Received: 18 June 2021 | Revised: 2 November 2021 | Accepted: 1 March 2022

DOI: 10.1002/asmb.2677

RESEARCH ARTICLE

WILEY

# Ranking customers for marketing actions with a two-stage Bayesian cluster and Pareto/NBD models

**Ignasi Puig-de-Dou**[1,2] | **Daniel González Ibáñez**[3] | **Xavier Puig**[2]

[1]Department of Operations, Innovation and Data Sciences, ESADE Business School, Barcelona, Spain

[2]Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain

[3]Department of Marketing Research and Quantitative Methods, ESIC Business & Marketing School, Barcelona, Spain

**Correspondence**
Xavier Puig, Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona. Spain.
Email: xavier.puig@upc.edu

**Abstract**
Modelling customer behaviour to predict their future purchase frequency and value is crucial when selecting customers for marketing activities. The profitability of a customer and their risk of inactivity are two important factors in this selection process. These indicators can be obtained using the well-known Pareto/NBD model. Here we cluster customers based on their purchase frequency and value over a given period before applying the Pareto/NBD model to each cluster. This initial cluster model provides the customer purchase value and improves the predictive accuracy of the Pareto/NBD parameters by using similar individuals when fitting the data. Finally, taking the outputs from both models, the initial cluster and Pareto/NBD, we present some recommendations to classify customers into interpretable groups and facilitate their prioritisation for marketing activities. To illustrate the methodology, this paper uses a database with sales from a beauty products wholesaler.

**KEYWORDS**
Bayesian model, customer base analysis, customer lifetime value, marketing, model-based clustering, pareto/NBD model

## 1 | INTRODUCTION

A customer's profitability and propensity to become inactive are two important factors companies consider when ranking customers for marketing activities, especially for retention activities.

Despite the impact of customer profitability and propensity to become inactive on a company's overall results, many businesses still use basic rules of thumb to estimate these indicators. Although managerial heuristics can work as well as stochastic models in some situations,[1,2] a more formal assessment of customers can improve decisions on how to allocate scarce company resources more efficiently. As Lehmann et al.[3] highlighted, there has to be a balance between rigour and relevance. Here we consider both approaches using statistical models embedded within management knowledge.

Ideally, a company wants to channel resources toward more profitable customers. The traditional metric used to assess a customer's future profitability is the so-called customer lifetime value (CLV). CLV is the discounted value of all profits obtained from a customer during their future lifetime relationship with the company. The literature contains several methods for calculating the CLV for a specific customer, or the aggregate if estimating the present value for a customer group (e.g., References 4–9).

The aim of this article is to provide businesses with a tool to assess profitability and propensity to churn for a specific customer in noncontractual settings, where customers can stop interacting with the company without any prior notice.

The ultimate goal is to determine how best to spend marketing resources on their customer base. For example, for marketing retention activities, resources could be prioritised toward customers with a high profitability and a high risk of prompt inactivity.

The probability of a customer being inactive at a given moment and their expected number of purchases in the future, both required when calculating CLV, can be estimated from the Pareto/NBD model proposed by Schmittlein et al.[10] The Pareto/NBD model has been used extensively in the past (see, for instance, Reference 11 or 12). It has also been applied following a Bayesian approach[13–15] which can be used to compute the uncertainty measures for its different estimators, such as the probability of being active.

The CLV estimation also requires the average ticket value per customer. This value is usually obtained by bringing in another model, such as the Normal–Normal mixture model,[11] the Gamma–Gamma mixture model[5,16] or the Log-normal model.[4]

Segmenting customers before applying the Pareto/NBD model can lead to a greater degree of homogeneity in the transactions and/or dropout rates for each group. This reduces the uncertainty regarding their future behaviour and ultimately results in more accurate estimates from the model. In this sense, for example, Fader et al.[17] showed how segmenting customers based on a simple threshold for their first purchase value before applying the Pareto/NBD model substantially improved its outputs, such as transaction and dropout rates, and hence the CLV value. In fact, deterministic cluster segmentation is sometimes performed beforehand, for example into customer cohorts defined by the time of first purchase, acquisition channel or type of business (see Reference 11 or 18).

The main contribution of this paper is the application of an initial cluster model before running the Pareto/NBD model for two reasons. First, it groups customers into homogeneous clusters based on their purchase frequency and value for an improved Pareto/NBD parameter estimation. Second, it estimates each customer's average ticket value, which is required for the CLV calculation. We also propose a customer classification for selecting customers for specific, future marketing actions. However, the selection of the actual actions and an assessment of their effectiveness is beyond the scope of the current paper. These three steps are implemented using a database of all the sales from a beauty product wholesaler over a 2-year period.

The models in this paper are analysed from a Bayesian perspective to take advantage of its modularity and flexibility. Rossi and Allenby[19] present a good introduction and review of Bayesian analysis applied to marketing problems. Jen et al.[20] and Baesens et al.[21] contain some examples of Bayesian models for calculating purchase frequency using hierarchical models and Bayesian neural networks. Borle et al.[22] also used a Bayesian approach to improve estimates of the Pareto/NBD model using demographic variables when available.

