Journal of King Saud University – Computer and Information Sciences 34 (2022) 1785–1792

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

RFM model for customer purchase behavior using K-Means algorithm

P. Anitha *, Malini M. Patil

Department of ISE, JSS Academy of Technical Education, Bengaluru-560060, Karnataka, India. Visveswaraya Technological University, Belgaum-590018, Karnataka, India

ARTICLE INFO

Article history:
Received 23 July 2019
Revised 16 December 2019
Accepted 18 December 2019
Available online 25 December 2019

Keywords:
Recency
Frequency
Monetary
Silhouette coefficient
Business intelligence
Segmentation

ABSTRACT

The objective of this study is to apply business intelligence in identifying potential customers by providing relevant and timely data to business entities in the Retail Industry. The data furnished is based on systematic study and scientific applications in analyzing sales history and purchasing behavior of the consumers. The curated and organized data as an outcome of this scientific study not only enhances business sales and profit, but also equips with intelligent insights in predicting consumer purchasing behavior and related patterns. In order to execute and apply the scientific approach using K-Means algorithm, the real time transactional and retail dataset are analyzed. Spread over a specific duration of business transactions, the dataset values and parameters provide an organized understanding of the customer buying patterns and behavior across various regions. This study is based on the RFM (Recency, Frequency and Monetary) model and deploys dataset segmentation principles using K-Means Algorithm**.** A variety of dataset clusters are validated based on the calculation of Silhouette Coefficient. The results thus obtained with regard to sales transactions are compared with various parameters like Sales Recency, Sales Frequency and Sales Volume.

1. Introduction

In the light of data segmentation, customers are divided into set of individuals with distinct similarities. Some of the attributes relevant to customer segmentation are gender, age, lifestyle, location, purchase and income behavior. Such attributes are mainly categorized based on the historical purchasing behavior that can lead to a specific outcome, for example, an increase in sales and the profit for the company.

In the ever-growing competition and increasing complexity of business environment, segmentation and its systematic study improves customer loyalty and enhances enterprise-level for long lasting relationship by widening profitable customer database (Khalili-Damghani et al., 2018). The two most prominent types of segmentations used in K-Means Algorithm are the Qualitative and Quantitative insights. In the scope of the current study, Quantitative insight is used for the purpose of segmentation clustering Well-defined customer segmentation helps in effective allocation of marketing resources, enables the companies to target the specific group of customers and also helps in building healthy long-term relationship with the customers. The major industries wherein customer segmentation and for data mining can be applied the Retail Industry (Han et al., 2011), because it requires a vast amount of data on sales, transportation, consumption ratio, redelivery service and many others. Also, Retail data mining helps in identifying and effectively mapping customer behavior and related patterns during the entire life-cycle of business transactions. This ultimately, leads to improved customer service, effective sales and distribution strategies and many more (Han et al., 2011). This work mainly focuses on tracking the historical purchasing behavior of customers with the aim to find maximum amount of sale possible in the specific area. Based on the statistical results and indicators, companies in the retail industry can design various sales and marketing strategies like promotional campaigns, extending seasonal discounts or floating sales enabling coupons to increase the sales and improve customer retention.

To achieve the above objectives, customer clustering and segmentation is carried out using the K-Means algorithm. It is based on RFM values for different regions. RFM can be defined as segmentation of customer analysis which not only gives information on frequent purchasing pattern of the customer, but also recent purchase and the profit obtained (Hu and Yeh, 2014). Initially the clusters are evaluated using Silhouette Analysis for Recency Vs.

* Corresponding author.
E-mail addresses: anitha.palakshappa@gmail.com (P. Anitha), drmalinimpatil@gmail.com (M.M. Patil).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2019.12.011
1319-1578/

Monetary using K-Means for varying number of Clusters. This is followed by the Silhouette Analysis of Frequency Vs. Monetary, using the K-Means for different number of Clusters.

Silhouette Analysis is a prototype based method to evaluate or validate Clusters. The validity can be either be cohesion, or separation, or a combination of both. In the present work, Silhouette Coefficient combines both cohesion and separation.

2. Literature review

Group-specific marketing is common and much needed from the traditional mass marketing perspective. Customer segmentation is a part of various activities under Customer Relationship Management(CRM) value chain (Kolarovszki et al., 2016; Khalili-Damghani et al., 2018; Khajvand et al., 2011). The author has proposed a novel modeling in the field of postal area using the multi-dimensional segmentation. This design of CRM is useful in companies of postal services (Kolarovszki et al., 2016).

The proposed hybrid method is to predict the new entities for the customer centric companies. This study is based on the methods like decision tree approach, clustering, rule extraction and many more. The author has used K-Means for predicting future transactions based on the historical behavior of the customers under various segmentation. To facilitate this, the hybrid feature selection for filtering making and decision making method is used. The current study can also be applied for Telecommunication and Insurance industries to predict the sales volume and projecting business profit of a company. This method is very effective not only in predicting profitable customers, but also in identifying behavior of new customer. (Khalili-Damghani et al., 2018).

