

Improving Direct Marketing Activities Effectiveness Using Analytical Models: RFM vs. Logit Model on a Casino Case

TJAŠA TABAJ PUŠNAR

Slovenia

tjasa.tabaj@gmail.com

DANIJEL BRATINA

University of Primorska, Slovenia

danijel.bratina@fm-kp.si

This research deals with the development and implementation of a large-scale analytics framework for improving segmentation and targeting of a service firm's direct marketing activities. The aim of the framework is to create a direct marketing response model using customers' demographics and other behavioural data (such as past response to direct marketing activities) from a casino (gambling industry). Prior to this research the company was using Recency-Frequency-Monetary (thereafter referred as RFM model). The statistical model used in our research, based on logit regression, significantly improves the accuracy of direct marketing activities as well as provides insight on relevant customers' characteristics that affect choice. We believe the results are a show-case of combining large, disaggregate, individual-level datasets with marketing analytics solutions for improving response to the marketing-communication mix. As per our knowledge in the time of writing this paper no such complete set of demographical and behavioural determinants have been used in direct marketing effectiveness analysis in the casino industry. Findings in this paper can be used by the company to considerably improve fine-tuning of target segments in their direct marketing activities, other industries (currently using RFM for direct marketing activities target group selection) can also benefit.

Key words: direct marketing effectiveness, RFM, logit modeling

<https://doi.org/10.26493/1854-4231.13.323-334>

Introduction and Theoretical Framework

Direct marketing activities, when not correctly targeted, can have higher than necessary cost turning an otherwise profitable promotion activity into a loss. A variety of direct marketing optimization

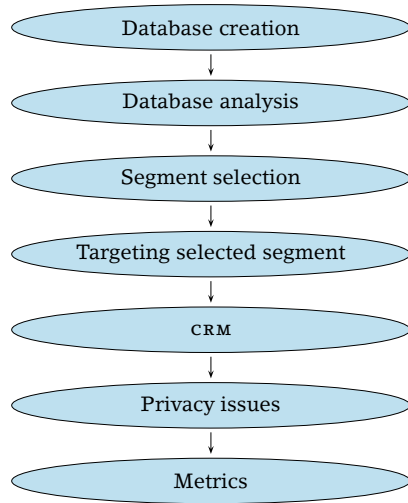


FIGURE 1
A Model of Customer
Relationship Management

models have been developed in the past helping target profitable current and potential customers (Bult and Wansbeek 1995; Shepard 1999; Malthouse 2001). With recent greater emphasis on importance of customer relationship management from academia and practitioners and increased availability of large quantities of data (due to gains informatisation technology and associated statistical methods) the development of such tools has been accelerated producing a wide set of useful tools. A theoretical model of CRM has been introduced by Winer (2001) and is represented in figure 1.

Winer argues that ideal datasets should include customer's purchase history, sales person contact with the single customer, customer demographics and his responses to different marketing activities (Winer 2001, 92). The process of CRM can be monetarily expressed with Customer Lifetime Value (thereafter as CLV) which calculates the forecasted discounted net cash-flows a customer creates to the company and where the sum of all customers' LVs is defined as Customer Equity (Wierenga 2008, 292). It is in this (long-term) perspective that a company should view its cost associated with acquiring new customers (and not through a single transaction). Most companies decide on running a CRM program based on the assumptions that a segment exists that is ready and willing to tighten its bonds with the company to establish a permanent relationship based on loyal purchases, that these customers are highly profitable and that a company can affect their loyalty through various approaches (Musek Lešnik 2008, 142). Such activities aim to increase consumers'

loyalty, increased frequency and monetary value of transactions as well as creating positive attitudes towards the company and/or its brands (Sharp and Sharp 1997, 474). Several strategical and tactical activities follow the concept of CRM. One can find related topics in every CRM literature.

