EDA on Consumer Behaviour Analysis in E-Commerce

Chikkeri Chinmaya
Dept. of Information Technology,
NIT Karnataka,
Surathkal, Mangalore
chikkerichinmaya.211it017@nitk.edu.in

Umesh
Dept. of Information Technology,
NIT Karnataka,
Surathkal, Mangalore
umesh.211it073@nitk.edu.in

Vishwa Mohan Reddy G
Dept. of Information Technology,
NIT Karnataka,
Surathkal, Mangalore
vishwa.211it082@nitk.edu.in

Vismay P
Dept. of Information Technology,
NIT Karnataka,
Surathkal, Mangalore
vismayp.211it083@nitk.edu.in

***Abstract*—The Exploratory Data Analysis (EDA) project delves into understanding customer behavior and predicting churn in an e-commerce setting using a comprehensive dataset. The project employs systematic data preprocessing, outlier detection, visualization techniques, and insightful analysis to derive actionable insights. It covers various aspects of the dataset, including numeric and categorical variables, aiming to uncover correlations between customer actions and churn rates. The analysis offers valuable findings, suggesting potential strategies to mitigate churn based on observed customer behaviors. The EDA project conducts a comprehensive analysis of an e-commerce dataset, aiming to unravel nuanced insights into customer behavior and potential factors influencing churn within the platform. Through meticulous data cleaning, identification of outliers, and in-depth visualizations using various statistical techniques, the study uncovers intriguing correlations between customer actions and the likelihood of churning. Key findings showcase that certain customer behaviors, such as engagement with push notifications and saving credit card information, exhibit a notable impact on reducing churn rates. These insights offer actionable strategies for enhancing customer engagement and trust, emphasizing the potential significance of personalized marketing approaches and user experience improvements to mitigate churn in e-commerce environments.**

***Index Terms*—Exploratory Data Analysis, E-commerce, Customer Behavior, Churn Prediction, Data Preprocessing, Outlier Detection, Visualization Techniques, Actionable Insights, Customer Engagement, Data Analysis, Business Strategy.**

I. INTRODUCTION

In today's competitive landscape of e-commerce, understanding and predicting customer behavior play pivotal roles in the sustained success of online businesses. This Exploratory Data Analysis (EDA) project embarks on a comprehensive investigation into customer actions and churn prediction within the realm of e-commerce platforms. Customer churn, referring to the phenomenon of customers discontinuing their engagement or usage of a service, poses a significant challenge for e-commerce entities. Identifying factors contributing to churn and deciphering patterns in user behaviors can offer valuable insights to strategize customer retention initiatives and improve overall business performance.

The proliferation of data in e-commerce domains presents an opportune moment to harness the power of data analytics in deciphering customer actions. This project aims to delve into a robust dataset sourced from an e-commerce platform, encompassing a diverse array of variables capturing customer interactions, preferences, and transactional behaviors. By leveraging systematic EDA techniques, including data preprocessing, outlier detection, and insightful visualizations, this analysis endeavors to unearth underlying correlations between customer behaviors and the propensity to churn. These findings aim to elucidate actionable strategies for businesses seeking to bolster customer engagement and reduce churn rates, thereby enhancing user retention and fostering sustainable growth in the competitive e-commerce landscape.

Understanding customer behavior dynamics in the context of e-commerce is pivotal for businesses to adapt and thrive in an ever-evolving digital marketplace. By delving into this rich dataset, this EDA project seeks to unravel patterns, unveil correlations, and distill actionable insights that can empower e-commerce enterprises to tailor strategies, optimize user experiences, and fortify relationships with customers, ultimately driving sustained business growth and success.

II. LITERATURE SURVEY

Customer behavior analysis within e-commerce has undergone significant evolution due to technological advancements and a deeper understanding of psychological, cultural, and technological factors shaping consumer choices. The following synthesis of research papers offers distinct insights into various facets of customer behavior analysis:

Charlie Custer (2023): Emphasizes the significance of customer behavior analysis in enhancing products, increasing conversions, and reducing churn. The paper covers process

overviews, tools, and case studies, focusing on effective analysis techniques such as audience segmentation and qualitative data validation [1].

