Journal of Business Thought, *DOI: 10.18311/jbt/2024/44468, Vol 15, 2024*

A Study on Micro-Segmentation of Retail Customers Using K-Means Clustering

Divya Mehta[1] and Sanjeewani Sehgal[2*]

[1]Shaheed Bhagat Singh College, University of Delhi, New Delhi – 110017, India; divyamehta@sbs.du.ac.in
[2]Cluster Innovation Centre, University of Delhi, New Delhi – 110007, India; ssehgal@cic.du.ac.in

Received: 12 June 2024 **Accepted:** 30 September 2024 **Published:** 18 December 2024

Abstract

This study aims to identify clients that share similar traits and develop a new micro-segmentation strategy. Drawn from two marketing theories i.e., customer relationship and personalisation and using the Recency Frequency Monetary (RFM) technique, clusters from the K-means technique are created to predict the behaviour of the best and least contributing retail customers. Transactional data was extracted from a Business to Customer (B2C) hyperretail store in India comprising 10, 20, 284 transactions done by 2140 regular customers taking into account their recency, frequency and total spending. Based on RFM metric values across three heterogeneous segments, customers were characterised as toppers, moderated and churners. Analysis reveals that the most valuable customers have RFM scores as HHH (high recency, high frequency and high monetary value). These are the most loyal customers and retailers cannot afford to lose them. In micro-segmentation, stores should also prioritise retaining customers who have a recent shopping experience (medium recency) but do so infrequently, while spending larger sums. This can be achieved through tailored marketing strategies. Implications stand for both offline and online retail businesses to understand customer behaviour and tailor-made marketing strategies.

Keywords: Business Analytics, Clustering Retail, RFM Analysis, Segmentation
JEL Classification: C38, L81, M31

1. Introduction

In the last ten years, the retail landscape has undergone significant transformations, driven by factors such as the rise of e-commerce, market saturation and shifts in consumer shopping habits. The revolution of businesses is now confronted with technology and competition from category-dominating retailers, discount stores and factory outlets which poses a formidable challenge. According to Grewal *et al.* (2018), identification by the retailer of who, what, where and why behind the shopping experience is going through a metamorphosis. Hence, managers must focus on developing and embracing marketing strategies that are highly effective in rekindling the waning interest of their customers, affected by intense market competition. The rising disposable incomes, favourable demographics, changing lifestyles, the expansion of the middle class and strong penetration potential in urban and rural markets are the key drivers of India's retail growth. The market is seeing more variety and quantity of items, but when the client base grows, resources may occasionally be squandered on the wrong target market (Calvo-Porral & Levy-Manin, 2019). Organisations will need to adopt an unconventional or inventive approach to their strategic decision-making by utilising cutting-edge information technologies to serve customers. Therefore, understanding client wants and meeting those needs when delivering products gives a competitive edge to the organisation.

Businesses engage in a variety of tasks, including developing advertising campaigns and providing pertinent products. Customers will eventually gravitate towards shopping centres that provide them with differentiated services and their preferred selection

*Author for correspondence

of goods because of homogeneity. Shop managers are encouraged by this element to create a strategy that sets them apart from rivals (Diallo *et al*., 2018; Hiremath *et al*., 2023). However, many marketers find it challenging to select appropriate customer categories for organising marketing initiatives (Mohammadian & Makhani, 2016). Marketing managers may want to reconsider their consumer engagement techniques such as customised and personalised offerings, if they know the customer behaviour of different segments (Hosseini & Shabani, 2015). If advertising and promotional campaigns are not able to target their audience identified by the similarities within each group and uniquely differentiated from each other, the real purpose of campaigns will be defeated. Hence, the segmentation process aids in analysing customers wants and purchasing patterns as well as in aiding in decision-making regarding evolving market conditions and rivalries (Manjunath & Kashef, 2021). Behavioural customer segmentation is commonly carried out using the RFM model (Parvaneh *et al.*, 2014; Hajipour & Esfahani, 2019; Abbasimer & Shabani, 2021).

From the above discussion, it is evident that target audience segmentation is adopted extensively as a marketing strategy for understanding customer behaviour. In a similar context, Delen and Ram (2018) focussed on major challenges and opportunities which future businesses will face while applying business analytics tools, especially in marketing contexts such as product pricing, product campaign optimisation and identifying potential customers and customers with high attrition rates. However, the rationale to empirically conduct the study on market segmentation emerges because existing literature uses target audience segmentation as a standalone practice and there appears to be a need to bridge the gap between scholarly advancements and practical use (Dibb & Wensley, 2022; Goller *et al.,* 2002*)*. There is a plethora of studies (Dolnicar, 2022; Kumar & Ayodeji, 2021; Olson *et al.*, 2021) discussing the possible benefits of market segmentation and hence, this raises questions about whether the possible payoffs of segmenting target audience that the studies discuss are just notional or if actual data may support them. Further to this, most of the research on market segmentation conducted till today focuses on several bases of segmentation and how to create or assess segments. However, there is in general a lack of connection between the application of the segmentation techniques and the theoretical aspect of marketing. Lastly, since we know that the resources of the firm are meagre, it gets to the basic to bring the resources available to the company to the important customers which will lead to the increase in the profitability of the firm. Hence, it is quite enthralling to uncover those customers that will have a direct visible contribution towards the financial advancement of the firm. This leads to the need for micro-segmentation of the customers. There has been a lack of literature focussing on the micro-segmentation of customers in India. This crevice in the existing literature acted as a motivator to microsegment the customers. Considering this, the present study delves into two phases, wherein the primary objectives are, firstly, to identify the clients that share similar traits, and then categorise them based on their purchasing patterns and secondly, to conduct a detailed empirical analysis of each cluster by mapping customers, which results in the creation micro-segments within the customer base.

