

Patel, Ronit; Perret, Jens K.; Samunderu, Eyden

Article

Unbundling CRM – A RFMC Perspective

Research Journal for Applied Management (RJAM)

Provided in Cooperation with:

International School of Management (ISM), Dortmund

Suggested Citation: Patel, Ronit; Perret, Jens K.; Samunderu, Eyden (2022) : Unbundling CRM – A RFMC Perspective, Research Journal for Applied Management (RJAM), ISSN 2701-6633, BoD – Books on Demand GmbH, Norderstedt, Vol. 3, Iss. 1, pp. 37-52

This Version is available at:

<https://hdl.handle.net/10419/324723>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Patel, Ronit; Perret, Jens K.; Samunderu Eyden

Unbundling CRM – A RFMC Perspective

Abstract

Customer lifetime value (CLV) has been gaining increasing importance in many areas but the (RFM – recency, frequency, monetary) segmentation framework is the most commonly used due to the availability of customer data as well as ease of calculation. One of the drawbacks of the RFM segmentation framework is that it is unable to identify behavior correctly in the face of customers that show hot and cold periods of visiting and buying (binge buying). This kind of phenomenon is called clumpiness (C).

One of the empirical findings of this study is that such a kind of behavior is quite prevalent and clumpy customers generate (if not significantly) more revenues than customers who are non-clumpy. Even if clumpiness does not account for high revenue customers, this study shows by use of a correlation and cluster analysis that the clumpiness statistic significantly adds to the RFM framework making the establishment of an RFMC framework immanent.

The clumpiness phenomenon is studied by performing a detailed empirical analysis on a representative UK based retailer's dataset. Regression and neural network analysis were implemented to identify the main drivers for clumpiness. Cluster Analysis was used to identify customers that have similar behavioral characteristics and to sort them in small groups so that targeting a group of customers is possible. Thus, clumpiness can detect customers that are high potential and high risk in nature, which were previously unseen.

1 Introduction

The retail sector is essential to the world economy, as it provides a substantial amount of employment to skilled and unskilled labor. The retail industry has been growing from the 19th century onwards and with the advancement of new e-commerce, there has been immense as well as tough competition. While the industry continues to undergo constant change at a rapid rate, amid all changes what remains constant are the consumers and their growing demand. Consumers are becoming more powerful, with the decision of purchase lying within them. To stay competitive in the market, many retailers have shifted their investment as well as business strategies over the past decades. Organizations have moved from growth via new stores to developing new business models.

To adjust to changing conditions, organizations have to adopt customer-centric strategies, helping the organization to retain loyal customers and increase repeat purchases as well as the monetary value thereof. To maintain relationships with their customers, organizations have invested heavily in IT infrastructure by creating databases that can store a large amount of customer as well as transaction-related data. For each customer, many data points are collected, ranging from name to gender to transaction amount, etc., allowing to analyze the customer's complete purchasing history (so-called

360° customer view). However, the information obtained is seldom integrated into other business functions, such as customer relationship management (CRM) or Enterprise resource planning (ERP).

In this context, this study aims to develop a methodology to support Analytical CRM in the retail industry, such that the relationship between organizations and their customers can be reinforced. This research focuses on customer identification, attraction, development, and retention by applying data mining techniques to extract knowledge from large databases to build a customer lifetime value (CLV), model.

Traditionally, recency, frequency, and monetary value (RFM) models have been and are still used by many organizations to determine the CLV of a customer. But this methodology doesn't fully explain the critical aspects of customer behavior. Zhang et al. (2015) propose a new individual-level measure of a customer's history known as clumpiness (C), which is defined as the degree of nonconformity to equal spacing. Hence, this study continues the work by Zhang et al. (2015) and provides a framework for analyzing customer data.

In a first step the analysis of Zhang et al. (2015) is replicated for a UK based retailer's dataset. In particular using Monte Carlo Simulations the distribution of the clumpiness indicator is estimated to develop a statistical test for clumpiness. For a number of significance levels the test is then applied to the dataset to determine the customers reporting clumpiness in their shopping behavior. Based on this classification of test is conducted whether a significant difference exists between clumpy and non-clumpy customers in regard to the revenue they generate, with the test showing that clumpiness as an indicator does not add to the detection of more valuable customers. A succeeding correlation analysis, however, determines that the clumpiness statistic as such contributes to the RFM framework and it is not simply a new concept which is already covered by the three existing components. This result is later backed by cluster analysis as well. Based on the assumption that an RFMC framework is superior to an RFM framework using an analytical hierarchical process is used to determine weights for all four components of the new RFMC framework. Finally, cluster analysis is used to test the robustness of the previous results regarding the value added by clumpiness to the RFM framework and the determination of weights and using an artificial neural networks approach impact factors of customers ranked as clumpy are determined.