Critics of loyalty programmes argue that their high cost seldom create increased revenue, there is hardly any empirical evidence on increased consumption (and thus frequency). Wide spread of such activities dilutes their effectiveness, failure to convert non-users to users or increase loyalty within current customer base (Leenheer et al. 2007, 31).

Further increase of computing power and development of statistical tools have contributed to new (often real-time) data gathering called data mining, that involves searching for patterns in a large set of incomplete, unfiltered data. Data mining has more than once contributed to the development of unsought relationships resulting in breakthrough findings (Wang in Sing'oei 2013). Direct marketing analytics support has also benefited from data mining techniques (i.e. mass parallel processing or symmetric multi-processing; Churchill and Iacobucci 2002).

Overview of Existing Models for Measuring Direct Marketing Effectiveness

As mentioned earlier the most widely used model for optimal direct mail targeting is the RFM and its various modifications (Malt-house 2001). Bult and Wansbeek (1995) specify a profit function. By setting marginal costs equal to marginal revenues they determine which household should be receiving the offering. Improved RFM models namely AID (Automatic Interaction Detection) and CH-AID (Chi Square AID) (Kass 1976) have been proposed to cope with limited variables selection used by RFM (only Recency, Frequency and Monetary) by adding demographics. All these early models have arbitrarily determined factors. In late 80ties regression models with statistical started to emerge (i.e. general customer purchase model; Bauer 1988). A popular model from Bansalben (Nash 1992) – gains chart analysis is based on a three-stage process. The initial stage analyses the test mailing response from target population by means of regression analysis to determine how household characteristics affect response by giving each household a probability that in future they will respond to marketing activities. The next step orders the population by the factors derived in the 1st stage. Last stage consists in dividing households with similar index value into groups of equal

sizes and calculating the average response rate of the group. A cut-off point is arbitrarily determined and household above this point targeted.

Due to its ease of use, the RFM model is still being most widely used (Wollen 2017). Wheaton Group (see <https://www.wheatongroup.com>) argues that with increasing sizes of databases, RFM becomes inefficient and requires constant re-segmenting, especially in situations where a large heterogeneous group of products with different profitability are being offered. Our research shows how a casino (with a very large database of direct marketing events) can improve their targeting accuracy using multiple logit regression. A set of statistically relevant factors is also given (not possible with RFM). Profitability is not calculated in this paper but can easily be added to the model. With statistical tools available management can set-up a predictive tool in real time.

Research Set-up and Dataset

The objective of direct marketing models is to select a subset of the population that is going to receive a marketing communication with the aim of maximising client response and minimising cost. Thus, a direct marketing model must be able to predict the client response to the received offer based on the data that the company has gathered about the client (Taghva, Bamakan, and Toufani 2011). One of the most widely used segmentation methods for direct marketing support is the RFM analysis for clustering (Aggelis and Dimistris 2005). RFM is a classifying mechanism based on three sets of factors, namely time since last purchase, frequency of purchases and average value of purchase. Every customer is given a score upon which his participation in the next direct marketing campaign is based (Hosseini, Maleki, and Taghva 2010). RFM assumes that higher ranking customers are more responsive to direct marketing activities and that there is increased probability they would buy (Yeh, Yang, and Ting 2009) There are several versions of setting the three grouping factors levels. The most commonly used is a 5 levels/dimension, resulting in 125 segments (Olson and Chae 2012).

The expected contribution of a customer is $\mu_i\pi_i$, where μ_i represents the expected sales from customer i and π_i is the probability that the i customer will respond to the communication.

A customer is profitable when $\mu_i\pi_i > c_i$, where c_i is the projected cost for sending the message through to the customer i .