The concept of customer segmentation can be applied for beauty and healthcare companies, which indirectly leads to CRM. The author used two approaches for segmentation, where one is based on RFM and second one is extended RFM by addition of count item parameter. Sales and marketing strategies are explained by calculating Customer Lifetime value using weighted RFM (Khajvand et al., 2011).

Multiple analysis that is based on integration of CRM and RFM model is essential for exploring CRM in large scale data (Song et al., 2017). RFM model is employed to predict the supply quantity per month by clustering the customers using K-Means algorithm. Each group is distinguished using CHAID decision trees based on attribute values (You et al., 2015).

The relationship between consumer behavior and order fulfillment in the field of marketing and operations is identified using various marketing tools, which enhances the consumer service levels (Nguyen et al., 2018). Clustering technique is used to group the retailers using RFM model based on Electronic Funds Transfer at Point of Sale (EFTPOS) in businesses (Singh et al., 2014).

Data analytics approach is proposed for customer segmentation based on the customer visit to the store, collected from the overall sales data. Also, feature selection approach is proposed, which takes product taxonomy as input and categories of customers as output (Griva et al., 2018; Hu and Yeh, 2014).

In the current study, the RFM analysis is executed using transactional dataset for evaluating customers on their purchase behavior and analyzing the same using unsupervised algorithms like K-Means and Fuzzy C – Means. Also, the author has introduced a novel idea of selection of centroids in K-Means Algorithm and comparing the results (Christy et al., 2018).

Customer segmentation requires descriptive variables for identifying behavioral patterns. However, in some domains, descriptive variables are not adequate. Considering this, the author has proposed the segmentation method to solve the problem leading to better performance using data mining methods (Murray et al., 2017).

K-Means and self-organizing SOM algorithms are used to cluster the customer characteristics using RFM model for insurance dataset. It helps in identifying the customer needs, their understanding and characteristics (Qadadeh and Abdallah, 2018). The study suggests that fuzzy and SOM based clustering methods are more efficient compared to the traditional methods (Arunachalam and Kumar, 2018). Sequential exploratory design is adopted which combines both qualitative and quantitative methods (Arunachalam and Kumar, 2018).

Customer segmentation can also be applied for e-pharmacy clients to increase the retention of e-customers and it is one of the prerequisite for the essential CRM in the area of customer loyalty (Patak et al., 2014). Due to high volatility of the market, customer segments change over a period of time. The author has introduced the concept of 'Stream Clustering' as a tool and has proposed a new algorithm to overcome this problem. An important aspect in Stream Clustering is to identify the new clusters or emerging clusters and replacing the older ones (Carnein and Trautmann, 2019).

Swarm based algorithms like Flower Pollination Algorithm, Black Hole Algorithm, Bat Algorithms and others are proposed to overcome the problem of slow convergence for larger datasets. Also Comparative analysis of algorithms is carried out using four performance parameters (Kaur et al., 2019). A framework is designed, where e-commerce research is categorized or segmented into three phases. In each phase, the author has gather the issues reported by the practitioners and has suggested solutions to overcome them. Conceptual framework consists of service relationships, business models and technologies (Yoo and Jang, 2019).

Data mining technique called Clustering Approach can also be used to address various road-blocks in the manufacturing and marketing problems in fashion industry. Needless to say, segmentation is very important for finding the patterns of customer preferences (Brito et al., 2015).

Business Intelligence refers to the intelligent technologies that help in improving the business performance. Sometimes, this concept difficult to apply for small size companies, due to high cost, limited availability of resources and many other factors. In this view, the author has introduced not only he Business Intelligence role for major companies, but also how implement the concept for small size enterprises to increase their profitability and productivity (D'Arconte, 2018).

To retain the online-gaming customers, the author has proposed an innovative model of segmentation. Various features such as performance, engagement and social interactions are considered to segment the players (Fu et al., 2017). Data driven approach is used to cluster the retail products based for the market basket data. Results are compared using k-means, SOM and hierarchical clustering approaches (Holý et al., 2017).

Impact of Big data on CRM is reviewed based on the critical success factors. Results shows three contributions like past reviews, five propositions and previous contributions (Zerbino et al., 2018). Naïve Bayes and neural networks classification approach is used to design a data mining CRM framework to enhance the decision making for retaining the customers (Bahari and Sudheep Elayidom, 2015). Streams of research in multiple fields like marketing, operations, information management and other areas are linked to develop an integrated framework (Sheng et al., 2017).

3. Methods and models

This section elaborates on the proposed objective, algorithm used and the experimental framework for the desired outcome of the study.

