Taxonomy of Marketing Strategies Using Bank Customers' Clustering

Azarnoush Ansari[1] & Arash Riasi[2]

[1] Department of Management, University of Isfahan, Isfahan, Iran

[2] Institute for Financial Services Analytics, University of Delaware, Newark, DE, USA

Correspondence: Arash Riasi, Alfred Lerner College of Business & Economics, Amstel Ave., Newark, DE, 19716, USA. Tel: 1-(302)-898-6249. E-mail: riasi@udel.edu

Abstract

Purpose–The goal of this study is to identify the main clusters of bank customers in order to help commercial banks to better identify their customers and design more efficient marketing strategies.

Design/methodology/approach – Data from 250 bank customers were analyzed by using two-step scalable clustering.

Findings – Five different clusters of bank customers were identified, namely, favorite customers, creditworthy customers, non-creditworthy customers, passers, and friends. The findings indicated that disparate clusters of bank customers are extremely different based on their loan amount, default risk, account balance, degree of loyalty and profitability for the bank.

Practical implications – The differences which were observed between these five clusters of bank customers accentuate the importance of customer clustering and market segmentation in the financial services industry. Customer clustering can help financial institutions to augment their competitiveness by shifting from traditional marketing strategies to target marketing and segmentation-based marketing approaches.

Originality/value - The most important contribution of this study is the incorporation of a wide range of factors that can potentially affect customer clustering in the analysis, whereas, the majority of previous studies only focused on a limited number of variables in order to determine the customer clusters. Specifically, the customer clustering in this study was performed by using demographic variables, profitability, loan amount, default risk, account balance, loyalty, account type, account closure history, customer location, and account currency.

Keywords: customer clustering, market segmentation, taxonomy, target marketing, financial services marketing

1. Introduction

Marketing strategy is concerned with effective and harmonic allocation of resources in order to achieve organizational goals in a specific market for a particular product/service. Therefore, a large portion of strategic marketing decisions are focused on identifying target markets and performing market segmentation. In order to improve product/service quality and to augment the degree of competitiveness, companies should identify the key needs of their target customers and should develop a strategy in order to satisfy these needs. The most salient issue in designing a marketing strategy is the fact that all customers are not the same and should not be treated similarly. Studies in services marketing have shown that in most cases, firms should not provide services to all customers in a similar fashion. Therefore, customer segmentation and customer relationship management are among the most important determinants of business viability (Ansari & Riasi, 2016a). In fact, customer relationship management helps the firms to enhance the customer value and enables them to retain their valuable customers.

Due to the extreme competition among companies and also due to the diversity of the products and services in the markets, identifying and analyzing consumers' behavior, and also adopting the best possible marketing strategies have become inseparable components of customer relationship management. By dividing the customers into different clusters, companies will be able to decide how to effectively allocate their limited resources to different groups of consumers based on their value. Companies try very hard to better identify their customers by analyzing their behavior, but the important point is that consumer behavior analysis requires a variety of industry-specific considerations. Particularly, analyzing the consumer behavior in financial services

industry is a very challenging task due to the heterogeneity of the customers and differences in customers' expectations (Ansari & Riasi, 2016b). One of the most useful techniques for analyzing the consumer behavior in financial services industry is customer clustering. By using clustering techniques, customers are divided into homogenous clusters in which customers with similar needs and characteristics are grouped together (Ghazanfari et al., 2010). After identifying the needs and values of its customers, a company can provide better services to its clients which will in turn lead to enhanced customer satisfaction and higher degrees of perceived service quality (Moslehi et al., 2012). This enhanced customer satisfaction and relationship commitment will also have a positive impact on the firm's brand loyalty and brand awareness (Ansari & Riasi, 2016b, 2016c; Fornell, 1992; Fornell & Wernerfelt, 1987; Gallarza & Saura, 2006; Kim et al., 2008; Kumar et al., 2013; Parasuraman et al., 1991; Pont & McQuilken, 2005; Reichheld & Sasser, 1990; Reynolds & Beatty, 1999; Salmones et al., 2009; Veloutsou et al., 2004).

According to Paswan et al. (2012) marketing strategy should be conceptualized as a mechanism which is used for competing in the marketplace. Every company might have customers that do not generate adequate profits; in other words, marketing strategies used for targeting certain groups of a firm's customers might be very costly compared to the revenues which are generated by them. On the other hand, there are always some more profitable groups of customers. The companies are always striving to retain the latter group of customers by offering them various incentives. Identifying these disparate groups of customers and their specific needs can help the firms to improve their profitability and competiveness. By using customer clustering, firms can better identify the behavior of their customers and design better marketing strategies. These relationships accentuate the need for using configuration theory in studying the consumer behavior. Configurations represent a number of specific and separate attributes that are only meaningful when used collectively (Dess et al., 1993; Rosenberg, 1968; Miller, 1987). They are finite in number and represent a uniquely integrated set of dynamics (Mintzberg. 1973; Miller, 1986, 1987). According to Dess et al. (1993), by using configurations, researchers can express complicated and interrelated relationships among different variables without resorting to artificial oversimplification of the phenomenon of interest. In other words, configurations are a useful tool for providing rich and complex descriptions of organizations (Dess et al., 1993; Hambrick, 1983; Miller & Friesen, 1977; Mintzberg, 1978). Previous researchers suggested that configuration models should be divided into taxonomies and typologies (Miller & Friesen, 1984; Meyer et al., 1993). While both offer multidimensional views of organizations, they are different with regard to their underlying objectives and key characteristics (Bozarth & McDermott, 1998). According to Bozarth and McDermott (1998), typologies are multidimensional models for ideal types and provide a generalizable grand theory and middle-range theories applicable to individual types. On the other hand, taxonomies are classification systems that categorize phenomena into mutually exclusive and exhaustive sets and are used for generating insight or to advance a predictive task (Bozarth & McDermott, 1998).

