

A Statistical Analysis of Customer Behaviour and Sales Patterns in Organized Retail: A Study of D-Mart and Supermarket in Jaysingpur City.

Riza Naushad Bagwan¹, Shraddha Namdev Tejam¹, Prachi Suresh Borgave¹,
Vishwamohini Yuvraj Jadhav¹, Gitanjali A. Gadhari²

Students, Department of Statistics, Rajashri Chh. Shahu College, Kolhapur, India¹

Assistant Professor, Department of Statistics, Rajashri Chh. Shahu College, Kolhapur, India²

rizabagwan4@gmail.com¹, shraddhatejam@gmail.com², prachiborgave23@gmail.com³,

vishwamohinijadhav@gmail.com⁴, gitanjaligadhari01@gmail.com

Abstract: *With a competitive retail market, it is incredibly important to understand the behaviours of customers in order to improve their overall satisfaction and the sales performance of our stores. The goal of this study is to understand customer behaviours and sales trends that occur in a supermarket or smart store in Jaysingpur city focusing specifically on D-Mart and local supermarket. The data used was gathered from customers via a structured questionnaire prepared using Google Forms. Different statistics and analysis of the data were conducted, such as graphical representation, clustering, market basket analysis, and RFM (Recency Frequency and Monetary). The goal was to identify purchasing patterns and segment customers based on their buying behaviour, shopping frequency, preferred product types, and preferred spending ways. By identifying these key factors, retailers will have a better understanding of what customers want, which can have a positive impact on the development of marketing strategies, product placements, and promotional activities. The results of this study can provide the management of supermarket with the means to make decisions based on facts rather than hypothetical assumptions, ultimately allowing for more satisfied customers and increased sales performance.*

Keywords: Customer Behaviour, Sales Analysis, Supermarket Retailing, Market Basket Analysis, RFM Analysis, Clustering Techniques, Consumer Purchasing Patterns.

I. INTRODUCTION

A. Background of the study

Today's retail industry is extremely competitive, making it crucial for retailers to understand their sales performance and how customers behave when shopping in order to stay ahead of the competition [2][3]. Retailers need to analyze customer purchasing patterns in order to attract, keep and satisfy their customers [1]. In this regard, D-Mart and supermarket serve an important function in organized retailing in providing large assortments of consumer goods under one-roof, which allows organized retailers to have a better opportunity to measure consumer buying behaviour [5].

Organized retailing has greatly changed how consumers purchase their everyday necessities in India. D-Mart and other large retailers such as supermarket offer a large selection of merchandise, such as groceries, household goods, clothing, and other consumer goods at competitive prices [6]. In Jaysingpur City, D-Mart serves as a leading retail provider by supplying customers with convenience and cost savings through one-stop shopping, which makes shopping much more time efficient and has contributed to changing consumer buying habits [3].

Shopping behaviour of customers at retail outlets is influenced by various factors like demographics, purchasing behaviours, store design, promotions and total shopping experience [9]. Furthermore, the growing use of technology in the form of digital payment methods, self-checkout technological advancements and data driven marketing have also



changed the nature of retailing as well as influenced customer behaviour. Retailers must continually respond by developing strategic changes to ensure continued competitiveness and profitability.

This research will study customer behaviour and performance in D-Mart and supermarket located at Jaysingpur City. The purpose of this research is to examine shopping behaviours, preferences for purchasing and sales patterns as ways to better understand how customers affect retail performance. Results of this research will assist retailers/managers with improving operational processes such as inventory availability, store design, quality of service and marketing techniques.

B. Problem statement

The growth of organized retailing is growing at a rapid pace in India; however, retailers are still facing some challenges in understanding their customers' behaviour and in making decisions about how they're using the sales data that is produced. Retailers must be able to identify shopping patterns, determine the different customer groups that buy from them, and find out what types of relationships exist between the products that they sell and how their customers purchase those products, in order to improve their sales performance and increase the satisfaction level of their customers.

