RETAIL CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING IN SPSS

1. Background and Problem Statement

A mid-sized retail chain in the U.S. was experiencing stagnation in repeat purchases despite steady footfall and online traffic. The marketing team lacked clarity on customer diversity and could not tailor promotions effectively. Without knowing who their high-value, one-time, or disengaged customers were, resources were spread thin across generic campaigns. To address this, the company decided to use SPSS to perform customer segmentation using K-Means clustering on transactional data to inform loyalty, retention, and upselling strategies.

2. Objectives

- To segment customers based on RFM (Recency, Frequency, Monetary) metrics using K-Means Clustering in SPSS
- To identify distinct behavioral clusters such as high-value loyal customers, low-frequency spenders, and recent one-time buyers
- To assist the marketing team in designing personalized email campaigns and loyalty rewards
- To generate an interpretable report that links customer profiles to specific business actions

3. Methodology

3.1 Dataset Description

- Source: 12-month point-of-sale (POS) and e-commerce transactional data
- Variables used:
    - **Recency**: Days since last purchase
    - **Frequency**: Total number of transactions per customer
    - **Monetary Value**: Total spend in USD
    - **Customer ID** (for identification, not included in clustering)

3.2 Data Preparation in SPSS

- Removed incomplete or duplicate records

- Standardized R, F, and M using Z-scores to avoid scale bias
- Checked skewness and kurtosis to verify normal distribution assumptions
- Verified outliers using boxplots

3.3 K-Means Clustering

- Chose K=4 after evaluating the elbow plot from ANOVA between-groups sum of squares
- Iterated K-means clustering multiple times to ensure consistency
- Cluster centers evaluated and named based on dominant RFM characteristics

4. Results

Cluster Profiles:

- **Cluster 1 – High-Frequency Spenders (22%)**
  - Low Recency (recent purchases), High Frequency, High Monetary
  - Suggested as core loyal segment
- **Cluster 2 – Dormant High Spenders (15%)**
  - High Recency, Low Frequency, High Monetary
  - Previously active, potential to be re-engaged
- **Cluster 3 – Casual Buyers (40%)**
  - Moderate across all RFM metrics
  - Targetable with volume discounts or bundle offers
- **Cluster 4 – One-Time Low Spenders (23%)**
  - High Recency, Low Frequency, Low Monetary
  - Cost-sensitive or disengaged segment

5. Interpretation and Insights

- The majority of spending was concentrated in Clusters 1 and 2, who together made up only 37% of customers but contributed 68% of the total revenue.
- Cluster 4 customers rarely returned; a personalized post-purchase communication strategy is required to understand their disinterest.

- Cluster 2 had the highest value potential if successfully re-engaged.
- Visualizations (dendrograms, cluster scatter plots, and RFM score distributions) supported clear differentiation.

6. Recommendations

- Launch a **loyalty program** targeting Cluster 1 to increase retention
- Deploy **reactivation emails** with special coupons for Cluster 2
- Offer **educational content or discounts** for Cluster 4 to reduce churn
- Use cluster labels as campaign segments in the company's CRM system

7. Future Work

- Integrate demographic data (e.g., age, ZIP code) to refine clusters
- Include channel data (online vs. offline) to identify behavioral differences
- Run a quarterly update to monitor shifts in cluster memberships
- Test the performance of tailored campaigns linked to each segment

8. Stakeholder Relevance

Academic:

- Applicable for teaching K-Means clustering, RFM segmentation, and consumer behavior analysis
- Useful in courses on marketing analytics, business intelligence, and applied statistics

Corporate:

- Helps retail businesses improve targeting, retention, and budget allocation
- Enables integration of clustering insights into marketing platforms and loyalty engines