*Journal of Infrastructure, Policy and Development* 2024, 8(7), 4426.
https://doi.org/10.24294/jipd.v8i7.4426

Article

# Application of data mining technique for customer purchase behavior via Extended RFM model with focus on BCG matrix from a data set of online retailing

**Soma Gholamveisy[1,*], Seyedamirmasoud Homayooni[2,*], Milad Shemshaki[3], Sogand Sheykhan[4], Payam Boozary[4], Hamed Ghorban Tanhaei[4], Nasrin Akbari[5]**

[1] Department of Industrial Engineering, Islamic Azad University South Tehran Branch, Tehran 1584743311, Iran
[2] Department of Digital Electronic System, Iran University Science and Technology, Tehran 13114-16846, Iran
[3] Department of Business and Law, University of East London, London E1541, United Kingdom
[4] Department of Management, Science and Technology, AmirKabir University of Technology, Tehran 15875-4413, Iran
[5] Department of International Law, Islamic Azad University of Mashhad, Neyshabour 9319974144, Iran
*** Corresponding authors:** Soma Gholamveisy, info@researchcenter-soma.com; Seyedamirmasoud Homayooni, Amirmasoudhomayouni@gmail.com

**CITATION**

Gholamveisy S, Shemshaki M, Homayooni S, et al. (2024). Application of data mining technique for customer purchase behavior via Extended RFM model with focus on BCG matrix from a data set of online retailing. Journal of Infrastructure, Policy and Development. 8(7): 4426. https://doi.org/10.24294/jipd.v8i7.4426

**ARTICLE INFO**

Received: 26 January 2024
Accepted: 4 April 2024
Available online: 31 July 2024



**Abstract:** This study explores the integration of data mining, customer relationship management (CRM), and strategic management to enhance the understanding of customer behavior and drive revenue growth. The main goal is the use of application of data mining techniques in customer analytics, focusing on the Extended RFM (Recency, Frequency, Monetary Value and count day) model within the context of online retailing. The Extended RFM model enhances traditional RFM analysis by incorporating customer demographics and psychographics to segment customers more effectively based on their purchasing patterns. The study further investigates the integration of the BCG (Boston Consulting Group) matrix with the Extended RFM model to provide a strategic view of customer purchase behavior in product portfolio management. By analyzing online retail customer data, this research identifies distinct customer segments and their preferences, which can inform targeted marketing strategies and personalized customer experiences. The integration of the BCG matrix allows for a nuanced understanding of which segments are inclined to purchase from different categories such as "stars" or "cash cows," enabling businesses to align marketing efforts with customer tendencies. The findings suggest that leveraging the Extended RFM model in conjunction with the BCG matrix can lead to increased customer satisfaction, loyalty, and informed decision-making for product development and resource allocation, thereby driving growth in the competitive online retail sector. The findings are expected to contribute to the field of Infrastructure Finance by providing actionable insights for firms to refine their strategic policies in CRM.

**Keywords:** data mining; customer purchase behavior; Extended RFM; BCG matrix

## 1. Introduction

Nowadays, information technology has a big impact on all aspects of business, so large companies spend millions of dollars each year using e-Commerce tools and spending money on e-Commerce (Gholamveisy et al., 2023). Because attracting and retaining customers is an important and strategic resource. In recent years, customer behavior has been a well-researched topic in retail decision-making and marketing. Customers' purchase decisions are influenced by various sociological and psychological aspects, which are well-discussed (Khade, 2016). The number of purchases you can make depends on whether you shop alone, with a companion, or with your parents or spouse (Gull and Pervaiz, 2018). It describes how technology is used to influence customer behavior in-store, including how behavioral techniques are used in the world of retail and financial services (Larsen et al., 2017). Therefore, one of the most important ways to increase financial profitability in purchasing is to identify and attract more customers through big data analysis. In this regard, big data analytics are crucial for retail decision-making, especially when data are related to consumer behavior.

The proposed research topic focuses on advancing the understanding of customer purchase behavior in online retailing through the application of data mining techniques. Specifically, it aims to extend the traditional Recency, Frequency, Monetary (RFM) model by incorporating additional variables that capture customer engagement and interaction nuances. The integration of the Boston Consulting Group (BCG) matrix with this Extended RFM model will provide a more dynamic and strategic approach to customer segmentation. So, the contribution of this paper is the Extended RFM model via a data mining technique that combines four key variables. By analyzing these variables, businesses can gain valuable insights into customer behavior and make informed decisions regarding marketing strategies and customer segmentation. In the context of online retailing, the BCG matrix is a tool used to analyze a company's product portfolio. It categorizes products into four quadrants based on their market growth rate and relative market share. The quadrants are labeled as stars, cash cows, question marks, and dogs. This matrix helps businesses identify which products have the highest growth potential and which ones may require further investment or divestment. By combining the Extended RFM model with the BCG matrix, businesses can gain a comprehensive understanding of customer purchase behavior and product performance. This approach allows companies to identify high-value customers who have the potential to become loyal brand advocates. Additionally, it helps businesses allocate resources effectively by focusing on products with high growth potential.

Research objectives:

1) To develop an Extended RFM model that includes additional behavioral metrics such as count day.
2) To apply data mining techniques, including clustering algorithms (e.g., K-means) to segment customers based on the Extended RFM model.
3) To map the Extended RFM customer segments onto the BCG matrix, creating a hybrid analytical framework that links customer value with product strategy.
4) To validate the effectiveness of this hybrid framework in predicting customer purchase patterns and identifying strategic business opportunities in an online retail dataset.