RFM has several flaws, mostly originating from its arbitrariness when selecting segment's criteria. It fails to predict the expected

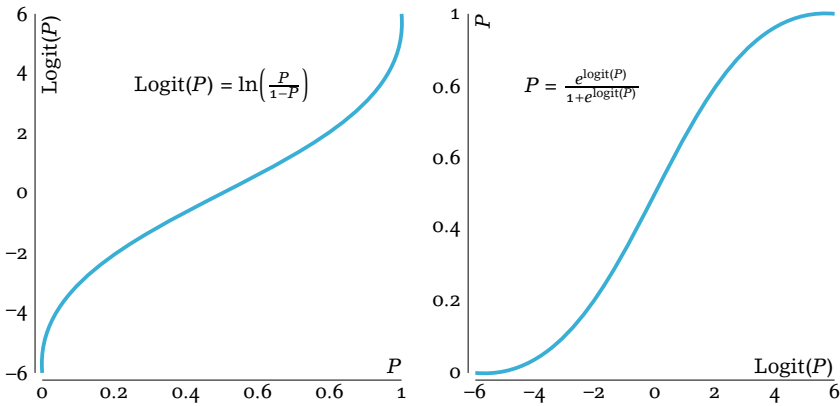


FIGURE 2 Logit Regression S-Shape Curve

number of buying customers in the future, it needs to be adjusted to specific industry needs and requires repeated trial-error fine tuning. To address some of the RFM disadvantages (mainly arbitrariness when selecting weights for the three criteria groups) we apply logistic regression on sample data from a service industry (gambling) and compare it with the current RFM model they are using to assess their direct marketing effectiveness.

Logit (or probit) regression analysis has been widely used in marketing research. Its main advantages are conceptual simplicity, it analytically defines weights (importance) for different criteria as opposed to RFM arbitrariness (Coussement, Harrigan, and Benoit 2015) and it is faster (Simonof 2016).

The logit model is a S-shaped success probability regression model with discrete dependent and various independent variables. It is modelled as:

$$\ln \frac{p}{1-p} = \sum_{i=0}^n \beta_1 X_i \tag{1}$$

where p represents the probability of success and $\sum_{i=0}^n \beta_1 X_i$ the vector of independent variables. Solving for p we get the S-shaped curve (see figure 2):

$$p = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \tag{2}$$

The regression is solved by using OLM or GLM. Regression coefficients are tested using Wald statistics, which has a chi-square distribution:

TABLE 1 Direct Marketing Activities

Month	2014		2015	
	(1)	(2)	(1)	(2)
January	22	SMS, e-mail	22	SMS, e-mail
February	30	mail	28	mail, SMS in e-mail
March	28	SMS, e-mail	23	SMS, e-mail
April			15	SMS, e-mail
May	27	SMS, e-mail	29	mail
June	29	mail	23	mail, SMS and e-mail
July	25	mail, SMS, e-mail	28	mail, SMS, e-mail
August			29	SMS, e-mail
September	25	SMS, e-mail	34	mail
October	28	mail	28	SMS, e-mail
Average	26.6		26.0	

NOTES Column headings are as follows: (1) response rate, (2) channel.

$$Wald = \left(\frac{\beta_i}{SE\beta_i} \right)^2. \quad (3)$$

By adjusting the used parameters in the set of independent variables used in the regression we can fine-tune our model and compare it to the currently used RFM techniques.

Data Set

Our data consists of a collection of 2 years direct marketing communications to 18.000 clients and including their responses. In marketing, such a large database of responses is rather exceptional than common. In our case it has been made possible by a law requiring compulsory registration of the customers at casino entrance and tracking of their activities (game play) in the casino.

We use a combination of demographical and behavioural variables as the independent set of variables. Response rate was roughly 25%. Table 1 shows types of communications per month and their response rates (there were no activities in November, December 2014 and September, December 2015).