Note, that the analysis take place based on an empirical foundation and not on a set of a priori assumptions. Thus, it can be seen as a case study on the process of analyzing purchasing data regarding the RFMC framework. The study concludes in the fifth and final chapter with a discussion of the practical implication of this study.

2 Combining CRM and Customer Analytics

Customer Relationship Management (CRM) on the one hand is a strategic process of recognizing the most profitable customers which an organization can serve and how they can develop interactions (Payne & Frow, 2005). It is divided into three components: Strategic, Operational, and Analytical. Analytical CRM – the focus of this study - consists of the Information management process, which focuses

on the development, creation, and exploitation of customer-related data for the organization's strategy. It focuses on the mining of customer data for accomplishing the objective of a company's strategy and operational goals (Iriana & Buttle, 2007).

Customer analytics on the other hand is a process by which data mining techniques are used to extract information regarding customer behavior via profiling, segmentation, clustering or the development of predictive models. Grouping can help organizations to optimize their campaigns and target the right customers with the right offers. Customer analytics helps organizations in acquiring customer insights.

The essence of acquiring customer knowledge is knowing who the customers are (customer profiling) with the capability to identify strategically significant customers or groups of customers. Customer Analytics predicts the possible actions that are likely to be taken by customers based on their previous behavior patterns. It can guide staff that has direct contact with customers as to which offers can improve their buying rate, satisfaction, and make real-time recommendations on the best offers available. Moreover, customer analytics can help organizations to allocate their resources differently for individual customers based on their customer lifetime value (Kumar & Reinartz, 2018, 50 ff.).

The use of customer analytics has a positive impact on an organization's performance. With the use of customer analytics, extensive use of data and models and fact-based management drive decision and actions, where data and models are defined at the customer level. However, most analytical models developed focus on customer transactions alone (Bijmolt et al., 2010). Therefore, organizations need to develop customer metrics that they can trace over a span of time. Customer behavior needs to be continuously observed in order to recognize behavior patterns and trends and to distinguish any abnormal behavior or emerging patterns. Monitoring and tracking should be based on predefined criteria to supervise on what and how to observe, and what results to anticipate.

To propose an effective Customer Relationship Management (CRM) - Data Mining framework for the prediction of customer behavior in the area of retail - one can refer to the process introduced by Bahari and Elayidom (2015).

3 The RFMC Framework

While the previous chapter motivated in general the necessity of CRM Analytics and related methods it has to be detailed in how far especially loyalty management can profit data-driven analytical tools. In this regard the classic RFM framework approach is detailed. While the RFM framework is an established marketing tool since more than two decades already (Hughes, 1996) studies like Kaul (2020), Maraghi et al. (2020) and Sun et al. (2020) illustrate the relevance it still plays in current research as well as in current research endeavors. Nevertheless, The framework approach did not remain unchanged over the years. Studies like Zhang et al. (2015), Josep (2020) and Oztaysi and Kavi (2020) show that the framework is continuously advanced still.

Oztaysi and Kavi (2020) introduce the use of a fuzzy k-means algorithm as part of the framework as such and is representative for studies that advance the RMF approach via the introduction of more sophisticated tools that are applied in the course of the basic framework itself. Josep (2020) on the other hand introduces the aspect of activating customers as an additional aspect considered in the

course of the framework. It is representative for the class of studies that shift the focus of the pure RFM framework into including additional information that promise insights on customer behavior. By focussing on customer activation, the aspect added by Josep (2020), the study, however, is extended by a factor not of original customer behavior but by the potential customer reactions. Finally, Zhang et al. (2015) incorporates the purely customer driven concept of clumpiness (Zhang et al., 2013) into the RFM framework. It is for this reason that the present study follows the approach initiated by (Zhang et al., 2015).

To alleviate the overall understandability, this chapter introduces the RFM framework as such followed by a deduction of the concept of clumpiness and a discussion of its general benefits in the context of the RFM framework.

3.1 The RFM Framework

The RFM model incorporates the three aspects recency, frequency, and monetary value. It employs three metrics to evaluate customer behavior and customer value and is frequently used in practice.

1. Recency – How much time has passed since the customer’s last order.
2. Frequency - How often a customer orders from the company in a certain pre-defined period.
3. Monetary value - The amount that the customer spends on average per transaction.

The general idea of RFM is to group customers based on their RFM scores. The resulting groups of customers are associated with their purchase behavior, e.g. the possibility of a customer reacting to a marketing campaign. RFM closely follows the transition matrix path in that it also traces customer behavior over a period of time in what is called a state space. The higher the RFM score of the customer, the more valuable he is for the organization. In recent studies, authors suggested using a weighted RFM score - instead of an RFM score with equal weights. They assigned weights to R, F, and M. depending upon the characteristics of different industries.