### 1.1. Background

The RFM (Recency/Frequency/Monetary) model is a segmentation framework used in the retailing industry to quantify consumer values. It helps in understanding customer behavior and making market decisions (Chen et al., 2023). The RFM model has been applied in various industries, including mobile phone sales, where it has helped identify target customers and analyze consumption habits (Taşabat et al., 2023). However, the RFM model has limitations and may not consider all relevant parameters. To address this, modifications have been proposed, such as adding the "economic" variable to create the RFMS model, which aims to improve customer relations and classification accuracy (Zhang, 2023). Another modification includes adding diversity (D) as a fourth parameter to the RFM model, which enhances the prediction of customer behavior and helps companies identify customers who will respond positively (Martínez et al., 2021). In the context of online learning, the RFM model has been used to analyze student engagement and construct a model based on online learning behavior. It has been extensively used in many different domains, particularly marketing. The RFM model has been linked with data mining techniques such as Self Organizing Maps (SOM) (Gholamveisy, 2021a), K-means, and Decision Tree C4.5 to facilitate consumer segmentation and classification. Additionally, the k-means clustering algorithm has been used to determine clusters and patterns in customer data sets retrieved from e-Commerce platforms. Logistic regression models have been developed to predict customer clusters based on product price, ratings, and discounts (Asmat et al., 2023).

### 1.2. Customer purchase behavior

Customer purchase behavior is a complex process influenced by various factors such as product features, decision-making processes, and external motivators (Kazmi et al., 2021). Understanding customer behavior in different domains, such as virtual worlds and tourism, requires analyzing the impact of technology and the unique characteristics of the industry (Bali et al., 2023). In the context of smartphones, customers consider factors such as processing speed, camera quality, and storage when making a purchase decision (Solomon et al., 2012). During the COVID-19 outbreak, companies have focused on analyzing customer buying behavior to provide exceptional customer service and personalized marketing plans (Sharma et al., 2022). Techniques like clustering, association rule mining (Gholamveisy, 2021b), and logistic regression are used to segment customers, recommend products, and validate clustering operations. Overall, customer purchase behavior is influenced by a combination of individual preferences, market dynamics, and technological advancements, making it a multi-dimensional and evolving field of study.

### 1.3. Related work

For many years, the science of data mining and customer relationship management has been used in many researches. Today, many commercial organizations are using it to increase profits, financial functions, and efficiency. Some of these articles are mentioned below, and the rest are mentioned in **Table 1**.

Chen et al. (2023) implemented customer relationship management by utilizing an RFM model along with K-means clustering, and their experimental findings suggest that this proposed model is an effective method for analyzing customer value. Additionally, Khalili-Damghani et al. (2018) introduced a hybrid soft computing technique that incorporates rule extraction for decision tree methodology and clustering to predict the segmentation of new customers of customer-centric companies; this approach was applied in two case studies in the fields of telecommunications and insurance, respectively, to predict potential profitable leads and outline the most influential features available to customers during such prediction. Using the RFM model and K-means algorithm, different data set clusters are validated through the calculation of the silhouette model (Anitha and Patil, 2022), in research uses business intelligence to identify new clients by providing corporate entities in the retail industry pertinent and timely data. The information provided is based on methodical research and scientific methods for examining customer purchase patterns and sales history. This research used k-means data set segmentation techniques and is based on the RFM (Recency, Frequency, and Monetary) model. The outcomes for sales transactions are therefore compared using several criteria, including sales volume, sales frequency, and sales recency. A decision tree classification model has been used to mine online buying behavior data and estimate online shopping features in another research, a client might be a part of multiple clusters by grouping RFM analysis as a novel framework to uncover better customer consumption behavior. and be connected to various allegiances and contributions about various features of acquired goods (Cui, 2023). **Table 1** shows the other research in the finance and banking industry

**Table 1.** Other researches in the finance and banking industry.

| Reference | Technique used | Case study | The features examined | Years |
|---|---|---|---|---|
| (Ben Ncir et al., 2023) | Genetic algorithm | Banking industry | Customer behavior prediction | 2023 |
| (Chayjan et al., 2020) | Partial least squares structural equation modeling | Banking sector | Customer behavior | 2016 |
| (Gholamiangonabadi et al., 2019) | Neural Network (NN), Radial Basis Function (RBFNN), Regression (GRNN), (MLPNN) Support Vector Machine (SVM) | Banking industry | Customer churn prediction | 2019 |
| (Audzeyeva et al., 2012) | Markov chain | Multi-service financial organization | Forecasting customer behavior | 2012 |
| (Bellou and Andronikidis, 2008) | Statistical method | Banking sector | Customer service behavior | 2008 |
| (Vasiljeva et al., 2021) | Statistical method | Banking sector | Customer service behavior | 2021 |
| (Corboș et al., 2023) | Bibliometric analysis | Sustainable fashion | Consumer behavior | |
| (Corbos et al., 2023) | Statistical method | Household appliances | Customer buying trust | |

According to **Table 1**, no research using the method mentioned in this article was not observed. This research has the potential to bridge the gap between traditional customer segmentation models and the complex realities of e-commerce behavior, providing a robust framework for leveraging data mining techniques in strategic business applications BCG matrix integration: Map the resulting clusters onto the BCG matrix framework to categorize customers into segments such as “Stars,” “Cash Cows,” “Question Marks,” and “Dogs” based on their purchasing patterns and potential value.

