

Understanding E-Commerce Sales Performance: Insights for Strategic Business Decision-Making

Most Mahmuda Akter¹, Abir Hasanath Antu², Sasha Shehara Liyanage³
^{1,2,3}Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

doi: <https://doi.org/10.37745/ejbir.2013/vol14n81430>

Published October 18, 2026

Citation: Akter M.M., Antu A.H., Liyanage S.S. (2026) Understanding E-Commerce Sales Performance: Insights for Strategic Business Decision-Making, *European Journal of Business and Innovation Research*, 14(8),14-30

Abstract: *This study examines sales performance in a synthetic e-commerce dataset and evaluates the business relevance of an order-level revenue model. The analysis covers 5,000 transactions with information on product category, region, quantity, price, discount, payment method, delivery time and customer rating. Descriptive statistics, grouped revenue summaries, visualisations and Pearson correlations are combined with an ordinary least squares regression model. Recorded total revenue is 5,109,775.74 monetary units, and average revenue per order is 1,021.96. Electronics accounts for the largest revenue share at 35.81%, while Beauty has the highest average order value at 1,059.28. Card payments represent 45.40% of transactions. Revenue has positive correlations with unit price and quantity, whereas delivery time and customer rating have little linear association with revenue. The recorded regression results indicate a test R^2 of 0.870 and mean absolute error of 219.76. However, displayed transactions suggest that revenue is derived from quantity, price and discount, limiting the model's independent predictive value. The study demonstrates how managers can interpret sales dashboards while distinguishing descriptive patterns, accounting relationships and causal claims. Its managerial recommendations are proposals for validation with real business data rather than evidence of proven commercial effects.*

Keywords: e-commerce, sales performance, business analytics, average order value, linear regression, synthetic data

INTRODUCTION

Retail transaction data can support business decisions when sales measures are interpreted in their commercial context. Dekimpe [1] examines the opportunities and challenges created by big data analytics in retailing, while Verhoef et al. [2] place data capabilities within the wider process of digital transformation. For e-commerce managers, total revenue, order counts and average order value answer different questions. A category may generate the highest revenue because it receives more orders, even when its average transaction value is lower. Consequently, a useful business intelligence approach combines these measures before drawing conclusions about product or regional performance.

This study adopts a business administration perspective, connecting sales analysis with product portfolio management, regional performance, payment options and promotion evaluation. It evaluates order-level revenue modelling against the commercial meaning and construction of the target variable. The contribution is an integrated interpretation of revenue concentration, transaction value and model validity rather than a new prediction algorithm.

The research objectives are to describe the distribution of sales across product categories and regions, assess the relationships between revenue and recorded order attributes, and evaluate whether the existing revenue model provides useful business information. Three questions guide the study: How do total revenue and average order value vary across categories and regions? Which recorded variables are associated with order revenue? What conclusions can managers reasonably draw from the analysis?

LITERATURE REVIEW

Dekimpe [1] examined retailing and retail research in the age of big data analytics. The study discusses the opportunities available to retailers and the methodological difficulties involved in translating extensive data into business knowledge. This perspective supports using transaction summaries to inform managerial questions while recognising that revenue rankings do not, by themselves, explain profitability or customer preferences.

Verhoef et al. [2] presented a multidisciplinary account of digital transformation and identified the organisational changes associated with digital technologies and business models. Their work connects analytical capabilities with strategic and operational decisions. The present analysis applies this decision orientation to product performance, regional sales and payment methods, without claiming that a descriptive dashboard alone constitutes organisational transformation.

Becker and Jaakkola [3] developed an integrative understanding of customer experience through a review of its conceptual foundations. Their findings emphasise the breadth of the customer experience construct and the need for consistent definitions. This is relevant because a transaction rating and delivery duration capture selected aspects of service experience rather than a complete measure of satisfaction or loyalty.

Huang and Rust [4] proposed a strategic framework for artificial intelligence in marketing, linking different forms of intelligence to marketing research, strategy and action. The framework suggests that analytical tools should be selected according to their managerial purpose. Here, grouped sales reporting provides a basis for performance assessment, while linear regression is evaluated as a baseline for order revenue rather than an autonomous marketing decision system.

Verma et al. [5] reviewed artificial intelligence applications in marketing and outlined directions for future research. Their study brings together the uses of AI within marketing activities and demonstrates the need to understand both the application and its decision context. For the present study, this supports distinguishing a useful business question from the technical task of fitting a model to available variables.

Vlačić et al. [6] reviewed the evolving role of artificial intelligence in marketing and developed a research agenda concerning adoption, use, acceptance, ethics and competencies. The review indicates that successful application involves organisational and human considerations alongside technology. Sales analytics therefore requires managers who can interpret outputs and recognise the limits of the data before changing commercial policies.

Mariani, Perez-Vega and Wirtz [7] integrated research on artificial intelligence across marketing, consumer research and psychology. Their systematic review connects technological applications with consumer responses and identifies future research directions. It provides a broader setting for transaction analytics, although the present records cannot test psychological mechanisms or the effects of AI on customer decisions.

Olabode et al. [8] investigated big data analytics capability and market performance, considering the roles of disruptive business models and competitive intensity. Their research treats analytical capability as part of a wider organisational and competitive setting. This reinforces the need to evaluate business outcomes beyond a model score. The present synthetic data illustrate performance reporting but do not measure firm capabilities or market performance.

