International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-June-2026, ISSN (Online): 2395-566X

Exploring Behavioural Patterns in Transaction Data: A Data-Driven Study

Mayuri Dongre [1], Arshiya Sahare[2] , Sarang Dumbhare [3]
[1] Department of Master in Computer Application, GHRCEM, Nagpur, India
Email: mayuri.k.dongre@gmail.com
[2]Department of Master in Computer Application, GHRCEM, Nagpur, India
Email: arshiya.sahare.mca@ghrcemn.raisoni.net
[3]Department of Master in Computer Application, GHRCEM, Nagpur, India
Email: mayuri.rangari@ghrua.edu.in

Abstract — In the age of digitalisation, a lot of transaction information is produced online, and it is significant to understand customer behaviour and market trends. This paper aims at examining the behavioural patterns in transaction data based on a data-driven approach. The information is gathered using web scraping on Flipkart, primarily in the electronic products categories of mobiles, headphones, smart watches, speakers, accessories with the help of Selenium WebDriver and Python. The obtained data is saved in the CSV format and processed with Python libraries, such as Pandas and NumPy, that allow cleaning data, eliminating duplicates, missing values, and categorizing products. Additional analysis is conducted to establish customer preferences, expenditure trends and product demand trends. The end results are presented in visual representations in the form of dashboards and reports to aid in improved business decision-making. This research assists in interpreting the behaviour of transactions and is useful in the data-driven strategies.

Keywords— Transaction Data, Behaviour Analysis, Web Scraping, Selenium, Python, Data Analytics, Customer Insights

I. INTRODUCTION

The In the modern fast-paced digital world, the use of online transactions has taken the form of a crucial aspect of conducting business. E-commerce systems produce a massive amount of data on transactions per second that holds valuable data on customer preferences, buying behaviors, and market trends. The analysis of this data assists the businesses to comprehend the customer behaviour and make effective strategic decisions.

Transaction information is comprised of product names, price, rating and reviews, categories and interaction between customers. This data when collected and analysed correctly can indicate significant behaviour patterns that aid in determining what products are demanded, how customers make buying decisions, and what influences their decisions. By monitoring the trends of repeated transactions, companies can determine customer areas of interest, detect abnormal behaviour and enhance the recommendation systems. This renders the analysis of tranThese are priceless lessons that could be applied to generate customer satisfaction and business performance.

A major aspect of modern data analytics is the analysis of behavioural patterns as it can help organisations to make predictions about the trends and understand consumer needs more.saction a valuable instrument in the research and industry.

As online shopping sites like Flipkart, Amazon, and other e-commerce sites continue to expand, Web scraping has become a viable way to gather real-time data associated with transactions.

Web scraping allows scraping of formatted data directly off of websites automatically, saving manpower and giving more data to play with. The primary tool in this project is selenium WebDriver that will be used to web scrape and gather information regarding transactions associated with products.

Electronic products that are collected comprise mobiles, headphones, smart watches, power banks, tablets, and speakers. These were chosen as these categories are popular consumer goods whose prices and customer preference differ and are commonly purchased. Through these classification, the project will examine product popularity and buyer behaviour in the electronics market.

Once data were collected, preprocessing is required to secure quality and accuracy of the data. Raw scraped data has many duplicates, missing values and inconsistent formatting. Data is cleaned, organized, and grouped into meaningful categories using Python libraries like Pandas and NumPy.This preprocessing phase enhances the accuracy of the analysis and prepares the data to explore more.

Once the data is cleaned, analytical techniques are applied in order to find out the behavioural patterns including their preferred type of products, their price, their rating and how many times they have posted a review. These lessons can be used to see how customers can relate to products and what factors can influence their purchasing behavior. This kind of analysis can help businesses to enhance pricing strategies and product planning.

The other aspect of this project that is of importance is visualization because complex data is easily interpreted through the use of graphics. The tools such as Power BI are utilized to design dashboards and charts that effectively show both the patterns of transactions and customer insights. These visual reports enable the decision-makers to easily determine trends and make relevant business decisions.

This research aims primarily to uncover behavioural patterns of transaction data using a fully data-driven approach, which encompasses web scraping, preprocessing, analysis, and visualization. This research adds to the comprehension of the customer behaviour in online shopping settings and shows that data analytics can be used to make wiser and more efficient business decisions.

II. LITERATURE REVIEW

A number of studies have also emphasized the value of the analysis of transaction data to comprehend client purchasing behaviour and market trends. According to Kumar and Reinartz (2018) [1], the data on transactions can provide valuable information about customer preferences, spending habits and decision-making tendencies which could be utilized by the business to improve their strategies and communicate with customers. The article by them points out how data analytics are becoming a more important part of modern business.