To achieve these objectives, we identify the following research issues to be addressed:

- Which customer segments are considered most valuable for designing the upcoming promotion offers?
- Which set of customers have reduced their visit and buying frequency?
- How to communicate with the customer segment with similar attributes?
- What can be the strategies regarding how to retain the best customers and win back lost customers?
- Which micro group should be targeted given the limited resources?

The study is organised in the following manner. Firstly, it presents the conceptual background of related theories based on fundamental traits and in the context of segmentation. The source of the data set and the approach are presented in the subsequent section. The analysis, findings and tactics that the store can use to keep and expand its client base are all presented in the fourth section. Finally, the study concludes by listing its implications along with some of its shortcomings.

2. Literature Overview

This section provides a theoretical understanding of customer relationship marketing as a governing philosophy for the selection of techniques for customer

segmentation. It reviews the literature on customer relationship marketing and personalisation, along with that on the RFM Model and K-means algorithm.

2.1 Theory of Customer Relationship Marketing

According to Gundlach and Wilkie (2009, p.260), "Marketing is the activity, set of institutions and processes for creating, communicating, delivering and exchanging offerings that have value for customers, clients, partners and society at large." It emphasises that the responsibility for nurturing customer relationships is dependent on how well the customer data is managed by businesses and to reflect upon the changes over a period of time, to boost the company's competitive edge, as outlined by Chakravorti (2009). In the present study, we are acknowledging the theory of customer relationship marketing (Payne & Frow, 2005; Moller & Halinen, 2010) which is closely related to the RFM model for segmenting customers. Understanding relationship marketing theory, with respect to customer centricity, delves marketers to segment, target and position their offerings based on their customer's shopping behaviour over a period. This strategy advises prioritising long-term customer interactions and is more efficient than depending on a sequence of possibly unconnected, one-time transactions (Zinkhan, 2002). It is widely discussed in related studies that marketing actions can have an indirect, peripheral or unintentional impact on marketing relationships (Eiriz & Wilson, 2006). Hence, the approaches companies employ for their product or service offerings mainly rely on the extraction of valuable customer information. This serves the ultimate goals of maximising value via personalisation, gaining customer loyalty and ultimately, securing substantial profits for the companies. It becomes essential for companies to predict customer behaviour and establish long-term relationships with their customers not only to reap benefits for their own sake but also to deliver value in return (Cheng & Chen, 2009).

2.2 Theory of Personalisation

Personalisation involves tailoring the shopping experience to customers based on data collected by marketers (Dangi & Malik, 2017). The theory involves crafting customer experiences that are pertinent, distinctive and convenient, to establish a mutually advantageous connection between the marketer and the customer. Likewise, Vesanen (2007) underscores that contemporary personalisation lacks a strict definition and encompasses diverse interpretations, ranging from aligning the visual appearance of a message to tailoring its content and customising the product. Past research has employed various terms to describe this concept, such as customisation (McIntosh *et al.,* 2010), one-to-one marketing (Kahan, 1998) and customerisation (Wind & Rangaswamy, 2001). However, the terms "personalisation" and "customisation" have been used interchangeably in some literature, with varying opinions on their distinction.

In recent times, there has been a growing fascination within retail marketing and data mining sectors with the idea of creating personalised models of customer behaviour (Pires *et al.* 2006). These custom-tailored models of customer behaviour become integral components of customer profiles and are later employed to provide personalised product and service recommendations to customers (Jiang & Tuzhilin, 2009). Hence, understanding customer requirements and building strong customer relationships is pivotal, but meeting those requirements in more personalised and adopting unique ways is a key factor ensuring the success of the business operation and critical for gaining more profits in the new business era.

2.3 Recency-Frequency-Monetary Value (RFM) Model

A well-known and reliable method for assessing clients based on their purchasing patterns is RFM analysis (Spring *et al.*, 1999; Hughes, 2005). The RFM model, extensively used in customer value analysis, is a method for predicting future consumer conduct based on past customer activity. The RFM model has been used extensively in customer value analysis. It has been employed by numerous academics to carry out customer segmentation (Jonker *et al*, 2004; Chen *et al.*, 2012; Khajvand & Tarokh, 2011; Wei *et al.*, 2012). Three main components form the basis of an RFM model: 1) How recently a customer dealt with a brand; 2) How frequently they did business with a brand; and 3) How much they paid for the products of a brand. It has been extensively utilised as a well-liked instrument of customer value analysis in determining client lifetime value and in analysing client behaviour and segmenting customers (Chen *et al.,* 2012).

Using the RFM concept, customer profiling is done to identify potential customers. For each transaction, different weights are specified to produce weighted

association rules through mining (Cho *et al.*, 2013). Past research has shown that companies can gain a lot from using RFM, including higher response rates, cheaper order costs, and higher profit (Wei *et al.*, 2012). It is proposed as a method that combines customer segmentation and targeting for market campaign strategies (Chan, 2008). To illustrate, in an Indian study, the RFM approach was used to investigate the relationship between marketing campaigns and segmentation of their target audience in the banking industry. The study came to its conclusion in two phases, the first of which included the RFM-based clusters and the second of which included demographic data that divided each cluster into new clusters (Kumar *et al.*, 2012).

