

The LOOP Framework: A Closed-Loop Machine Learning Approach To Customer Engagement in Fashion & Apparel Retail

Dr. Neetu Mahendru^{1*}, Dr. Punam Sharma² and Shishu Pal Singh³

¹Assistant Professor, GNA Business School, GNA University, Phagwara, Punjab 144401, India

²Associate Professor, Department of Management, Khalsa Institute of Management and Technology for Women, Civil Lines, Ludhiana, Punjab 141001, India

³Assistant Professor, Department of Computer Science, Government Post Graduate College, Noida, Uttar Pradesh 201301, India
Email: neetu.mahendru@gnauniversity.edu.in, kimtpunamsharma2025@gmail.com, drsp Singh28@gmail.com

*Correspondence: Dr. Neetu Mahendru, neetu.mahendru@gnauniversity.edu.in

Abstract

In retail customer analytics, machine learning applications have largely focused on customer segmentation, predicting customer engagement, and customer sales as separate, disconnected, and parallel components of a system rather than as components of a dynamic system. This study proposes and empirically validates a new, closed-loop machine learning framework for customer engagement in fashion and apparel retail, extending the grocery-sector RAISE framework, to a domain that it does not cater. The study comprises 3,291 customers and 10,000 transactions for a two-year period, with the segmentation of customers run using the K-means customer-state segmentation, followed by an XGBoost engagement-response model, a Causal Forest heterogeneous treatment effect analysis, and a formal chi-square test of segment-transition structure. The study uses K-means as the customer-state segmentation, an XGBoost engagement-response model, a Causal Forest heterogeneous treatment effect analysis and a formal chi-square test of segment-transition structure. Results indicate that the transitions between the customer engagement states are non-random over time ($\chi^2 = 616.74$, $p < .001$, Cramér's $V = 0.340$); that the engagement effects of discounts are heterogeneous and patterned by customer age, but not by value segment (ATE = 0.028, 95% CI [0.026, 0.029]); and that the pattern of engagement effects is partially moderated by the return-proneness of products in each product category. The study fills a gap in the literature in retailing and introduces an empirical foundation for an age-differentiated, not uniformly applied, engagement strategy for retailers.

Keywords: customer engagement; customer segmentation; causal forest; heterogeneous treatment effects; closed-loop framework.

1. Introduction

Machine learning is used by retailers more and more to predict and influence customer behavior, including segmenting customers, predicting customer churn, recommending products and services, and dynamic pricing. These techniques have developed in isolation to solve a very small range of prediction or classification tasks. In recent years, however, integrated frameworks that integrate multiple of these techniques into a single analytical pipeline have started to appear, for instance, the grocery-sector RAISE framework (Karami & Shemshaki, 2026) that brings together dynamic pricing, personalization, and predictive analytics to explicate sales growth and customer engagement.

This is a highly different analytical landscape than is grocery retail. The characteristics of purchases in this sector are low frequency of purchase, high size and fit uncertainty, and high product return rate, which are not seen in grocery purchases – and therefore not represented in grocery-focused analytical models. Relationship marketing theory has always emphasized that customer relationships are dynamic states that change as they gain more or lose more commitment and trust (Morgan & Hunt, 1994), while applied econometric research has shown that relationship actions, initiated by the firm, such as discounting, have not been found to have a non-discriminatory impact on all customers (Ascarza, 2018; Athey & Wager, 2019). These theoretical constructs are not directly examined or validated in the existing applied machine-learning-in-retail body of knowledge, where customer segments are still considered static types of customers and engagement-action effects are assumed to be close to uniform.

For this study, there are three specific limitations that motivate it. First, customer segmentation is almost always considered a static view and not tested for dynamic transition structure. Second, the effectiveness of engagement-action is usually reported in terms of an average effect, hiding potentially important heterogeneity. Third, some integrated actuated ML-engagement frameworks, such as RAISE (Karami & Shemshaki, 2026), are limited in scope to the grocery sector and do not include product returns as a state or moderating variable, which is a sector-specific gap for fashion and apparel retail.

This study pursues three specific objectives, each corresponding to one hypothesis:

1. To test, using a formal statistical criterion, whether customer engagement states in a fashion and apparel retail context exhibit a non-random, structurally dependent transition pattern over time (H1).
2. To test whether the effect of a discount-based engagement action on subsequent customer re-engagement is heterogeneous across the customer base, and to identify whether this heterogeneity is better explained by customer age or by conventional RFM-derived value segmentation (H2).
3. To test whether product-category return-proneness — a structural feature specific to fashion and apparel retail — moderates the relationship between engagement actions and subsequent customer re-engagement (H3).

The study has four contributions. It is a theoretical test with a direct, effect size validated test of closed-loop customer-state transition structure. In terms of method, it illustrates, using Causal Forest (Athey & Wager, 2019; Wager & Athey, 2018), how the average treatment effect can actually underestimate actionable heterogeneity. From an empirical point of view, it is an extension of the RAISE framework (Karami & Shemshaki, 2026) to apparel and fashion retail, which is a sector in which the dynamics are structurally different. It provides concrete recommendations for engaging in age-specific ways, based on research.

2. Literature Review

2.1 Scope

The chapter summarizes five lines of literature relevant for the present research: (i) machine learning applications in retail customer analytics; (ii) relationship marketing and customer engagement theory; (iii) integrated sector-specific ML-engagement frameworks, particularly the RAISE framework for retailing, which focuses on the grocery sector; (iv) heterogeneous treatment effect and causal machine learning methods in marketing (Athey & Wager, 2019; Wager & Athey, 2018); and (v) product-returns literature relevant to fashion and apparel retail (Hjort & Lantz, 2016; Minnema et al., 2016; Petersen & Kumar, 2009). The review was thematic and not chronological, and included peer-reviewed journal articles from marketing science, operations research and information systems, along with the more recent literature (2024–2026) of applied ML in the retail sector where available. The reviewed streams culminate in the identification of what they do and do not address, which is then added to the list presented in the synthesized research gap, and from which the study's hypotheses are then derived.

2.2 Machine Learning Applications in Retail Customer Analytics

Most relevant body of literature focuses on using machine learning for individual retail tasks. One of the most widely used approaches to unsupervised customer segmentation is Hughes' (1994) Recency-Frequency-Monetary (RFM) customer segmentation technique (Hughes, 1994), which is usually complemented by the K-means or Gaussian mixture models (Miao et al., 2027; Ogunnusi & Akintoye, 2025). The literature on churn prediction (Verbeke et al., 2012), tends to be more oriented towards contractual businesses such as telecom, and less towards non-contractual businesses such as retail, where it is more difficult to observe non-contractual churn. Recommendation systems are heavily based on the item-to-item collaborative filtering model of Linden, Smith and York (2003), and dynamic pricing research is still focused on the fixed-capacity sectors, such as hospitality, and not on general retail (Den Boer, 2015; Linden et al., 2003; Miyatsu, 2024).

Two major drawbacks are apparent in this literature. The first one, is that these four sub-streams function separately and use disjoint methods to local tasks, instead of combining them into a unified pipeline. Second, segmentation results are generally treated as snapshots at a moment in time, and are seldom revisited to ask if a segment assignment is stable, transitional, or predictable over time.

