Customer Segmentation Based On Recency Frequency Monetary Model: A Case Study in E-Retailing[1]

Araştırma Makalesi/Research Article

İnanç KABASAKAL

İşletme Bölümü, Ege Üniversitesi, İzmir, Türkiye
inanc.kabasakal@ege.edu.tr
(Geliş/Received:28.05.2019; Kabul/Accepted:20.01.2020)
DOI: 10.17671/gazibtd.570866

***Abstract*—** Marketing studies have often drawn attention to the importance of customers for businesses that aim to endure in a harsh competitive environment. Customer Relationship Management (CRM) has been a prominent approach in marketing management that aims to improve relationships with customers. A practical implication of the CRM approach is the analysis of customer data to extract value for businesses, as well as customers. In this context, customer segmentation has been a useful task that helps to group customers with similar attributes and designate better-tailored marketing strategies for customer groups. Among a variety of approaches for customer segmentation, Recency Frequency Monetary (RFM) Model stands out as an easy-to-adopt and effective technique. Based on three dimensions regarding the sales data, the RFM Model depends on scoring customers with different approaches. In this study, a prototype software is introduced that helps to apply the RFM technique with two scoring approaches. Moreover, the sales data obtained from an e-retailer has been analyzed for clustering using the prototype software, and clusters discovered with RFM variants were compared using cluster evaluation metrics. Finally, the segments were presented along with relevant offers for marketing strategies.

***Keywords*—** Customer Segmentation, RFM Model, Analytical CRM

Güncellik Sıklık Parasallık Modeline Dayalı Müşteri Bölümlendirme: E-Perakende Sektöründe Bir Uygulama

***Özet*—** Pazarlama alanındaki çalışmalarda, yoğun rekabet ile başa çıkmaya çalışan işletmeler açısından müşterilerin önemine sıklıkla dikkat çekildiği görülmektedir. Pazarlama yönetimi bağlamında öne çıkan bir yaklaşım olan Müşteri İlişkileri Yönetimi (MİY), işletmeler ile müşterileri arasında kurulan ilişkilerin geliştirilmesini amaçlamaktadır. Müşteri verisinin işletme ve müşterileri için değer yaratmak üzere analiz edilmesi, MİY'in uygulamada gereksinimlerinden birisi olarak ifade edilebilir. Bu bağlamda müşteri bölümlendirme, benzer niteliklere sahip müşteri gruplarının ortaya çıkarılarak grup odaklı pazarlama stratejilerinin uyarlanması için yararlı bir işlev görmektedir. Müşteri bölümlendirme için ortaya konulmuş çeşitli yaklaşımlar arasında RFM Modeli, etkin ve kolay uyarlanabilir olmasıyla öne çıkmaktadır. Müşterilerin satış verisine ilişkin 3 farklı boyut üzerinden sıralanmasına dayanan yöntem, sıralamada kullanılan puanlama biçimine göre çeşitli yaklaşımlara konu olmaktadır. Bu çalışmada, RFM yöntemini iki farklı puanlama yaklaşımı ile yürütmek üzere geliştirilmiş bir prototip yazılım tanıtılmaktadır. Bir e-perakende işletmesinden alınan satış verisi sözü edilen yazılım ile incelenmiş, RFM modeline ilişkin her iki puanlama yöntemi ile bölümlendirme yapılmış, bulgular veri madenciliği bağlamında değerlendirme ölçütleri ile karşılaştırılmıştır. Son olarak ortaya çıkarılan müşteri bölümleri sunulmuş ve seçilen gruplara yönelik öneriler sıralanmıştır.

***Anahtar Kelimeler*—** Müşteri Bölümlendirme, RFM Modeli, Analitik CRM

[1] This paper is a substantially revised and extended version of the paper entitled "RFM Model for Segmentation in Retail Analytics: A Case Study" presented in the International Data Science & Engineering Symposium held in Karabük, Turkey between the dates May 2-3, 2019.

1. INTRODUCTION

The success or failure of businesses relies on a variety of factors. As the primary source of income for businesses, customers might be described as valuable, intangible assets for businesses that directly correspond to financial outcomes [1]. In this regard, businesses need to avoid losing customers and try to acquire new ones. Pfeifer [2] highlighted the well-known comparison of customer acquisition costs against customer retention costs and underlined that acquiring new customers is five times higher than maintaining relationships with existing customers. As emphasized in such findings, Customer Relations Management (CRM) has been embraced as an essential strategy for businesses.

Implementing CRM in businesses involves practices and methodology to make use of customer data. In such a data-oriented business environment, businesses try to exploit existing data sources and make data-driven decisions for competitive advantage [3]. The analysis of customer data helps businesses to figure out relationships between data elements, describe significant events, and provide predictions [4]. For a customer-driven strategy, marketers try to divide total markets and identify customer segments based on geographic, demographic, psychographic, and behavioral variables [5]. Accordingly, it could be noted that clustering models in data mining have often been borrowed to analyze customer behaviors in customer segmentation.

Recency Frequency Monetary (RFM) model is a popular technique for customer segmentation based on the analysis of purchase behaviors. The model analyzes purchase records and represents each transaction by three dimensions. In RFM, all customers are scored individually and ranked according to each dimension. As a result, customer groups with similar purchase patterns are identified. Each group is represented with a 3-digit RFM combination that summarizes a purchase pattern typical within members' transaction history.