2.4 K-Means Algorithm

Cluster analysis, a data mining approach, has been widely employed in a variety of industries, including manufacturing, marketing, travel, the health sector, the fashion sector and others (Brito *et al.*, 2015). Cluster analysis divides a data set into segments based on data similarity using partitions (Duan *et al.*, 2017). In recent times, K-means clustering has found applications in climatic clustering, as evidenced by its utilisation in a case study involving Iran, a nation characterised by diverse climatic regions (Sadeghi *et al.*, 2022). Moreover, Das *et al.* (2023) formed clusters while considering K-means clustering based on the estimated Bitcoin's extreme value returns. Its popularity in this context stems from its advantages compared to other methods, such as its straightforward approach to cluster creation, adaptability to new datasets and guaranteed convergence of clusters. Recently, studies such as Salminen *et al.* (2023), Wei *et al.* (2012), Tabianan *et al.* (2022), Sheshasaayee and Logeshwari (2017) found that K-means clustering is the most frequently used algorithm for Lifetime Value (LTV) and RFM analysis. Liu and Shih (2004) also suggested a combination of various hybrid methods for product recommendations, drawing on Customer Lifetime Value (CLV) and buying behaviour. Furthermore, a three-dimensional strategy to enhance consumer behaviour, customer happiness, and CLV was developed by He X and Li (2016). Very few studies like Shirole *et al.* (2021) and Anitha *et al.* (2022) propose business intelligence to analyse past sales data and consumer behaviour to find future customers in the said sector by using RFM and the K-means algorithm. Additionally, Salminen *et al.* (2023) found that multiple algorithms yield more effective outcomes. Therefore, amalgamating several algorithms could help overcome the weaknesses unique to each algorithm and contribute to the development of stronger and more distinct customer segments.

3. Data and Methodology

3.1 Data

The data of an Indian retail store as a sampling case is gathered from a secondary source. The store is a well-known India-based hypermarket catering to the daily needs of customers. The final data is extracted from 1st January 2022 to 31st December 2022. To ensure confidentiality, the name of the brand and the store has been masked. The data consists of the variables of regular customers visiting the store. For the calculation of RFM parameters, transactional data is used. The number of transactions for the year of the store is 10,20,284. These transactions are for a total of 2140 customers.

3.2 Methodology

There are certain pre-conditions which are enumerated as a four-step procedure to conduct the analysis. Given the significant variations in the value ranges of the three RFM model indicators-time elapsed since the last purchase, purchase frequency, and total expenditures on purchases. The min-max normalisation approach (see formula (1)) is used to standardise the data and produce the first standardised dataset to reduce the influence of numeric values on the classification outcomes

$$stdx_{ij} = \frac{x_{ij} - \min\left(x_j\right)}{\max\left(x_j\right) - \min(x_j)} \qquad (1)$$

here x_{ij} denotes the j^{th} index of the i^{th} sample.

RFM Score- a marketing framework for client segmentation based on behaviour was calculated. Firstly, recency signifies how long has it been since a consumer last engaged in an activity or made a purchase from the business. The gap between the end of the year, which is December 31, 2022, and the most recent time a customer visited a physical store is how recency was calculated. Secondly, frequency depicts how frequently a customer buys anything from the store or pays a bill during a specific timeframe. Frequency was determined as the total of each customer's "Sale_trans_no" during their

visit to the physical store and thirdly, monetary value is an attribute which shows how much a consumer spends at the business over a specific period. The total value of net sales for each customer who entered the physical retail store was used to compute the monetary value.

Further RFM values are each given a weight, which results in a weighted score for everyone in the database. Each customer will get a weight between 1 and 3 for each parameter. The assignment of weights is based on the judgement with regard to this database; for instance, in this particular dataset recency, frequency and monetary value were weighed into 3 quantiles. As a result, this use of RFM is frequently referred to as judgment-based RFM (Baier *et al.*, 2002) and a K-means algorithm was created. Therefore, this step is one of the essential stages in the procedure.

Clustering the customers by K-means algorithm, where K-means is used which is a well-known, non-supervised, recursive clustering method that starts with fixed K initial clusters. The Euclidean distance from each data point to the cluster centroid is calculated. The data objects are subsequently allocated to the cluster whose centroid C_i is closest to the target data object. The new cluster centre for the next iteration is then determined by averaging all the data objects in each cluster. Until the cluster centre stops shifting, or the allotted number of iterations has been achieved, the process is repeated. The data objects in the space and the Euclidean distance between the cluster centre are calculated using the following formula:

$$d(x,C_i)=\sqrt{\sum_{i=1}^{n}(x_j-C_{ij})} \quad (2)$$

where x refers to the data object,

C_i refers to the i^{th} cluster centre,

n refers to the dimension of the data object,

and x_j and C_{ij} refer to the j^{th} attribute value of x and C_i, respectively.

The selection of the value of k for the number of clusters has a substantial effect on the clustering outcomes. The optimal value k is usually identified using an elbow plot. The k value associated with the elbow-shaped plot represents the point where the relationship between SSE and k reflects the actual cluster number for the data. The sum of squared errors, or SSE, is the key indicator of choosing the actual number of clusters on the elbow plot (see Figure 1) and is represented by the following formula:

$$SSE=\sum_{f=1}^{k}\sum_{x\varepsilon C_i}\left|d(x_iC_i)\right|^2 \quad (3)$$

Here, C_i represents the i^{th} cluster and SSE denotes the clustering error of all samples, acting as an indicator of clustering quality. Finally, the micro-segmentation features are again aligned inside each cluster to produce more precise and uniform segments.

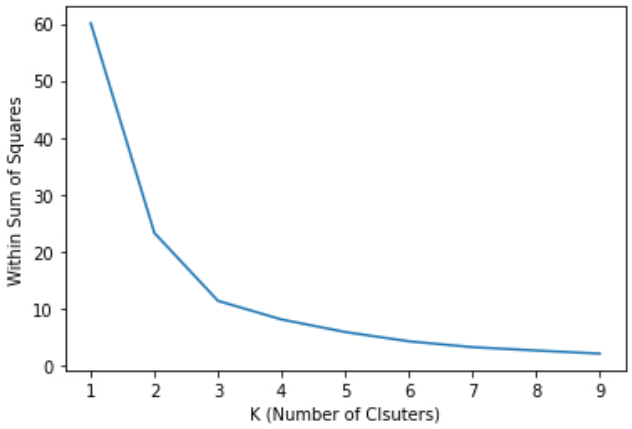


Figure 1. Elbow plot.