3.1. Selection of clustering algorithm

A cluster is understood as a conceptually meaningful group of objects that have common characteristics. Clustering can be used for customer segmentation for additional analysis. The literature survey reveals (Qadadeh and Abdallah, 2018; Arunachalam and Kumar, 2018) that one of the applications of K-means is customer segmentation. K-Means clustering algorithm is a prototype based partition clustering technique that finds the user specified number of clusters, which are represented by their centroids. K-Means is computationally faster and performs well on large datasets compared to other clustering methods. Another advantage of using K-Means is that the algorithm requires only one input parameter 'K' than other algorithms. Also it decreases the rate of misclassification of data. One of the major applications of K-Means is customer segmentation. The present work uses K-Means algorithm.

3.2. Proposed methodology

The proposed methodology can be broadly divided into 4 steps as shown in the Fig. 1. The corresponding details are explained as below:

Step 1: Exploratory Analysis and Data Preprocessing.
Exploratory data analysis (EDA) refers to initial exploration of data in order to extract or discover the patterns with the help of statistics or graphical representations. In this activity, EDA helps in identifying unique customers, percentage of orders by top 10 or more, information about the data, mismatch in description, stock code and to check null values. Further, data preprocessing is applied to identify and remove missing customer identification number, negative transactions and so on.
Step 2: a) Execution of RFM Analysis
After data is preprocessed, check for recent transactions, frequency and the amount spent by the customers. In order to create recency variable, decide the reference date - that is one day prior to the last transaction. RFM analysis is a very popular customer segmentation and identifiable technique in database marketing (Christy et al., 2018). It is significant especially in Retail Industry. Each customer under RFM is scored based on three factors.

- *Recency:* It refers to the number of days before the reference date when a customer made the last purchase. Lesser the value of recency, higher is the customer visit to a store.
- *Frequency:* It is the period between two subsequent purchases of a customer. Higher the value of Frequency, more is the customer visit to the company.
- *Monetary:* This refers to the amount of money spent by a customer during a specific period of time. Higher the value, more is the profit generated to the company.

Step 2: b) K-Means algorithm is applied using Euclidean distance metric to partition the customers for RFM values. K-Means is used twice to analyze the amount obtained for Recent and Frequent transactions as mentioned below:

i) To partition the customers based on the amount generated with recent transactions.
ii) To group the customers on the amount generated with frequent transactions.

Step 3: a) Calculation of Silhouette Score
Clusters obtained in step 2.b) are evaluated using silhouette score, which analyzes how well the resulting clusters are separated. It lies in a range of $[-1, +1]$. If the value is near to +1, then objects (customers) are grouped far away from neighboring clusters, whereas if it is -1, then objects (customers) might have been assigned to a wrong cluster or preprocessing of data is not correct.
Step 4: Evaluation of clusters
Let K = number of clusters. Silhouette values are compared for K = 3 and K = 5 to identify the optimal clusters based on the value. After the analysis, compare the sales recency with sales

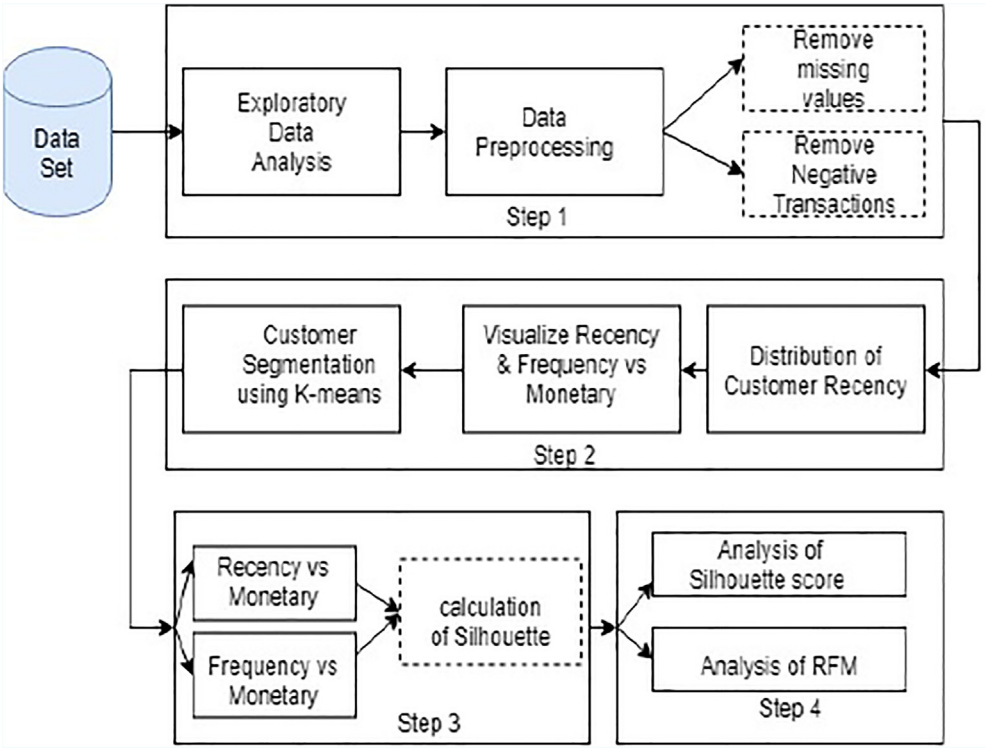

Fig. 1. Steps of Proposed Methodology.

amount and sales frequency with sales amount from one cluster to another cluster respectively. This helps in identifying the group of customers having highest sales recency, sales frequency and the sales amount.

