A customer segmentation approach in commercial banks

Conference Paper *in* AIP Conference Proceedings · October 2018

DOI: 10.1063/1.5064881

2 authors:

Vesela Mihova
University of Ruse Angel Kanchev

Velisar Pavlov
University of Ruse Angel Kanchev

A Customer Segmentation Approach in Commercial Banks

V. Mihova[a)] and V. Pavlov[b)]

University of Ruse, 8 Studentska str., 7017 Ruse, Bulgaria

Corresponding author: [a)]vmicheva@uni-ruse.bg
[b)]vpavlov@uni-ruse.bg

Abstract. The results of various recent analyses of commercial banking trends show that proper classification of borrowers is fundamental for the development of a successful business. The increasing competition in both banking system and non-bank institutions requires the use of modern market strategies and individual customer approaches. One of the main priorities in the banking sector is to improve customer segmentation and take it into account in the design and distribution of new products. A common tool to improve competitiveness is the designing of a special range of products and services targeting loyal customers or offering them special discounts for existing products. This represents the so-called "loyalty program", which includes the issuance of various types of cards for such customers. Three clusters (segments) of loyal borrowers: "platinum," "gold," and "silver," are identified in the present work, using K-means clustering. A database of 100 borrowers from a commercial bank branch that took secured consumer loans is analyzed. The clients are defined as loyal based on their credit history (they have less than 3 missed payments for the last year). Three variables are used for their segmentation. Initially, the initial segmentation variables are taken as input data for the analysis, and further study on standardized segmentation variables is carried out. The potential segmentation strategies are formulated depending on the leading segmentation variable. A comparative analysis of the results of both methods examined (with initial and standardized segmentation variables) and of a two-step clustering (obtained in a previous study from one of the authors) is made within each of the strategies. It is specified which type of cluster analysis suits best to each of the strategies.

INTRODUCTION

An analysis of the characteristics and dynamics of the volume of bank credit portfolios for the period 2007-2015 of the Association of Banks in Bulgaria (ABB) shows qualitative changes and improvements in the lending terms after the global financial crisis [1]. According to the analysis the normative changes, the changes in the interest rates and the dropping of certain types of fees and commissions in 2014 and 2015 have led to significant market turbulence and increased consumer interest. The analysis also points out that the increasing competition in both the banking system and the non-bank financial institutions requires the use of modern market strategies and individual customer approaches.

An important part of the modern market strategies is the development of statistical models, used to estimate this risk associated with the approval of people applying for certain products (loans, credit cards, *etc.*). These models are used to assess how the behavior of the approved customers (good or bad customers) depends on the data submitted in the application form for the product. The resulting relation could be used to assess new customers even at the time of application. This methodology is described by the authors in [2, 3]. Risk models also have a key role in general insurance [4, 5, 6, 7].

Another important part of the modern market strategies includes customer segmentation. The importance of customer segmentation is confirmed by many different surveys. An example of such confirmation are the results of a global survey of retail banking trends in which senior bank managers rank current retail banking priorities for their institutions [8]. These results show that one of the main priorities (and a condition for a successful future) is to improve customer segmentation and take it into account in the design and distribution of new products.

Market segmentation provides banking management with the opportunity to develop appropriate strategies for more effective investment based on in-depth knowledge of customer groups [9]. One of the purposes of segmentation is to determine the attitude of individual customer groups to a particular product or service. It enables financial institutions to regulate the supply of services in line with current and potential market users and to develop long-term market strategies. Segmentation of the banking market is nowadays a key factor for the

Application of Mathematics in Technical and Natural Sciences
AIP Conf. Proc. 2025, 030003-1–030003-8; https://doi.org/10.1063/1.5064881
Published by AIP Publishing. 978-0-7354-1745-8/$30.00

development of a successful business [10].

Customer segmentation is a complex process requiring both knowledge and skills, as well as experience in the sale of financial products, market savvy and intuition. The purpose of segmentation is not in the identification of any user groups in a given market, but rather in the search for such users that have or may have significantly different financial service requirements [9].