Fedushko and Ustyianovych [9] applied cohort analysis to e-commerce customer behaviour in a COVID-19 case study. Their work shows how grouping customers over time can address business questions concerning awareness and loyalty. It suggests a useful extension to order-level reporting: tracking subsequent purchases within customer cohorts. Such a longitudinal analysis is outside the current reported results.

Alves Gomes and Meisen [10] reviewed 105 publications on customer segmentation for personalised targeting in e-commerce. They identified a process involving data collection, customer representation, segmentation and targeting, with RFM analysis and k-means frequently used. Their findings support a structured approach to customer analysis. Product and regional groups in the present study are reporting categories, not the outcome of an algorithmic customer segmentation procedure.

Salminen et al. [11] reviewed algorithmic customer segmentation and examined how algorithms, parameters and evaluation methods are used. The review found limited involvement of subject matter experts in assessing the usefulness of generated segments. This highlights the importance of evaluating outputs against managerial needs. The present analysis accordingly interprets revenue

Publication of the European Centre for Research Training and Development -UK
totals alongside order volume and average order value rather than treating a numerical ranking as a sufficient decision rule.

Herhausen et al. [12] outlined recent progress and future directions for machine learning in marketing and presented a guiding framework for its applications. Their discussion supports placing prediction within a defined marketing task and recognising implementation challenges. In order revenue modelling, this requires specifying what information is available when a prediction is made and whether the target is an unknown outcome or a calculation from known transaction terms.

Cruz and Rosário [13] reviewed emerging themes and research gaps in data-driven decision-making in marketing. Their bibliometric and thematic approach connects analytical technologies with marketing practices and research. This recent review motivates a business-oriented interpretation of sales data, where the value of analysis depends on the decisions it can support and the evidence available to justify those decisions.

GAPS IDENTIFIED IN THE EXISTING WORKS

The reviewed literature addresses retail analytics, digital transformation, customer experience, artificial intelligence and segmentation. These studies provide established foundations; they do not justify claiming that integrated sales analysis is absent from prior research. The specific issue addressed here is how a compact set of transaction measures can produce different managerial interpretations, and how an apparently strong revenue model should be assessed when revenue may already be determined by recorded commercial terms. The study therefore combines category, regional and payment reporting with an examination of the business meaning of the model target. Its contribution is a transparent illustrative analysis rather than a novel forecasting method or evidence of a market-wide behavioural effect.

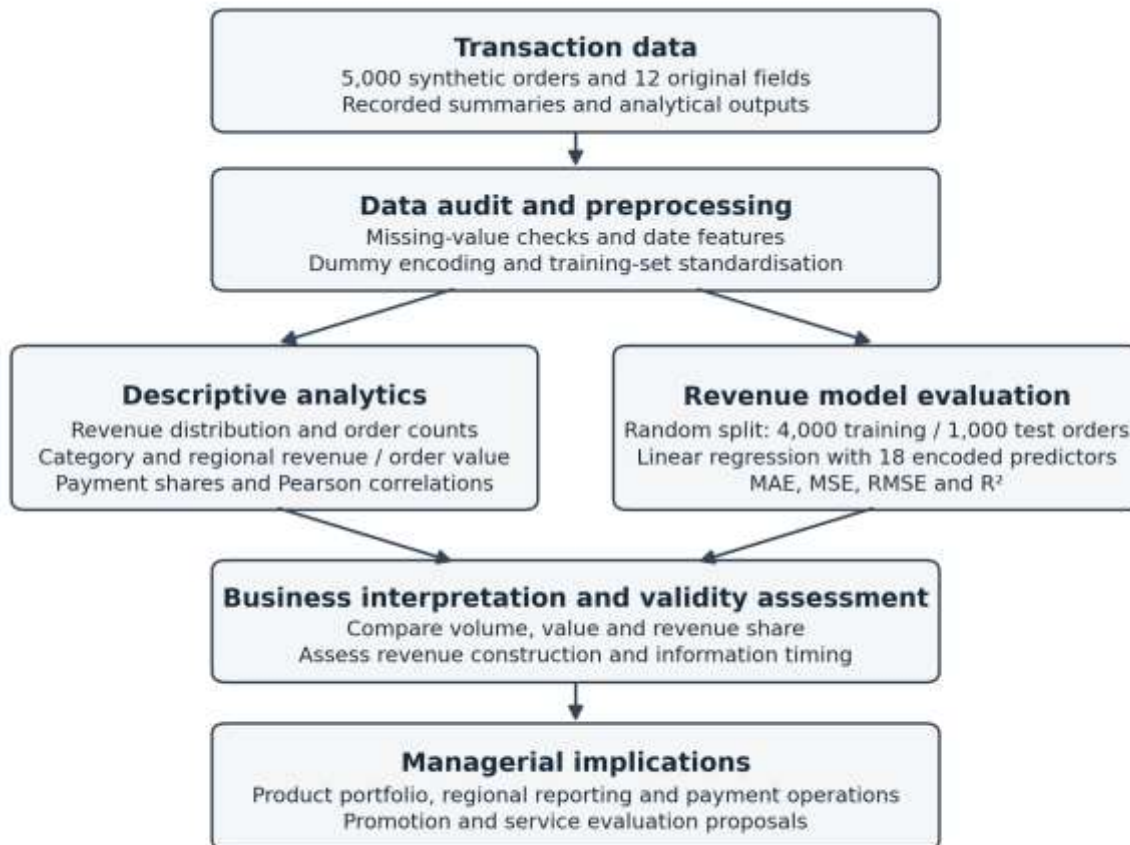
METHODOLOGY

Research Approach

The study adopts a quantitative exploratory approach to e-commerce sales analysis. A business intelligence perspective is used to connect order-level information with product portfolio management, regional performance assessment and payment operations. Descriptive findings are interpreted separately from predictive evaluation and causal claims. The analytical workflow covers structural auditing, preprocessing, descriptive analysis, visualisation, regression evaluation and managerial interpretation.

