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
2211-9736/

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 scoring'. In this method, monetary score is expected to create almost equal amounts of sales in each quintile. In addition, this method enables to group the customers who has similar behaviours (Miglautsch (2000). Some of the academics suggested the use of weighted RFM method, as an alternative, which weights to R, F, and M indicators depending on characteristics of the industry. For example, some researchers suggest placing the "highest weighting on the Frequency, followed by the Recency, with the lowest weighting on the Monetary; some others offer to give the most value to Monetary, and the least value to Recency" (Khajvand et al., 2011). Weighted-RFM method examines "the relative importance of the RFM variables via AHP (Analytic Hierarchy Process) algorithm, while the non-weighted RFM method does not" (Wei, Lin, & Wu, 2010). Similar versions of RFM are: "Timely RFM (TRFM) which deals with the product periodicity; RFD (Recency, Frequency, Duration) which measures the duration website visits; RML (Recency, Monetary and Loyalty) which adapts RFM into annual transaction environments; RFR (Recency, Frequency, Reach) which was proposed for social graph, such as Recency — last post, Frequency — total number of posts, Reach — networks, friends; FRAT (Frequency, Recency, Amount and Type of goods) which aims to improve the segmentation by taking into account the categories of bought products" (Birant, 2011, p. 94).

In addition to these models, Chen, Kuo, Wu, and Tang (2009) offered minimum and maximum cut-off points for each RFM indicator. For Recency (R), these limits are named as $R_{time\_minimum}$ and $R_{time\_maximum}$, which show the distance between the subject sample and the specified beginning date. For example, if a subject sample purchased in the last 200–270 days, $R_{time\_minimum}$ will be: 200 and $R_{time\_maximum}$ will be: 270. Similarly, Monetary (M), which shows amount of spend, has to be classified inside $M_{\_minimum}$ and $M_{\_maximum}$. Frequency (F) is the percentage of sum series of the recency and monetary limits.

Another method of RFM model is the usage of original data instead of coded numbers. In this method, Recency (R) shows the time since the latest purchase was made, Frequency (F) indicates the total number of purchases in this period, and Monetary (M) is the total amount of all purchases. For R, time period starts from the earliest purchase and starts at 1, and continues by plus 1 for each day until the most recent date. For example, in the case of the time period between 1st January and 31st March 2015: 1 for 1st January; 2 for 2nd January, and 91 for 31st March should be given as R values. F value is the customers' total occurrences of purchase in this period. M value is determined by the total amount of purchases. Depending on the means, clusters which have RFM scores higher than average are shown by '↑', and lower than average are shown by '↓'. In the result, members of the cluster which is symbolized by R↑F↑M↑ are named as the '*Loyal Customer*'. Members of the cluster with R↓F↓M↓ symbol are the '*Lost Customers*', and R↑F↓M↓ shows the '*New Customers*' cluster. Finally, members of the cluster which is symbolized by R↑F↑M↓ are named as the 'potential customers' (Wei et al., 2013).

In previous studies, RFM analysis was employed in the tourism literature, for examining destination revisiting intention and tourists' loyalty (Oppermann, 2000; Wong, Chen, Chung, & Kao, 2006; Jang & Feng, 2007; Assaker, Vinzi, & O'Connor, 2011), investigating customer value relying on pre-purchase motivations (Lumsden, Beldona, & Morrison, 2008), analysing purchasing behaviours at hotel duty-free shops (Weng, Ruey-Kei, Wang, & Su, 2006/2007), segmenting (Wong, Chen, Chung, & Kao, 2006), and generating demand forecasts. However, application of RFM model in the studies about hotel marketing and CRM fields are very scarce. Although Morrison, Bose, and O'Leary (2000) conducted a RFM model in the hotel sector, the data set they used was obtained from the marketing database of an American financial institution which provides credit card services. In addition, they used customer demographic, socio-economic, psychographic, purchase and credit behaviour characteristics to individual RFM models by regression analysis.

4. Methodology

Hu and Yeh (2014) recommend the researchers to use RFM indicators obtained from the customer membership cards or login accounts. Therefore, in this study, the authors decided to use original customer data, obtained from a hotel chain's membership card system and CRM database. In the RFM model of the present study, Recency (R) indicates the time since the latest purchase was made; Frequency (F) shows the total number of purchases in this period, and Monetary (M) reflects the total amount of all purchases, by following one of the previously explained RFM methods. In this method, 'loyal', 'lost', 'new', and 'potential' customers can be easily distinguished from each other in terms of CRM, since customer groups are individually symbolized. In addition, RFM analysis is conducted by using periodic data which enabled the researchers to compare the differences between high and low season customers visiting the hotels.

4.1. Sample of the study

The sample hotel chain has a membership card system based on loyalty points earned by the total overnights in the hotel chain. 'One point for one overnight' rule was followed for calculating the loyalty points of the cards to be given to the customers who stayed at least one time in the hotels for 10 days or more. Five types of membership were then formed according to total overnights of the customers, starting from 2011. 'Non-members' are the customers who stayed more than 10 nights, but have been only once in the hotels. 'Silver' cards are given to the customers who both had been in the hotels more than once and stayed for minimum 30 nights in total. 'Silver Plus' cards are given to customers who have been more than once and made a minimum 60 overnights; minimum 100 overnights were awarded by 'Platinum' cards.