In the city of Jaysingpur, both the Super Market and D-Mart have consistently attracted large amounts of customers, yet there has been a lack of research conducted with regards to how consumer demographics, consumer purchasing habits, and consumer experiences in the store effect sales outcomes. Retailers that do not have a thorough understanding of consumer demographics, consumer purchasing habits, and consumer experiences in their stores may find it very difficult to create effective marketing strategies, manage their inventories effectively, and improve their customer engagement; therefore, in order to identify the key elements that are affecting retail performance, it is important for retailers to analyze both customer behaviour and sales and to identify the key factors, which are driving retail performance.

The aims of our study are to determine busy days/times for customers; this can help retailers better understand how customer shopping trends affect businesses ,To look at the association rules related to products purchased together, **To identify distinct customer segments based on demographics and spending behaviour for targeted marketing and strategic decision-making** , Identify customers that are loyal, provide high-value purchases, and are at risk of losing and to analyze demographics and purchasing patterns with the Chi-Square test.

LL. LITERATURE REVIEW

Retail data analysis helps supermarket and retailers understand consumer behaviour and increase profitability through improved sales efficiencies. Statistical and data-mining techniques, such as pattern matching using graphical analysis, and listing data in a specified manner using clustering techniques, are used by many researchers to identify unique consumer purchase patterns and different consumer segments. For example, visual representations are a common method used within retail analytics to help store owners visualize and comprehend their customers' purchasing behaviours, as well as the trends of sales. According to Montgomery, Peck, and Vining (2021), plots such as bar graphs, pie charts, and related graphical representations are beneficial to the efforts of summarizing relatively large datasets, as well as identifying points or events that may have occurred in relation to those datasets.

The use of customer segmentation, through clustering techniques, is another relatively new development in marketing research that helps retailers achieve the above goals. Research conducted by Sinha & Jain (2015) included the use of the K-Means clustering algorithm to identify customer insights and classify customers according to their purchasing behaviours and demographic characteristics. In each case, clustering was used to identify which segments offered high (high-value), medium (moderate), and low (low) contributions to retailers, thereby aiding retailers in developing their targeted marketing strategies and driving customer retention rates.

Association rule mining is another critical technique for analyzing product purchase relationships known as Market Basket Analysis. An example of using this technique is performed by Shelke, Dharaskar and Thakare (2017) who utilized association rule mining on supermarket transaction data to identify which products tend to be purchased together. Retailers can leverage market basket analysis to determine the combinations of products that are commonly bought and



then use this information to improve store layout, cross-selling strategies and promotional offers based on the results obtained by calculating support, confidence and lift values.

RFM (Recency, Frequency, Monetary) analysis is a common technique for segmenting customers and analyzing customer loyalty. Fader, Hardie and Lee (2005) indicated that through RFM analysis, a retailer can identify their customer segments including loyal, high-value and at-risk customers by analyzing the behaviour of customers regarding their purchases. The use of RFM analysis enables retailers to design targeted marketing strategies for each customer segment and, in turn, help retain their customers.

Chi-Square tests are widely used in retail research as a statistical method to evaluate the relationship between two categorical variables. For example, Agresti (2018) reports Chi-Square tests can help analyse the extent to which variables (age group, shopping frequency, and promotional purchases) can be shown to be statistically related.

Overall, the previous literature has shown that statistical methods and data mining techniques, such as graphical representation, clustering algorithms, market basket analysis, RFM analysis and Chi-Square tests, can be effective tools in understanding the behaviour of customers and developing effective sales strategies in the retail environment.

III. DATA AND METHODOLOGY

A. Research Design

This study employed a cross-sectional research design, which is where the researcher collects data from the sample at one specific point in time. The purpose of the study is to assess customer behaviour and sales trends for both D-Mart and one of the local supermarket in Jaysingpur city. Specifically, the researcher wanted to know the shopping habits of customers, their spending habits, their preferred stores, as well as other factors that influence their purchasing decisions.

B. Study Area and Population

This study was conducted in the city of Jaysingpur, Kolhapur district, Maharashtra. The target population is made up of customers who shopped at D-Mart and one of the local supermarket during the survey.

According to information from both stores, there were approximately 1400 customers total for both stores; 800 were customers of D-Mart and 600 were customers of the local supermarket.