The currently used RFM model classifies as potentially profitable (will probably engage in profitable activity for the casino after receiving the direct marketing communication) 25% of the customers base. Communication offers included a monetary value (to be spent in the casino), based on past expenditures and frequencies of visits. Some potentially profitable customers are filtered out due to incomplete data (missing address, phone number or e-mail). The

TABLE 2 Demographic Variables Descriptive Statistics

Factor		<i>N</i>	%
Gender	0 Men	10249	56.9
	1 Women	7751	43.1
Nationality	1 Slovenians	741	4.2
	2 Italians	14472	80.4
	3 Austrians	2108	11.7
	4 Chinese	566	3.1
	5 Other nationalities	113	0.6
Distance	1 Closer than 300 km	14077	78.2
	2 300 to 500 km	1873	10.4
	3 More than 500 km	2050	11.4

company distinguishes 3 segments of customers based on recency (undisclosed criteria):

- Regular customers,
- Revitalisation segment (customers that hasn't visited for a short period of time), and
- Lost segment.

The first two segments are further split into (based on frequency):

- Low frequency segment and
- High frequency.

The resulting segments are further split into 21 segments based on average game bet they did in the past (detailed data has not been disclosed by the company).

RFM does not measure which segments has responded better to the communication sent, although this could be done post fest, the company does not do this kind of analysis. The logit model focuses on demographical and behavioural (only used in RFM) characteristics of customers namely: gender, age, nationality, distance from casino (demographical variables), monetary value of offer, type of service, average spending on games, bet size, number of visits, channel of communication and percentage of past responses to direct communications (behavioural variables). The dependent variable is whether the offer sent was used or not. Tables 2 and 3 show descriptive statistics for independent variables.

The success rate was 29.7%. Beside conceptualizing the model, we have also tested some demographical/behavioural variables interdependency before constructing the logit model.

TABLE 3 Behavioural Variables Descriptive Statistics

Item	Average	SD	Min	Max
Age	54.80	15.50	18.00	98.00
No. visits	28.14	50.20	1.00	699.00
Total bet	109.00	404.00	0.60	28.15
Average bet	10.60	50.80	0.02	2314.00
Incentive sent through marketing	15.53	26.15	5.00	750.00
Percentage of responses	31.18	0.33		

NOTES $N = 18000$.

Initial findings show that women (playing automates) play statistically different games than men (playing at tables), there is a statistically significant correlation between age and number of visits per week ($r = 0.159$, sig. = 0.000) proving that elder people visit casinos more often. Not surprisingly, distance from casino affects the number of visits (group 1 – 0–300 km distance visited the casino 31 times in two years, group 2 – 300–500 km 17 and group 3 – above 500 km 18.5 times), although group 2 and 3 show adverse results. We attribute this to the fact that very distant high frequency/high monetary customers were transported free of charge to the casino. The average value played on machines was around 4 times smaller than at tables. While all nationalities show similar percentages of successful responses (from 30.4% for Slovenians to 34.7% for Austrians), Chinese exhibit only 12.1%. For the casino it would be worth researching further why. Response rates were 31% for e-mail and smss, while for classic mail they dropped to 15%. The latter is also the costliest way of distributing communication.

Results and Discussion

Our model was constructed using 2 steps. In the initial step we included all the demographical variables we had available. The 2nd model then only included statistically significant variables from the initial step. Nationality was recoded to (Italian/non-Italian) as this group represented 80% of the total sample. The initial model is shown in table 4.

Gender, nationality, game strength and past visits are not statistically significant. It is interesting that behavioural factors – number of visits and money spent do not affect direct marketing responses. This might suggest that there is a strong loyal customer base that ignores the incentives given by the casino.