The population of the research consists of 5939 hotel customers who visited the hotels at least once, stayed more than 10 nights, either were given the hotel membership cards or not, in the period of 15.04.2014–15.04.2015. Distribution of the membership cards is shown in Fig. 1.

The authors, unfortunately, could not use the whole population (5939) in the analyses, since data registered to the hotel CRM system was not adaptable to the statistics programme used in this study. The size of the sample for a population that contains 6000 people is calculated to be 362 with 95% reliability level and ±0.05 error margin (Bartlett, Kotrlik, & Higgins, 2001). Since the population has some subgroups, stratified sampling method was preferred. Stratified sampling enables

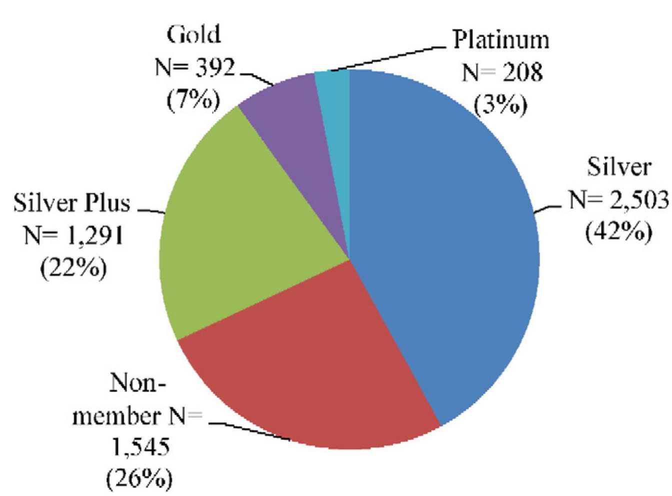


Fig. 1. Distribution of the hotel membership cards.

the researchers to select the samples randomly from a population with specified borders where subunit groups exist (Yıldırım & Şimşek, 2005). Accordingly, in the following phase, the authors one by one selected and re-coded the customer data in a separate work sheet. As a result of this procedure, sample of this study consisted of 369 hotel customers who were selected by stratified sampling method from the dataset relying on each subgroup's percentage in total (Table 1).

5. Results

5.1. Demographics of the customers

Demographics of the hotel customers are shown in Table 2. There were 60.4% female (f = 223), and 39.6% male (f = 146) customers. The age groups were: 35.2% aged 35–44 (f = 130), 19.2% aged 55–64 (f = 71). The majority of the customers were the Russians (36%) and Germans (19.2%). Most of the customers were couples (37.7%; f = 139) with children (48.2%; f = 178).

5.2. RFM analysis indicators

R (Recency) value in this study indicates the most recent stay of the customers. The data set was limited to the period 15.04.2014–15.04.2015 for RFM analysis. Thus, beginning with 1 from 15.04.2014, each of the dates are numbered respectively, up to 366 for 15.04.2015. F (Frequency) value is determined by the customers' total number of visits in this period. M (Monetary) is calculated by dividing the total amount of expenses of the customers into the number of people who stayed in the same room. Thus, monetary indicator in this study just consists of room price that was paid by the customers. Monetary values in Turkish Liras (TL) or US dollars were converted to Euros (EUR) (Table 3).

5.3. Self organizing maps (SOM) and K-means

Self organizing maps (SOM) is a technique used as data visualizing tool where numbers of the clusters are not definite (Oğuzlar, 2005). K-means is one of the most popular algorithms used in cluster analysis (Tsiptsis & Chorianopoulos, 2009, p. 85), mostly in the areas of data mining, and statistical data analyses (Cheng & Chen, 2009). For conducting the cluster analysis in this study, cluster numbers had to be explored by using SOM to be used later in K-means algorithm. By using Viscovery SOMine software, and following SOM-Ward-Clusters method, eight clusters were obtained. It was visually obvious that each of the clusters had their own boundaries and inner structures (Fig. 2).

Beside to SOM findings, possible combinations of the clusters were also found to be 8 (2 × 2 × 2), by using K-means method with R, F and M values. R, F, M Values were categorized in two groups with R, F, M '↓' for under average, and R, F, M '↑' for above average (Sohrabi & Khanlari, 2007). In the following phase, RFM scores were symbolized by '↓', if R, F or M values are lower than average, and by '↑', if R, F or M values are above average (Table 4).

Table 1
Sampling according to membership card types.

	Silver	Silver Plus	Gold	Platinum	Non-member	Total
Population	2503	1291	392	208	1545	5939
Share in total population (%)	42%	22%	7%	3%	26%	100%
Sample size	156	80	24	13	96	369

Table 2
Demographics of the Customers ($N = 369$).