2. Conceptual Development

2.1 Market Segmentation and Customer Clustering

Amid market upheaval, companies are spending more money on improving their marketing strategies in order to enhance their competitiveness in the marketplace (Riasi & Pourmiri, 2015; Riasi, 2015a, 2015b). These marketing strategies are primarily focused on customer segmentation, increasing the level of customer satisfaction, and customer retention. In order to determine their preferences and thus improve marketing decision support, businesses can significantly benefit from analyzing large volumes of customer data which they have collected (Liu & Shih, 2005). Dividing customers into disparate clusters based on their similar buying behaviors, needs, demographics, and other parameters can lead to significant growth in profitability because it helps the companies to optimize their marketing practices. According to Zeithaml et al. (2001), marketing success is equivalent to generating maximum profits from a company's total set of customers when managerial resources are allocated to the groups of customers that can be cultivated most efficiently by the company. Identifying these groups of customers requires advanced customer clustering and market segmentation. Smith (1956) believed that market segmentation is based upon developments on the demand side of the market. Additionally, he believed that market segmentation represents a rational and precise adjustment of product and marketing effort to customer requirements (Smith, 1956). Customer classification and clustering enable the firms to group similar customers together and help managers to better understand the customers' needs; because it is much easier to identify and analyze the characteristics of groups of customers rather than studying each customer individually (Ansari & Riasi, 2016a).

2.2 Market Segmentation and Target Marketing

Market segmentation and target marketing are closely related concepts and they are sometimes used

interchangeably. Target marketing refers to the identification of a group of buyers sharing common needs or characteristics that a firm intends to serve (Kotler et al., 1991; Aaker et al., 2000). Target marketing has been the driving force behind the success of various well-known companies. In fact, target marketing provides the basis of a user positioning approach, which is a paramount branding strategy closely associated with a specific consumer (Aaker et al., 2000). Despite the similarities between market segmentation and target marketing, there is a big difference in the objectives of these two approaches. Market segmentation is concerned with dividing the market into different subsections based on the customers' expected response to marketing practices; while target marketing focuses on identifying the most interesting parts of the market and designing marketing strategies for customer acquisition in these markets. Therefore, in order to have a successful segmentation-based marketing strategy, a firm should follow three steps: The first step is market segmentation, in which the market is divided into different segments based on the characteristics of the customers and their needs. In the second step, different market segments are targeted using appropriate marketing strategies. In other words, the firm should investigate the customers in each segment and determine the best marketing approach based on the characteristics of that particular market segment and the firm's available resources. Finally, the last step is to determine the market position. In this step, the firm should be able to examine its competitiveness in each market segment by analyzing the customers' attitude toward the firm's brand in comparison to their attitude toward competing brands.

According to Gavett (2014), companies tend to make big mistakes in the market segmentation process and particularly when they start thinking about it: First of all, many companies who adapt segmentation-based marketing tend to uncover existing segments rather than creating their own segments. Second, many companies think that market segmentation should only focus on demographic characteristics and try to define market segments based on demographics. This is a big mistake, because in many cases, customers who have similar demographic characteristics may require different marketing approaches in order to be attracted to a brand. This indicates that companies should be aware that demographics and segmentation are two different things. Third, many companies forget to determine their ultimate objective from market segmentation and as a result of this, they end up spending a lot of money on performing unnecessary market segmentation. Gavett (2014) believes that before starting the market segmentation, firms should ask themselves why they want to segment and what decision they will make based on the information obtained from the process of market segmentation. Therefore, it is salient to consider these three pitfalls before starting to think about market segmentation and allocating a budget for this purpose. After considering these potential pitfalls, companies should decide whether they want to start segmenting by needs or behaviors. After figuring out whether the firm has the right brands, appropriate value proposition, and the right product line, then segmentation can be initiated (Gavett, 2014).

2.3 Characteristics of a Useful Segmentation

According to Boespflug (2013) market segments should be identifiable, reachable, significant, relevant, and understood properly in order to be useful. He believes that behind every successful firm there is a good market segmentation strategy that guides the company during every stage of the commercialization process. According to Boespflug (2013) a company that has a useful segmentation model will be able to increase its competitiveness and profitability. Gavett (2014) believes that a useful segmentation should have six characteristics: First, it should be identifiable, meaning that the characteristics of the customers in each market segment should be easily measurable. Second, it should be substantial, meaning that the segments should not be very small because small market segments are very costly to target. Third, it should be accessible, meaning that the segments should be created in such a way that the firm will be able to reach them via communication and distribution channels. Fourth, it should be stable, meaning that the characteristics used for segmenting the customers should not change very quickly in order to allow strategic planning. Fifth, it should be differentiable, meaning that people who are in the same segment should have similar characteristics and needs and their attributes should be different from those who belong to other segments. Sixth, it should be actionable, meaning that the firms should be able to provide products/services to the specified market segments.

3. Literature Review

Customer segmentation and clustering can be performed according to disparate criteria, including but not limited to profitability (Zeithaml et al., 2001), customer behavior (Ghazanfari et al., 2010; Gough & Sozou, 2005; Neal, 1998), degree of customer loyalty (Ansari & Riasi, 2016a; Cheng & Chen, 2009; Khajvand & Tarokh, 2011; McCarty & Hastak, 2007; Wei et al., 2012), purchase frequency (Ansari & Riasi, 2016a; Khajvand & Tarokh, 2011; McCarty & Hastak, 2007), purchase volume (Ghazanfari et al., 2010; Zeithaml et al., 2001), demographics (Dehghanpour & Rezvani, 2015; Gough & Sozou, 2005) and etc.

Wang et al. (2014) proposed a hierarchical analysis structure for customer clustering in order to optimize the logistics network. In thei study, the customers' characteristics were represented by using linguistic variables under major and minor criteria. In the next step, fuzzy integration methodology was used to map the sub-criteria into the higher hierarchical criteria based on the trapezoidal fuzzy numbers.

Newstead and D'Elia (2010) studied the concept of customer classification in the auto insurance industry by using log-linear Poisson regression analysis. They analyzed various variables including condition at the time of crash, vehicle type, crash injury severity, state, demographics, and color of the car and found that there is a clear statistically significant relationship between vehicle color and crash risk. Particularly, their results showed that compared to white vehicles, a number of dark colors were associated with higher crash risk. Their findings indicated that auto insurance companies should cluster their customers based on the color of their vehicles.

