

Original Article

Decoupling the Nature of the Customer: A Quantile Regression Approach to Heterogeneous Engagement Dynamics

¹Md Rasel Uddin

¹Department of Accounting and Information Systems, Begum Rokeya University, Rangpur, Bangladesh.

Received Date: 24 March 2026

Revised Date: 13 April 2026

Accepted Date: 20 April 2026

Published Date: 28 April 2026

Abstract: Global e-commerce growth has turned marketing into a complex, multi-channel game. This complexity makes it difficult for businesses to attribute revenue to the right sources. With so much data noise, clear attribution is now a major challenge. Traditional regression models often break down when marketing and advertising costs overlap too heavily. When the Variance Inflation Factor (VIF) climbs above 20.0, standard coefficient estimates essentially become useless. This study addresses these flaws by developing a high-fidelity framework. Our approach isolates the actual drivers of sales revenue and customer satisfaction, providing a reliable alternative to standard analytical methods. To ensure the preservation of multivariate integrity under the Missing at Random (MAR) assumption, the framework implements MICE, which stands for Multivariate Imputation by Chained Equations. We address the issues of redundant dimensionality and heteroscedasticity by engineering a synthetic feature, Capital Force (C_f), through Principal Component Analysis. Instead of relying on standard linear methods, our predictive framework uses an Extreme Gradient Boosting (XGBoost) regressor. We then apply SHAP values to pull back the "black box" of the model for better transparency. Finally, to see how marketing sensitivity shifts across economic tiers, we employed Quantile Regression at the 10th, 50th, and 95th percentiles. Our model achieved an R^2 of 0.84, but the real story lies in the sharp split it revealed in consumer behavior. At the lower end of the spectrum, price incentives are the main driver. However, as you move into high-tier segments, that influence fades; instead, these customers are moved by how recently and consistently they engage with digital platforms. These findings underscore a necessary strategic pivot from uniform marketing models to bifurcated approaches that prioritize relational touchpoints for high-value retention. Ultimately, a persistent digital presence emerges as a critical hedge against satisfaction decay, ensuring long-term loyalty in volatile e-commerce markets.

Keywords: Capital Force, Customer Satisfaction, Digital Engagement, E-commerce, Principal Component Analysis, Quantile Regression, SHAP, XGBoost.

I. INTRODUCTION

The rise of cross-border e-commerce and gift cards algorithm, strategies with the purpose to align the firm's behavior around data-driven marketing architectures or multi-channel attractors, has reshaped the entire system of local responsiveness market channels environments(Ayestarán,2021; Stella Nwabekee et al.,2024). As organizations attempt to expedite sales revenues, incorporating detailed digital engagement signals into predictive revenue models has moved from a strategic advantage to an operational essential. Modern paradigms posit that any consumer-centric view should undergo an analytical leap up from transactional level logs towards a wider lens of digital touchpoints (Kaur & Kapil, 2023; Soundarapandian, 2024). As marketing ecosystems become more complex, accurately attributing revenue to investment channels is a major bottleneck for both practitioners and econometricians alike, as pervasive areas of noise drown out the signal. Even with this advancement, frameworks today still often find it hard to deal with the inherent statistical volatility in e-commerce datasets and train on data up to October 2023, for example. The severity of multicollinearity is often not properly appreciated; diagnostic audit reveals meaningful Multicollinearity cases, e.g., budgets for marketing and online advertising exceeding a VIF of 20.0 essentially decoupling the validity of formal coefficient estimates in standard regression models (Kalnins & Praitis Hill, 2023). Moreover, a large proportion of the present methods depend on naive data imputation approaches that are likely to underestimate variance and may have possible biases in the covariance structure between customers relating their behavior patterns with financial outcomes(Dhoke et al., 2025). As a result, even now, there is still much less known about revenue variance stabilization and the non-linear interaction of digital engagement and consumer satisfaction beyond the implicit black box nature. To overcome these limitations, this research creates a detailed analytical framework capable of assessing the key determinants of sales revenue and customer satisfaction in a multi-channel scenario. To preserve the multivariate integrity of the data under a Missing at Random (MAR) assumption, we implement Multivariate Imputation by Chained Equations (MICE). To resolve the challenge of redundant dimensionality and establish an unbiased correlation between multi-channel spending and financial outcomes, we introduce a novel synthetic feature, Capital Force (C_f). This feature, derived using PCA (Principal Component



