Customer Segmentation Hybrid Recommender for E-commerce
Chien-Wei WENG
January 2026

Abstract

E-commerce grocery platforms face the challenge of personalizing recommendations across diverse customer behaviors. This project investigates whether customer segmentation improves recommendation quality compared to global models. Using the Instacart dataset (206,209 users, 49,688 products, 3.3M orders), we apply K-means clustering with PCA dimensionality reduction to identify five distinct customer segments. We compare three recommendation approaches—Baseline (popularity), Collaborative Filtering (SVD), and Hybrid (CF + Content-Based)—in both global and segment-specific configurations. Rigorous temporal splitting prevents data leakage. Statistical testing (paired t-tests, $n = 2{,}000$ stratified users) shows segment-specific models significantly outperform global counterparts: CF achieves **+213.8%** F1@5 improvement ($p < 0.05$), Hybrid achieves **+48.4%** ($p < 0.001$). Results demonstrate that behavioral segmentation measurably enhances recommendation quality in sparse, high-dimensional grocery transaction data.

1. Introduction

Research Question: Does explicit customer segmentation—grouping users by behavioral similarity—improve recommendation accuracy by tailoring predictions to distinct shopping profiles?

Business Relevance: Personalized recommendations drive conversion rates and customer lifetime value. Segment-aware models enable targeted marketing strategies, optimized inventory management, and improved customer satisfaction.

2. Methodology

2.1 Dataset

Source: Instacart Market Basket Analysis (Kaggle)

Scale: 206,209 users, 49,688 products, 3.3M orders

Temporal Split: Last order (N) for test, second-to-last (N-1) for validation, all prior orders for training—ensuring no future information leakage

2.2 Phase 1: Customer Segmentation

Feature Engineering (163 features):

- **RFM Metrics:** Recency (days since last order), Frequency (total orders), Monetary (total items purchased)
- **Behavioral Patterns:** Average basket size, reorder ratio, order interval, preferred shopping day and hour
- **Category Preferences:** Department and aisle purchase distributions (21 departments, 134 aisles)

Dimensionality Reduction: PCA reduces 163 features to 99 components (80.43% variance retained) to address curse of dimensionality.

Clustering: K-means algorithm ($K = 5$) selected via silhouette score (0.022) and Davies-Bouldin index (4.15). Weak cluster separation reflects behavioral spectrum in grocery data rather than discrete groups.

Segments Identified:

Segment	**Size**	**Characteristics**
Power Users	45.0%	High frequency (21.8 orders), large baskets (10.6 items), fresh food focus
Bulk Shoppers	36.0%	Low frequency (10.6 orders), largest baskets (11.1 items), variety seekers
Routine Snackers	13.4%	Regular purchases (15.9 orders), beverage/snack focused (26.0%/22.4%)
Household Essentials	4.4%	Utilitarian (11.3 orders), household goods (20.6%)
Alcohol Enthusiasts	1.1%	Niche, alcohol-specialized (43.8%)

Table 1: Customer Segments

2.3 Phase 2: Recommendation Models

Three Approaches (Global vs Segment-Specific):

(1) Baseline: Popularity-based ranking (overall or per-segment top-N products)

(2) Collaborative Filtering (CF): SVD matrix factorization with log-transformed purchase frequencies:

$$r_{ui} = \log(1 + \text{frequency}_{ui})$$

Default hyperparameters (100 factors, 20 epochs) to avoid tuning overhead.

(3) Hybrid: Linear combination of CF and Content-Based Filtering:

$$\text{score}_{\text{Hybrid}}(u, i) = \alpha \cdot \text{score}_{\text{CF}}(u, i) + (1 - \alpha) \cdot \text{score}_{\text{CBF}}(u, i)$$

where $\alpha = 0.5$ (equal weighting). CBF uses binary product feature vectors (department, aisle) with cosine similarity.