Research Analytics Framework

Figure 1 presents the research analytics architecture. The workflow converts transaction records into two complementary forms of evidence: descriptive business measures and an evaluated revenue model. Both outputs are considered together before managerial implications are developed.



Commercial recommendations require validation with real transactions and cost data.

Figure 1. Descriptive analysis and revenue model evaluation converge in a business interpretation stage, where sales patterns and model validity are assessed before managerial recommendations are proposed.

The descriptive branch identifies the distribution of sales across products, regions and payment methods. The modelling branch assesses held-out order revenue using linear regression. The interpretation stage compares total revenue with order counts and average order value and examines whether the revenue target reflects an accounting calculation. This stage also considers whether predictors would be available when a business forecast is required. The architecture therefore links analytical outputs with product, regional, payment and service questions while preserving the

Publication of the European Centre for Research Training and Development -UK
distinction between observed patterns and proposed interventions. It describes the research workflow rather than an implemented real-time business intelligence platform.

Dataset Description

The study examines a publicly available synthetic transaction dataset [14]. Each observation represents a completed order. The analytical records contain 5,000 orders and 989 distinct customer identifiers. The 12 original variables comprise order and customer identifiers, order date, product category, region, quantity, unit price, discount, payment method, delivery days, customer rating and revenue. Because currency and geographic meanings are not established in the analytical records, financial values are reported in monetary units and regions retain their original labels.

The analysis combines recorded descriptive summaries, regression outputs and archived visualisations. Revenue shares and average order values are calculated from group totals and order counts. The reported model results have not been independently reproduced from the complete raw transaction file. Consequently, inference is limited to the available analytical records; statistical significance, model uncertainty and the revenue-generation mechanism remain unverified.

Table 1. Dataset characteristics

Characteristic	Recorded value
Orders	5,000
Original variables	12
Distinct customer IDs	989
Product categories	4
Regions	4
Payment methods	3
Missing values	0 in each original field
Recorded date range	1 January 2022 to 9 September 2035

Data Cleaning and Preprocessing

Dates are converted to datetime, and year, month, day, weekday and quarter are extracted. The future-dated range means that these records should not be presented as an observed historical market period. The date distribution must be considered part of the synthetic dataset design until its generation process is documented. There is insufficient evidence to claim real seasonality or a real forecast horizon.

Data Analysis Techniques

Exploratory analysis consists of descriptive statistics, category and regional totals, payment shares, histograms, scatterplots, boxplots and a numeric correlation matrix. Average order value is calculated as group revenue divided by group order count. Revenue share is group revenue divided by total revenue, multiplied by 100. Pearson correlations measure linear association and are not interpreted as causal effects.

An interquartile-range screening identifies 67 revenue observations above the upper threshold of 3,257.43. Although an intermediate filtering step retains 4,933 orders, the final modelling sequence restores the complete 5,000-order dataset. All orders are therefore included in the reported regression results. High-value orders are retained because a large transaction value alone is insufficient evidence of data error.

Revenue Model and Evaluation

For regression, revenue is the target. Order ID, raw order date and customer ID are excluded. Categorical variables are converted into dummy indicators with the first level omitted. After encoding, the model uses 18 predictors. The data are split randomly into 4,000 training and 1,000 test observations using `random_state = 42`. Numerical predictors are standardised using training-set statistics and the same transformation is applied to the test set. An ordinary least squares LinearRegression model is fitted. The fitted model is linear regression; no XGBoost comparison is reported.

Model evaluation uses MAE, MSE, RMSE and R^2 . MAE describes the average absolute prediction error in monetary units; RMSE gives greater weight to larger errors. R^2 compares squared errors with variation around the test-set mean. No cross-validation, confidence intervals, temporal evaluation or customer-grouped validation is available. The random split therefore assesses held-out orders within this synthetic sample, not future sales forecasting.

Analytical Tools

The analytical workflow uses Python. Pandas and NumPy support data manipulation and numerical summaries, Matplotlib and Seaborn produce the recorded visualisations, and scikit-learn provides the standardisation, train-test split and linear regression procedures. The current study interprets the archived outputs; independent re-execution requires the complete raw transaction file.

RESULTS AND DISCUSSION

Overall Sales Profile

Total recorded revenue is 5,109,775.74 monetary units. The mean order value is 1,021.96, compared with a median of 796.65. Revenue ranges from 11.21 to 4,119.33, with a standard deviation of 825.58.

The mean exceeding the median indicates that higher-value orders pull the average upwards. Mean quantity is 4.04 units and mean unit price is 308.42. The mean discount fraction is approximately 0.18, equivalent to 18%. Mean delivery time is 6.12 days and mean customer rating is 2.97 on the recorded 1-to-5 scale.