4. Empirical Setting

For the creation of the initial dataset, anomalous data is eliminated. RFM scores are calculated and clusters are created using the K-means clustering technique. The elbow plot (Figure 1) indicates that the reduction in SSE is not significant after k more than 3. Therefore, selecting $k=3$ would have positive results.

In addition to the Elbow plot, the Silhouette Visualiser (Figure 2) graphically assesses the density and separation

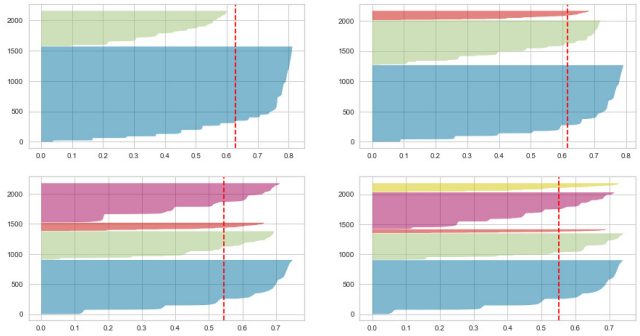


Figure 2. Silhouette visualisation of K means in different clusters.

between clusters by displaying the silhouette coefficient for each sample on a per-cluster basis. The values with scores close to +1 denote considerable separation in cluster assignment.

Table 1 displays the mean values within clustering based on RFM. According to the evaluation findings, three clusters with the corresponding central tendency values that is mean and customer counts are the number of customers in that particular cluster.

The following nomenclatures are created for simplicity as the clusters are based on average values of recency, frequency, and monetary value. The proposed insights are based on a clear interpretation of RFM that comprises distinctive and intrinsic features of the data set.

Table 1. Mean values of variables in clusters in RFM-based clustering.

Clusters	Mean of Recency	Mean of Frequency	Mean of Monetary	Count of Customers	Percentage of Customers
1	5.59	26.68	77,5502	1261	59.2
2	23.44	18.91	68,060	729	33.7
3	54.94	17.3	57,512	150	7

Table 2. Description of R, F and M features according to thresholds.

Low (L)	Customers having value ≤ Q1 of the respective feature
Medium (M)	Customers having value ≥ Q1 and ≤ Q3 of respective feature
High (H)	Customers having value > Q3 of the respective feature

Table 3. H, M and L matrix with customer's summary.

			Monetary			
	Recency	**Frequency**	**H**	**M**	**L**	**Total**
Active	H	H	116	100	17	233
		M	46	143	70	259
		L	12	37	37	86
At Risk	M	H	90	103	15	208
		M	109	318	132	559
		L	32	124	141	297
In Active	L	H	35	16	1	52
		M	75	135	60	270
		L	20	94	62	176
	TOTAL	**535**		**1070**	**535**	**2140**

Table 4. Attributes of shopping behaviour.

Score	Attribute	Description
HH	VIP or Loyalists	Segments with "H" Monetary and Frequency
HHH	Champions	RFM score as HHH
LHH	Risk of Loosing	H in Monetary, H in Frequency and L in Recency
MM	Potential Loyalists	M in Monetary and Frequency
LLL	Lost customers:	The ones with a score LLL

Interpreting and understanding such clusters will help in developing or modelling customer-centric strategies in a much better way.

- Cluster 1 which includes 59.2 per cent of customers, comprising 1261 top customers who have recently made purchases with an average interval of five days between transactions. The average monetary value of this cluster is Rs. 7,75,502. Customers in this cluster are topper performers. They exhibit high frequency and recency in their purchases, classifying them as champions, regular buyers and promising customers.
- Cluster 2 (33.7 per cent customers) consists of 729 regular shoppers with an average purchase cycle of 23 days, visiting every 19 days and spending an average of Rs. 68,060. They exhibit moderate frequency and recency, including new, promising, dormant and at-risk customers.
- Cluster 3 (7 per cent customers) includes 150 buyers who purchase on average every 55 days, with visits spaced 17 days apart and an average spend of Rs. 57,512. This segment consists of churned customers with low frequency and recency.

Table 5. **Customer mapping.**

		Frequency	Cluster	Monetary			Total
				H	M	L	
Recency	H	H	1	116	100	17	233
		M	1	46	143	70	259
		Total		162	243	87	492
	M	H	1	68	75	13	156
			2	22	28	2	52
				90	103	15	208
		M	1	61	202	93	356
			2	48	116	39	203
				109	318	132	559
		L	1	15	69	87	171
			2	17	55	54	126
				32	124	141	297
		Total		231	545	288	1064
	L	H	2	32	10	1	43
			3	3	6		9
				35	16	1	52
		M	2	58	92	38	188
			3	17	43	22	82
				75	135	60	270
		L	2	15	61	41	117
			3	5	33	21	59
				20	94	62	176
		Total		130	245	123	498
Grand Total				535	1070	535	2140

Further, to make the recommendations more specific, the customers are separated by RFM variable values and divided into three groups The standard cut-offs for R, F and M are established based on quartiles (Tables 2 and 3).

Segmentation based on RFM scores: In RFM analysis, a customer's recent visit (recency) is prioritised more heavily, with frequency and monetary value considered next in importance. The highest-value segment consists of customers with "H" in both monetary value and frequency, followed by the next best group marked with "M" in these areas, while all others with "L" in frequency and monetary value are given lower priority. Accordingly, different promotional activities can be tailored for these segments based on their RFM scores aligned with marketing objectives and budget. Therefore, the segments with shopping behaviour characteristics are as listed in Table 4.

