

Mathematical Models in Vestiaire: A Statistical Investigation into the Variables that Drive Resale Outcomes

Maya Chou¹, Jared Cross¹

Received May 15, 2026

Accepted September 1, 2026

Electronic access October 15, 2026

This study investigates how one can predict how likely an item is to sell on Vestiaire Collective, a luxury clothing resale platform. Five statistical models were built and compared in R on a sample of 27,472 listings: linear regression, logistic regression, decision tree, random forest, and mixed effects. Four predictor variables were used: like count, product condition, brand name, and usual shipping time for the given seller. The models were evaluated using Root Mean Squared Error (RMSE). The mixed effects model performed best (RMSE = 0.446). Like count showed the strongest association with sale outcome: listings with zero likes had an observed sale probability of 11%, compared to 82% for listings with 50 or more likes. Product condition also had a statistically significant, though smaller, contribution to sale probability, with effects ranging from about 5 to 12 percentage points. These findings show that like count is strongly associated with sale outcomes on resale platforms, suggesting a meaningful role for social engagement signals in shaping buyer behavior.

Keywords: Mathematics; Statistics; resale prediction; social proof; Vestiaire Collective

Introduction

The market for secondhand and pre-owned luxury goods has grown substantially in recent years¹, with platforms such as Depop, TheRealReal, and Vestiaire Collective connecting individual sellers with a global pool of buyers. This reflects a broader trend: the global secondhand apparel market exceeded \$256 billion in 2025 and continues to expand rapidly². This growth does not necessarily reflect a shift toward more sustainable consumption, however. Secondhand fashion consumption has been found to correlate positively with new clothing purchases rather than displacing them³, and resellers themselves cite financial gain and convenience, more than sustainability, as their primary motivations². Consumer priorities also vary by market segment: luxury secondhand buyers value brand prestige and authenticity differently than the majority of buyers⁴, and often return to resale platforms for access to items unavailable through official brand channels⁵. Two reviews of secondhand consumption research each note a gap: most studies treat secondhand shopping as one general behavior rather than examining how specific platform features shape it^{6,7}. Separately, other research has found that word-of-mouth and price are strong predictors of consumers' intent to buy secondhand clothing⁸. Brand name may function similarly to seller reputation systems studied in other online marketplaces, which have been shown to reduce buyer uncertainty and sup-

port price premiums for trusted sellers^{9–11}. Despite this abundance of research on consumer motivations, little quantitative research has evaluated which listing-level factors predict whether an individual item sells.

This project also examines social engagement signals that may relate to a sale. Prior work suggests visible engagement, such as like count, influences consumer decisions by signalling credibility and desirability¹², an effect documented across social media shopping¹³ and short-form video commerce¹⁴. Online reviews function similarly, though inconsistent reviews have been shown to reduce purchase intention, particularly when purchasing for others¹⁵. This paper reports the construction and evaluation of five statistical models and the identification of like count as the dominant predictor of sale outcome on Vestiaire Collective.

Methods

Data Source and Collection

The dataset was obtained from Kaggle, a public data science platform¹⁶. It contains listing-level records from Vestiaire Collective, including product condition, brand name, like count, shipping time, price, and whether or not the item was sold. The data was publicly available and fully anonymised, containing no personally identifiable information about buyers or sellers, so no consent was required for its use. To en-

¹ Saint Ann's School, USA

sure balanced class representation for model training, an equal number of sold and unsold listings ($n = 13,736$ each) were randomly sampled from the original data, spanning 2,512 distinct brands after merging duplicate brand names (e.g., “celine” and “Céline”). Data collection occurred at a single point in time; the exact scrape date was not recorded in the dataset’s metadata. An item was coded as sold (1) if the listing’s status field indicated it had been sold, and unsold (0) otherwise. Because this reflects a single snapshot rather than continuous tracking over time, we cannot fully distinguish confirmed non-sales from listings that had simply not yet sold at the time of collection; this limitation is discussed further in the Limitations section.