Ahmad Ghandour (2015): Spotlights data-driven approaches in e-commerce, emphasizing the role of big data analytics and machine learning. Real-time analysis is highlighted for adapting to evolving customer behaviors, including applications like customer profiling and churn prediction [2].

Milad Zam et al. (2022): Explores the psychological factors impacting e-commerce transactions, identifying cognitive biases, trust, and decision-making processes as key influencers. The review calls for acknowledging psychological aspects to enhance user experience and conversion rates [3].

Artem Bielozorov et al. (2019): Explores the influence of user emotions on content personalization, discussing strategies like recommendation systems and targeted marketing to bolster customer satisfaction, sales, and loyalty [4].

Adam Vrechopoulos et al. (2003): Focuses on mobile commerce, exploring differences in user behavior between desktop and mobile platforms. Factors like screen size and user interface are identified as influencing mobile shopping behavior [5].

Federica Codignola et al. (2021): Takes a cross-cultural perspective, underscoring cultural influences on e-commerce customer behavior. The paper stresses understanding cultural nuances for businesses operating in global markets [6].

Erwin Halim and Rizal Haqo Karami (2020): Investigates the impact of social media influencers on e-commerce, discussing their influence on consumer choices, product discovery, and purchase decisions [7].

E.W.T. Ngai and A. Gunasekaran (2007): Provides a comprehensive review of mobile commerce research, highlighting specific nuances in customer behaviors on mobile platforms and the importance of optimizing the mobile shopping experience [8].

Fitore Jashar and Visar Rrustemi (2017): Explores the impact of social media on consumer behavior through a case study. Discusses how platforms like Instagram and TikTok influence consumer choices and brand perception [9].

Fei-Fei Cheng et al. (2019): Presents a cross-cultural review of user interface design's influence on consumer perceptions, emphasizing the need for cultural considerations in marketing strategies [10].

Krishna S. Dhir (2016): Focuses on consumer behavior analysis in the hospitality industry, exploring factors like customer service and online reviews in building brand loyalty [11].

Weng Marc Lim et al. (2022): Provides insights into emerging trends in consumer behavior analysis, discussing the impact of data analytics, AI, and digital transformation [12].

Navpreet Saini and Anupama Vashisht (2018): Explores trends and challenges in understanding consumer behavior in e-commerce, discussing topics like mobile commerce and pandemic-induced changes [13].

Ratih Hadiantini et al. (2021): Examines the importance of consumer satisfaction in e-commerce, highlighting elements like website design's influence on shopping decisions [14].

Zhang Xiaoyan and Thillai Raja Pertheban (2023): Presents a comparative study on factors affecting consumer trust in cross-border e-commerce [15].

Hong Huang et al. (2018): Focuses on consumer behavior analysis within mobile e-commerce apps, discussing the impact of mobile app design, push notifications, and in-app purchases [16].

This review provides a holistic understanding of customer behavior analysis in e-commerce, encompassing data-driven approaches, psychological factors, cross-cultural influences, social media impact, mobile commerce, and machine learning. Each paper contributes unique insights, forming a comprehensive overview of the evolving landscape of customer behavior in the digital era.

III. Methodology

A. Data Collection and Preparation

The initial step involved collecting a comprehensive dataset from an e-commerce platform, containing various customer-related variables such as transactional data, session duration, product interactions, location codes, and churn indicators. The dataset underwent rigorous data cleaning to handle missing values, ensure data consistency, and uniform formats. Additionally, columns were renamed, and categorical variables were encoded for subsequent analysis.

B. Exploratory Data Analysis (EDA)

The EDA process began with a thorough exploration of descriptive statistics within the dataset, utilizing summary statistics to comprehend the distribution of numeric variables and identify potential outliers. Various visualizations such as histograms, box plots, and heatmaps were employed to reveal patterns, detect outliers, and visualize relationships between variables. This comprehensive analysis facilitated a deeper understanding of the dataset's characteristics, uncovering insights into customer behavior metrics and their potential impact on churn.