3.3. Mathematical model

Clustering using K-means algorithm is a method of unsupervised learning used for data analysis. This algorithm identifies 'K' centroids from the dataset 'D' and assigns the non- overlapping data points to each of the nearest clusters. The intra-cluster distance is maximum compared to inter-cluster distance in K-means algorithm. Since it is an iterative approach, data points are moved to different clusters, based on the centroids calculation.

As per the pseudo algorithm shown in Fig. 2, the mathematical model for the manual calculation of silhouette for an object is given below. Consider K clusters of which each cluster contains variable objects. Since K-Means is applied twice in the present experiment, objects are clustered based on a customer transaction data for recency vs monetary and frequency vs monetary values.

$$K = \{\{(p_1, q_1), (p_2, q_2) \dots (p_x, q_x)\}, \{(p_1, q_1), (p_2, q_2) \dots (p_y, q_y)\}, \dots \{(p_1, q_1), (p_2, q_2) \dots (p_z, q_z)\}\}$$

where,

K = number of clusters, (p,q) = object in a cluster.

Identify any point, for example $\{p_1, q_1\}$ in cluster 1. Objects in a cluster represents RFM values. Calculate the average distance from $\{p_1,q_1\}$to all objects of the same cluster (intra distance value a_i). Calculate average distance from $\{p_1,q_1\}$to the objects of other clusters as given in the Eq. (1).

$$\sqrt{\sum_{i=1}^{n}(p1 - pi) + (q1 - qi)^2} \quad (1)$$

Repeat the same procedure for other clusters and find the minimum average distance from {p1, q1} of cluster 1 to cluster 2,3…. m (y(i)). Find the silhouette coefficient s(i) for cluster 1 using the following Eq. (2).

$$s(i) = \begin{cases} 1 - \frac{a_i}{b_i}, a_i < b_i \\ 0, a_i = b_i \\ \frac{a_i}{b_i} - 1, a_i > b_i \end{cases} \quad (2)$$

where,

a_i is the minimum average distance from object $\{p_1, q_1\}$to all other objects in the same cluster.
b_i is the minimum average distance from $\{p_1, q_1\}$to all other clusters, which does not contain $\{p_1, q_1\}$.

comparably calculate silhouette values for cluster 2,3…n by repeating the above steps. The cluster with highest silhouette value is the best as per the evaluation method. Compute the mean silhouette value of all objects to evaluate for whole cluster.

3.4. Data set description used in the analysis

Description of synthetic dataset is shown in Table 1. Real time dataset is collected from a logistics company in India. A fifteen-day customer transactions are collected from one of the physical Retail Store and is shown in Table 2.

4. Experiments, results and discussions

The proposed methodology is implemented on the synthetic dataset of one-year customer transactions obtained UCI repository (Chen et al., 2012). The dataset consists of 8492 instances of information on customer purchase from 1-12-2010 to 09-12-2011 with eight attributes. Missing values, negative transactions, mismatch in stock code and description are handled using data preprocessing. For the modified dataset, apply RFM analysis and K-Means clustering. The same methodology is applied for Real time dataset.

Input:
M: Dataset with 'n' instances
K: clusters in number

Output:
Dataset partitioned into 'K' clusters
Algorithm:

1. Choose arbitrarily 'k' random points from M as the cluster centers
2. **repeat**
3. Reassign each object to the clusters based on calculation of mean value.
4. Revise the cluster means that is recalculate the mean value of each cluster
5. **Until**
6. there **is** no change in the clusters obtained
7. Evaluation using silhouette coefficient: calculate average distance from objects in the same cluster and calculate average distance from objects to all other clusters
8. Calculate silhouette coefficient as below:

$S_i = (b_i - a_i) / \max(a_i, b_i)$ for $a_i > b_i$

Where,
S_i represents silhouette coefficient
a_i is average distance from i^{th} object to all other objects in a cluster.
b_i is average distance from i^{th} object to any cluster not containing the object. Calculate the minimum such value with respect to all the clusters.

Fig. 2. Algorithm -K means.

Table 1
Dataset Description: Online Retail.