Variables	Groups	Frequency	Percentage (%)
Gender	Male	146	39.6
	Female	223	60.4
	18–24	14	3.8
	25–34	36	9.8
Age	35–44	130	35.2
	45–54	64	17.3
	55–64	71	19.2
	65 and above	54	14.6
	Russian	133	36.0
Nationality	German	71	19.2
	French	16	4.3
	British	16	4.3
	Turkish	16	4.3
	Other	117	31.9
	Single	34	9.2
Travel companion	Couple	139	37.7
	With children	178	48.2
	Group of seniors	18	4.9
	Odeon	62	16.8
	Pegas	34	9.2
	TUI	34	9.2
	Thomas cook	27	7.3
	DER touristic	24	6.5
Travel organizer	Intourist style	19	5.1
	Diana	16	4.3
	Enda	15	4.1
	Exim	13	3.5
	Other tour operators	101	27.3
	Individual	24	6.5

5.4. Results of the RFM analysis

RFM analysis results about the cluster characteristics according to highest observation numbers are shown in Table 5. In this table, clusters which have loyalty points below average ($\overline{X} = 25.3$) are shown with '↓', while equal or higher than average are shown with '↑' symbols. For the first cluster, values of R ($\overline{X} = 212$), F ($\overline{X} = 2.5$) and M ($\overline{X} = 2073$) were above average R, F, and M values and this cluster, therefore, is symbolized by R↑F↑M↑. Customers of this cluster were accommodated in the hotel chain more than once in a year. Thus, this cluster is named '*Loyal Customers*'. '*Loyal Customers*' spend more than average, visit the hotels more frequently and mostly stay in the low season. '*Loyal Customers*' are mostly Russian couples (35.8%) in the age group of 35–44 (35.7%), who prefer to stay in standard rooms. They make reservations either individually or via travel agencies. In the last three years, these customers made an average of 57 overnight stays in the hotel chain. '*Loyal Customers*' are an important segment, whose loyalty tendencies should be sustained by some company offerings and campaigns that are specially generated for them.

The second cluster is symbolized by R↓F↑M↑ and named '*Loyal Summer Season Customers*'. Customers in this cluster seem to prefer summer season for their travel as the R value is lower than the average, and made high spends in more than one stay within selected time period. '*Loyal Summer Season Customers*' are mostly Russian (58.4%) females (58.3%), in the age group of 35–44 (58.3%). These customers make reservations via travel agencies, travel with their children and prefer standard rooms for accommodation. An average 37 overnight stays have been undertaken by these customers since 2011. If the hotel company

Table 3
RFM indicators.

RFM indicators	Minimum	Maximum	$\overline{X}$	St. dev.
R (Recency)	2	363	141.42	90.653
F (Frequency)	1	5	1.13	0.427
M (Monetary)	105	6899	1030	2530

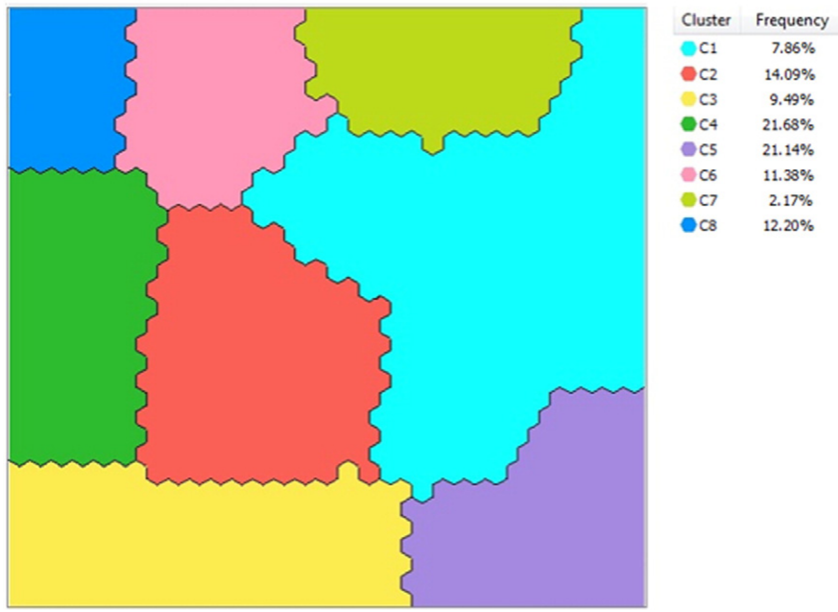


Fig. 2. Clusters offered by SOM.

suggests attractive prices and makes promotions to the customers in this cluster, some part of this demand may be transferred to winter season to achieve effective yield management.

The symbol of the third cluster is R↓F↑M↓, since they did not have a high spend, although they had multiple stays in the hotel chain in a year. Total spend (M) might have been affected by the calculation method used in this study. While per person spends were found by dividing the total room prices into number of the people who stayed in the same room, obtained amounts might have been negatively affected when most of the customers stayed in suites or standard rooms with their families and friends. This cluster is named as '*Collective Buying Customers*'. '*Collective Buying Customers*' are equally male and female, and in the age group of 35–54 (100%). They prefer to travel with their children (100%). The majority of these customers are Romanian and Turkish, and accommodated in suite-type rooms. Reservations are made through travel agencies or individually. According to loyalty points earned by the customers in this cluster, they made 12 overnight stays since 2011 in the chain hotels. '*Collective Buying Customers*', stay for a short time period, with their families, in the summer time.