C. Outlier Detection and Handling

Outliers were identified using statistical techniques such as interquartile range (IQR) and visualization-based methods across numeric features. Differentiating 'true outliers' representing natural variations from potential data errors was a priority, considering their relevance in the e-commerce domain. Contextual evaluation determined appropriate handling strategies.

D. Correlation Analysis and Churn Prediction

Correlation analysis was conducted to unveil potential relationships between customer behaviors (e.g., engagement with promotions, app usage) and churn rates. Statistical tests and visualizations determined the strength and significance of these relationships. Additionally, preliminary churn prediction models using machine learning algorithms were explored to forecast potential churn patterns based on identified features.

E. Deriving Actionable Insights and Recommendations

Insights from EDA, outlier analysis, and correlation studies were synthesized to derive actionable recommendations for e-commerce businesses. Strategies focusing on user engagement enhancement, marketing campaign optimization, user experience improvement, and building customer trust were proposed. These recommendations aimed to empower businesses in formulating targeted retention strategies to mitigate churn and cultivate lasting customer relationships.

IV. Results and Discussion

The EDA process revealed critical insights into customer behavior and its association with churn within the e-commerce dataset.

A. Data Collection and Preparation

The dataset retrieved from the e-commerce platform contained various customer-centric variables, including transactional data, session duration, product interactions, location codes, and churn indicators. Rigorous data cleaning ensured data integrity by handling missing values and standardizing data formats. Categorical variables were encoded for subsequent analysis, facilitating a comprehensive exploration of customer behaviors.

B. Exploratory Data Analysis (EDA)

Descriptive statistics unveiled valuable insights into the distribution of numeric variables. Visualizations, including histograms, box plots, and heatmaps, provided a detailed understanding of customer behavior metrics. Notably, the distribution of session duration exhibited outliers, with sessions below 1 minute considered potential outliers. Additionally, variables such as promotion clicks, average order value, and app transactions displayed significant variations, contributing to a deeper understanding of customer interactions.

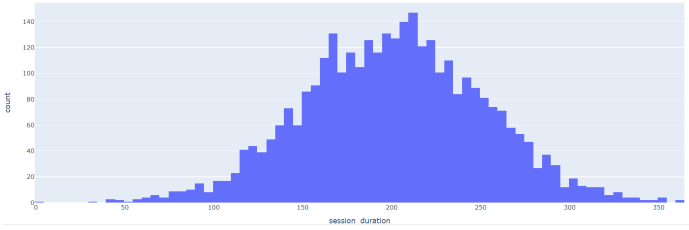

Fig. 1. Histogram representing Session Duration

C. Outlier Detection and Handling

The outlier detection process identified potential outliers across numeric features. Careful consideration distinguished 'true outliers' representing natural variations from potential data errors. Analysis revealed that certain extreme values, such as very high desktop sessions or significantly low average order values, represented genuine variations in customer behavior within the e-commerce environment.

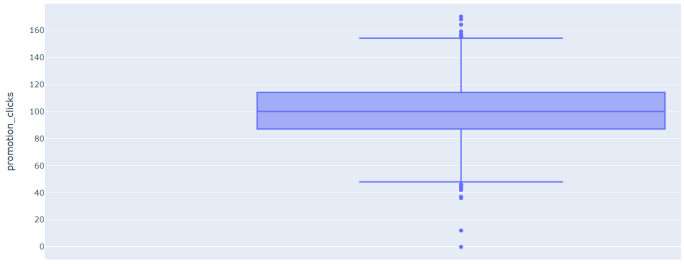

Fig. 2. Box Plot representing Promotion Clicks

The box plot indicates that anything above 155 and any data point below 46 is an outlier

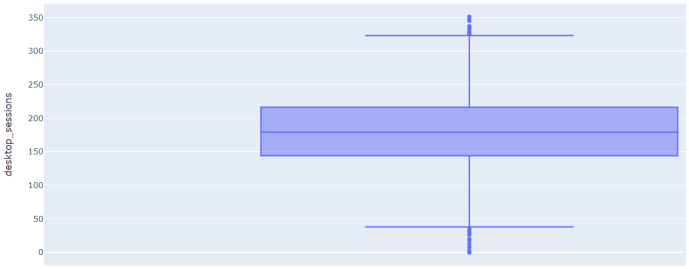

Fig. 3. Box Plot representing Desktop Session outlier

The box plot suggests that anything above 325 and anything below 35 is an outlier.