## 2. Method

Considering the dataset, the combined approach in this investigation can considered in three phases, and the structure of the proposed approach is shown in **Figure 1**. The software used in this article is MATLAB and rapid miner to use preprocessing data.

- Data preprocessing;
- Segmentation of the Extended RFM model with the k-means model;
- Combining the categories with the shared growth matrix.

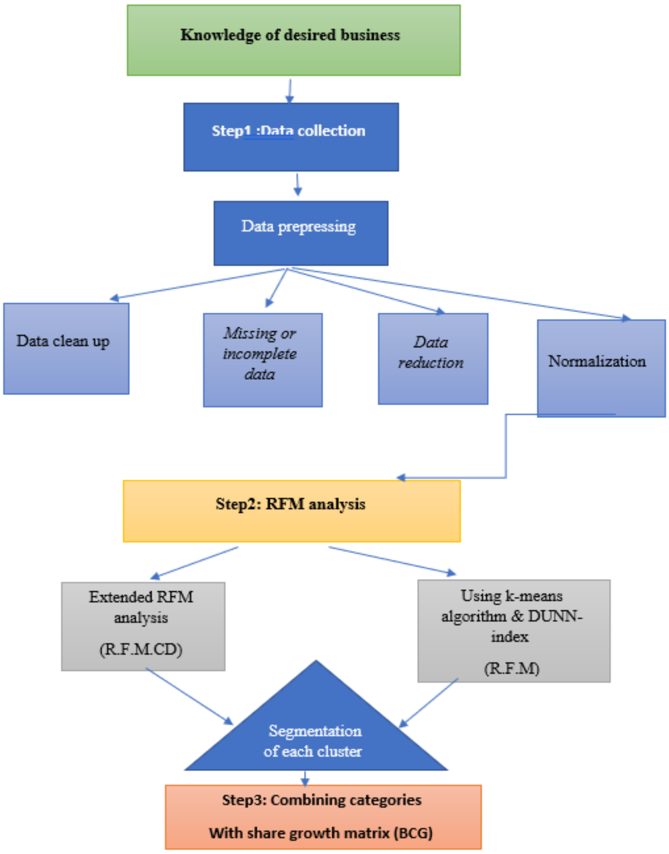


**Figure 1.** The structure of the proposed approach.

The steps of research based on the schematic representation of the suggested methodology as depicted in figure.

Step 1: data collection

Step one involves using statistical methods or visual tools to detect patterns in data. This process aids in recognizing distinct customers, calculating the percentage of orders from the top customers, and identifying data inconsistencies and null values. Issues such as missing customer IDs, negative transactions, and more are detected and addressed during data pre-processing.

Step 2: RFM analysis

After pre-processing data, conduct an RFM Analysis to assess recent transactions, purchase frequency, and customer spending amounts. Furthermore, select a reference date, ideally the day before the latest transaction, to determine the customer recency variable. RFM analysis is widely used in database marketing for customer segmentation and identification, especially in retail. Customers are scored on three key metrics: recency, frequency, and monetary value, as previously described.

Step 2-1: Extended RFM analysis

The variable count days has been added to the model as a new factor, and it is considered as a new variable and its effect on the optimal clusters.

Step 2-2: Using the k-means algorithm to cluster the customers

The consumer segmentation process uses the clustering algorithm. In this research, two modeling techniques for customer clustering are presented. in this research. Extended RFM parameters are only incorporated into the clustering process in the first method. Clustering of k-means is used to divide the customer.

## 2.1. Cluster evaluation to obtain the optimal k

The effectiveness method must be assessed because each clustering technique produces unique results. Different evaluation standards for clustering algorithms may be categorized into internal criteria and unsupervised evaluation standards.

To evaluate using these criteria, finding the quality of the clustering operations using the dataset's data is necessary. Optimally maximizing and minimizing the intra-cluster distance is the most crucial task of a clustering algorithm. Maximizing each cluster's density and minimizing the distance between groups are the goals of maximizing and minimizing intra-cluster space, respectively. All non-observer evaluation criteria are comparable in maximizing the density and clustering factors. Moreover, various indicators are introduced. We employ the Dunn index in this study. As an internal benchmark index for evaluation clustering, introduces the index (Ncir et al., 2021). The process of this index is as follows.

$$D = \min_{1 \leq i \leq} \left[ \min_{i+1 \leq j < k} \left[ \frac{\text{dist}(a_i, a_j)}{\max\limits_{i \leq j \leq k} \text{dim}(a_j)} \right] \right]$$

where,

$\text{dist}(a_i, a_j)$ is the internal cluster distance between cluster $a_i$ and $a_j$.

where,

$$\text{dist}(a_i, a_j) = \min_{y_i \in a_i, y_j \in a_j} (y_i, y_j) \tag{1}$$

$\text{d}(y_i, y_j)$ is the distance between data $y_i \in a_i$ and $a_j$, diam($a_i$) is the diameter of the cluster $a_i$.

$$\text{Diam}(a_i) = \max_{y_{i1}, y_{j2} \in a_i} (y_{i1}, y_{j2})$$

An optimal value of the k maximizes the Dunn's index (Luna-Romera et al., 2016).