In this study, the use of RFM will be presented along with a case study of an e-retailer from Turkey. From a data mining perspective, RFM for customer segmentation was interpreted as a specific clustering task that requires further assessment. With such regard, this study differentiates from most studies by utilizing an assessment over the RFM clusters through three similarity measures for clustering: cohesion, separation, and the Silhouette coefficient. Additionally, a prototype software is introduced to demonstrate RFM clusters and present RFM-based customer segments.

2. CUSTOMER RELATIONS MANAGEMENT

Customer Relations Management (CRM) is a philosophy in business management that involves customer-focused strategies to acquire and retain customers while improving customer value and customer loyalty [6]. CRM is often regarded as an essential approach in improving customer satisfaction, customer acquisition, customer retention, and profitability. In this regard, it could be argued that CRM strategy emphasizes the importance of customers for businesses and promotes customer-centric practices in marketing.

In prior research, CRM has been explored with its operational implications and analytical practices, just as its strategic aspect. Venturini and Benito [7] defined CRM as "an IT-enabled business strategy" that requires capturing, storing, and analysis of customer data. Farquad et al. [8] noted that Analytical CRM corresponds to analyses over customer data with data mining and machine learning methods to extract useful findings such as identification of customers with high importance or predictions on customer churn.

2.1. Analytical CRM

For businesses, attracting and developing profitable relations with their most valuable customers requires operations that require analytical tools and techniques. In various decisions such as recommending relevant products for customers, launching the appropriate campaigns for target segments, or bundling products; analysis of operational data is utilized frequently. In this regard, it could be emphasized that the implementation of CRM often relies on quantitative methods.

Cheng and Chen [9] mentioned that the success of CRM processes requires the effective utilization of IT tools. The use of algorithms and statistical methods over customer data to accomplish segmented or individualized marketing decisions is mostly related to Analytical CRM. Bahari and Elayidom [10] noted that analyzing customer data with data mining techniques is essential to understand customers, to implement a competitive CRM strategy, and to increase customer value. According to Liu [11], Analytical CRM for target marketing is essential to implement customer-focused strategies, and ideally requires focusing on customer data that include the entire customer journey. In this regard, customer data depends on various interactions established with customers. However, most interactions do not demonstrate customers' purchase intentions as precise as the prior purchase records.

The purchase history is an essential resource to support tactical decisions in marketing, and an important asset for Analytical CRM. For instance, as in the algorithms asserted by Agrawal and Srikant [12], basket data can be mined to discover frequently-purchased products, and generate recommendations accordingly. Additionally, sales transactions might help to forecast demand. Within the data mining context, clustering methods are helpful for Analytical CRM tasks such as customer segmentation and classification.

2.2. Customer Segmentation

One of the most crucial benefits of customer analytics is the ability to identify customer segments that group customers with similar attributes. Elliott, Scionti, and Page [13] noted that a customer-focused approach in marketing requires understanding customers, and identifying customer segments provides insights about the common intentions among customers.

An intuitive and widely used approach for customer segmentation makes use of demographic variables. A marketing manager might classify customers into segments such as seniors, children, or metropolitan citizens and target one or more of such groups in marketing plans. On the other hand, the segmentation of customers considering their purchase behaviors is a common practice for marketers rather than segmentation that merely rely on demographics.

Literature inquiry for customer segmentation in marketing research hosts various behavioral approaches that depend on the customers' actions, rather than their attributes. For instance, the approach proposed by Lin [14] analyzes the brands of products chosen by customers, and utilizes brand preferences for segmentation. Furthermore, in a study that explores segmentation approaches employed in Migros Türk, Cooil et al. [15] reported the following among the variables used for behavioral clustering:

- Location of purchase behavior
- Purchases of food versus non-food
- Method of payment
- The intensity of communication among customer and staff
- Time of purchase
- Channel used in the purchase (phone, website, kiosk, etc.)

Another typical criterion for behavioral segmentation can be noted as profitability. In terms of relationship marketing, the relationships with customers have a cost and provide revenue. Accordingly, a segmentation approach over the relationship cost & revenue results in four segments [16] that correspond to distinctive levels of profitability. Just as the customer profitability criterion, customer lifetime value is another measure that indicates the forecasted profits for a customer in the future. Xu and Walton [17] underlined that the profitability and lifetime value differ among customers, and such difference necessitates the segmentation of customers to figure out the most valuable customers.

Among a variety of customer segmentation techniques, RFM is an easy-to-use technique that reduces purchase behaviors into a limited set of variables, and identifies customer groups of members with similar purchase patterns.

3. RFM TECHNIQUE FOR CUSTOMER SEGMENTATION

Recency, Frequency, and Monetary (RFM) Technique is a widely recognized method to identify the characteristics of customers concerning purchase patterns. The technique analyzes previous basket transactions and identifies customer groups that involve members who resemble each other in terms of purchase decisions. In this manner, RFM might be classified as a behavioral segmentation model that might provide interesting information in marketing decisions.

In RFM, the purchase behaviors are represented as a combination of three dimensions that basically depend on the timing and sum of transactions. In particular, the purchase history is analyzed with an assessment over those dimensions, and each customer is rated accordingly. The three dimensions mentioned for RFM analysis are introduced in Table 1:

Table 1. Dimensions of the RFM Model

Dimension	**Description**
Recency	The duration since the date of last purchase
Frequency	The total count of purchases
Monetary	The average amount purchased

The first step in RFM analysis involves scoring and ranking customers according to three attributes separately. For each dimension, the highest 20% is ranked as 5 (best), where the lowest 20% is ranked as 1 (worst) over a scale of 1-5. As an exception, the high values in recency denote an undesirable condition. Accordingly, the lowest 20% is ranked as 5 and the highest 20% is ranked as 1 for the recency attribute.

