

Unlocking Growth: Data-Driven Customer Segmentation & Product Association

Heeje Yoo – August 13, 2025

Executive Summary

We applied K-Means clustering and the Apriori algorithm to customer transaction data, identifying five distinct segments and strong product pairings. The results offer clear strategies to increase retention, raise average order value, and deliver personalized experiences for measurable revenue growth.

Methodology

Phase 1: **Customer Segmentation** – Engineered Recency, Frequency, Monetary (RFM) features; scaled data; selected optimal k via Elbow & Silhouette; applied K-Means to create five segments.

Phase 2: **Market Basket Analysis** – Used Apriori to detect product co-purchases, ranked by support, confidence, and lift to highlight cross-sell and bundling opportunities.

Customer Segments

- **Champions** – Most valuable: frequent (8.6x), high spend (\$2,287), recent purchases (10 days).
- **Rising Stars** – Recent (17 days), moderate spend (\$1,185), lower frequency.
- **Potential Loyalists** – High spend (\$1,817), frequent (7.3x), but last purchase ~55 days ago.
- **Needs Attention** – Previously active, inactive for 101 days.
- **At Risk** – Longest recency (103 days), lowest spend/frequency.

Top Product Associations (Lift)

1. Prod 8 → Prod 1 (1.82)
2. Prod 3 → Prod 18 (1.68)
3. Prod 15 → Prod 1 (1.61)
4. Prod 9 → Prod 2 (1.61)
5. Prod 27 → Prod 19 (1.60)

Recommendations

- **Targeted Marketing**
 - Champions: Loyalty rewards, early access.
 - Rising Stars: Follow-up discounts to boost frequency.
 - Potential Loyalists & Needs Attention: “We miss you” campaigns tied to past purchases.
- **Smart Bundling**
 - Highlight frequent pairs (e.g., 8+1) online and in-store.
 - Bundle discounts to lift order value.
 - Proximity placement in retail settings.

Conclusion

Focusing on high-value segments and leveraging product affinities enables a shift from mass marketing to personalized, high-ROI engagement, increasing lifetime value and transaction size.

Appendix