The BCG matrix, created by the Boston Consulting Group, is a strategic planning tool that assists businesses in their long-term planning and decision-making about product portfolios. By assessing their range of products, companies can determine in which areas to invest, discontinue, or develop, thereby identifying growth opportunities. The BCG matrix is also called the Growth/Share Matrix.

The portfolio planning matrix, often referred to as the BCG matrix, is widely recognized as the most renowned and straightforward tool for strategic portfolio planning. It emphasizes the importance of maintaining a well-rounded product mix within a company's range. The BCG matrix has gained significant prominence in the business world and is regarded as one of the most influential strategic tools ever developed (Hambrick et al., 1982). The BCG matrix for this paper is shown in **Figure 2**.

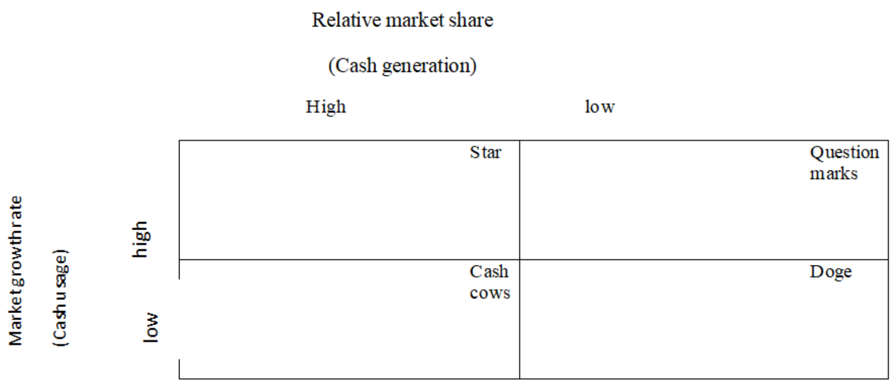


**Figure 2.** The BCG matrix.

1) Dogs: These products are characterized by low growth rates and market shares.
2) Question marks or problem child: These products are in high-growth markets with low market shares.
3) Stars: These products thrive in high-growth markets and enjoy substantial market shares.
4) Cash cows: These are products that operate within low-growth markets but possess significant market shares

## 2.2. Data description

Business intelligence systems and mathematical models for decision-making lead to accurate and effective results if the input data are reliable. Occasionally, data collected from the main sources and data booths have defects that need to be identified and corrected by analysts.

In this stage, which is the most basic stage, the data in the online retailing database is prepared for data mining operations and analyses. The data recorded provides another perspective suitable for unsupervised problems such as customer pouch behavior from online retailing The data set is mostly adequate for creating indicators of the Extended RFM model and labeling purchases via the clustering k-means algorithm and segmenting which has not been observed in any research. This data allows us to communicate more accurately with customers and improve sales strategies using the BCG matrix method. The Online Retail II data set encompasses all the transactions happening for a UK-based and registered, non-store online retail (Data period; 1 December 2009 and 9 December 2011).

- Attribute data:
  - Invoice number: A unique 6-digit number assigned to each trade. Numbers starting with "c" indicate cancelation.
  - Stock/item code: A unique 5-digit number assigned to each separate product or item.
  - Description: The name of the product or item.
  - Quantity: The amount of each product included in the transaction.
  - Invoice date: The date and time assigned to the invoice when the transaction occurred.
  - Unit price: The price per unit of product in British pounds.
  - Customer ID: A unique 5-digit number assigned to each customer.

- Country: The name of the country where the customer resides.

In this paper, data are prepossessed via data clean-up, missing or incomplete data, data reduction, data conversion, and outlier data.

Consistent with the RFM theory, three variables in identifying customer behavior are introduced as crucial communication variables in most articles. However, the variables are derived from the previous stage of preprocessing of the data set.

- Recency: the number of days since a customer last made a purchase. The more frequently a customer visits a store, the less prominence recency has.
- Frequency: When a customer makes two more purchases after the first. However, the more customers come into the store, the higher the frequency value.
- Monetary: This refers to the sum of money a customer spent during a particular period.
- Count day: The number of different days that the purchase occurs.

## 3. Discussion & finding

Within this section, we analyze the data performance of its competitors across 344 datasets using the k-means algorithm. Unlike previous prior studies, we assess the outcomes of cluster using both objective function measures and the quality measures of cluster. It should be emphasized that cluster quality measures depend on predetermined labels, while the objective function measures can operate independently without such labels.

To normalize the data, we standardized min-max in the range of (0, 1) in rapid miner software for all values of three features. Frequency, Recency, Monetary, and count day and satisfaction variable are used in the second step.

In this step, first, we apply the Dunn-index to obtain the best optimal number of clustering (k). The best optimal k in this investigation is four clusters. Therefore, four groups are selected to segmenting customer purchase behaviors in k-means clustering. Results are shown in **Table 2** and **Figures 3** and **4**.

