De-duplicated User-item Metrics for E-commerce Conversion Funnel Diagnosis: Method, Validation, and Operational Audit Cues

Aiwei Yang*
University of International Business and Economics, Beijing 100029, China
*Corresponding email: 493093145@qq.com

Abstract

Consumer-to-consumer (C2C) e-commerce sellers routinely face high exposure but low conversion. Existing research, however, relies largely on qualitative descriptions or small-scale surveys and rarely exploits behavioral logs to pinpoint where buyers drop out. Using a public Taobao user-behavior data set from 2017 as a methodological benchmark, this study constructs a de-duplicated user-item conversion framework and cross-checks it against an event-count metric. Under the de-duplicated metric, only 5.640% of interested user-item pairs culminate in purchase - well below the 36.450% implied by raw event ratios - and 75.600% of completed purchases bypass the shopping cart entirely. Conversion peaks during 10:00-13:00 and troughs during 19:00-23:00; the top 10.000% of categories account for 67.900% of sales. The bottleneck is therefore a mismatch between traffic timing and category structure, not shopping-cart abandonment. The framework itself is portable across periods, although the exact time labels must be recalibrated as the platform environment changes. This study anchors its contribution in reusable structural diagnosis rather than in point predictions for any single period.

Keywords

Consumer-to-consumer e-commerce, User behavior, Conversion diagnosis, User-item pair, Operational audit

Introduction

"High exposure, low conversion" is the chronic predicament of sellers in C2C e-commerce: Products are listed successfully but receive little attention or few orders. The bottleneck may not be whether they are seen, but why viewers do not buy. Most studies of second-hand platforms remain at the level of qualitative description or small-scale surveys. They lack quantitative diagnosis based on real behavioral logs and therefore cannot pinpoint where exactly buyers drop out.

This paper is a methodological diagnostic study based on historical large-scale data. We use the full Taobao user-behavior data from November 25th to December 3rd, 2017, as a methodological benchmark. The data set is one of the largest publicly accessible C2C behavioral logs, with complete fields and clear definitions, and is sufficient to support robust de-duplicated estimates. On this basis we construct a reusable de-duplicated user-item conversion framework, empirically test the basic behavioral structure of C2C e-commerce, and use the metric to trace the true view-interest-purchase path. This ensures that funnel ratios are no longer inflated by repeated visits.

The analysis shows that the shopping cart, long thought to be a pivotal step, is far less central in the data than convention assumes. Instead, overlooked time-of-day structure and duration distribution are the hidden levers shaping conversion. The contribution lies in revealing time-invariant structural facts and providing a reusable diagnostic method, not in a real-time claim about the current industry.

The study makes three contributions: It establishes robust procedures for sampling, cleaning, and funnel calibration; it provides a fully replicable analytical workflow; and it distills the findings into reusable operational audit cues rather than operational tactics.

From a platform-governance perspective, C2C transactions involve non-standardized goods and information asymmetry, so the gap between "being seen" and "being bought" is wider than in new-product e-commerce. Prior C2C research indicates that information asymmetry over product authenticity and the unresolved economics of used-product reuse leave supply-side conversion efficiency insufficiently quantified - a black box that needs to be opened [1,2].

The 2017 sample is treated through "de-historicization" lens: Its historical nature is acknowledged, but it is positioned as a vehicle for methodological validation rather than a real-time operational snapshot. Specific timing and scale conclusions must be recalibrated as the platform environment evolves and should not be extrapolated across periods.

Literature review

Trust and information asymmetry are the most frequently cited barriers in second-hand transaction research. Moriuchi and Takahashi show that perceived value, trust, and engagement jointly shape buyer behavior in the C2C online secondary marketplace, establishing trust as a central driver of participation [3]. Kas et al. further show that in used-goods markets, reputation and platform-mediated promises sustain transaction trust, so users rely on platform-level assurance more than on individual counterparties [4]. We synthesize this as follows: the primary hurdle in second-hand transactions is trust rather than mere traffic volume. This provides the theoretical basis for distinguishing "whether buyers dare to buy" from "whether transactions can be completed".