The measurement of CLV via the RFM framework is one of the most common methods (Khayvand et al., 2011). RFM has been used for many years in direct marketing to predict customer behavior and to generate customer segments. If an organization has a database of its customers, which includes their purchase history, the firm can perform an RFM analysis at practically no cost. From a behavioral perspective, the RFM method evaluates the relationship between the organization and its customers (Shih & Liu, 2003). Customers who purchased recently, frequently, and spent large amounts of money are more inclined to react to mailings and therefore represent more attractive candidates for future marketing campaigns (Coussement et al., 2014).

3.2 Clumpiness

According to Zhang et al. (2013), Clumpiness is defined as the degree of nonconformity to equal spacing and demonstrates an essential component in a better understanding of recognizing a profitable customer. To better capture clumpiness, Zhang et al. (2013) provides a rich analysis of this issue and proposes a new class of clumpiness measures that are shown to have larger statistical power in exten-

sive simulations under a wide variety of data-generating mechanisms (alternative hypotheses as compared to the null of stationarity). Although details are provided in Zhang et al. (2013), the central properties of such measures are:

- Minimum – The measure should be at its minimum if the events are equally spaced.
- Maximum - The measure should be at its maximum if all the events are gathered together.
- Continuity - Shifting event times by a small amount should only change the measure by a small amount.
- Convergence. As events move closer (further apart), the measure should increase (decrease).

These four features provide a comprehensive description of any clumpiness measure.

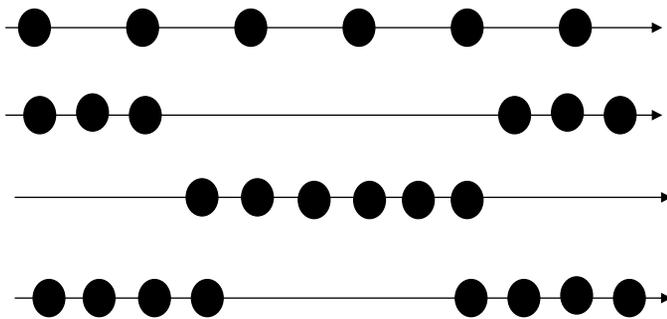


Figure 1: Exemplary Cases of Customer Behavior

Source: Zhang et al. (2015)

Figure 1 shows possible situations of customer behavior. While the first and third situation are not considered clumpy the second and fourth are.

According to Zhang et al. (2015), a clumpiness measure is a formal statistical test that is needed to determine the importance of an observed clumpiness value. Assuming the null hypothesis is random sampling without replacement, where n (the number of events) and N (number of trials) are known Monte Carlo simulation is applied and the z-table, the table of critical values, can be estimated. To measure clumpiness, H_p , an individual level entropy measure (metric) from the class of convex and symmetric IET statistics is chosen. It can be written as:

$$H_p = 1 + \frac{\sum_{i=1}^{n+1} \log(x_i) * x_i}{\log(n+1)}$$

with $x_i = N+1$ at the i -th event (this is done to control the length of the observation period).

In an information system, entropy is a measure of uncertainty. It quantifies the expected value of the information contained in a message. Although it is intended to represent a different story, its characterization and profound indication of disorder can provide a great measure for clumpiness.

A distinguishing feature of using this type of clumpiness measure is that a given percentage variation in any of the IETs has the same effect on the measure itself. Hence, it “normalizes” the ranges being aggregated (Zhang et al., 2013). Since the clumpiness measure is calculated by using scaled inter-event times (i.e., dividing by $N+1$), it is conceptually invariant to the scaling of time units. However, the timeframe should not be too long; otherwise, the data points would be aggregated, and information

about inter-event times (IETs) would be wasted. On the other hand, a too-short timeframe would give rise to over-fitted patterns.

Steps required for computation of clumpiness

1. Convert the individual-level (purchase/visitation) data into incidence data (0/1).
2. Compute the inter-event times (IETs), specifically x_i, x_{i-1}, \dots , where the x_i represent the time occurrence of i -th visitation.
3. Rescale the IETs. Divide IETs by $N+1$.
4. Compute the entropy given by the formula

$$H_p = 1 + \frac{\sum_{i=1}^{n+1} \log(x_i) * x_i}{\log(n+1)}$$

5. Create a cut-off table (critical values) using Monte Carlo simulation at different significance levels and find the corresponding value based on frequency
6. Determine the significance of clumpiness by comparing the clumpiness values with the critical cut-off point. If H_p is larger than the critical value then the customer is considered as a clumpy customer else the customer is not clumpy.

4 Empirical Analysis

Since clumpiness is a metric based approach, Zhang et al. (2015) hints that the customer lifetime value framework should be transformed from the traditional RFM model to a newer model of RFMC. Where the C, i.e. clumpiness, adds additional predictive power to the RFM model in recognizing valuable customers. Hence, this study proceeds with the recommendation suggested by (Zhang et al., 2015)

After checking whether clumpiness is present in the dataset that is used in this study and determining how many customers are clumpy it is tested whether they are generating higher revenues than non-clumpy customers. Secondly, a weighted RFMC model where the weights are determined through an analytical hierarchical process (AHP) is conceived. Subsequently, data mining techniques such as clustering are implemented to identify groups of customers that are valuable to an organization. Finally, artificial neural networks are used to check which of the variables of the RFMC framework, if any, is the most important in predicting clumpy behavior per se.