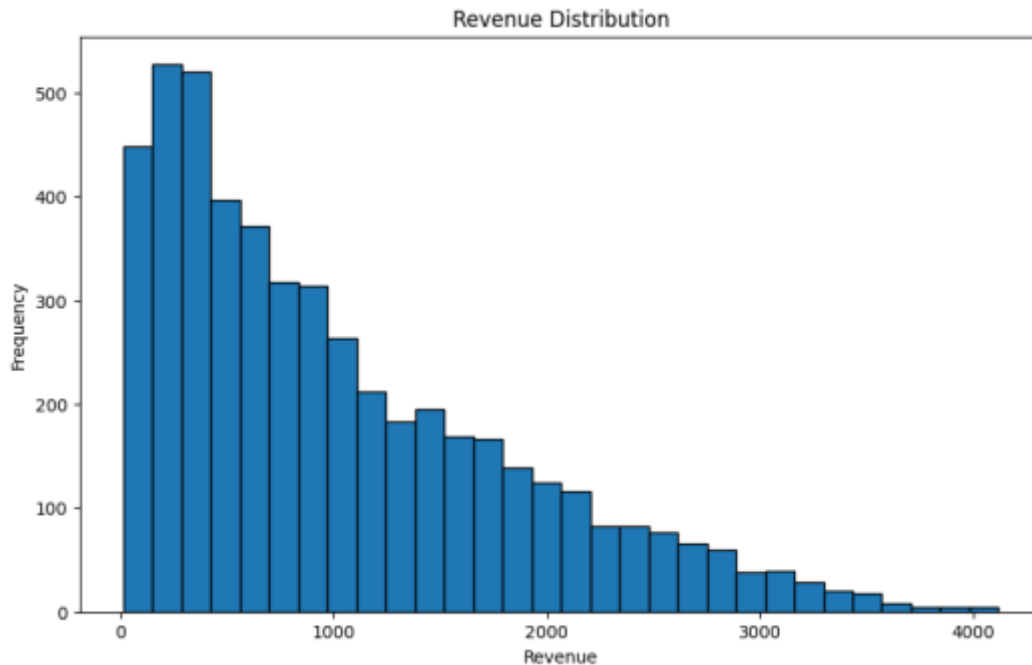


Figure 2. Order revenue is concentrated at lower values with a long upper tail, consistent with the mean exceeding the median.

Figure 2 supports reporting both the average and median rather than treating the mean as a complete description of a typical order. In a business dashboard, an order-value distribution would also help distinguish regular orders from unusually large transactions. The observed upper tail alone does not demonstrate fraud or recording error.

Product Category Performance

Table 2. Sales performance by product category

Category	Orders	Revenue	Share %	Average order value
Electronics	1777	1,829,899.22	35.81	1,029.77
Clothing	1531	1,531,931.72	29.98	1,000.61
Home	969	982,083.92	19.22	1,013.50
Beauty	723	765,860.88	14.99	1,059.28

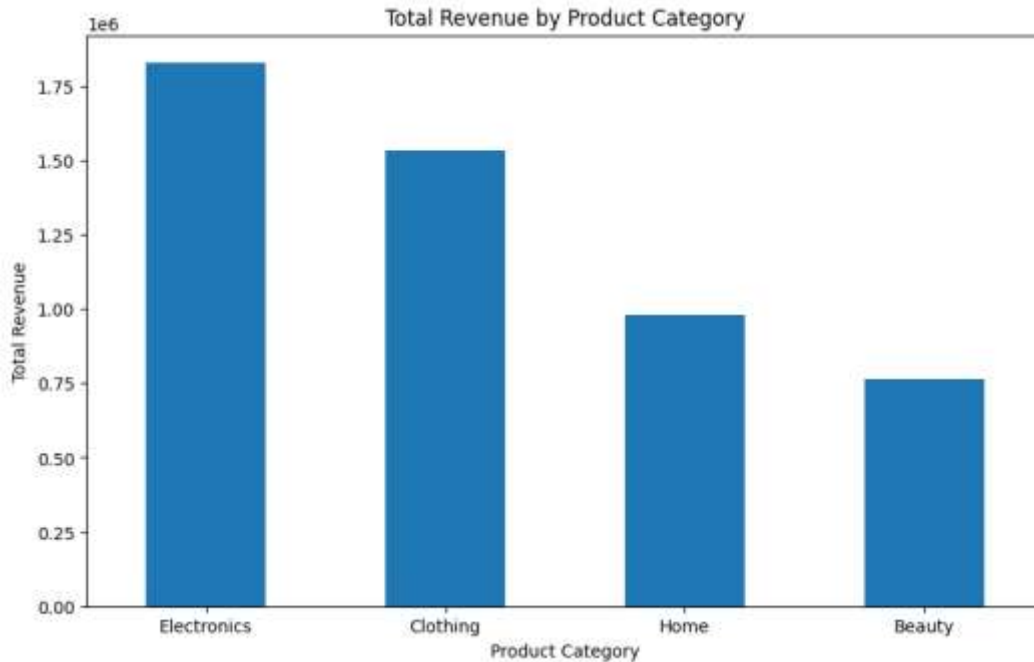


Figure 3. Electronics contributes the most total revenue and Beauty the least; the totals do not indicate average order value or profit.