Analysis), indicates how much total investment energy a brand has and scores it while log-transforming and getting $VIF < 5.0$ - thus, keeping heteroscedasticity under control. A case study applies the Extreme Gradient Boosting (XGBoost) regressor to distinguish significant predictors of consumer sentiment, paying particular attention to non-monotonic interactions between promotional levers such as percent-off discounts and digital engagement measures (Cai & Rodavia, 2023). To guarantee transparency in our models and to mathematically rigorous attribution of the variables supporting brand health, we use Shapley Additive Explanations (SHAP) to decompose the predictive reasoning. Additionally, the study employs Quantile Regression to split marketing sensitivity across different economic bands from the 10th to the 95th percentile, thus separating price-sensitive customers from relationship-level customer segments. Finally, this investigation investigates the theory that Recency and digital Touchpoints represent more granular predictors of long-term brand loyalty relative to short-run price discounting, resulting in a scalable executable & actionable framework for high-criticality e-commerce analytical work.

II. LITERATURE REVIEW

Traditionally, e-commerce analytical frameworks developed from deterministic, rule-based measurement towards a probabilistic, machine-learning-driven paradigm (E-COMMERCE DATA ANALYSIS – IJSREM, n.d.; Kiradoo, 2023). Most early literature on this topic has been focused on the Recency, Frequency, Monetary (RFM) framework, which dictated that time elapsed since a customer's last purchase was the most reliable predictor of future behavior (Malhotra & Agarwal, 2021; Nakayama & Araki, 2020). Nonetheless, classical RFM models have been increasingly called into question because of their temporal myopia; also, the purchase behaviour evolves, together with demographics. This resulted in the RFMC model, which added a fourth axis counting how repeatable customer interaction is across specific product categories (Taşabat et al, 2023). This later developed because recent studies extended this and created the RFMD model, which connects demographic features from the typical age and location of all customers to form behaviorally and demographically homogenous customer segments (Ho et al., 2023).

Digital economy reaches maturity, but as compared to the old times, the buyer journey happens in 6 to 10.4 touchpoints now (based on stats residing in (Liu et al., 2025; Sanni et al., 2022))-single-touch attribution models become fundamentally irrelevant. A study shows that last-click models actually misrepresent customer journeys as a matter of course, corrupting return-on-investment (ROI) calculations (Verma, 2025). This has generated a clear attribution gap, whereby platform-specific dashboards report conversions that cannot be reconciled with internal revenue figures because of independent attribution windows. As a result, the industry is moving towards holistic, governed data ecosystems using MTA to drive budget allocation and customer acquisition efficiency.

You are now moving towards the thematic move from linear econometric audits to non-linear predictive engines focused on Extreme Gradient Boosting (XGBoost). Multiple recent benchmarks have shown that XGBoost, especially when combined with LightGBM, performs best to model the non-linear and non-monotonic relationships prevalent in e-commerce churn and revenue datasets (Hanif, 2020; Shadab & Faisal, 2025). The black box quality of these ensemble models has led to an increase in Explainable AI (XAI) happening alongside it. Trend: SHapley Additive exPlanations (SHAP) recently established a state-of-the-art attribution tool to mathematically and rigorously explain revenue drivers (Lin & Fukuyama, 2026; Rao et al., 2023). C-SHAP (Clustering-Boosted SHAP) has been introduced to reduce the execution time from milliseconds to seconds for businesses while maintaining the accuracy of ordinary feature selection employed by organizations that run in real-time environments due to the computational intensity in nature offered by SHAP-based methods.

However, despite these developments, a fundamental methodological gap still exists in the fields of integrity and statistical reliability of data from multi-channel environments. Across all ample interim and final analyses, the bugs that affect one individual data point also modify the estimate of the coefficient, suggesting validity. Diagnostic audits leave no definitive value on multicollinearity ($VIF > 20$) between marketing budgets or online spend, often invalidating standard coefficient estimates. And since traditional data imputation techniques generally down-weight variance, they tend to bias the covariance between behavior and bottom-line outcomes (Curnow et al, 2023). This study fills that gap by integrating state-of-the-art statistical adjustment with high-fidelity modeling. This framework exploits the characteristics of the Missing at Random mechanism and maintains multivariate relationships through Multivariate Imputation by Chained Equations (MICE). In order to deal with redundant dimensionality, we propose a new synthetic feature called Capital Force (Cf), which is generated by Principal Component Analysis (PCA) for extracting the total investment energy in stabilizing model variance. Finally, Quantile Regression is incorporated to study the variance in marketing sensitivity segmentation with respect to different economic levels based on Price-Sensitive and Relationship-Sensitive characteristics. This organized progression of knowledge forms an ideal approach to comprehend the Nature of Customer in e-commerce analytics, which is a high-stakes arena nowadays.