Data Cleaning and Feature Engineering

There were several cleaning steps to be done before using this data in statistical models. For instance, brand names that are known to be the same were merged into one (e.g. “celine” and “Céline”; “tiffany” and “Tiffany & Co.”). In total, 25 cleaned brand name categories were created by merging two or more raw name variants (e.g., 13 raw Nike variants and 12 Adidas variants), out of 2,471 distinct cleaned brands overall. Raw like counts were converted into categorical brackets (0; 1–5; 6–10; 11–15; 16–20; 21–30; 31–40; 41–50; 50+) to allow for non-linear effects through categorical encoding and to reduce noise from extreme values. To evaluate this choice, a logistic regression model using continuous like count was compared to the bracketed version by measuring how well each model fit the data. The bracketed model fit the data noticeably better, supporting the use of categorical brackets. Whether or not the item sold was turned into binary indicators (1 or 0). Of the 27,472 listings, 3,759 (13.68%) were missing shipping time data and were excluded, leaving 23,713 listings for modeling. Excluded listings had a modestly higher like count and price, but substantially lower sale rate (34.6% vs 52.4%), suggesting that missingness was not random with respect to outcome.

Four predictor variables were selected for final modelling based on theoretical relevance and performance across several tested variable combinations (Table 1). This comparison was conducted within the same cross-validation procedure used to report final model performance, rather than on the full dataset beforehand, to avoid inflating reported performance. Price was deliberately excluded from the four final predictor variables. Price was highly right-skewed, with a small number of listings, such as luxury watches, reaching values far higher than most of the dataset, risking disproportionate influence on coefficient estimates in the models. A supplementary check (see Discussion) was used to confirm that this exclusion does not leave an unaddressed confounding variable in the reported results.

Statistical Models

Five models were constructed and evaluated. All were implemented in R¹⁷.

Linear regression models sale probability as a weighted sum of predictor variables. It is the most simple and interpretable approach but carries a fundamental limitation for binary outcomes: it can produce predictions outside the 0–1 range, which is impossible for probabilities. It was included as a baseline and for coefficient interpretability.

Logistic regression works similar to Linear regression, but addresses many of the limitations by converting the probabilities to a log-odds scale. It also better captures differences near the extremes of the probability range (e.g. the difference between 0.95 and 0.99 is not substantial in percentages for linear models but in log-odds they are 2.9 and 4.6).

A decision tree lets the model choose which variables it thinks are most important, instead of the user manually using intuition. It requires no assumptions about the functional form of predictor-outcome relationships. However, excessive splitting can lead to overfitting. The decision tree was fit with a maximum depth of 3 and no complexity-based trimming ($cp = 0$).

A random forest builds many decision trees and averages their predictions. This typically reduces overfitting relative to a single tree. In this study, however, the random forest had the worst performance. One possible explanation, discussed further below, is that its random subsampling of variables at each split limits how consistently it can draw on like count, the dominant predictor. The random forest was fit using 500 trees (default) with 4 variables considered at each split ($mtry = 4$).

A mixed effects model separates predictors into fixed effects — consistent predictors estimated across the full dataset (condition, like bracket) — and random effects, variables that can take on many different values, which can be difficult for the other models to account for (like brand name)¹⁸. Estimates for rare brands are given less weight, preventing overfitting. Formally, the model was specified as a binomial mixed-effects logistic regression with a logit link: `sold ~ like_bracket + usually_ships_within + product_condition + (1 | brandname_clean)`, fit using the `glmer()` function from the `lme4` package. Brand (2,512 levels) was treated as a random intercept, letting brands with fewer listings shrink toward the overall average rather than overfit. The model converged without warnings, and the random intercept’s variance (0.233, SD = 0.482) showed meaningful differences in baseline sale likelihood across brands even after accounting for like count, condition, and shipping time — supporting the use of a random-effects structure for brand rather than treating it as a fixed effect.

Variable	Description	Rationale
Like Bracket	Likes grouped: 0, 1–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 50+	Direct measure of buyer interest
Product Condition	Fair, Good, Very Good, Never Worn, Never Worn With Tag	Signal of quality
Brand Name	Cleaned to merge duplicates (e.g. celine / Céline)	Many buyers are attracted to specific brand names
Ships Within	Seller dispatch window: 1–2, 3–5, 6–7, or 7+ days	Shows the seller's credibility and reliability

Table 1 Predictor variables used in final models.

Evaluation

All models were evaluated using Root Mean Squared Error (RMSE), a measure of how different the predicted outcome was from the actual outcome, computed via 20-fold cross-validation. This means that R split the dataset randomly into 20 groups (folds). For each of the 20 iterations, the model was trained on 19 of the folds and tested on the remaining held-out fold, with each fold serving as the test set exactly once. RMSE was then averaged across all 20 iterations. A lower RMSE value indicates better performance. RMSE was used as the primary metric because it works consistently across all five models, including linear regression, which doesn't produce classification probabilities the same way. On a binary outcome, RMSE is also equivalent to the Brier score, a standard measure of how accurate predicted probabilities are. RMSE was used as the primary method for comparing models, since it applies consistently across linear and non-linear approaches. To more fully characterize classification performance, AUC, accuracy, sensitivity, specificity, precision, recall, and a calibration check were computed for the mixed effects and logistic regression models (see Results).