Electronics contributes the largest revenue share, followed by Clothing, Home and Beauty. Electronics also has the most orders, with 1,777 transactions, or 35.54% of the sample. However, Beauty has the highest average order value at 1,059.28, compared with 1,029.77 for Electronics, 1,013.50 for Home and 1,000.61 for Clothing. Its lower total revenue therefore coexists with a higher average transaction value.

This comparison demonstrates why total revenue alone should not determine category priorities. The Electronics total partly reflects its larger order count. Beauty may deserve further investigation despite its smaller overall contribution. Neither ranking establishes profitability, customer preference in the wider market or the return from additional marketing expenditure because cost, exposure and margin data are absent.

Regional Sales Performance**Table 3. Sales performance by region**

Region	Orders	Revenue	Share %	Average order value
West	1289	1,345,582.16	26.33	1,043.90
North	1276	1,281,508.45	25.08	1,004.32
South	1208	1,246,640.90	24.40	1,031.99
East	1227	1,236,044.23	24.19	1,007.37

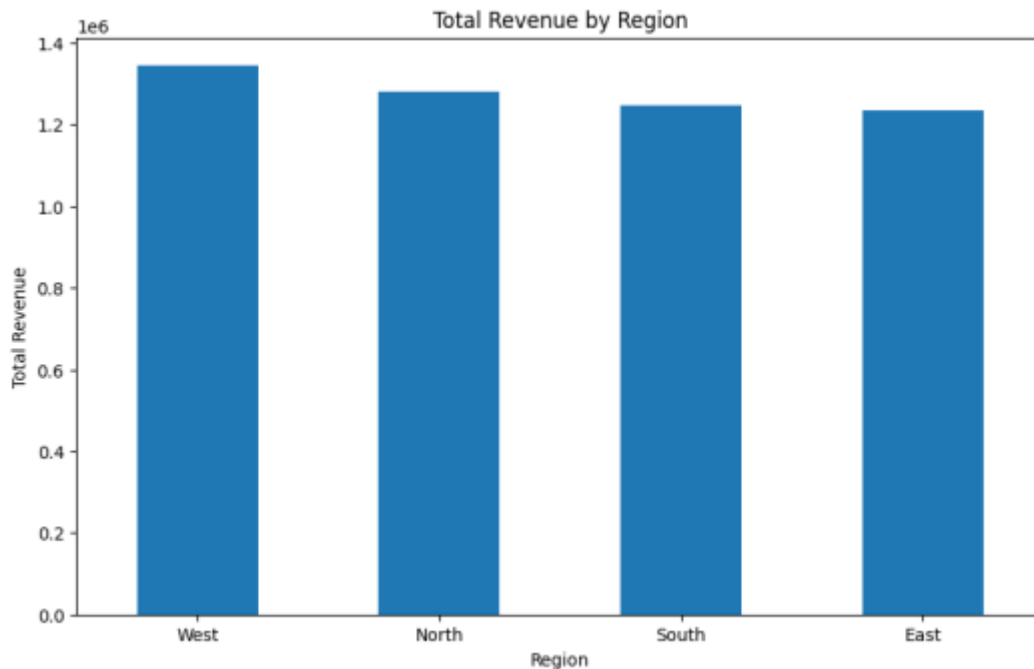


Figure 4. West has the highest total revenue and East the lowest, while the four regional totals remain relatively close.

West records the highest revenue at 1,345,582.16, equivalent to 26.33% of the total. North contributes 25.08%, South 24.40% and East 24.19%. Average order value is 1,043.90 in West, 1,031.99 in South, 1,007.37 in East and 1,004.32 in North. South therefore ranks below North in total revenue but above it in average order value. Regional comparisons should distinguish order volume from transaction value.

These regional shares are relatively close and do not identify a dramatically dominant market. No conclusions can be drawn about regional penetration or marketing efficiency without visitor counts,

Publication of the European Centre for Research Training and Development -UK
customer population, advertising expenditure and delivery costs. The original region labels do not support country-specific recommendations.

Payment Methods

Table 4. Payment method distribution

Payment method	Orders	Share %
Card	2,270	45.40
Cash on delivery COD	1,774	35.48
Wallet	956	19.12

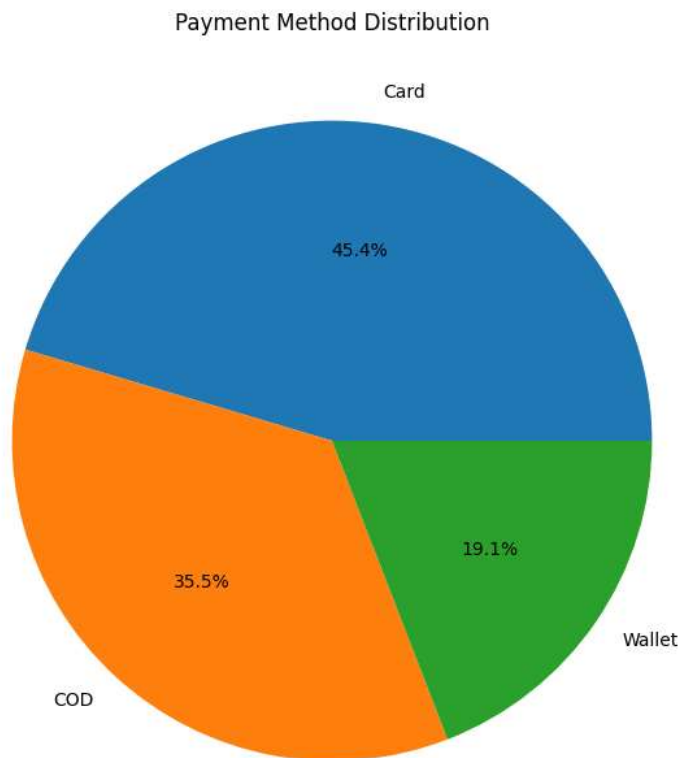


Figure 5. Card accounts for the largest payment share at 45.40%, followed by COD at 35.48% and Wallet at 19.12%.

