ORIGINAL

Analyzing Consumer Behavior in E-Commerce: Insights from Data-Driven Approaches

Análisis del comportamiento del consumidor en el comercio electrónico: perspectivas de enfoques basados en datos

Danish Anwar[1] ✉, Md. Faizanuddin[1] ✉, Faisal Rahman[2] ✉, Rajeshwar Dayal[3] ✉

[1]Post Graduate Department of Commerce & Business Management, Veer Kunwar Singh University, Ara, India.
[2]University Department of Management, MMH A & P University, Patna, India.
[3]Department of Computer Applications, International School of Management, Patna, India.

ABSTRACT

Introduction: understanding modern shopping behaviors and developing effective marketing strategies have made e-commerce consumer analysis a critical tool for businesses.
Objective: this study delves into consumer behavior patterns by analyzing a comprehensive Kaggle dataset centered on online retail activities, which includes demographic details, purchasing trends, and user engagement metrics. Utilizing advanced data science methodologies—such as data preprocessing, exploratory data analysis, and predictive modeling—the research uncovers valuable insights into customer segmentation, spending behaviors, and strategies to enhance customer retention.
Method: the analysis reveals significant trends, such as the impact of seasonal variations on online shopping, the role of demographic factors in shaping product preferences, and the effectiveness of personalized marketing in boosting conversion rates.
Result: these findings underscore the importance of tailoring marketing efforts to align with consumer needs and preferences. Additionally, the study highlights the necessity for e-commerce businesses to adopt continuous data monitoring and implement robust analytical frameworks to remain competitive in an ever-evolving digital marketplace.
Conclusion: this research provides actionable recommendations for e-commerce stakeholders, emphasizing the integration of data-driven insights into their strategies. By leveraging these insights, businesses can optimize their marketing approaches, improve customer satisfaction, and ultimately drive growth in the highly competitive online retail landscape.

Keywords: Consumer Behavior; E-Commerce Stakeholders; Marketing Strategies.

RESUMEN

Introducción: comprender los comportamientos de compra modernos y desarrollar estrategias de marketing efectivas ha convertido el análisis del consumidor en el comercio electrónico en una herramienta crítica para las empresas.
Objetivo: este estudio profundiza en los patrones de comportamiento del consumidor mediante el análisis de un conjunto de datos integral de Kaggle centrado en actividades de retail en línea, que incluye detalles demográficos, tendencias de compra y métricas de engagement de los usuarios. Utilizando metodologías avanzadas de ciencia de datos—como preprocesamiento de datos, análisis exploratorio y modelos predictivos—

la investigación revela insights valiosos sobre segmentación de clientes, hábitos de gasto y estrategias para mejorar la retención.

Método: el análisis identifica tendencias clave, como el impacto de las variaciones estacionales en las compras en línea, la influencia de factores demográficos en las preferencias de productos y la efectividad del marketing personalizado para aumentar las tasas de conversión.

Resultados: los hallazgos resaltan la importancia de adaptar las estrategias de marketing a las necesidades y preferencias de los consumidores. Además, el estudio subraya la necesidad de que las empresas de e-commerce adopten un monitoreo continuo de datos y marcos analíticos robustos para mantenerse competitivas en un mercado digital en constante evolución.

Conclusión: esta investigación proporciona recomendaciones prácticas para los actores del comercio electrónico, enfatizando la integración de insights basados en datos en sus estrategias. Al aprovechar estos hallazgos, las empresas pueden optimizar sus enfoques de marketing, mejorar la satisfacción del cliente e impulsar el crecimiento en el competitivo entorno del retail en línea.

Palabras clave: Comportamiento del Consumidor; Stakeholders de E-Commerce; Estrategias de Marketing.

INTRODUCTION

The rapid expansion of e-commerce has revolutionized the global retail industry, reshaping how consumers interact with brands and make purchasing decisions. Today, online shopping platforms allow customers to browse, compare, and buy products from virtually anywhere, at any time, creating a seamless and highly accessible shopping experience. This shift has not only altered consumer expectations but has also generated vast amounts of data, capturing every aspect of user behavior—from browsing patterns and product preferences to transaction histories and feedback. For businesses, this data represents a goldmine of opportunities to better understand their customers, optimize operations, and craft targeted marketing strategies. As a result, e-commerce consumer analysis has emerged as a critical discipline, bridging the fields of data science, marketing, and business strategy to unlock actionable insights that drive growth and innovation (Chen et al., 2021).[1] This study focuses on analyzing a comprehensive Kaggle dataset specifically designed to explore online consumer behavior. The dataset includes a wide range of variables, such as customer demographics, product categories, transactional details, and user engagement metrics. By leveraging advanced data science techniques, including descriptive analytics, predictive modeling, and machine learning, this research seeks to address several key questions that are central to understanding and optimizing e-commerce operations.