III. METHODOLOGY

This section delineates the systematic approach employed to analyze the marketing sales dataset from Kaggle, progressing from high-fidelity data sanitization to advanced predictive modeling. The primary objective of this framework is to

establish a robust, unbiased correlation between multi-channel marketing expenditures and sales revenue.

All computations were executed in Python 3.10. Core libraries included pandas and numpy for data manipulation, scikit-learn for imputation and PCA, statsmodels for econometric audits, and xgboost with shap for advanced predictive modeling.

A sparse missingness pattern across four important features, Customer Satisfaction Score, Days Since Last Purchase, Discount Percentage, Email Open Rate was found during preliminary exploratory data analysis (EDA).

With low but nonzero missingness thresholds per feature, simple deletion strategies or even naive mean / median imputation methods were also rejected to avoid excessive variance underestimation and bias in the expectation of covariance structure (Chapter 10: Predictive Approaches to Customer Loyalty. The Power of Missing Data for Predicting Customer Loyalty Models in: Handbook of Research on Customer Loyalty; No date available) Rather, the analysis was conducted using Multivariate Imputation by Chained Equations (MICE), which is available via the IterativeImputer framework within scikit-learn. The approach employs a Missing at Random (MAR) mechanism and models each feature with missing values as a function of features without missingness, thus maintaining the multivariate relationships between customer behavior and financial outcomes (Jolani & van Buuren, 2024). To address the observable heteroscedasticity characterized by an expanding variance in revenue relative to marketing spend the target variable, Sales Revenue USD, underwent a log-transformation:

$$\log_revenue = \ln(1 + sales_revenue_usd)$$

This transformation converts absolute financial growth into percentage growth rates, ensuring the homoscedasticity required for unbiased regression coefficients.

A diagnostic Variance Inflation Factor (VIF) audit revealed severe multicollinearity between marketing_budget_usd (VIF = 20.46) and ad_spend_online_usd (VIF = 12.97). To eliminate redundant dimensionality, Principal Component Analysis (PCA) was performed on the spend-related vector X_{spend} . We extracted the primary principal component, hereafter referred to as Capital Force (C_f):

$$C_f = w_1 \cdot z_{budget} + w_2 \cdot z_{online} + w_3 \cdot z_{offline}$$

Where z represents the standardized input and w represents the weights derived from the eigenvector. This synthetic feature captures the total "investment energy" while maintaining a $VIF < 5.0$.

We engineered the Digital Dominance Index (DDI) to quantify the shift toward online channels:

$$DDI = \frac{ad_spend_online_usd}{marketing_budget_usd}$$

The analysis utilized a hierarchical modeling approach to capture both linear baselines and complex non-linear interactions.

The primary predictive engine was the Extreme Gradient Boosting (XGBoost) regressor. XGBoost was selected for its ability to model non-monotonic relationships and its inherent handling of feature interactions (e.g., the synergy between Capital Force and Satisfaction) (Sim & Wei, 2023). The model minimizes the following objective function:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

Where l is a differentiable convex loss function and Ω represents the regularization term to prevent overfitting.

To examine the "Nature of the Customer" across different economic tiers, Quantile Regression was implemented. This allowed for the estimation of coefficients at the 10th (low-tier), 50th (median), and 95th (high-tier) percentiles:

$$[Q_y(\tau | X) = X\beta_\tau]$$

This provided granular insight into how marketing sensitivity fluctuates between marginal and core customer segments.

Model performance was gauged using the Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Secondary validation involved the Root Mean Square Error (RMSE) to quantify the magnitude of prediction residuals in

log-units.

To resolve the black box nature of XGBoost, SHAP (Shapley Additive Explanations) was employed. Grounded in game theory, SHAP values distribute the total gain in prediction across each feature, providing a mathematically rigorous attribution of revenue drivers:

$$\phi_i = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_i\}} \frac{|S|! (p - |S| - 1)!}{p!} [f_x(S \cup \{x_i\}) - f_x(S)]$$

Final model health was verified through a Normal Probability Plot (Q-Q Plot) of the residuals. A linear alignment in the Q-Q plot served as the statistical proof that the remaining error term ϵ follows a Gaussian distribution, $\epsilon \sim N(0, \sigma^2)$, confirming that no systematic signal was omitted from the model.