C. Sample Size Determination

The sample size for this study was calculated using Yamane's formula (1967). The formula is stated below:

$$N = 1400 \text{ (the total number of customers surveyed)}$$

$$e = 0.05 \text{ (the margin of error).}$$

Based on this formula, a sample size of 311 respondents would be required; however, in order to increase reliability and representation, the researcher collected 400 survey responses.

D. Data Collection

Primary data collection for this study includes visitors at D-Mart and a selected supermarket of Jaysingpur through a structured questionnaire introduced through Google Forms and personally administered. Data collection occurred over a 7-day period to ensure customers' responses were gathered within a set length of time and at consistent intervals. In order to accurately reflect the changes in shopping behaviours at various times throughout any given day, the survey was administered over 4 different daily shifts which are, morning, afternoon, evening and night.

Random samples of customers visiting the stores during each shift were solicited for participation. All respondents who were approached were fully informed regarding the nature of the study prior to participating. Provided in the questionnaire were questions regarding customer demographics, shopping frequency, average spending amounts, overall customer satisfaction with their selected store, product categories purchased and the specific days/times of their visitation to the store.



Utilising different data collection techniques and data collection across both multiple days and shift times resulted in a diverse sample of respondents thus providing a greater degree of reliability and representivity of the data collected for the purpose of conducting statistical analyses.

E. Data Processing

The collected data were pre-processed, including cleaning and more broadly organizing them into many different software programs such as Microsoft Excel and Python. In brief, the many steps in this process included:

- Missing data were addressed
- Any incomplete responses were deleted
- Coding categorical types of data
- Making categorical data into one-hot encoded categorical variables for use with machine learning systems
- Identifying outliers using statistical techniques

F. Statistical Analysis

The collected data were then analyzed using the appropriate statistical and machine learning techniques as follows:

- Descriptive statistics (i.e., mean, frequency distribution, and percentage distribution)
- Chi-square testing for measuring the relationship between two categorical variables
- K-mode Clustering to group customers together based on similar demographic characteristics
- Association Rule Mining to determine the combination of products purchased together and how often this takes place
- RFM Analysis to classify customers based on recency, frequency, and monetary value of their purchasing patterns

G. Software used

- The analysis used many different software tools, which include:
- Python for statistical modelling and machine learning
- Microsoft Excel for data cleaning as well as basic analytical processes of the data

