Original Research Article

Pricing, Perception, and Popularity: Modeling the Impact of Discounts and Categories on Product Ratings and Engagement in E-Commerce Platforms

Vito Hsu

Arcadia High School, 180 Campus Dr, Arcadia, CA 91006, United States

ABSTRACT

Online shopping has changed how people buy products by offering more access, choices, and convenience. Still, success in e-commerce depends heavily on pricing, discount strategies, and customer response. This project examines how product price and discount level influence customer satisfaction (ratings) and customer engagement (review counts) using a public Kaggle dataset from 2022 that includes pricing, discount fractions, ratings, review counts, and product categories. Three models were used: a multiple linear regression for ratings, a Negative Binomial regression for review counts, and a joint model that estimates both together. The results show that a one–standard deviation increase in log price raises average product ratings by 0.047 points, but lowers rating counts by 0.171 on the log scale—meaning higher-priced items are rated better but reviewed less. Larger discounts show the opposite pattern; a one–standard deviation increase in discount fraction decreases ratings by 0.053 and reduces review counts by 0.152, suggesting that heavy markdowns may hurt both perceived quality and engagement. Category differences also appeared: Electronics products received significantly more reviews (0.672) but slightly lower ratings (–0.127), while Home & Kitchen and Office Products performed worse across both outcomes. The joint model provided the most stable estimates and helped connect the relationships between satisfaction and engagement. Overall, the study demonstrates that pricing and discount strategies strongly shape how customers rate and interact with products online.

Keywords: Online Shopping; Pricing; Discounts; Customer Satisfaction; Product Ratings; Review Counts; Product Categories; Regression Models

INTRODUCTION

E-commerce has grown rapidly and changed the way people shop and how businesses operate. It offers convenience, wider product selection, and often lower prices, benefits that traditional stores cannot always provide (1). For people living in rural or remote areas, online shopping gives access to goods they might not otherwise find locally (2). At the same time, e-commerce has enabled faster logistics, lower startup costs, and more data-driven strategies, helping companies reach more customers and scale their operations (3).

Pricing and discounts are especially important in shaping customer perceptions. Very deep discounts can create doubts about a product's quality, while moderate discounts can generate positive feelings and increase perceived value (4). Trust in the platform and product also plays a key role: studies show that e-trust increases

Corresponding author: Vito Hsu, E-mail: vitosstyles@gmail.com.

Accepted November 22, 2025
https://doi.org/10.70251/HYJR2348.36685692

perceived value and satisfaction, which in turn affects purchasing behavior and likelihood of leaving reviews (5).

Despite the importance of these factors, few studies examine how pricing strategies affect both customer satisfaction, measured through ratings, and customer engagement, measured through review counts, especially across multiple product categories (6). To address this gap, this study uses a public Kaggle dataset containing product prices, discount percentages, rating scores, review counts, and product categories (7).

This research applies a multiple linear regression to estimate how pricing and discounting relate to average ratings, a Negative Binomial regression to predict review counts, and a joint model to capture the relationship between satisfaction and engagement. The goal is to understand how pricing and discounts influence customer behavior and to provide practical insights for online retailers and platform designers to maintain trust and maximize value. By linking pricing strategy with customer response data, this study contributes to a better understanding of what drives long-term success in e-commerce (8).

While prior research has examined pricing, discounts, and trust separately, few studies have systematically investigated how these factors simultaneously influence both customer satisfaction (ratings) and engagement (review counts) across multiple product categories (6). This study addresses this gap by combining pricing data, discount information, and category effects to provide a holistic view of consumer behavior in e-commerce.

Few studies have jointly modeled ratings and rating counts using both linear and count models in a unified framework, which can capture the interdependence between customer satisfaction and engagement. By integrating these outcomes, this study provides more accurate estimates and a richer understanding of how pricing and discounts shape consumer behavior.