4.1 Description and Pre-processing of the Dataset

The dataset that is used in this research work was obtained from Kaggle, an online community that allows users to publish datasets, explore, and build models in a web-based data science environment. The dataset contained transactions between 01/12/2010 and 09/12/2011 from a UK based and registered non-store online retailer. The company sells unique all occasion gifts, and many of the organization's customers are wholesalers. The variables included in the dataset are Invoice No, Quantity, InvoiceDate, Unit Price, Customer ID, Country to which they belonged. Since this data stems from a UK retailer, it is safe to assume that the unit prices are measured in Pounds Sterling. However, as long as all purchases are in the same currency the calculations are not affected.

The dataset contains 541,909 transactions. Since some entries for the unit prices and the quantity were less than or equal to zero, those transactions were eliminated. After performing this cleaning task, the number of transactions is reduced to 168,631.

While the time of purchase is not required for determining the CLV of a customer through RFMC Analysis the date of purchase is. Additionally, for RFMC analysis, another variable is required that is the total price. The total price results as the product of quantity and unit price. Hence, both variables are replaced by the total price. As sales were recorded product-wise but an invoice usually consists of different positions to generate the total costs of an order all positions with the same invoice number were summed up resulting in a total of 7,961 orders for the years 2010 and 2011. As the majority of orders occurred in 2011 which is covered in full as compared to 2010 this study limited to only these orders yielding a final set of 7,070 orders in 2011 by 2,821 unique customers.

4.2 Clumpiness as a Valuable Complement to the RFM Framework

While Table 1 summarizes the core statistics for the four indicators of the RFMC framework, Table 5 in the appendix summarizes the critical values for the clumpiness statistic in dependence on the frequency score. Using this z-table in conjunction with the clumpiness scores of the 2,821 different customers depending on different margins of error Table 2 is constructed showing the number of clumpy customers as well as the average worth of their orders.

Table 1: Core Statistics for RFMC

	R	F	M	C
Mean	151.8245	2.5041	458.6717	0.2538
Standard Deviation	98.8776	3.5469	3213.0330	0.2238
Minimum	22.0000	1.0000	5.9000	-0.0012
Maximum	364.0000	83.0000	168,469.6000	0.9689

Source: Own calculations

Table 2: Classification of clumpy and non-clumpy customers

Significance Level	Number of - Clumpy customers	Number of - Non-clumpy customers	Average revenue of Clumpy customers	Average revenue of Non-Clumpy customers	% increase in revenue	% clumpy customers	% non-clumpy customer
90%	411	2410	522.72	447.75	14%	17.06	82.94
95%	263	2558	620.41	442.04	29%	10.28	89.72
99%	85	2736	662.69	452.33	32%	3.11	96.89
99.5%	52	2769	781.11	457.82	41%	1.88	98.12

Source: Own calculations

The table clearly indicates that clumpiness is present in the implemented dataset even in the very conservative case of allowing only for an α of less than 0.5% as margin of error. It is this conservative

case that is of particular interest as in this case no clumpy customers with a frequency of 1 exist anymore. (Customers with a frequency of 1 skew the analysis as in their case clumpiness is solely determined by the date of the purchase.) The table also motivates that on average clumpy customers generate higher revenues as compared to non-clumpy customers.

Using a Shapiro-Wilk-test to check for normality in the clumpiness statistic little does it surprise that the clumpiness scores in the sample are not normally distributed but significantly rightwards skewed. Considering that this characteristic will hold for the population as well a U-test instead of a t-test is used to compare the revenues via clumpy and non-clumpy customers. It turns out that although the values for the two groups as reported in Table 2 considerably differ, these differences are not significantly different from zero (accepting a margin of error of 5%).

Considering the results from Table 1 that all statistics are measured on different scales, it is imperative for the ongoing analysis to z-standardize them first. To account for severe outliers from the outset all customers whose z-standardized score exceed the range from -3 to +3 are excluded from the analysis reducing the dataset to only 2,735 customers.