The next step in the RFM Model involves defining segments based on RFM combinations. In this step, segments are formulated as the combinations of scores in each of the three attributes. For instance, a segment with code 515 involves the customers that purchase in large sums and have purchased recently; however, the total count of purchases by such customers is lower than usual. Accordingly, the most valuable customers are in the segment: 555, which represents the best combination in all RFM dimensions.

The advantage of utilizing an RFM Model for segmentation is the ease of use. Madeira [18] argued that the representation of purchase behavior patterns with fewer variables is a strength of the RFM model, contrary to the large variable count of variables needed in a demographic segmentation approach. On the other hand, Peker et al. [19] argued that taking the recency attribute as the only indicator for future transactions might be misleading; thus, the authors proposed another model that incorporates an alternative variable for periodicity.

The RFM Model provides several benefits for marketers. Wei, Lin, and Wu [20] remarked that the practical use of the model is customer segmentation with the purpose of identifying valuable customers and improving response rates in direct marketing campaigns. Also, Aggelis and Christodoulakis [21] noted that customer segmentation with the RFM technique identifies customers that are more likely to respond to offers, and helps to estimate the profitability of customers according to their segments. Besides, Wei et al. [22] argued that noted that marketing strategies for businesses might involve customization of products and services according to the RFM segments, and applied the RFM technique for customer segmentation in a veterinary hospital for such purpose.

Sarvari et al. [23] underlined the importance of targeting customer segments with relevant offers and argued that the use of RFM combinations along with demographic variables leads to better results in customer segmentation. Additionally, the authors demonstrated a rule-based analysis and concluded that rules with RFM variables have greater support and accuracy.

An additional use case for the RFM Model is the calculation of Customer Lifetime Value (CLV). CLV is a forward-looking measure [24] of profits to be obtained from a customer. The objective in lifetime value calculation can be described as the identification of most important customers for a business. In RFM, segments with relatively high scores for R, F, and M are considered highly important. In addition, Altan [25] used RFM-based scoring for identifying and targeting more valuable customers in the airline industry.

In prior research several studies have been found [26] [27] that have proposed CLV calculation models through scoring and prioritization of RFM dimensions. In such calculations, it was noted that CLV calculations depend on a summation after the weighting of RFM variables. Additionally, a more recent model by Srivastava [28] utilized the RFM model for clustering and compared the segment populations across years.

4. CASE STUDY

In this section, a custom software prototype for RFM analysis and customer segmentation is introduced. Basket data obtained from an e-retailer was analyzed with the RFM segmentation method, using the prototype software. The customer segmentation approach being utilized is presented along with the results.

4.1. Methodology

The primary objective of this study is to implement a behavioral segmentation strategy to identify customer groups with similar purchase patterns. For this purpose, an RFM-based technique is utilized for customer segmentation through the study. Customer segmentation corresponds to a clustering task in data mining where the objective is to identify groups of similar customers according to purchase behaviors. Therefore, our methodology involves the clustering of customers over the analysis of basket data.

In data mining literature, clustering techniques are used in tasks that aim to group similar data together. As noticed in [29], the objective of clustering is to create non-similar classes where interclass similarity is high. Accordingly, the RFM-based clusters identified through the study have also been evaluated using cluster evaluation metrics.

4.2. Cluster Evaluation

If a clustering problem involves a priori knowledge of the correct solution, the usual approach for evaluating a clustering task is to compare resulting clusters with the correct ones [31]. In contrast, the unavailability of correct clusters in such problems requires similarity measures. Clustering techniques utilize similarity measures such as the sum of squared errors, cohesion, and separation to measure the compactness of clusters [32].

For any data point, cohesion measure identifies the similarity to its cluster, and the separation measure demonstrates the dissimilarity to the data in other clusters. In clustering, cohesion and separation are both taken into consideration with the Silhouette Coefficient, which is another measure proposed by Kaufman and Rousseeuw [33]. This measure corresponds to a value within the range [-1, 1], and signifies how much an object belongs to a cluster by comparing cohesion and separation [29].

The Euclidean distance method is a common way to measure the dissimilarity between data with continuous variables [30]. Accordingly, the distance between two points of *i* and *j* can be calculated as follows:

$$d(i,j) = \sqrt{(r_i - r_j)^2 + (f_i - f_j)^2 + (m_i - m_j)^2} \qquad (1)$$

where r_i, f_i, and m_i correspond to normalized values for each one of the RFM dimensions. For a data point *i*, let $cohesion(i)$ represent the average dissimilarity of node *i* from the nodes within the same cluster, c_i with k members. Accordingly, cohesion measure in our problem might be formulated as the following:

$$cohesion(i) = \frac{1}{k-1} \sum_{j_n \in c_i \, where \, j_n \neq i} d(i, j_n) \qquad (2)$$

The formula can be modified for measuring the average distance from data points on other clusters to obtain the separation. The separation for a point *i* in the cluster c_i corresponds to its average distance to points in all other clusters. Such points can be represented as $\acute{c}_i = \mho - c_i$ where $\mho$ corresponds to the whole dataset with a total of *u* data points. Accordingly, separation for the point *i* can be calculated as:

$$separation(i) = \frac{1}{u-k} \sum_{j_n \in \acute{c}_i} d(i, j_n) \qquad (3)$$

Using both measures, Silhouette Coefficient for a clustering solution can be calculated as follows [34]:

$$SC = \frac{1}{u} \sum_{i \in U} \frac{separation(i) - cohesion(i)}{\max\{separation(i), cohesion(i)\}} \quad (4)$$

As noted in [33], a higher score in Silhouette Coefficient indicates the superiority of a clustering solution that has comparably high similarity within clusters and low similarity across clusters.