In spite of R ($\overline{X} = 278.1$) and F ($\overline{X} = 2$), values of the fourth cluster were above average, M ($\overline{X} = 763$) value was found to be lower than the average M. This cluster is symbolized as R↑F↑M↓ and named '*Winter Season Customers*'. Customers in this cluster are the people who were accommodated in the hotel chain in the winter season, more than once and had a high spend. '*Winter Season Customers*' travel in couples and make their reservations either individually or via travel agencies.

Table 4
RFM Scores of the clusters.

Clusters	N	Recency $\overline{X}$	Frequency $\overline{X}$	Monetary $\overline{X}$	RFM Scores
1	14	212.0	2.5	2,073 EUR	R↑ F↑ M↑
2	12	96.9	2.3	1,732 EUR	R↓ F↑ M↑
3	2	93.5	2.0	698 EUR	R↓ F↑ M↓
4	9	278.1	2.0	763 EUR	R↑ F↑ M↓
5	133	78.7	1.0	669 EUR	R↓ F↓ M↓
6	81	92.0	1.0	1,837 EUR	R↓ F↓ M↑
7	95	247.3	1.0	560 EUR	R↑ F↓ M↓
8	23	171.4	1.0	1,351 EUR	R↑ F↓ M↑
Total	**369**	**141.42**	**1.13**	**1,030 EUR**	

They are middle-aged (mostly 35–44 years old) (66.7%) customers who are accommodated in standard-type rooms. The majority of these customers are German and Russian. Average loyalty points of the customers in this cluster show that '*Winter Season Customers*' made 24 overnight stays in the hotel chain. Although this group stayed in the hotels more than once in a year, their profitability potential is low as they stay for a shorter time and spend less than other customers.

In the fifth cluster, all R ($\overline{X} = 78.7$), F ($\overline{X} = 1$) and M ($\overline{X} = 669$) values were found below average, which, for this reason, was symbolized as R↓F↓M↓. This cluster also contains the highest observation number (36.0%; N = 133). These are the customers who stayed in the hotel chain once and had a low spend in the summer season. They are named '*Lost Customers*'. '*Lost Customers*' are identified to be middle-aged (35–44 years old) (36.1%), travelling with their children (52.7%) and staying in standard rooms. Most of the customers are Germans (20.3%), who organize their travel either from central reservation systems or travel agencies. The average loyalty points that were calculated starting from 2011 for this cluster show that the customers made 23 overnight stays in the chain hotels. The average loyalty point of this cluster is low. Therefore, customers' frequency of travel has to be increased by attractive pricing and marketing efforts.

The sixth cluster has R ($\overline{X} = 92$) and F ($\overline{X} = 1$) values lower than average, but M ($\overline{X} = 1837$) value higher than the average M. Although the customers in this cluster stayed in the hotels once, they had a high spend in the summer time. They have high potential of purchase and impact on profitability of the hotels. Thus, the sixth cluster is symbolized as R↓F↓M↑, and named '*High Potential Customers*'. They make reservations through travel agencies, travel with their children and stay in standard rooms. The majority of the customers are middle-aged (35–44 years old) (50.6%), female (58.05) Russians (79.0%). Since many of the platinum card owners positioned in this group prefer to stay in private villas, they show high potential of spending. The average loyalty point of the cluster is 28 overnight stays, calculated starting from 2011. Members of this cluster should be offered special service packages and special interest programmes in the winter season so that the company may keep them visiting chain hotels and meet their specific expectations.

Customers of the seventh cluster stayed in the hotel chain once, and made limited spend in the selected period. This cluster is symbolized as R↑F↓M↓ and named as '*New Customers*'. '*New Customers*', who are mostly male (72.6%), 45–54 years old (21.1%), travel in couples, and

Table 5
Characteristics of the clusters.