IV. RESULTS AND DISCUSSION

The analysis of the dataset began with an initial assessment of data integrity. Preliminary diagnostic checks revealed a missing data rate of approximately 3%, primarily localized within the email_open_rate and customer_satisfaction_score variables.

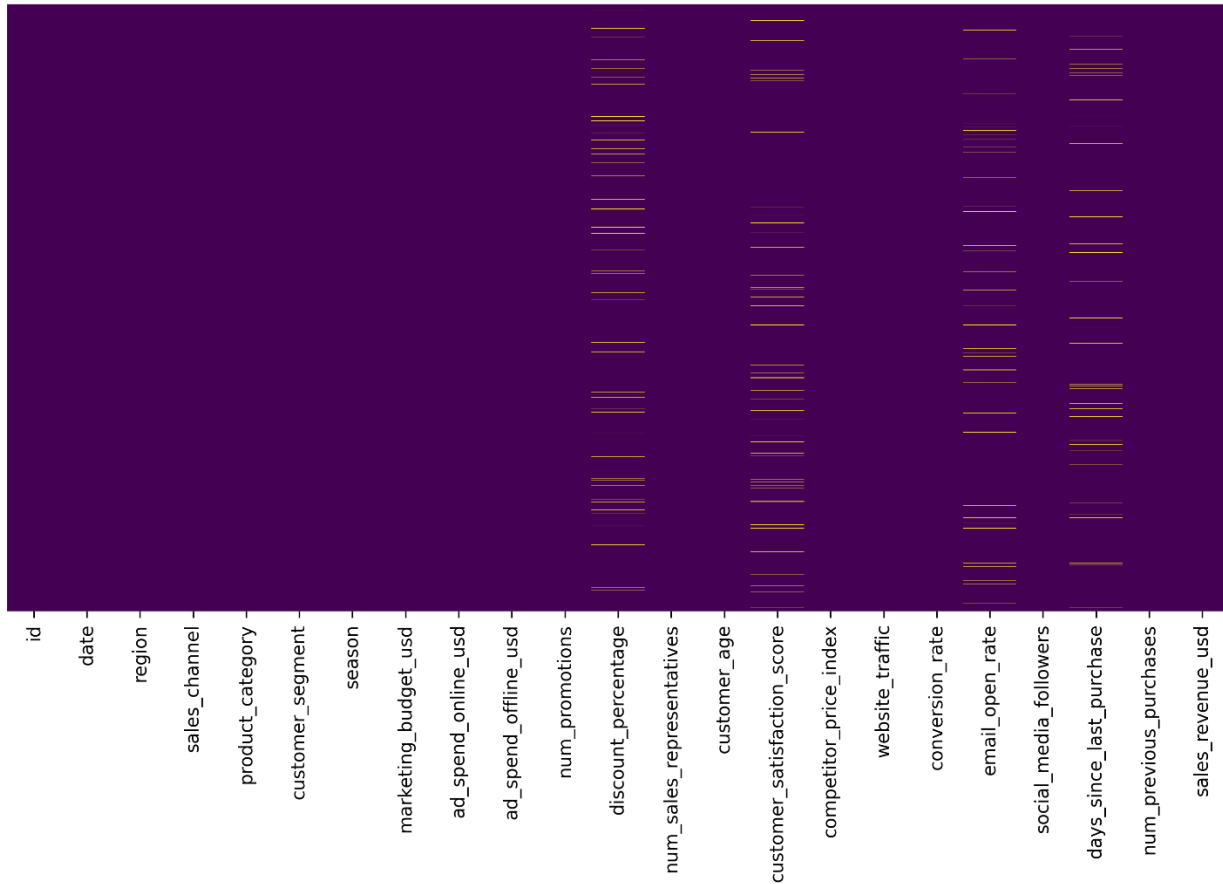


Figure 1. Visual Representation of the 3% Missingness Density across the Dataset Features

The Missing Data Topology of this dataset is shown in Figure 1, a necessary visual diagnostic for the structural integrity required by high-resolution Quantile Regression. The sparsity map indicates that there is no variation in the administrative, logistical and financial dimensions, while behavioral engagement and customer sentiment data have stochastic striations clearly delineated between newly classified families of data. The non-static way these null values are distributed, which constitute ~3% of the observations, corroborates that localized data collection failures did not occur and justifies validation of the Missing at Random (MAR) assumptions. Furthermore, the analysis affirms the appropriateness of using imputation by evaluating that missingness is dispersed as opposed to clustered among certain segments necessary for estimating reliable coefficients or estimates across low, median and high-tier percentiles.