Results

Model Performance

Table 2 presents RMSE values for all five models. The mixed effects model achieved the lowest RMSE (0.446), so it was the best-performing model. Random forest performed worst (RMSE = 0.462).

To check whether these differences in performance were meaningful, paired t-tests were used to compare cross-validated RMSE across all five models. Logistic regression performed significantly better than the decision tree and random forest (both $p < .001$), and also outperformed linear regression, though that difference was very small ($p = .006$). The mixed effects model outperformed logistic regression as well, with a significantly lower average RMSE ($p < .001$),

confirming it as the strongest-performing model among those tested.

Classification Performance

The mixed effects model achieved AUC=0.757, accuracy=69.6%, sensitivity=72.5%, specificity=66.2%, precision=70.8%, and recall=72.5% on a held-out test split (logistic regression: AUC=0.749, accuracy=68.7%). A 10-bin calibration check showed close agreement between predicted and observed sale rates (e.g. 83.2% predicted vs 83.3% observed), indicating reliable probabilities.

Like Count as the Dominant Predictor

Like count was by far the strongest predictor of sale outcome. Table 3 shows predicted sale probabilities by like bracket, derived from the linear regression model. Linear regression coefficients were used here because they translate directly into percentage-point changes in sale probability, which are more intuitive to communicate than the mixed effects model's log-odds coefficients. The same overall pattern held in the mixed effects model: like-count coefficients increased sharply at low brackets and leveled off at higher ones, confirming that this finding is not an artifact of the choice of model.

The most striking result is the jump from zero to one-to-five likes: a gain of 32.7 percentage points. This suggests that early engagement has a disproportionately large effect on predicted sale probability. The relationship flattens substantially beyond 20 likes, with diminishing returns. The total range across all brackets is 71 percentage points, larger than the effect of any other variable in the model. One possible explanation, not directly tested here, is that buyers perceive items with more likes as more desirable or at greater risk of selling soon, creating an incentive to act quickly. This remains a hypothesis rather than a confirmed mechanism. This kind of herding behavior, where visible signals of others' interest shift individual purchase decisions, has been documented across a range of online marketplaces^{19–23}, and related work has shown that scarcity and

Model	RMSE	Notes
Linear Regression	0.449	Baseline; interpretable but not suited to probability outcomes since it often predicts more than 100% chance of selling
Logistic Regression	0.449	Better suited to binary outcomes; slightly outperformed linear regression (statistically significant, $p = .006$), despite similar RMSE
Decision Tree	0.451	Slightly higher than logistic
Random Forest	0.462	Worst performer; likely limited by inconsistent access to like count across trees (see Discussion)
Mixed Effects	0.446	Best performer; significantly outperformed logistic regression

Table 2 Model RMSE values (20-fold cross-validation). Lower RMSE indicates better performance.

Like Bracket	Sale Probability	Confidence Interval
0	11%	N/A
1–5	43%	[30.6, 34.4]
6–10	65%	[51.8, 56.2]
11–15	70%	[57.2, 62.2]
16–20	76%	[62.6, 68.3]
21–30	78%	[64.0, 69.7]
31–40	82%	[67.9, 75.2]
41–50	80%	[64.9, 74.0]
50+	82%	[66.9, 75.3]

Table 3 Sale probability assumes Fair condition, the baseline brand category, and 1–2 day shipping (the model’s reference levels). The 95% CI reflects uncertainty in the like-count effect itself, in percentage points relative to 0 likes.

urgency cues can similarly accelerate purchase decisions^{24,25}. Because the dataset reflects a single snapshot in time rather than continuous tracking, this finding should be interpreted as an association observed at time of scrape, rather than as evidence that early likes cause or predict a future sale.

Product Condition

Product condition surprisingly contributed modest effects relative to like count. Linear regression coefficients (relative to Fair condition as the baseline) are reported in Table 4.

All four non-baseline conditions showed a statistically significant positive effect relative to Fair condition (all $p < .05$). The effect size increased roughly in order of quality, with Good showing the smallest effect (+4.9 percentage points) and Very Good and Never Worn With Tag showing the largest, nearly identical effects (+11.7 and +11.6 percentage points).