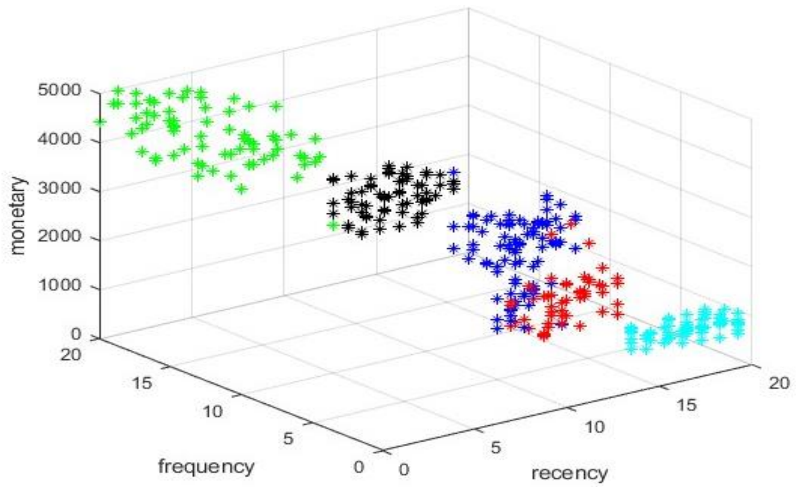


**Figure 3.** User clustering scatterplot based on k-means algorithm.

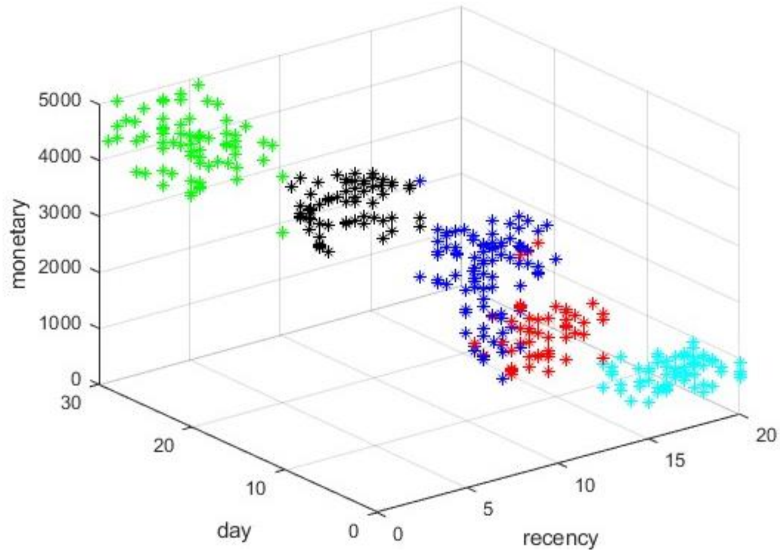


**Figure 4.** User clustering scatterplot based on K-means algorithm (new factor).

**Table 2.** The result of k-means clustering.

| **Cluster** | ***R*** | ***F*** | ***M*** | **Cd** |
|---|---|---|---|---|
| C1 | 23 | 12 | 15,000 | 14 |
| C2 | 81.48 | 22.6 | 905.96 | 92 |
| C3 | 41 | 26.67 | 173.66 | 38 |
| C4 | 58.68 | 10.14 | 1658.93 | 80 |

The summary of the value Dunn-index for each and k-means clustering are shown in **Table 3** and **Figure 5**, respectively.

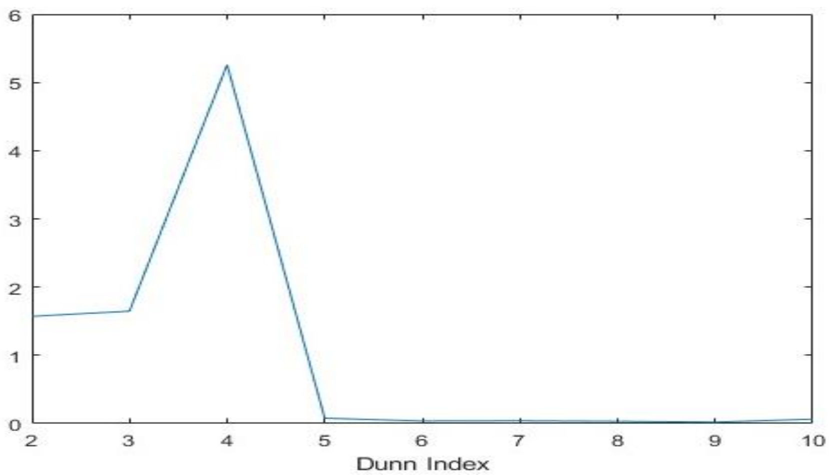


**Figure 5.** Clustering evaluation with Dunn index (2–10).

**Table 3.** The value of the Dunn index for each cluster.

| **Cluster** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|
| **DUNN-index** | 1.57 | 1..64 | 1.23 | 1.07 | 1.03 | 1.04 | 1.03 | 1.02 | 1.06 |

## 3.1. Explain the cluster

Since the customer identification number field is considered ineffective in clustering, we may compare customers' behavior in the online shopping process. The results of the comparison of procurement behavior are the following:

Cluster 1: this group has high *M* and *F*, but the *R* is extremely low and includes the largest number of customer members. However, they cover the firm with good and permanent customers with two factors: *F*&*M* are above the average. and in 14 days, the purchase happened.

Cluster 2: in this cluster, the group members have three factors: *R*, *M*, and *F* above the average, in addition to the fact that they have always bought and brought a high income to the firm. It should be said that this cluster includes the firm's excellent customers and in 92 days, the purchase happened.

Cluster3: in this cluster, having three factors *R*, *F*, and *M* below the average includes customers who have gone to the competitor, and in 38 days, the purchase happened

Cluster 4: this group with only a high factor *M* above the average is considered firm customers in this cluster. It shows that these customers have turned to the firm recently. In 80 days, the purchase happened.