CLUSTER NAME	RFM SCORES	N (%)	GENDER (%)	AGE GROUP (%)	NATIONALITY (%)	TRAVEL COMPANION (%)	ROOM TYPE (%)	CARD TYPE (%)	LOYALTY POINTS		
									↑ ↓	$\bar{X}$	St.Dev.
'Loyal Customers'	R↑ F↑ M↑	**14** (3.79%)	**Male and female** (50.0%-50.0%)	**35-44** (35.7%)	**the Russian** (35.8%)	**Couple** (57.2%)	**Standard** (92.9%)	**Silver Plus and Platinum** (35.7%-35.7%)	↑	57.4	35,3
'Loyal Summer Season Customers'	R↓ F↑ M↑	**12** (3.25%)	**Female** (58.3%)	**35-44** (58.3%)	**the Russian** (58.4%)	**With children** (83.4%)	**Standard** (100.0%)	**Silver and Silver Plus** (33.3%-33.3%)	↑	36.8	17,7
'Collective Buying Customers'	R↓ F↑ M↓	**2** (0.05%)	**Male and female** (50.0%-50.0%)	**35-44 and 45-54** (50.0%-50.0)	**Romanian and Turkish** (50.0%-50.0%)	**With children** (100%)	**Suite** (100%)	**Silver and Silver Plus** (50.0%-50.0%)	↓	12.0	0,0
'Winter Season Customers'	R↑ F↑ M↓	**9** (0.02%)	**Male** (55.6%)	**35-44** (66.7%)	**German and the Russian** (33.3%-33.3%)	**Couple** (66.7%)	**Standard** (100%)	**Silver** (77.8%)	↓	23.8	11,4
'Lost Customers'	R ↓F↓ M↓	**133** (36.0%)	**Male** (66.2%)	**35-44**(36.1%)	**German** (20.3%)	**With children** (52.7%)	**Standard** (90.2%)	**Silver** (60.9%)	↓	22.7	14,3
'High Potential Customers'	R↓ F↓ M↑	**81** (21.9%)	**Female** (58.0%)	**35-44** (50.6%)	**the Russian** (79.0%)	**With children** (81.5%)	**Standard** (81.5%)	**Non-member** (35.8%)	↑	28.1	20,1
'New Customers'	R↑ F↓ M↓	**95** (25.7%)	**Male** (72.6%)	**45-54** (21.1%)	**German** (31.6%)	**Couple** (60.0%)	**Standard** (88.4%)	**Non-member** (42.1%)	↓	21.2	13,9
'Winter Season High Potential Customers'	R↑ F↓ M↑	**23** (6.23%)	**Male** (60.9%)	**55-64** (34.8%)	**the Russian** (73.9%)	**Couple** (43.5%)	**Standard** (87.0%)	**Silver** (34.8%)	↑	25.3	10,3

stay in standard rooms. These customers generally do not have a loyalty card. Most of these customers are of German nationality and make their travel arrangements via travel agencies. The average loyalty points of the group indicate that '*New Customers*' made 21 overnight stays in average, since 2011. Low average loyalty points show that the customers in this cluster have to be motivated to visit and to spend more by some customer relationship management strategies.

The eighth cluster consists of the customers who made long stays in the winter season. The symbol of the cluster has to be R↑F↓M↑, which is named '*Winter Season High Potential Customers*'. They are mostly males (60.9%), 55–64 years old (34.8%) and Russian (73.9%). They travel in couples and prefer to stay in standard rooms. Average loyalty points of the cluster is 25 overnight stays in the hotel chain. These are the senior customers who are travelling in the winter season and staying in the hotels for long periods of time. The high loyalty points indicate the importance of these cluster members. Thus, hotel managers may consider generating customer relationship management strategies targeting to this cluster, which own different characteristics from the summer season customers.

6. Conclusion

While extensive literature shows the role of CRM on hotel business, not all data mining techniques have been employed by the academics and practitioners with the purpose of customer profiling. For example, RFM model is an ignored data mining technique in the hotel management context. The main contribution of the current study is fulfilling this gap in the literature. The study examines profile of the valuable customers of a five-star hotel chain, which is located in Antalya, Turkey. Results show that majority of the hotel membership card owners are the middle aged (35–44; 35.2%), couples (37.7%) with children (48.2%). This is a typical situation for resort hotels which are attracting tourists who mostly travel for holiday purpose. The sample hotel chain serves the Russian and German tourists in majority. As previously noted by Newell (1997), RFM analysis efficiently profiled the most profit-generating hotel customers in this study as well. Results show that eight clusters exist, that are important for CRM policies and practices of the sample hotel chain. Each of the clusters were named depending on their characteristics as the: '*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*'.

In principle, the customers with high RFM scores are 'usually the most highly responsive to promotions, the most likely to repurchase, the most profitable, and vice versa' (Hamzehei, Fathian, Gholamian, & Farvaresh, 2011). However, each of the RFM indicators should be considered individually as well depending on what it really indicates. According to Cheng and Chen's (2009) argument, the high R and F values indicates the repeat purchase tendency and buying potential; and the high M value shows the profitability of the customers. RFM results of this study show that '*Loyal Customers*' (N = 14; 3.79%) of the subject hotel chain have the highest R, F, and M values. This finding confirms the relationships amongst repeat purchase, loyalty and profitability. If a market segment has already become the repeat and frequent customers of a hotel, they may be expected to become less sensitive to price increases. If the customers are highly satisfied with the hotel services or products, they will show loyalty and tend to response positively to CRM efforts of the company.

In spite of that, '*Collective Buying Customers*' (N = 2; 0.05%), '*Winter Season Customers*' (N = 9; 0.02%), '*Lost Customers*' (N = 133; 36.0%), and '*New Customers*' (N = 95; 25.7%) have M values lower than the average. Although they might have frequently and recently visited the hotels, their potential of buying high-priced services and products is low. In CRM, companies need to target the market segments which have high customer lifetime values. Thus, it is clear from the results that five-star resort hotels are facing difficulties in attracting the market segments which have high spending potentials and customer lifetime values.