The rest of this article is organised as follows: Section 2 describes the Bayesian clustering, the Pareto/NBD model and the outputs derived thereof. Section 3 presents the results and their practical application to the study case. Finally, we conclude the paper with a discussion section.

## 2 | DESCRIPTION OF THE MODELS

The Pareto/NBD model estimates the probability of a customer being active at the end of an observation period and their expected number of future purchases. This information, together with the average customer ticket value, provides their future revenue stream. Customer homogeneity increases the model's predictive power; which is the reason for the initial customer clustering, which also brings the extra benefit of estimating the average ticket value. The Pareto/NBD model is run on each of the resulting clusters independently, so information sharing among customers only occurs for customers with similar purchasing habits.

### 2.1 | Bayesian cluster model

The initial cluster model groups customers into a small number of clusters, each with a distinctive purchasing pattern. While assigning each customer to a cluster, the model also estimates the average ticket value.

We start with the assumption that the number of purchases per month for a given customer follows a Poisson distribution and their monthly expenditure a Normal one. The main problem with this approach is when there are months with no customer activity.

If we take $x_{ij}$ as the number of purchases made by customer $i$ in month $j$ and $y_{ij}$ as the total expenditure obtained from customer $i$ in month $j$, then they can be modelled as:

$$x_{ij} \mid \theta_1, \dots, \theta_K, S_i \sim \text{Poisson}\left(\theta_{S_i}\right), \tag{1}$$

and

$$y_{ij} \mid \eta_i, \sigma, x_{ij} \sim \text{Normal}\left(x_{ij}\eta_i, \sqrt{x_{ij}}\sigma\right), \tag{2}$$

where $S_i$ is a latent categorical variable denoting the customer's cluster and $K$ the total number of clusters. $S_i$ takes values over the range $\{1, \dots, K\}$ for customers $i = 1, \dots, n$ and is such that $S_i = c$ whenever the $i$th customer belongs to the $c$th cluster. $S_i$ is assumed to follow a discrete distribution with $P(S_i = c|\pi_1, \dots, \pi_K) = \pi_c$ for $c = 1, \dots, K$. Hence, $\pi_c$ becomes the proportion of customers belonging to the $c$th cluster. $\theta_c$ is the expected number of purchases per month for customers in cluster $c$. $\eta_i$ is the expected ticket value for the $i$th customer. This parameterisation contemplates the fact that the variance of the total expenditure per month increases when there are more purchases, $V(y_{ij} \mid \eta_i, \sigma, x_{ij}) = x_{ij}\sigma^2$, as one would expect. $y_{ij}$ is the sum of the values of $x_{ij}$ independent purchases with an expected ticket value of $\eta_i$ and a variance of $\sigma^2$. Therefore, the expectation and variance are $x_{ij}\eta_i$ and $x_{ij}\sigma^2$, as per the expectation and variance of the sum of random variables. This is also why we choose a Normal distribution for $y_{ij}$, because its sum also follows a Normal distribution, which is a desirable property for the model.

Dependence on the average ticket value for customers in the same cluster can now be incorporated by letting $\eta_i$ follow a common distribution for each cluster,

$$\eta_i \mid m_1, \dots, m_K, \gamma, S_i \sim \text{logNormal}\left(\log\left(m_{S_i}\right), \gamma\right), \tag{3}$$

where $m_c$ becomes the median of the purchase value for customers in cluster $c$ and $\gamma$ the standard deviation of the $\log(\eta_i)$. Note that the distributions for $y_{ij}$ and $x_{ij}$ are related by the parameter $S_i$.

The subsequent value expected for $\eta_i$ will be used as an estimate of the average ticket value when calculating the CLV. This estimate will shrink towards the cluster average, with the effect being more prevalent for customers who make fewer purchases.

This model formulation accommodates the large number of zeros in the data. When customer $i$ does not make any purchases in month $j$ ($x_{ij} = 0$), the variable $y_{ij}$ follows a Normal distribution with a mean and standard deviation equal to zero (in fact, a degenerated variable that takes the value 0). If the customer makes $x_{ij}$ purchases, it follows a Normal distribution with a mean of $x_{ij}\eta_i$ and standard deviation of $\sqrt{x_{ij}}\sigma$. This parameterisation solves the problem of months with no customer activity.

In general, it is preferable to work with relatively few clusters for greater cluster stability and business applicability. The meaning of clusters can be achieved by imposing some constraints on the cluster model parameters. The number of clusters and their meaning should be set using business knowledge as the threshold between low and high purchase frequencies and value changes for different companies.

For example, in the case study analysed in Section 3, the company we worked with considered it convenient to set the number of clusters at five and with the following definitions:

- Cluster 1: low frequency and low ticket value,
- Cluster 2: high frequency and low ticket value,
- Cluster 3: low frequency and medium ticket value,
- Cluster 4: high frequency and medium ticket value and
- Cluster 5: low frequency and high ticket value.