METHODS AND MATERIALS

Data Description

The dataset for this project came from a public source on Kaggle (7). It includes customer reviews and product details taken from Amazon. This dataset was chosen because it fits well with what the study aims to explore. It offers a mix of numbers and written reviews across different product categories, which makes it great for analyzing online shopping behavior. The original data had over a thousand entries and sixteen variables, such as product names, prices, discounts, ratings, number of reviews, and links to the items. Having both the numbers and the text helps give a full picture of how people shop and rate products online.

Since this research focuses on how prices and product categories affect ratings, the data was cleaned and narrowed down to five main parts: rating, rating_count, category, actual_price, and discount_percentage. These categories show both customer opinions (like ratings and review counts) and money-related factors (like prices and discounts) that could change how people see a product's quality. The final version of the data made it easier to run predictions and build regression models. Table 1 lists the main statistics for each variable, including the average, median, standard deviation, and range of values.

Table 1. Descriptive Statistics of Variables Used

Numerical Variables

Variable	mean	median	std	min	max
actual_price	5472.214625	1690	10906.8112	39	139900
discount_percentage	47.69415808	50	21.6193928	0	94
rating	4.096907216	4.1	0.28970896	2	5
rating_count	18314.28797	5179	42841.1973	2	426973

Categorical Variables

Category	Frequency
Electronics	526
Computers & Accessories	451
Home & Kitchen	447
Office Products	31

Methods

As previously outlined, the central objective of this study is to quantify the relationship between two key response variables, product rating and rating count, and three major predictors: actual price, discount percentage, and product category. These relationships are investigated using a series of regression models suited to the statistical properties of the data. Specifically, a multiple linear regression model is first employed to estimate the average rating across products, followed by a negative binomial regression model to capture the dynamics of rating counts. Finally, a joint modeling framework is introduced to integrate the two outcomes and account for shared latent heterogeneity. The following subsections describe each of these models in detail.

Multiple Linear Regression Model

To analyze how product ratings respond to variation in pricing and categorical variables, this study first applies a multiple linear regression model. This choice is motivated by the approximately normal distribution of the rating variable and the continuous nature of the primary predictors. The model estimates the product rating as a linear function of the actual price, the discount percentage, and dummy variables representing product categories (9).

Formally, the model is specified as:

$$R_i = \beta_0 + \beta_1 P_i + \beta_2 D_i + \sum_{k=1}^{K-1} \gamma_k C_{ik} + \varepsilon_i \tag{1}$$

where:

- R_i is the product rating for item i,
- Pi is the actual price,
- Di is the discount percentage,
- Cik denotes dummy variables for product categories (with K total categories and one omitted as reference),
- β0, β1, β2, γk are model coefficients,
- εi~N(0,σ2) is the error term.

This model allows for direct interpretation of the marginal effects of pricing and categorical features on consumer-perceived product quality. The primary advantages of this modeling choice lie in its simplicity, interpretability, and ease of implementation (10).

Negative Binomial Regression Model

While the linear model is suitable for rating prediction, the rating count variable is discrete and non-negative, taking integer values that represent the number of user engagements. Initial analysis of the data reveals that rating counts are not only count-based but also exhibit overdispersion, meaning the variance significantly exceeds the mean, a condition that violates the assumptions of simpler models like Poisson regression. To address this, a negative binomial regression model is employed, which introduces an extra parameter to account for the overdispersion and thus offers a more flexible structure for modeling count data.

The negative binomial model is formulated as:

$$log(\mu_i) = \alpha_0 + \alpha_1 P_i + \alpha_2 D_i + \sum_{k=1}^{K-1} \delta_k C_{ik} \tag{2}$$

where:

- μ_i is the expected rating count for item i,
- The actual rating count, $Y_i \sim N(\mu_i, \theta)$ where θ is the dispersion parameter,
- α_0, α_1, α_2, δ_k are model coefficients,
- The other variables are defined as before.