The largest customer group was '*Lost Customers*', who have R, F, and M scores lower than the average. Surprisingly, majority (60.9%) of the '*Lost Customers*' own Silver membership card, which is given the customers who made more than one hotel stay and overnighted for minimum ten days. However, using the number of hotel stays and overnights as the main indicators of customer profiling are shown to be insufficient, while these customers have not high potential of lifetime value in the context of CRM. Therefore, hotel managers should set up the requirements of obtaining Silver card, so that the right customers are identified by data analyses. A similar upscaling can be performed for the Silver Plus, Gold, and Platinum membership cards.

'*New Customers*' are the second largest customer group. Although, they are not the members of hotel card system yet, and their spendings are lower than average M; data belong to these group of customers may be checked periodically both for understanding their behavioural changes, and also for following their potentials. Similarly, '*High Potential Customers*' (N = 81; 21.9%) who do not have any membership cards yet, need to be followed closely in the long term.

7. Managerial implications

Obtaining information about customer characteristics, and analysing the previously recorded data for generating relationship marketing practices and CRM strategies became vital for the hotel companies due to worldwide competition. RFM analysis, in particular, enables hotel managers to 'discover the potential of profitable customers by the observation of their past behaviours' (Li, Lin, & Lai, 2010). The current study show that majority of the customers registered into CRM system of the sample hotel chain are the Russian tourists (N = 133; 36.0%), followed by German tourists (N = 71; 19.2%). According to RFM analysis, 79.0% of the '*High Potential Customers*', and 73.9% of the '*Winter Season High Potential Customers*' are constituted from the Russian tourists, while '*New Customers*', and '*Lost Customers*' are mostly comprised of German tourists (31.6%, and 20.3%, respectively). Interestingly, the customers who already have hotel membership cards, such as '*Loyal Customers*', '*Loyal Summer Season Customers*', '*Winter Season Customers*' and '*Collective Buying Customers*' represent minor segments in the total sample. Thus, hotel managers should also consider the size and market potential of customer groups for generating the most accurate service strategies.

In general, a complete customer profile is divided into two sections, which are: 'virtual' and 'behavioural'. While virtual profile is configured by the information obtained from the system such as the customer's name, gender, and birth date; behavioural profile describes the customer actions and generally contains the operation-based data (Adomavicius & Tuzhilin, 2001). In this wise, RFM model offers both virtual and behavioural segmentation of the customers relying on various variables important for CRM implementations. In the sample hotel chain, preferences, spendings and behaviours of the customers should be followed and registered by including more number of contact points into CRM system. Later, managers of the hotels may give accurate decisions about planning, budgeting, pricing, marketing and other issues by easy to understand RFM results. They may also compare the effectiveness of CRM implementations on various market segments in the selected periods (such as high and low seasons of the destination).

The more information registered to hotel CRM systems, the more detailed customer profiles can be generated by data mining techniques. Therefore, as noted by Hendler and Latour (2008), CRM programmes in the hotel companies has to be user-friendly and technologically sophisticated by allowing essential data collection from service points such as central reservations, front office and room service, so that integrated raw information can be converted into useful knowledge (Luck & Stephenson, 2009). In this way, hotels may ensure the success in CRM efficiency. Hotel managers, therefore, should aim to generate close cooperations amongst the CRM programme designers and the staff working at all operational points.

8. Limitations

This study has some limitations like any others. For example, minor clusters (Cluster 3 and Cluster 4) were obtained — similar to some previous studies conducted in the banking and hairdressing services (Sohrabi & Khanlari, 2007; Wei et al., 2013) — as a result of selecting low number of customer data out of hotel CRM system. That result shows the necessity of designing hotel CRM systems which may easy to adapt to the analytic programmes. Moreover, just using room price paid as monetary indicator of RFM analysis, limited the possibility of highlighting the total spendings of the customers. In particular, membership card systems of the hotels should enable to register detailed information about the customer spendings such as spa and wellness, extra payments, and à-la-carte restaurant visits. In addition, data analysis of this study was limited by one year data (15.04.2014–15.04.2015). That made it difficult to compare summer and winter season customers' profiles, since the Recency came closer to winter term. In the future studies, separate RFM analyses are suggested to be conducted for the summer and winter seasons so that customer characteristics can be compared to each other. Finally, study findings cannot be generalized to whole hotels operating in Antalya, since the sample hotel chain serves domestic and international customers who mostly travel with the purpose of holidaying. The customers who prefer to stay in the city centre hotels may have different RFM scores than holiday customers.

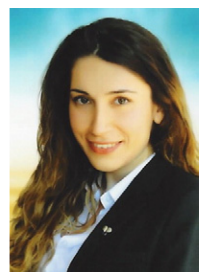
Acknowledgement

This paper is based on the Master of Social Sciences dissertation of Aslıhan Dursun, which is accepted by Akdeniz University, Social Sciences Institute, in the Tourism Management Programme in 2015.