Regarding the drivers of willingness to buy, Wang and Sun identify the principal factors influencing trading on second-hand C2C platforms and stress the role of platform-level mechanisms in shaping purchase behavior [5]. Goldstein et al. model the conversion funnel as a staged progression in which shopper progress is tracked across discrete steps; this funnel logic supplies a methodological analogy for our conversion-diagnosis framework [6]. The shared insight is that conversion can be decomposed into an observable sequence of stages rather than treated as a single outcome.

Platform studies usually start from business models. Wang et al. examine how the economics of used-product reuse are governed within supply chains, while Li et al. show that asymmetric information about product authenticity on C2C platforms is a central friction that assurance services can mitigate; both emphasize the pivotal role of platform-mediated reputation in second-hand transactions. Pang et al. analyze how consumer-to-consumer resale interacts with new-product demand through buyers' utility dependence, clarifying the demand-side structure of resale platforms [7]. Nadeem and Salo demonstrate that value co-creation materially shapes consumer responses in the sharing economy, providing theoretical support for layered operations within platform ecosystems [8]. Collectively, these findings show that platform structure itself shapes the boundary conditions of conversion.

On the data side, it is noted that the Taobao user-behavior data set comes from the Alibaba Cloud Tianchi open platform (data set ID 649, released in 2018). It is public, reproducible, and has been used in multiple user-behavior studies. Methodological research on extending and preparing behavioral data sets for reproducible analysis remains limited. This study contributes a transparent, log-based procedure that others can reuse [9]. In sum, existing literature shows that trust determines the threshold of a transaction, while the present study reveals that timing and structure determine conversion efficiency. The two are complementary: Trust concerns whether buyers dare to buy; timing and structure concern whether transactions can be completed.

Existing studies mainly answer "who is more likely to buy" but rarely answer "where the buying process breaks off". This study fills the gap with behavioral logs: Trust and risk perception determine whether buyers dare to buy, while time-of-day and category structure determine whether transactions can be completed. Together they form a complete explanation of transaction outcomes.

Methodology

It should be noted that the diagnostic framework is applicable across periods. The framework rests on three mathematical facts that do not change over time. First, the hierarchical structure of behavioral sequences (view-interest-purchase) is an observable order in which users act on items, independent of time. Second, the logic for determining whether a purchase occurs is based on state transitions of user-item pairs and does not depend on traffic scale in a particular year. Third, the bias introduced by using event counts to approximate conversion stems from repeated counting, not from the sample year. Thus, the 2017 Taobao data serve only as a validation vehicle to test whether the framework can stably reproduce structural facts; they are not a validity precondition of the framework itself. For this reason, specific values will drift as the environment changes, but the structural relationships revealed by the framework can be reused in any setting with behavioral logs.

A two-sample strategy is adopted. The full sample covers 100,150,807 raw records for aggregate estimation. A 2.000% user-level systematic sample (user_id mod 50=7) yields a subset of 2,034,389 records from 19,770 users for de-duplicated path analysis. Because user_id is randomly assigned by the platform, modulo sampling can be treated as approximately random.

Conversion is measured through two contrasting metrics. The event-count metric simply sums events of each type. The user-item-pair metric first de-duplicates records by (user_id, item_id) and then checks whether the pair has viewed, shown interest in, or purchased the item, thereby identifying the true sequence. Placing the two metrics side by side exposes the distortion caused by raw event ratios. Diagnosis covers only item pairs with at least one view; zero-exposure items are excluded.

At the category level, the transaction distribution is characterized using quantiles and Pareto concentration. Time lag is measured by the cumulative distribution of hours between cart addition and purchase. Time-of-day patterns are aggregated by hour in Beijing time. Statistical tests include Pearson correlation and two-sample t-tests, with significance set at 0.050. The entire workflow is implemented in Python (pandas, pyarrow, scipy), and the scripts are archived with the paper.

Technically, the full scan reads the 3.67 GB file in 64 MB blocks via pyarrow to avoid memory overflow. The 2.000% sample is generated by modulo sampling on user_id and stored as sample_2pct.parquet. The side-by-side design of the two metrics is intended to make the bias hidden in event ratios visible. Because user_id is randomly assigned, the sample is approximately random, and the de-duplicated results are stable and replicable.

Data and preprocessing

Data come from the publicly available “Taobao User Behavior Dataset” on Alibaba Cloud Tianchi. The data contain five fields: user_id, item_id, item_category, behavior_type (pv/fav/cart/buy), and timestamp (Unix seconds), as listed in Table 1. There are 100,150,807 raw records.

