



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 4 (2026)



ijerst.editor@gmail.com
editor@ijerst.com

E COMMERCE GROWTH ENGINE

Shivam R. Ghate, Mayur M. Bhairam, Dipak Dohale, Mohit Ghate, Dr. Nikita R. Hatwar, Prof.. Renuka Kulat

Department of Information Technology Priyadarshini College of Engineering, Nagpur, India

Academic Year 2026–2027

Abstract—E-commerce platforms generate large volumes of session, page-view, transaction, product, and refund data, but operational decisions often require these heterogeneous sources to be interpreted together. This paper presents an interactive analytics framework that integrates six related e-commerce datasets into a reproducible pipeline for marketing performance, conversion behavior, product and refund analysis, and customer lifecycle analysis. The implementation uses Python, Pandas, Plotly, and Streamlit. Data are cleaned and joined using session, order, user, product, and order-item keys, after which business metrics are computed and exposed through an interactive dashboard. On 472,871 website sessions and 32,313 orders, the framework reports a 6.83% orders-per-session conversion rate, 44.76% bounce rate, USD 1.94 million gross order revenue, and USD 85.34 thousand in recorded refunds. Channel analysis shows that gsearch generated the largest traffic and revenue contribution, while bsearch produced a higher conversion rate. Product analysis identifies differences in refund exposure, and lifecycle analysis shows a low observed repeat-ordering rate of 1.86% with a mean inter-order gap of about 34 days. The contribution is a practical, end-to-end analytics architecture that connects acquisition, navigation, transactions, refunds, and repeat behavior in one decision-support interface while explicitly documenting metric definitions and analytical limitations. Electronic commerce (e-commerce) has emerged as a major driver of economic growth by transforming traditional business practices into digitally connected systems. The increasing use of smartphones, internet connectivity, digital payment systems, artificial intelligence, and online marketplaces has changed the way consumers purchase goods and services. E-commerce enables businesses to reach a wider customer base, reduce certain operational costs, improve customer convenience, and develop new business models. In India, initiatives such as Digital India, Unified Payments Interface (UPI), and the Open Network for Digital Commerce (ONDC) have supported the expansion of digital commerce, particularly among micro, small, and medium enterprises (MSMEs). However, cybersecurity threats, data privacy concerns, logistical difficulties, intense competition, and unequal digital literacy remain important challenges. This paper examines the major factors driving e-commerce growth, its economic and social impact, the challenges faced by businesses, and future opportunities. The study adopts a descriptive research approach based on secondary information from government publications and industry reports. It concludes that sustainable e-commerce growth requires technological innovation, consumer trust, efficient logistics, digital inclusion, and responsible business practices.

Index Terms—e-commerce analytics, web analytics, conversion analysis, customer retention, interactive visualization

I. INTRODUCTION

E-commerce systems create a continuous stream of behavioral and transactional data. Website sessions and page views describe acquisition and navigation, while orders, order items, products, and refunds describe commercial outcomes. Web analytics can support marketing performance measurement, but its usefulness depends on the selection, processing, and organizational interpretation of metrics [2]. Similarly, the customer journey spans multiple touch points and therefore requires a cross-stage view rather than isolated measurements [3].

A practical challenge is that these data are commonly stored in separate tables and analyzed with different tools. A marketing analyst may inspect traffic and conversion, a product team may inspect sales and refunds, and a retention team may inspect repeat orders. Without a common analytical layer, relationships among these views can be difficult to examine consistently.

Recent research also demonstrates the value of integrating browsing patterns with transaction information for purchase analysis and segmentation [4].

This work addresses the following research question: *How can heterogeneous e-commerce interaction and transaction data be integrated into an interactive analytics framework that provides consistent, reproducible, and decision-oriented views of acquisition, conversion, product performance, refunds, and customer lifecycle behavior?*

The paper makes three contributions. First, it defines an end-to-end data model and processing workflow linking six operational datasets. Second, it implements reusable analytical functions and an interactive dashboard using open Python technologies. Third, it evaluates the framework on a multi-year e-commerce dataset and reports operational metrics with explicit definitions and limitations rather than treating dashboard visualizations as predictive models.

II. RELATED WORK

Marketing analytics has evolved toward data-rich environments in which organizations combine behavioral and operational measurements to support decisions [1]. Järvinen and Karjaluoto showed that web analytics can improve the measurability of digital marketing, while also emphasizing the importance of how metrics are selected and processed [2]. These findings motivate a framework in which traffic, conversion, and revenue are examined together.