The log-link function ensures that predicted rating counts are positive, while the dispersion parameter θ relaxes the equidispersion constraint of the Poisson model. This makes the negative binomial model particularly suitable for consumer engagement data, where a few products may receive disproportionately large numbers of ratings. A key strength of this approach is its robustness to variance inflation in the dependent variable. However, this flexibility also introduces estimation complexity and can make the interpretation of coefficients less intuitive compared to the linear model. Still, the negative binomial model provides a statistically sound framework for understanding how price and discount strategies influence user participation metrics such as review frequency.

Joint Modeling Framework

Although the two preceding models address the rating and rating count separately, both outcomes arise from the same products and are influenced by shared unobserved factors. To capture this dependency, a joint modeling framework is introduced. The joint model was implemented in R using the rstanarm package (11), which supports joint estimation of linear and count outcomes in a Bayesian framework.

The structure of the joint model is as follows:

$$\begin{aligned} \{R_i &= \beta_0 + \beta_1 P_i + \beta_2 D_i + \sum_{k=1}^{K-1} \gamma_k C_{ik} + \varphi_i log(\mu_i) \\ &= \alpha_0 + \alpha_1 P_i + \alpha_2 D_i + \sum_{k=1}^{K-1} \delta_k C_{ik} + \omega_i \end{aligned} \tag{3}$$

where:

- φ_i and ω_i are the error terms for the rating and rating count equations, respectively,
- A key assumption is that $\mathrm{Cov}(\varphi_i, \omega_i) \neq 0$, capturing the shared latent heterogeneity across products.

The correlation between these error terms was modeled using a bivariate normal distribution, allowing the model to capture unobserved factors, like brand reputation or product popularity, that simultaneously influence ratings and review counts.

This modeling strategy allows for information sharing across the two outcomes, potentially improving the efficiency and accuracy of parameter estimates for both models. By jointly modeling consumer satisfaction and engagement, the approach captures richer behavioral patterns than could be obtained from independent models. Nonetheless, joint models come with increased complexity, both computationally and in terms of model specification. Estimation requires assumptions about the joint distribution of error terms, commonly a bivariate normal distribution, which must be validated carefully. Despite these challenges, the joint framework is particularly well suited to e-commerce contexts, where multiple response variables often reflect underlying consumer preferences and perceptions in tandem.

RESULTS

Multiple Linear Regression Results for Reviewing Scores

The multiple linear regression results presented in Table 2 reveal several important relationships between pricing, discounting, and customer ratings across e-commerce product categories. The baseline intercept of 4.195 indicates that, on average, products in the reference category (*Computers & Accessories*) receive relatively high customer satisfaction scores, suggesting an overall positive perception of online products. Category coefficients demonstrate meaningful variation: both *Electronics* (–0.127, $p < 0.001$) and *Home & Kitchen* (–0.176, $p < 0.001$) products are rated significantly lower than the baseline category, implying that consumers tend to be more critical of these goods—possibly due to higher expectations regarding performance, durability, or ease of use. In contrast, *Office Products* show a small positive, though marginally significant, deviation (0.102, $p = 0.061$), indicating relatively stable or slightly better satisfaction levels.

The continuous predictors highlight the dual influence of price and discount on perceived quality. The standardized log-price variable (z_log_price) has a positive and statistically significant coefficient (0.047, $p < 0.001$), confirming that higher-priced products generally receive higher ratings. This pattern aligns with consumer behavior theories suggesting that price often serves as a heuristic for quality—buyers may interpret premium pricing as an indicator of superior performance or brand credibility. Conversely, the standardized discount fraction (z_disc_frac) shows a negative and significant effect (–0.053, $p < 0.001$), meaning that larger discounts are associated with slightly lower ratings (4). This finding supports the notion that excessive markdowns can reduce perceived value or raise concerns about product authenticity, shelf life, or quality control.