Table 1. Data dictionary.

Field	Meaning	Values / Type
User_id	User identifier	Integer
Item_id	Item identifier	Integer
Item_category	Item category	Integer
Behavior_type	Behavior type	pv / fav / cart / buy
Timestamp	Behavior time	Unix seconds

Cleaning follows the official observation window: 00:00 on November 25th, 2017, to 23:59 on December 3rd, 2017, Beijing time. After this step, 55,576 out-of-window records (0.056%) are removed, leaving 100,095,231 valid records, as shown in Table 2. No missing values or duplicate primary keys are found at the field level, confirming data reliability. The four behavior_type values are distributed reasonably: pv 89.575%, cart 5.525%, fav 2.886%, buy 2.014%, reflecting typical user interactions.

Table 2. Cleaning steps and record counts.

Step	Records	Note
Raw records	100,150,807	Full data set
Removed out-of-window	55,576	All pv, 0.055%
Valid records after cleaning	100,095,231	Used in analysis
pv / cart / fav / buy	89,660,688 / 5,530,446 / 2,888,258 / 2,015,839	Distribution by behavior type

To operationalize “high exposure, low conversion”, “high exposure” is mapped to an item being viewed (pv) and "low conversion" to a user-item pair ending without purchase. Under the de-duplicated metric, the 2.000% sample contains 1,544,947 distinct user-item pairs. Of these, 1,445,341 had at least one view and constitute the diagnosed universe, among which 165,560 showed interest (favorite or cart) and 38,284 eventually purchased. This mapping keeps conversion rates from being inflated by repeated visits and stays closer to real transactions. Of course, mapping "viewed but not purchased" to "low conversion" may include users who were interested but had not yet ordered; the results should therefore be interpreted as a lower bound on listing conversion, not as a precise failure rate.

For variable operationalization, timestamp is converted from Unix seconds to Beijing time and restricted to the official observation window. The four behavior_type values map directly to the view-favorite-cart-purchase chain.

Note that the data set lacks price and seller-reputation fields, so the conclusions hold only at the category level. They should be understood as the marginal difference due to timing and structure given product and seller conditions, not as a causal claim about individual transactions.

Results

Dual-metric funnel: The shopping cart is overstated

Under the event-count metric, buy/cart =36.450%, i.e., an add-to-cart-purchase conversion rate of 36.450%. However, the user-item-pair metric (2.000% sample) reveals the conversion rates, as shown in Table 3 and Figure 1: only 5.640% of interested pairs end in purchase, and only 2.649% of viewed pairs end in purchase.

More importantly, 75.600% of completed purchases never generated an interest signal (neither favorite nor cart) and moved directly from view to purchase. The shopping cart is far from a central step in C2C transaction paths; diagnoses that rely on it will systematically overestimate conversion efficiency.

Table 3. Dual-metric conversion funnel comparison.

Metric	Indicator	Value
Event count	Page views (pv)	89,660,688
Event count	Cart additions (cart)	5,530,446
Event count	Favorites (fav)	2,888,258
Event count	Purchases (buy)	2,015,839
Event count	Cart-to-purchase (buy/cart)	36.450%
User-item pair	Viewed pairs	1,445,341
User-item pair	Interested pairs (fav/cart)	165,560
User-item pair	Purchasing pairs	38,284
User-item pair	Interest-to-purchase	5.640%
User-item pair	View-to-purchase	2.649%
User-item pair	Purchases with no interest signal	75.600%