Web scraping has emerged as a popular method to gather real-time information on online websites. Mitchell (2018) [2] described that Selenium and BeautifulSoup are web scraping tools that help to extract structured information on websites in an automated manner. This technique is very useful in e-commerce studies to collect product information such as price, customer rating and reviews without involving human labor.

A study by Chen et al. (2019) [3] concentrated on behavioural pattern mining in online transaction systems and discovered that buying habits of customers could be forecasted based on the repetitive purchase behaviour. Their results indicated that the identification of these patterns assists organizations to predict the demand of their products as well as enhance inventory management systems.

Preprocessing of the data is another important phase in the analysis of transaction data as raw data is usually full of duplicates, incomplete records and discrepancies. According to Han, Kamber, and Pei (2017) [4], data cleaning and transformation enhance the quality of analysis and enhances the reliability of findings. They emphasize in their work the importance of preprocessing as the basis of proper data mining. Product categorization is a crucial component in the identification of market segments in the e-commerce analytics field. The research conducted by Aggarwal (2020) [5] has shown that the classification of products by category (mobiles, accessories, and audio devices) contributes to showing the trends and customer interests category-specifically more clearly. This enhances the behavioural analysis.

Transactions data also is greatly interpreted through visualization methods. Few (2019) [6] clarified that dashboards and graphical reports simplify the complex datasets and ease of understanding behavioural insights by the decision-makers. Power BI and other visualization tools are often used to display the pattern of transactions in the business world.

Another important area is anomaly detection in transaction behaviour. Chandala, Banerjee and Kumar (2009) [7] talked about how abnormal behaviour can be revealed through unusual transaction patterns which could be a sign of fraud risk, sudden changes in the market or abnormal behaviour. Such anomalies can be detected to help in the security of data and enhancement of monitoring systems.

In spite of the existing literature on transaction analytics, few studies have been conducted to incorporate web scraping, preprocessing, behavioural analysis and visualization into a single framework specifically in e-commerce product transaction data. This study fills that gap by formulating a full-fledged data-driven methodology on the study of behavioural trends in scraped transaction data on online shopping datasets.

III. METHODOLOGY

The approach taken by this study is a systematic data-driven approach to analyse behavioural patterns in transaction data obtained at e-commerce platforms. It begins by scraping the web, where products information, such as names, prices, ratings and reviews, are scraped out of Flipkart using Selenium WebDriver in Python. The data received is saved in a CSV format to be managed and further analyzed in the structured manner [8].

Data Collection and Storage: The scraped transaction data is categorized as datasets of product categories which include mobiles, headphones, smart watches, speakers, and accessories. CSV storage is also compatible with Python, SQL, and analytics to process it efficiently [9].

Preprocessing of the Data: Pandas and NumPy are used to clean the raw data by eliminating duplicates, processing missing data, and fixing the inconsistencies in the data. The products are then classified into meaningful groups to enhance the accuracy of analysis and identification of patterns [10]
.
Behavioural Analysis: Exploratory Data Analysis (EDA) is used to determine customer preference, pricing behaviour, popularity of a product and buying behaviour. This action helps in the discovery of hidden behavioural trends in the records of transactions [11].

Visualization and Insights: The processed results are visualized using Charts, graphs, dashboards with the assistance of Power BI and Python visualization libraries. Such visual interpretations aid in proper business decision making and forecasting of trends [12].

IV. KEY TECHNOLOGIES

This study incorporates some of the key technologies to gather, process, and analyze the data on transactions effectively. The technologies can be used to automate data extraction and data cleaning of raw data and generate value-based behavioural insights.

Python Programming Language: Python is the primary programming language to be used in this project due to its simplicity and good data analysis capabilities. It can be employed to handle web scraping, preprocessing, analysis, and visualization tasks [13].

Selenium WebDriver: Selenium WebDriver will be employed to carry out automated web scraping of Flipkart. It assists in the extraction of product information like names, prices, rating and reviews right out of dynamically-generated web pages [14].

Pandas and NumPy Libraries: Pandas and NumPy are used to clean data, fill in gaps, remove duplications, and organize data into clean formats that are simpler to analyze [15].

csv Data Storage: The data scraped will be saved as CSV file, which can easily and effectively manage large volumes of transaction data and can be easily integrated with analysis software [16].

Power BI / Visualization Tools: Visualization tools will be used to construct charts, graphs and dash boards that will easily display behavioural patterns and customer insights that will be interpreted more effectively [17].

Future Scope

The Machine learning implementation:

In future this research can be enhanced by using more sophisticated machine learning and artificial intelligence tools to narrow down the accuracy of prediction of behaviour patterns. Some of the algorithms that could be used to improve on identifying customer preferences are predictive analytics, clustering and recommendation systems. The technologies will enable companies to predict buying patterns and make wiser data-driven choices [18].