Zeithaml et al. (2001) introduced the concept of customer pyramid which is a methodology that enables a company to augment its profits by customizing its responses to distinct customer profitability tiers. According to the customer pyramid, consumers can be divided into four different segments, namely, platinum tier, gold tier, iron tier, and lead tier. Customers in the platinum tier are typically the most profitable customers of the company, who can be considered as heavy users of the product/services who are committed to the company. The gold tier includes the customers who are less profitable than those in the platinum tier and are not as loyal to the firm even though they are still considered as heavy users of the products/services of the firm. The important characteristic of these customers is that they want price discount which will limit the profit margins of the firm. Customers in the iron tier purchase products/services at the volume that is needed to utilize the firm's capacity but since they have relatively low loyalty and profitability for the firm they are not qualified for special treatment. Finally, the customers in the lead tier are the ones that are not at all profitable for the firm and even complain about the firm to others. These customers are very costly for the company and targeting them is not a good strategy.

Neal (1998) proposed a multidimensional segmentation approach based on three variables, namely, customer values, energy expenses, and customer energy management priorities. In this model, customer values were derived from an adaptive conjoint exercise which yielded eight segments based on value. Additionally, energy expenses were used to divide customer into five expenditure groups. Finally, customer energy management priorities were used to cluster the customer into seven different segments.

Ghazanfari et al. (2010) studied market segmentation in apparel industry using clustering algorithms. To analyze the clusters and to measure the value of the customers in each cluster, they used RFM (recency, frequency, and monetary value) model. The RFM is a very popular model for customer value analysis and it has been used by many scholars in order to perform market segmentation (Cheng & Chen, 2009; Khajvand & Tarokh, 2011; McCarty & Hastak, 2007). Since RFM analyzes the behavior of the consumers, it can be considered as a behavior-based model (Wei et al., 2012; Yeh et al., 2008).

Wei et al. (2012) adopted self-organizing maps (SOM) methodology in order to extend the RFM model. SOM is a neural network method which can be used for clustering problems, visualization, and market screening (Fish & Ruby, 2009; Hanafizadeh & Mirzazadeh, 2011; Hsu et al., 2009; Hung & Tsai, 2008; Kiang & Fisher, 2008; Wang, 2001; Wei et al., 2012). SOM also has the advantage of automatically detecting strong features in large data sets (Wei et al., 2012). Wei et al. (2012) used an extended model which is called LRFM (length, recency, frequency, and monetary value) to perform customer segmentation in a children's dental clinic in Taiwan.

Ansari and Riasi (2016a) combined fuzzy c-means clustering and genetic algorithms to cluster the customers of steel industry. They divided the customers into two clusters by using the variables of the LRFM model. After comparing the performance of the combined algorithm (i.e., fuzzy c-means clustering and genetic algorithms) with fuzzy c-means clustering, they found that the combined algorithm had a lower mean squared error (MSE) but a higher run time.

4. Methodology

The concept of customer segmentation and clustering in the banking industry has become an interesting area of research in recent years. Customers are now demanding higher quality services and more customized products from their banks (Popli & Vadgama, 2012), therefore, it is necessary for the banks to perform customer clustering in order to enhance the quality of their services. Since bank customers vary in their complexity and characteristics, they have become an interesting subject for many studies in the field of services marketing. The data set for this study contains information about different characteristics of 250 loan applicants of a commercial bank located in Isfahan province of Iran. The data were collected during a one-month period (from April 20, 2015 to May 20, 2015).

Fraley and Raftery (1998) define cluster analysis as "partitioning data into meaningful subgroups, when the number of subgroups and other information about their composition may be unknown". According to Fraley and Raftery (1998), clustering techniques range from heuristic methods to more formal procedures that are based on statistical models and usually follow either a hierarchical strategy or a strategy in which observations are relocated among tentative clusters. The studies by Gordon (1981), Hartigan (1975), Kaufman and Rousseeuw (1990), McLachlan and Basford (1988) and Murtagh (1985), provide a good introduction into clustering algorithms.

In this study, two-step cluster analysis using scalable cluster algorithm is performed using IBM SPSS Statistics 23. The main advantage of using this technique is the ability of the two-step cluster analysis to analyze both continuous and categorical variables. In the first step, which is known as pre-clustering, the observations are divided into various small clusters. In this step each observation is first considered as an independent cluster and then the algorithm considers adding new observations to each cluster and merges some of the small clusters. Pre-clustering is performed by generating a data structure called clustering feature (CF) tree, which contains the cluster centroids. The cluster feature CF_j of cluster j is:

$$CF_j = \{N_j, S_{Aj}, S_{Aj}^2, N_{Bj}\} \tag{1}$$

In equation 1, N_j represents the number of observations in cluster j, S_{Aj} is the sum of continuous attributes of the observations in cluster j, S_{Aj}^2 is the sum of the squared continuous variables of observations in cluster j, and $N_{Bj} = (N_{Bj1}, N_{Bj2}, ..., N_{BjK_B})$ is a $\sum_{k=1}^{K_B}(L_k - 1)$ –dimensional vector where the kth sub-vector is of $(L_k - 1)$ dimension, given by $N_{Bjk} = (N_{jk1}, ..., N_{jkL_k-1})$ in which N_{jkl} represents the number of observations in cluster j whose k^{th} categorical attribute takes the lth category, $l = 1, ..., L_k - 1$ (Chiu et al., 2001).