4.3. RFM Segmentation Tool

As mentioned in the introduction, the study introduces a prototype software tool in which the RFM technique was implemented. The prototype software was developed by the author in C# using the ASP.NET framework and Microsoft Visual Studio 2017. The prototype software takes a data source as the input for basket data and performs RFM analysis. As a result, the software presents the customer segments that correspond to RFM combinations.

The RFM implementation for the study currently supports two types of data sources: CSV files and SQLite data files. In CSV files, each row should define a purchase record, and comma-separated columns should list purchase details in a specific order. Accordingly, each record should be represented in a row, including the following details with the specific order: basket identifier, customer identifier, date of the purchase, and total money spent. Additionally, our prototype supports SQLite databases. SQLite is a 'C-language library that implements a fast, reliable, and fully-featured SQL database engine' [35]. The SQLite website claims that SQLite is still the most widely deployed database engine globally in 2019, due to its use in mobile devices, web browsers, and various operating systems by default [36]. Furthermore, the prototype is intended to support NoSQL databases that offer higher scalability and performance than relational databases for larger datasets [37].

The prototype software has a native User Interface (UI) for Microsoft Windows. Within the software, each RFM analysis is declared as an individual project with a particular dataset. Accordingly, when a project is created, basket records in the dataset are chronologically grouped and listed as in Figure 1.

According to the UI in Figure 1, selecting a day from the list induces a date filter, and the transaction list located on the right side of the window lists basket details for transactions on the selected date. Moreover, the UI displays cumulative statistics such as the count or the total of monthly transactions and average payments.

When the 'Customer Segments' button is clicked, another UI for customer segmentation is launched as in Figure 2. The 3-digit combinations listed at the left-side indicate RFM clusters and total count of customers within each cluster. Moreover, the RFM clusters are grouped into RFM-based segments.

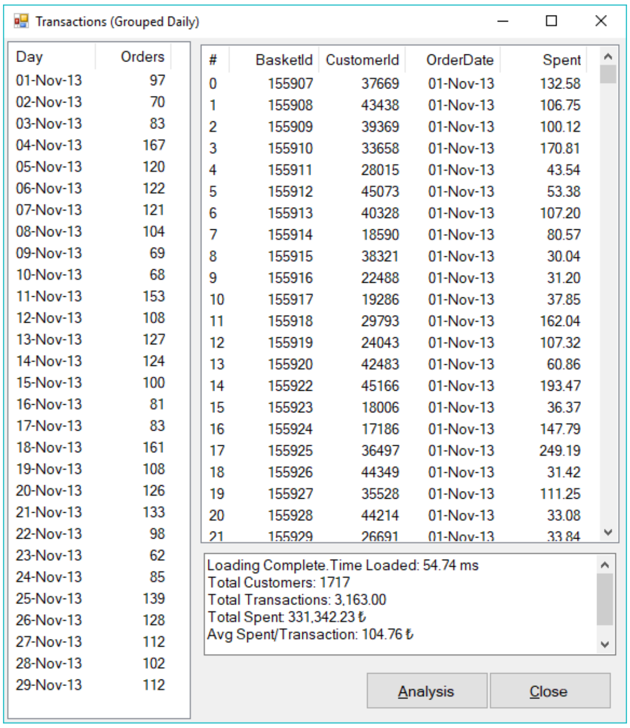


Figure 1. The UI for Transactions in Prototype Software

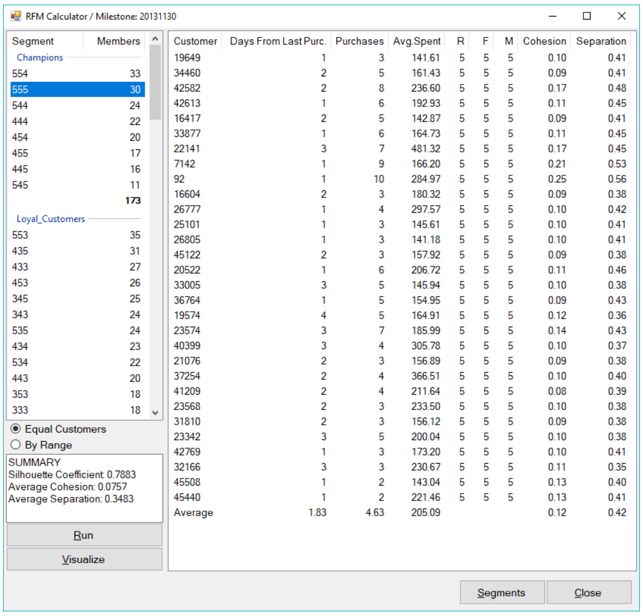


Figure 2. RFM Segments in Prototype Software UI

The user interface provides an option that help to customize how the scores in RFM combinations are to be obtained. The implementation of RFM-based segmentation with those options will be detailed further in Sections 4.5 and 4.6.

4.4. Data Preprocessing

In this study, customer segmentation was performed over the basket data obtained from an online retailer. The e-retailer (http://www.adepo.com) had been founded in 2001 and had offered groceries, beverages, cosmetics, and housekeeping products until the termination of its operations in December 2015. Specifically, the dataset involved records of 3163 transactions that were ordered from and delivered in İzmir, Turkey. All orders had taken place in November 2013.