Card is the most common payment method among completed orders, while cash on delivery represents over one-third of transactions. Wallet accounts for the smallest share. These figures describe the sample's payment mix; they do not show that card payments cause conversion or that

Publication of the European Centre for Research Training and Development -UK
wallet users spend less. Payment-specific order values, payment failures, cancellations and collection costs are not evaluated in the recorded analysis.

Relationships with Order Revenue

Table 5. Linear associations with order revenue

Variable	Pearson correlation with revenue
Unit price	0.6780
Quantity	0.6236
Discount	-0.1393
Delivery days	0.0054
Customer rating	0.0124

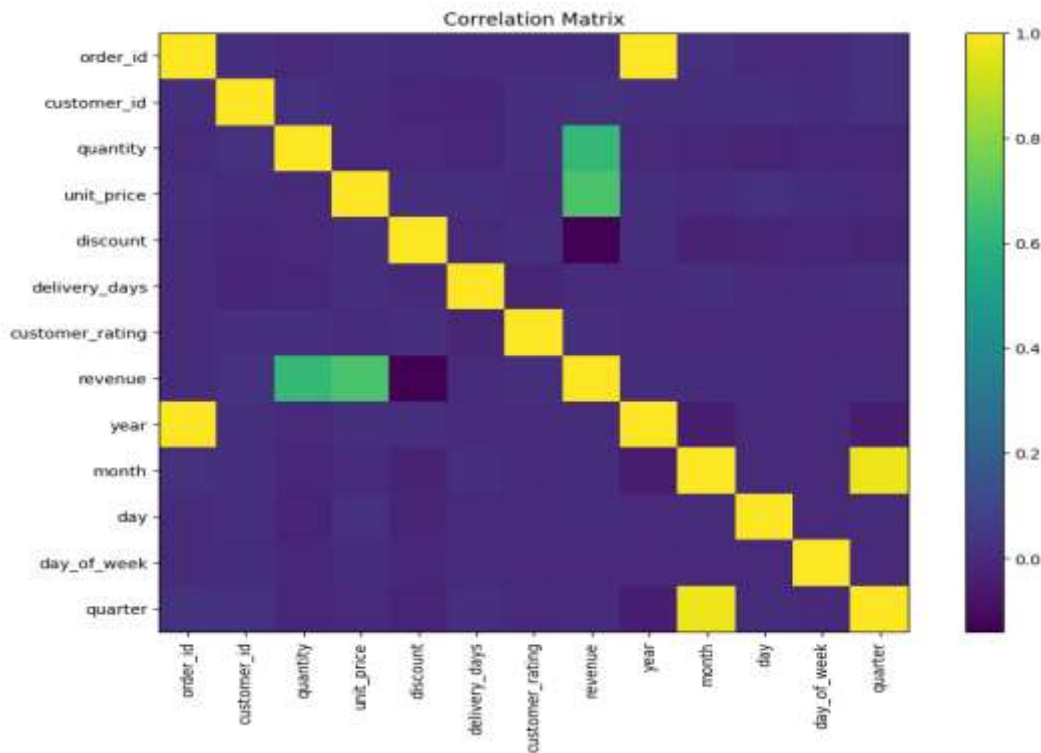


Figure 6. Revenue is positively associated with unit price and quantity, while delivery days and customer rating show little linear association.

Unit price and quantity have the largest positive linear correlations with revenue among the original commercial variables. Discount has a smaller negative correlation. Delivery time and customer rating are close to zero in their linear correlations with revenue. These results do not establish that delivery

Publication of the European Centre for Research Training and Development -UK
quality is irrelevant to retention or satisfaction. Those outcomes are not tested here, and the contemporaneous rating may be recorded after the purchase.

The full correlation matrix also includes identifiers and engineered dates. Such fields should not be treated as meaningful customer characteristics merely because they appear in a numeric matrix. Month and quarter are strongly related because both represent calendar position. Individual revenue correlations do not substitute for model coefficients, tests of significance or evidence of causal mechanisms.