In Table 5, customers are classified into RFM clusters (1,2,3) and 27 segments generated through data compression (H, M, L). The table shows that some segments fall into multiple clusters, such as MHH, which is part of both cluster 1 and cluster 2, but there are only 22 customers in the same segment of MHH of cluster 1 as

Table 6. **Mean values of MLH segment.**

Segment	Customer Count	Cluster	Recency	Frequency	Monetary
MLH	15	1	9	13	1.27
	17	2	18	12	1.1

Table 7. Six differentiated profiles of customers using a mapped approach of RFM segments and k Means clustering.

Customer/ RFM Segment	Interpretation	Customer's Activity	Marketing Strategy
Champions (Cluster_1)	High recency, frequency and monetary value	Bought recently (0-4 days) often buy (>27 visits a year) and spend.the most (more than INR 90,000 a year)	Incentivise the customer
Dedicated Customers (Cluster_1)	High frequency and monetary value, but low recency	Spend good money (30K-90K) (15-26 visits a year)	Upselling
Potential customers (Cluster_1)	Low recency and frequency, but high monetary value	Recent customers with significant purchases	Provide membership and suggest additional products
Recent Customers (Cluster_1)	High recency, but low frequency and monetary value	Bought most recently but not often	Provide on-boarding support
Promising (Cluster_2)	No prior visits and spent very less	Recent shoppers but haven't spent much	Create brand awareness, and offer free trials
Lost (Cluster_3)	Low recency, frequency, and monetary value	Lowest recency, (>40 days) frequency (10-14 visits) lowest amount spent	Revive interest with a reach-out campaign, ignore otherwise

compared to 68 customers in cluster 2. Such a customer profile mapping can benefit stores by focusing on the groups of customers to be concentrated on due to the limited resources available to them while considering the business's objectives of profit maximisation. The store using this technology can also target a customer group based on bespoke marketing and personalised experiences.

Table 6 shows that the MLH segment which includes 32 customers (15 in cluster 1 and 17 in cluster 2), is characterised by medium recency, low frequency, and high monetary value according to the available RFM segment output. The company wants to target the customers with appropriate marketing strategies. Even though the company is unable to reach all 32 customers due to budget constraints. The reality is, however, different and the store can reach only fifty per cent of its customers.

5. Results and Discussion

In the current study, a transactional data set of 2140 consumers of a retail store from January 2022 to December 2022 is used to analyse the segmentation of customers by RFM. From the analysis, the distribution of recency, frequency and monetary metric values is done across the three clusters. The three clusters are separate from each other, which indicates a good heterogeneous mix of clusters. In cluster 1, all customers are categorised as toppers and champion customers who are described with their regularity, loyalty, and ongoing emotional relationship with their shopping points. This cluster widens the scope for the retailers to identify and retain their potential customers. Cluster 2 has customers who visit a retail store not very frequently, but they are known to stores with their few recent purchases. These customers require the stores' attention and need to be handled with care and enhanced purchase experience by the retail store. Cluster 3 consists of customers with the least monetary spending and making no major transactions.

In the next stage, a total of 45 micro-segments are created, which are regrouped into 6 customer profiles based on recency, frequency and monetary values (Table 7). These nomenclatures are created for simplicity and grouping of common characteristics of micro-segments. The following table is a summary of 6 profiles of the customers which provides them with a general overview of RFM segmentation and marketing strategy.

Finally, the results indicate micro-segmentation can be possible with clustering based on RFM, based on the research questions posited in the first section. It entails retailers enhancing their marketing mix strategies while giving due consideration to the purchasing behaviour of customers (Arora *et al.,* 2008). Based on an in-depth analysis, it is derived that the most valuable customers

have RFM scores as HHH. These are the most loyal customers and retailers cannot afford to lose them. Such customers are the ones who have made recent purchases from the store (0-4 days), often buy (>27 visits a year) and spend the most (more than Rs. 90,000 a year). This segment requires the undivided focus of the marketing department of any retail store. For their retention, these customers should be rewarded with special offers, loyalty programs, and other incentives to keep them coming back. Though this cluster shows the potential market and the valuable ones, businesses still need to be cautious about not losing this segment at any cost, as there is a subgroup that is spending good money (Rs. 30,000 to Rs. 90, 000) with the respective store often, also making 15-26 visits a year. Stores need to spend more in terms of promotion as these customers are responsive to advertising. In this subgroup, there may be chances of brand switching because of fierce competition and their interest may fade away if they are not engaged by the retailers. The company should try to increase this micro segment's frequency of purchase with targeted offers and promotions and should try to keep them engaged with regular communications and promotions. Hence, businesses need to monitor the number of times the same customer makes purchases, the frequency of their recent visits and the amount of money spent with a store. These customers have the potential to be very valuable, but they need to be nurtured. In the literature about customer relationship management, relationship marketing is studied as the best model to achieve the needs of customers and to enhance the strength with customers for the company (Payne and Frow, 2017). To reduce costs and increase profits, reinforcement of excellent customer relationship marketing is required.

Further, the study has also identified customers who have reduced their visit and frequency of buying. In Table 3, their RFM score is LLL which reveals customers in this segment have low recency in terms of purchases (<40 days), low frequency (10-14 visits) and low monetary value (lowest amount spent). Hence, providing customers of this segment with a personalised experience will improve their buying decision process which improves their actual visits and maps actual buying decisions from just having intentions to buy (Jiang & Tuzhilin, 2009). For such a profile, retailers need to win customers back via renewals or newer products. To revive their interest, the marketing team needs to be vigilant and also should take feedback from customers. Their reach-out campaigns should be so designed to ensure that this segment is not lost to the competition.