Customer journey research emphasizes that customer experience is formed across multiple touch points and channels [3]. In e-commerce, clickstream and transaction data can be integrated to improve purchase prediction and customer segmentation [4]. Prescriptive churn research further shows that customer lifecycle decisions can benefit from business-oriented objectives rather than relying only on predictive discrimination [5].

Existing e-commerce studies also frequently focus on a single problem, such as purchase prediction, recommendation, fraud detection, or demand forecasting. For example, recent IEEE publications have investigated online purchase intention and e-commerce recommendation systems using machine learning [6], [7]. In contrast, the present work focuses on an integrated

TABLE I
DATASETS USED IN THE FRAMEWORK

Dataset	Records	Key	Primary use
Sessions	472,871	session ID	acquisition, bounce
Page views	1,188,124	session ID	navigation stages
Orders	32,313	order ID	revenue, conversion
Order items	40,025	order item ID	product sales
Refunds	1,731	order item ID	refund exposure
Products	4	product ID	product labels

descriptive analytics layer. Its goal is not to claim predictive superiority, but to provide a transparent foundation from which later predictive or prescriptive modules can be developed.

III. DATASET AND ANALYTICAL MODEL

The experimental data contain six CSV sources covering approximately March 2012 to March 2015, with refund records extending into April 2015. The sources are website sessions, website page views, orders, order items, refunds, and products. Table I summarizes the records used in the study.

The core entities are connected through session, user, order, product, and order-item identifiers. This permits session-level measures to be connected to order-level revenue and item-level product and refund measures. The dashboard distinguishes unique orders from order-item records; this distinction is important because one order can contain multiple items.

The principal metrics are defined as follows. Orders-per-session conversion is

$$C = \frac{N_{orders}}{N_{sessions}} \times 100. \quad (1)$$

Average order value is

$$AOV = \frac{R_{orders}}{N_{orders}}, \quad (2)$$

where R_{orders} is the sum of order revenue. Net revenue after recorded refunds is

$$R_{net} = R_{orders} - R_{refunds}. \quad (3)$$

For customer lifecycle analysis, an ordering user is a user associated with at least one order, while a repeat ordering user has more than one order. The observed repeat-user rate is

therefore

$$RR = \frac{N_{repeat\ users}}{N_{ordering\ users}} \times 100. \quad (4)$$

IV. PROPOSED FRAMEWORK

Fig. 1 shows the architecture. Raw CSV files are loaded through a cached data-access layer. Cleaning includes data-type normalization, timestamp parsing, duplicate checks, missing-value handling where applicable, and consistent key-based joins. The analytical layer then computes KPIs and grouped summaries. Plotly functions in a separate visualization module generate charts, while Streamlit exposes the results as four

Fig. 1. Proposed analytical architecture

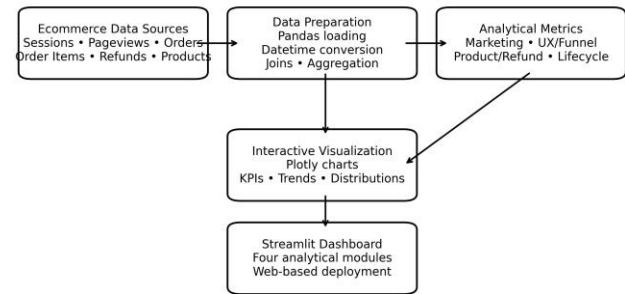


Fig. 1. End-to-end architecture of the e-commerce analytics framework.

campaign activity. The second examines page-stage reach and checkout navigation. The third links products, order items, sales, cost, and refunds. The fourth groups customers by ordering behavior and measures repeat activity, order gaps, and order-value distributions.

A design objective is reproducibility. Instead of manually producing individual charts, the implementation uses reusable functions in `visualization.py`. The Streamlit application loads the datasets once through a caching mechanism and passes cleaned tables to analytical functions. This reduces repeated data-loading operations and makes the dashboard easier to extend.

V. IMPLEMENTATION AND EXPERIMENTAL SETUP

The framework was implemented in Python using Pandas for tabular processing, Plotly for interactive visualizations, and

Streamlit for dashboard deployment. The deployed application provides a browser-based interface suitable for exploratory business analysis. The analysis was performed on the complete available records rather than a sampled subset.

