Tourism Management Perspectives 18 (2016) 153–160

Contents lists available at ScienceDirect

Tourism Management Perspectives

journal homepage: www.elsevier.com/locate/tmp

Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis

Aslıhan Dursun [a,*], Meltem Caber [b]

[a] Tasliburun Location, Belek Tourism Area, Belek, Antalya, Turkey
[b] Dumlupınar Boulevard, Campus, Antalya, Turkey

ARTICLE INFO

Article history:
Received 2 September 2015
Received in revised form 29 February 2016
Accepted 5 March 2016

Keywords:
Data mining
RFM analysis
CRM
Hotel customer

ABSTRACT

This study focuses on profiling profitable hotel customers by RFM analysis, which is a data mining technique. In RFM analysis, Recency, Frequency and Monetary indicators are employed for discovering the nature of the customers. In this study, the actual CRM data belong to three five-star hotels operating in Antalya, Turkey were used. Analysis results showed that 369 profitable hotel customers were divided into eight groups: '*Loyal Customers*', '*Loyal Summer Season Customers*', '*Collective Buying Customers*', '*Winter Season Customers*', '*Lost Customers*', '*High Potential Customers*', '*New Customers*', and '*Winter Season High Potential Customers*'. Majority of the customers (36%) were positioned at '*Lost Customers*' segment, who stay for shorter periods, spend less than other groups and tend to come to the hotels in the summer season. Results indicated that RFM effectively clusters the customers, which may lead hotel top managers to generate new strategies for increasing their abilities in CRM.

1. Introduction

Under tough global competition, managers should seize the opportunities that have high capability of returns on capital within a time frame. With this aim, data which were obtained during daily operations and stored within the warehouses for customer relationship management (CRM) purpose have to be transformed into useful knowledge (Ha & Park, 1998). The obtained knowledge from the data may also minimize managerial risks and increase the effectiveness of CRM strategies (Cheng & Chen, 2009). Thus, data mining methods maintain the identification of the hidden meaningful trends and the relationships inside data. Identification of the most profit-generating customers and segmentation of customers relying on variables stored in the datasets are quite vital, since previous studies in the services sector show that only 15% of the customers generate 45% revenue, and 70% of profit (Ivanovic, Mikinac, & Perman, 2011). In addition, it is evident that customer loyalty and profitability are correlated (Payne, Christopher, Clark, & Peck, 1999). Therefore, one of the main assumptions of CRM is; satisfying and creating long term relationships with profitable customers enhance the business success of the companies (Wu & Lu, 2012).

Garrido-Moreno and Padilla-Meléndez (2011) summarizes the key factors of a successful CRM implementation by a literature review as the: organizational factors, technology, customer orientation and CRM experience. In particular, hotels as major players (with total revenues of 457 billion US dollars, in 2011) of the global tourism sector (which totally contributes 6.44 trillion US dollars to the world economy, in 2011) (http://www.statista.com), have high capacity of technology usage, and mostly perform contemporary marketing strategies like CRM. Ivanovic et al. (2011) note that hotels "have a positive attitude regarding the implementation of CRM in the business, unlike others", and widely benefit from CRM systems especially for new product development purposes. Nowadays, many hotels proactively gather and register information about customer preferences into CRM systems (Sarmaniotis, Assimakopoulos, & Papaioannou, 2013). Hotel companies also target to understand the customers in order to generate customized products (Min, Min, & Emam, 2002). 'Customization' primarily requires having knowledge about customer preferences and behaviour (Adomavicius & Tuzhilin, 2001). Once the hotels begin to know and categorize the customers, they may develop appropriate products and marketing strategies for each group. As a result, hotel companies meet customer needs and demands, make the customers highly satisfied, and maintain their loyalties. In this way, both the efficiency of CRM efforts and the ability of the company in terms of competitiveness may be increased. For example, findings of a recent study by Wu and Lu (2012), which investigates the relationships amongst CRM, relationship marketing, and business performance in the sample of Taiwanese hotel sector, confirm that CRM implementations have a significant and positive influence on the hotels' relationship management effectiveness, and business performance. However, advanced analysis techniques are not adequately used in the hotel business yet, with the purpose of effectively profiling the customers by using comprehensive data collected via hotel CRM systems.