Following the clustering, the customer purchase behaviors of each cluster were prioritized and calculated based on an RFM model. Accordingly, prioritizing purchase is equal to the average of the variables' values of each problem. The clusters were classified into four grades: Gold, Platinum, Bronz, and Silver, in line with their average value index, which means that the clusters that have a close average value to each other were assigned the same value class (Alghanam et al., 2022). **Table 4** shows the segmentation of customer purchase behaviors for each cluster.

**Table 4.** Segmentation of customer purchase behaviors for each cluster.

| Number of clusters | Class code | Classification value |
|---|---|---|
| C1 | Class 1 | Gold |
| C2 | Class 4 | Platinum |
| C3 | Class 2 | Bronz |
| C4 | Class 3 | Silver |

### 3.2. Combining the categories obtained with the growth matrix: The contribution of Boston Consulting Group

This section uses the growth-share matrix to analyze the customer behavior regarding market share and growth created for the related firm. The "Growth-share matrix" concentrates on different departments of the firm and the type of profitability these departments have. Moreover, it is drawn on the base of the firm's growth rate and relative market share. However, because of the similarity among the concepts of liquidity and market growth, they are displayed on the chart's horizontal axis because the market share is correlated with recent purchases. They are also given on the vertical axis.

### 3.3. Further analysis of the four groups in line with the growth matrix, the contribution of the Boston Consultants Group

Platinum, which makes up the second cluster, is our excellent customer. This group is part of the matrix star group since they hold a high purchase share (newness of purchase) and contribute significantly to the firm's market growth (liquidity). Furthermore, it was advised that the business consider keeping these loyal customers. Moreover, with solutions such as particular discounts and incentive prizes, one may maintain the employees in this group can be permanently maintained if we research the kinds of products that customer produce, meet their deadlines for the delivery of orders, and expand services such as transportation.

Gold, who form our first cluster of good customers, is a group of 15 retailers related to supermarkets and sales centers. Until now, the firm mistakenly thought that retailers were not significant, but with this analysis, it became clear that most customers are in this category. Moreover, the class includes customers who bring good market growth (liquidity) to the firm but have a low share of recent purchases (recentness of purchase) and are placed in the milking cow group of the matrix. Since this cluster is the closest to ordinary individuals and regular consumers, there is a potential for advertising in this group. As **Table 4** displays, there is a high purchase frequency in the golden group, which is in addition to the growth of the Anqedingi market. A good purchase of the dairy cow group in the matrix gives the status of a permanent purchase from the firm. Banners and advertisements for this group, and even free ads for retailers and supermarkets, may make this brand better known to the main customers or the people and end consumers. Paying attention to this group is significant since they should be considered regular customers, and the firm should try to keep and satisfy these customers more than ever.

The Silver category that makes up our fourth cluster includes groups that have recently become customers of the firm because they have a high share of purchases but bring low liquidity growth for the firm. However, the firm's goal should be to work with these new customers because people in this group may become better, and if the firm's products are well known, these customers become regular and good customers. The cluster is placed in the uncertain group, and as a result of low liquidity growth, they may be drawn to the dead dog group, but it has the potential to improve and become a star group or milk cow.

The bronze category, our third pleasure, does not benefit the firm. They have bought several times before but have not heard from them for a long time. In fact, after procurement from the firm several times, this group has concluded that the firm probably does not meet their needs and that the products and services are not what they want. They wanted and probably found the answer to their needs in competing companies. This topic is likely to be the subject of another research project on this topic. This group of people is fugitives from the firm.

For this reason, the bronze group is included in the group of dead dogs, which consists of fugitive customers, and the firm must carefully find the reason for the fugitives of these customers and, in line with proper analysis, find a way to deal with them.

Further analysis of the four groups in line with the growth matrix, the contribution of the Boston Consultants Group is shown in **Figure 6**.

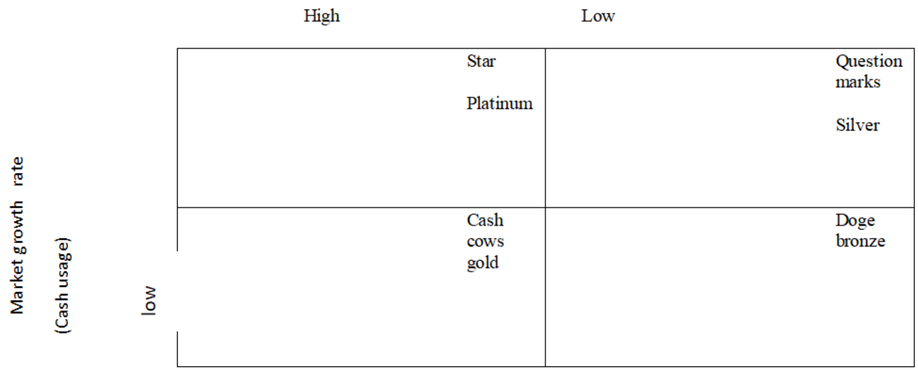


**Figure 6.** Growth share-matrix.

## 4. Discussion

In the field of customer analytics, data mining techniques are widely employed to gain insights into customer purchase behavior. One such technique is the Extended RFM (Recency, Frequency, Monetary Value) model, which allows businesses to segment their customers based on their purchasing patterns. This model takes into account not only the recency, frequency, and monetary value of purchases but also incorporates additional variables such as count day.