Revenue Model Performance

Table 6. Revenue model evaluation

Metric	Test result
Mean absolute error	219.76
Mean squared error	87,915.64
Root mean squared error	296.51
R squared	0.8695

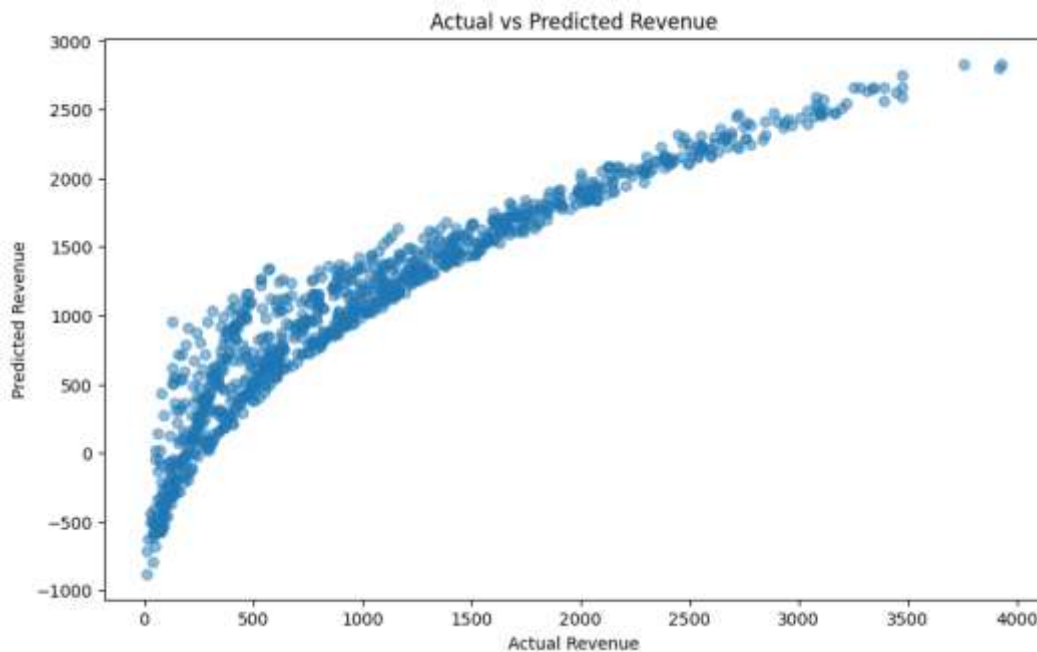


Figure 7. The predictions follow the general revenue trend, but high-value orders tend to be underestimated and some low-value orders receive negative predictions.

The recorded test R^2 is approximately 0.870. This represents a reduction in squared error relative to a test-mean reference, not 87% classification accuracy. MAE indicates an average absolute error of approximately 219.76 monetary units. RMSE is larger than MAE, showing that larger errors contribute more strongly to the squared-error metric. Figure 7 presents the stored actual-versus-predicted scatterplot; the available analysis does not provide residual tests or uncertainty estimates.

A fundamental limitation is the apparent construction of revenue. For the first displayed order, $7 \times 373.65 \times (1 - 0.28)$ equals 1,883.196, which rounds to the recorded revenue of 1,883.20. This suggests the accounting relationship $\text{revenue} = \text{quantity} \times \text{unit price} \times (1 - \text{discount})$. It has not been checked across all observations. If it holds throughout the dataset, a formula-based calculation is the appropriate benchmark and regression mainly approximates a known product of variables. The model result would not establish successful prediction of unknown future demand.

INTERPRETATION OF FINDINGS

The first research question is answered by the category and regional summaries. Electronics leads category revenue and order count, whereas Beauty leads average order value. West leads regional revenue, but the ranking of South and North changes when average order value replaces total revenue. These findings show how a business interpretation can improve on a simple bar-chart ranking.

The second question is answered descriptively by the correlations. Quantity and price are positively associated with revenue, discount is negatively associated, and delivery days and rating show little contemporaneous linear association. The likely accounting relationship makes the first three patterns unsurprising. A higher discount mechanically reduces net revenue for a fixed price and quantity. The analysis cannot establish whether promotions generate enough additional demand to offset that reduction.

The third research question concerns the decision value of the analysis. The summaries show the composition of sales, while the model assessment highlights the need to verify the commercial meaning of the target. These interpretations are consistent with the decision orientation of retail analytics [1] and machine learning in marketing [12]. A high model score cannot establish that a pricing, promotion or service policy will increase profit; that question requires an appropriate business evaluation design.

Business Implications

Product portfolio decisions should use order counts, average order value and revenue together. In a real business, Electronics would warrant monitoring because of its large contribution, while Beauty would warrant investigation because of its higher average order value. Managers would need category margins, stock availability and marketing costs before reallocating budget. The synthetic rankings themselves do not justify an immediate commercial change. This decision-oriented approach is consistent with the role of analytics in retail performance assessment [1].

Regional reporting should separate transaction volume from value. A useful follow-up dashboard would combine the present measures with customer acquisition cost, delivery cost and repeat purchasing. Regional campaigns should be tested against these measures rather than selected solely because West has the highest total revenue in the sample.

Promotion evaluation requires a comparison of discount levels with quantity, margin and demand. The small negative discount–revenue correlation is insufficient grounds to reduce discounts. A business could test promotions through a controlled experiment, measuring incremental gross profit rather than revenue alone. Revenue after discounts must be distinguished from profit after product and fulfilment costs. This proposed evaluation would connect transaction reporting with business decision-making [13].

The payment mix suggests that a comparable real business should assess the operational requirements of multiple payment options. Cash on delivery would merit review of cancellation rates and collection costs. Card and wallet payments would merit review of processing fees and checkout failures. None of these outcomes is established by the current dataset.