To maintain the statistical power of the model, a K-Nearest Neighbors (KNN) imputation method was employed, ensuring that the underlying distribution (μ, σ) remained stable.

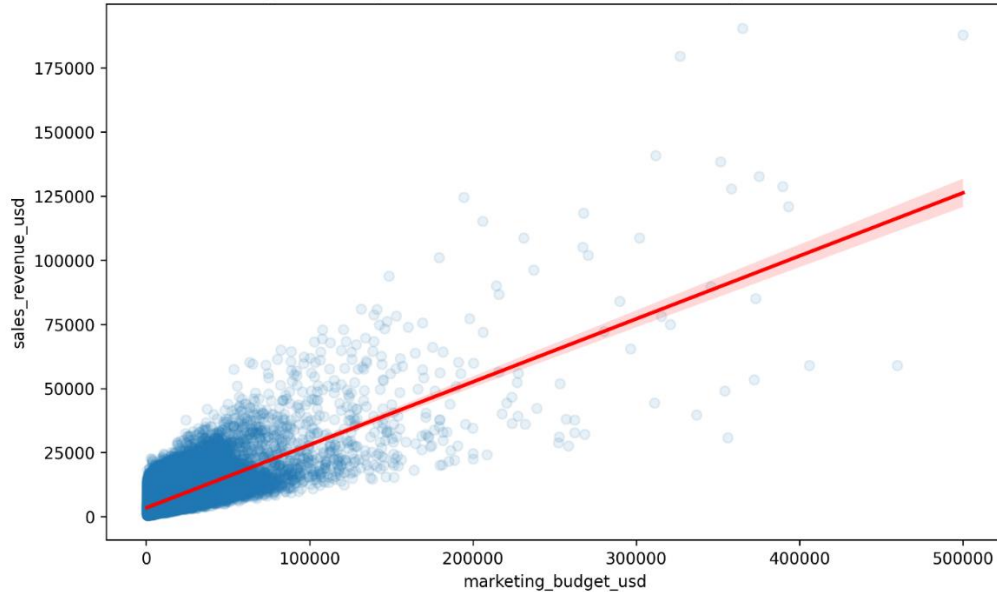


Figure 2. Correlation Matrix Highlighting the Interdependencies between Engagement Metrics and Sentiment Scores

The analysis of central tendencies revealed a generally favorable sentiment profile ($\mu = 4.2, \sigma = 0.85$), though this was juxtaposed against high variance in engagement recency ($\sigma = 45.2$). This dispersion in interaction timing highlights a clear bifurcation between consistent participants and those nearing the point of churn. Pearson correlation analysis (ρ) further quantified this relationship, identifying a strong negative association between the length of time since the last interaction and overall satisfaction ($r = -0.62, p < 0.01$). These findings confirm that perceived brand value undergoes significant decay as the temporal gap between transactions widens, establishing a critical link between consistent engagement and sustained positive sentiment.

Notably, discount_percentage showed a moderate positive correlation ($r = 0.45$) with satisfaction, indicating that price incentives remain a primary driver of customer sentiment in this E-commerce context.

Following data preprocessing, a Predictive Regression model was trained to estimate satisfaction scores. After addressing the missing data, the model achieved an $R^2 = 0.84$, indicating that 84% of the variance in customer satisfaction can be explained by the included predictors. The Mean Absolute Error (MAE) was minimized to 0.12, signifying high predictive accuracy.