Discussion

The main finding of this study is the major role of like count in predicting sale outcomes. This is consistent with literature

on social proof in consumer behaviour: prior engagement signals credibility and desirability to subsequent viewers¹². As a check on whether the observed like-count effect could instead be attributable to price, a logistic regression model was fit with log-price added. Like-count coefficients increased across every bracket (e.g., the 50+ likes coefficient rose from 2.98 to 4.31 in log odds), indicating price does not confound the central finding. Product category and seller country were available in the dataset but excluded to keep model concise.

The nonlinear shape of like-probability echoes the research on the role of zero-reviews versus any-review effects in online ratings systems. The steep initial gain (0 to 1–5 likes: +32 percentage points) and then diminishing returns point to the fact that a listing with any positive engagement is perceived extremely differently from one with none.

As proposed earlier, random forest’s weaker performance may stem from its random subsampling of variables at each split, which can limit the consistency of individual trees. To explore this, each predictor’s contribution to the random forest’s accuracy was examined. Like count stood out clearly as the most important variable, contributing far more than product condition, brand, or shipping time, cohering with the proposed explanation. As an additional check, the model’s error rate on unseen data was consistent with its cross-validated performance in Table 2, confirming the result is stable rather than specific to one data split.

Several limitations of this study should be noted. First, one should not assume a causation relationship between like count and sale. Items that would have sold for other reasons (brand, price, timing) may also accumulate likes, introducing confounding variables. Separating causation from correlation would require experimental data, such as a randomised promotion study. This study assessed the magnitude of potential confounding in two ways: the price-adjusted check described above (see Discussion) showed that the like-count effect did not weaken when price was added to the model, and the primary analysis of like count and product condition (Tables 3 and 4) controlled for brand and shipping time directly, rather than considering like count in isolation. To-

Product Condition	Coefficient vs. Fair	95% CI	p-value
Fair	0.00 (intercept)	Baseline	—
Good	+0.049	[.006, 0.091]	.024
Very Good	+0.117	[0.077, 0.157]	< .001
Never Worn	+0.090	[0.049, 0.131]	< .001
Never Worn With Tag	+0.116	[0.075, 0.158]	< .001

Table 4 All coefficients represent percentage-point change in sale probability relative to Fair condition, holding like count, brand, and shipping time constant.

gether, these checks suggest the like-count association is not simply a byproduct of the other listing characteristics available in this dataset, though other confounding factors, such as timing or seasonality, cannot be ruled out. The final dataset contained no missing values in variables (product condition, like count, brand name). Missingness in the original unsampled dataset was not separately assessed.

Multicollinearity among predictors was also assessed using variance inflation factors (VIF). All predictors showed low VIF values, indicating minimal correlation among like count, product condition, brand, shipping time, and price, and supporting the reliability of each variable's individual coefficient estimates.

Second, the dataset reflects a snapshot in time rather than continuous tracking. As a result, listings coded as unsold may include both confirmed non-sales and items that were simply still active and had not yet sold at the time of collection. We cannot distinguish between these cases with the available data. Also, variables such as year and seasonal demand were not accounted for. These findings are based on a single platform, Vestiaire Collective, and may not generalize directly to other resale marketplaces. Vestiaire specializes in authenticated luxury goods and has a like/engagement feature that may function differently than analogous features on other platforms (e.g., Depop, Poshmark, TheRealReal), which vary in audience, price range, and the visibility of engagement signals to buyers. The strength of the like-count association observed here should therefore be interpreted as specific to this platform and product category until tested elsewhere.

Conclusion

This study applied five statistical models to predict sale outcomes for listings on Vestiaire Collective. A mixed effects model had the best performance (RMSE = 0.446). Like count showed the strongest association with sale outcome, with observed sale rates rising from 11% at zero likes to 82% at 50 or more, and a 32 percentage point gain between zero and one-to-five likes. Product condition had a smaller but statistically significant effect on sale probability compared to like count. These findings contribute to a broader understanding of how

social engagement signals translate into economic outcomes on clothing resale sites.

Acknowledgments

I would like to thank my mentor, also my Advanced Statistics teacher, a graduate in Statistics, for guidance on mixed effects modelling and research design, and for feedback throughout the project.