For customer experience, a more useful follow-up would examine delivery time against customer rating and later repeat purchases. A near-zero delivery–revenue correlation does not rule out longer-term service effects. For modelling, the next step is to define when a prediction is made, exclude information unavailable at that moment, and compare against a simple business benchmark. Actual delivery days and post-purchase ratings would be unsuitable inputs for a pre-purchase forecast. Customer experience research [3] supports measuring broader service outcomes, while cohort analysis [9] offers a direction for tracking later customer behaviour. These extensions remain proposals for future research.

Limitations and Future Research

The synthetic dataset limits external validity. Its category frequencies, regional mix and variable relationships may reflect generation choices rather than actual market behaviour. Dates extending to 2035 prevent treating the analysis as a historical trading record. Missing values are absent in stored outputs, but this does not establish realistic data quality or verify the complete generation process.

The analysis relies on archived analytical outputs without the raw CSV or independent re-execution. The revenue formula requires a full-data check. There is only one model, one random split and no recorded statistical inference. Repeat customer IDs may occur in both sets; a customer-grouped evaluation would be appropriate if the objective were generalisation to new customers. A chronological split would be required for a genuine future forecasting question.

The data contain completed orders and therefore cannot estimate visitor conversion, cart abandonment or non-purchaser behaviour. Costs are absent, so revenue cannot be interpreted as profit.

Ratings are single recorded values rather than a validated multi-item measure of satisfaction. Neither the present correlation analysis nor the fitted regression identifies causal effects.

Future work should begin by verifying the dataset and defining a business outcome that is unknown at the prediction time. On real data, suitable extensions include repeat purchasing, later customer value, fulfilment delays or aggregate future demand. Linear regression and XGBoost could then be compared using appropriate held-out evaluation. Such comparisons are proposed research, not results of this paper.

CONCLUSION

The study demonstrates a business administration approach to e-commerce sales analysis using 5,000 synthetic orders. Electronics generates the highest category revenue, Beauty has the highest average order value, West leads regional revenue and card is the most common payment method. Revenue is positively associated with price and quantity, while delivery time and customer rating have little direct linear association with contemporaneous revenue. The recorded linear regression results indicate R^2 of 0.870, but an apparent accounting formula limits the interpretation of this score as independent predictive achievement. The principal contribution is therefore the business interpretation of descriptive measures and the careful separation of sales reporting, revenue calculation and genuine prediction. Commercial recommendations require validation with real transactions, cost information and an appropriate research design.

REFERENCES

- [1] Dekimpe, M. G. (2020). Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing*, 37(1), 3–14. DOI: <https://doi.org/10.1016/j.ijresmar.2019.09.001>
- [2] Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., and Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889–901. DOI: <https://doi.org/10.1016/j.jbusres.2019.09.022>
- [3] Becker, L., and Jaakkola, E. (2020). Customer experience: Fundamental premises and implications for research. *Journal of the Academy of Marketing Science*, 48, 630–648. DOI: <https://doi.org/10.1007/s11747-019-00718-x>
- [4] Huang, M.-H., and Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49, 30–50. DOI: <https://doi.org/10.1007/s11747-020-00749-9>
- [5] Verma, S., Sharma, R., Deb, S., and Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights*, 1(1), 100002. DOI: <https://doi.org/10.1016/j.jjime.2020.100002>

- [6] Vlačić, B., Corbo, L., Costa e Silva, S., and Dabić, M. (2021). The evolving role of artificial intelligence in marketing: A review and research agenda. *Journal of Business Research*, 128, 187–203. DOI: <https://doi.org/10.1016/j.jbusres.2021.01.055>
- [7] Mariani, M. M., Perez-Vega, R., and Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4), 755–776. DOI: <https://doi.org/10.1002/mar.21619>
- [8] Olabode, O. E., Boso, N., Hultman, M., and Leonidou, C. N. (2022). Big data analytics capability and market performance: The roles of disruptive business models and competitive intensity. *Journal of Business Research*, 139, 1218–1230. DOI: <https://doi.org/10.1016/j.jbusres.2021.10.042>
- [9] Fedushko, S., and Ustyianovych, T. (2022). E-Commerce Customers Behavior Research Using Cohort Analysis: A Case Study of COVID-19. *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1), 12. DOI: <https://doi.org/10.3390/joitmc8010012>
- [10] Alves Gomes, M., and Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 21, 527–570. DOI: <https://doi.org/10.1007/s10257-023-00640-4>
- [11] Salminen, J., Mustak, M., Sufyan, M., and Jansen, B. J. (2023). How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation. *Journal of Marketing Analytics*, 11, 677–692. DOI: <https://doi.org/10.1057/s41270-023-00235-5>
- [12] Herhausen, D., Bernritter, S. F., Ngai, E. W. T., Kumar, A., and Delen, D. (2024). Machine learning in marketing: Recent progress and future research directions. *Journal of Business Research*, 170, 114254. DOI: <https://doi.org/10.1016/j.jbusres.2023.114254>
- [13] Cruz, R. N., and Rosário, A. T. (2025). Data-Driven Decision-Making in Marketing: A Systematic Literature Review of Emerging Themes and Research Gaps. *Systems*, 13(12), 1114. DOI: <https://doi.org/10.3390/systems13121114>
- [14] abbas829. (n.d.). Ecommerce sales dataset [Synthetic dataset]. Kaggle. <https://www.kaggle.com/datasets/abbas829/ecommerce-sales-dataset> Accessed 10 October 2026. No DOI assigned.