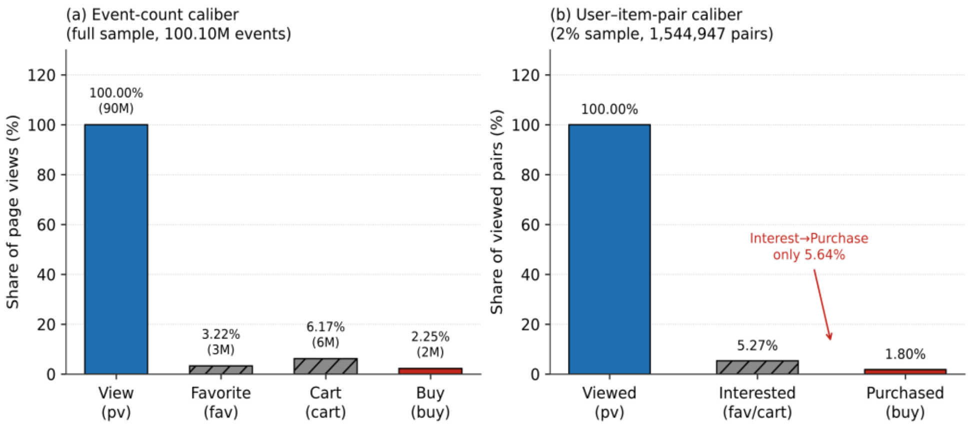

Figure 1. Dual-metric conversion funnel (event-count vs. user-item-pair metrics).

Time-lag structure after cart addition

Among the 6,606 pairs that both added to cart and purchased, the median time from cart addition to purchase is 19.950 hours, with the 25th percentile at 2.720 hours and the 75th percentile at 42.720 hours, as shown in Figure 2.

Purchases completed within 24.000 hours account for 58.500%, while only 23.700% occur within 2.000 hours. This indicates that cart addition exhibits an overnight-decision feature: Most purchases happen after the day of cart addition, leaving sellers a window for re-engagement.

This lag structure means that adding to cart is not equivalent to churn; it is the starting point of subsequent engagement.

The real conversion window lies between half a day and two days after cart addition, precisely when the platform can intervene.

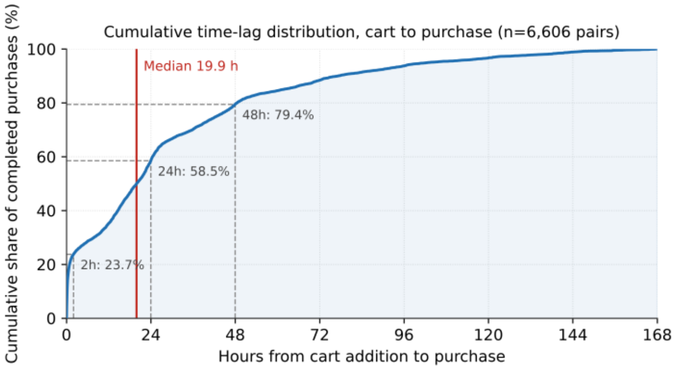

Figure 2. Cumulative distribution of cart-to-purchase time lag.

Time-of-day divergence: Traffic peak meets conversion trough

Hour-by-hour, page-view volume peaks at 21:00 (7.5384 million), yet the conversion rate at that hour is only 1.930%, clearly below the daily average of about 2.250%, producing a divergence between traffic peak and conversion rate. The conversion-rate peak occurs at 10:00 (2.950%), as shown in Figure 3.

Dividing the day into five identified bands, as listed in Table 4, the 10:00-13:00 band has the highest conversion rate (2.820%) and efficiency index (1.253). The 19:00-23:00 band alone accounts for 36.360% of page views but has only a 1.940% conversion rate and an efficiency index of 0.863. Traffic and conversion clearly diverge: In the evening, many view but few actually buy.

Table 4. Traffic and conversion efficiency by time band.

Time band	Share of views (%)	Share of purchases (%)	Conversion rate (%)	Efficiency index
0:00-5:00	7.590	5.750	1.700	0.757
6:00-9:00	11.410	10.750	2.120	0.943
10:00-13:00	19.470	24.410	2.820	1.253
14:00-18:00	25.180	27.710	2.470	1.101
19:00-23:00	36.360	31.380	1.940	0.863

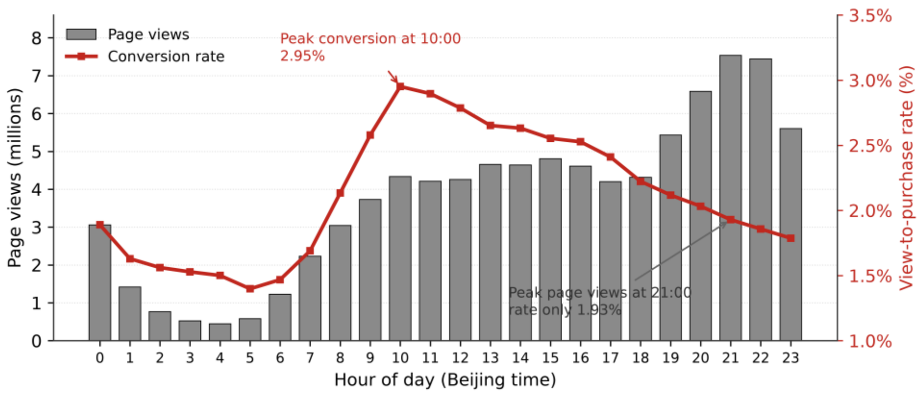

Figure 3. Hourly page views (bars) and conversion rate (line).