The workflow is: (1) load six datasets, (2) parse timestamps and standardize fields, (3) validate primary and foreign-key relationships, (4) calculate session, order, product, refund, and user-level measures, (5) aggregate metrics by source, campaign, product, and customer type, and (6) render interactive visualizations. All financial results are expressed in U.S. dollars

because the underlying price and cost fields are denominated in USD.

For acquisition analysis, sessions are grouped by traffic source. For product analysis, order items are joined with product

analytical views.

The first dashboard view evaluates marketing channels using sessions, orders, conversion, bounce behavior, revenue, and

metadata and refund records. For lifecycle analysis, orders are grouped by user and sorted by creation time; the difference between consecutive order dates is used to estimate the inter-order gap. Page-stage analysis uses unique sessions reaching URLs containing recognized page-stage patterns.

An important analytical safeguard is that page-stage reach is not interpreted as a strict monotonic funnel. In the underlying data, the unique-session count at the product page can exceed the count at the home page because users may enter the site through non-home landing paths. Consequently, page-stage

TABLE II
OVERALL EXPERIMENTAL RESULTS

Metric	Result
Website sessions	472,871
Page views	1,188,124
Unique orders	32,313
Order items	40,025
Order revenue	USD 1,938,509.75
COGS	USD 722,370.25
Refund amount	USD 85,338.69
Net revenue	USD 1,853,171.06
AOV	USD 59.99
Orders/session	6.83%
Bounce rate	44.76%
Repeat ordering users	591
Repeat-user rate	1.86%
Mean inter-order gap	34.00 days

counts are reported as navigation reach rather than as sequential conversion denominators.

VI. RESULTS

A. Overall Business Metrics

Across 472,871 sessions, the framework identifies 32,313 unique orders. Total order revenue is USD 1,938,509.75 and cost of goods sold is USD 722,370.25, producing USD 1,216,139.50 gross profit before recorded refunds. Refunds total USD 85,338.69, resulting in USD 1,853,171.06 net revenue after refunds. The resulting AOV is approximately USD 59.99 and orders-per-session conversion is 6.83%. The observed bounce rate is 44.76%.

These figures demonstrate the value of connecting traffic and transaction tables: acquisition volume alone does not represent commercial performance, and order revenue alone does not expose refund leakage or navigation behavior.

B. Marketing Channel Analysis

The channel results show different performance profiles. gsearch accounts for 316,035 sessions and 21,333 orders, with 6.75% conversion and USD 1,276,144.89 revenue. bsearch contributes 62,823 sessions and 4,519 orders, with a higher 7.19% conversion and USD 268,672.50 revenue. socialbook contributes 10,685 sessions and 343 orders, with 3.21% conversion and USD 22,259.33 revenue. An unattributed group accounts for 83,328 sessions, 6,118 orders, 7.34% conversion, and USD 371,433.03 revenue.

The results illustrate why both volume and efficiency should be considered. A channel can generate a large absolute number of orders because of traffic scale while another can show a higher conversion ratio with substantially less traffic. The dashboard therefore presents channel-level traffic, orders, conversion, and revenue together.

C. Product and Refund Analysis

The four products display different revenue and refund patterns. The Original Mr. Fuzzy generates the largest sales volume with 24,226 units and USD 1,211,057.74 in sales. Its recorded refund amount is USD 61,837.63, equivalent to 5.11%

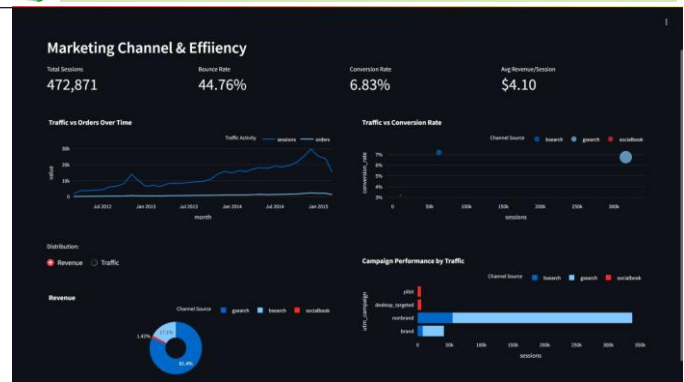


Fig. 2. Marketing analytics view showing traffic, orders, conversion, revenue distribution, and campaign activity.