When two clusters (e.g., cluster s and j) are merged, it means that two corresponding sets of data points are gathered together to form a new cluster. The $CF_{<s,j>}$ for the new cluster can be calculated as:

$$CF_{<s,j>} = \{N_j + N_s, S_{Aj} + S_{As}, S_{Aj}^2 + S_{As}^2, N_{Bj+} + N_{Bs}\} \tag{2}$$

As described earlier, in pre-clustering step, observations are scanned sequentially and for each of them a decision is made whether the observation should be merged with an existing dense region or not. A CF tree is constructed during this process to store the summary statistics of the dense regions. The CF tree contains various nodes where each node represents one of the inputs. Each leaf in the CF tree is considered as an input for a final subsystem. For each observation, the algorithm initiates from the start node and in each step the observation finds the closest entry in the node and travels to the closest child node. This process continues recursively and the observation descends along the CF tree until it reaches a leaf node. After reaching a leaf node, the observation finds the closest entry and it is absorbed to the closest entry if its distance with the closest entry is within a threshold value. If the distance is not within the threshold, then it starts as a new leaf entry in the leaf node (Chiu et al., 2001). There is an optional outlier-handling step which can be used in the process of constructing the CF tree. According to Chiu et al. (2001), the outliers in this process are observations that cannot be fitted into any of the clusters. Observations in a leaf entry are considered as outliers if the number of records in the entry is less than a certain fraction of the size of the largest leaf entry in the CF tree (Chiu et al., 2001). The default value for this fraction is 25% in SPSS. Additionally, before constructing the CF tree, the algorithm searches for the potential irregular observations and discards them. After the CF tree is constructed the algorithm reexamines these observations based on the degree to which they will add to the size of the tree. Finally, observations which are not appropriate to be added to the tree will be considered as outliers. If the CF tree becomes larger than the maximum size which is allowed, it should be reconstructed using a larger threshold. As a result, the new CF tree which will be obtained will have a smaller size and hence will have more room for incoming observations. At the end of step one a collection of dense regions is identified and is stored in the leaf nodes of the CF tree (Chiu et al., 2001). The second step is called the clustering step, in which the dense regions identified in step one are merged in order to create the final clusters. If the number of clusters is not known, the method can automatically calculate the number of clusters by using Bayesian information criterion (BIC) and Akaike information criterion (AIC). If the number of clusters is pre-specified, observations will be clustered using hierarchical clustering such that the observations in each cluster will be as similar as possible. In this algorithm, the distance between the clusters is calculated using a log-likelihood-based distance measure. Specifically, the distance between two clusters is based on the decrease in log-likelihood as a result of merging the two clusters. The distance $d(j, s)$ between clusters j and s is defined as:

$$d(j, s) = \xi_j + \xi_s - \xi_{<j,s>} \tag{3}$$

Where $\xi_v = -N_v \left[\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right]$ for v = s, j and $\xi_{<j,s>}$ is defined in a similar way. Also $\hat{E}_{vk}$ is the entropy of the kth categorical attribute in cluster v, K^A is the total number of continuous attributes, K^B is the total number of categorical attributes, $\hat{\sigma}_{vk}^2$ is the variance of the kth continuous attribute in cluster v, and N_v is the number of observations in cluster v (Chiu et al., 2001).

In order to make sure that the number of clusters determined by this algorithm were appropriate, analysis of variance (ANOVA) was used. After comparing the results for three, four, and five clusters it was found that using five clusters produces the best results. After performing the cluster analysis, the significance of difference between clusters was tested using discriminant function analysis. Additionally, in order to further analyze the differences between the identified clusters, post hoc analysis was performed.

5. Results

As described in the methodology section, five clusters were identified for the bank customers using a two-step scalable clustering algorithm. Demographic characteristics of each cluster are summarized in table 1 and the values of the categorical attributes are displayed in table 2. According to table 2, the majority of business accounts belong to customers in cluster one, while the majority of personal accounts are held by customers in cluster five. Furthermore, almost all clusters include customers without any account closure history, however the majority of the customers with account closure history belong to the fifth cluster. Moreover, customers with accounts denominated in foreign currency are only observable in the clusters one and two.

Table 1. Demographic characteristics of survey respondents

| Cluster | Gender | | Age | | | | Monthly Income | |
	Male	Female	Below 30	30-40	41-50	Above 50	Less than 10 Million Rial	10 Million Rial or Higher
1	15.8%*	22.0%	0%	23.4%	44.2%	0%	0%	35.0%
2	12.6%	16.6%	0%	31.2%	0%	0%	41.0%	8.6%
3	26.6%	17.3%	43.6%	20.4%	25.4%	39.1%	13.9%	20.5%
4	20.7%	0%	0%	0%	0%	60.9%	8.0%	25.9%
5	24.3%	44.1%	56.4%	25.1%	30.4%	0%	37.1%	10.0%
Total	100%	100%	100%	100%	100%	100%	100%	100%

 * Each number indicates the percentage of the customers in the entire data set which have characteristic m and belong to cluster i.

Table 2. Values of the categorical attributes

| Cluster | Account Type | | Account Closure History | | Account Currency | | Location | |
	Business	Personal	No	Yes	Rial	Foreign Currency	Isfahan City**	Other Cities***
1	62.7%*	0%	18.5%	6.5%	0%	67.3%	0%	62.7%
2	37.3%	0%	11.6%	0%	2.4%	32.7%	0%	37.3%
3	0%	32.3%	27.5%	0%	31.5%	0%	42.9%	0%
4	0%	14.3%	12.2%	0%	13.9%	0%	20.5%	0%
5	0%	53.4%	30.2%	93.5%	52.2%	0%	36.6%	0%
Total	100%	100%	100%	100%	100%	100%	100%	100%

* Each number indicates the percentage of the customers in the entire data set which have attribute value n and belong to cluster i.

** Isfahan is the largest city in Isfahan province, it is the provincial capital and a major financial hub.

*** Indicates the percentage of customers from other cities in Isfahan province.

Table 3 displays the mean and standard deviation of continuous attributes which were used in this study. It can be seen that the most profitable customers are those that belong to clusters one and three, whereas the customers in the fifth cluster have the highest default risk and are the least profitable customers.