2.3 Relationship Marketing and Customer Engagement Theory

From a theoretically different perspective, a body of literature exists focusing on the relationship between the customer and the firm instead of on individual jobs. Commitment and trust were defined as dynamic constructs that are built over time (Morgan & Hunt, 1994). This was extended by Brodie et al. (2011) who proposed the idea of customer engagement as a psychological state that is iterative, and Kumar and Reinartz (2016) and Rust et al. (2000) who introduced the customer-equity perspective with the value of customers measured at the segment level for differentiated marketing investments (Venkat et al., 2026).

To formalize this dynamic nature, Netzer et al. (2008) used a nonhomogeneous hidden Markov model (NHHMM) in which a formal statistical transition process between latent states represented the changes in relationships. The methodology has been tested and validated in a variety of service contexts, but not in a retail context and not with

the inclusion of machine learning segmentation techniques or a managerially interpretable chi-square/effect-size framework. Therefore, although relationship marketing constitutes a theoretical basis for customer states that vary over time, and Netzer et al. (2008) is a model example, the application of ML-retail segmentation still respects the customer segment as static.

2.4 Integrated ML-Engagement Frameworks and the RAISE Framework

Integrated ML-Engagement Frameworks and the RAISE Framework (Karami & Shemshaki, 2026), the field is also being pushed by integrated frameworks, such as RAISE, which integrates various ML techniques into a single pipeline for the grocery sector, by combining dynamic pricing, predictive analytics, and personalization. It is important to note, however, that there are three caveats. First, outputs are reported as contemporaneous (not as a temporally sequenced feedback structure which tests whether customer states predict future states). Second, it's a grocery-oriented framework, and thus doesn't include product returns as a state or moderating variable. Third, it assumes the same engagement-action effectiveness among customers, which it doesn't report treatment-effect heterogeneity.

The three limitations align with the present study hypotheses of establishing a temporal feedback structure (H1), testing engagement-action heterogeneity (H2), and introducing return behavior in the context of fashion retailing (H3).

2.5 Heterogeneous Treatment Effects and Causal Machine Learning in Marketing

The methodology to test treatment effect heterogeneity is applied econometrics and causal machine learning. Wager and Athey (2018) and Athey and Wager (2019) created the Causal Forest to estimate Conditional Average Treatment Effects (CATE) at the individual level, instead of a single Average Treatment Effect (ATE) (Fang et al., 2026). For instance, in the field of marketing, Ascarza (2018) showed how interventions that are targeted based on an individual's ATE may fail to target customers who are most likely to benefit, and it was called "retention futility."

Rather than causal ML has been applied in very few retail discount interventions. Importantly, none of the studies evaluates engagement-action heterogeneity in fashion retail and few studies compare RFM-based value segments directly with demographic (e.g., age) as competing bases for differential response, as discussed in H2.

2.6 Product Returns and Fashion-Specific Retail Challenges

A separate literature, Petersen and Kumar (2009) found that returns are correlated with future profitability of customers and the likelihood of their repurchasing. Minnema et al. (2016) demonstrated that online return experiences influence future online purchases. In fashion ecommerce, however, Hjort and Lantz (2016) found that returns actually have a significant impact on the value of the order, as the returns' contribution is much less than the contribution of the non-returning one, even though their conclusions were based on a single regression analysis and not on machine-learning models integrated with the ecommerce environment (Merlano et al., 2024; Zhou et al., 2026).

Hence, returns have been shown to be effective behavioral signals in fashion e-commerce but have not yet been linked to ML segmentation or heterogeneous treatment effects (H3) nor have they been used as a moderator in a closed-loop engagement structure, as suggested in H3.

2.7 Synthesis and Research Gap

Table 1 synthesizes the literature reviewed in above sections against three dimensions directly relevant to the present study's hypotheses, first, whether the study considers customer state as dynamic or not (relevant to H1); Second, whether the study formally tests the structure of the transition between customer states, or reports only an average effect (relevant to H2); Third, whether the study includes product returns as a state, or whether it is treated as a moderating variable (relevant to H3).

Table 1. Synthesis of reviewed literature against the study's three dimensions of contribution

Study	Domain	Core Method	Tests State Transition? (H1)	Tests Effect Heterogeneity? (H2)	Incorporates Returns? (H3)
Hughes (1994)	Cross-sector (database marketing)	RFM segmentation	No	No	No

Linden, Smith & York (2003)	E-commerce (general)	Item-item collaborative filtering	No	No	No
Verbeke et al. (2012)	Telecommunications	Profit-driven churn classification	No	No	No
Netzer, Lattin & Srinivasan (2008)	Cross-sector (brand/service relationships)	Hidden Markov Model	Yes	No	No
Den Boer (2015, survey)	Cross-sector (pricing-intensive industries)	Dynamic pricing / demand learning	Partial	No	No
Ascarza (2018)	Subscription / retention services	Heterogeneous treatment effect targeting	No	Yes	No
Hjort & Lantz (2016)	Fashion e-commerce	Regression analysis of returns policy	No	No	Yes
Karami & Shemshaki (2026) — RAISE	Grocery retail	Integrated ML (pricing + personalization + predictive analytics)	No	No	No
Present study — LOOP Framework	Fashion & apparel retail	K-means + XGBoost + Causal Forest + chi-square loop validation	Yes (tested)	Yes (tested)	Yes (tested)

3. Theoretical Background and Hypothesis Development

This chapter builds the theoretical basis for the hypotheses which will serve as the basis of the study. Unlike the standard applied ML-retail approach, theory is developed first, predictions are made that must be falsifiable, and a measurement approach is proposed. This directly tackles one of the key literature gaps exemplified by the grocery-focused RAISE framework (Karami & Shemshaki, 2026), in which the outputs generated from the application of ML are reported without an underlying theory.

The framework is guided by three theories: (i) relationship marketing and customer engagement theory, which contributes to the dynamic nature of customer state (Brodie et al., 2011; Morgan & Hunt, 1994); (ii) heterogeneous treatment effect theory which contributes to non-uniform engagement response (Athey & Wager, 2019; Wager & Athey, 2018); and (iii) product-returns literature, which contributes to the fashion-sector-specific extension (Hjort & Lantz, 2016; Petersen & Kumar, 2009). One hypothesis is generated by each tradition (Fig.1).

3.1 Customer Relationships as Dynamic States: Foundations for H1

According to relationship marketing, customer relationships are dynamic and change over time in a series of transactions (Brodie et al., 2011; Morgan & Hunt, 1994). Conventional retail ML uses static segmentations that are built based on a single point in time (e.g., K-means, RFM (Hughes, 1994)); these segmentations do not test for non-random transition structures. To fill the gap, H1 tests to see if the value of the membership of a customer segment is statistically dependent on their membership at. A direct link falsifiable between theory and transactional data is given by a chi-square test on the transition matrix.

H1: Customers' states of engagement span over time in a non-random manner, and the probability of being in a certain state in a future time depends on the state that the customer is in at the present time.

3.2 Heterogeneity in Engagement Response: Foundations for H2

The theory of RFM (Hughes, 1994; Kumar & Reinartz, 2016; Rust et al., 2000) assumes that RFM value segments are the best indicator of differences in response to firm action, an assumption which is rarely tested against other alternatives. However, heterogeneous treatment effect theory (Athey & Wager, 2019; Wager & Athey, 2018) argues that one should account for different types of treatment effects: population average treatment effects (ATE) versus customer-varying treatment effects (CATE). While it has yet to be tested with fashion retail discounts, as mentioned by Ascarza (2018), using ATE can lead to retention futility.