* Corresponding author.
E-mail addresses: aslican.dursun@gmail.com (A. Dursun), meltemcaber@akdeniz.edu.tr (M. Caber).

http://dx.doi.org/10.1016/j.tmp.2016.03.001

The purpose of the present study, therefore, is profiling profitable hotel customers by using a data mining technique, called RFM (Recency, Frequency and Monetary) analysis. RFM model is a simple technique (Hughes, 1994) for "defining valuable customers as those simultaneously having high Recency, Frequency, and Monetary values" (Hu & Yeh, 2014). As noted by some academics (Khajvand, Zolfaghar, Ashoori, & Alizadeh, 2011), RFM model may be the most powerful and simplest technique for generating knowledge from CRM data (Kahan, 1998; McCarty & Hastak, 2007). It had been, therefore, a widely used method by the researchers in many areas. As summarized by some of the academics (Olson, Cao, Gu, & Lee, 2009; Wei, Lee, Chen, & Wu, 2013; Hu & Yeh, 2014), RFM analysis is used for the identification of profit-generating customers, segmentation of the customers in terms of CRM, generation of new products or services, measurement of the customer lifetime value areas in the finance, telecommunication, electronic, online, and travel companies, meat production retailers, and many other areas. However, to the knowledge of the authors, RFM analysis was not employed before by the academics with the purpose of generating valuable hotel customers' profile. Thus, the current study will be the first attempt in the literature which investigates the profiles of the profitable customers in the content of CRM, by the use of RFM analysis.

Following to this section, paper continues with the presentation of customer profiling and data mining principles. In the next section, the RFM model and its methods are explained. The Methodology section presents the sample and measurement tool of the study. Results of the RFM analysis are followed by the conclusion, where findings of the study are summarized and managerial implications are offered.

2. Customer profiling and data mining

Demographics, socioeconomic, or geographic characteristics of the customers are the traditionally and widely used variables for market segmentation (Frochot & Morrison, 2000). In the tourism literature, segmentation criterions that are generally employed have been travel expenditures (Mudambi & Baum, 1997; Mok & Iverson, 2000), travel motivations (Cha, McCleary, & Uysal, 1995; Pesonen, 2012), destination activities (Dotson, Clark, & Dave, 2008), benefit seeking attitudes (Frochot & Morrison, 2000; Koh, Yoo, & Boger, 2010), industry data (Chung, Oh, Kim, & Han, 2004), and technology readiness index (Victorino, Karniouchina, & Verma, 2009). The academics usually adapt quantitative approaches for profiling and segmenting customers such as factor analysis, conjoint analysis, linear regression or logistic regression analysis, discriminate analysis (Chung et al., 2004), neural networks, and CHAID (Chi-Squared Automatic Interaction Detection) (Bowen, 1998).

In the contemporary hotel management, customer preferences, behaviour and profiles are understood by analyses of the data, gathered together from several customer–employee contact points and recorded into CRM systems. By filtering and extracting the necessary data from the data warehouses or databases, and originating meaningful forecasts for the future (Savaş, Topaloğlu, & Yılmaz, 2012), data mining actually tries to answer the question 'what will happen?' (Ünal, 2011). Data mining works like one actor of a wider process known as 'knowledge discovery' that consists of several stages to be followed for filtering out the meaningful results (Rygielski, Wang, & Yen, 2002). The most frequently used data mining methods are categorization, clustering, connotation rules, regression analysis, and sequence analysis. In addition to these, rule-based reasoning, genetic algorithms, decision trees, fuzzy logics, inductive training systems, RFM analysis, and other statistical methods are used with the aim of data mining by the researchers (Cheng & Chen, 2009). In the next section, detailed information about RFM analysis is presented, which is used in the current study with the purpose of customer profiling.