Therefore, we set the values of the parameters according to the following restrictions: $\theta_1 = \theta_3 = \theta_5 < \theta_2 = \theta_4$ and $m_1 = m_2 < m_3 = m_4 < m_5$. Besides taking advantage of company experience, this approach also avoids identifiability and estimation problems usually associated with finite mixture models.

Finally, we had to define the parameters' prior distributions. Setting five clusters, we chose a prior distribution that assumes the weights $(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$ are Dirichlet (1,1,1,1,1), which corresponds to a uniform distribution on the simplex space. The inverse of $\sigma^2$ and $\gamma^2$ are real nonnegative values and, by default, they are typically assumed to be Gamma(0.01, 0.01), which is the standard noninformative choice for dispersion parameters. As the posterior distribution cannot be computed analytically, we used a Markov chain Monte Carlo implementation to update and simulate from the model.

## 2.2 | Pareto/NBD model with clusters

One of the most appealing features of the Pareto/NBD model is the small amount of information required to estimate customer lifetime and probability of churn. For each customer $i$ in a specified observation period $T_i$, it only needs the number of purchases made during the observation period, $x_i$, and the length of time between their first and last transactions, $t_i$. In marketing terms, $x_i$ would be the customer's frequency and $T_i - t_i$ is their recency. The customer tracking time $T_i$ starts from his/her first purchase within the period of observation and ends at the same time for all customers.

Except for the cluster grouping, the model takes the same assumptions as those stated in Schmittlein et al.[10]:

(i) Customer lifetime is characterised by two stages: each customer is active for a certain period and then they become permanently inactive at a given time.

(ii) Let the unknown time at which customer $i$ becomes inactive be denoted by $\tau_i$. If customer $i$ is still active at $Ti$ (so $\tau_i > T_i$), then the number of purchases per time period (in our case a time period is 1 day) follows a Poisson distribution with expected value $\lambda_i$, hence:

$$x_i \mid \lambda_i, \tau_i > T_i \sim \text{Poisson}\left(\lambda_i T_i\right). \tag{4}$$

(iii) Heterogeneity in transaction rates across customers from the same cluster $c$ follows a Gamma distribution with shape parameter $r_c$ and inverse scale parameter $\alpha_c$:

$$\lambda_i \mid r_c, \alpha_c \sim \text{Gamma}\left(r_c, \alpha_c\right), \tag{5}$$

leading to a different negative binomial distribution (Poisson–Gamma mixture) for the number of purchases for each cluster.

(iv) Customer's unobserved "lifetime" $\tau_i$, after which they become permanently inactive, is exponentially distributed with dropout rate $\mu_i$:

$$\tau_i \mid \mu_i \sim \text{exponential}\left(\mu_i\right). \tag{6}$$

(v) The heterogeneity in dropout rates across customers from the same cluster $c$ follows a Gamma distribution with shape parameter $s_c$ and inverse scale parameter $\beta_c$:

$$\mu_i \mid \beta_c, s_c \sim \text{Gamma}\left(s_c, \beta_c\right), \tag{7}$$

leading to a different Pareto distribution (Exponential–Gamma mixture) for the dropout rate for each cluster.

(vi) The transaction rate $\lambda_i$ and the dropout rate $\mu_i$ vary independently across customers.

The prior chosen for each of the model hyperparameters, $r_c$, $\alpha_c$, $\beta_c$, and $s_c$ for $c = 1 \ldots K$, was Gamma(0.01, 0.01), which provides a very diffuse prior.

## 2.3 | Outputs of the models

From the initial cluster model, we obtained the average purchase value per customer. From the Pareto/NBD model, we obtained the probability of a customer still being active at the end of the observation period and the discounted expected transactions (DET) per customer. The DET is the expected discounted number of transactions for the customer's remaining lifespan at the end of the observation period given their past purchasing activity with the company. These values can be used to calculate the CLV, which, along with the probability of being active, will help target customers for future marketing actions.

As explained by Schmittlein et al.,[10] the probability of customer $i$ being active at the end of the observation period $T_i$ is:

$$P\left(\tau_i > T_i | \lambda_i, \mu_i, x_i, t_{i,} T_i\right) = \frac{1}{1 + \left(\frac{\mu_i}{\lambda_i + \mu_i}\right)\left[e^{(\lambda_i + \mu_i)(T_i - t_i)} - 1\right]}. \tag{8}$$

The conditioned DET can be found in Fader et al.[5] and for each customer $i$ is computed as:

$$DET_i = \frac{\lambda_i}{\mu_i + \delta} \cdot \frac{1}{1 + \left(\frac{\mu_i}{\lambda_i + \mu_i}\right)\left[e^{(\lambda_i + \mu_i)(T_i - t_i)} - 1\right]}, \tag{9}$$

where $\delta$ is the continuously compounded discount rate used to account for the effect of time on future revenues, in turn, this is a function of the yearly discount rate $d$, assumed to be 2% in our case study. In our study, the data were recorded on a daily basis so the continuous discount rate has to be factored by the number of days in a year and $\delta$ is computed as $\log(1 + d)/365$.