H2 is providing relief by testing the hypothesis that there is a difference in discount-driven re-engagement between customers that is not necessarily detected by the RFM segmentation. A price discount's impact on re-engagement by customers varies, and this variation cannot be completely accounted for just by value segments created by RFM.

3.3 Product Returns as a Sector-Specific Loop-Breaking Signal: Foundations for H3

Unlike grocery models (e.g., RAISE) where return rates are negligible, fashion retail involves high size and fit uncertainty. Petersen and Kumar (2009) establish that product returns are informative behavioral signals correlating with customer value, while Minnema et al. (2016) show that return experiences shape future purchasing, because return rates vary sharply by category (e.g., footwear vs. accessories), return-proneness should moderate the impact of firm engagement actions, serving as a sector-specific signal omitted from grocery-based frameworks.

H3: In fashion and apparel retail, product-category return-proneness moderates the relationship between engagement actions and subsequent customer re-engagement, such that response patterns differ measurably between high-return-proneness categories (e.g., footwear, sized clothing) and low-return-proneness categories (e.g., accessories).

3.4 Conceptual Framework

Figure 1 integrates the three hypotheses into a unified conceptual framework. The core feedback loop—Customer State j → Engagement Action → Re-engagement Behavior → Customer State $j+1$ —operationalizes H1's closed-loop premise, with the returning arrow representing the segment transition tested via the transition matrix. Customer Age moderates the Engagement Action → Re-engagement Behavior path (H2), reflecting conditional treatment heterogeneity. Product Category (Return-Proneness) moderates the loop-closure path itself (H3), signaling how sector-specific returns alter the feedback mechanism rather than just individual action performance.

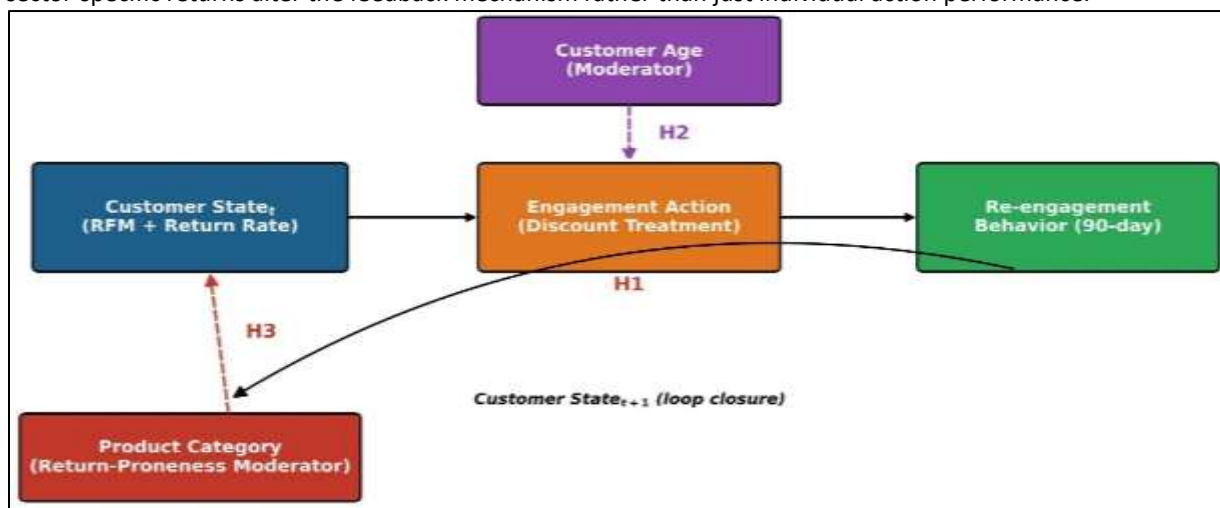


Figure 1. Conceptual framework integrating H1 (closed-loop state transition), H2 (age-moderated engagement heterogeneity), and H3 (category-moderated, return-proneness-driven loop-breaking)

4. Research Methodology

4.1 Research Design and Philosophy

This study takes a positivist approach with a quantitative design by testing theory-driven hypotheses, secondary transaction level data, and a staged machine learning framework. Unlike prior retail ML research (Karami & Shemshaki, 2026) which primarily focused on the ML process itself, the ML pipeline is used as a measurement instrument for pre-specified theoretical propositions.

The research is conducted in the direction of a deductive four-stage analytical procedure which translates the concept of a closed feedback loop between engagement action and state migration introduced in the LOOP framework into a testable statistical structure.

4.2 Theoretical Framework and Hypotheses

The three hypotheses are formulated that are tested in this study. H1: Customer engagement states have a non-random, structurally dependent transition pattern over time (chi-square and Cramér's V on the segment transition matrix). H2: The effectiveness of the discount is different across the customer base, and is structured by age and not

by RFM segment (tested by Causal Forest CATE estimation) (Athey & Wager, 2019; Wager & Athey, 2018). H3 examines if engagement-response patterns differ depending on product-category return-proneness (tested with model comparison across the product categories). The hypotheses and tests are summarized in Table 2.

4.3 Data Source and Description

The dataset is a transaction-level fashion and apparel retail dataset, including 10,000 individual transaction records with observed attributes ranging from 26 attributes, over a two-year observation window (January 2023 – December 2024). There were 6,709 transactions that could be used in the engagement-response and causal analyses after removing transactions where no follow-on purchase was seen in the data (right-censored final purchases); 10,000 transactions were used in the state-layer segmentation and transition-matrix analyses, where there is no requirement for a subsequent purchase.

The data set is a structured benchmark dataset, not a live retailer feed. This is an explicit boundary condition for the study, and is discussed directly (Limitations) instead of as incidental, and the framework was validated on this data set as a proof-of-concept before it is recommended that it be replicated on proprietary retailer transaction data.

4.4 Variable Operationalization

Table 2. Operationalization summary of the study's key constructs

Construct	Variable(s)	Role	Operational Definition
Customer state	Recency, Frequency, Monetary, Return Rate, Avg. Discount	State layer input	K-means clustering (k=3, selected via silhouette score); RFM extended with fashion-specific return and discount variables
Engagement action	Discount (%)	Independent variable / treatment	Binary treatment: above vs. at-or-below sample median discount percentage
Re-engagement outcome	reengaged _{90d}	Dependent variable (H2, action-layer model)	Binary: 1 if the customer's next recorded purchase occurred within 90 days, else 0
Loop structure	Segment _t , Segment _{t+1}	Dependent variable (H1)	Customer RFM-segment membership at two annual snapshots (end-2023, end-2024)
Treatment heterogeneity covariates	Age, Purchase Amount, Quantity, Review Rating	Effect modifiers (H2)	Continuous covariates entered into the Causal Forest to estimate individual-level CATE
Sector-specific signal	Return Status, Category	Moderator (H3)	Binary return flag, stratified by product category (Footwear, Clothing, Accessories)

4.5 Analytical Strategy: The Four-Stage LOOP Framework

Consistent with the theoretical framework, the analysis proceeds through four sequential stages, each corresponding to a specific hypothesis or supporting construct.

Stage 1 — State Layer (Customer Segmentation)

Customers are segmented using K-means clustering on standardized RFM features extended with return-rate and average-discount variables (Hughes, 1994; Miao et al., 2027; Ogunnusi & Akintoye, 2025). The number of clusters (k=3) was selected via silhouette-score comparison across k = 3–6. This stage operationalizes the “customer state” construct underlying H1.