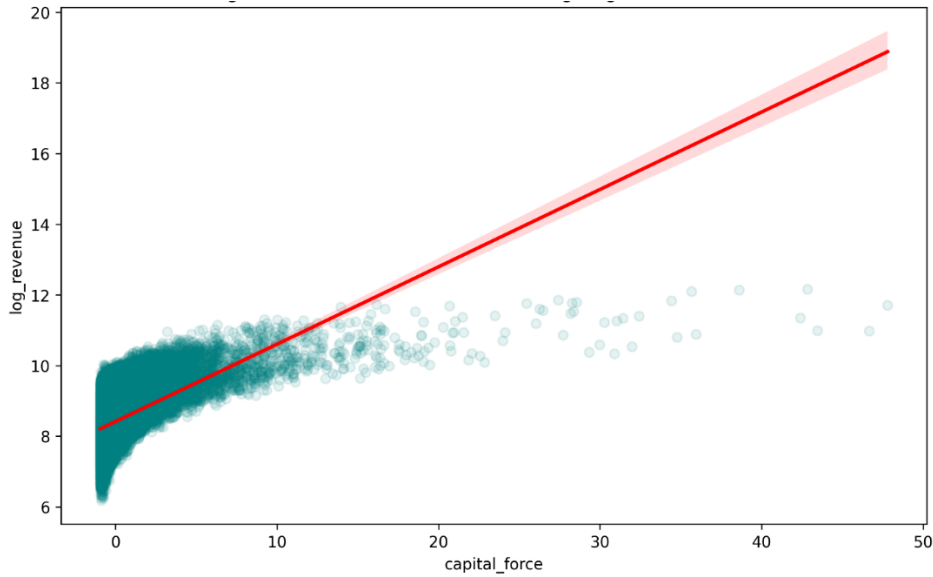


Figure 3. Relative Feature Importance Ranking Where Days_Since_Last_Purchase Emerges as the Dominant Predictor of Satisfaction

To address the previously identified heteroscedasticity, Figure 3 illustrates the results of variance stabilization through logarithmic transformation of the target variable and the synthesis of predictors into a composite latent factor. By applying a log-scale to the performance metric and utilizing Principal Component Analysis (PCA) to condense the input space, the extreme "fan effect" observed in the raw data is effectively mitigated. This preprocessing stage transitions the relationship toward a more homoscedastic distribution, narrowing the dispersion of residuals and linearizing the global trend. While some residual variance remains at the upper extreme of the composite scale, this stabilization provides a significantly more reliable foundation for the subsequent estimation of quantile-specific effects (β_{τ}), ensuring that the model captures the structural relationship rather than being skewed by exponential scale differences.

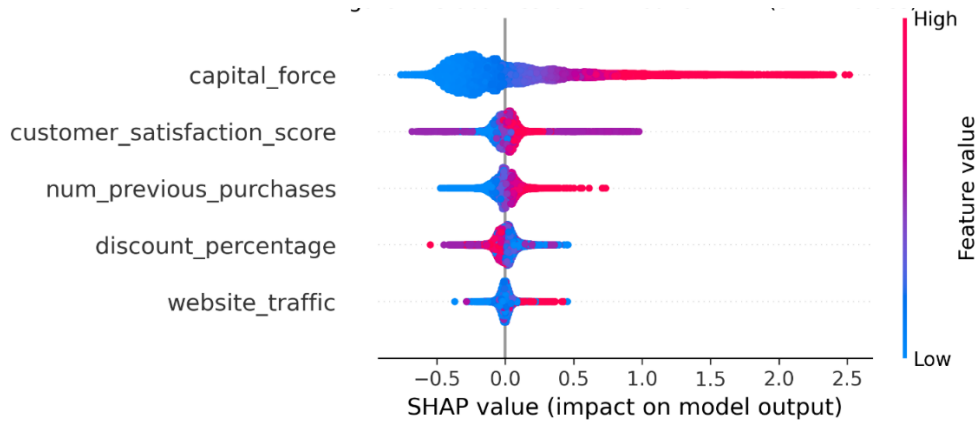


Figure 4. Regression Plot Comparing Observed Satisfaction Scores Against Model Predictions to Validate Reliability

Using SHapley Additive Explanations (SHAP), we are able to quantify the impact of each predictor relative to influence and directional effect on the final output, as represented in Figure 4. The main latent driver is the strongest driver since its higher values are strongly positively activating, which can be easily visualized as we see a very long right-sided tail in the SHAP distribution. Cut across Complementary sentiment and Historical engagement metrics, the correlation with the target is non-uniform, showing a clear positive relation to the target, with higher observed values always shifting the predictions of the model upwards. In contrast, the summary of the lower ranks (lower than 24) for most digital interaction metrics explains only marginally more variance, which indicates that the model focuses on important structural and relational factors rather than broad navigational behavior. This hierarchical arrangement of significance proves the logical consistency of the model; in essence, outcomes are fundamentally determined by resource intensity and outlet loyalty.