In its basic form, the CLV estimates the expected current value of a customer's future profit stream with the company.[5,7] The net margin term involved in the CLV calculation may be unavailable because it is unknown or changes over time or for products and/or customers. Therefore, since our data do not reflect customer net margin but rather sales revenue, we replaced net margin with sales value. To avoid confusion, we are going to call this magnitude customer lifetime value of sales (CLVS). This paper will estimate the CLVS per customer as:

$$\mathrm{CLVS}_i = \eta_i \times DET_i, \tag{10}$$

where $\eta_i$ is the average ticket value estimated from the initial cluster model.

As the CLVS estimates the present value of all the sales the customer will generate in the future, we used it as our ranking score when classifying clients in terms of their monetary relevance. The probability of being active just after the end of the observation period can be considered as a measure of the possibility of losing a client. We used this information as our ranking score for the likelihood of losing a customer. With these two scores, a company can decide how to prioritise resources based on the sales value and the risk of losing each customer.

In Section 3.3 we illustrate the classification of these scores in a two-way table based on managerial criteria and show how it can help companies rank customers for future marketing activities.

## 3 | CASE STUDY

### 3.1 | Description of the data

Our working database consists of 25,594 sales made by a beauty products wholesaler over a 24-month period, from June 2016 to May 2018, to its 1467 customers (beauty salons) who completed at least two purchases during the observation period. The tracking time for each customer starts from their first purchase within the period.

Figure 1 shows all customers according to their mean number of monthly purchases and average ticket value. The vertical axis shows the buying frequency in monthly purchases, which ranges from 0.08 to 4, that is, it includes customers who make purchases once a year, while others buy things weekly. The median and mean purchase frequency are 0.67 and 0.85, respectively. The first and third quartiles are 0.43 and 1.04, corresponding approximately to one purchase every 2 or 1 month, respectively. The horizontal axis shows the average ticket value for each customer. The median and mean are €706 and €1582, respectively, and the first and third quartiles are €446 and €1357.

Considering the large variability of customers with respect to the two dimensions shown in Figure 1, our suggestion of performing clusters on customers before implementing the Pareto/NBD model seems sensible.