Table 3. Mean and standard deviation of continuous attributes

Cluster	Profitability (percent)		Loyalty (years)		Loan Amount (million Rial)		Account Balance (million Rial)		Default Risk (percent)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	69.3	2.9	10.0	3.4	153.0	12.0	20.0	0	10.7	2.9
2	51.2	3.5	3.0	0	47.0	5.0	30.0	0	50.6	3.5
3	70.5	3.2	1.3	0.4	147.0	8.0	10.0	0	28.2	3.2
4	59.9	3.4	1.0	0	81.0	9.0	20.0	0	43.1	3.4
5	39.9	1.7	4.0	0	20.0	0	30.0	3.0	64.2	1.7
Overall	58.1	2.9	3.9	0.8	89.6	6.8	22.0	0.6	39.4	2.9

As mentioned earlier, analysis of variance (ANOVA) was performed to analyze the differences between the clusters. The ANOVA results revealed that the F value is significant for all continuous variables ($p < 0.05$) which indicates that all clusters are significantly different.

Table 4. ANOVA results

Attribute	Cluster Root Mean Square	DF*	Error Root Mean Square	DF	F Value	P-Value
Profitability	13538.0	5	7.399	249	1830.0	<0.001
Loyalty	267.1	5	2.029	249	131.6	<0.001
Loan Amount	1792.0	5	0.868	249	2066.0	<0.001
Account Balance	31.2	5	0.018	249	1699.0	<0.001
Default Risk	13435.0	5	7.983	249	1722.0	<0.001

* DF stands for degrees of freedom.

By using discriminant function analysis, a confusion matrix was created which is displayed in table 5. The matrix displays the number of customers who have been correctly or incorrectly clustered by the algorithm. Under the perfect conditions, only the diagonal of the confusion matrix should have numbers in it; therefore, the confusion matrix displayed here is very close to the perfect condition, because it has only one misclassified observation.

Table 5. Confusion matrix

		Predicted Cluster				
		1	2	3	4	5
Actual Cluster	1	56	0	0	0	0
	2	0	22	0	0	0
	3	0	0	52	0	0
	4	0	0	0	23	0
	5	0	0	1	0	66

The ANOVA results indicated that the overall difference between the clusters was significant, but it did not indicate which specific clusters differed. Post hoc tests can be used in order to confirm where the differences occurred between the clusters. Since post hoc tests should only be run when the ANOVA results have been shown to be significant, therefore they can be used to further analyze the data for this study. Since the data set for this study met the homogeneity of variances assumption, Tukey's HSD test was used for post hoc analysis. The results of the post hoc analysis are summarized in Table 6.

Table 6. Results of post hoc analysis

Attribute	Cluster		Mean Difference	Standard Error	Significance
Profitability		2	29.92	0.732	<0.001
	1	3	8.87	0.585	<0.001
		4	19.39	0.722	<0.001
		5	39.41	0.574	<0.001
	2	3	-21.05	0.692	<0.001
		4	-10.53	0.811	<0.001
		5	9.51	0.683	<0.001
	3	4	10.52	0.681	<0.001
		5	30.54	0.522	<0.001
	4	5	20.02	0.672	<0.001
Loyalty		2	5.00	0.384	<0.001
	1	3	6.73	0.306	<0.001
		4	7.00	0.378	<0.001
		5	4.00	0.301	<0.001
	2	3	1.73	0.362	<0.001
		4	2.00	0.425	<0.001
		5	2.00	0.403	<0.001
	3	4	0.27	0.231	<0.001
		5	-2.73	0.273	<0.001
	4	5	-3.00	0.405	<0.001
Loan Amount		2	136.41	0.251	<0.001
	1	3	46.31	0.201	<0.001
		4	92.42	0.247	<0.001
		5	163.22	0.197	<0.001
	2	3	-90.11	0.237	<0.001
		4	-44.13	0.278	<0.001
		5	26.81	0.234	<0.001
	3	4	46.00	0.233	<0.001
		5	116.91	0.179	<0.001
	4	5	70.92	0.231	<0.001
Account Balance		2	-10.00	0.036	<0.001
	1	3	10.00	0.029	<0.001
		4	0.45	0.032	<0.001
		5	-9.61	0.029	<0.001
	2	3	20.00	0.034	<0.001
		4	10.00	0.041	<0.001
		5	20.00	0.038	<0.001
	3	4	-10.00	0.034	<0.001
		5	-19.62	0.026	<0.001
	4	5	-9.61	0.033	<0.001
Default Risk		2	-29.90	0.732	<0.001
	1	3	-8.87	0.585	<0.001
		4	-19.40	0.722	<0.001
		5	-39.40	0.574	<0.001
	2	3	21.05	0.692	<0.001
		4	10.53	0.811	<0.001
		5	-9.50	0.683	<0.001
	3	4	-10.50	0.681	<0.001
		5	-30.50	0.522	<0.001
	4	5	-10.60	0.621	<0.001

6. Specification of the Clusters

The five clusters which were identified using two-step scalable clustering are fully specified in this section.

First cluster or "favorite customers": all of the customers in this cluster are business account holders and the majority of them are privately-held companies. These customers are extremely profitable (mean profitability =

69.3%) for the bank. They are called "favorite customers" because they have the highest average loan amount (mean loan amount = 153 million Rial) compared to the other four clusters. On average the customers in this cluster do business with their bank for roughly 10 years, and rarely close their bank accounts. Furthermore, they do not obtain financing from other banks and have a relatively low default risk (mean default risk = 10.7%). The average monthly income for the primary account holder in this cluster is relatively high. Almost 44.2% of the primary account holders in this cluster are between 41 and 50 years old. Moreover, long-term saving accounts and checking accounts denominated in foreign currencies are highly preferred by these customers. They usually start their business with the bank by opening a checking account but extend their interactions with the bank over time by opening long-term saving accounts and obtaining financing.

Second cluster or "non-creditworthy customers": similar to the favorite customers (i.e., first cluster) the non-creditworthy customers are also business account holders, however the majority of the account holders in this group are publicly-held companies. These customers usually receive small loans (average loan amount = 47 million Rial) and have a very high default risk (mean default risk = 50.6%). All of the primary account holders in this group are between 30 and 40 years old and have relatively low monthly incomes. The combination of high default risk, low monthly income, and lack of experience of primary account holders have contributed to low degrees of profitability for these customers. Finally, these customers have average degrees of loyalty and never close their bank accounts.