Table 2. Detailed Modeling Parameters for the Multiple Linear Regression Model

term	coef	std_err	t	p_value
Intercept	4.195	0.014	303.048	0.000
C(category) [T.Electronics]	-0.127	0.019	-6.529	0.000
C(category) [T.Home&Kitchen]	-0.176	0.020	-8.958	0.000
C(category) [T.OfficeProducts]	0.102	0.054	1.878	0.061
z_log_price	0.047	0.008	5.731	0.000
z_disc_frac	-0.053	0.008	-6.802	0.000

From a practical and policy perspective, these results underscore the importance of strategic pricing management within competitive e-commerce environments. Retailers and platform designers should aim to maintain balanced discount strategies, leveraging price incentives to stimulate demand while avoiding signals that could erode product reputation. Transparent pricing histories and value-based promotion policies can strengthen consumer trust and protect brand equity. At the same time, policymakers and analysts may use such evidence to guide fair-pricing frameworks and algorithmic oversight that prevent manipulative discounting practices. Ultimately, the regression outcomes demonstrate that sustainable e-commerce performance depends not only on sales volume but also on maintaining the perceived integrity and value of products through thoughtful pricing and promotional strategies.

Negative Binomial Regression Results for Rating Counts

Table 3 presents the estimated parameters from the Negative Binomial regression model, which predicts *rating count*—a measure of customer engagement—based on product category, standardized log price (z_log_price), and standardized discount fraction (z_disc_frac). The intercept value of 9.707 indicates a high baseline log-count of ratings for the reference category (*Computers & Accessories*), reflecting that these products generally attract substantial consumer attention and review activity. Category effects display clear differences across product types. Products in the *Electronics* category show a positive and significant coefficient (0.672, $p < 0.001$), suggesting that they tend to receive more ratings than the baseline group, likely due to their widespread popularity and frequent customer interaction. In contrast, both *Home & Kitchen* (–0.961, $p < 0.001$) and *Office Products* (–1.691, $p < 0.001$) exhibit significantly lower engagement levels, indicating that these categories attract fewer reviews, possibly due to lower purchase frequency or less consumer motivation to provide feedback.

Table 3. Detailed Modeling Parameters for the Negative Binomial Regression Model

term	coef	std_err	z	p_value
Intercept	9.707	0.071	136.607	0.000
C(category) [T.Electronics]	0.672	0.099	6.772	0.000
C(category) [T.Home&Kitchen]	-0.961	0.102	-9.411	0.000
C(category) [T.OfficeProducts]	-1.691	0.281	-6.014	0.000
z_log_price	-0.171	0.048	-3.598	0.000
z_disc_frac	-0.152	0.044	-3.457	0.001

The continuous predictors—price and discount—reveal critical behavioral insights regarding consumer engagement dynamics. The negative coefficient for z_log_price (–0.171, $p < 0.001$) indicates that higher-priced products generally receive fewer ratings, even after controlling for category differences (12). This finding implies that as product prices rise, purchase frequency and user interaction decline, it is likely to reflect smaller customer bases for premium goods. Similarly, the discount variable (z_disc_frac) also carries a negative coefficient (–0.152, $p = 0.001$), meaning that products with deeper discounts tend to accumulate fewer ratings. This pattern may indicate that heavy discounting attracts more transactional, price-sensitive buyers who are less inclined to leave reviews, or that consumers perceive steeply discounted items as less credible, thereby lowering post-purchase engagement (4).

From a managerial and policy standpoint, these findings suggest that consumer engagement is sensitive to both pricing structure and category characteristics. Retailers should recognize that while discounts can drive short-term sales, excessive or poorly targeted markdowns may dampen long-term engagement metrics such as reviews and customer feedback—key elements of algorithmic visibility on e-commerce platforms. Moderating discount frequency and coupling promotions with value-driven messaging could sustain both sales and review activity. Moreover, policy designers and platform administrators should consider integrating category-specific engagement models into recommendation and ranking algorithms to ensure fair exposure across diverse product types. Overall, the Negative Binomial results highlight that sustaining active consumer participation in online marketplaces requires not only competitive pricing but also balanced strategies that preserve perceived value and encourage authentic user interaction.