To get a better impression of the structure of the four variables of the RFMC model correlations between them are calculated. Due to the issue with normality which persists with the R, F and M variables as well a Spearman test is implemented here:

Table 3: Correlations between the parts of the RFMC framework

	R	F	M	C
R	-	0.449 (0.000)	0.117 (0.000)	0.027 (0.145)
F	0.449 (0.000)	-	0.141 (0.000)	0.109 (0.000)
M	0.117 (0.000)	0.141 (0.000)	-	0.031 (0.103)
C	0.027 (0.145)	0.109 (0.000)	0.031 (0.103)	-

Source: Own calculations

Table 3 reports for all pairwise correlations the correlation coefficient as well as the significance level in parentheses. While the three parts of the original RFM framework are highly correlated with each other, the clumpiness variable is highly correlated only with frequency but not with the other two. This can be interpreted in such a way that clumpiness integrates into the CLV framework but at the same time provides additional valuable insights not yet covered by the original framework. This insight could be replicated via a factor analysis design strengthening it in this regard. A factor analysis, however, is not explicitly conducted at this point. As mentioned above it cannot be assured that the four variables are normally distributed across the sample or the population.

4.3 Determination of RFMC Weights

While there have been various combinations, and weightings proposed, judging and weighting of RFMC variables remains highly subjective (Shih & Liu, 2003). This study uses the Analytical Hierarchy Process (AHP) to evaluate the weight of each RFMC variable by ranking them according to the importance for all customers. This process is explained in detail in Shih and Liu (2003).

Before the AHP can be applied the dataset needs to be re-standardized as the AHP requires non-negative values. The formula that is used for normalization in this context is:

$$X' = \frac{X - X_s}{X_l - X_s}$$

Where X' is the value after standardization, X is the value that is related to the variable, X_l , and X_s are the maximum and minimum values of the variable.

Using the approach described by Shih and Liu (2003) yields a 4x4 (each column or row representing one variable of the RFMC framework) matrix of pairwise comparisons. The matrix is then column-normalized by dividing the column score by the row score. After column-wise normalization and averaging across rows (which in this case are all identical) the relative weights can be calculated for each customer. Averaging the relative weights for all customers results in the following values:

$$w_R = 0.38 \quad w_F = 0.15 \quad w_M = 0.15 \quad w_C = 0.32$$

This shows not only that clumpiness is a critical part of the RFMC framework – as suspected from the correlation analysis already – but with an average relative weight of 0.32 it is the second most important aspect and close in its weight to the first most important part recency.

4.4 Clustering as a Stability Test

The AHP in the previous section resulted in individual relative weights for each customer. Thus, to determine which customers behave similarly a clustering analysis in regard to the determined weights is conducted. This allows to detect groups of customers that share similar behavior patterns. In this study a k-means approach as implemented in the SPSS 23 package is used for clustering. As the k-means algorithm requires an ex-ante fixed number of clusters an analysis is conducted for two and three clusters. Since the two cluster solution does not provide suitable answers only the results for the three cluster approach are discussed below. Using more than three clusters at this stage has been deemed irrelevant as the three clusters approach already provides satisfactory answers. The cluster centers of the three cluster solution are summarized in Table 4

Table 4: Cluster Centers - 3 Cluster Solution

Cluster	1	2	3
R	0.0527	0.3389	0.1987
F	0.0052	0.0297	0.0091
M	0.0096	0.0102	0.0087
C	0.1629	0.0883	0.0323

Source: Own calculations

Of the 2,735 valid cases 380 are assigned to cluster 1, 1,148 to cluster 2 and the remaining 1,207 to cluster 3. To additionally assure that the clusters not only with regard to all four parts of the RFMC framework considered at the same but for each part by itself represent a powerful separation a Kruskal-Wallis-H-test has been performed. The four H-tests all turned out highly significant showing that the three clusters represent groups that are distinctly different with regard to all four aspects of the RFMC framework. In all four cases the difference between the groups has been highly significant. With

a significance level of only 0.008 the monetary variable is the weakest of the four variables of the RFMC framework. Considering the results in Table 4 this result does not surprise.

Cluster 1 represents those customers that are mainly described by their clumpy behavior. In cluster 2 there are those customers that have a high RFM score. Finally, in Cluster 3 there are primarily customers with a high weight for recency. While clusters 2 and 3 would have been covered by the RFM framework. The presence of the small but still pronounced Cluster 1 with roughly 14% of all customers stresses once more the relevance of clumpiness as an essential part in a comprehensive RFMC framework. Comparing the results in Table 2 and Table 4 it shows that the size of the cluster of primarily clumpy customers roughly coincides with the group of clumpy customers at a margin of error of 90%. The arguments given above in favor of using on the group defined by using a margin of error of 0.5% can thus be considered a very conservative one.

4.5 The RFMC Framework as a Predictor for Clumpy Customers

To test whether adding the clumpiness statistic to the RFM framework extending it to the RFMC framework helps in detecting clumpy customers an artificial neural network (ANN) approach has been considered. The advantage of ANNs over rigid regression models is that they act more flexible with regard to non-linearities in the underlying relation. Additionally, as shown in Table 3 the four variables are strongly correlated with each other. In a regression model this would have resulted in a severe problem with multi-collinearity. Also, the absence of normality in all four variables does not pose a critical problem in an ANN approach.