The efficiency index (period conversion rate divided by daily average conversion rate) shows that morning traffic-to-conversion efficiency is about 1.452 times that of the evening. Blindly increasing spending in the evening is equivalent to pouring budget into a conversion trough; moving the same budget to the morning yields higher returns.

Extreme category concentration

Among 3,167 categories with at least one cart addition, the P90/P10 ratio of transaction rates reaches 10.688, and the top 10.000% of categories contribute 67.900% of transactions (the top 20.000% contribute 80.840%), as reported in Table 5 and Figure 4.

In addition, 18 categories have zero cart additions yet an average purchase rate of 20.980%, far higher than the normal-category rate of 3.970%.

Why these zero-cart categories still generate purchases requires attribution to price and search terms, so generic reactivation strategies should not be applied directly. Category-level correlation (n=2,357, r=0.630, p<0.001) shows that cart heat is positively associated with transactions, but 16 categories show high interest yet low transaction volume ("pseudo-hot" categories), indicating that part of the traffic is not effectively converted.

Table 5. Category transaction quantiles and Pareto concentration.

Indicator	Value
Categories with cart > 0	3,167
P10 / P25 / P50 / P75 / P90 / P99 transaction rate (%)	0.770 / 1.610 / 3.220 / 5.530 / 8.230 / 14.050
P90 ÷ P10	10.688
CR1 / CR5 / CR10 / CR20 (%)	26.950 / 54.330 / 67.900 / 80.840
Categories with zero cart additions	18
Average purchase rate in zero-cart categories (%)	20.980
Average purchase rate in normal categories (%)	3.970

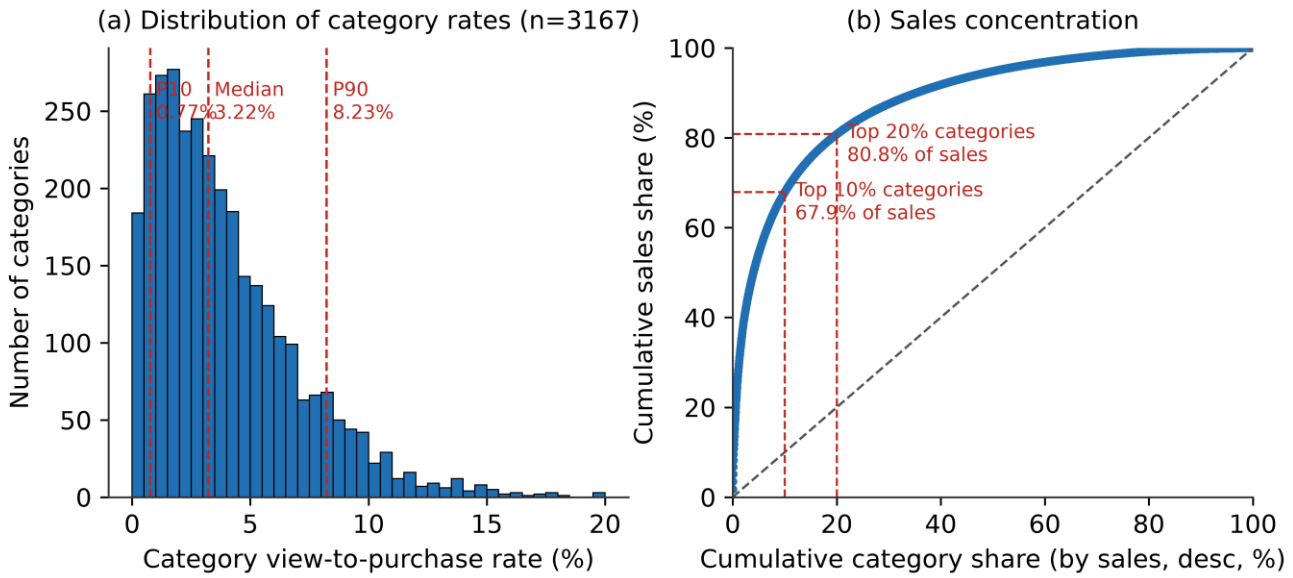

Figure 4. Category transaction distribution and Pareto concentration.