The basket data involves the following attributes:

- Basket Identifier (Integer)
- Customer Identifier (Integer)
- Order Date (Date)
- Invoice in Turkish Liras (Decimal)

The basket data obtained in a spreadsheet has subsequently been moved into an SQLite database for analysis. Such migration corresponds to a basic ETL (Extract, Transform, and Load) process that involves extraction and transformation of data migrating data [38] of different formats such XML files, flat files, and data streams by the use of data cleansing, conversion and aggregation steps [39]. Accordingly, the SQLite database for the study was prepared after cleansing and conversion of basket data in the spreadsheet.

4.5. Customer Segmentation based on RFM Model

Customer segmentation with RFM Model requires ranking customers in RFM attributes. By definition, a cluster or a segment is set to contain elements that are similar to each other, and not similar to other elements out of their segment. In this regard, segments defined with the RFM model are expected to involve members that have common purchase patterns.

For segmentation based on RFM Model, measures for RFM dimensions should be calculated at first. In the prototype software, the measures of recency, frequency, and monetary dimensions are calculated as described below:

- The recency is measured by counting the days after the last purchase by a customer. Inactive customers that have purchased long ago would have a high number of days from their last purchase, and the recency dimension of RFM corresponds to a low score.
- The frequency measure equals to the count of purchase activities by a customer. A high number in purchase count often refers to high profitability. A high frequency calculated for a customer segment thus indicates a high score.
- The monetary measure is the average of payments when purchasing. A high value usually signifies an active customer, and corresponds to a high score.

Upon the calculation of RFM dimensions, the customers and scored individually and ranked. RFM-based scoring is detailed in the next subsection.

4.6. Scoring Customers in RFM Model

Scoring a customer for an RFM dimension requires a cumulative comparison over the whole customer list. In particular, customers are sorted from the highest to the lowest for each RFM attribute individually. As a result of these steps, a sorted list of customers will be obtained for each of the three RFM dimensions. Subsequently, the lists are split into five groups. As the last step, each group is given a distinct score with an integer value in [1, 5]. Repeating this procedure for all RFM dimensions leads to a maximum of 5^3 segments formulated with 3-digit codes such as 555 and 543.

In the method detailed above, multiple techniques might be applied for splitting customers based on RFM measures. In particular, there are a variety of techniques for scoring RFM dimensions, and each one has its advantages and disadvantages [40] that require caution when choosing one. Creating groups of segments with equal customer counts [41] is a standard method in RFM segmentation. Apart from the values calculated for RFM dimensions, the rank for each customer determines the RFM segment in this approach. An alternative approach is finding the minimum and the maximum value for a specific RFM dimension, and equally splitting the distance between min and max values into five ranges of equal length. In this approach, the values determined for RFM dimensions of a customer generates the RFM segment. The software prototype was implemented to split customers with both approaches. As an alternative approach, the RFM model with K-Means clustering proposed by Anitha and Patil [42] might be adopted where each dimension data is clustered with the K-Means technique separately. However, the repetition of K-Means clustering for each dimension consecutively might also undermine the strengths of RFM technique such as simplicity and ease-of-use.

A problem in the first approach is the possibility of the inconsistent ranking of customers virtually with identical scores. For instance, two customers with total spending of 100.00$ against 100.25$ might be scored as M=2 and M=3. The reason is the determination of score merely with respect to the order of the customer on the list. In such a scenario, it might be argued that both customers have spent nearly equal money, and they should be placed in the same segment. Such insight might lead to the second approach mentioned for segmentation. However, the second approach also has limitations. A possible could be described as the probable shortage of members in the customer segments at the extremes (1 and 5). For instance, the density of samples is higher at the values close to the mean value in a normal distribution. In this example, a range that surrounds the average values would lead to large groups with excessive members.

5. FINDINGS

In the study, 3163 purchase records by a total of 1717 customers were analyzed with the RFM technique. The customer segments were identified using the prototype software developed for the study. The calculation of recency measure depends on the total count of days after the last purchase. In this regard, the date of analysis was taken as the last day in our available dataset, 30^{th} of November in 2013.

The first step in RFM segmentation adopted in study is to analyze purchase history to conduct clustering based on RFM dimensions. Among two distinct alternatives covered in the previous section, the first approach was chosen that led to equally split clusters of customers. As a result, the total count of distinct RFM combinations was found as 81.

Table 2 lists the RFM combinations along with the corresponding customer counts:

Table 2. RFM combinations

RFM Score	Group Size	RFM Score	Group Size	RFM Score	Group Size
111	69	343	25	352	12
115	67	345	24	542	12
112	67	435	24	241	11
114	61	325	23	541	11
225	59	551	22	354	10
224	56	533	21	243	10
113	52	443	20	355	9
221	49	321	20	545	9
223	39	433	20	332	9
222	38	341	19	245	8
453	34	544	19	351	7
452	34	353	18	212	7
532	34	455	18	142	6
534	32	322	18	253	6
535	31	333	18	143	6
555	30	445	17	141	6
552	29	244	16	211	6
432	28	344	16	144	5
431	28	342	16	213	5
554	27	335	16	214	5
553	27	334	16	215	4
323	27	242	15	251	3
444	26	442	15	252	3
454	26	543	15	255	3
331	26	441	14	145	2
531	26	324	14	151	1
451	25	434	14	155	1

When the second option was picked, only 20 RFM combinations were identified. Moreover, the top four of the most populous segments involved 1236 customers out of 1717 (~72%). In this sense, the results obtained through the second approach had been found limited.