Third cluster or "creditworthy customers": the customers in this cluster are personal account holders with very high monthly incomes. Since these customers are the most profitable (mean profitability = 70.5%) account holders they should be the target of marketing strategies. The majority of the customers in this cluster are highly educated engineers and physicians. Since these customers have a relatively low default risk (mean default risk = 28.2%) and are highly profitable for the bank, they receive high amounts of financing (mean loan amount = 147 million Rial). These customers are mostly interested in long-term saving accounts and governmental or private bonds that are distributed by the commercial banks. All of these customers hold bank accounts denominated in domestic currency and have no account closure history. Finally, it was found that there were more male account holders (26.6% of total male account holders in the sample) in this group than female account holders (17.3% of total female account holders in the sample).

Fourth cluster or "passers": the customers in this group are the least loyal (mean loyalty = 1 year) account holders and have very low account balances (mean account balance = 20 million Rial). Additionally, they are more inclined toward short-term saving accounts and checking accounts and are always trying to obtain relatively large loans. The majority of these customers are employees of publicly- and privately-held companies and are extremely sensitive to interest rates. These customers have a high default risk (mean default risk = 43.1%) and are extremely willing to change their banks if a new bank offers them loans with lower interest rates.

Fifth cluster or "friends": these customers have the highest default risk (mean default risk = 64.2%) mainly due to their extremely low monthly incomes. Therefore, they are willing to accept high interest rates. The profitability of these customers is extremely low (mean profitability = 39.9%) for the banks and as a result the banks do not give them loans with large denominations (mean loan amount = 20 million Rial). Although these customers are not very profitable for the banks, they have high degrees of loyalty (mean loyalty = 4 years) and are willing to advertise for the bank services which they have received when talking with their friends or family. This cluster is labeled as "friends" due to these customers' high loyalty and willingness to advertise for the bank services. Another characteristic of the customers in this cluster is their relatively low education. More than half of the customers in this group had a high school degree or below which makes them the least educated cluster. Finally, the account holders in this cluster are mostly interested in short-term saving accounts and checking accounts.

7. Discussion and Conclusions

According to Mass et al. (2008), customer value in financial services industry is a synonym for customer equity, and the true sense of customer value is the benefits of a product, service or relationship as perceived by the customer. However, Mass et al. (2008) believe that the importance of customer value to the financial services industry is seldom realized. In order to evaluate the value of bank customers and to determine their impact on the bank's performance, it is necessary to identify their key characteristics by using customer clustering. Using customer clustering enables the banks to identify their most profitable customers and to design marketing strategies for each group of customers based on their attributes.

This study used two-step scalable clustering algorithm in order to divide the customers of banking industry into different clusters. Two-step clustering has various advantages over similar algorithms: First, two-step clustering

has the ability to analyze both categorical and continuous variables. Second, it is relatively efficient when used for analyzing big data sets. However, since this algorithm discards observations with missing values, it may not be the best choice when the data set contains a lot of missing values.

The findings indicated that loan amount, profitability, default risk, account balance, and customer loyalty had the most salient role in determining the customer clusters. This is consistent with Zeithaml et al. (2001) who suggested that customer profitability is one of the most important attributes which can be used for customer segmentation and clustering. These findings are also in line with Newstead and D'Elia (2010) who found that the customer risk is a primary factor for customer clustering in the insurance industry. Furthermore, consistent with Ansari and Riasi (2016a), Cheng and Chen (2009), Khajvand and Tarokh (2011), McCarty and Hastak (2007), and Wei et al. (2012) it was found that customers' degree of loyalty can be used as a principal determinant of customer clusters. The most important contribution of this study is the incorporation of a wide range of factors that can potentially affect customer clustering in the analysis, whereas, the majority of previous studies only focused on a limited number of variables in order to determine the customer clusters. Specifically, the customer clustering in this study was performed by using demographic variables, profitability, loan amount, default risk, account balance, loyalty, account type, account closure history, customer location, and account currency. Analyzing disparate factors in order to perform customer clustering enables the managers to obtain more precise customer clusters.

The results revealed that 56 customers belonged to the first cluster, 22 customers belonged to the second cluster, 52 customers belonged to the third cluster, 23 customers belonged to the fourth cluster, and 67 customers belonged to the fifth cluster. According to the results of post hoc analysis (table 6), all five clusters are different from each other based on the profitability attribute. Additionally, the customers in the first cluster (i.e., favorite customers) were found to be extremely different from those in the fifth cluster (i.e., friends) when compared on the basis of loan amount. The significance of other differences between the clusters are displayed in table 6. Figure 1 depicts the differences between the five customer clusters which were identified in this study.
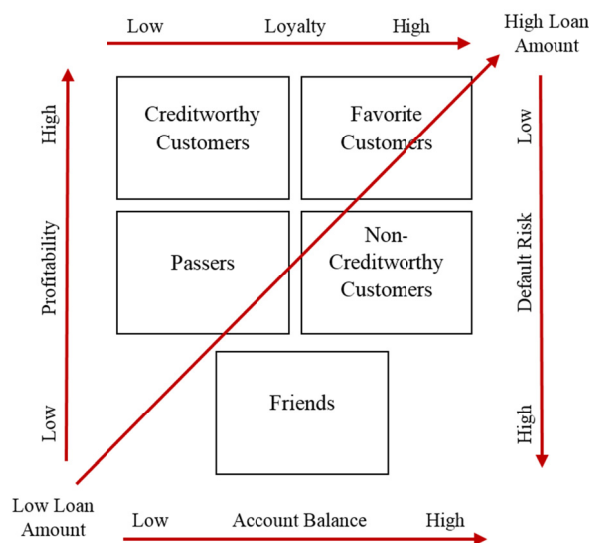

Figure 1. Differences between different clusters of bank customers

The differences which were observed between these five clusters of bank customers accentuate the importance of customer clustering and market segmentation in the financial services industry. Almost all financial institutions collect tons of valuable customer data every day but the majority of them do not have the necessary resources for analyzing these precious data. In addition to the lack of resources for analyzing these customer data, a large number of financial institutions do not perform customer clustering because they simply ignore the importance of market segmentation. It is crystal clear that performing market segmentation and customer clustering can help financial institutions to better understand the actual needs of their customers and to provide better services to them. Customer clustering can help financial institutions to augment their competitiveness by shifting from traditional marketing strategies to target marketing and segmentation-based marketing approaches.