These questions include:

1. Customer Segmentation: How can customers be effectively grouped based on factors such as demographics, purchasing frequency, and product preferences? Identifying distinct customer segments enables businesses to tailor their offerings and marketing strategies to meet the unique needs of each group.

2. Spending Patterns: What factors influence variations in key metrics such as average order value (AOV) and customer lifetime value (LTV)? Understanding these drivers can help businesses identify high-value customers and develop strategies to maximize revenue.

3. Predictive Modeling: Can machine learning algorithms accurately forecast future purchasing behavior or predict customer churn based on historical data? Predictive insights are essential for proactive decision-making and improving customer retention.

4. Marketing Insights: Which marketing interventions or personalization tactics are most effective in enhancing customer retention and boosting conversion rates? By analyzing the impact of different strategies, businesses can allocate resources more efficiently and improve overall marketing effectiveness.

The structure of this paper is designed to provide a thorough exploration of these questions. It begins with a literature review that contextualizes the study within existing research on e-commerce consumer behavior and data analytics. Next, the methodology section outlines the data science techniques employed, including data preprocessing, exploratory data analysis, and the application of machine learning models. The data analysis section presents the findings, highlighting key trends and patterns, while the discussion interprets these results in the context of their practical implications for e-commerce businesses.

Literature review

The evolution of e-commerce over the last two decades has been nothing short of transformative, driven by advancements in technology, the widespread adoption of smartphones, and the development of robust global logistics networks (Laudon & Traver, 2020).[2] In its early stages, online marketplaces were relatively simplistic, offering limited product selections and rudimentary user interfaces. However, today's e-commerce platforms

are sophisticated ecosystems that leverage personalized recommendations, targeted advertising, and seamless payment solutions to enhance the shopping experience. These innovations have not only elevated consumer expectations but have also intensified competition, making data analytics a critical tool for businesses aiming to retain and grow their market share (Brynjolfsson & McAfee, 2017).[3]

Understanding consumer behavior in the digital marketplace is a complex endeavor, influenced by a multitude of factors such as trust, convenience, perceived value, and the quality of user interface design (Gefen et al., 2003).[4] Theoretical frameworks like the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) have been instrumental in explaining how perceived ease of use and usefulness drive the adoption of e-commerce platforms (Venkatesh et al., 2012).[5] Beyond initial adoption, recent research has shifted focus toward sustaining long-term customer engagement. Studies have explored the impact of loyalty programs, personalized email campaigns, and social proof mechanisms—such as customer reviews and ratings—on consumer decision-making and retention (Chen et al., 2021).[6]

The advent of big data analytics has further revolutionized the e-commerce landscape, enabling platforms to collect and analyze vast amounts of user-generated data. Recommender systems, for example, utilize techniques like collaborative filtering and content-based filtering to provide tailored product suggestions that align with individual customer preferences (Aggarwal, 2016).[7] Additionally, machine learning models—including logistic regression, random forests, and neural networks—have become indispensable tools for predicting customer churn, forecasting purchase behavior, and segmenting users based on engagement metrics (Ahmed et al., 2020).[8] These advanced analytical methods empower businesses to optimize various aspects of their operations, from inventory management and dynamic pricing to highly targeted marketing campaigns.

Despite the significant progress in e-commerce analytics, several gaps remain in the existing body of research. While many studies focus on improving the performance of algorithms, they often overlook the practical challenges of integrating data-driven solutions into established business processes (Wamba et al., 2020).[13] Furthermore, there is a lack of comprehensive studies that address the entire analytics pipeline— from data cleaning and exploratory analysis to predictive modeling and actionable marketing insights—within a unified framework. This paper seeks to address these gaps by providing a holistic approach to e-commerce consumer analysis, combining technical rigor with practical applications to deliver actionable strategies for businesses navigating the competitive digital marketplace.

METHOD

For this study, we selected the *E-commerce Customer Behavior Dataset* from Kaggle. This dataset includes user IDs, demographic attributes (e.g., age, location), browsing sessions, product categories, order histories, and marketing touchpoints. Each record captures essential details such as transaction timestamps, order values, product subcategories, and any applied promotional codes.[14]

https://www.kaggle.com/datasets/salahuddinahmedshuvo/ecommerce-consumer-behavior-analysis-data.[14]