IV. GRAPHICAL ANALYSIS

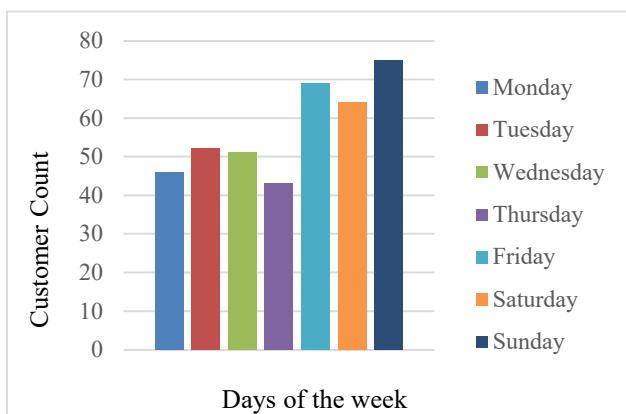


Fig 1. Day-wise customer count

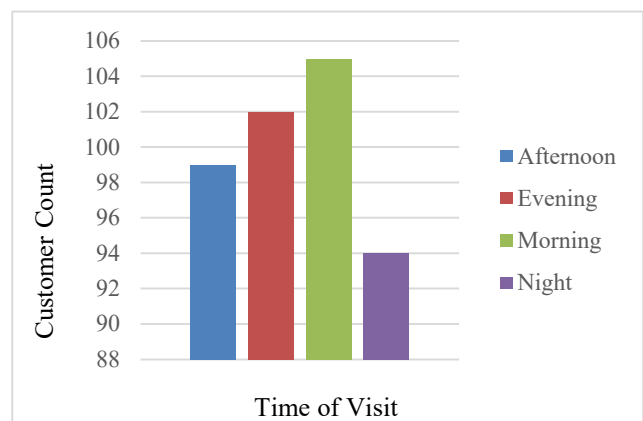


Fig 2. Time-wise customer count



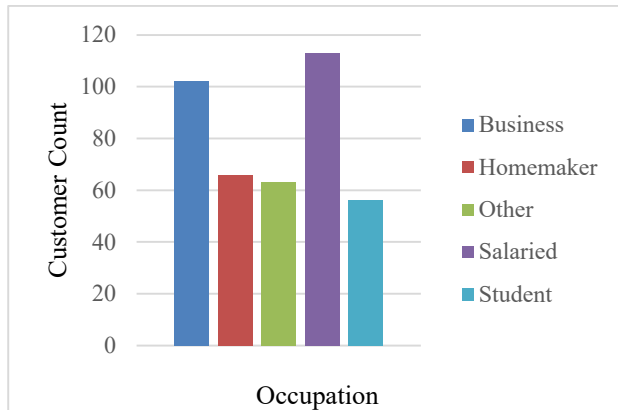


Fig.3 Customer Count by occupation

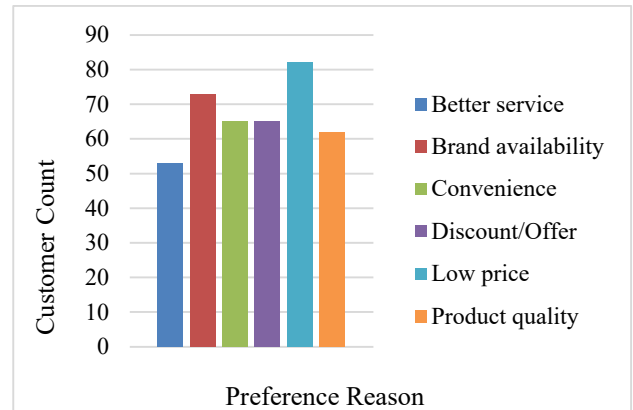


Fig.4 Reason for preference

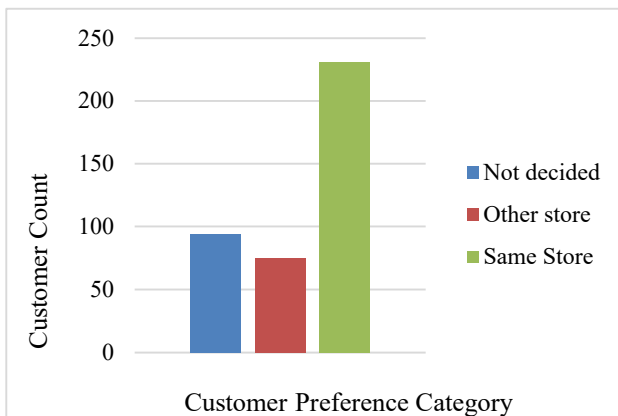


Fig. 5 Distribution of customers based on store preference

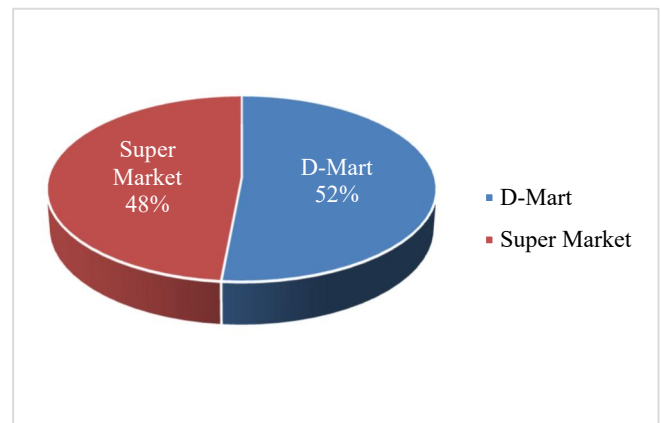


Fig. 6 Preference of shopping place

Fig. 1 shows that, the customer visits are highest on Sunday with 75 customers and lowest on Thursday with 43 customers, indicating peak activity on weekends.

Fig. 2 shows that, the store is busiest in the Morning with 105 customers and least crowded at Night with 94 customers, showing a slight preference for early visits.

Fig. 3 shows that, the store is most popular among Salaried individuals, while Students make up the smallest portion of the customer base.