2.4 Phase 3: Evaluation

Test Setup:

- Stratified sampling: 2,000 test users (400 per segment)
- Metrics: Precision@K, Recall@K, F1@K for $K \in \{5, 10, 20\}$
- Comparison: Global vs Segment-Specific models (6 total configurations)
- Statistical Testing: Paired t-tests on per-user F1 scores

3. Results

3.1 Model Performance

Model	Global F1@5	Segment F1@5	Improvement	Significance
Baseline	0.00985	0.01157	+17.5%	$p > 0.05$
Collaborative Filtering	0.00046	0.00146	**+213.8%**	$p < 0.05$
Hybrid (CF + CBF)	0.00157	0.00233	+48.4%	$p < 0.001$

Table 2: Model Performance Comparison (Test Set, $n = 2{,}000$)

3.2 Key Findings

Segment-Specific Models Consistently Win:

- **CF:** 9/9 metrics improved (100% success rate, statistically significant)
- **Hybrid:** 9/9 metrics improved (100% success rate, $p < 0.001$)
- **Baseline:** 7/9 metrics improved (78% success rate, not statistically significant)

Statistical Validation: Hybrid F1@20 shows +59.5% improvement with $p < 0.001$, confirming segmentation benefits are not due to random chance.

Absolute Performance Ranking: Baseline segment-specific (F1@20: 0.0127) outperforms CF (0.0019) and Hybrid (0.0026) in absolute terms—consistent with known patterns in grocery domains where popularity baseline captures strong repeat purchase behavior.

4. Business Impact & Deployment

4.1 Segment-Specific Strategies

- **Power Users (45%):** VIP loyalty programs, prioritize fresh food inventory, exclusive benefits
- **Bulk Shoppers (36%):** Large order discounts, weekly essentials reminders, combo deals
- **Routine Snackers (13%):** Subscription box opportunities, personalized snack recommendations
- **Household Essentials (4%):** Competitive pricing emphasis, cross-sell opportunities
- **Alcohol Enthusiasts (1%):** Curated alcohol selections, pairing suggestions

4.2 Deployment Recommendations

- Deploy **segment-specific Hybrid model** (best balance of performance and complexity)
- **A/B test** against popularity baseline in production
- Monitor segment distribution over time for model retraining
- Balance popularity with personalization to maintain diversity

5. Conclusion & Limitations

Core Finding: Customer segmentation measurably enhances recommendation quality across all model types. Segment-specific models consistently outperform global models (9/9 metrics for CF and Hybrid), validating the hypothesis that behavioral segmentation improves personalized e-commerce recommendations.

Limitations & Context:

- **Modest absolute scores** (F1@5: 0.002-0.011) due to large catalog (49,623 products), extreme sparsity (99.89%), and diverse preferences—typical challenges in grocery recommendation
- **Popularity dominance:** Simple baseline outperforms complex models in absolute terms due to strong repeat purchase patterns in grocery shopping
- **Test sample:** 2,000 users (computational constraints, stratified sampling ensures representativeness)
- **Fixed hybrid weights:** $\alpha = 0.5$ for all segments

Contribution: Demonstrated consistent 48-213% relative improvements through segmentation, validating segment-aware personalization despite modest absolute scores.

Future Work: Temporal features (time-of-day, seasonality), richer product features (brands, nutrition, price), per-segment weight optimization, sequential recommendation models.

Technical Skills: K-means clustering, PCA, SVD matrix factorization, statistical hypothesis testing (paired t-tests), temporal validation, Python (scikit-learn, scikit-surprise, Pandas, NumPy).

Appendix: Visualizations

All visualizations are available in the GitHub repository at `results/figures/`:
`https://github.com/wengchienwei/customer-segmentation-hybrid-recommender`

Phase 0: Data Analysis

- `product_analysis.png` - Product catalog distribution across departments and aisles
- `temporal_patterns.png` - Order patterns by day of week and hour of day

Phase 1: Customer Segmentation

- `phase1_clustering/cluster_size_distribution.png` - Distribution of customers across 5 segments
- `phase1_clustering/cluster_profiles_radar.png` - Multi-dimensional segment profiles
- `phase1_clustering/rfm_analysis.png` - RFM comparison across segments
- `phase1_clustering/behavioral_patterns.png` - Segment behavioral metrics

Phase 3: Evaluation Results

- `phase3_evaluation/global_vs_segment_comparison.png` - Performance comparison
- `phase3_evaluation/f1_improvement_by_k.png` - F1 improvements by K
- `phase3_evaluation/performance_by_segment.png` - Per-segment performance