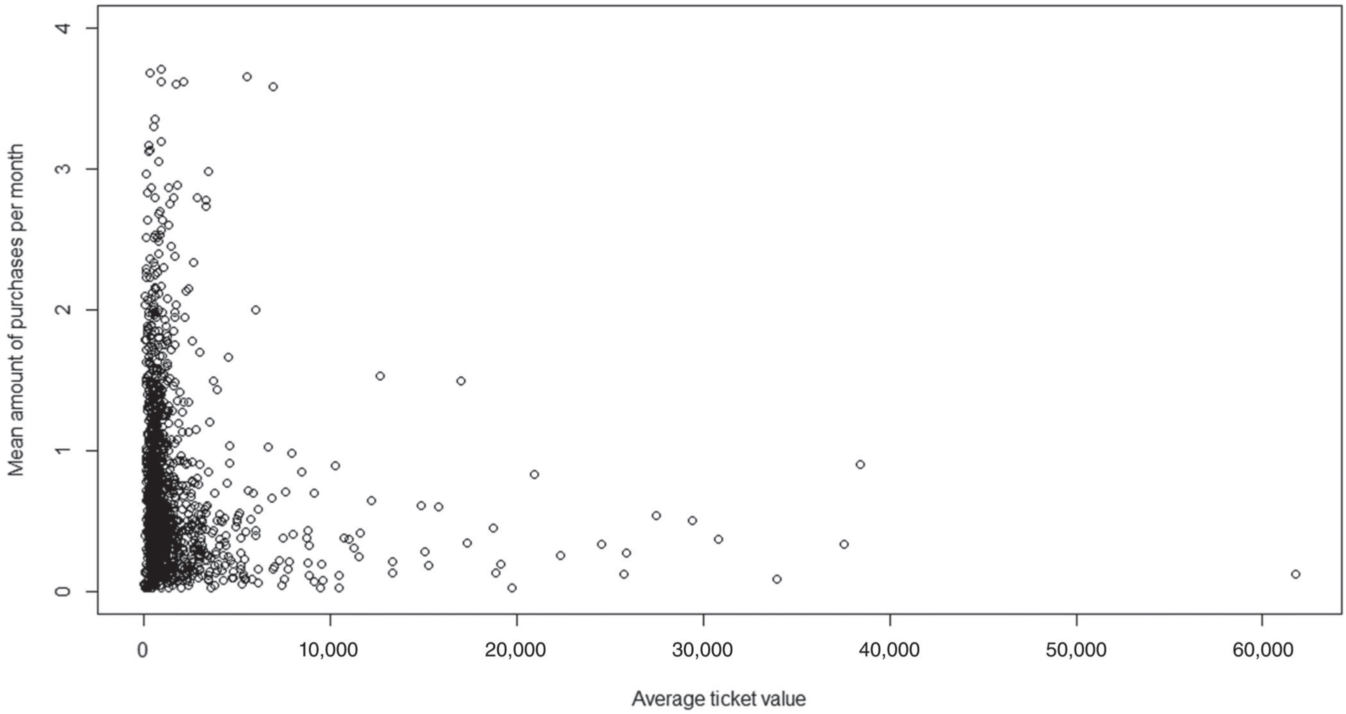


**FIGURE 1** Average ticket value, in euros, vs. mean number of monthly purchases. Each point represents a customer

## 3.2 | Cluster model results

The purpose of the cluster model was to assign customers to interpretable clusters with homogeneous purchasing features. As stated in Section 2.1, the number of clusters and their meaning has to be set using company-specific knowledge, as the meaning of low/high frequency and value depends on each business. In our case, the company considered that it makes sense to use five clusters with the following descriptions: Cluster 1 – "low frequency and low ticket value", Cluster 2 – "high frequency and low ticket value", Cluster 3 – "low frequency and medium ticket value," Cluster 4 – "high frequency and medium ticket value" and Cluster 5 – "low frequency and high ticket value." The parameters for the clusters were chosen so they would be separable and in line with the company's knowledge of its customers. When applying the analysis to other businesses and sectors, it would have to be adapted to the nature of their customers. In our case, the parameters chosen were $\theta_1 = \theta_3 = \theta_5 = 0.5$ purchases/month, $\theta_2 = \theta_4 = 1.5$ purchases/month, $m_1 = m_2 =$ €500, $m_3 = m_4 =$ €1500 and $m_5 =$ €10,000. As there were no customers with a high purchase frequency and high ticket value (see Figure 1), we did not create a cluster for this type of behaviour.

The main result derived from the cluster model was the allocation of each customer to a single cluster. The mode of the posterior distribution of $S_i$ indicated the most likely cluster for customer $i$. This value was used to allocate a customer to one cluster. The cluster model also provided the average customer purchase value ($\eta_i$) needed for the CLVS calculation.

The cluster model results are shown in Table 1. Over half of customers, 60.6%, belong to cluster 1 for low frequency and low ticket value. Clusters 2, 3, and 4 represent 23.2%, 8.9%, and 2.9% of customers, respectively. Finally, 4.4% of customers belong to cluster 5 with a low frequency and high ticket value. It is important to note that while clusters 4 and 5 represent a small percentage of customers, they represent a significant proportion of the company revenue.

## 3.3 | Pareto/NBD model results

After clustering the customer sample, we fitted the Pareto/NBD model independently to each cluster to obtain more accurate estimates. To validate the assumption that prior clustering improves the estimates from the Pareto/NBD model,

**TABLE 1** Cluster model summary results: number of customers (and % over total) and cluster total customer lifetime value of sales (CLVS) (and % over total)

| Cluster | Customers | CLVS (euros) |
|---|---|---|
| 1 | 889 (60.6%) | €21,366,278 (22.0%) |
| 2 | 340 (23.2%) | €27,808,452 (28.7%) |
| 3 | 131 (8.9%) | €8,556,029 (8.8%) |
| 4 | 43 (2.9%) | €16,804,730 (17.3%) |
| 5 | 64 (4.4%) | €22,514,042 (23.2%) |

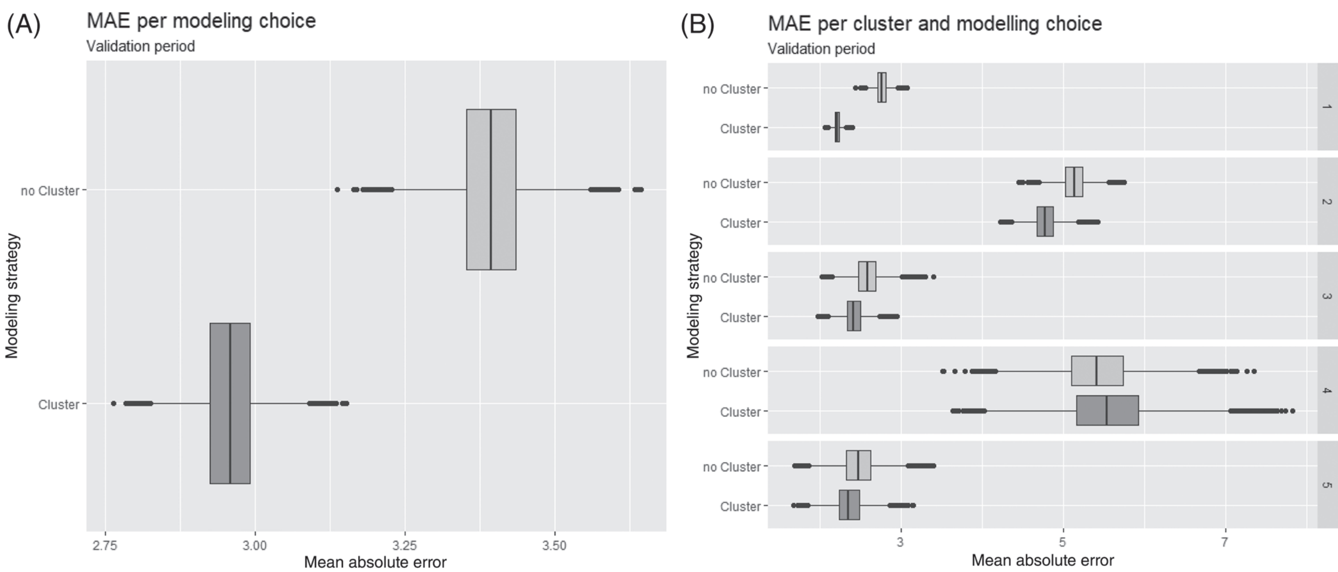


**FIGURE 2** (A) Sample box-plots for 20,000 observations of the mean absolute error (MAE) posterior distribution from the Pareto/NBD model with and without clusters. (B) MAE stratified by clusters