Apart from a mere comparison of cluster counts, both clustering approaches have also been assessed using cluster evaluation metrics. The comparison of average cohesion, average separation, and Silhouette Coefficient metrics have been demonstrated in Table 3. As the results suggest, clusters created by equally splitting customers have lower (better) inner-class similarity. Despite the higher separation achieved by the second approach, higher Silhouette Coefficient calculated for the first approach indicated that splitting clusters with equal numbers was the better option in our case.

Table 3. Cluster evaluation of alternative approaches

	Equal Customers	Equal Range
Silhouette Coefficient	**0.7919**	0.7496
Average Cohesion	**0.0747**	0.0929
Average Separation	0.3483	**0.3927**

To use the findings in marketing decisions, tailoring descriptive names for RFM combinations might be beneficial. Tsiptsis and Chorianopoulos [41] stressed that giving proper names for segments helps to describe the matching customer profiles, and facilitates the conversation for the practitioners. From this perspective, it could be noted that segments with proper descriptors help multiple decision-makers to stay on the same page.

In a study that integrates RFM analysis into multiple data mining tasks, Birant [43] identified customer segments with clustering techniques and named those segments based on their average RFM scores in comparison to other segments. Accordingly, the author introduced segments with RFM patterns; namely, best customers (R↑F↑M↑), valuable customers (R↑F↑M↑), shoppers (R↑F↑M↓), first-timers (R↑F↓M↓), churned (R↓F↑M↑), frequent visitors (R↓F↑M↓), spenders (R↓F↓M↑) and uncertain (R↓F↓M↓).

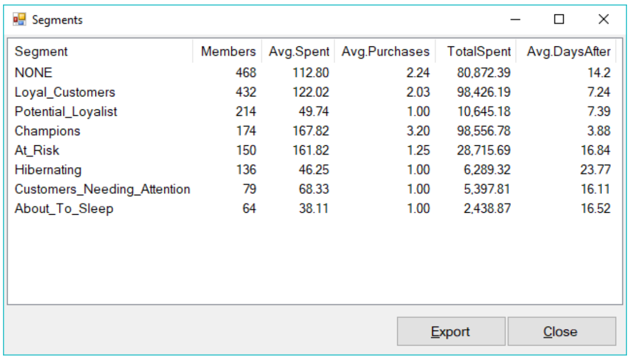
Putler, an online vendor for analytical services, has recommended 11 segments [44] each of which correspond to multiple R, F, M combinations as listed below:

Table 4. Customer segments labeled by RFM dimensions

Customer Segment	R	F and M
Champions	4-5	4-5
Loyal Customers	2-5	3-5
Potential Loyalist	3-5	1-3
Recent Customers	4-5	0-1
Promising	3-4	0-1
Customers Needing Attention	2-3	2-3
About To Sleep	2-3	0-2
At Risk	0-2	2-5
Can't Lose Them	0-1	4-5
Hibernating	1-2	1-2
Lost	0-2	0-2

In a case study to identify the RFM segments, Sutresno et al. [45] have adopted the segment labels proposed by [44]. With similar consideration, the RFM combinations in Table 2 have been merged to obtain the popular customer segments listed in Table 4.

In the prototype software developed for the study, the RFM combinations are grouped under corresponding segments. The prototype software developed for the study involves another user interface that opens by clicking the 'Segment' control in Figure 2. This user interface lists the customer segments and the count of members as shown in Figure 3.



Segment	Members	Avg.Spent	Avg.Purchases	TotalSpent	Avg.DaysAfter
NONE	468	112.80	2.24	80,872.39	14.2
Loyal_Customers	432	122.02	2.03	98,426.19	7.24
Potential_Loyalist	214	49.74	1.00	10,645.18	7.39
Champions	174	167.82	3.20	98,556.78	3.88
At_Risk	150	161.82	1.25	28,715.69	16.84
Hibernating	136	46.25	1.00	6,289.32	23.77
Customers_Needing_Attention	79	68.33	1.00	5,397.81	16.11
About_To_Sleep	64	38.11	1.00	2,438.87	16.52

Figure 3. The User Interface for the RFM Segments in the Prototype Software

The results indicate that 1249 of 1717 customers have been assigned into a segment. The list involves 7 of the 11 segments proposed in [44] and [45]. Furthermore, the following four segments have been expected but unmatched according to the results, possibly due to the limited range of purchase dates in our data.

- Recent Customers
- Promising
- Can't Lose Them
- Lost

As noted by Han and Kamber [29], the data that might not fall into one of the existing clusters are often labeled as outliers. Accordingly, the remaining customers might be noted as the outliers. In Figure 3, the list also reports such customers under a specific group label: 'None'. It should be noted that such customers actually are involved in an RFM cluster that does not match a particular customer segment described in [44] and [45].