To realize the ANN model a multilevel perceptron is estimated. The variables resulting from Table 2 - at a margin of error of 10% as well as of 0.5% - that measure whether a customer is clumpy (value of 1) or not (value of 0) are used as dependent variables and the four indicator variables measuring the z-standardized score for recency, frequency, monetary value and clumpiness are used as independent variables. Additionally, an ANN with only a single layer has been estimated as trials with using two layers revealed that it does not significantly decrease the relative error.

In this context the discussion of Table 2 can be recalled in that a high clumpiness score not necessarily results in clumpy behavior if the frequency score is 1. Thus, this part of the overall analysis is not self-evident. This can also be evidenced from the results of the estimation of the ANN. In this study a learning design for estimating the ANN has been selected where roughly 70% of the sample are used to estimate the synaptic weights while the remaining 30% are using to test the quality of the estimated model. This approach not only allows to get an estimate of the out-of-sample performance of the resulting model it also assures – as the assignment of cases to the estimation and the testing set is randomized - that a lucky assignment generates biased results. However, to get a representative impression of the true quality of the results the analysis has to be performed multiple times with the results of all results being averaged. In the context of this study for each of the two dependent variables the estimation has been repeated twenty times.

When using the more relaxed version of clumpiness, i.e. the variant with a 10% margin of error, the relative error of the models turns out to be 1.075% for the estimation set and 1.005% for the testing set. It is interesting to note that the model actually works better out-of-sample than in-sample. With

regard to the four variables of the RFMC framework clumpiness and frequency play the most important role, whereas in most cases clumpiness is the most important variable.

In contrast, if the more conservative version of clumpiness is used, i.e. the variant with a 0.5% relative error, decreases to 0.805% for the estimation set and to 0.785% for the testing set. Again, the model performs slightly better out-of-sample than in-sample. With regard to the RFMC framework it is again the frequency and clumpiness that report the major importance, however, where before clumpiness has been the most important variable, here it is frequency.

That clumpiness and frequency both play a major role in determining whether a customer is truly clumpy seems only reasonable following the discussion of Table 2 where it has been stated that clumpiness can also be an effect of infrequency, in particular of one-time shoppers. Thus, the results of the $\alpha = 0.5\%$ case might be more relevant than those of the $\alpha = 10\%$ case as in the 0.5% case no clumpy customers report a frequency score of 1 anymore. While an error of 1% or less can be considered marginal at best it still might signify that between 5% to 10% of the clumpy customers remain undetected as the model on average generates more false negatives than false positives; in particular in the testing sets. Nevertheless, an overall error of roughly 1% shows that the RFMC framework has forecasting quality and can significantly help companies in detecting highly valuable clumpy customers.

5 Conclusions

5.1 General Insights

CRM is an important issue to improve an organization's competitiveness in today's changing environment in the retail industry. This study contributes to the field of CRM by contributing innovative analytical models backed by data mining techniques. The models introduced herein are adapted to real-world problems motivated by the difficulties faced in identifying customer behavior. This study is an example of application-driven theory, which is considered of the highest importance in the field of customer relationship management and data analytics. The contributions of the study can be summarized as follows:

- Replication of previous studies on clumpiness to ensure the validity of the implemented dataset.
- Segmentation of customers according to them being clumpy or not does not help in the determination of a customer group that generates significantly higher revenues.
- Exemplification that the clumpiness statistic though correlated with the frequency statistic does provide additional insights and can thus be used to expand the RFM framework.
- The model makes use of an analytical hierarchical process (AHP) to determine the weights of the four components of the RFMC framework.
- Identification of customer groups that behave in similar patterns by making use of clustering techniques.
- Identification of the important variables concerning clumpy customers through artificial neural networks.

Summarizing, aside from adding to the literature on clumpiness, the study provided an empirically driven outline on analyzing purchasing data in the context of the expanded RFMC framework.

5.2 Managerial Applications

Since clumpiness is a phenomenon of identifying customer behavior, in campaign building it suggests the retailer perform a trigger-based dynamic campaign. For example, a customer could be shopping only on the days that the retailer is providing huge discounts, so the clumpiness methodology could help the retailer to identify such customers and then target them accordingly. This could help the retailer to check whether clumpy customers respond to the trigger based dynamic campaign. To deepen insights on such a customer, the retailer could identify the categories that the customer has purchased during the past discount day and target the customer with cross-selling or up-selling categories.

Since clumpiness is observing the buying pattern of the customer, it makes clumpiness an important parameter for retailers' loyalty programs. According to (Kumar, 2009, 16), the Loyalty program, in contrast to other marketing efforts, has more effect on average purchase frequency than on penetration. From the empirical analysis, it was observed that frequency was one of the main drivers of clumpiness; hence clumpiness is an important factor when deciding customer loyalty. This is the main reason why loyalty programs are most engaging with existing and heavy buyers.