TABLE III
PRODUCT SALES AND REFUND EXPOSURE

Product	Units	Sales (USD)	Refund/Sales
Original Mr. Fuzzy	24,226	1,211,057.74	5.11%
Forever Love Bear	5,796	347,702.04	2.23%
Birthday Sugar Panda	4,985	229,260.15	6.04%
Hudson River Mini bear	5,018	150,489.82	1.28%

of sales. Forever Love Bear records USD 347,702.04 sales and a 2.23% refund-to-sales ratio. Birthday Sugar Panda records USD 229,260.15 sales and the highest refund-to-sales ratio among the four products at 6.04%. Hudson River Mini bear records USD 150,489.82 sales and the lowest refund-to-sales ratio at 1.28%.

D. Customer Lifecycle Analysis

The lifecycle view identifies 31,696 ordering users, of whom 591 placed more than one order. This produces an observed repeat-user rate of 1.86%. The mean gap between consecutive orders is approximately 34 days. The low repeat-user proportion should be interpreted as a descriptive property of the observed dataset, not as a causal conclusion about customer satisfaction or loyalty.

The dashboard also enables comparison of new and repeat customer order values and product-level repeat-purchase behavior. Such views can support subsequent retention experiments, but the current study does not estimate causal treatment effects or customer lifetime value.

VII. DISCUSSION

The experimental results indicate that an integrated descriptive layer can expose relationships that are difficult to see when acquisition, transactions, refunds, and customers are analyzed independently. The framework provides four complementary perspectives: traffic efficiency, navigation reach, commercial product performance, and customer lifecycle behavior.

From a methodological perspective, the main value is traceability. Each reported KPI has a defined denominator and source table, and the distinction between orders and order items prevents a common aggregation error. The framework

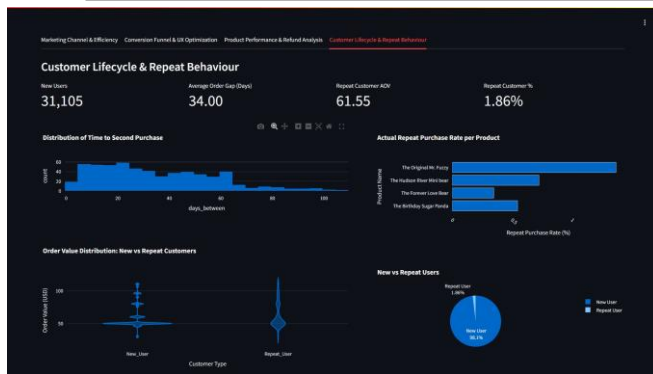


Fig. 3. Customer lifecycle view showing inter-order timing, product repeat behavior, order-value distributions, and new-versus-repeat users. also separates descriptive analytics from predictive claims. This is important because dashboard correlations alone cannot establish causality.

The findings also suggest several practical analytical directions. Channel efficiency can be investigated using controlled campaign comparisons; product refund exposure can be decomposed by product and order characteristics; and the observed 34-day mean inter-order gap can be used to construct candidate retention windows. More advanced models could then predict purchase propensity or churn using browsing and transaction features, an area supported by prior research [4], [5].

VIII. LIMITATIONS

The study has five principal limitations. First, the data represent one historical e-commerce environment and therefore may not generalize to other businesses. Second, the analysis is descriptive and does not establish causal relationships. Third, the observed repeat-user rate depends on the available observation window and user identifiers. Fourth, source attribution includes an unattributed traffic group, which limits channel interpretation. Fifth, page-stage reach is not a strict sequential funnel because users can enter at different URLs; consequently, stage counts are not presented as a conventional monotonic funnel conversion.

In addition, the current implementation does not include machine-learning prediction, controlled experimentation, customer lifetime value estimation, or real-time streaming. These omissions are deliberate: the objective of this paper is to establish a transparent analytical foundation rather than to claim performance for models that were not evaluated.

IX. FUTURE WORK

Future development will extend the descriptive framework with time-series sales forecasting, product recommendation, customer segmentation, and near-real-time event processing. A machine-learning extension can use session-level and transaction-level features to predict purchase probability, while a retention module can estimate customer value and recommend intervention windows. Model evaluation should use time-aware validation and business metrics in addition to standard statistical

measures. A real-time architecture could replace periodic CSV loading with an event stream and analytical warehouse while retaining the same metric definitions.