we compared the results from the two approaches: (1) applying the Pareto/NBD model to the total set of customers[10] or (2) applying to each cluster of customers after clustering. To this end, we split the dataset into two periods, the training period from June 2016 to May 2017, and the test period, from June 2017 to May 2018. Only customers with at least one purchase in the training period were considered. This included 1349 customers, or 92% of the total.

We compared the prediction accuracy of the two approaches in terms of the mean absolute error (MAE) on the predicted number of transactions per customer during the test period. The MAE posterior distribution in Figure 2A indicates a better fit for the clustered Pareto/NBD model than for the unclustered one. The average MAE in the number transactions completed in the test period (1 year) when using the cluster-based Pareto/NBD model was 2.93, which is lower than the 3.39 obtained when using the unclustered model. Contrary to non-Bayesian model comparisons, based on goodness-of-fit tests, the posterior distribution in Figure 2 captures the degree of the uncertainty embedded in the final conclusion.

Models were also compared by stratifying per cluster. Figure 2B shows that the goodness-of-fit was better for the cluster-based Pareto/NBD model than the unclustered one, except for in cluster 4, where both are very similar. Also, as expected, the prediction error was larger in clusters 2 and 4, where customer purchase frequency was higher.

After demonstrating the overall better fit of the cluster-based Pareto/NBD model, we can use the whole set of customers 2-year data to predict their CLVS. Therefore, the following results correspond to the complete dataset of 1467 customers.

To illustrate the results obtained from the models, Table 2 shows the relevant information for two customers randomly selected from each cluster. For instance, the first customer in cluster 1 has a low probability of being active as a long period has passed between their last purchase and the end of the observation period. This is very different from the other customer taken from this cluster who had a high probability of being active as their last purchase was closer to the end of

**TABLE 2** Number of transactions ($x$) during the observation period, recency ($T - t_x$) and observation time ($T$) in days. Point estimation of the probability of being active at the end of the observation period ($P$) and the 95% credible interval (95% CI) and customer lifetime value of sales (CLVS) for two randomly selected customers from each cluster

| Customer ID. | Cluster | $x$ | $T - t$ | $T$ | $P$ | 95% CI | CLVS |
|---|---|---|---|---|---|---|---|
| 10535 | 1 | 4 | 385 | 707 | 0.22 | (0.01–0.67) | €3713 |
| 10117 | 1 | 9 | 119 | 693 | 0.85 | (0.68–0.95) | €17,449 |
| 11628 | 2 | 23 | 119 | 714 | 0.42 | (0.08–0.81) | €19,345 |
| 11963 | 2 | 33 | 7 | 700 | 0.99 | (0.97–1.00) | €63,272 |
| 20051 | 3 | 1 | 318 | 515 | 0.45 | (0.04–0.87) | €14,133 |
| 10035 | 3 | 11 | 51 | 665 | 0.92 | (0.79–0.98) | €124,375 |
| 14799 | 4 | 24 | 72 | 700 | 0.78 | (0.41–0.96) | €319,343 |
| 15000 | 4 | 55 | 7 | 567 | 0.99 | (0.96–1.00) | €297,470 |
| 30395 | 5 | 10 | 105 | 707 | 0.81 | (0.51–0.96) | 151,830€ |
| 15196 | 5 | 8 | 14 | 281 | 0.95 | (0.86–0.99) | 1,690,759€ |

the observation period. The effect of recency on the probability of being active differs between clusters as they are different in terms of purchasing frequencies.

### 3.4 | Business deployment

As mentioned in Section 1, customer potential profitability and the probability of becoming inactive are basic metrics for prioritising marketing actions.

We used the CLVS to rank customers in terms of their monetary value relevance. We took the probability of being active at the end of the observation period as our ranking score for the likelihood of losing a customer. Companies can use these two scores to prioritise their decisions customer by customer. So, it seems reasonable to combine both factors in the prioritisation procedure, that is, the CLVS and probability of being active.

In our case study, the company's main objective was customer prioritisation for marketing retention actions. There may be very little value in focusing on a customer with a high risk of leaving but low CLVS and, similarly, it might be unwise to spend extra resources on customers with a significant CLVS but low risk of prompt inactivity.

We propose the following criteria for flagging customers according to their purchasing behaviour. First, classify customers into three groups: high, medium and low value. In our case, we defined the high-value group as the customers who accounted for the first 60% of the total CLVS, the medium-value group included the next 30% and the low-value group the final 10%. With these limits, 258 customers (17.6% of the total) were classified as high-value, 663 (45.2%) were considered medium-value and the remaining 546 (37.2%) fell into the low-value group.