Data Analysis
Distribution of purchase across different income levels

This analysis is often visualized using bar charts, histograms, or pie charts, where the x-axis represents different income levels and the y-axis represents the total purchases or the number of transactions. This visual representation makes it easier to see patterns and trends in consumer behavior across different income groups. Analyzing the distribution of purchases across different income levels helps businesses and researchers understand consumer behavior, market dynamics, and economic conditions.
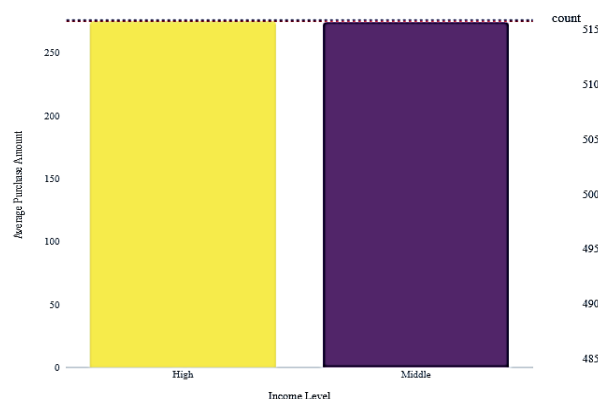

Figure 1. Purchase across different Income Level.[14]

This analysis helps businesses understand how discount sensitivity affects consumer spending behavior across different shopping environments, allowing them to make informed decisions about pricing and promotions.
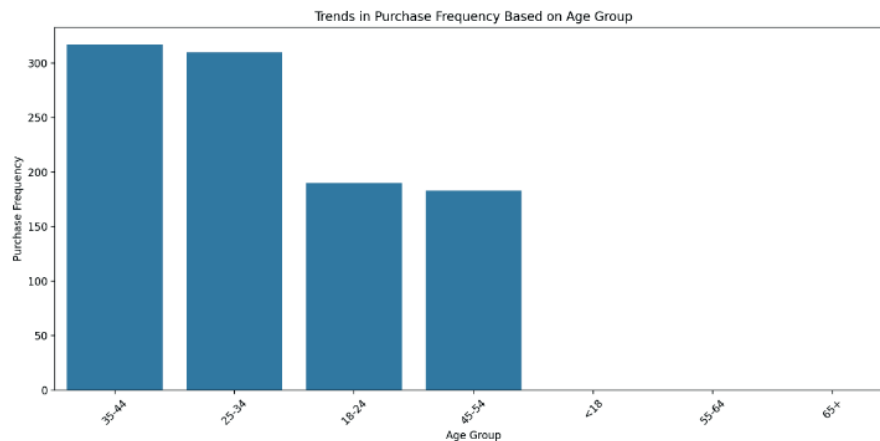

Figure 2. Purchase frequency based on age group. [14]

Customer satisfaction vs Time spent on product research (Correlation: 0,01)

The analysis has been successfully conducted, revealing a correlation coefficient of 0,01 between Customer Satisfaction and Time Spent on Product Research. This value indicates a slight positive relationship, implying that as customers spend more time researching products, their satisfaction levels may show a marginal increase. However, due to the extremely low correlation value, it can be concluded that there is no meaningful or significant relationship between these two variables. The correlation between Customer Satisfaction and Time Spent on Product Research is 0,01, reflecting a minimal positive association. While this suggests that increased research time might slightly enhance satisfaction, the near-zero correlation value confirms that the relationship between the two variables is negligible and lacks practical significance.
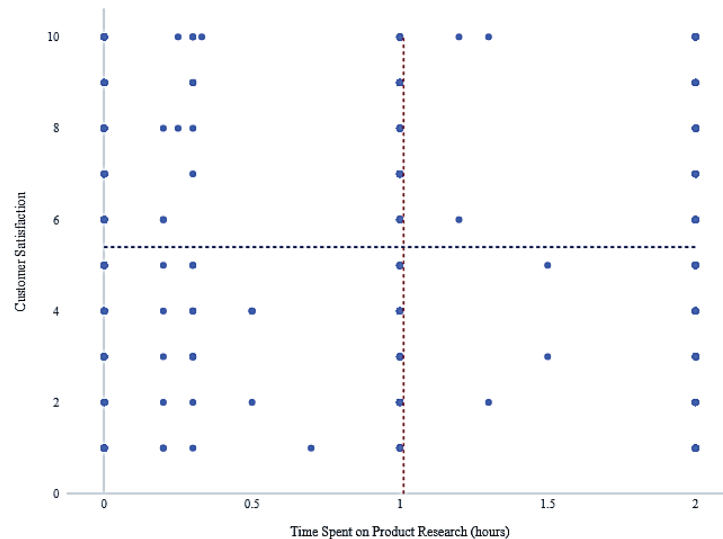

Figure 3. Time spent on product research. [14]

Impact of Discount Sensitivity on Purchase Amount Across Different Purchase Channels

When analyzing the impact of discount sensitivity on purchase amounts across different purchase channels. This analysis helps businesses understand how discount sensitivity affects consumer spending behavior across different shopping environments, allowing them to make informed decisions about pricing and promotions.