The principal findings are:

1. **Digital accessibility is a major growth factor.** Smartphones and internet connectivity make online shopping available to a wider population.
2. **Digital payments support transaction efficiency.** Reliable payment systems simplify online purchasing and can increase consumer confidence.
3. **E-commerce creates opportunities for smaller businesses.** Digital platforms can expand market reach without requiring a large physical retail network.
4. **Technology improves operational performance.** AI, analytics, cloud computing, and automation can support demand forecasting, inventory control, and customer service.
5. **Trust remains essential.** Data protection, secure payments, transparent product information, and effective complaint resolution influence long-term customer relationships.
6. **Sustainable growth requires inclusion and efficiency.** Businesses must address digital literacy, delivery costs, competition, and environmental concerns to ensure that the benefits of e-commerce are broadly shared.

Impact of E-Commerce on the Indian Economy

India's digital ecosystem provides a favourable environment for e-commerce expansion. Rising digital adoption, real-time payments, improved logistics, and increasing participation from smaller cities are contributing to the development of online commerce.

According to NITI Aayog's *Trade Watch* publication, India's e-commerce market reached approximately US\$120 billion in 2024. The publication also discusses India's position among the world's major e-commerce markets and the potential for further growth.

The Open Network for Digital Commerce represents another development in India's digital marketplace ecosystem. According to the Government of India's Economic Survey 2024–25, ONDC recorded more than 14 million transactions in November 2024, with participating sellers and service providers spread across more than 600 cities and towns. These figures refer to the period reported in the survey and should not be interpreted as current transaction totals.

The impact of e-commerce in India includes:

- Wider market access for local manufacturers and retailers.

- Increased opportunities for small enterprises and entrepreneurs.
- Improved access to products and services in smaller cities.
- Potential growth in domestic trade and cross-regional commerce
- Development of digital payment and logistics services.
- Greater opportunities for data-driven business operations.

X. CONCLUSION

This paper presented an interactive e-commerce analytics framework that integrates website sessions, page views, orders, order items, products, and refunds into a common decision-support workflow. On the evaluated dataset, the framework measured traffic and conversion, quantified product sales and refund exposure, and characterized repeat ordering behavior. The implementation demonstrates that a lightweight Python stack can provide a reproducible and interpretable analytical layer without requiring a proprietary business-intelligence plat-form. The most important contribution is the explicit connection between data sources, metric definitions, visual analytics, and limitations. This foundation can support later predictive and prescriptive extensions while keeping the analytical process transparent and auditable.

ACKNOWLEDGMENT

The authors thank Dr. Nikita R. Hatwar, Department of Information Technology, Priyadarshini College of Engineering, Nagpur, for academic guidance and support during the project.

REFERENCES

- [1] M. Wedel and P. K. Kannan, "Marketing analytics for data-rich environments," *Journal of Marketing*, vol. 80, no. 6, pp. 97–121, 2016, doi: 10.1509/jm.15.0413.
- [2] J. Ja'rvinen and H. Karjaluo, "The use of Web analytics for digital marketing performance measurement," *Industrial Marketing Management*, vol. 50, pp. 117–127, 2015, doi: 10.1016/j.indmarman.2015.04.009.
- [3] K. N. Lemon and P. C. Verhoef, "Understanding customer experience throughout the customer journey," *Journal of Marketing*, vol. 80, no. 6, pp. 69–96, 2016, doi: 10.1509/jm.15.0420.
- [4] S. Kim, W. Shin, and H. W. Kim, "Predicting online customer purchase: The integration of customer characteristics and browsing patterns," *Decision Support Systems*, vol. 177, Art. no. 114105, 2024, doi: 10.1016/j.dss.2023.114105.
- [5] Y. Feng, Y. Yin, D. Wang, J. Ignatius, T. C. E. Cheng, M. Marra, and Y. Guo, "Enhancing e-commerce customer churn management with a profit- and AUC-focused prescriptive analytics approach," *Journal of Business Research*, vol. 184, Art. no. 114872, 2024, doi: 10.1016/j.jbusres.2024.114872.
- [6] "Machine Learning Models for Predicting Online Shoppers Purchase Intention," in *Proc. 2026 5th Int. Conf. on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballari, India, 2026, doi: 10.1109/ICDCECE68961.2026.11646006.
- [7] "AI-Driven Hybrid Recommendation System for E-Commerce," in *Proc. 2026 4th Int. Conf. on Inventive Computing and Informatics (ICICI)*, Bengaluru, India, 2026, doi: 10.1109/ICICI68773.2026.11580790.
- [8] "AI-Driven Supply Chain Optimization with Predictive Analytics for Enhanced Efficiency and Scalability in E-Commerce," in *Proc. 2026 13th Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2026, doi: 10.23919/INDIACom70271.2026.11526274.