From a marketing manager's perspective, discovering customer segments might be the initial and essential step towards further decisions. Each labeled segment might be targeted individually with a specific set of marketing actions. As an example, Table 5 describes several customer segments along with recommended actions:

Table 5. Targeting RFM segments for marketing decisions

Segment	Clusters	Description / Actions
Champions	554: 32 555: 31 544: 26 444: 23 454: 19 445: 17 455: 16 545: 10 Total:174	• Most valuable and active customers • Offer new products, send discount coupons to keep on sales
Customers needing attention	223: 42 144: 37 Total:79	• Valuable customers that have not purchased recently • Try promotions to trigger membership renewals • Send mail messages, coupons
Potential Loyalist	432: 41 431: 36 331: 26 323: 25 532: 22 321: 19 322: 18 531: 18 332: 9 Total:214	• Active customers with less frequent purchases with smaller totals than loyal customers • Offer membership benefits to convert such customers into loyal ones

It should be noticed that the decisions for segments were just listed in accordance with common sense. In fact, the formulation of strategies for segments requires managerial effort and expertise by practitioners. Moreover, prioritization of segments and actions might differ occasionally based on the organizational resources, strategies, and constraints.

6. CONCLUSION

CRM has been a popular strategy that transforms businesses with a customer-centric perspective. Rather than mass marketing, such perspective emphasizes targeting customer segments or customers as individuals. In this regard, identification of customer segments is an essential task to implement CRM.

RFM Analysis is a useful and practical technique that is used for behavioral customer segmentation. The technique relies on three dimensions that represent different aspects of purchase behaviors. The technique and various approaches in the scoring of RFM dimensions were discussed in detail in the study. More importantly, cluster evaluation metrics were utilized to compare RFM-based clusters. Thus, our methodology helps to evaluate different RFM variations with individual scoring and splitting strategies and highlights the better option.

Finally, this paper presents a prototype software that conducts RFM analysis with two variants and provides cluster evaluation metrics to choose the more convenient one for a particular dataset. The use of our software was explored through a case study. Purchase data obtained from an e-retailer was analyzed for customer segmentation. Despite the limited size of transactions obtained, the analysis over the dataset was sufficient to demonstrate a sample RFM-based customer segmentation. The resulting RFM segments were presented along with member statistics in the results. In addition, the study provides a demonstration to identify conventional customer segments such as champions and loyalists, through the use of combined RFM clusters.

For businesses, the use of RFM or other segmentation methods has often been promoted by researchers. Such analyses are often accomplished through CRM software. However, businesses with the purpose of implementing CRM solutions often come up against obstacles [46], such as the lack of resources, experience, uncertainty, and technological infrastructure. Furthermore, despite the potential benefits of e-CRM solutions, Harrigan et al. [47] underlined several drawbacks of dedicated CRM software for small businesses and advised the use of more straightforward solutions to maintain relations with customers and perform necessary analyses. From this point of view, designing and using compact software in CRM might contribute to businesses as a strategic tool for competition. On the other hand, assessment of long-term outcomes of segment-based decisions might require a more detailed focus in further studies.

REFERENCES

[1] S. Gupta, D. R. Lehmann, “Customers as assets”. *Journal of Interactive Marketing*, 17(1), 9–24, 2003.

[2] P. E. Pfeifer, “The optimal ratio of acquisition and retention costs”, *Journal of Targeting, Measurement and Analysis for Marketing*, 13(2), 179–188, 2005.

[3] F. Provost, T. Fawcett, “Data science and its relationship to big data and data-driven decision making”, *Big Data*, 1(1), 51–59, 2013.

[4] G. Phillips-Wren, A. Hoskisson, “An analytical journey towards big data”, *Journal of Decision Systems*, 24(1), 87-102, 2015.

[5] P. Kotler, G. Armstrong, **Principles of Marketing**, 14th Edition; Pearson Education, USA, 2012.

[6] E. Chablo, “The importance of marketing data intelligence in delivering successful CRM”, **Customer Relationship Management**, Editor: SCN Education B.V., Vieweg+ Teubner Verlag, Wiesbaden, 57–70, 2000.

[7] W. T. Venturini, Ó. G. Benito, “CRM software success: a proposed performance measurement scale”, *Journal of Knowledge Management*, 19(4), 856-875, 2015.

[8] M. A. H. Farquad, V. Ravi, S. B. Raju, “Churn prediction using comprehensible support vector machine: An analytical CRM application”, *Applied Soft Computing*, 19, 31-40, 2014.

[9] C. H. Cheng, Y. S. Chen, “Classifying the segmentation of customer value via RFM model and RS theory”, *Expert Systems with Applications*, 36(3), 4176–4184, 2009.

[10] T. F. Bahari, M. S. Elayidom, “An efficient CRM-data mining framework for the prediction of customer behavior”, *Procedia Computer Science*, 46, 725-731, 2015.

[11] C. H. Liu, “A Conceptual Framework of Analytical CRM in Big Data Age”, *International Journal of Advanced Computer Science and Applications*, 6(6), 194-152, 2015.

[12] R. Agrawal, R. Srikant, “Fast algorithms for mining association rules”, **In Proceedings of the 20th International Conference of Very Large Data Bases**, Santiago, Chile, 487–499, September 12-15, 1994.

[13] Internet: K. Elliott, R. Scionti, M. Page, The confluence of data mining and market research for smarter CRM, https://www. researchgate.net/publication/228851712, 20.04.2019.

[14] C. F. Lin, “Segmenting customer brand preference: demographic or psychographic”, *Journal of Product & Brand Management*, 11(4), 249–268, 2002.

[15] B. Cooil, L. Aksoy, T. L. Keiningham, “Approaches to customer segmentation”, *Journal of Relationship Marketing*, 6(3-4), 9–39, 2008.

[16] K. Storbacka, “Segmentation based on customer profitability - retrospective analysis of retail bank customer bases”, *Journal of Marketing Management*, 13(5), 479–492, 1997.