Second, define the criteria for flagging each customer in every group depending on the two aforementioned scores. Our proposal, developed in collaboration with business experts, is outlined in Table 3. Customers flagged as green require no action. Yellow customers are candidates for either close monitoring or making a move. Finally, red customers require immediate action. The alarm levels are set more severely for high-value customers with the same probability of being inactive as lower value customers. This is because if they became inactive, they would have a higher negative impact on company revenue. The criteria used in Table 3 were chosen according to the company's objectives but they can be modified to suit other companies' goals.

After applying these criteria to our customer sample, Table 4 shows how many customers were flagged differently depending on their CLVS classification (high, medium or low). According to these guidelines, the company has 74 customers (5.0%) in need of immediate action, 93 (6.3%) who should be monitored closely and 1300 (88.6%) requiring no action at the moment.

The company outlook seems quite good as the most important customers in terms of CLVS are mainly green. The percentage of the company's future sales (CLVS) with red and yellow flags are just 3.0% and 6.2%, respectively. Thus, 90.8% of the company's total CLVS of €97 million is, at this moment, relatively safe.

TABLE 3 Criteria for flagging different types of customers based on their value and probability of being active

| | | Probability | Active | | |
|---|---|---|---|---|---|
| | <0.5 | 0.5–0.7 | 0.7–0.8 | 0.8–0.9 | 0.9–1 |
| High value | Red | Red | Red | Yellow | Green |
| Medium value | Red | Red | Yellow | Green | Green |
| Low value | Red | Yellow | Green | Green | Green |

TABLE 4 Number of customers in each flagged group per value category according to the criteria in Table 3

| | | Flag | | |
|---|---|---|---|---|
| | Green | Yellow | Red | Total |
| High value | 228 | 22 | 8 | 258 |
| Medium value | 622 | 30 | 11 | 663 |
| Low value | 450 | 41 | 55 | 546 |
| Total | 1300 | 93 | 74 | 1467 |

The percentage of CLVS flagged as green could be taken as a key performance indicator of the effectiveness of marketing retention actions. Other indicators, such as the percentage of customers with a significant risk (red and/or yellow), could be monitored to assess the outcomes of those actions. We believe the proposed flag system is quite intuitive, both when choosing which customers to prioritise and as a quick summary of the short-term business performance.

The categorisation in this study helps companies prioritise customers for marketing actions based on their goals. If a company wanted to identify inactive but profitable customers that should be reactivated, the selection could focus on yellow and red customers in the high-value group, and yellow customers in the medium one. If they wanted to remove inactive and low-value customers from their customer base, they could remove red customers from the low-value group. If they wanted to identify the best customers for a loyalty program offer, they could select green customers from the high-value set.

To assess the impact of changes in cluster parameters on the final results, we performed a sensitivity analysis using different values of $m$ and $\theta$. Changing these parameters by 20% of their present value only changed, in the worst case, the customer flag assignment by 1.7%. This ensures that reasonable fluctuations in the choice of cluster parameters does not significantly change the final conclusions.

## 4 | DISCUSSION

We have proposed a method for sorting customers according to their value and risk of churn in order to help prioritise marketing actions. The procedure involves, first of all, using a cluster model to segment customers based on their purchase frequency and ticket value. Second, we estimated the probability of being active and the CLVS for each customer in the same cluster using the Pareto/NBD model. Third, we classified customers into three groups (green, yellow and red) based on their probability of becoming inactive and their CLVS to facilitate the company's marketing decisions.

The initial clustering step improved the results obtained from a stand-alone Pareto/NBD model. It also allowed to estimate the customer average ticket value, which is necessary for the CLVS calculation. This variable is not included in the classic Pareto/NBD model regardless of its relevance when characterising customer purchasing behaviour. Segmenting customers into more homogeneous groups gives better estimates of model parameters. In fact, deterministic clustering is sometimes applied to customer cohorts based on the time of their first purchase, purchasing channel or type of business (see, for example, Reference 11; or Reference 18). It should also be noted that the customer average ticket value must be estimated using another model if the clustering step is not carried out first.[4,11,16]

Once customers have been prioritised for potential marketing actions, the next step is to perform actions on those customers where the campaign would be more profitable and then assess their impact. Usually in a contractual setting, a campaign's effectiveness is measured by means of A/B tests. Ascarza[23] and Lemmens and Gupta[24] have published relevant studies on this topic. However, further research is needed for noncontractual settings, especially in the case of low purchase frequencies, as in our study. Recent publications highlight the importance of the interpurchase time[25,26] and, in particular, how monitoring interpurchase time could help identify the best time to carry out retention actions for each specific customer.[27,28] Considering the type of purchasing time pattern in conjunction with the purchase frequency and value could improve customer clustering, CLVS predictions and other estimates, and ultimately improve resource allocation decisions. This perspective should be considered in future research.

## ACKNOWLEDGMENTS

We would like to thank Xavier Tort-Martorell from the Statistics Department at UPC for his valuable comments. We would also like to thank the journal's reviewers for their comments and recommendations.