D. Correlation Analysis and Churn Prediction

Correlation analysis delved into relationships between customer behaviors and churn rates. Strong correlations were observed between reduced churn rates and specific behaviors, such as customer engagement with push notifications and saving credit card information. The analysis highlighted that customers with activated push status and saved credit card information exhibited lower churn rates, signifying the potential impact of these behaviors on customer retention.

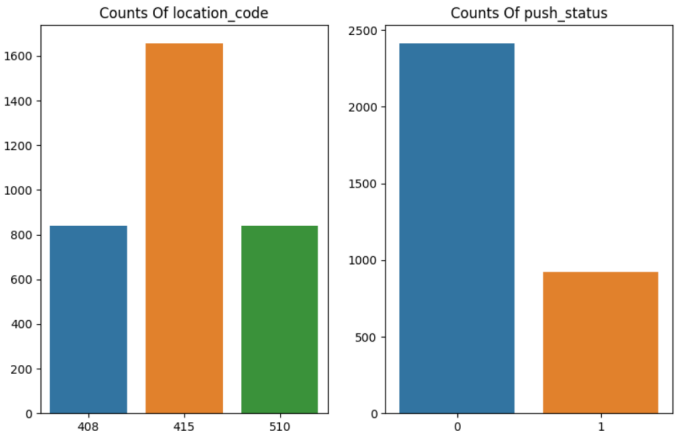

Fig. 4. Bar Graphs representing Churn vs Category

- Majority of customers are in location code 415.

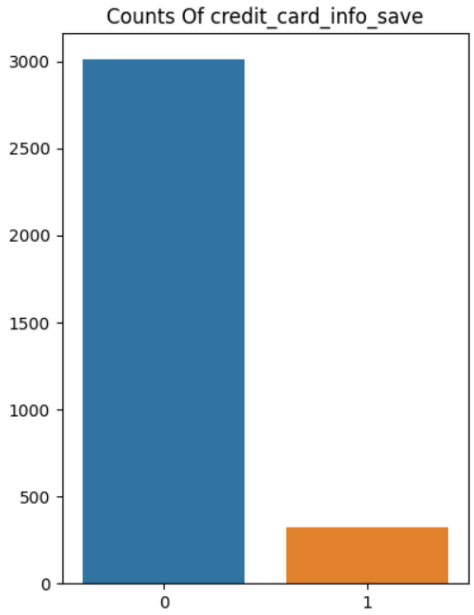

Fig. 5. Bar Graphs representing Churn vs Category (Second Graph)

- Majority of customers don't allow push notifications.
- Majority of customers don't store their credit card information on the site.

E. Deriving Actionable Insights and Recommendations

Synthesizing insights from the EDA, outlier analysis, and correlation studies led to actionable recommendations for e-commerce businesses. Strategies focusing on personalized marketing, improving user experiences, and encouraging customer trust through credit card information storage were proposed. These insights aimed to aid businesses in formulating targeted retention strategies to minimize churn and foster lasting customer relationships.

We've now taken a closer look at the relationship between the categorical variables and the dependent variable, churn. Here are some of the key observations from our analysis:

Location code doesn't seem to be a strong predictor of churn, as the percentage of customers who churn is similar across different locations. Customers with push notifications enabled are less likely to churn, suggesting that encouraging customers to activate push notifications may help reduce churn. The presence of saved credit card information on the site appears to be a strong indicator of a customer's likelihood to remain with the business. Customers who have saved their credit card information on the site are much less likely to churn than those who haven't. This correlation makes sense, as saving credit card information on the site demonstrates trust in the brand, which is likely to lead to greater customer loyalty. Overall, these insights provide valuable information for the business on how to minimize customer churn by focusing on certain customer behaviors.