Weekends do not convert better

Average daily page views on weekends are 17.731% higher than on weekdays, yet weekend conversion (2.120%) is lower than weekday conversion (2.370%), as shown in Table 6.

Thus the weekend traffic dividend does not translate into a transaction dividend, corroborating the time-of-day divergence finding.

Table 6. Weekend vs. weekday comparison.

Indicator	Weekend	Weekday	Test
Average daily page views	10,871,947.500	9,234,579.600	Weekend +17.731% over weekday
Conversion rate (%)	2.120	2.370	Weekend −0.250 percentage points vs. weekday

Discussion

These results point to a counterintuitive conclusion: The shopping cart is a weak signal, while time and category structure are the real drivers of conversion. Treating the 36.450% event ratio as conversion efficiency creates the illusion that "adding to cart is halfway to purchase". The de-duplicated metric corrects this illusion, echoing Kas et al., who find that platform-mediated reputation and promises - rather than the attributes of the specific object - sustain transaction trust in used-goods markets.

The implication for smart marketing is not fixed tactics but three auditable structural signals. Time-of-day divergence signals that "when to reach" is closely tied to conversion efficiency. The fact that 75.600% of purchases bypass the cart suggests that recall metrics centered on cart addition may overstate effectiveness. Category concentration signals the risk of traffic being siphoned to the head of the distribution - the last point resonates with Nadeem and Salo, who show that platform-level value co-creation shapes how users engage with and concentrate within platform ecosystems.

Boundaries should be noted. The data set is a Taobao all-category behavioral log; second-hand categories are only one of many high-frequency scenarios. Extending the conclusions to pure second-hand platforms requires category-specific data. The data window is only nine days, and seasonal or promotion effects are not included. Conclusions should be treated as structural diagnosis rather than precise prediction.

The historical nature of the sample must also be considered. The 2017 data window was dominated by "human-seeks-goods (search-based discovery)" search logic. Lin et al. show that emotion and perception mediate impulse buying in live-streaming shopping through a stimulus-organism-response mechanism [10]. This indicate that the current environment has moved toward "goods-find-human (interest-based discovery)" live- streaming and short-video interest commerce, with structural shifts in traffic allocation and decision paths. Therefore, the specific time labels (e.g., 10:00-13:00 as a conversion highland) should not be applied directly to the current environment; they must be recalibrated against new traffic-allocation and content-distribution mechanisms.

In this sense, the study's core contribution is not the timeliness of any set of percentages but time-invariant structural facts and a reusable methodology. First, the shopping cart is only a weak signal in real transactions (75.600% of purchases bypass it), and this does not depend on traffic scale in a specific year. Second, conversion efficiency is jointly determined by timing and category structure and can diverge from traffic scale; this logic can be transferred to any setting with behavioral logs. Third, the framework that contrasts the de-duplicated user-item-pair metric with the event-count metric provides a reproducible diagnostic baseline. Anchoring the contribution in the stability of structure and method avoids using a historical snapshot to assert conclusions about an evolving industry.

In other words, if managers rerun this framework on current behavioral logs, they obtain up-to-date "time-slot efficiency indices" and "category concentration measures". The audit value lies not in the historical numbers themselves but in the rolling re-calculation of this diagnostic.

Conclusion

This study has constructed and validated a C2C e-commerce conversion-diagnosis framework based on a de-duplicated user-item-pair metric, using the 2017 Taobao public data set for empirical testing and validity assessment. The core conclusion is that the transaction bottleneck lies not in the shopping cart but in time-of-day mismatch and category imbalance. The contribution is anchored in time-invariant structural facts and reusable diagnostic methods.

The study has two limitations. First, the data window is short (November 25th to December 3rd, 2017, nine days) and does not cover promotional periods. Second, it lacks price and seller-reputation fields, so within-category differences cannot be explained. Future work can introduce multi-period data and seller-side variables to test the causal effects of structural relationships further. It can also recalibrate specific time labels in live-streaming and short-video settings.