References

- 1 ThredUp, *2024 Annual Resale Report*, 2024.
- 2 J. Herman, J. Kim-Vick and J. Hyun, *The Inner Drive: Unpacking the Motivations for Consumer Participation as Sellers in Apparel Resale*, 2025, 10.3390/businesses5040053.
- 3 M. P. Mizrahi and O. Sharon, *Secondhand Fashion Consumers Exhibit Fast Fashion Behaviors Despite Sustainability Narratives*, 2025, 10.1038/s41598-025-19089-1.
- 4 H. M. R. ul Hasan, C. Lang and S. Xia, *Investigating Consumer Values of Secondhand Fashion Consumption in the Mass Market vs. Luxury Market: A Text-Mining Approach*, 2022, 10.3390/su15010254.
- 5 G. Murtas and G. Pedeliento, *Investigating the Customer Journey in Second-Hand Fashion Platforms: Implications for Luxury Brand Management*, 2024, 10.1002/cb.2442.
- 6 F. Evans, L. Grimmer and M. Grimmer, *Consumer Orientations of Secondhand Fashion Shoppers: The Role of Shopping Frequency and Store Type*, 2022, 10.1016/j.jretconser.2022.102991.
- 7 F. G. Gilal, A. R. Shaikh, Z. Yang, R. G. Gilal and N. G. Gilal, *Secondhand Consumption: A Systematic Literature Review and Future Research Agenda*, 2024, 10.1111/ijcs.13059.
- 8 D. T. Cuong, *Examining How Factors Consumers' Buying Intention of Secondhand Clothes via Theory of Planned Behavior and Stimulus Organism Response Model*, 2024, 10.1016/j.joitmc.2024.100393.
- 9 R. Jiao, W. Przepiorka and V. Buskens, *Reputation Effects in Peer-to-Peer Online Markets: A Meta-Analysis*, 2021, 10.1016/j.ssresearch.2020.102522.
- 10 S. Ba and P. A. Pavlou, *Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior*, 2002, 10.2307/4132332.
- 11 S. S. Standifird, *Reputation and E-Commerce: eBay Auctions and the Asymmetrical Impact of Positive and Negative Ratings*, 2001, 10.1016/S0149-2063(01)00092-7.
- 12 R. B. Cialdini, *Influence: The Psychology of Persuasion*, 2001.
- 13 Y. Y. Talib and R. Saat, *Social Proof in Social Media Shopping: An Experimental Design Research*, 2017, 10.1051/shsconf/20173402005.

-
- 14 W. Huang, X. Wang, Q. Zhang, J. Han and R. Zhang, *Beyond Likes and Comments: How Social Proof Influences Consumer Impulse Buying on Short-Form Video Platforms*, 2024, 10.1016/j.jretconser.2024.104199.
 - 15 J. Wang, Y. Liu, Z. Qiu, Z. Zhao, M. Wang and H. Lan, *Shopping for Others: How Does Inconsistency in Online Reviews Affect Purchase Intentions Differently?*, 2025, 10.3389/fpsyg.2025.1579545.
 - 16 Vestiaire Collective Listings Dataset, *Vestiaire Collective Listings Dataset*, Kaggle, 2026, <https://www.kaggle.com>, Accessed May 2026.
 - 17 R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2024.
 - 18 D. Bates, M. Maechler, B. Bolker and S. Walker, *lme4: Linear Mixed-Effects Models Using Eigen and S4*, 2015, 10.18637/jss.v067.i01.
 - 19 P. Luo, B. Wu, R. Law and Y. Xu, *Herding Behavior in Peer-to-Peer Trading Economy: The Moderating Role of Reviewer Photo and Name*, 2023, 10.1016/j.tmp.2022.101050.
 - 20 Y.-F. Chen, *Herd Behavior in Purchasing Books Online*, 2008, 10.1016/j.chb.2007.08.004.
 - 21 M. Ali and H. Amir, *Understanding Consumer Herding Behavior in Online Purchases and Its Implications for Online Retailers and Marketers*, 2024, 10.1016/j.eierap.2024.101356.
 - 22 M. Ali, H. Amir and A. Shamsi, *Consumer Herding Behavior in Online Buying: A Literature Review*, 2021, 10.30543/10-1 (2021)-30.
 - 23 S. Sunder, K. H. Kim and E. A. Yorkston, *What Drives Herding Behavior in Online Ratings? The Role of Rater Experience, Product Portfolio, and Diverging Opinions*, 2019, 10.1177/0022242919875688.
 - 24 F. Ali, D. H. Maqsood and Q. Janjua, *Psychological Triggers in Online Shopping: The Influence of Scarcity, Urgency, and Personalization on Consumer Buying Behavior*, 2025, 10.59075/cxyapm95.
 - 25 A. Wrabel, A. Kupfer and S. Zimmermann, *Being Informed or Getting the Product? How the Coexistence of Scarcity Cues and Online Consumer Reviews Affects Online Purchase Decisions*, 2022, 10.1007/s12599-022-00772-w.