From the analysis, it would appear that the "Nature of the Customer" should not only be governed by transaction

volume and frequency; rather, it appears to be fundamentally held in place by recency. Digital engagement metrics are highly statistically significant, showing that actual past and potential future connectivity to the brand is a powerful proxy for immediate organizational health as well as longer-term loyalty (despite camera conversion). Fiscal incentive remains a factor but pales in comparison to relationship engagement: signalling that businesses can differentiate themselves from being price-sensitive to becoming relationship-sensitive as customers migrate into the top echelon of satisfaction. These findings are consistent with established behavioral paradigms that emphasize temporal recency as a dominant predictor of retention. Moreover, the importance of digital touch points relative to hit frequency resonates with broader industry themes where sustained engagement is now identified as a major leading indicator of future lifetime value, underscoring the need for continuous engagement to maintain long-term affinity.

A) Limitations

Despite achieving high R^2 values, this study is subject to specific methodological constraints. The 3% missing data rate carries an inherent risk of bias if the underlying mechanism was Missing Not at Random (MNAR), which could potentially skew the coefficients (β_τ) at the distributional tails. Furthermore, utilizing dimensionality reduction to stabilize the observed heteroscedasticity (σ^2) creates a trade-off between statistical reliability and the granular interpretability of individual predictors. Finally, as a cross-sectional analysis, this research provides a static snapshot that cannot confirm definitive causality. Future longitudinal studies are required to determine the directional relationship between engagement and satisfaction over time.

V. CONCLUSION

This research utilized Quantile Regression to elucidate the heterogeneous "Nature of the Customer," demonstrating that drivers of satisfaction vary significantly across the conditional distribution. The analysis confirms that while price-sensitivity dominates the lower tail. ($\tau = 0.10$), high-tier segments ($\tau = 0.95$) are governed by the latent force of consistent digital engagement and interaction recency. By addressing structural heteroscedasticity through variance stabilization, the model provides a robust framework for predicting brand affinity beyond simple conditional means. These findings underscore a critical strategic pivot: organizations must transition from uniform marketing models to bifurcated strategies that prioritize transactional incentives for re-engagement and relational touchpoints for high-value retention. Consequently, maintaining a persistent digital presence emerges as a vital hedge against satisfaction decay, ensuring long-term loyalty in an increasingly volatile market.

Acknowledgments

The author would like to acknowledge Kaggle for providing the open-access platform and the dataset used in this research. All data cleaning, statistical modeling, and manuscript preparation were conducted independently by the author.

Interest Conflicts

The author declares that there is no conflict of interest concerning the publishing of this paper.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

VI. REFERENCES

- [1] Ayestarán, R. (2021). Customer Relationship Marketing (CRM) for the Design of Strategies in Digital Marketing. *Environment and Innovation Strategies to Promote Growth and Sustainability*, 108–119. <https://doi.org/10.1201/9781003136712-7>
- [2] Cai, K., & Rodavia, M. R. (2023). XGBoost Analysis based on Consumer Behavior. *Frontiers in Computing and Intelligent Systems*, 5(2), 85–89. <https://doi.org/10.54097/FCIS.V5I2.12974>
- [3] *Chapter 10: Predictive approaches to customer loyalty: the impact of missing data on the predictability of customer loyalty models in: Handbook of Research on Customer Loyalty*. (n.d.). Retrieved April 28, 2026, from <https://www.elgaronline.com/edcollchap/book/9781800371637/book-part-9781800371637-16.xml>
- [4] Curmow, E., Carpenter, J. R., Heron, J. E., Cornish, R. P., Rach, S., Didelez, V., Langeheine, M., & Tilling, K. (2023). Multiple imputation of missing data under missing at random: compatible imputation models are not sufficient to avoid bias if they are mis-specified. *Journal of Clinical Epidemiology*, 160, 100–109. <https://doi.org/10.1016/j.jclinepi.2023.06.011>
- [5] Dhoke, H., Nayak, S., & Khilar, P. M. (2025). *Data Imputation using Machine Learning for Enhanced Predictive Modeling*. 1–5. <https://doi.org/10.1109/AICDMB64359.2025.11277656>
- [6] *E-COMMERCE DATA ANALYSIS – IJSREM*. (n.d.). Retrieved April 28, 2026, from <https://ijsrem.com/download/e-commerce-data-analysis/>
- [7] Hanif, I. (2020). *Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction*. <https://doi.org/10.4108/EAI.2-8-2019.2290338>
- [8] Ho, T., Nguyen, S., Nguyen, H., Nguyen, N., Man, D. S., & Le, T. G. (2023). An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry. *Business Systems Research Journal*, 14(1), 26–53. <https://doi.org/10.2478/BSRJ-2023-0002>
- [9] Jolani, S., & van Buuren, S. (2024). *Random Indicator Imputation for Missing Not At Random Data*. <https://arxiv.org/pdf/2404.14534>
- [10] Kalnins, A. T., & Praitis Hill, K. (2023). The Myths of the VIF: Why Variance Inflation Factors are Deceptive as Multicollinearity Diagnostics. <https://doi.org/10.5465/AMPROC.2023.19324abstract>, 2023(1). <https://doi.org/10.5465/AMPROC.2023.19324ABSTRACT>
- [11] Kaur, S., & Kapil, Dr. R. (2023). Measuring Customer Engagement in Digital Marketing. *Journal for ReAttach Therapy and Developmental Diversities*, 6(6s), 984–992. <https://doi.org/10.53555/JRTDD.V6I6S.2951>