Fig. 4 shows that, low price is the main reason customers prefer the store, followed by brand availability. Better service is the least selected reason compared to the other factors.

Fig. 5 shows that, the most customers prefer to continue shopping at the same store, showing strong customer loyalty. Only a smaller number of customers are undecided or prefer switching to another store.

Fig. 6 shows that, the D-Mart are preferred by 52% of respondents, slightly more than Supermarket by 48% of respondents, showing nearly equal preference.



V. STATISTICAL ANALYSIS

A. Association Rule

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an item set occurs in a transaction.

An association rule is an implication of the form, $X \Rightarrow Y$

“If X then Y” where X & Y are item set. X is called the antecedent and Y is called the consequent.

Item set: Set of attributes that are together in a transaction.

Support(S)= P(a transaction contains X and Y) here, support is 0.002

Confidence(C) = P(a transaction containing X also contains Y) here, confidence is 0.7

Lift: Lift measures how much more likely the consequent is purchased when the antecedent occurs compared to random chance.

Table 1: Association Rule

Antecedents	Consequents	Support	Confidence	Lift
(Packaged food, Cleaning, Household)	(Clothing)	0.0200	0.888889	1.964395
(Personal care, Beverages, Packaged food)	(Household)	0.0200	0.888889	1.871345
(Beverages, Other)	(Personal care)	0.0225	0.818182	1.628223
(Electronics, Cleaning, Household)	(Personal care)	0.0225	0.818182	1.628223
(Personal care, Cleaning)	(Beverages)	0.0375	0.789474	1.611171
(Groceries, Electronics, Clothing)	(Packaged food)	0.0250	0.769231	1.493652
(Electronics, Personal care, Cleaning)	(Household)	0.0225	0.750000	1.578947
(Cleaning, Electronics)	(Personal care)	0.0300	0.750000	1.492537
(Household cleaning, Clothing)	(Packaged food)	0.0225	0.750000	1.456311

From Table 1, the analysis reveals that Packaged food, Cleaning, and Household items are the strongest predictors of other purchases. Customers who buy these three items together are 89% likely to also buy Clothing, as shown by the highest confidence in the table. This indicates a strong link between buying daily home essentials and fashion items in a single trip.

B. Customer Segmentation

Clustering is a popular and effective unsupervised learning method for market segmentation. It groups customers into k clusters based on similarities, using centroids to represent each cluster. This helps marketers identify distinct customer segments for targeted strategies. We use K-Mode clustering to group customers based on Age, Gender, Amount Spent, and Monthly Income.



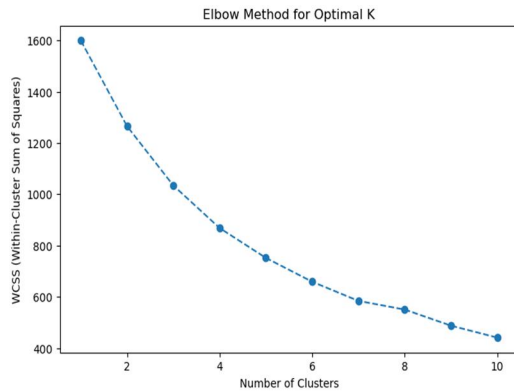


Fig. 7 Elbow method for optimal cluster

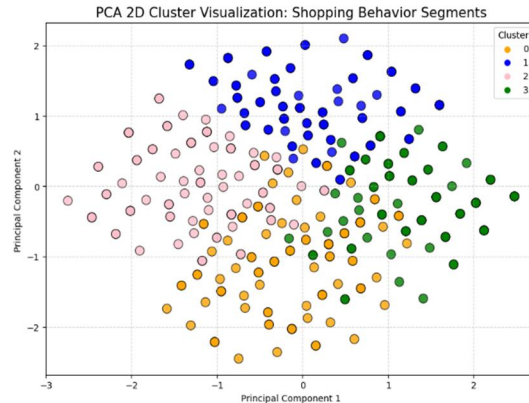


Fig. 8 PCA cluster visualization

Fig. 7 shows that, optimal number of clusters for segmenting customers based on Age, Gender, Amount Spent, and Monthly Income is $k = 4$.

