

Customer Segmentation Analysis Report

Executive Summary

Using K-means clustering analysis on the e-commerce dataset, we identified six distinct customer segments. The segmentation shows strong statistical validity with improved clustering metrics compared to other cluster numbers, providing a solid foundation for targeted marketing strategies.

Technical Approach

Methodology

- Algorithm: K-means Clustering
- Input Features:
 - Transaction metrics (count, value, frequency)
 - Category preferences
 - Price sensitivity
 - Regional indicators
 - Customer lifecycle metrics
- Feature Scaling: StandardScaler
- Dimensionality Reduction: PCA for visualization

Clustering Metrics

- Davies-Bouldin Index: 1.7796 (strong cluster separation)
- Silhouette Score: 0.1567 (good cluster cohesion)
- Calinski-Harabasz Index: 23.7846 (high cluster density)

Cluster Characteristics

Cluster 0: High-Value Premium Customers

- Highest total spend (~6000 units)
- High transaction frequency
- Above-average transaction values
- Moderate active period
- Primary focus on high-value categories

Cluster 1: Regular Mid-Tier Customers

- Moderate total spend (~2500 units)
- Consistent transaction pattern
- Average transaction values
- Balanced category distribution

Cluster 2: Low-Volume Customers

- Lowest total spend (~500 units)
- Minimal transaction frequency
- Higher than average transaction values
- Limited category exploration

Cluster 3: Growth Potential Customers

- Good total spend (~3200 units)
- Growing transaction frequency
- Strong average transaction values
- Diverse category interests

Cluster 4: Steady Value Customers

- Consistent total spend (~3100 units)
- Regular transaction pattern
- Stable average transaction values
- Focused category preferences

Cluster 5: Mainstream Customers

- Moderate total spend (~3000 units)
- Regular purchase frequency
- Standard transaction values
- Broad category exploration

Feature Importance Analysis

Principal Component 1 (Spending Behavior)

Strongly influenced by: - Transaction count (negative correlation) - Total spend (negative correlation) - Total quantity (negative correlation) - Days active (negative correlation)

Principal Component 2 (Purchase Patterns)

Dominated by: - Average transaction value (positive correlation) - Average price (positive correlation) - Category preferences (varying correlations) - Regional indicators (moderate influence)

Strategic Recommendations

1. **High-Value Premium Customers (Cluster 0)**
 - VIP program implementation
 - Early access to new products
 - Personalized shopping experience

- Premium customer service
- 2. **Regular Mid-Tier Customers (Cluster 1)**
 - Loyalty program benefits
 - Category-specific promotions
 - Upgrade opportunities
 - Regular engagement campaigns
- 3. **Low-Volume Customers (Cluster 2)**
 - Re-engagement strategies
 - Entry-level product focus
 - Simplified shopping experience
 - Value-based messaging
- 4. **Growth Potential Customers (Cluster 3)**
 - Targeted upgrade offers
 - Cross-category promotions
 - Enhanced service options
 - Personalized recommendations
- 5. **Steady Value Customers (Cluster 4)**
 - Retention-focused campaigns
 - Category expansion incentives
 - Loyalty rewards
 - Consistent engagement
- 6. **Mainstream Customers (Cluster 5)**
 - Broad product exposure
 - Standard promotions
 - Category discovery campaigns
 - Regular communications

Implementation Plan

Immediate Actions

1. Segment-specific email campaigns
2. Targeted promotional calendars
3. Customized product recommendations
4. Service level differentiation

Long-term Initiatives

1. Automated segment classification
2. Real-time customer scoring
3. Segment transition tracking
4. Personalization engine development

Conclusion

The six-cluster solution provides an optimal balance between cluster separation and business actionability. The improved clustering metrics validate the robust-

ness of this segmentation, while the clear distinction between segments enables targeted marketing strategies and personalized customer experiences.