## DATA AVAILABILITY STATEMENT

The data is not openly shared as it reflects the detailed revenue from an existing company. However, the authors are willing to share it upon request as long as the requestor accepts not to openly re-share it with others.

## ORCID

*Ignasi Puig-de-Dou* https://orcid.org/0000-0003-1747-8564
*Daniel González Ibáñez* https://orcid.org/0000-0002-9553-0252
*Xavier Puig* https://orcid.org/0000-0001-6525-0498

## REFERENCES

1. Artinger F, Petersen M, Gigerenzer G, Weibler J. Heuristics as adaptive decision strategies in management. *J Organ Behav*. 2014;36:33-52.
2. Wübben M, von Wangenheim F. Instant customer base analysis: managerial heuristics often εget it rightε. *J Market*. 2008;72:82-93.
3. Lehmann DR, McAlister L, Staelin R. Sophistication in research in marketing. *J Market*. 2011;75:155-165.
4. Borle S, Jain DC, Singh SS. Customer lifetime value measurement. *Manag Sci*. 2008;54:100-112.
5. Fader PS, Hardie BG, Lee KL. RFM and CLV: using iso-value curves for customer base analysis. *J Market Res*. 2005;42:415-430.
6. Fader PS, Hardie BGS. How to project customer retention. *J Interact Market*. 2007;21:76-90.
7. Gupta S, Lehmann DR. Customers as assets. *J Interact Market*. 2003;17:9-24.
8. Malthouse EC, Blattberg RC. Can we predict customer lifetime value? *J Interact Market*. 2005;19:2-16.
9. Glady N, Baesens B, Croux C. A modified Pareto/NBD approach for predicting customer lifetime value. *Exp Syst Appl*. 2009;36:2062-2071.
10. Schmittlein DC, Morrison DG, Colombo R. Counting your customers: who are they and what they will do next? *Manag Sci*. 1987;33:1-24.
11. Schmittlein DC, Peterson RA. Customer base analysis: an industrial purchase process application. *Market Sci*. 1994;13:41-67.
12. Reinartz WJ, Kumar V. On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing. *J Market*. 2000;64:17-35.
13. Abe M. εCounting your customersε one by one: a hierarchical Bayes extension to the Pareto/NBD model. *Market Sci*. 2009;28:541-553.
14. Ma SH, Liu JL. The MCMC approach for solving the Pareto/NBD model and possible extensions. Proceedings of the 3rd International Conference on Natural Computation (ICNC 2007), Vol. 2, 2007:505-212; IEEE, Washington, DC.
15. Mzoughia MB, Borle S, Limam M. A MCMC approach for modeling customer lifetime behavior using the COM-Poisson distribution. *Appl Stoch Models Bus Ind*. 2017;34:113-127.
16. Colombo R, Jiang W. A stochastic RFM model. *J Interact Market*. 1999;13:2-12.
17. Fader PS, Hardie BGS, Kinshuk J. Estimating CLV using aggregated data: the Tuscan lifestyles case revisited. *J Interact Market*. 2007;21:55-71.
18. Blattberg R, Getz G, Thomas J. *Customer Equity*. Harvard Business School Press; 2001.
19. Rossi PE, Allenby GM. Bayesian statistics and marketing. *Market Sci*. 2003;22:304-328.
20. Jen L, Chou C-H, Allenby GM. A Bayesian approach to modeling purchase frequency. *Market Lett*. 2003;14:5-20.
21. Baesens B, Viaene S, Van den Poel D, Vnathienen J, Dedene F. Bayesian neural network learning for repeat purchase modelling in direct marketing. *Eur J Operat Res*. 2002;138:191-211.
22. Borle S, Singh SS, Jain DC, et al. Analyzing recurrent customer purchases and unobserved defections: a Bayesian data augmentation scheme. *Cust Need Solut*. 2016;3:11-28.
23. Ascarza E. Retention futility: targeting high-risk customers might be ineffective. *J Market Res*. 2018;55:80-98.
24. Lemmens A, Gupta S. Managing churn to maximize profit. *Marketing Sci*. 2020;39:849-1031.
25. Platzer M, Reutterer T. Ticking away the moments: timing regularity helps to better predict customer activity. *Market Sci*. 2016;35:779-799.

26. Zhang Y, Bradlow ET, Small DS. Predicting customer value using clumpiness: from RFM to RFMC. *Market Sci*. 2015;34:195-208.
27. Dew R, Ansari A. Bayesian nonparametric customer base analysis with model-based visualizations. *Market Sci*. 2018;37(2):216-235.
28. Holtrop N, Wieringa J. Timing customer reactivation interventions; June 18, 2020: https://ssrn.com/abstract=3443422 or 10.2139/ssrn.3443422



**How to cite this article:** Puig-de-Dou I, González Ibáñez D, Puig X. Ranking customers for marketing actions with a two-stage Bayesian cluster and Pareto/NBD models. *Appl Stochastic Models Bus Ind*. 2022;1-11. doi: 10.1002/asmb.2677