In the context of online retailing, the application of data mining techniques using the Extended RFM model can provide valuable information for businesses. By analyzing customer data from online retail platforms, businesses can identify different customer segments and understand their purchasing behavior. This information can then be used to tailor marketing strategies and personalize customer experiences, ultimately leading to increased customer satisfaction and loyalty.

One specific area of focus when applying the Extended RFM model to online retailing is the BCG (Boston Consulting Group) matrix. The BCG matrix is a strategic tool that helps businesses analyze their product portfolio and make informed decisions about resource allocation. It categorizes products into four quadrants based on their market growth rate and relative market share: stars, cash cows, question marks, and dogs. By integrating the BCG matrix into the Extended RFM model, businesses can gain a deeper understanding of customer purchase behavior about their product portfolio. This allows them to identify which customer segments are more likely to purchase products from different quadrants of the BCG matrix. By leveraging this information, businesses can develop targeted marketing campaigns for each customer segment, focusing on promoting specific products or categories that align with their purchasing preferences. Additionally, the integration of the BCG matrix with the Extended RFM model can also help businesses identify potential opportunities for product development or improvement based on customer feedback and preferences.

## 5. Conclusion

This study has effectively segmented the firm's customer base into four distinct clusters—Platinum, Gold, Silver, and Bronze—each with unique characteristics and implications for business strategy. The Platinum cluster represents highly valuable customers who are loyal and contribute significantly to market growth; thus, the firm is advised to retain these customers through targeted incentives and by ensuring product quality, timely delivery, and enhanced services. The Gold cluster, previously underestimated, consists of retailers with high purchase frequency and good market growth potential. This group should be engaged through strategic advertising and promotional efforts to solidify their status as regular customers.

The Silver cluster, comprising new customers with high purchase shares but low contribution to liquidity growth, presents an opportunity for the firm to nurture these relationships. By increasing product awareness and satisfaction, these customers have the potential to become either stars or cash cows for the firm. Finally, the Bronze cluster appears to be a less beneficial segment, with a history of purchases but no recent engagement. This indicates a possible mismatch between the firm's offerings and the group's needs, suggesting a need for further research into this disengagement.

In conclusion, the application of data mining techniques, specifically the Extended RFM model with a focus on the BCG matrix, can provide valuable insights into customer purchase behavior in the context of online retailing. By analyzing customer data and segmenting customers based on their purchasing patterns, businesses can tailor their marketing strategies and improve customer satisfaction and loyalty. The integration of the BCG matrix further enhances this analysis by providing a strategic perspective on product portfolio management and resource allocation. the study provides a clear roadmap for the firm to prioritize customer segments and tailor strategies to maintain and grow its customer base effectively. By focusing on the value provided to each cluster and addressing their specific needs, the firm can enhance customer loyalty, market share, and overall profitability. Ultimately, these insights can help businesses make informed decisions and drive growth in the competitive online retail industry.

**Author contributions:** Conceptualization, SG, SH and SS; methodology, SG and SS, PB; software, SS and SG; validation, HGT, SS, MS and PB; formal analysis, HGT; investigation, SG; resources, NA; data curation, PB; writing—original draft preparation, SS; writing—review and editing, PB; visualization, HGT; supervision, SG; project administration, PB; funding acquisition, SS. All authors have read and agreed to the published version of the manuscript.

**Conflict of interest:** The authors declare no conflict of interest.

## References

Alghanam, O. A., Al-Khatib, S. N., & Hiari, M. O. (2022). Data Mining Model for Predicting Customer Purchase Behavior in E-Commerce Context. International Journal of Advanced Computer Science and Applications, 13(2). https://doi.org/10.14569/ijacsa.2022.0130249

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), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011

Asmat, F., Suryadi, K., & Govindaraju, R. (2023). Data mining framework for the identification of profitable customer based on recency, frequency, monetary (RFM). In: AIP Conference Proceedings. https://doi.org/10.1063/5.0130290

Audzeyeva, A., Summers, B., & Schenk-Hoppé, K. R. (2012). Forecasting customer behaviour in a multi-service financial organisation: A profitability perspective. International Journal of Forecasting, 28(2), 507–518. https://doi.org/10.1016/j.ijforecast.2011.05.005

Bali, S., Bali, V., Gaur, D., et al. (2023). A framework to assess the smartphone buying behaviour using DEMATEL method in the Indian context. Ain Shams Engineering Journal, 102129. https://doi.org/10.1016/j.asej.2023.102129

Bellou, V., & Andronikidis, A. (2008). The impact of internal service quality on customer service behaviour. International Journal of Quality & Reliability Management, 25(9), 943–954. https://doi.org/10.1108/02656710810908098

Ben Ncir, C. E., Ben Mzoughia, M., Qaffas, A., et al. (2023). Evolutionary multi-objective customer segmentation approach based on descriptive and predictive behaviour of customers: application to the banking sector. Journal of Experimental & Theoretical Artificial Intelligence, 35(8), 1201–1223. https://doi.org/10.1080/0952813x.2022.2078886

Chayjan, M. R., Bagheri, T., Kianian, A., et al. (2020). Using data mining for prediction of retail banking customer's churn behaviour. International Journal of Electronic Banking, 2(4), 303. https://doi.org/10.1504/ijebank.2020.114770

Chen, Y., Liu, L., Zheng, D., et al. (2023). Estimating travellers' value when purchasing auxiliary services in the airline industry based on the RFM model. Journal of Retailing and Consumer Services, 74, 103433. https://doi.org/10.1016/j.jretconser.2023.103433