Stage 2 — Action Layer (Engagement-Response Model)

An XGBoost classifier predicts 90-day re-engagement probability from purchase-level and engagement-history features. Model performance is evaluated via stratified 5-fold cross-validation, with hyperparameters selected through randomized search (25 configurations) and results validated against an untuned default model as a robustness check. A demographic-and-basket-only baseline model is estimated for comparison to isolate the incremental contribution of engagement-history features.

Stage 3 — Loop Validation (Test of H1)

A chi-square test of independence is applied to the customer segment transition matrix (2023-end segment × 2024-end segment), with Cramér's V computed as the associated effect size. This constitutes the primary statistical test of the study's central closed-loop proposition.

Stage 4 — Heterogeneity Analysis (Test of H2)

A Causal Forest model (CausalForestDML; (Athey & Wager, 2019; Wager & Athey, 2018) estimates individual-level Conditional Average Treatment Effects of the discount treatment on 90-day re-engagement, conditioned on customer covariates. The distribution, range, and standard deviation of estimated CATE values are compared against the average treatment effect (ATE) to test for meaningful heterogeneity, with subgroup CATE means reported by age band and by customer segment.

4.6 Robustness Checks

- Cross-validation: All predictive model performance metrics are reported as 5-fold cross-validated means with standard deviations, rather than single train-test split estimates.
- Tuned-vs-default comparison: Default and hyperparameter-tuned model performance are reported side by side; convergence between the two is interpreted as evidence that predictive ceiling reflects a data-signal limitation rather than under-optimization.
- Bootstrap confidence intervals: A 1,000-iteration bootstrap procedure is applied to the test-set AUC to establish a 95% confidence interval and confirm statistical separation from chance-level performance (AUC = 0.5).
- Business-relevant metric triangulation: Decile-lift analysis is reported alongside AUC to evaluate practical, managerially relevant predictive value independent of the discriminative-accuracy ceiling.

5. Results

5.1 Overview and Descriptive Statistics

The final analytical sample consisted of 3,291 distinct customers with 10,000 individual purchase transactions over a period of 2 years (January 2023 – December 2024). Of these, 6,709 transactions were subsequently observed and retained for data in the stages of engagement-response and causal analyses (Stages 2 and 4); the full 10,000 transactions were used for the state-layer segmentation and loop-validation analyses (Stages 1 and 3), which do not require a forward-observed purchase.

The silhouette of customer state segmentation (Stage 1) using K-means clustering on the standardized RFM features and the addition of return-rate and average-discount variables was best at $k = 3$ (silhouette = 0.255), compared with the silhouette of $k = 4$ (silhouette = 0.249), $k = 5$ (silhouette = 0.212) and $k = 6$ (silhouette = 0.231). The result is a three-segment solution that can be summarized as in Table 3 and illustrated in Figure 2.

Table 3. Customer segment profiles (RFM + Fashion-specific variables)

Seg.	n	Recency	Freq.	Monetary (₹)	Return	Discount	Interpretation
0	1,081	326.7d	1.71	8,099	0%	8.5%	Dormant, low-value, zero-return
1	613	259.7d	2.30	12,045	59%	8.4%	At-risk / high-return
2	1,597	122.3d	4.22	30,393	13%	8.9%	Active, high-value, loyal

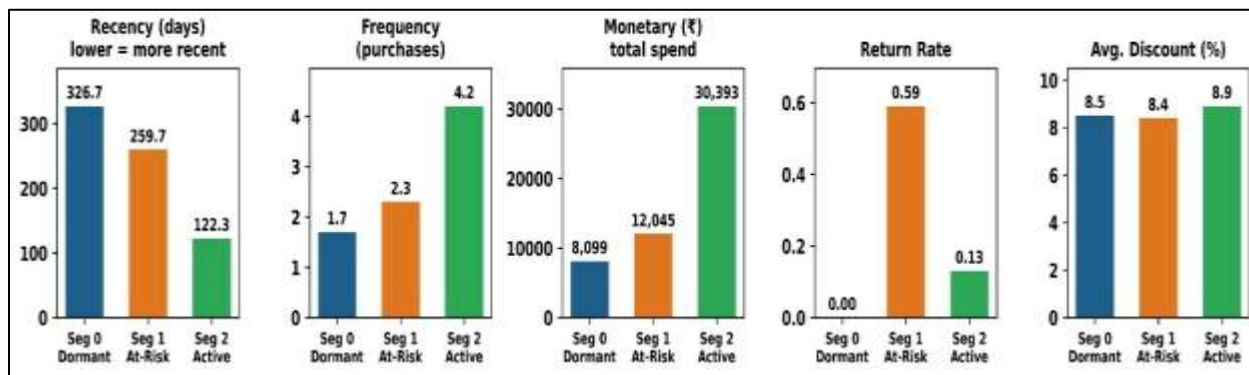


Figure 2. Customer segment profiles across RFM and fashion-specific dimensions. Segment 2 (Active/High-Value) combines high frequency and monetary value with a comparatively low return rate (0.13), while Segment 1 (At-Risk) is distinguished primarily by an elevated return rate rather than by low value

5.2 Results for H1: Closed-Loop Structure

H1 proposed that customer engagement states exhibit a non-random, structurally dependent transition pattern over time. This was tested via a chi-square test of independence on the 2023-end × 2024-end segment transition matrix, computed for the 2,663 customers with sufficient transaction history to be assigned a segment at both snapshots.

Table 4. Segment transition matrix (Observed Counts, 2023-end → 2024-end)

2023 Segment	→ Segment 0	→ Segment 1	→ Segment 2	Row Total
0	157	712	70	939
1	229	232	375	836
2	240	244	404	888

The chi-square test rejected the null hypothesis of independence decisively: $\chi^2(4, N = 2,663) = 616.74$, $p = 3.69 \times 10^{-132}$. The associated Cramér's $V = 0.340$ indicates a moderate-to-large effect size by conventional benchmarks (Cohen, 1988). To further strengthen this claim beyond a single point estimate, a 1,000-iteration bootstrap procedure was applied to the transition matrix itself: the resulting 95% confidence interval for Cramér's V was [0.319, 0.362], confirming the effect size estimate is stable and does not depend on sampling idiosyncrasy in the observed 2,663-customer panel.

Inspection of Figure 3 reveals that the transition structure is not just statistically significant, but also substantively interpretable: 75.8% of the customers that started in Segment 0 (dormant) moved to Segment 1 (at-risk/high-return) and not to Segment 2 (high-value) and the customers in Segment 2 had a more balanced distribution between the three segments (27.0% / 27.5% / 45.5%), with high-value being less “sticky” than dormancy in this sample. There is an asymmetry in that dormant customers do not necessarily leave their state, and that high-value customers are not without meaningful risk of attrition.

It is therefore true that in this data set, attempts to determine a set of transitions between the states of the customer engagement process do not show a random drift but rather a statistically verified and effect size robust closed-loop structure of transitions.