Fig. 8 shows that, different groups (clusters) based on their characteristics.

Table 2: Customer Segment Profile

Cluster	Gender	Age Group	Monthly Income	Average Spending
0	Male	46–60	10000–30000	<500
1	Female	35–45, 60+	80000+	3500+
2	Female	46–60	<10000	<500
3	Male	26–35, 60+	80000+	2500–3500, 3500+

From Table 2,

Cluster 0: This group mainly contains male customers aged 46–60 with low to medium income (₹10,000–₹30,000). Their average spending is very low (below ₹500). These customers are low spenders.

Cluster 1: This cluster mostly includes female customers aged 35–45 and above 60 with very high income (₹80,000+). Their spending is also very high (above ₹3500). These customers are high-value customers.

Cluster 2: This group contains female customers aged 46–60 with very low income (below ₹10,000) and very low spending (below ₹500). These customers are budget or low purchasing customers.

Cluster 3: This cluster mainly has male customers aged 26–35 and above 60 with very high income (₹80,000+). Their spending is high (₹2500 – ₹3500 or more). These customers are high-income active spenders.

C. Recency Frequency Monetary (RFM) Analysis

RFM analysis provides valuable customer insights by segmenting them based on Recency, Frequency, and Monetary value of their purchases, revealing who your best customers are (high R, F, M scores) and predicting future behaviour, allowing for personalized marketing, better resource allocation, and improved retention by identifying loyal customers and those at risk of churning.

Recency (R): How recently did a customer buy? (Recent buyers are more likely to buy again).

Frequency (F): How often do they buy? (Frequent buyers are more loyal).

Monetary (M): How much do they spend? (Higher spenders contribute more revenue).



By scoring our customer by recency, frequency and monetary value, we able to identify 3 major segments for targeting. Fig. 9 shows that, most customers belong to the Medium Value segment (41%), which means many customers have average purchasing behaviour. About 33% of customers are Low Value customers, indicating that a large group of customers spend less or purchase less frequently. Only 26% of customers are High Value customers, but these customers are very important because they spend more and contribute more to revenue.

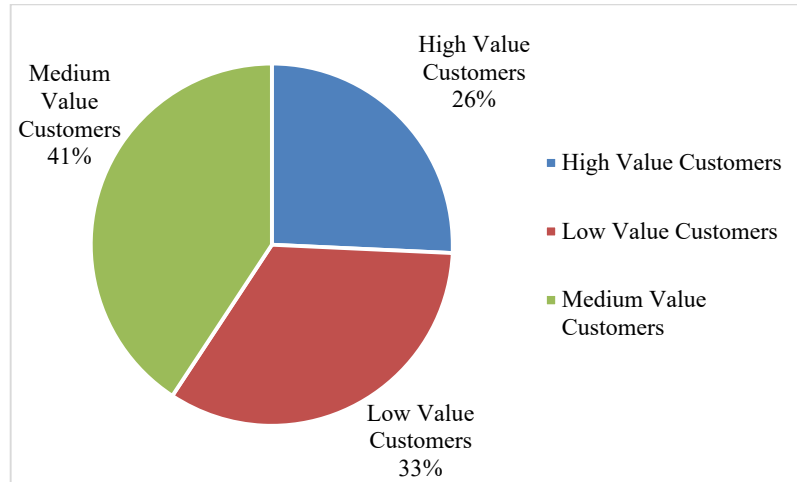


Fig. 9 Customer types distribution

D. Chi- square Test

1. To check the dependency between Age and Shop Frequency.

H_0 : There is no statistically significant association between Age and Shop Frequency.

H_1 : There is a statistically significant association between Age and Shop Frequency.

Under H_0 , Chi-Square Statistic: 26.701, p-value: 0.04493 and Significance Level: 0.05

Conclusion:

Since, P-value < 0.05, we reject H_0 .

Therefore, there is significant difference between Age and Shop Frequency.

2. To check the dependency between Promo Purchase Frequency and Average Spending.

H_0 : There is no statistically significant association between Promo Purchase Frequency and Average Spending.

H_1 : There is a statistically significant association between Promo Purchase Frequency and Average Spending.

