to define dominant characteristics. These included average spend, order count, discount use, and recency.

4. Results and Cluster Definitions

Cluster	Description	Size	Key Traits
C1	High-Spenders & Loyalists	21%	High spend, frequent purchases, low return rates
C2	Price-Sensitive Deal Seekers	32%	High use of discounts, low order value, responsive to promos
C3	Occasional Browsers	29%	Long gaps between orders, mid-range order values
C4	High Return Rate & Low Loyalty	18%	Moderate spending but frequent returns and low repeat rate

Key Insights

- 53% of the customer base was either highly price-sensitive or at risk of churn
- Only 21% formed the revenue-driving loyal customer segment
- Customized campaigns are required for C2 and C4 to reduce discount dependency and return rate, respectively

5. Recommendations

- Launch a tiered loyalty program targeting Cluster C1 to enhance retention
- Redesign promotional strategy for C2 by introducing bundled offers instead of percentage-based discounts
- Offer tailored re-engagement content for C3 via email and push notifications
- Introduce stricter return policies or product education for C4 to reduce losses from repeated returns

6. Deliverables

- SPSS .sav file containing the final cluster assignments
- Cluster profiling report in APA format
- Excel dashboard highlighting key metrics by segment
- Guidance document on integrating clustering routine into quarterly customer review process

7. Stakeholder Relevance

Corporate:

- Valuable for marketing, product, and CRM teams for designing targeted offers and enhancing customer experience
- Applicable for campaign-level performance tracking per cluster

Academic:

- A practical case in unsupervised learning using SPSS for MBA and Data Analytics courses
- Demonstrates real-world application of K-Means, cluster validation, and profiling techniques