Multi-Platform Data Analysis:

The project is now based on the Flipkart transaction data, but it can be extended to various e-commerce platforms in future including Amazon, Myntra, Snapdeal and more. The comparison of the transaction behaviour on various platforms will give a wider market view and will assist in the analysis of the preferences of customers on a larger scale. This expansion will improve the overall usefulness of the research [19].

Real-Time Transaction Monitoring:

Real-time transaction monitoring systems can be adopted as one of the biggest improvements that can be made in the future. It is possible to upgrade the system to allow live stream of transactions rather than merely analysing historical data that is stored. This will assist businesses to identify abnormal buying behavior, abrupt demand variations and anomalies related to fraud more effectively in real time [20].

Sentiment Analysis Customer Review:

Another important future enhancement is the addition of Natural Language Processing (NLP) for sentiment analysis of customer reviews. Through text reviews, the system will be able to comprehend customer opinions, level of satisfaction and emotional feedback on products. This will give more insights into behaviour than the numeric transaction information [21].

Interactive BI Dashboard:

Through the project, it can also be developed into a fully functioning intelligent business dashboard where automated reports and dynamic graphs as well as trend forecasting can be generated in real time. These dashboards may aid managers and

businesses to make fast decisions, strategic planning, and forecasting the market more effectively [22].

Scalability and Growth of Big Data:
The system can be built in future to manage much larger datasets with cloud computing and big data technologies like Hadoop or Spark. This will render the model scalable to enterprise level transaction analysis where millions of records are handled a day across world markets [23].

V. CONCLUSION

This paper demonstrates how a data-driven approach can be effectively used to learn the behavioural trends in transaction data that is collected in e-commerce websites. Webscraping with Selenium and Python allowed extracting pertinent product information efficiently and Pandas and NumPy served to clean and process the data and categorize it to analyze it properly. Exploratory data analysis methods enabled the study to determine the critical trends in customer preferences, product demand, and pricing trends [23].

The results of this study are very insightful on customer buying behaviour and the significance of analysing data of transactions in business decisions. The implementation of visualization tools also contributed to a better understanding of complex data sets and findings were more comprehensible and feasible. In general, this project demonstrates that the integration of web scraping, preprocessing, analysis, and visualization can assist any business in making wiser, data-driven decisions in the future about its further development [24].

REFERENCES

1. V. Kumar and W. Reinartz, Customer Relationship Management: Concept, Strategy, and Tools, 3rd ed. Berlin, Germany: Springer, 2018
2. R. Mitchell, Web Scraping with Python: Collecting More Data from the Modern Web, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2018.
3. Y. Chen, X. Liu, and J. Zhang, "Behavioral Pattern Mining in E-commerce Transactions," International Journal of Data Analytics, vol. 6, no. 2, pp. 45–53, 2019.
4. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2017.
5. C. C. Aggarwal, Data Mining: The Textbook. Cham, Switzerland: Springer, 2020.
6. S. Few, Information Dashboard Design: Displaying Data for At-a-Glance Monitoring, 2nd ed. Burlingame, CA, USA: Analytics Press, 2019.
7. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," ACM Computing Surveys, vol. 41, no. 3, pp. 1–58, 2009.
8. SeleniumHQ, "Selenium WebDriver Documentation," [Online]. Available: https://www.selenium.dev/documentation/
9. Flipkart Official Website, "E-commerce Product Data Source," [Online]. Available: https://www.flipkart.com/
10. W. McKinney, Python for Data Analysis, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2018.
11. T. VanderPlas, Python Data Science Handbook. Sebastopol, CA, USA: O'Reilly Media, 2016.
12. Microsoft, "Power BI Documentation," [Online]. Available: https://powerbi.microsoft.com/
13. Python Software Foundation, "Python Documentation," [Online]. Available: https://www.python.org/doc/
14. Beautiful Soup Documentation, "Beautiful Soup 4," [Online]. Available: https://www.crummy.com/software/BeautifulSoup/
15. NumPy Developers, "NumPy User Guide," [Online]. Available: https://numpy.org/doc/
16. Pandas Documentation, "Pandas User Guide," [Online]. Available: https://pandas.pydata.org/docs/
17. QuillBot, "Paraphrasing and Writing Assistance Tool," [Online]. Available: https://quillbot.com/
18. IEEE Author Center, "IEEE Paper Formatting Guidelines," [Online]. Available: https://journals.ieeeauthorcenter.ieee.org/
19. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.
20. C. C. Aggarwal, Data Mining: The Textbook. Cham, Switzerland: Springer, 2015.
21. F. Chollet, Deep Learning with Python. Shelter Island, NY, USA: Manning Publications, 2017.
22. A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.
23. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, vol. 41, no. 3, pp. 1–58, 2009.
24. S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.