High-frequency behavior tends to increase the share of requirements asked by the customers and therefore tends to increase the effort spent on them. An increase in loyalty could also help the retailer to improve the brand penetration level. This means that the brand will have a higher average purchase frequency than would be expected, given the level of penetration. Finally, the retailer's main aim is to increase the profitability of the company, and since clumpy customers are more profitable in nature (i.e., clumpy customers generate 40% additional revenue as compared to non-clumpy customers). Therefore, it is important to integrate clumpy customers in the loyalty program.

5.3 Limitations and Outlook

The main limitation of the present study is that it is based solely on purchasing data for a single year from a single UK retailer. Thus, any generalization of the study might be precluded from the outset. Since the study successfully replicated part of previous studies and their results for US retailers (Zhang et al., 2015), it is, however, highly likely that similar results will present themselves if purchasing data from other retailers for from other countries is implemented. Nevertheless, a replication of this study in different sectoral or regional contexts is highly advised. Additionally, the RFM framework, even though mainly implemented in retail, is a general purpose CRM approach which requires the replication of this study as well in the context of different industries.

Since there is no correct way of predicting whether customers will return in the future, it would be appropriate to build a churn model for the clumpy customer that could help the UK based retailer to identify all potential churners. The Bayesian learning or Markov model could be an excellent methodology to develop such a new churn prediction model. For future research, it would be worthwhile to extend the work developed in this study to support the company's direct marketing initiatives. It would be appropriate to create a model to predict customer's responses to direct marketing campaigns.

Additionally, it would be insightful to analyze the buying patterns of clumpy customers. It could be possible that there are some categories that are brought by the clumpy customers in huge categories, and these categories are influencing the clumpy behavior of the customer. Since the UK based retailer

considered herein is selling its products to international customers, it would be interesting to identify whether international customers or domestic customers are more clumpy in their behavior. This would help the retailer to target its marketing actions more efficiently.

As future research, it would be interesting to compare the different customer lifetime value (CLV) models to identify which CLV model would be best suited for a UK based retailer. The retailer could then develop a retention model (loyalty program) that can predict accurately to what level and how much a customer will consume in a visit, day, or a year. This ability can aid the retailer with developing strategies to retain clumpy customers by making them feel like a privileged customer; privilege could lead to the customer being more loyal. Such types of loyalty programs for the retailer would help to provide valued customers with valuable rewards, all with the intent of keeping them being clumpy in nature and spending more money with the organization over the competitors. The information gathered would result in a lifetime loyalty and thereby profit the organization in the long term.

The research methodologies proposed in this study could be applied to other business environments in which customer relationship management (CRM) is crucial to guarantee an organization's competitiveness and success in the business environment. Fields that could be studied in the future cover, for example, banking, insurance, media & entertainment, and telecommunications, etc. Finally, the study focused on a UK online retailer. It still remains to be analyzed in how far the results of this study hold for offline retailers from different sectors and countries.

References

- Bahari, T. F., & Elayidom, M. S. (2015). An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour. *Procedia Computer Science*, 46, 725–731.
- Bijmolt, T.H.A., Leeflang, P.S.H., Block, F., Eisenbeiss, M., Hardie, B.G.S., Lemmens, A., & Saffert, P. (2010). Analytics for Customer Engagement. *Journal of Service Research*, 13(3), 341–356.
- Coussement, K., van den Bossche, F.A.M., & Bock, K. W. de (2014). Data Accuracy's Impact on Segmentation Performance: Benchmarking RFM Analysis, Logistic Regression, and Decision Trees. *Journal of Business Research*, 67(1), 2751–2758.
- Hughes, A. M. (1996). Boosting Response with RFM. *Marketing Tools*, 5, 4–10.
- Iriana, R., & Buttle, F. (2007). Strategic, Operational, and Analytical Customer Relationship Management. *Journal of Relationship Marketing*, 5(4), 23–42.
- Josep, A. V. (2020). Predicting Customer Behavior With Activation Loyalty per Period. From RFM to RFMAP. *Esic Market Economics and Business Journal*, 51(3), 609–637.
- Kaul, S. (2020). Customer Segmentation Through Rfm (Recency, Frequency, Monetary) Model In Retail Sector. *Solid State Technology*, 63(2), 8918–8931.
- Khayvand, M., Zolfaghar, L., Ashoori, S., & Alizadeh, S. (2011). Estimating Customer Lifetime Value Based on RFM Analysis of Customer Purchase Behavior: Case Study. *Procedia Computer Science*, 3, 57–63.