6. Implications

- *Customer segmentation with accuracy:* The study contributed to segmenting retail customers into distinct groups based on their RFM characteristics with accuracy. This helps retailers not only understand different customer behaviours but also tailor their marketing strategies, promotions and product offerings best suited to each segment. Standardising features help ensure that all dimensions contribute equally to the clustering process or purchase history, which can enhance the accuracy of customer segmentation. Identifying an optimal number of clusters is critical and techniques such as the elbow method and silhouette score can help determine the appropriate number of clusters that best represent the data used under segmentation.
- *New era retail stores:* The methodology can be used for online retail just as effectively as it can be used for traditional brick-and-mortar retail. It is particularly well-suited for online retail due to the abundance of data that e-commerce businesses can collect. Online retail allows for highly personalised marketing and product recommendations based on customer behaviour. To illustrate, there are many new businesses such as online fashion retail (e.g., Myntra.com; Nykaa.com), food and grocery delivery (e.g., Zomato.com; Blinkit.com and Country Delight. com) are taking the help of these tools for analysing customer data for tailoring customer recommendations. These recommendations will be helpful in personalised messages for effective communication. These e-businesses can set up automated marketing campaigns, like email marketing, retargeting, and personalised product recommendations. As online retail provides a wealth of data on the customer journey, from the initial website visit to the final purchase, RFM analysis, can help identify critical touchpoints in the customer journey and optimise the user experience.

- *Customer Lifetime Value (CLV):* The analysis in this study can help in estimating the lifetime value of customers in each segment. Retailers can prioritise their resources and marketing efforts on segments with higher CLV potential. CLV is a crucial metric in marketing and business that estimates the total value a customer is expected to generate for a company throughout their entire relationship. It's a forward-looking metric that helps businesses understand the long-term financial impact of acquiring and retaining customers. It is based on the idea that some customers are more valuable over time due to their buying habits and loyalty. So, this type of analysis can guide managers to make marketing and customer retention efforts to high CLV segments while implementing strategies to increase CLV for at-risk or low-value segments. This allows businesses to allocate their marketing budget appropriately while giving due consideration to being efficient and effective while designing marketing campaigns.
- *Competitive Advantage to Retailers:* Retailers can develop loyalty programs and other initiatives depending upon the unique and differentiated needs of customers who are more likely to appeal to those segments. It entails better decision-making about a variety of issues, including pricing and product development for targeted segments. This can lead to increased customer loyalty, which can translate into repeat business and higher profits. Micro-segmentation helps in improved product assortment while minutely predicting the true purchasing patterns of different customer segments, retailers can develop their strategies to know which products will create demand in the future and which ones are to be discarded. This can help retailers to improve their product assortment and ensure that they are carrying the products that their customers want.

7. Conclusion

The prime objective of this study was to devise an appropriate marketing strategy for hyperstores catering to meeting the daily customer requirements of different types of customers. Depending upon the shopping traits of customers like the number of visits, the time between their different visits and the total amount spent, an in-depth empirical analysis by customer mapping leads to the generation of micro-segments of customers. With this, marketing thinkers and practitioners such as retailers can identify the major areas to focus on to implement effective marketing and promotional tactics. Thus, RFM along with the K-means clustering algorithm is found suitable for creating clusters and micro-segments based on the shopping behaviour of customers of one of the largest Indian retail stores. The detailed analysis was found to be more pragmatic because of its dyadic approach of considering revenue-based and purchasing complexity-driven methods. Therefore, it enables retail marketers to undertake an easier execution of customer segmentation based on clustering.

8. Limitations and Future Recommendations

The research gives a holistic view of the customers that the firm must largely focus on, yet the study has its limitations. For instance, the stores additionally will also have to pay attention to which products, are bought often, and ensure their availability at times (seasons) advantageous to the store. It is possible to study a variety of other business criteria to create a more successful marketing strategy such as promoting popular products during different seasons and adopting appropriate marketing strategies for particular events. The study also has methodological limitations and hence, other techniques such as Self Organising Map (SOM), Fuzzy C mean, Latent Class Analysis Model and Gaussian Mixture Model (GMM) techniques can be used alongside K-means that will help in the creation of more robust and distinct customer segments (Spoor, 2022). Future researchers can also explore that customer segmentation is done while taking cost as a constraint which is an essential attribute for meeting business goals more objectively (Kumar *et al.*, 2020; Joun & Kim, 2023).

9. References

Abbasimehr, H., & Shabani, M. (2021). "A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series

techniques", *Journal of Ambient Intelligence and Humanised Computing, 12*(1), 515-31. https://doi.org/10.1007/s12652-020-02015-w

Anitha, P., & Patil, M. M. (2022). "RFM model for customer purchase behavior using K-Means algorithm", *Journal of King Saud University: Computer and Information Sciences, 34*(5), 1785-92. https://doi.org/10.1016/j.jksuci.2019.12.011

Arora, N., Dreze, X., Ghose, A., Hess, J. D., Iyengar, R., Jing, B., Joshi, Y., Kumar, V., Lurie, N., Neslin, S., Sajeesh, S., Syam, M. S. N., Thomas, J., & Zhang Z. J. (2008). "Putting one-to-one marketing to work: Personalization, customization, and choice". *Marketing Letters*, *19*, 305-21. https://doi.org/10.1007/s11002-008-9056-z

Baier, M., Ruf, K.M., & Chakraborty, G. (2002). "Contemporary database marketing: concepts and applications. Racom Communications, Evanston.

Brito, P.Q., Soares, C., Almeida, S., Monte, A. & Byvoet, M. (2015). "Customer segmentation in a large database of an online customized fashion business", *Robotics and Computer-Integrated Manufacturing, 36*, 93-100. https://doi.org/10.1016/j.rcim.2014.12.014

Calvo-Porral C. & Lévy-Mangin, J. P. (2019). "Profiling shopping mall customers during hard times". *Journal of Retailing and Consumer Services, 48,* 238-46. https://doi.org/10.1016/j.jretconser.2019.02.023

Chakravorti, S. (2009), "Extending customer relationship management to value chain partners for competitive advantage". *Journal of Relationship Marketing,* 8, 299-312. https://doi.org/10.1080/15332660903344636

Chan, C. C. H. (2008). "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer". *Expert System with Applications, 24*, 2754-62. https://doi.org/10.1016/j.eswa.2007.05.043

Chen, Y. S., Cheng, C. H., Lai, C. J., Hsu, C. Y., & Syu, H. J. (2012). "Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment". *Computers in Biology and Medicine, 42*(2), 213-21. https://doi.org/10.1016/j.compbiomed.2011.11.010 PMid:22177941.