This insight should be addressed by the casino when targeting customers to avoid giving incentives to those who would anyway be

TABLE 4 First Model Regression Coefficients

Item	B	SN	Wald	Sig.	Exp(B)
Age	-0.005	0.002	7.026	0.008	0.995
Distance	-0.099	0.044	4.987	0.026	0.905
Game type	0.173	0.074	5.487	0.019	1.189
Average bet	-0.002	0.001	10.190	0.001	0.998
Incentive	0.007	0.001	34.666	0.000	1.007
Mail	0.179	0.048	13.686	0.000	1.196
Past response rate	6.015	0.100	3589.036	0.000	409.691
Gender	0.053	0.049	1.132	0.287	1.054
Nationality (Italian)	0.086	0.063	1.906	0.167	1.090
Game strength	0.000	0.000	1.009	0.315	1.000
No. of visits	0.000	0.001	0.099	0.753	1.000
Constant	-3.230	0.124	675.972	0.000	0.040

TABLE 5 Regression Coefficients of the Final Logit Model

Item	B	SN	Wald	Sig.	Exp(B)
Age	-0.004	0.002	5.756	0.016	0.996
Distance	-0.092	0.044	4.373	0.037	0.912
Game type	0.189	0.071	7.130	0.008	1.208
Average bet	-0.003	0.001	16.996	0.000	0.997
Incentive	0.006	0.001	37.115	0.000	1.006
Channel	0.176	0.048	13.426	0.000	1.193
Past response rate	6.022	0.090	4456.921	0.000	412.204
Constant	-3.175	0.119	710.613	0.000	0.042

coming to the service. The final model is thus shown in table 5 or in equation script of type:

$$\text{Logit}(p) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_kx_k = X\beta, \tag{4}$$

as:

$$\begin{aligned} \text{Logit}(p) = & -3.175 + 6.022 \times \textit{past response rate} + 0.176 \times \textit{channel} \\ & + 0.006 \times \textit{incentive} - 0.003 \times \textit{average bet} \\ & + 0.189 \times \textit{game type}, \end{aligned}$$

where $\text{Logit}(p) = p/(1 + p)$ and p is the probability of successful communication or used coupon. We could also add distance and age to the final model that combined with these regression coefficients proved to be statistically significant but small (0.092 for distance and 0.004 for age).

The major factor affecting future behaviour is Past response behaviour, followed by game type and Channel used. Based on this research one could simplify the model to only past response rate as other factors do not contribute much to the success of direct marketing campaign.

Testing the Logit Model

The model was tested with Omnibus test for validity and Cox & Snells *R*-square as well as Nagelkerke *R* to determine variance explained. Significance is 0.000, which means that at the negligible risk we may reject the hypothesis, that the zero model is better than the last produced. We accept the hypothesis that the elaborated model is better. With the final model we can explain between 41.4 and 58.9% of the variance.

The model has correctly predicted 84.5% percent of responses. The model is more accurate for predictions of customers who won't use the offer. It correctly predicts 91.9% customers. But it's worse in predicting the announcements of customers who will use the offer, the percentage of correctly predicted is 66.9.

The equation that we got from modelling sample data can be applied to each customer in the database. We can calculate the probability of response, which is a combination of predictive variables: age, consumer distance from the casino, type of game, average bet, free play offer amount, direct marketing channel – mail and prior response to free play offer.

Limitations and Further Research

This research has been done using the available determinants as regressors, which is not the best practice. For a better insight of determinants affecting consumer behaviour, a qualitative research (focus groups, pilot research, more testing on smaller samples) as a priori would suggest a better set of determinants. The study is limited to a single casino and can thus be biased by specific properties of their customer base, meaning generalization is not straightforward. Due to data privacy RFM model couldn't have been analysed (i.e. determining which selection criteria would give the best results) and profitability evaluation has been made impossible.

Our research show how logit could be applied to decision making in direct marketing activities. The model we built for a casino company has a joint prediction probability of 84.5%. In terms of industry research, a deeper analysis of performance between various types of RFM models and logit should be done, before the company switches

to the use of the logit model. The company reports a big segment of lost customers (as per RFM model) which would require a broad understanding of their behaviour and a set of activities to bring them back. None of suggested models address this issue. Situational (those that only become loyal given a certain incentive from the company) and true loyalist should also be addressed separately, as company is losing money on the second group which would come to the service regardless of the direct marketing support received.