Corbos, R. A., Bunea, O. I., & Breazu, A. (2023). Influence of online consumer reviews on the sales of large household appliances: a survey in Romania. Electronic Commerce Research. https://doi.org/10.1007/s10660-023-09758-6

Corbos, R. A., Bunea, O.-I., & Triculescu, M. (2023). Towards Sustainable Consumption: Consumer Behavior and Market Segmentation in the Second-Hand Clothing Industry. Amfiteatru Economic, 25(Special 17), 1064. https://doi.org/10.24818/ea/2023/s17/1064

Cui, J. (2023). Deep Mining Algorithm of Online Purchase Behavior Data Based on Decision Tree Model. Journal of Testing and Evaluation, 51(3), 1398–1407. https://doi.org/10.1520/jte20220094

Gholamiangonabadi, D., Shahrabi, J., Hosseinioun, S. M., et al. (2019). Customer Churn Prediction Using a New Criterion and Data Mining; A Case Study of Iranian Banking Industry. In: Proceedings of the International Conference on Industrial Engineering and Operations Management.

Gholamveisy, S. (2021a). Discovering hidden cluster structures in citizen complaint call via Som and association rule technique. Journal of Mechanics of Continua and Mathematical Sciences, 16, 79–92. https://doi.org/10.26782/jmcms.2021.07.00007

Gholamveisy, S. (2021b). Gasoline consumption prediction via data mining technique. Journal of Mechanics of Continua and Mathematical Sciences, 16, 74–84. https://doi.org/10.26782/jmcms.2021.09.00007

Gholamveisy, S., Momen, A., Hatami, M., et al. (2023). The effect of perceived social media marketing activities on brand loyalty. Apuntes Universitarios, 13(3), 105–118. https://doi.org/10.17162/au.v13i3.1374

Gull, M., & Pervaiz, A. (2018). Customer Behavior Analysis Towards Online Shopping using Data Mining. 2018 5th International Multi-Topic ICT Conference (IMTIC). https://doi.org/10.1109/imtic.2018.8467262

Hambrick, D. C., MacMillan, I. C., & Day, D. L. (1982). Strategic Attributes and Performance in the BCG Matrix—A PIMS-Based Analysis of Industrial Product Businesses. Academy of Management Journal, 25(3), 510–531. https://doi.org/10.2307/256077

Kazmi, S. H. A., Ahmed, R. R., Soomro, K. A., et al. (2021). Role of Augmented Reality in Changing Consumer Behavior and Decision Making: Case of Pakistan. Sustainability, 13(24), 14064. https://doi.org/10.3390/su132414064

Khade, A. A. (2016). Performing Customer Behavior Analysis using Big Data Analytics. Procedia Computer Science, 79, 986–992. https://doi.org/10.1016/j.procs.2016.03.125

Khalili-Damghani, K., Abdi, F., & Abolmakarem, S. (2018). Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries. Applied Soft Computing, 73, 816–828. https://doi.org/10.1016/j.asoc.2018.09.001

Larsen, N. M., Sigurdsson, V., & Breivik, J. (2017). The Use of Observational Technology to Study In-Store Behavior: Consumer Choice, Video Surveillance, and Retail Analytics. The Behavior Analyst, 40(2), 343–371. https://doi.org/10.1007/s40614-017-0121-x

Luna-Romera, J. M., del Mar Martinez-Ballesteros, M., Garcia-Gutierrez, J., et al. (2016). An approach to silhouette and dunn clustering indices applied to big data in spark. In: Proceedings of the Advances in Artificial Intelligence: 17th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2016; 14–16 September 2016; Salamanca, Spain.

Martínez, R. G., Carrasco, R. A., Sanchez-Figueroa, C., et al. (2021). An RFM Model Customizable to Product Catalogues and Marketing Criteria Using Fuzzy Linguistic Models: Case Study of a Retail Business. Mathematics, 9(16), 1836. https://doi.org/10.3390/math9161836

Ncir, C. E. B., Hamza, A., & Bouaguel, W. (2021). Parallel and scalable Dunn Index for the validation of big data clusters. Parallel Computing, 102, 102751. https://doi.org/10.1016/j.parco.2021.102751

Sharma, A., Pratap, A., Vyas, K., et al. (2022). Machine Learning Approach: Consumer Buying Behavior Analysis. 2022 IEEE

Pune Section International Conference (PuneCon). https://doi.org/10.1109/punecon55413.2022.10014928
Solomon, M., Russell-Bennett, R., & Previte, J. (2012). Consumer behaviour: Pearson Higher Education AU. In: Australia.
Taşabat, S. E., Özçay, T., Sertbaş, S., & Akca, E. (2023). A new RFM model approach: RFMS. In: Industry 4.0 and the Digital Transformation of International Business. Springer; pp. 143–172.
Vasiljeva, T., Kreituss, G., & Kreituss, I. (2021). The implications of customer behaviour for banking service management: evidence from Latvia. Reliability and Statistics in Transportation and Communication. In: Proceedings of the 20th International Conference on Reliability and Statistics in Transportation and Communication, RelStat2020; 14–17 October 2020; Riga, Latvia.
Zhang, Y. (2023). Cluster analysis of perceptual demands of users' internet consumption behaviours based on improved RFM model. International Journal of Web Based Communities, 19(1), 15. https://doi.org/10.1504/ijwbc.2023.128408