Sl. No	Name of the Attribute	Type of the Attribute	Description of the Attribute
1	Invoice Number	Nominal	Six-digit number uniquely assigned for each transaction.
2	StockCode	Nominal	Five-digit unique number assigned to each distinct product.
3	Description	Nominal	Name of the product
4	Quantity	Numeric	Quantities of each product per transaction
5	InvoiceDate	Numeric	Date and time of each transaction generated of x attribute
6	UnitPrice	Numeric	Product price per unit of
7	Customer Id	Numeric	Five digit unique number assigned to a customer.
8	Country	Character	Name of the country

Table 2
Dataset Description-Real time.

Sl. No	Name of the Attribute	Type of the Attribute	Description of the Attribute
1	Bill Number	Nominal	Three-four digit unique number assigned for each transaction.
2	StockCode	Numeric	Six -digit unique number given to each product.
3	Description	Nominal	Product name
4	Quantity	Numeric	Quantities of each product per transaction
5	Date	Numeric	Date of each transaction generated
6	UnitPrice	Numeric	Product price per unit
7	Customer Id	Numeric	Five digit unique number assigned to a customer
8	Customer name	Nominal	Name of the customer
9	Location	Nominal	Address of the Customer

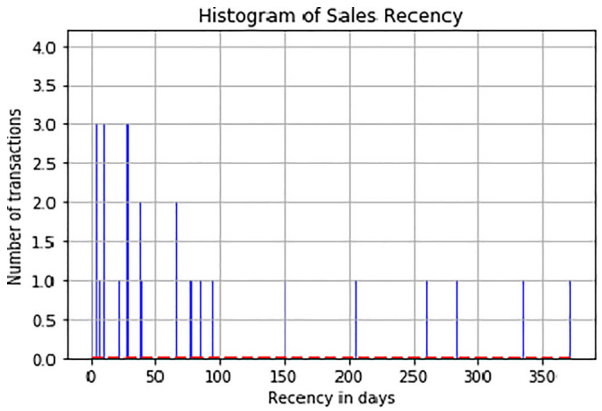

Fig. 3. Visualization of Sales Recency.

Fig. 3 represents balanced distribution of sales Recency with a moderate frequent number of transactions and a fairly uniform number of sales in recent transactions. Similarly, Frequency and monetary values are calculated and the result of RFM for first five customers is shown in Table 3.

Table 3
RFM values.

Index	Customer ID	Recency	Amount	Frequency
1	12413.0	67.0	758.101	38
2	12437	2.0	4951.411	200
3	12441.0	367.0	173.551	11
4	12488.0	10.0	1298.661	55
5	12489.0	336.0	334.931	15

Table 4
Silhouette score for K = 3 and K = 5 clusters.

No of Clusters	Recency_log	Frequency_log	Amount_log
0	161.191479	16.761959	291.852580
1	11.373230	209.371490	5316.800437
2	20.323096	48.877509	894.321423
Silhouette Score for number of cluster 3 is 0.362159752			
0	210.331020	10.265222	170.590293
1	20.309899	95.243198	1913.827582
2	10.779640	280.086131	7456.390843
3	81.290783	32.515424	619.014428
4	3.847593	41.650355	708.639718
Silhouette Score for number of cluster 5 is 0.3490755342			

From Table 4, RFM log is calculated for K = 3 and K = 5.It is observed that the results of silhouette score matrix for K = 5 is less optimal compared to K = 3. The value of silhouette nearer to +1 represents optimal comparatively to other clusters. Assignment of customer to different clusters are shown in Table 5.

The visualization of silhouette plot for K = 3 and K = 5 is shown in Figs. 4 and 5 respectively. Similar analysis is applied for Recency and Monetary values as shown in Fig. 6. Cluster analysis from Table 4, shows that Segmentation of customers for K = 3 is more optimal than K = 5. Box plot for K = 3 is shown in Figs. 6a–c.

According to Fig. 6a, cluster 1 is having the highest sales recency. Cluster 2 is having a highest sales frequency as shown in Fig. 6b and it is interesting to note that cluster 2 is also having the highest sales amount compared to other clusters. Similar analysis based on RFM for real time data is shown in Table 6.

5. Conclusion and scope of future work

Customer segmentation based on the buying pattern of customers though strategically important, is an equally challenging task. Customer retention is another major concern for both online and physical enterprises. In the present work, the RFM model is implemented for synthetic and real datasets, to analyze customer segmentation. Also, clusters are evaluated using Silhouette Analysis for K-Means clustering algorithm with different number of clusters. Based on the Silhouette Score, the Sales Recency, Sales Frequency and Sales Monetary can be analyzed and an optimal solution is found.

Table 5
Assignment of cluster labels.