Joint Modeling Results for both Rating Scores and Counts

Table 4 presents the results of the joint modeling framework, which simultaneously estimates product *ratings* using a multiple linear regression component and *rating counts* using a Negative Binomial component. This integrated structure allows both customer satisfaction (rating) and engagement (rating count) to be modeled in parallel while accounting for their shared underlying influences (11). The intercepts for both submodels—4.195 for the linear component and 9.707 for the Negative Binomial component—indicate high baseline values, consistent with generally positive consumer feedback and strong engagement levels across the dataset. Each submodel reveals significant effects across category, pricing, and discount variables, confirming that these features jointly influence how consumers evaluate and interact with products on e-commerce platforms.

Within the categorical variables, consistent directional patterns emerge. For the *rating* component, the coefficients for *Electronics* (–0.127, $p < 0.001$) and *Home & Kitchen* (–0.176, $p < 0.001$) are negative, suggesting slightly lower satisfaction scores relative to the reference category (*Computers & Accessories*), while *Office Products* (0.102, p = 0.061) shows a marginally positive tendency. In the *rating count* component, however, the signs diverge: *Electronics* displays a strong positive association (0.672, $p < 0.001$), whereas *Home & Kitchen* (–0.961, $p < 0.001$) and *Office Products* (–1.691, $p < 0.001$) exhibit lower engagement levels. These dual outcomes reflect the multidimensional nature of consumer behavior—where satisfaction and participation are shaped by overlapping yet distinct category-level perceptions and motivations.

The continuous predictors further illustrate this complexity. The standardized log-price variable (z_log_price) shows a positive and significant effect on ratings (0.047, $p < 0.001$) but a negative and significant effect on rating counts (–0.171, $p < 0.001$), indicating that higher-priced products are perceived more favorably yet receive fewer reviews. Similarly, the standardized discount fraction (z_disc_frac) has a negative impact on both outcomes—(–0.053, $p < 0.001$) for ratings and (–0.152, p = 0.001) for rating counts—suggesting that large discounts may simultaneously reduce perceived quality and dampen post-purchase engagement. These aligned but distinct patterns underscore how pricing and promotional signals jointly shape both evaluative and participatory dimensions of consumer response.

In terms of fit, the joint model maintains strong statistical reliability, as indicated by low AIC and BIC values (398.935 and 430.631 for the linear component; 30211.894 and 30248.874 for the Negative Binomial component) and an acceptable dispersion parameter (NB_α = 2.070). The moderate R^2 value (0.089) in the linear submodel suggests that although pricing and category factors meaningfully explain variations in ratings, additional behavioral or contextual variables—such as product age, brand reputation, or advertising exposure—may further refine prediction accuracy. The Negative Binomial's overdispersion parameter confirms that count data variability exceeds Poisson expectations, validating the appropriateness of this joint modeling approach.

Collectively, the joint estimation highlights how integrating satisfaction and engagement into a unified analytical framework yields a more holistic understanding of consumer behavior in e-commerce contexts. By simultaneously capturing how customers *evaluate* and *interact* with products, this approach provides a richer, multi-dimensional foundation for decision-making in pricing policy, product positioning, and consumer analytics.

Comparison of Individual and Joint Modeling Results

The comparison among Tables 2, 3, and 4 reveals consistent but nuanced differences between the isolated models (the standalone multiple linear regression for *rating* and the standalone Negative Binomial model for *rating count*) and the joint modeling framework that estimates both outcomes simultaneously. In terms of coefficient magnitude and direction, the joint model preserves the same qualitative patterns observed in the individual regressions—negative effects for discounts,

Table 4. Detailed Modeling Parameters for the Joint Multiple Linear Regression and Negative Binomial Models

term	Linear_coef	Linear_p	NegBin_coef	NegBin_p
C(category)[T.Electronics]	-0.127	0.000	0.672	0.000
C(category)[T.Home&Kitchen]	-0.176	0.000	-0.961	0.000
C(category)[T.OfficeProducts]	0.102	0.061	-1.691	0.000
Intercept	4.195	0.000	9.707	0.000
z_disc_frac	-0.053	0.000	-0.152	0.001
z_log_price	0.047	0.000	-0.171	0.000

Modeling Performance

Model	AIC	BIC	R2_or_PseudoR2	NB_alpha
Linear (rating)	398.935	430.631	0.089	
NegBin (rating_count)	30211.894	30248.874		2.070

mixed category effects, and opposite price influences between rating and rating count—but introduces subtle refinements in the size and precision of several coefficients. These refinements stem from the joint estimation's ability to capture shared latent relationships between satisfaction (rating) and engagement (rating count), which isolated models treat independently.