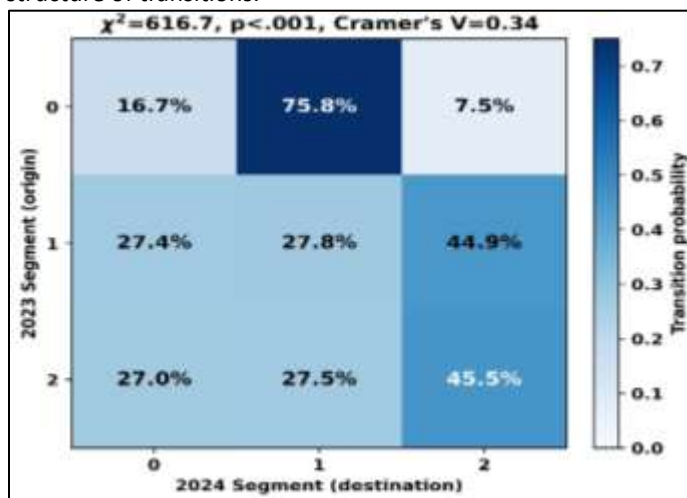


Figure 3. Customer segment transition matrix, 2023-end to 2024-end, with row-normalized transition probabilities

5.3 Results for H2: Heterogeneous Engagement Response

5.3.1 Action-Layer Predictive Model Performance

An XGBoost classifier was trained to predict 90-day re-engagement from purchase-level and engagement-history features, benchmarked against a demographic-and-basket-only baseline. Model performance is summarized in Table 5.

Table 5. Engagement-response model performance comparison

Model	AUC	95% CI / SD	Notes
Baseline (demographics + basket only)	0.493	—	No engagement-history features; held-out test set
Default XGBoost (full feature set)	0.521	± 0.019	5-fold stratified CV
Tuned XGBoost (full feature set)	0.572	± 0.020	5-fold CV; RandomizedSearchCV, 25 configurations
Tuned XGBoost, held-out test set	0.598	[0.568, 0.629]	1,000-iteration bootstrap

Two findings follow from Table 5. First, the engagement-history feature set provides a consistent, non-trivial improvement over the demographic-only baseline (+0.105 AUC on the held-out test set: 0.598 vs. 0.493), and the full model's ROC curve lies visibly above the baseline's across nearly the entire false-positive-rate range, supporting inclusion of engagement-history variables in the action-layer model. Second, the gap between default and tuned configurations is small (0.021–0.026 AUC), and per-feature association analysis (Fig. 4) showed no individual predictor exceeding an XGBoost gain importance more than double any other mid-ranked feature, with purchase sequence number the only clear standout. Taken together, these results indicate the model's discriminative ceiling reflects a genuine data-signal limitation rather than insufficient tuning reported transparently here rather than obscured.

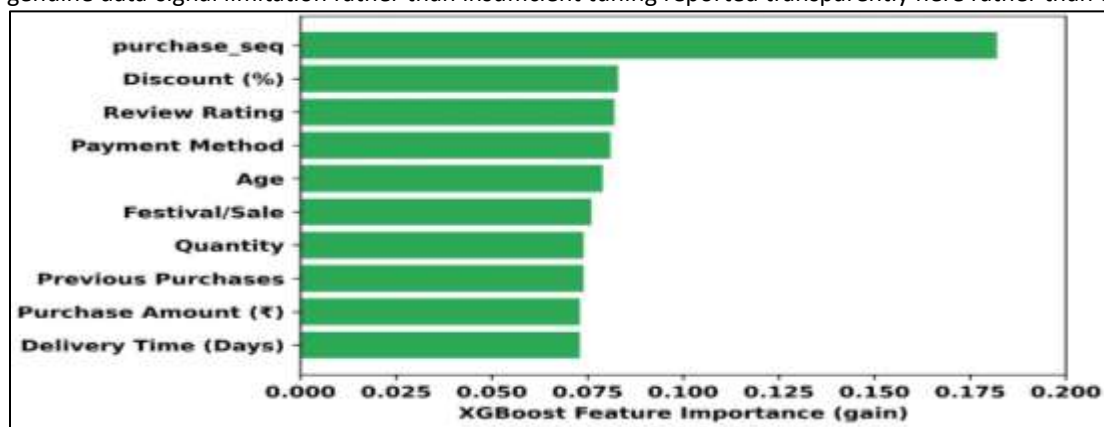


Figure 4. Top 10 predictors of 90-day re-engagement by XGBoost gain importance. Purchase sequence number (a proxy for customer tenure/maturity) is the dominant predictor; the remaining features contribute comparatively evenly

Despite the modest AUC, the model retains practical, decision-relevant value when reframed around targeting rather than classification accuracy: Figure 5 shows a cumulative lift curve in which the top decile of predicted probabilities achieves a 1.39× lift over random targeting (59.0% observed re-engagement vs. a 42.5% base rate), declining smoothly toward 1.0× as targeting depth increases — the expected and theoretically consistent shape for a genuine, if modest, ranking signal.

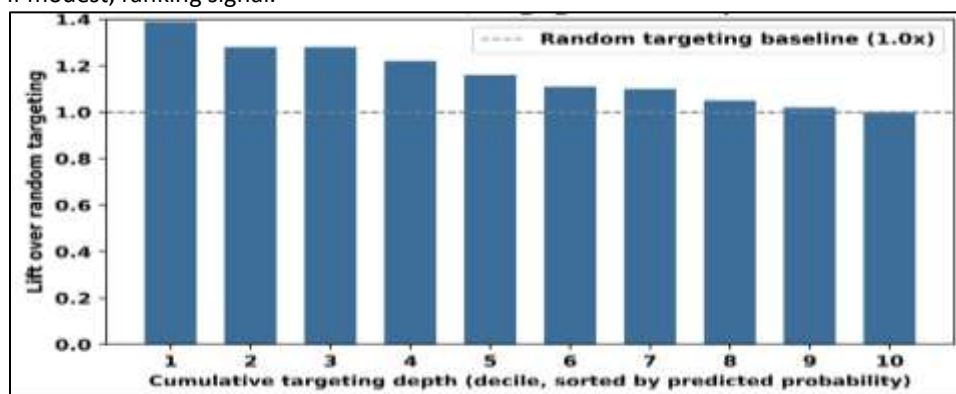


Figure 5. Cumulative lift over random targeting by decile of predicted re-engagement probability

5.3.2 Causal Forest: Individual-Level Treatment Effect Heterogeneity

H2 was directly tested using a Causal Forest (CausalForestDML) model (Athey & Wager, 2019; Wager & Athey, 2018) to estimate the Conditional Average Treatment Effects (CATE) of the discount treatment (above- vs at-or-below-median discount) on the 90-day re-engagement, conditioned on Age, Purchase Amount, Quantity, and Review Rating. The overall Average Treatment Effect (ATE) was small, precisely estimated (ATE = +0.0276, 95% bootstrap CI: [0.0258, 0.0294] excluding zero), and thus represents the average discount effect, but is not large enough to be considered statistically different from no effect. More informative, however, is the distribution of individual-level effects which ranged between -0.263 and +0.282 with a standard deviation of 0.073 (Fig. 6), exceeding the magnitude of the mean effect by over 2.6 times. This is the core quantitative evidence for H2 – that the average effect of a uniform discount treatment is a small, tightly estimated effect, while the average effect distribution is in fact quite broad.