V. Conclusion and Future Scope

The conducted Exploratory Data Analysis (EDA) on the e-commerce dataset unveiled profound insights into customer behavior and its correlation with churn rates. The systematic examination of various customer-centric variables unearthed crucial patterns, outliers, and correlations that can significantly influence business strategies focused on customer retention. Identification of key behaviors such as engagement with push notifications and saving credit card information as robust indicators of reduced churn rates underscores the criticality of targeted initiatives to bolster user engagement and cultivate trust within e-commerce platforms. Additionally, the nuanced differentiation between 'true outliers' depicting natural variations and potential data errors offered a deeper comprehension of customer behavior dynamics inherent within the dataset.

To propel the analysis further, future endeavors could concentrate on predictive modeling employing machine learning algorithms to anticipate churn probabilities based on identified customer behaviors. This predictive modeling approach would empower businesses to proactively address potential churn risks and implement preemptive strategies. Segmentation analysis could be another avenue for exploration, focusing on categorizing customers based on behavior patterns and demographics to craft tailored marketing strategies and refine customer experiences. Integration of time-series analysis to trace the evolution of customer behaviors over time and the integration of external data sources such as social media or market trends could offer a comprehensive view beyond platform-specific interactions. Furthermore, the implementation and A/B testing of proposed strategies derived from the analysis could validate insights and fine-tune strategies for optimal results.

This comprehensive EDA lays the groundwork for sophisticated analyses and actionable strategies aimed at enhancing user retention and fostering sustainable growth for e-commerce enterprises. The identified key behaviors and potential future avenues pave the way for a more nuanced understanding of customer dynamics and the implementation of data-driven strategies to mitigate churn and cultivate enduring customer relationships.

References

[1] Custer, C. (2023). Customer Behavior Analysis In 2023.
[2] Ghandour, A. (2015). Big Data Driven E-Commerce Architecture.
[3] Zam, M., Rezasoltani, A., Ramezanian, H., Tavakoli, M. (2022). Effects Of Psychological Factors On Customer Behavior In E-Transactions.
[4] Bielozorov, A., Bezbradica, M., Helfert, M. (2019). The Role Of User Emotions For Content Personalization In E-Commerce: Literature Review.
[5] Vrechopoulos, A., Sideris, I., Doukidis, G., Mylonopoulos, N. (2003). The Critical Role Of Consumer Behaviour Research In Mobile Commerce.
[6] Codignola, F., Capatina, A., Lichy, J., Yamazaki, K. (2021). Customer Information Search In The Context Of E-Commerce: A Cross-Cultural Analysis.
[7] Halim, E., Karami, R.H. (2020). Information Systems, Social Media Influencers And Subjective Norms Impact To Purchase Intentions In E-Commerce.
[8] Ngai, E.W.T., Gunasekaran, A. (2007). A Review For Mobile Commerce Research And Applications.

[9] Jashar, F., Rrustemi, V. (2017). The Impact Of Social Media On Consumer Behavior – Case Study.
[10] Cheng, F.-F., Wu, C.-S., Leiner, B. (2019). The Influence Of User Interface Design On Consumer Perceptions: A Cross-Cultural Comparison.
[11] Dhir, K.S. (2016). Analysis Of Consumer Behaviour In The Hospitality Industry: An Application Of Social Judgement Theory.
[12] Lim, W.M., Kumar, S., Pandey, N., Verma, D., Kumar, D. (2022). Evolution And Trends In Consumer Behaviour: Insights From Journal Of Consumer Behaviour.
[13] Saini, N., Vashisht, A. (2018). Trends in E-Commerce and Consumer Behaviour in Contemporary Era.
[14] Hadiantini, R., Silalahi, H., Hendrayati, R.H. (2021). The Importance Of Consumer Satisfaction For E-Commerce Users: How That Affects Consumer Behavior On Consumer Satisfaction?
[15] Xiaoyan, Z., Pertheban, T.R. (2023). Exploring the Factors Affecting Consumer Trust in Cross-Border E-commerce: A Comparative Study.
[16] Huang, H., Zhao, B., Zhao, H. (2018). A Cross-Platform Consumer Behavior Analysis of Large-Scale Mobile Shopping Data.