RESULTS AND DISCUSSION

Our findings reinforce established theories of online consumer behavior, underscoring the centrality of trust, convenience, and personalization (Gefen et al., 2003). [9] The strong correlation between frequency and retention aligns with loyalty-driven models that emphasize habit formation and brand affinity. Furthermore, the significance of session duration for electronics supports the concept that complex product categories

necessitate extended information searches before purchase (Venkatesh et al., 2012).[10] While the dataset is comprehensive, it may not fully represent global e-commerce markets or specialized niches (e.g., luxury items, digital goods). Additionally, the analysis relied on short-term data (18 months), potentially missing macroeconomic or competitive shifts over longer periods. Finally, the personal nature of e-commerce data raises privacy concerns, emphasizing the importance of secure data handling and compliance with regulations like the General Data Protection Regulation (GDPR).

Figure 4. Purchase Amount Across Different Purchase Channels.[14]

CONCLUSION

This research demonstrates how data-driven methods can unveil critical insights into e-commerce consumer behavior, using a Kaggle dataset as a case study. Through a comprehensive methodology—encompassing data cleaning, exploratory analysis, feature engineering, and predictive modeling—we identified key drivers of consumer loyalty, spending patterns, and product preferences. Ensemble models like gradient boosting and neural networks proved especially adept at forecasting repeat purchases, highlighting the sophisticated nature of online consumer interactions. The study further emphasizes the managerial value of customer segmentation and the importance of bridging data analytics with marketing strategies. While methodological limitations and privacy concerns persist, the ongoing refinement of machine learning techniques, coupled with robust data governance, promises to elevate e-commerce analytics to new heights. Ultimately, the paper underscores that a holistic, evidence-based approach to consumer analysis can significantly improve user engagement, boost revenue, and create more meaningful customer relationships in the digital marketplace.

REFERENCES

1.   Ekman P, Friesen W V. Facial action coding system. Environ Psychol Nonverbal Behav. 1978;

2.   Schuller B, Batliner A. Computational paralinguistics: emotion, affect and personality in speech and language processing. John Wiley & Sons; 2013.

3.   Zeng F, Hu Z, Chen R, Yang Z. Determinants of online service satisfaction and their impacts on behavioural intentions. Total Qual Manag. 2009;20(9):953–69.

4.   Scherer KR. What are emotions? And how can they be measured? Soc Sci Inf. 2005;44(4):695–729.

5.   Calvo RA, D'Mello S. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Trans Affect Comput. 2010;1(1):18–37.

6.   Cui Y, Shi Y, Gu M, Li J. Exploring the knowledge mapping and research trends of cross-border e-commerce. South African J Econ Manag Sci. 2025;28(1):11.

7.   Laudon KC, Traver CG. E-commerce 2019: Business, technology, society. Pearson; 2020.

8.   Brynjolfsson E, Mcafee A. Artificial intelligence, for real. Harv Bus Rev. 2017;1:1–31.

9.   Gefen D, Karahanna E, Straub DW. Trust and TAM in online shopping: An integrated model. MIS Q. 2003;51–90.

10.   Venkatesh V, Chan FKY, Thong JYL. Designing e-government services: Key service attributes and citizens' preference structures. J Oper Manag. 2012;30(1–2):116–33.

11.   Chen J, He Y, Frey EC, Li Y, Du Y. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. arXiv Prepr arXiv210406468. 2021;

12.   Aggarwal CC. Recommender systems. Vol. 1. Springer; 2016.

13.   Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 2020;12(3):254.

14.   Wamba SF, Dubey R, Gunasekaran A, Akter S. The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism. Int J Prod Econ. 2020;222:107498.

15.   Ecommerce Consumer Behavior Analysis Data [Internet]. Available from: https://www.kaggle.com/datasets/salahuddinahmedshuvo/ecommerce-consumer-behavior-analysis-data

FINANCING

CONFLICT OF INTEREST
None.

AUTHOR CONTRIBUTIONS
Conceptualization: Danish Anwar, Md. Faizanuddin, Faisal Rahman, Rajeshwar Dayal.
Investigation: Danish Anwar, Md. Faizanuddin, Faisal Rahman, Rajeshwar Dayal.
Methodology: Danish Anwar, Md. Faizanuddin, Faisal Rahman, Rajeshwar Dayal.
Drafting - original draft: Danish Anwar, Md. Faizanuddin, Faisal Rahman, Rajeshwar Dayal.
Writing - review and editing: Danish Anwar, Md. Faizanuddin, Faisal Rahman, Rajeshwar Dayal.