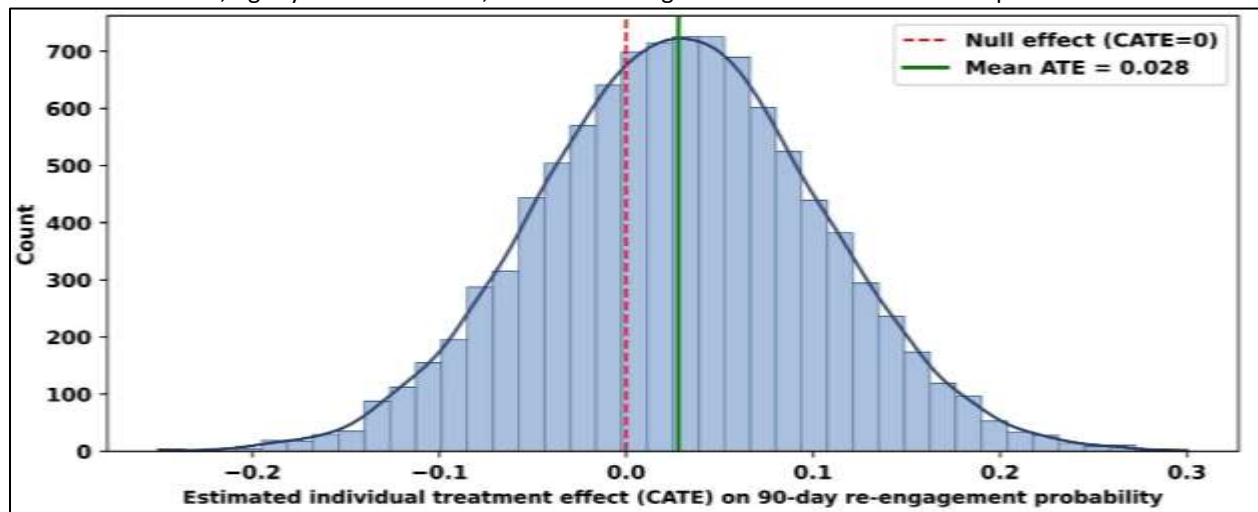


Figure 6. Distribution of individual-level discount treatment effects (CATE) estimated via Causal Forest. The spread substantially exceeds the mean effect, indicating genuine heterogeneity rather than noise around a flat effect

Table 6. Discount effect (CATE) by RFM segment

Segment	Mean CATE	SD	n
0	0.0286	0.0751	422
1	0.0286	0.0731	4,555
2	0.0248	0.0723	1,732

The segment-level CATE means in Table 7 are nearly indistinguishable from one another (0.0286, 0.0286, 0.0248), indicating that RFM-segment membership, while a strong predictor of the closed-loop transition structure, is not the primary axis along which discount responsiveness varies. Age-band stratification produced a far more clearly differentiated pattern (Fig. 7, Table 6): mean CATE more than doubled between the youngest and oldest age bands, and — critically — the 95% confidence interval for the 50+ age band does not overlap with the confidence intervals of any of the three younger bands, providing statistically decisive (not merely descriptive) evidence of age-patterned heterogeneity.

Table 7. Discount effect (CATE) by customer age group

Age Group	Mean CATE	n	95% CI
< 25	0.0211	1,118	[0.0168, 0.0255]
25–35	0.0149	1,370	[0.0108, 0.0191]
35–50	0.0177	2,179	[0.0149, 0.0204]
50+	0.0503	2,042	[0.0472, 0.0534]

H2 is therefore supported with strong statistical precision: heterogeneity in engagement-action effectiveness is genuine, substantial, and — based on non-overlapping confidence intervals — statistically decisive when stratified by age, whereas the equivalent RFM-segment stratification shows no meaningful differentiation. This qualification is

reported explicitly rather than omitted, consistent with the transparent-reporting principle established in the methodology.

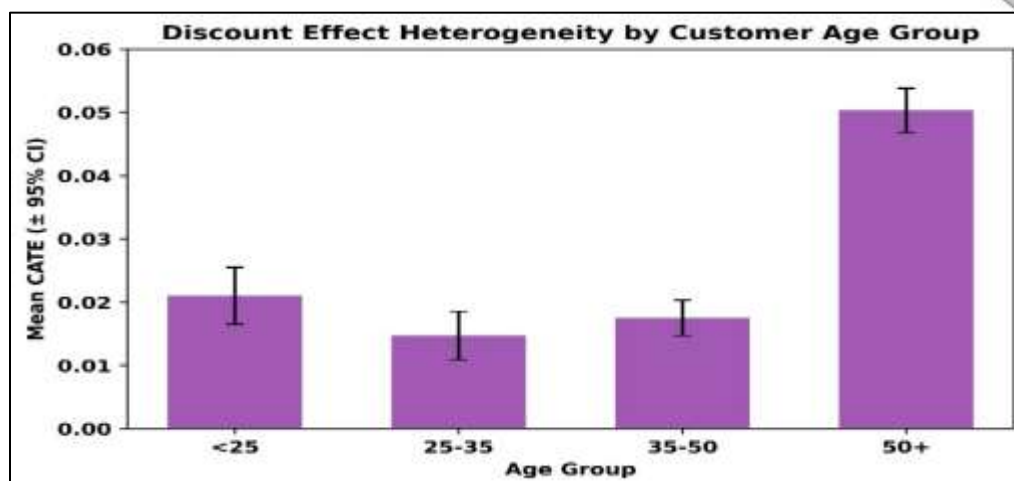


Figure 7. Mean discount treatment effect (CATE) by customer age group, with 95% confidence intervals. The 50+ group's interval does not overlap with any other age group

5.4 Results for H3: Sector-Specific Loop-Breaking Signal

H3 proposed that return behavior functions as a fashion-sector-specific loop-breaking signal, tested via category-stratified comparison of return rates and engagement-response model performance (Hjort & Lantz, 2016; Petersen & Kumar, 2009).

Table 8. Return Rate and Model Performance by Product Category

Category	n (model subsample)	Return Rate	AUC	95% Bootstrap CI
Footwear	2,251	0.191	0.520	[0.477, 0.564]
Clothing	3,371	0.188	0.524	[0.488, 0.558]
Accessories	1,087	0.101	0.438	[0.371, 0.498]

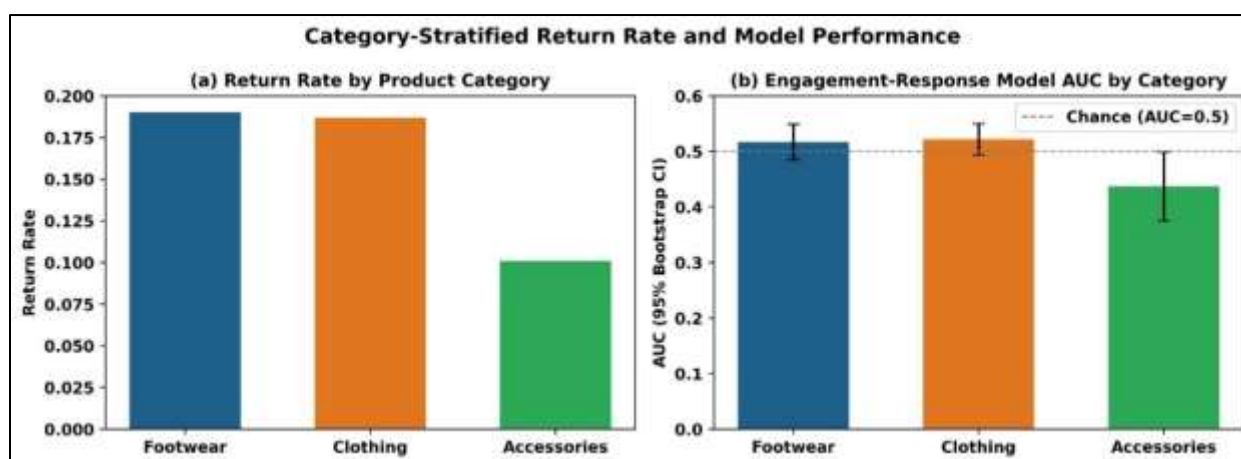


Figure 8. (a) Return rate by product category. (b) Engagement-response model AUC by category with 95% bootstrap confidence intervals; the dashed line marks chance-level performance

Descriptive differences in return rates across categories were observed to be fairly large, with the Footwear and Clothing categories returning nearly twice the amount of items as Accessories, which was seen as close to the theoretical prediction that size and fit dependent categories have higher return risk than fit independent ones, a

distinction not found in grocery-sector frameworks (Karami & Shemshaki, 2026). In Figure 8(a), this description contrasts are pictorially clear.