Customer ID	Recency	Amount	Frequency	Recency_log	Frequency_log	Amount_log	Cluster5 labels	Cluster3 labels
12413.0	67.0	758.101	38	4.204693	3.637586	6.630817	3	2
12437.0	2.0	4951.411	200	0.693147	5.298317	8.507428	2	1
12441.0	367.0	173.551	11	5.905362	2.397895	5.156472	0	0
12488.0	10.0	1298.661	55	2.302585	4.007333	7.169089	1	2
12489.0	336.0	334.931	15	5.817111	2.708050	5.813925	0	0

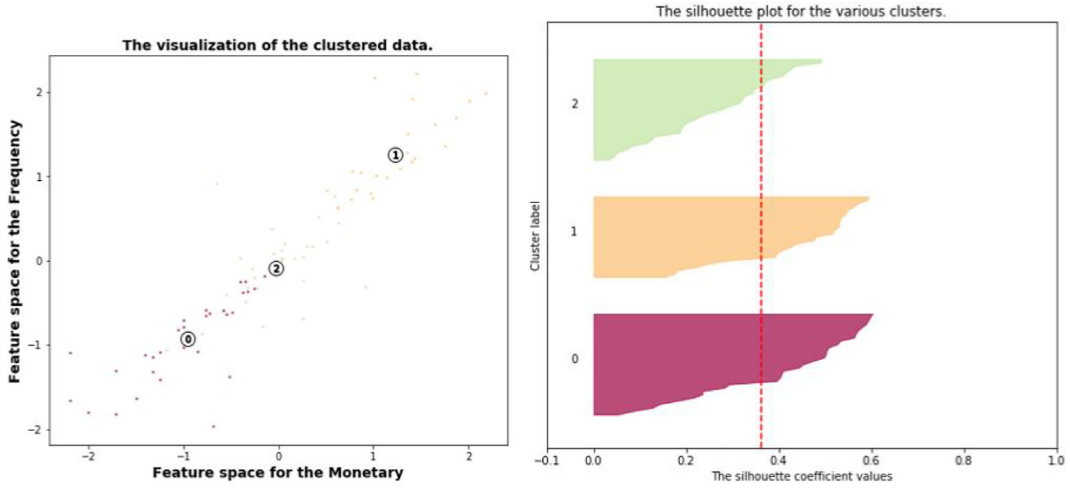

Fig. 4. Visualization of Cluster K = 3 for Frequency vs Monetary and corresponding silhouette coefficient values.

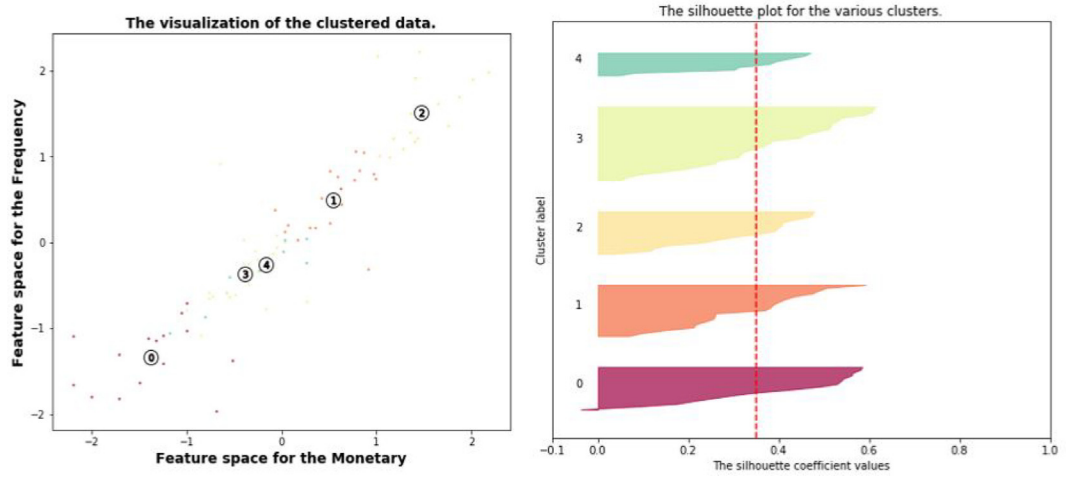

Fig. 5. Visualization of Cluster K = 5 for Frequency vs Monetary and corresponding silhouette coefficient values.