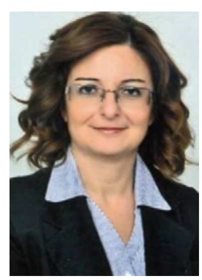
References

Adomavicius, G., & Tuzhilin, A. (2001). Using data mining methods to build customer profiles. *Computer*, *34*(2), 74–82.
Assaker, G., Vinzi, E., & O'Connor, P. (2011). Examining the effect of novelty seeking, satisfaction, and destination image on tourists' return pattern: A two-factor, non-linear latent growth model. *Tourism Management*, *32*(4), 890–901.
Băcilă, M. -F., Rădulescu, A., & Marar, I. L. (2012). *RFM based segmentation: An analysis of a telecom company's customers, marketing from information to decision (marketing from information to decision)5. (pp. 52–62)*, 52–62.
Baecke, P., & Van den Poel, D. (2011). Data augmentation by predicting spending pleasure using commercially available external data. *Journal of Intelligent Information Systems*, *36*(3), 367–383.
Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning and Performance Journal*, *19*(1), 43–49.
Birant, D. (2011). Data mining using RFM analysis, in Knowledge oriented applications in data mining, Funatsu, K., and Hasegawa, K., (Editors). pp. 91–108, InTech: Rijeka, Croatia.
Bowen, J. T. (1998). Market segmentation in hospitality research: No longer a sequential process. *International Journal of Contemporary Hospitality Management*, *10*(7), 289–296.
Cha, S., McCleary, K. W., & Uysal, M. (1995). Travel motivations of Japanese overseas travellers: A factor-cluster segmentation approach. *Journal of Travel Research*, *34*(1), 33–39.
Chen, Y. -L., Kuo, M. -H., Wu, S. -Y., & Tang, K. (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns. *Electronic Commerce Research and Applications*, *8*, 241–251.
Cheng, C. -H., & Chen, Y. -S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*, 4176–4184.
Chung, K. -Y., Oh, S. -Y., Kim, S. -S., & Han, S. Y. (2004). Three representative market segmentation methodologies for hotel guest room customers. *Tourism Management*, *25*, 429–441.
Dotson, M. J., Clark, J. D., & Dave, D. S. (2008). Travel motivations and destination activities of young travelers of different ages: An empirical study in the United States. *International Journal of Management*, *25*(4), 665–672.
Elsner, R., Krafft, M., & Huchzemeier, A. (2003). Optimizing Rhenania's mail-order business through dynamic multilevel modeling (DMLM). *Interfaces*, *33*(1), 50–66.
Fitzpatrick, M. (2001). Statistical analysis for direct marketers—In plain English. *Data Management*, *64*(4), 54–56.
Frochot, I., & Morrison, A. M. (2000). Benefit segmentation: A review of its applications to travel and tourism research. *Journal of Travel & Tourism Marketing*, *9*(4), 21–45.
Garrido-Moreno, A., & Padilla-Meléndez, A. (2011). Analyzing the impact of knowledge management on CRM success: The mediating effects of organizational factors. *International Journal of Information Management*, *31*, 437–444.
Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the intranet for database marketing. *Expert Systems with Applications*, *15*, 1–31.
Hamzehei, A., Fathian, M., Gholamian, M., & Farvaresh, H. (2011). A new methodology to study customer electrocardiogram using RFM analysis and clustering. *Management Science Letters*, *1*(2), 139–148.
Hendler, F., & Latour, K. A. (2008). A qualitative analysis of slot clubs as drivers of casino loyalty. *Cornell Hospitality Quarterly*, *49*, 105–121.
Hosseini, S. M., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*, 5259–5264.
http://www.statista.com/statistics/247264/total-revenue-of-the-global-hotel-industry/ (Accessed in 21.11.2015).