The fact that the current study used data from customers of a single bank is the main limitation of this study. Further research is required in order to confirm the robustness of the relationships which were observed. For instance data from different commercial banks in different countries can be compared and the existence of the suggested five customer clusters can be examined. However, the authors believe that the five clusters of bank customers which were introduced in this study, will probably exist in other commercial banks with minor modifications.

References

Aaker, J. L., Brumbaugh, A. M., & Grier, S. A. (2000). Nontarget Markets and Viewer Distinctiveness: The Impact of Target Marketing on Advertising Attitudes. *Journal of Consumer Psychology*, *9*(3), 127-140.

Ansari, A., & Riasi, A. (2016a). Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms. *Working Paper*

Ansari, A., & Riasi, A. (2016b). Modelling and Evaluating Customer Loyalty Using Neural Networks: Evidence from Startup Insurance Companies. *Future Business Journal, 4.*

Ansari, A., & Riasi, A. (2016c). An Investigation of Factors Affecting Brand Advertising Success and Effectiveness. *International Business Research*, *9*(4), 20-30. http://dx.doi.org/10.5539/ibr.v9n4p20

Boespflug, S. (2013). Successful Marketing Segmentation = More Effective and Profitable Marketing. Retrieved from http://www.askpivot.com/successful-marketing-segmentation/

Bozarth, C., & McDermott, C. (1998). Configurations in manufacturing strategy: A review and directions for future research. *Journal of Operations Management*, *16*(4), 427-439. http://dx.doi.org/10.1016/S0272-6963(98)00022-9

Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory", *Expert Systems with Applications*, *36*(3), 4176-4184. http://dx.doi.org/10.1016/j.eswa.2008.04.003

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 263-268).

Dehghanpour, A., & Rezvani, Z. (2015). The profile of unethical insurance customers: A European perspective. *International Journal of Bank Marketing*, *33*(3), 298-315. http://dx.doi.org/10.1108/IJBM-12-2013-0143

Dess, G. G., Newport, S., & Rasheed, A. M. (1993). Configuration research in strategic management: Key issues and suggestions. *Journal of Management*, 19(4), 775-795. http://dx.doi.org/10.1016/0149-2063(93)90027-K

Fish, K., & Ruby, P. (2009). An artificial intelligence foreign market screening method for small businesses. *International Journal of Entrepreneurship*, *13*, 65-81.

Fornell, C. (1992). A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, *56*(1), 6-21. http://dx.doi.org/10.2307/1252129

Fornell, C., & Wernerfelt, B. (1987). Defensive marketing strategy by customer complaint management: A theoretical analysis. *Journal of Marketing Research*, *24*(4), 337-346. http://dx.doi.org/10.2307/3151381

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, *41*(8), 578-588. http://dx.doi.org/10.1093/comjnl/41.8.578

Gallarza, M. G., & Saura, I. G. (2006). Value dimensions, perceived value, satisfaction and loyalty: An investigation of university students' travel behavior. *Tourism Management*, *27*(3), 437-452. http://dx.doi.org/10.1016/j.tourman.2004.12.002

Gavett, G. (2014). What You Need to Know about Segmentation. Retrieved from https://hbr.org/2014/07/what-you-need-to-know-about-segmentation/

Ghazanfari, M., Malek Mohammadi, S., & Alizadeh, S. (2010). Customer Segmentation in Clothing Exports Based on Clustering Algorithm. *Iranian Business Studies Quarterly*, *14*(56), 59-86.

Gordon, A. D. (1981). *Classification: Methods for the Exploratory Analysis of Multivariate Data*. Chapman & Hall, London, U.K.

Gough, O., & Sozou, P. D. (2005). Pensions and retirement savings: Cluster analysis of consumer behaviour and attitudes. *International Journal of Bank Marketing*, *23*(7), 558-570. http://dx.doi.org/10.1108/02652320510629917

Hambrick, D. C. (1983). An empirical typology of mature industrial-product environments. *Academy of Management Journal*, *26*(2), 213-230. http://dx.doi.org/10.2307/255971

Hanafizadeh, P., & Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method–ADSL market of a telecommunication company. *Expert Systems with Applications*, *38*(1), 198-205. http://dx.doi.org/10.1016/j.eswa.2010.06.045

Hartigan, J. A. (1975). *Clustering Algorithms*. Wiley, New York City, NY.

Hsu, S. H., Hsieh, J. P. A., Chih, T. C., & Hsu, K. C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, *36*(4), 7947-7951. http://dx.doi.org/10.1016/j.eswa.2008.10.065

Hung, C., & Tsai, C. F. (2008). Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. *Expert Systems with Applications*, *34*(1), 780-787. http://dx.doi.org/10.1016/j.eswa.2006.10.012

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data*. Wiley, New York City, NY.

Kiang, M. Y., & Fisher, D. M. (2008). Selecting the right MBA schools–An application of self-organizing map networks. *Expert Systems with Applications*, 35(3), 946-955. http://dx.doi.org/10.1016/j.eswa.2007.08.053

Kim, K. H., Kim, K. S., Kim, D. Y., Kim, J. H., & Kang, S. H. (2008). Brand equity in hospital marketing. *Journal of Business Research*, *61*(1), 75-82. http://dx.doi.org/10.1016/j.jbusres.2006.05.010

Khajvand, M., & Tarokh, M. J. (2011). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science*, *3*, 1327-1332. http://dx.doi.org/10.1016/j.procs.2011.01.011

Kotler, P., Armstrong, G., & Starr, R. G. (1991). *Principles of Marketing*. Prentice-Hall, Englewood Cliffs, NJ.

Kumar, V., Dalla Pozza, I., & Ganesh, J. (2013). Revisiting the satisfaction–loyalty relationship: Empirical generalizations and directions for future research. *Journal of Retailing*, *89*(3), 246-262. http://dx.doi.org/10.1016/j.jretai.2013.02.001

Liu, D. R., & Shih, Y. Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, *42*(3), 387-400. http://dx.doi.org/10.1016/j.im.2004.01.008

Maas, P., Graf, A., & Bieck, C. (2008). *Trust, transparency and technology: European customers' perspectives on insurance and innovation*. IBM Global Services, Somers, NY.

McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, *60*(6), 656-662. http://dx.doi.org/10.1016/j.jbusres.2006.06.015

McLachlan, G. J., & Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York City, NY.

Meyer, A. D., Tsui, A. S., & Hinings, C. R. (1993). Configurational approaches to organizational analysis. *Academy of Management Journal*, *36*(6), 1175-1195. http://dx.doi.org/10.2307/256809

Miller, D. (1986). Configurations of strategy and structure: Towards a synthesis. *Strategic Management Journal*, *7*(3), 233-249. http://dx.doi.org/10.1002/smj.4250070305

Miller, D. (1987). Strategy making and structure: Analysis and implications for performance. *Academy of Management Journal*, *30*(1), 7-32. http://dx.doi.org/10.2307/255893

Miller, D., & Friesen, P. H. (1977). Strategy−making in context: Ten empirical archetypes. *Journal of Management Studies*, *14*(3), 253-280. http://dx.doi.org/10.1111/j.1467-6486.1977.tb00365.x

Miller, D., & Friesen, P. H. (1984). *Organizations: A Quantum View*. Prentice-Hall, Englewood Cliffs, NJ.

Mintzberg, H. (1973). Strategy-making in three modes. *California Management Review*, *16*(2), 44-53. http://dx.doi.org/10.2307/41164491

Mintzberg, H. (1978). Patterns in strategy formation. *Management Science*, *24*(9), 934-948. http://dx.doi.org/10.1287/mnsc.24.9.934

Moslehi, N. (2012). *Using LRFM model for clustering customers based on their life cycles in order to improve customer relationship management* (Master's thesis). Ferdowsi University of Mashhad, Mashhad, Iran.

Murtagh, F. (1985). Multidimensional clustering algorithms. *CompStat Lectures*, Physica Verlag, Heidelberg, Germany.

Neal, W. D. (1998). Multidimensional Segmentation", *Canadian Journal of Marketing Research*. Retrieved from http://www.sdrnet.com/article2.html

Newstead, S., & D'Elia, A. (2010). Does vehicle colour influence crash risk? *Safety Science*, *48*(10), 1327-1338. http://dx.doi.org/10.1016/j.ssci.2010.05.001

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Understanding customer expectations of service. *Sloan Management Review*, *32*(3), 39-48.

Paswan, A. K., Guzmán, F., & Blankson, C. (2012). Business to business governance structure and marketing strategy. *Industrial Marketing Management*, *41*(6), 908-918. http://dx.doi.org/10.1016/j.indmarman.2011.10.011

Pont, M., & McQuilken, L. (2005). An empirical investigation of customer satisfaction and loyalty across two divergent bank segments. *Journal of Financial Services Marketing*, *9*(4), 344-359. http://dx.doi.org/10.1057/palgrave.fsm.4770165

Popli, G. S., & Vadgama, C. (2012). Role of service marketing in the growth of banking sector in India. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2055075

Reichheld, F. F. (1992). Loyalty-based management. *Harvard business review*, *71*(2), 64-73.

Reynolds, K. E., & Beatty, S. E. (1999). Customer benefits and company consequences of customer-salesperson relationships in retailing. *Journal of Retailing*, *75*(1), 11-32. http://dx.doi.org/10.1016/S0022-4359(99)80002-5

Riasi, A. (2015a). Barriers to international supply chain management in Iranian flower industry. *Management Science Letters*, *5*(4), 363-368. http://dx.doi.org/10.5267/j.msl.2015.2.005

Riasi, A. (2015b). Competitive Advantages of Shadow Banking Industry: An Analysis Using Porter Diamond Model. *Business Management and Strategy*, *6*(2), 15-27. http://dx.doi.org/10.5296/bms.v6i2.8334

Riasi, A., & Pourmiri, S. (2015). Effects of online marketing on Iranian ecotourism industry: Economic, sociological, and cultural aspects. *Management Science Letters*, *5*(10), 915-926. http://dx.doi.org/10.5267/j.msl.2015.8.005

Rosenberg, M. (1968). *The logic of survey analysis*. New York City, NY: Basic Books.

Salmones, M. M. G., Perez, A., & Bosque, I. R. (2009). The social role of financial companies as a determinant of consumer behavior. *International Journal of Bank Marketing*, *27*(6), 467-485. http://dx.doi.org/10.1108/02652320910988339

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *The Journal of Marketing*, *21*(1), 3-8. http://dx.doi.org/10.2307/1247695

Veloutsou, C., Daskou, S., & Daskou, A. (2004). Are the determinants of bank loyalty brand specific? *Journal of Financial Services Marketing*, *9*(2), 113-125. http://dx.doi.org/10.1057/palgrave.fsm.4770146

Wang, S. (2001). Cluster analysis using a validated self-organizing method: Cases of problem identification. *Intelligent Systems in Accounting, Finance and Management*, *10*(2), 127-138. http://dx.doi.org/10.1002/isaf.198

Wang, Y., Ma, X., Lao, Y., & Wang, Y. (2014). A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization. *Expert Systems with Applications*, *41*(2), 521-534. http://dx.doi.org/10.1016/j.eswa.2013.07.078

Wei, J. T., Lin, S. Y., Weng, C. C., & Wu, H. H. (2012). A case study of applying LRFM model in market segmentation of a children's dental clinic. *Expert Systems with Applications*, *39*(5), 5529-5533. http://dx.doi.org/10.1016/j.eswa.2011.11.066

Yeh, I. C., Yang, K. J., & Ting, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, *36*(3), 5866-5871. http://dx.doi.org/10.1016/j.eswa.2008.07.018

Zeithaml, V. A., Rust, R. T., & Lemon, K. N. (2001). The customer pyramid: Creating and serving profitable customers. *California Management Review*, *43*(4), 118-142. http://dx.doi.org/10.2307/41166104

Copyrights