- Kumar, V. (2009). *Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty*. Pearson Prentice Hall.
- Kumar, V., & Reinartz, W. J. (2018). *Customer Relationship Management: Concepts, Strategy, and Tools*. Springer.
- Maraghi, M., Adibi, M. A., & Mehdizadeh, E. (2020). Using RFM Model and Market Basket Analysis for Segmenting Customers and Assigning Marketing Strategies to Resulted Segments. *Journal of Applied Intelligent Systems & Information Sciences*, 1(1), 35–43.
- Oztaysi, B., & Kavi, M. (2020). Fuzzy RFM Analysis: An Application in E-Commerce. In C. Kahraman, S. C. Onar, B. Oztaysi, I. U. Sari, S. Cebi, & A. C. Tolga (Eds.), *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference (1225–1232)*. Springer.
- Payne, A., & Frow, P. (2005). A Strategic Framework for Customer Relationship Management. *Journal of Marketing*, 69(4), 167–176.
- Shih, Y.-Y., & Liu, C.-Y. (2003). A Method for Customer Lifetime Value Ranking - Combining the Analytic Hierarchy Process and Clustering Analysis. *Journal of Database Marketing & Customer Strategy Management*, 11(2), 159–172.
- Sun, G., Xie, X. F., Zeng, J., Jiang, W., Lin, M., & Huang, Y. X. (2020). Using improved RFM Model to classify Consumer in big data Environment. *International Journal of Embedded Systems*, 14(1), 54–64.
- Zhang, Y., Bradlow, E. T., & Small, D. S. (2013). New Measures of Clumpiness for Incidence Data. *Journal of Applied Statistics*, 40(11), 2533–2548.
- Zhang, Y., Bradlow, E. T., & Small, D. S. (2015). Predicting Customer Value Using Clumpiness: From RFM to RFMC. *Marketing Science*, 34(2), 195–208.

Appendix:

Table 5: Cut-off table (z-table) for different significance level at different frequencies

Frequency	Cut-off value - 95%	Cut-off Value - 90%	Cut-off value - 99%	Cut-off Value - 99.5%
1	0.8335	0.7055	0.9511	0.9728
2	0.5846	0.4842	0.7701	0.8173
3	0.4698	0.4045	0.6363	0.6866
4	0.4098	0.3501	0.5336	0.5874
5	0.3663	0.3151	0.4722	0.5132
6	0.3320	0.2897	0.4292	0.4658
7	0.3070	0.2696	0.3898	0.4216
8	0.2877	0.2543	0.3594	0.3936
9	0.2674	0.2406	0.3416	0.3682
10	0.2568	0.2285	0.3217	0.3482
11	0.2457	0.2193	0.3018	0.3257
12	0.2337	0.2103	0.2881	0.3077
13	0.2249	0.2024	0.2771	0.2968
14	0.2168	0.1958	0.2638	0.2825
15	0.2103	0.1895	0.2541	0.2716
16	0.2045	0.1843	0.2467	0.2635
17	0.1977	0.1801	0.2387	0.2540
18	0.1921	0.1748	0.2287	0.2433
19	0.1867	0.1703	0.2231	0.2377
20	0.1839	0.1668	0.2178	0.2322
23	0.1706	0.1570	0.2004	0.2122
24	0.1681	0.1544	0.1963	0.2081
25	0.1637	0.1503	0.1916	0.2035
27	0.1579	0.1455	0.1836	0.1943
28	0.1548	0.1428	0.1797	0.1891
30	0.1508	0.1396	0.1748	0.1837
32	0.1457	0.1348	0.1671	0.1761
34	0.1411	0.1311	0.1626	0.1726
35	0.1385	0.1295	0.1590	0.1667
37	0.1362	0.1263	0.1542	0.1615
49	0.1174	0.1106	0.1328	0.1388
66	0.1004	0.0948	0.1110	0.1155
83	0.0873	0.0830	0.0962	0.1002

Source: *Own calculations*

Authors

Ronit **Patel**, M.A. studied Engineering at the Dwarkadas J Sanghvi College of Engineering in Mumbai and International Business at the International School of Management in Dortmund. Before his post-graduate studies Mr. Patel worked with an Indian Consultancy as senior data analyst. Currently he is employed by a major German insurance company focussing on project management in IT service lines. His research interests include data analytics in the context of customer relationship management.



Prof. Dr. Jens K. **Perret** studied business mathematics and economics at the University of Wuppertal. He received his doctorate in economics from the University of Wuppertal with a thesis on the knowledge society in Russia. For more than a decade Mr. Perret worked at the European Institute for International Economic Relations and at the Chair of Macroeconomic Theory and Policy at the University of Wuppertal. He was a lecturer at the Technical University of Kaliningrad. Since September 2016 he holds a professorship for Economics and Statistics at the International School of Management in Cologne.



Prof. Dr Eyden **Samunderu** is a leading expert in the area of air transport with extensive consulting experience. He holds a PhD in Strategic Management and is a Program Director for a Master's program at the International School of Management in Dortmund Germany. He is also sitting as an Board Advisor at Africa Aviation Partners (London). His research interests include: aviation, multimarket contact, strategic groups and competition dynamics, container logistics and maritime economics.