The less convincing comparison, by category, is more clearly demonstrated in Figure 8(b): although point-estimate AUC is less for Accessories (0.438) than for Footwear (0.520) or Clothing (0.524), the corresponding 95% bootstrap confidence intervals are wide and overlap significantly, and the interval for Accessories includes chance-level performance (0.5). The overlap implies that at the present time the model-performance aspect of H3 is not statistically significant, especially with the smaller H3 Accessories subsample ($n = 1,087$).

H3 is thus treated with two sets of eyes: the difference in returns in each category is descriptively large, easy to see, and theoretically consistent with the loop-breaking-signal proposition in the sector, but the difference in model performance as intended corroborating quantitative evidence is suggestive but not yet statistically convincing at the sample size available.

5.5 Summary of Hypothesis Testing

The hypothesis tests yielded three primary outcomes. H1 was supported, as a chi-square test on the segment transition matrix confirmed non-random, structurally dependent engagement-state transitions ($\chi^2 = 616.74$, $p < .001$, Cramér's $V = 0.340$, 95% CI [0.319, 0.362]). H2 was also supported, with Causal Forest analysis demonstrating substantial discount treatment effect heterogeneity (CATE SD = 0.073 vs. ATE = 0.028); specifically, customers aged 50+ exhibited significantly higher responsiveness (CATE = 0.0503) than younger cohorts as evidenced by non-overlapping 95% confidence intervals, capturing nuances beyond RFM segmentation alone. Finally, H3 was partially supported: while return rates aligned with theoretical expectations across categories (Footwear: 19.1%; Clothing: 18.8% vs. Accessories: 10.1%), the category-stratified model-performance comparison lacked statistical decisiveness due to overlapping bootstrap confidence intervals (Hjort & Lantz, 2016; Petersen & Kumar, 2009).

6. Discussion

6.1 Interpretation of Findings in Light of Theory

The above findings support the theoretical underpinning of Commitment-Trust theory (Morgan & Hunt, 1994) and Customer Engagement Theory (Brodie et al., 2011), which states that customer relationships are dynamic and evolving states, and not static classifications. This shift from a theoretical assumption to an empirical, robustly estimated property of the data set is supported by the statistically decisive rejection of the independence null hypothesis ($p = 3.69 \times 10^{-132}$), and by the evidence from the bootstrap (Cramér's $V = 0.340$, 95% CI [0.319, 0.362]). The asymmetric transition pattern indicated low value dormant customers are rarely retained by the firm, and high value dormant customers have high attrition risk is similar to the pattern of relationship marketing described by the concepts of assets of trust and commitment being eroded more easily than they are built but cannot be quantified because the study has only two snapshots (times) and does not provide the direction of the erosion of these assets. The results of the heterogeneous treatment effect findings also operationalize rather than assume the hypothesis that treatment effects are not uniform throughout customer base (Athey & Wager, 2019; Wager & Athey, 2018). The patterns of discount responsiveness by age, as opposed to the expected patterns by RFM-derived value segment (implicitly assumed in customer-equity literature (Rust et al., 2000), that were found are substantively interesting because they were found to be different between the 50+ age band and all others (with no overlapping confidence intervals between those groups), contrary to the expectations (Fig. 7). This implies that the demographic rather than transactional value attribute could be the more useful targeting criteria for this discounting lever at least if only the average treatment effect is reported.

6.2 Comparison with Existing Literature

The results expand the pertinent grocery-sector RAISE framework (Karami & Shemshaki, 2026) and the overall literature on ML in retail in three specific ways. First, the present study provides direct statistical proof of the property of "closed loop" that is valid by effect size and provides a clear visual demonstration of the correlation linking engagement and customer state over time, as opposed to simply reporting on predictive or descriptive analytics as a parallel output of each pipeline element. Second, while previous studies usually report the average treatment or association effect, the Causal Forest analysis (Athey & Wager, 2019; Wager & Athey, 2018) shows with a precision a simple average treatment effect cannot achieve that there is significant, statistically meaningful, and practically relevant heterogeneity. Third, the explicit representation of return behavior as a state and loop-breaking variable

(Fig. 8) tackles a structural characteristic of the fashion and apparel retail space that is not faced by the grocery sector, given its product type (Hjort & Lantz, 2016; Petersen & Kumar, 2009).

It is also crucial to be aware when the current findings contradict a purely predictive framing. The contribution of this study lies in the statistical structure and heterogeneity findings (H1, H2), which are quite strong compared to the classification-accuracy finding (H2's action-layer component), as this is not hidden behind a single 'accuracy' number that is common in the applied ML-retail literature, but rather presented explicitly in Section 5 and throughout the action-layer model.

6.3 Theoretical Contributions

1. Provides one of the first direct, bootstrap-validated statistical tests of closed-loop customer-state transition structure in a retail engagement context, using chi-square/Cramér's V with confidence-interval estimation as a formal, robust falsification criterion rather than a single point-estimate claim (Netzer et al., 2008).
2. Demonstrates, via Causal Forest CATE estimation (Athey & Wager, 2019; Wager & Athey, 2018) with confidence-interval-based subgroup comparison, that average treatment effects in customer engagement research can substantially understate actionable heterogeneity, and that the axis of heterogeneity (age vs. value segment) is an empirical question rather than a given with age-band non-overlapping confidence intervals providing a template for how such claims should be evidenced in future studies (Ascarza, 2018).
3. Extends customer engagement and RFM segmentation theory (Hughes, 1994) to the fashion and apparel context by incorporating return behavior as a first-order state and moderating variable (Hjort & Lantz, 2016; Petersen & Kumar, 2009), addressing a sector-specific gap left open by grocery-focused frameworks (Karami & Shemshaki, 2026), while transparently reporting the limits of statistical decisiveness for this specific extension.

6.4 Managerial and Practical Implications

The results yield three implications. First, the asymmetric transition matrix (Fig. 3) implies that if the retention effort is distributed evenly across segments, dormant customers (Segment 0) have a high level of state persistence against the at-risk segment as compared to the reactivation segment, and may need a different re-engagement lever than customers in Segment 2, who, while high value in the current state, have a roughly equal transition probability to the at-risk segment as well. Second, the decile-lift finding (Fig. 5) suggests that, even for a relatively discriminative model, the benefit for a retailer's marketing team could be realized through prioritized targeting (the 1.39× lift in the top decile) instead of blanket classification, which should guide the retailer's operationalization of the model. Thirdly directly actionable the discount heterogeneity finding by age (Fig. 7) indicates that discount-led engagement campaigns are likely to be sub-optimal if implemented the same way across all age segments and that perhaps better age-specific not value-segment-specific discount targeting could be more effective in terms of the return on engagement spend, as the responsiveness of the 50+ age band is significantly higher, and more than double, than for all the other age bands observed here (Ascarza, 2018; Athey & Wager, 2019).