[17] M. Xu, J. Walton, “Gaining customer knowledge through analytical CRM”, *Industrial Management & Data Systems*, 105(7), 955–971, 2005.

[18] S. A. C. Madeira, **Comparison of target selection methods in direct marketing**, Master’s Thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico, 2002.

[19] S. Peker, A. Kocyigit, P. E. Eren, “LRFMP model for customer segmentation in the grocery retail industry: A case study”, *Marketing Intelligence & Planning*, 35(4), 544–559, 2017.

[20] J. T. Wei, S. Y. Lin, H. H. Wu, “A review of the application of RFM model”, *African Journal of Business Management*, 4(19), 4199–4206, 2010.

[21] V. Aggelis, D. Christodoulakis, “Customer Clustering Using RFM Analysis”, **In Proceedings of the 9th World Scientific and Engineering Academy and Society International Conference on Computers**, Athens, Greece, 2–7, July 14-16, 2005.

[22] J. T. Wei, S. Y. Lin, Y. Z. Yang, H. H. Wu, “The application of data mining and RFM model in market segmentation of a veterinary hospital”, *Journal of Statistics and Management Systems*, 22(6), 1049-1065, 2019.

[23] P. A. Sarvari, A. Ustundag, H. Takci, “Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis”, *Kybernetes*, 45(7), 1129-1157, 2016.

[24] V. Kumar, “CLV: the databased approach”, *Journal of Relationship Marketing*, 5(2-3), 7–35, 2006.

[25] D. Altan, **Güncellik/Sıklık/Parasallık (RFM) Analizi ile Hedef Kitle Seçimi: Hava Yolu Sektöründe Bir Uygulama**, Master’s Thesis, Hacettepe University, Graduate School of Social Sciences, 2019.

[26] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh, “Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study”, *Procedia Computer Science*, 3, 57–63, 2011.

[27] B. Sohrabi, A. Khanlari, “Customer lifetime value (CLV) measurement based on RFM Model”, *Iranian Accounting & Auditing Review*, 14(47), 7–20, 2007.

[28] R. Srivastava, “Identification of Customer Clusters using RFM Model: A Case of Diverse Purchaser Classification”, *International Journal of Business Analytics and Intelligence*, 4(2), 45–50, 2016.

[29] J. Han, M. Kamber, **Data Mining Concepts and Techniques**, Morgan Kaufmann Publishers, San Francisco, USA, 2006.

[30] Z. Ceylan, S. Gürsev, S. Bulkan, “İki Aşamalı Kümeleme Analizi ile Bireysel Emeklilik Sektöründe Müşteri Profilinin Değerlendirilmesi”, *Bilişim Teknolojileri Dergisi*, 10(4), 475–485, 2017.

[31] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Pérez, I. Perona, “An extensive comparative study of cluster validity indices”, *Pattern Recognition*, 46(1), 243-256, 2013.

[32] O. Maimon, L. Rokach, **The Data Mining and Knowledge Discovery Handbook**, Springer, USA, 2005.

[33] Leonard Kaufman, Peter J. Rousseeuw, **Finding Groups in Data: An Introduction to Cluster Analysis**, Wiley, USA, 1990.

[34] Z. Halim, J.H. Khattak, "Density-based clustering of big probabilistic graphs", *Evolving Systems*, 10(3), 333-350, 2019.

[35] Internet: SQLite – What is SQLite?, https://www.sqlite.org, 25.04.2019.

[36] Internet: SQLite – Most widely deployed and used database engine", https://sqlite.org/mostdeployed.html, 25.04.2019.

[37] Y. Gökşen, H. Aşan, "Veri büyüklüklerinin veritabanı yönetim sistemlerinde meydana getirdiği değişim: NOSQL", *Bilişim Teknolojileri Dergisi*, 8(3), 125–131, 2015.

[38] M. Fuchs, W. Höpken, M. Lexhagen, "Big data analytics for knowledge generation in tourism destinations – A case from Sweden", *Journal of Destination Marketing & Management*, 3(4), 198–209, 2014.

[39] Internet: S. Sayad, An introduction to data science: data preparation, https://www.saedsayad.com/data_preparation.htm, 20.04.2019.

[40] J. R. Miglautsch, "Thoughts on RFM scoring", *Journal of Database Marketing & Customer Strategy Management*, 8(1), 67–72. 2000.

[41] K. K. Tsiptsis, A. Chorianopoulos, **Data mining techniques in CRM: inside customer segmentation**, John Wiley & Sons, 2009.

[42] P. Anitha, M.M. Patil, "RFM model for Customer Purchase Behavior using K-Means Algorithm", *Journal of King Saud University-Computer and Information Sciences*, In Press, 2019.

[43] D. Birant, "Data mining using RFM analysis", **Knowledge-oriented applications in data mining**, Editor: Funatsu, K., IntechOpen, 91-108, 2011, Available: https://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-using-rfm-analysis.

[44] Internet: Putler Analytics – RFM analysis for successful customer segmentation, https://www.putler.com/rfm-analysis, 26.04.2019.

[45] S. A. Sutresno, A. Iriani, E. Sediyono, "Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan", *Jurnal Teknik Informatika dan Sistem Informasi*, 4(3), 443–440, 2018.

[46] T. Jackson, "CRM: From art to science", *Journal of Database Marketing & Customer Strategy Management*, 13(1), 76–92, 2005.

[47] P. Harrigan, E. Ramsey, P. Ibbotson, "Critical factors underpinning the e-CRM activities of SMEs", *Journal of Marketing Management*, 27(5-6), 503–529, 2011.