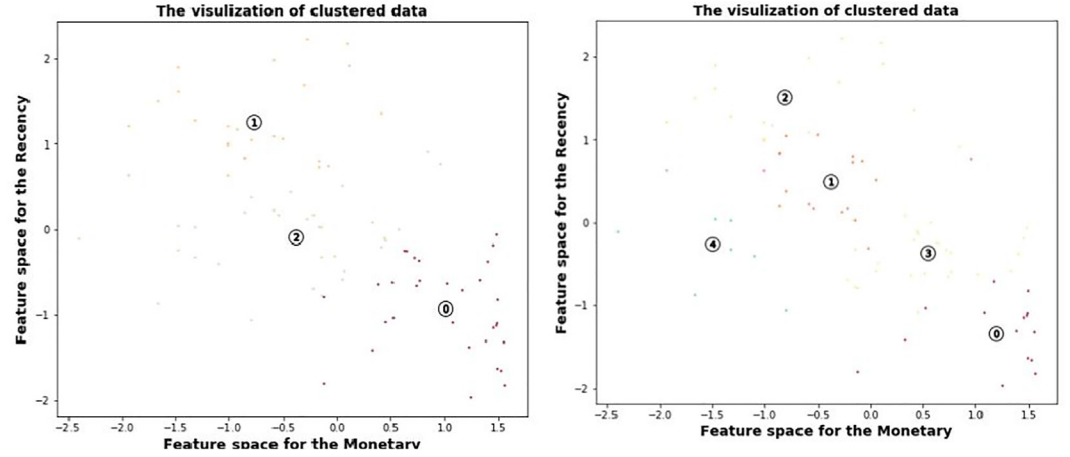

Fig. 6. Visualization of Recency vs Monetary for K = 3 and K = 5 clusters.

The scope of future work in this area lies in the study and analysis of specific categories of products, for example, Mobile Phones and Accessories. Various other business parameters such as the most preferred product or the most effective sales technique during as specific event, or some threshold parameters in different regions can be studied for designing effective business enhancement. Such advancements and deliberations in this area will help the enterprises to improve businesses by offering promotions and designing innovative strategies that can prove cutting edge against the competitors.

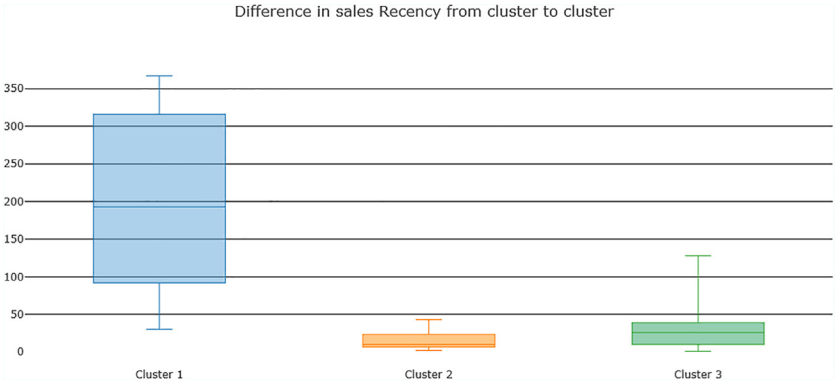

Fig. 6a. Visualization of difference in sales recency

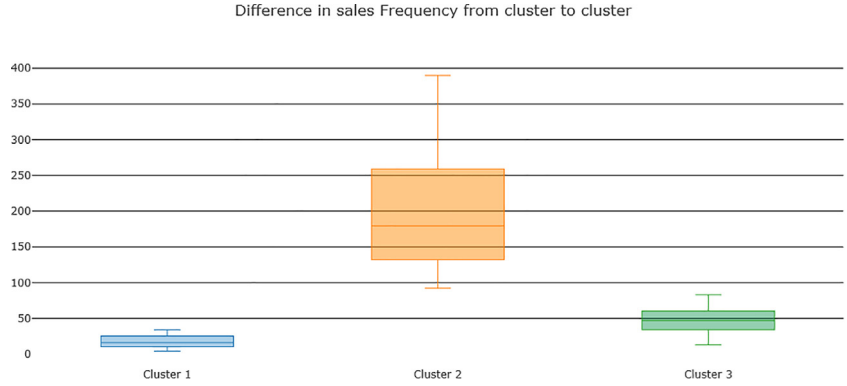

Fig. 6b. Visualization of difference in sales frequency

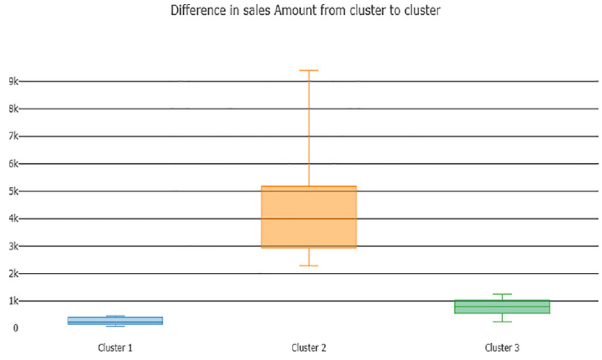

Fig. 6c. Visualization of difference in sales amount.

Table 6
Silhouette score for real time data.

No. of Clusters	Recency_log	Frequency_log	Amount_log
0	10.591432	46.571136	7395.997897
1	2.000000	115.430465	24945.010775
2	1.000000	1312.355135	612006.400895
Silhouette Score for number of cluster 3 is 0.362052482			
0	6.633250	107.144762	4.006409e+04
1	1.414214	297.060600	1.413740e+05
2	10.458643	32.268573	4.026655e+03
3	1.414214	1.298395e+04	1.298395e+04
4	1.000000	5172.000000	1.228960e+06
Silhouette Score for number of cluster 5 is 0.381312431			

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors wish to acknowledge JSS Academy of Technical Education, Bengaluru, for providing the facilities to carry out the research work.