3. RFM analysis

RFM analysis is a well-known (Hu & Yeh, 2014), behavioural-based data mining method, which extracts the customer profile by using few numbers of criterions, and by reducing the complexity of analysis (Kaymak, 2001). In RFM analysis, customer data are classified by Recency (R), Frequency (F) and Monetary (M) variables (McCarty & Hastak, 2007). Recency shows the length of time since the latest purchase (such as days or months); Frequency is the number of purchases in a period; and Monetary indicates the total amount of spending in a period (Wei et al., 2013; Hosseini, Maleki, & Gholamian, 2010). Previous studies show that "the bigger the values of R and F are, the more likely the customers are going to produce a new trade with the company; the bigger M is, the more likely the customers are going to buy more services or products of the company" (Cheng & Chen, 2009). Exceptionally, RFM indicators are adaptable to measure customer values and to segment customers in different services areas such as finance, telecommunication, electronic commerce, etc. In a recent literature review by Wei, Lin, and Wu (2010), it is noted that RFM enables the practitioners "to observe customer behaviour, to segment customers, to estimate the response probability for each offer type, to calculate customer value and customer lifetime value and to evaluate on-line reviewers".

RFM analysis has both advantages and disadvantages. The main advantages of RFM are: (1) being a powerful tool for assessing customer lifetime value, which is also available to be combined with frequent pattern mining techniques (Hu & Yeh, 2014); (2) being considered as "a basis for a continuing stream of techniques to improve customer segmentation" (Elsner, Krafft, & Huchzemeier, 2003), and (3) being effective in predicting response and boosting company profits in a short term (Baecke & Van den Poel, 2011). The main disadvantages are: (1) its insufficiency for generating successful marketing programmes by using only three criterions, while some other attributes such as customers' income, lifestyle, and product variation are ignored and not included to the analysis (Fitzpatrick, 2001); (2) high correlations between Frequency and Monetary values (Olson et al., 2009), (3) ignorance of the potential and non-profit customers, and (4) varying importance of the RFM indicators from one industry to another (Băcilă, Rădulescu, & Marar, 2012). Improvement efforts of RFM model by the researchers has been significantly increased in the recent years. For example, for increasing the number of the indicators of RFM analysis, Cheng and Chen (2009) proposed the use of RFMTC model (Recency, Frequency, Monetary value, time since first purchase, and Churn probability). However, RFM model results outperformed to RFMCI model.

In principle, RFM analysis can be performed by following one of the several approaches. In the traditional application of RFM, all dimensions of the customer data is analysed, and later coded by dividing it into five categories. It is called as 'the customer quintile method'. By coding, each customer is compared with all the others depending on the parameters used. Then, for each customer a score is calculated. For Recency (R), purchase dates are decreasingly ranged. Of the most recently purchasing customers, 20% are numbered 5, the next 20% are numbered 4, and so on. Similarly, both for Frequency and Monetary, data are ranged decreasingly. All of the customers are coded to 555, 554, 553, … 111 by 125 different versions ($5 \times 5 \times 5$). In this way, the data base is divided into 125 equal clusters. The customers who have the highest RFM scores are generally the company's most profitable customers (Hosseini et al., 2010; Wei et al., 2013). Miglautsch (2000) notes the advantage of this method for projecting customer behaviour, if segmentation schemes are generated periodically. However, he also argues that the main disadvantage of the customer quintile method is its tendency of "grouping together customers who have vastly different buying behaviour (at the top) and arbitrarily break apart customers who have identical behaviour (at the bottom)".

Some academics recommended to sort of the customers by generating cutoffs on percentage behaviour for solving sensitivity problem of the customer quintile method. It is known as the 'behaviour quintile