6.5 Limitations

Four main limitations constrain the empirical results. First, the single benchmark dataset analyzed across two annual snapshots (end-2023 to end-2024) validates the methodology and proves non-random state transitions (H1), but continuous trajectories and causal mechanisms require continuous, multi-period replication. Second, the action-layer model reaches a modest AUC ceiling (0.55–0.60); robustness checks confirm this stems from data-signal constraints rather than model under-specification, bounding individual-level targeting precision despite Causal Forest and lift-based compensations. Third (H2), although individual-level treatment effects vary substantially ($SD = 0.073$ vs. $ATE = 0.028$), discount CATE varies weakly across RFM segments (means: 0.0286, 0.0286, 0.0248), proving treatment heterogeneity is far more strongly patterned by customer age than by RFM segment alone. Fourth (H3), while category return-proneness shows descriptive differences in loop behavior, overlapping bootstrap confidence intervals prevent conclusive model-performance confirmation, leaving category-level return moderation as a hypothesis for multi-retailer replication.

6.6 Directions for Future Research

Future research should address these empirical opportunities through four targeted extensions. First, replicating the closed-loop transition test (H1) on continuous, multi-period longitudinal retailer data will trace dynamic migration paths beyond the two-snapshot non-randomness baseline (Netzer et al., 2008). Second, extending Causal Forest heterogeneity analysis (H2) to additional engagement levers (e.g., communication timing, loyalty program enrollment) will test whether age-patterned treatment responsiveness generalizes across action types (Athey & Wager, 2019; Wager & Athey, 2018). Third, re-evaluating H3 using larger multi-category, multi-retailer datasets will provide the sample size required for statistically decisive category-stratified confidence intervals, particularly for

lower-volume segments like accessories (Hjort & Lantz, 2016; Petersen & Kumar, 2009). Fourth, employing qualitative or mixed-methods approaches will clarify the underlying customer experience journeys that differentiate persistent dormant segments from high-value attrition-risk segments.

6.7 Ethical Considerations

In this study, an anonymized secondary benchmark dataset has been used to ensure privacy of the customers, as the data has been anonymized by removing personally identifiable information, using non-traceable identifiers. Secondary data analysis does not require formal informed consent procedures because there is no intervention; however, all data ethics are respected in that usual precautions protect and reveal the provenance of the data and no attempt is made to re-identify customers.

7. Conclusion

This study aimed to investigate (1) if customer engagement states in fashion and apparel retailing are truly closed loop; (2) if the engagement-action effects are substantively distinct; and (3) if product-category return-proneness plays a moderating role in the process that cannot be explained by models from the grocery sector. The LOOP framework revealed statistically significant data for a non-random and closed loop transition structure (H1) with 3,291 fashion retail customers, where discount-driven engagement effects were directionally heterogeneous and stronger in terms of customer age than in terms of conventional value segmentation (H2), and were partially supported with directionally consistent data for return-proneness moderation (H3). In sum, these results bring together the concepts of relationship marketing and heterogeneous treatment effect theory into a neglected retail setting, and provide a statistically sound foundation for age-specific, as opposed to uniformly applied, engagement strategy. The study also shows that closed-loop customer dynamics should be formally tested, not assumed, before they are incorporated into predictive or engagement systems; in other words, that the dynamics should be tested before the dynamics of predictive and engagement systems are tested. These conclusions would be even more strengthened and generalized by longitudinal, multi-retailer replication.

Competing Interests

The authors declare that they have no competing interests.

Consent for Publication

Not applicable.

Ethics Approval and Consent to Participate

Not applicable.

Availability of Data and Materials

Not applicable.

References

1. Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1), 80–98. <https://doi.org/10.1509/jmr.16.0163>
2. Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies*, 5(2), 37–51. <https://doi.org/10.1353/obs.2019.0001>
3. Brodie, R. J., Hollebeek, L. D., Jurić, B., & Ilić, A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research*, 14(3), 252–271. <https://doi.org/10.1177/1094670511411703>
4. Den Boer, A. V. (2015). Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20(1), 1–18. <https://doi.org/10.1016/j.sorms.2015.03.001>
5. Fang, C., Zhou, Y.-W., Feng, X., & Wang, K. (2026). The value of product-specific first-meet coupons: Enhancing customer acquisitions and sales through causal forest. *Omega*, 141, Article 103528
6. Hjort, K., & Lantz, B. (2016). The impact of returns policies on profitability: A fashion e-commerce case. *Journal of Business Research*, 69(11), 4980–4985. <https://doi.org/10.1016/j.jbusres.2016.04.064>
7. Hughes, A. M. (1994). *Strategic database marketing*. Probus Publishing.
8. Karami, A., & Shemshaki, M. (2026). RAISE framework: Robust analytical insights for sales and engagement. *Journal of Revenue and Pricing Management*. <https://doi.org/10.1057/s41272-026-00579-4>

9. Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing*, 80(6), 36–68. <https://doi.org/10.1509/jm.15.0414>
10. Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76–80. <https://doi.org/10.1109/MIC.2003.1167344>
11. Merlano, E. F., Frei, R., Zhang, D., Murzacheva, E., & Wood, S. (2024). Consumer perspectives on interventions to combat fraudulent product returns in omnichannel fashion retail. *International Journal of Physical Distribution & Logistics Management*, 54(10), 969–1001. <https://doi.org/10.1108/IJPDLM-02-2024-0082>
12. Miao, T., Dou, Z., & Mai, S. (2027). Customer segmentation using big data: Implications for retail marketing. *Journal of Retailing and Consumer Services*, 94, Article 104966.
13. Minnema, A., Bijmolt, T. H. A., Gensler, S., & Wiesel, T. (2016). To keep or not to keep: Effects of online customer reviews on product returns. *Journal of Retailing*, 92(3), 253–267. <https://doi.org/10.1016/j.jretai.2016.03.001>
14. Miyatsu, K. (2024). The optimal pricing models: Dynamic pricing personalization based on machine learning. In 2024 IEEE International Conference on Big Data (BigData) (pp. 5147–5153). IEEE. <https://doi.org/10.1109/BigData62323.2024.10825186>
15. Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(3), 20–38. <https://doi.org/10.1177/002224299405800302>
16. Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing Science*, 27(2), 185–204. <https://doi.org/10.1287/mksc.1070.0294>
17. Ogunnusi, O. S., & Akintoye, K. A. (2025). From segmentation to prediction: A unified RFM–machine learning model for online retail analytics. *International Journal of Research and Innovation in Applied Science*, 10(11), 1267–1280.
18. Petersen, J. A., & Kumar, V. (2009). Are product returns a necessary evil? Antecedents and consequences. *Journal of Marketing*, 73(3), 35–51. <https://doi.org/10.1509/jmkg.73.3.035>
19. Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2000). *Driving customer equity*. Free Press.
20. Venkat, R., Carter-Rogers, K., Fullerton, G., & Morales, M. (2026). Customer engagement in a retail setting: An examination of antecedents and outcomes. *International Journal of Retail & Distribution Management*, 54(4), 361–376.
21. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1), 211–229. <https://doi.org/10.1016/j.ejor.2011.09.031>
22. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>
23. Zhou, K., Zhang, H., Fang, Z., & Qin, Y. (2026). The impact of consumers' strategic return behavior on review bias: Evidence from Chinese online fashion retail. *Journal of Retailing and Consumer Services*, 88, Article 104550.