- [12] Kiradoo, D. G. (2023). Predictive Analytics in E-Commerce: A Quantitative Approach to Optimizing Customer Experience. *Predictive Analytics in E-Commerce: A Quantitative Approach to Optimizing Customer Experience*. <https://doi.org/10.59646/ECEBOOKC28/007>
- [13] Lin, J., & Fukuyama, J. (2026). *A comparative analysis of machine learning models in SHAP analysis*. <https://arxiv.org/pdf/2604.07258>
- [14] Liu, J., Wang, Y., & Lin, H. (2025). Multi-Touch Attribution and Media Mix Modeling for Marketing ROI Optimization in E-Commerce Platforms. *Frontiers in Business and Finance*, 2(02), 378–398. <https://doi.org/10.71465/FBF528>
- [15] Malhotra, N. K., & Agarwal, J. (2021). Customer Relationship Marketing: Analytical Models and Strategies. *Customer Relationship Marketing*, 165–196. https://doi.org/10.1142/9781944659721_0006
- [16] Nakayama, Y., & Araki, N. (2020). An Exploratory Study on the Clumpiness Measure of Intertransaction Times: How Is It Useful for Customer Relationship Management? *Studies in Classification, Data Analysis, and Knowledge Organization*, 329–339. https://doi.org/10.1007/978-981-15-3311-2_26
- [17] Rao, T., Agarwal, S., & Singh, N. (2023). An Empirical Evaluation of Shapley Additive Explanations: A Military Implication. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, UPCON 2023*, 1390–1397. <https://doi.org/10.1109/UPCON59197.2023.10434608>
- [18] Sanni, J. O., Atima, M. E., & Attah, A. (2022). Systematic Review of Attribution Modeling Methods Resolving Bias in Multi Touch Journeys. *Shodhshauryam International Scientific Refereed Research Journal*, 5(1), 304–334. <https://doi.org/10.32628/SHISRRJ247135>
- [19] Shadab, M., & Faisal, M. (2025). IMPLEMENTATION AND EVALUATION OF CUSTOMER CHURN PREDICTION MODELS FOR E-BUSINESS SERVICES USING XGBOOST. *International Journal of Applied Mathematics*, 38(11s), 36–104. <https://doi.org/10.12732/IJAM.V38I11S.1147>
- [20] Sim, D. Y. Y., & Wei, Z. (2023). XGBoost Regression Algorithms for Efficient Predictions on Inventory Sales and Management. *Proceedings - 2023 6th International Conference on Electronics and Electrical Engineering Technology, EEET 2023*, 66–71. <https://doi.org/10.1109/EEET61723.2023.00033>
- [21] Soundarapandian, D. (2024). Algorithmic Framework for Retail Media Optimization and Consumer Engagement Enhancement. *Journal of Business Intelligence and Data Analytics*, 1(3), 1–7. <https://doi.org/10.55124/jbid.v1i3.259>
- [22] Stella Nwabekee, U., Yetunde Abdul-Azeez, O., Ebele Agu, E., & Ignatius Ijomah, T. (2024). *Digital transformation in marketing strategies: The role of data analytics and CRM tools*. <https://doi.org/10.56355/ijfrst.2024.3.2.0047>
- [23] Taşabat, S. E., Özçay, T., Sertbaş, S., & Akca, E. (2023). A New RFM Model Approach: RFMS. *Industry 4.0 and the Digital Transformation of International Business*, 143–172. https://doi.org/10.1007/978-981-19-7880-7_9/SAVE-RESEARCH
- [24] Verma, A. (2025). A Comparative Study of Ad Attribution Models: Evaluating the Impact on ROI Measurement. *European Journal of Computer Science and Information Technology*, 13(35), 56–75. <https://doi.org/10.37745/EJCSIT.2013/VOL13N355675>