It also needs to be stressed that the model was built on a single casino customer base and it is yet to be tested on other casinos to eliminate specific casino properties.

References

- Aggelis, V., and C. Dimistris. 2005. 'Customer Clustering Using RFM Analysis.' <http://www.wseas.us/e-library/conferences/2005athens/csc/papers/497-433.pdf>
- Bauer, C. L. 1988. 'A direct Mail Customer Purchase Model.' *Journal of Direct Marketing*, 2:16-24.
- Bult, J. R., and T. Wansbeek. 1995. 'Optimal Selection for Direct Mail.' *Marketing Science* 14 (4): 378-94.
- Churchill, G. A., and D. Iacobucci. 2002. *Marketing Reserch: Methodological Foundations*. Mason, OH: South-Western Thomson Learning.
- Coussement, K., P. Harrigan, and D. F. Benoit. 2015. 'Improving Direct Mail Targeting through Customer Response Modeling.' *Expert Systems With Application* 42:8403-412.
- Hosseini, S. M., A. G. Maleki, and M. R. Taghva. 2010. 'Cluster Analysis using Data Mining Approach to Develop CRM Methodology to asses the Customer Loyalty.' *Expert Systems with Applications* 37:5259-64.
- Leenheer, J., H. J. van Heerde, T. H. A. Bijmolt, and A. Smidts. 2007. 'Do Loyalty Programs Really Enhance Behavioral Loyalty? An Empirical Analysis Accounting for Self-Selecting Members.' *International Journal of Research in Marketing* 24:31-47.
- Kass, G. V. 1976. 'Significance Testing in Automatic Interaction Detection.' Doctoral dissertation, University of Intwatersand, South Africa.
- Malthouse, E. 2001. 'Assessing the Performance of Direct Marketing Scoring Models.' *Journal of Interactive Marketing* 15 (1): 49-62.
- Musek Lešnik, K. 2008. *Od zadovoljstva potrošnikov do programov zvestobe*. Koper: Fakulteta za management.
- Nash, E. L. 1992. *Predictive Modelling*. New York: McGraw Hill.
- Olson, D. L., and B. K. Chae. 2012. 'Direct Marketing Decision Support through Predictive Customer Response Modeling.' *Decision Support Systems* 54:443-51.

- Sharp, B., and A. Sharp. 1997. 'Loyalty Programs and Their Impact on Repeat-Purchase Loyalty Patterns.' *International Journal of Research in Marketing* 14 (5): 473–86.
- Shepard, D. 1999. *The new Direct Marketing*. New York: Irwin.
- Simonof, J. S. 2016. 'Logistic Regression: Modeling the Probability of Success.' <http://people.stern.nyu.edu/jsimonof/flases/2301/pdf/logistic.pdf>
- Taghva, M. R., M. H. S. Bamakan, and S. Toufani. 2011. 'A Data Mining Method for Service Marketing: A Case Study of Banking Industry.' *Management Science Letters*. 1:253–62.
- Wang, J., and L. Sing'oei. 2013. 'Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing.' *IJCSI International Journal of Computer Science Issues*. 10 (2): 198–203.
- Wierenga, B. 2008. *Handbook of Marketing Decision*. New York: Springer.
- Winer, Russel S. 2001. 'A Framework for Customer Relationship Management.' *California Management Review* 43 (4): 89–105.
- Wollen, R. 2017. 'A Modern Approach to RFM Segmentation.' <https://cdn2.hubspot.net/hub/184373/file-41856256-pdf/docs/modern-approach-to-rfm-segmentation-ebook.pdf>
- Yeh, I-C., K.-J. Yang, and T.-M. Ting. 2009. 'Knowledge Discovery on RFM Model Using Bernoulli Sequence.' *Expert Systems with Applications* 36:5866–71.



This paper is published under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (cc BY-NC-ND 4.0) License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).