Cheng, C. H., & Chen, Y. S. (2009). "Classifying the segmentation of customer value via RFM model and RS theory". *Expert systems with applications, 36*(3), 4176-84. https://doi.org/10.1016/j.eswa.2008.04.003

Cho, Y. S., Noh, S. C., Moon, S. C. (2013). "Weighted Mining Association Rules Based Quantity Item with RFM Score for Personalized u-Commerce Recommendation System", in Park, J.J. (H., Arabnia, H.R., Kim, C., Shi, W., Gil, JM. (Ed.s), GPC 2013: *Lecture Notes in Computer Science, 7861. Springer*. Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38027-3_39

Dangi, H., & Malik, A. (2017). "Personalisation in marketing: an exploratory study". *International Journal of Internet Marketing and Advertising, 11*(2), 124-36. https://doi.org/10.1504/IJIMA.2017.084079

Das D., Kayal P., Maiti M. (2023). "A K-means clustering model for analyzing the Bitcoin extreme value returns". *Decision Analytics Journal, 6*, 100152. https://doi.org/10.1016/j.dajour.2022.100152

Delen, D., & Ram, S., (2018). "Research challenges and opportunities in business analytics. *Journal of Business Analytics, 1*(1), 2-12. https://doi.org/10.1080/2573234X.2018.1507324 https://doi.org/10.1080/2573234X.2018.1507324

Diallo, M. F., Diop-Sall, F., Djelassi, S., & Godefroit-Winkel, D. (2018). How shopping malls service quality affects customer loyalty across developing countries: The moderation of the cultural context. *Journal of International Marketing, 26*(4), 69-84. https://doi.org/10.1177/1069031X18807473

Dibb, S. & Wensley, R. (2002). "Segmentation analysis for industrial markets: problems of integrating customer requirements into operations strategy". *European Journal of Marketing, 37,* 231-251. https://doi.org/10.1108/03090560210412773

Dolnicar, S. (2022). " Market segmentation for e-tourism", in Xiang, Z., Fuchs, M., Gretzel, U., Höpken, W. (Ed.s). Handbook of e-Tourism. Springer, Cham. https://doi.org/10.1007/978-3-030-48652-5_53

Duan, Q., Yang, Y. L., & Li, Y. (2017). "Rough k-modes clustering algorithm based on entropy". *IAENG International Journal of Computer Science, 44*(1), 13-8.

Eiriz, V., & Wilson, D. (2006). "Research in relationship marketing: antecedents, traditions and integration". *European Journal of Marketing,* 40(3/4), 275-91. https://doi.org/10.1108/03090560610648057

Goller, S., Hogg, A. & Kalafatis, S. P. (2002). "A new research agenda for business segmentation". *European Journal of Marketing, 36*(1/2), 252-71. https://doi.org/10.1108/03090560210412782

Grewal, D., Motyka, S., & Levy, M. (2018). "The evolution and future of retailing and retailing education". *Journal of Marketing Education, 40*(1), 85-93. https://doi.org/10.1177/0273475318755838

Gundlach, G. T., & Wilkie, W. L. (2009). "The American marketing association's new definition of marketing: Perspective and commentary on the 2007 Revision". *Journal of Public Policy & Marketing, 28*(2), 259-64. https://doi.org/10.1509/jppm.28.2.259

Hajipour, B., & Esfahani, M. (2019). "Delta model application for developing customer lifetime value". *Marketing Intelligence & Planning, 37*(3), 298-309. https://doi.org/10.1108/MIP-06-2018-0190

He X., & Li, C. (2016). "The research and application of customer segmentation one-commerce websites" at 6th International Conference on Digital Home (ICDH), Guangzhou, p. 203-8. https://doi.org/10.1109/ICDH.2016.050

Hiremath, S., Panda, A., & Pasumarti, S. S. (2023). "An empirical investigation of customer characteristics on retail format selection-a mediating role of store image". *Journal of Indian Business Research. 15*(1), 55-75. https://doi.org/10.1108/JIBR-04-2022-0090

Hosseini, M., & Shabani, M. (2015). "New approach to customer segmentation based on changes in customer value". *Journal of Marketing Analytics, 3*(3), 110-21. https://doi.org/10.1057/jma.2015.10

Hughes, A. M. (2005). Strategic database marketing. McGraw-Hill, Chicago, USA Companies. https://dl.acm.org/doi/book/10.5555/1199134

Jiang, T., & Tuzhilin, A. (2009), "Improving personalization solutions through optimal segmentation of customer bases". *IEEE Transactions on Knowledge and Data Engineering, 21*(3), 305-20. https://doi.org/10.1109/TKDE.2008.163

Jonker, J. J., Piersma, N., & Van den Poel, D. (2004). "Joint optimization of customer segmentation and marketing policy to maximize long-term profitability". *Expert Systems with Applications, 27*(2), 159-68. https://doi.org/10.1016/j.eswa.2004.01.010

Joung, J., & Kim, H. (2023). "Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews". *International Journal of Information Management, 70*, 102641. https://doi.org/10.1016/j.ijinfomgt.2023.102641

Kahan, R. (1998). "Using database marketing techniques to enhance your one-to-one marketing initiatives". *Journal of Consumer Marketing, 15*(5), 491-3. https://doi.org/10.1108/07363769810235965

Khajvand, M., & Tarokh, M. J. (2011). "Estimating customer future value of different customer segments based on adapted RFM model in retail banking context". *Procedia Computer Science, 3*, 1327-32. https://doi.org/10.1016/j.procs.2011.01.011

Kumar, V., & Ayodeji, O. G. (2021). "E-retail factors for customer activation and retention: An empirical study from Indian e-commerce customers". *Journal of Retailing and Consumer Services, 59*, 102399. https://doi.org/10.1016/j.jretconser.2020.102399

Kumar, M. V., Chaitanya, M. V., & Madhavan, M. (2012). "Segmenting the banking market strategy by clustering". *International Journal of Computer Applications, 45*, 10-15.