In the category variables, the joint model maintains nearly identical coefficient directions but exhibits slight stabilization in their magnitudes and standard errors. For example, *Electronics* remains negatively associated with ratings (–0.127) but positively associated with rating counts (0.672), while *Home & Kitchen* continues to show consistent negative effects across both outcomes (–0.176 and –0.961, respectively). The *Office Products* category, previously marginally significant in the linear model for ratings ($p = 0.061$), remains weakly positive in the joint structure while retaining a strong negative association with rating counts (–1.691, $p < 0.001$). These patterns suggest that joint modeling improves parameter stability, particularly for smaller subgroups or marginal effects, by leveraging the shared information structure between the two dependent variables.

The continuous predictors—price and discount—illustrate the clearest practical distinction between isolated and joint estimation. Across both modeling approaches, higher prices are associated with higher ratings but fewer reviews, while larger discounts decrease both ratings and engagement. However, the joint model slightly moderates these relationships, reflecting an adjustment for cross-outcome dependencies. Specifically, the coefficients for z_log_price (0.047 for rating; –0.171 for rating count) and z_disc_frac (–0.053 for rating; –0.152 for rating count) retain their direction but exhibit improved consistency and smaller residual variance compared to their isolated counterparts. This stabilization indicates that by simultaneously accounting for how satisfaction and engagement interact, the joint model provides a more balanced and interpretable representation of consumer response to pricing signals.

From a practical and managerial standpoint, these differences carry meaningful implications. The isolated models independently explain customer satisfaction and engagement, offering clear but separate insights for pricing and marketing strategy. In contrast, the joint model delivers a holistic behavioral perspective, enabling decision-makers to assess how pricing actions may simultaneously influence both perceived quality (rating) and visibility (rating counts). For instance, a price increase might enhance perceived quality but reduce engagement; the joint framework helps quantify this trade-off within a unified system. This integrated perspective is particularly valuable for dynamic pricing and recommendation algorithms, which must optimize both sales conversion and reputation metrics concurrently.

In summary, while the isolated models effectively identify key determinants of rating and engagement, the joint model enhances analytical robustness and interpretive depth by acknowledging the inherent connection between how consumers *evaluate* and *interact with* products. Methodologically, it provides a more efficient use of information, and practically, it offers a decision framework that better aligns with the multifaceted objectives of e-commerce platforms—balancing satisfaction, visibility, and sustainable brand value.

CONCLUSION

This study investigated how pricing and discount structures influence both customer satisfaction (as reflected in product ratings) and customer engagement (as measured by rating counts) in the e-commerce environment. Using a publicly available Kaggle dataset that included key variables such as *actual price*, *discount percentage*, *rating*, *rating count*, and *product category*, three models were developed: a multiple linear regression model for ratings, a Negative Binomial model for rating counts, and a joint model integrating both outcomes. The descriptive statistics demonstrated substantial variation in price levels, discount magnitudes, and engagement intensity across categories such as *Electronics*, *Computers & Accessories*, and *Home & Kitchen*.

The modeling results consistently revealed that pricing signals and discount magnitudes are critical determinants of online consumer behavior. Higher prices were positively associated with higher product ratings but negatively related to rating counts, indicating that premium products are perceived as higher quality yet attract fewer reviews. Conversely, larger discounts were negatively associated with both ratings and engagement, implying that excessive markdowns may erode perceived quality and reduce post-purchase participation. Category effects further emphasized the heterogeneity of consumer responses: *Electronics* products attracted significantly more reviews but slightly lower ratings, while *Home & Kitchen* and *Office Products* tended to receive fewer and lower ratings overall. The joint model confirmed that modeling satisfaction and engagement together yield more stable, interpretable results by accounting for shared behavioral mechanisms. Collectively, these findings highlight that consumer evaluation and engagement are interdependent processes influenced by both economic signals and category context, offering quantitative evidence of the price–perception–popularity relationship in digital marketplaces.