Hu, Y. -H., & Yeh, T. -W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information. *Knowledge-Based Systems*, *61*, 76–88.
Hughes, A. M. (1994). *Strategic database marketing.* Chicago, USA: Probus Publishing Company.
Ivanovic, S., Mikinac, K., & Perman, L. (2011). CRM development in hospitality companies for the purpose of increasing the competitiveness in the tourist market. *UTMS Journal of Economics*, *2*(1), 59–68.
Jang, S. C., & Feng, R. (2007). Temporal destination revisit intention: The effects of novelty seeking and satisfaction. *Tourism Management*, *28*, 580–590.
Kahan, R. (1998). Using database marketing techniques to enhance your one-to-one marketing initiatives. *Journal of Consumer Marketing*, *15*(5), 491–493.
Kaymak, U. (2001). Fuzzy target selection using RFM variables. *IFSA World Congress and 20th NAFIPS International Conference*, *2*, 1038–1043.
Khajvand, M., Zolfaghar, K., 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.
Koh, S., Yoo, J. J. -E., & Boger, C. A., Jr. (2010). Importance-performance analysis with benefit segmentation of spa goers. *International Journal of Contemporary Hospitality Management*, *22*(5), 718–735.
Li, Y. -M., Lin, C. -H., & Lai, C. Y. (2010). Identifying influential reviewers for word-of-mouth marketing. *Electronic Commerce Research and Applications*, *9*(4), 294–304.
Luck, D., & Stephenson, M. L. (2009). An evaluation of the significant elements of customer relationship management within the hotel industry. *Tourism Today*, *9*, 7–26.
Lumsden, S. -A., Beldona, S., & Morrison, A. M. (2008). Customer value in an all-inclusive travel vacation club: An application of the RFM framework. *Journal of Hospitality & Leisure Marketing*, *16*(3), 270–285.
McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, *60*, 656–662.
Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing*, *8*(1), 67–72.
Min, H., Min, H., & Emam, A. (2002). A data mining approach to developing the profiles of hotel customers. *International Journal of Contemporary Hospitality Management*, *14*(6), 274–285.
Mok, C., & Iverson, T. J. (2000). Expenditure-based segmentation: Taiwanese tourists to Guam. *Tourism Management*, *21*, 299–305.
Morrison, A. M., Bose, G., & O'Leary, J. T. (2000). Can statistical modeling help with data mining: A database marketing application for U.S. hotels. *Journal of Hospitality & Leisure Marketing*, *6*(4), 91–110.
Mudambi, R., & Baum, T. (1997). Strategic segmentation: An empirical analysis of tourist expenditure in Turkey. *Journal of Travel Research*, *36*(1), 29–34.
Newell, F. (1997). *The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry*. New York, USA: McGraw-Hills Companies Inc.
Oğuzlar, A. (2005). A new approach in cluster analysis: Self-organizing maps (Kohonen webs). *Journal of Economics and Administrative Sciences*, *19*(2), 93–107.
Olson, D. L., Cao, Q., Gu, C., & Lee, D. (2009). Comparison of customer response models. *Service Business*, *3*, 117–130.
Oppermann, M. (2000). Tourism destination loyalty. *Journal of Travel Research*, *39*, 78–84.
Payne, A., Christopher, M., Clark, M., & Peck, H. (1999). *Relationship marketing for competitive advantage.* Butterworth Heinemann: Oxford, UK.
Pesonen, J. A. (2012). Segmentation of rural tourists: Combining push and pull motivations. *Tourism and Hospitality Management*, *18*(1), 69–82.
Rygielski, C., Wang, J. -C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, *24*, 483–502.
Sarmaniotis, C., Assimakopoulos, C., & Papaioannou, E. (2013). Successful implementation of CRM in luxury hotels: Determinants and measurements. *EuroMed Journal of Business*, *8*(2), 134–153.
Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Data mining and its samples in Turkey, İstanbul Commerce University. *Journal of Faculty of Science*, *11*(21), 1–23.
Sohrabi, B., & Khanlari, A. (2007). Customer lifetime value (CLV) measurement based on RFM model. *Iranian Accounting & Auditing Review*, *14*(47), 7–20.
Tsiptsis, K., & Chorianopoulos, A. (2009). *Data mining techniques in CRM: Inside customer segmentation.* Chippenham, Wiltshire, UK: John Wiley and Sons, Ltd.
Ünal, S. (2011). *Customer relationships and knowledge technologies in the new economy.* İstanbul, Turkey: Beta Publishment.
Wei, J. -T., Lee, M. -C., Chen, H. -K., & Wu, H. -H. (2013). Customer relationship management in the hairdressing industry: An application of data mining techniques. *Expert Systems with Applications*, *40*, 7513–7518.
Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). A review of the application of RFM model. *African Journal of Business Management*, *4*(19), 4199–4206.
Weng, S. -S., Ruey-Kei, C., Wang, B. -J., & Su, S. -H. (2006/2007). The study and verification of mathematical modeling for customer purchasing behavior. *The Journal of Computer Information Systems*, *47*(2), 46–57.
Wong, J. -Y., Chen, H. -J., Chung, P. -H., & Kao, N. -C. (2006). Identifying valuable travelers and their next foreign destination by the application of data mining techniques. *Asia Pacific Journal of Tourism Research*, *11*(4), 355–373.
Wu, S. -I., & Lu, C. -L. (2012). The relationship between CRM, RM, and business performance: A study of the hotel industry in Taiwan. *International Journal of Hospitality Management*, *31*, 276–285.
Victorino, L., Karniouchina, E., & Verma, R. (2009). Exploring the use of the technology readiness index for customer segmentation. *Cornell Hospitality Quarterly*, *50*(3), 342–359.
Yıldırım, A., & Şimşek, H. (2005). *Qualitative research methods in social sciences.* Ankara, Turkey: Seçkin Publishing.



Aslıhan Dursun is currently CRM Manager of the Papillon Hotels Resort & SPA, located in Antalya, Turkey. She completed her M.Sc in the Tourism Management Programme at Akdeniz University, Social Sciences Institute. She began PhD education in 2015, at the same university, and programme.



Meltem Caber is associate professor at Akdeniz University, Tourism Faculty. She has several published articles and conference papers in the area of Tourism Management. Her research interests are tourism marketing, destination management and tourism economics.