References

Arunachalam, Deepak, Kumar, Niraj, 2018. Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data-driven decision-making. Expert Syst. Appl. 111, 11–34.

Bahari, T. Femina, Sudheep Elayidom, M., 2015. An efficient CRM-data mining framework for the prediction of customer behaviour. Procedia Comput. Sci. 46, 725–731.

Brito, Pedro Quelhas et al., 2015. Customer segmentation in a large database of an online customized fashion business. Rob. Comput. Integr. Manuf. 36, 93–100.

Carnein, Matthias, Trautmann, Heike, 2019. Customer segmentation based on transactional data using stream clustering. Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham.

Chen, Daqing, Sain, Sai Liang, Guo, Kun, 2012. Data mining for the online retail industry: a case study of RFM model-based customer segmentation using data mining. J. Database Mark. Customer Strategy Manage. 19 (3), 197–208. https://doi.org/10.1057/dbm.2012.17.

Christy, A., Joy, A., Umamakeswari, L. Priyatharsini, Neyaa, A., 2018. RFM ranking–an effective approach to customer segmentation. J. King Saud Univ.-Comput. Inf. Sci.

D'Arconte, Carmine, 2018. Business intelligence applied in small size for profit companies. Procedia Comput. Sci. 131, 45–57.
Fu, Xin et al., 2017. User segmentation for retention management in online social games. Decis. Support Syst. 101, 51–68.
Griva, Anastasia et al., 2018. Retail business analytics: customer visit segmentation using market basket data. Expert Syst. Appl. 100, 1–16.
Han, Jiawei, Pei, Jian, Kamber, Micheline, 2011. Data Mining: Concepts and Techniques. Elsevier.
Holý, Vladimír, Sokol, Ondřej, Černý, Michal, 2017. Clustering retail products based on customer behaviour. Appl. Soft Comput. 60, 752–762.
Hu, Ya-Han, Yeh, Tzu-Wei, 2014. Discovering valuable frequent patterns based on RFM analysis without customer identification information. Knowl.-Based Syst. 61, 76–88.
Kaur, Arvinder, Pal, Saibal Kumar, Singh, Amrit Pal, 2019. Hybridization of chaos and flower pollination algorithm over K-means for data clustering. Appl. Soft Comput. 105523.
Khajvand, Mahboubeh et al., 2011. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study. Procedia Comput. Sci. 3, 57–63.
Khalili-Damghani, Kaveh, Abdi, Farshid, Abolmakarem, Shaghayegh, 2018. Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case of customer-centric industries. Appl. Soft Comput. 73, 816–828.
Kolarovszki, Peter, Tengler, Jiří, Majerčáková, Margita, 2016. The new model of customer segmentation in postal enterprises. Procedia-Soc. Behav. Sci. 230, 121–127.
Murray, Paul W., Agard, Bruno, Barajas, Marco A., 2017. Market segmentation through data mining: a method to extract behaviors from a noisy data set. Comput. Ind. Eng. 109, 233–252.
Nguyen, Dung H., de Leeuw, Sander, Dullaert, Wout E.H., 2018. Consumer behaviour and order fulfilment in online retailing: a systematic review. Int. J. Manage. Rev. 20 (2), 255–276.
Patak, Michal et al., 2014. The e-pharmacy customer segmentation based on the perceived importance of the retention support tools. Procedia-Soc. Behav. Sci. 150, 552–562.
Qadadeh, Wafa, Abdallah, Sherief, 2018. Customers Segmentation in the Insurance Company (TIC) Dataset. Procedia Comput. Sci. 144, 277–290.
Sheng, Jie, Amankwah-Amoah, Joseph, Wang, Xiaojun, 2017. A multidisciplinary perspective of big data in management research. Int. J. Prod. Econ. 191, 97–112.
Singh, Ashishkumar, Grace Rumantir, Annie South, 2014. "Market Segmentation of EFTPOS Retailers." AusDM.
Song, M., Zhao, X., Haihong, E., Ou, Z., 2017. Statistics-based CRM approach via time series segmenting RFM on large scale data. Knowl.-Based Syst. 15 (132), 21–29.
Yoo, Byungjoon, Jang, Moonkyoung, 2019. A bibliographic survey of business models, service relationships, and technology in electronic commerce. Electron. Commer. Res. Appl. 33, 100818.
You, Zhen, Si, Yain-Whar, Zhang, Defu, Zeng, XiangXiang, Leung, Stephen CH, Li, Tao, 2015. A decision-making framework for precision marketing. Expert Syst. Appl. 42 (7), 3357–3367.
Zerbino, Pierluigi et al., 2018. Big data-enabled customer relationship management: a holistic approach. Inf. Process. Manage. 54 (5), 818–846.