Under H_0 , Chi-Square Statistic: 20.394, p-value: 0.00894 and Significance Level: 0.05

Conclusion:

Since, P-value < 0.05, we reject H_0 .

Therefore, there is significant difference between Promo Purchase Frequency and Average Spending.

VI. OVERALL CONCLUSION

A study that uses various statistical methods in analyzing customer behaviour and sales performance for both D-Mart and other supermarket. Customer visits are recorded to be the highest on Sundays and lowest on Thursdays; therefore there is more shopping activity on weekends. Morning is the time slot that has the most customers, while most customers are salaried employees and the least number of customers are students. The primary reason for shopping at the D-Mart and super market is because of low prices, while brand availability is cited as a reason for being at the store. Customer loyalty appears to be strong as there are many customers who will continue to shop in the same store. Association rule analysis indicates that customers who buy packaged food, cleaning, and household items tend to buy clothing items.



Customers are grouped into three clusters as low spenders, high value & budget customers. RFM a general rewrite using a different tone but maintaining the same meaning as the original text. There is a significant relationship between the age of customers and their shopping frequency, and between the frequency of promotional purchases by customers and their average spending amount.

REFERENCES

- [1] Khan, S. (2018). Factors affecting consumer buying behaviour and consumer preference towards organized retail outlets in India. *Journal of Advances and Scholarly Researches in Allied Education*, 15(5), 202-211.
- [2] Ahmad, Z. (2020). A study of the changing consumer buying behaviour in organised retailing in Lucknow city. *International journal of engineering technologies and management research*, 5(2), 92-99. .
- [3] Agarwal, A., & Singh, M. R. (2018). The relationship between retail experience, customer satisfaction, and behavioural intention: Exploring the consumer shopping behaviour in unorganized retail settings. *Indian Journal of Marketing*, 9-27. <https://doi.org/10.17010/ijom/2021/v51/i5-7/161647>.
- [4] Ranjan, R., & Singh, N. (2025). A Study on Consumer Behaviour towards Organised Retail Stores in Bhopal City. *International Journal of Innovations in Science, Engineering And Management*, 409-415. <https://doi.org/10.69968/ijsem.2025v4i2409-415>
- [5] Stylianou, T., & Pantelidou, A. (2025). A machine learning approach to consumer behaviour in supermarket analytics. *Decision Analytics Journal*, 16, 100600. <https://doi.org/10.1016/j.dajour.2025.100600>
- [6] Alfian, G., Octava, M. Q. H., Hilmy, F. M., Nurhaliza, R. A., Saputra, Y. M., Putri, D. G. P., ... & Syafrudin, M. (2023). Customer shopping behavior analysis using RFID and machine learning models. *Information*, 14(10), 551. <https://doi.org/10.3390/info14100551>
- [7] Lin, Y. C., Sung, B., & Park, S. D. (2024). Integrated systematic framework for forecasting China's consumer confidence: A machine learning approach. *Systems*, 12(11), 445. <https://doi.org/10.3390/systems12110445>
- [8] Hanif, A., & Noor, S. (2025). Machine Learning Analysis of Consumer Spending. *Journal of Social Impact Studies*, 3(1), 83-96.
- [9] Hossam, A., Ramadan, A., Magdy, M., Abdelwahab, R., Ashraf, S., & Mohamed, Z. (2024, May). Revolutionizing retail analytics: advancing inventory and customer insight with AI. In *2024 International Conference on Machine Intelligence and Smart Innovation (ICMISI)* (pp. 64-69). IEEE. DOI: [10.1109/ICMISI61517.2024.10580424](https://doi.org/10.1109/ICMISI61517.2024.10580424)
- [10] Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of marketing research*, 42(4), 415-430. <https://doi.org/10.1509/jmkr.2005.42.4.415>
- [11] Hughes, A. M. (2005). *Strategic database marketing*. McGraw-Hill Pub. Co..
- [12] Stylianou, T., & Pantelidou, A. (2025). A machine learning approach to consumer behavior in supermarket analytics. *Decision Analytics Journal*, 16, 100600. <https://doi.org/10.1016/j.dajour.2025.100600>
- [13] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