While the present study provides a strong empirical foundation, several avenues remain open for methodological and substantive enhancement. First, future research could expand the model by incorporating a broader set of predictors, such as brand reputation, seller rating, shipping speed, or visual presentation quality—factors known to influence consumer trust and purchase intent. Integrating temporal variables (e.g., seasonal price fluctuations, promotion timing, or review trends over time) would also allow dynamic modeling of consumer behavior. Second, the dataset could be extended to include additional product categories or larger sample sizes, improving generalizability across market sectors. Incorporating interaction terms between price and category may reveal whether pricing sensitivity differs across product types or brand tiers.

From an analytical standpoint, future studies might explore machine learning or hierarchical modeling frameworks that capture nonlinear relationships and category-level dependencies. For example, random-effects models or mixed neural networks could better represent variations across sellers or platforms. Additionally, employing textual analysis of reviews could enrich the interpretation of satisfaction drivers by linking linguistic sentiment features with quantitative ratings. Finally, for policy and managerial implications, e-commerce platforms should consider designing integrated pricing and feedback systems that optimize both short-term sales and long-term brand reputation, guided by predictive models that jointly evaluate satisfaction and engagement outcomes.

In conclusion, while the current study demonstrates the clear statistical relationship between pricing, discounts, and consumer responses, expanding the analytical scope and data richness will enhance both predictive accuracy and strategic relevance, enabling more precise, data-driven decision-making in the ever-evolving e-commerce ecosystem.

ACKNOWLEDGEMENT

The authors would like to express sincere gratitude to Kaggle for providing open-access e-commerce datasets that made this research possible. The availability of high-quality, publicly shared data has enabled rigorous quantitative analysis and fostered deeper insights into consumer behavior and pricing dynamics in online marketplaces.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Sharma R, Srivastva S, Fatima S. E-commerce and digital transformation: Trends, challenges, and implications. *Int J Multidiscip Res (IJFMR)*. 2023; 5: 1–9. https://doi.org/10.36948/ijfmr.2023.v05i05.7128
2. Zhang Q. E-commerce in rural areas, financial literacy, and elderly consumers. 2025. https://doi.org/10.1016/j.iref.2025.104533
3. Jusda Global. Business opportunities and challenges in e-commerce logistics. 2025.
4. Business Standard. Boosting growth: Integrate MSMEs and e-commerce to surge GDP globally. Business Standard. 2025.
5. Abrar K, Irfan S, Malik F. Impact of online reviews on consumer trust, perceived value, and purchase behavior in Pakistan's e-commerce market. *Journal of Business and Management Research*. 2024; 3 (3): 1009–1030.
6. International Trade Administration / Shopify. Global e-commerce statistics: Trends to guide your store. 2024.
7. The Devastator. E-Commerce Sales Dataset. 2024.
8. Khalikussabir K, Waris A, Wahono B. Perceived price, eTrust, eSatisfaction, and online repurchase intention. *JEMA: Jurnal Ilmiah Bidang Akuntansi dan Manajemen*.
9. Fratama F, Pardede R, Samara A, Susanti M. The influence of trust and perceived usefulness on purchase intention with price mediation. *International Journal of Economics (IJEC)*. 2023; 2 (2): 833–840. https://doi.org/10.55299/ijec.v2i2.660
10. Lu. Online shopping consumer perception. PeerJ Computer Science. 2024.
11. UNCTAD. Making e-commerce and the digital economy work for all. United Nations Conference on Trade and Development. 2024.
12. Zhao H, Sadiq MW, Larsen T. Impact of pricing and product information on consumer buying behavior. *Frontiers in Psychology.* 2021. https://doi.org/10.3389/fpsyg.2021.720151