Kumar, S., Singh, J., & Singh, O. (2020). "Ensemble-based extreme learning machine model for occupancy detection with ambient attributes". *International Journal of System Assurance Engineering and Management. 11*, 173-83. https://doi.org/10.1007/s13198-019-00935-1

Liu, D.R., & Shih, Y. Y. (2004). "Hybrid approaches to product recommendation based on customer lifetime value and purchase preference". *The Journal of System and Software, 77,* 181-91. https://doi.org/10.1016/j.jss.2004.08.031

Manjunath, Y. S. K., & Kashef, R. F. (2021). "Distributed clustering using multi-tier hierarchical overlay super-peer peer-to-peer network architecture for efficient customer segmentation". *Electronic Commerce Research and Applications, 47*, 101040. https://doi.org/10.1016/j.elerap.2021.101040

McIntosh, R. I., Matthews, J., Mullineux, G., & Medland, A. J. (2010). "Late customisation: issues of mass customisation in the food industry". *International Journal of Production Research, 48*(6), 1557-74. https://doi.org/10.1080/00207540802577938

Mohammadian, M., & *Makhani*, I. (2016). "RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies". *International Academic Journal of Accounting and Financial Management, 3*(6), 21-35. https://doi.org/10.9756/IAJAFM/V6I1/1910009

Olson, E. M., Olson, K. M., Czaplewski, A. J., & Key, T. M. (2021). "Business strategy and the management of digital marketing". *Business* Horizons, *64*(2), 285-93. https://doi.org/10.1016/j.bushor.2020.12.004

Parvaneh, A., Tarokh, M. J., & Abbasimehr, H. (2014). "Combining data mining and group decision making in retailer segmentation based on LRFMP variables". *International Journal of Industrial Engineering & Production Research, 25*(6), 197-206.

Payne, A., & Frow, P. (2005). "A strategic framework for customer relationship management". *Journal of Marketing, 69*, 167-76. https://doi.org/10.1509/jmkg.2005.69.4.167

Payne, A., & Frow, P. (2017). "Relationship marketing: looking backwards towards the future". *Journal of Services Marketing Vol. 31*(1), 11-5. https://doi.org/10.1108/JSM-11-2016-0380

Pires, G.D., Stanton, J. & Rita, P. (2006). "The internet, consumer empowerment and marketing strategies". *European Journal of Marketing, 40*(9/10), 36-949. https://doi.org/10.1108/03090560610680943

Sadeghi, M., Naghedi, R., Behzadian, K., Shamshirgaran, A., Tabrizi, M. R., & Maknoon, R. (2022). "Customisation of green buildings assessment tools based on climatic zoning and experts' judgement using K-means clustering and fuzzy AHP". *Building and Environment, 223*, 109473. https://doi.org/10.1016/j.buildenv.2022.109473

Salminen, J., Mustak, M., Sufyan, M. & Jansen, J. B. (2023) "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic

customer segmentation". *Journal of Marketing Analytics*, p. 1-13. https://doi.org/10.1057/s41270-023-00235-5

Sheshasaayee, A., & Logeshwari, L. (2017). "An efficiency analysis on the TPA clustering methods for intelligent customer segmentation'. In International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, *p.* 784-8. https://doi.org/10.1109/ICIMIA.2017.7975573

Shirole, R., Salokhe, L., & Jadhav, S. (2021). "Customer segmentation using RFM model and k-means clustering". *International Journal of Scientific Research in Science and Technology, 8*, 591-7. https://doi.org/10.32628/IJSRST2183118

Spoor, J.M. (2022). "Improving customer segmentation via classification of key accounts as outliers". *Journal of Marketing Analytics*, p. 1-14. https://doi.org/10.1057/s41270-022-00185-4

Spring, P., Leeflang, P. S., & Wansbeek, T. (1999). "The combination strategy to optimal target selection and offer segmentation in direct mail". *Journal of Market-Focused Management. 4*(3), 187-203. https://doi.org/10.1023/A:1009899802421

Tabianan, K., Velu, S., & Ravi, V. (2022). "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data". *Sustainability. 14*(12), 7243. https://doi.org/10.3390/su14127243

Vesanen, J. (2007). "What is personalization? A conceptual framework". *European Journal of Marketing. 41*(5/6), 409-18. https://doi.org/10.1108/03090560710737534

Wei, J. T., Lin, S. Y., Weng, C. C., & Wu, H. H. (2012). "A case study of applying LRFM model in market segmentation of a children's dental clinic". *Expert Systems with Applications,* 39(5), 5529-33. https://doi.org/10.1016/j.eswa.2011.11.066

Wind, J. & Rangaswamy, A. (2001). "Customerization: the next revolution in mass customization". *Journal of Interactive Marketing, 15*(1), 13-32. https://doi.org/10.1002/1520-6653(200124)15:1<13::AID-DIR1001>3.0.CO;2-#

Zinkhan, G. M. (2002). "Relationship marketing: Theory and implementation. *Journal of Market-Focused Management*, *5*(2), 83-89. https://doi.org/10.1023/A:1014031025271