Operational audit cues

Based on the structural findings above, the original "strategic recommendations" are downgraded into three operational audit cues rather than being offered as prescriptive strategic recommendations. They identify structural signals worth monitoring continuously and do not constitute operational tactics for the present; implementation must be combined with the current environment and controlled experiments.

The following cues use the 2017 sample only as a demonstration vehicle. In actual operations, auditors should replace the specific values in each cue with real-time metrics from current databases (e.g., a de-duplicated user-item-pair table for the last 30 days). The values below are used only to illustrate logic and should not be extrapolated.

Cue 1: Structural mismatch in time-of-day conversion efficiency. In this data set, the 10:00-13:00 band has the highest efficiency index (1.253) and the 19:00-23:00 band the lowest (0.863); traffic-to-conversion efficiency per unit differs by about 1.450 times. In practice, this should be recalculated with current period-by-period data and used to adjust the audit metric for ad placement, not as a direct instruction on when to spend more.

Cue 2: Weak mediation of the shopping-cart signal. In this data set, 75.600% of purchasing pairs never produced an interest signal such as cart addition or favorite; the cart is only a weak mediator. In practice, this should be recalculated with current de-duplicated user-item-pair data and used to evaluate whether recall metrics centered on cart addition overstate effectiveness, not to design recall scripts directly.

Cue 3: Category transaction concentration. In this data set, the top 10.000% of categories contribute 67.900% of transactions, and 18 categories have zero cart additions, showing obvious head-siphoning of traffic. In practice, this should be recalculated with current category-level transaction data and used to identify a list of “high-exposure, low-conversion” categories for governance auditing, not as a direct order to redistribute traffic.

On current relevance, the specific time labels above (e.g., 10:00-13:00 as a conversion highland and 19:00-23:00 as a trough) will become outdated as traffic-allocation mechanisms evolve. However, the three structural misalignment logics - “divergence between traffic scale and conversion efficiency”, “weak cart mediation” and “head siphoning in categories” - retain cross-period explanatory power. The practical audit focus is not on copying historical time bands but on using this framework as a baseline for rolling metric recalculation and structural monitoring in the current environment.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Li, L., Fang, X., Lim, Y. F. (2023) Asymmetric information of product authenticity on C2C e-commerce platforms: How can inspection services help? *Manufacturing & Service Operations Management*, 25(2), 631-647.

[2] Wang, M., Zhang, J., Cheng, T. C. E. (2024) Sharing or reselling? How does manufacturer handle used products? *Naval Research Logistics*, 71(5), 645-659.

[3] Moriuchi, E., Takahashi, I. (2022) The role of perceived value, trust and engagement in the C2C online secondary marketplace. *Journal of Business Research*, 148, 76-88.

[4] Kas, J., Corten, R., van de Rijt, A. (2023) Trust, reputation, and the value of promises in online auctions of used goods. *Rationality and Society*, 35(4), 387-419.

[5] Wang, L. L., Sun, H. (2023) Influencing factors of second-hand platform trading in C2C e-commerce. *Journal of Intelligent Management and Decision*, 2(1), 21-29.

[6] Goldstein, A., Oestreicher-Singer, G., Barzilay, O., Yahav, I. (2022) Are we there yet? Analyzing progress in the conversion funnel using the diversity of searched products. *MIS Quarterly*, 46(4), 2015-2054.

[7] Pang, C., Jiang, L., Li, G. (2024) Cannibalization or enhancement: Effects of consumer-to-consumer resale with consumers’ utility dependence. *International Journal of Electronic Commerce*, 28(3), 416-446.

[8] Nadeem, W., Salo, J. (2024) Does value co-creation matter? Assessing consumer responses in the sharing economy. *Information Technology &*

People, 37(3), 1279-1304.

[9] Heinrich, B., Hopf, M., Lohninger, D., Schiller, A., Szubartowicz, M. (2022) Something's missing? A procedure for extending item content data sets in the context of recommender systems. *Information Systems Frontiers*, 24(1), 267-286.

[10] Lin, S. C., Tseng, H. T., Shirazi, F., Hajli, N., Tsai, P. T. (2023) Exploring factors influencing impulse buying in live streaming shopping: a stimulus-organism-response (SOR) perspective. *Asia Pacific Journal of Marketing and Logistics*, 35(6), 1383-1403.