In the light of the above, a common tool to improve competitiveness is the designing of a special range of products and services targeting loyal customers or offering them special discounts for existing products. This represent the so-called "loyalty programs" (or reward programs). Loyalty programs are structured marketing strategies designed by merchants to encourage customers to continue to shop at or use the services of businesses associated with each program [11]. These programs include the issuance of various types of cards for such customers (platinum, gold, silver, etc.), similar to a credit card, debit card, or digital card that identify the card holder as a participant in a loyalty program. Loyalty cards (both physical and digital) relate to the loyalty business-model. By presenting such a card, customers typically earn the right either to a discount on the current purchase/service, or to an allotment of points that they can use for future purchases/services.

The banking sector has already immersed itself in work on loyalty programs. To increase the number of loyal customers, banks started experimenting with various methods and tools. They usually offer loyalty cards in partnership with another business – airline companies, automobile concerns, fuel manufacturers, etc. In fact, credit card points-earning programs are one of the most active sectors in the reward process.

A survey on emerging financial consumer trends (see [12]), conducted by CGI in the USA, Canada, France, Germany, Sweden and the UK, shows that number one in the list of bank customers' top needs belongs to loyalty rewards such as credit card points, discounts for purchases and other bonuses. It turned out that 81% of respondents expect to be valued for their total spending and rewarded for their loyalty.

Another study (see [13]) found that rewards and incentives can influence behavior as well as change brand perception. The research included several hundred U.S.A.-based consumers who had participated in some kind of loyalty, incentive, or reward program. Fifty percent of respondents reported that these programs drive their purchasing behavior. The study found that personalized incentives make a company stand out as smart (according to 42% of respondents), ambitious (35%), and unique (35%) as well as exciting (31%), caring (30%) and informative (30%). 56% of respondents said that receiving a personalized incentive would improve their consideration of the brand.

In an increasingly competitive environment where consumers have a huge array of banking options, are the banks successfully utilizing the full value of their loyalty programs? Forrester Consulting polled 150 directors and brand managers (including 50 within the banking industry) to better understand how those overseeing loyalty programs craft their strategy for optimum results [14]. The study found that loyalty marketers in the financial industry have high aspirations for their programs but are not universally satisfied with performance. However, those whose loyalty programs integrated multiple channels and high-engagement mechanisms reported greater satisfaction with their program performance.

Across the globe loyalty programs are increasingly finding the need to outsource strategic and operational aspects of their programs, given the size and complexity a loyalty program entails. Program managers are typically agencies with specialist skilled in loyalty consulting, creativity, communication, data analytics, loyalty software, and back end operations. They usually form customer segments based on industry practices, accumulated knowledge and statistical methods for data analysis, such as cluster analysis [15, 16, 17, 18].

In the present work we consider a loyalty program for commercial bank customers, which includes the issuing of three types of cards – "platinum," "gold," and "silver." In this context the loyal borrowers are divided into three clusters (segments), using *K*-means clustering.

PROBLEM FORMULATION

A database of 100 borrowers from a commercial bank branch that took secured consumer loans is used for the purposes of the study. The clients are defined as loyal based on their credit history (they have less than 3 missed payments for the last year).

The purpose of this analysis is to divide the 100 objects into 3 groups using the following 3 variables:

- *Loan Amount* (in euro);
- *Time with Bank* (in months);
- *Worst Status Last 12 Months* – shows the number of missed payments per customer in the last 12 months (0 – "0 missed payments," 1 – "1 missed payment," 2 – "2 missed payments").

Time with Bank and *Loan Amount* are continuous variables.

K-means clustering is used for the purposes of the study, with the initial segmentation variables taken for input data. The method records the distance of each unit to the individual cluster centers, and the nearest distance determines whether the unit belongs to the corresponding cluster [15]. Since it is a non-hierarchical clustering, the number of clusters is predetermined. Clusters centers are either known or evaluated on the data, and may remain constant or updated in the analysis process. This procedure requires the use of quantitative variables, therefore the variable *Worst Status Last 12 Months* is taken as a numeric variable.

Whenever we use a statistical procedure that calculates distances, we have to worry about the impact of the different units in which variables are measured [16]. Variables that have large values will have a large impact on the distance compared to variables that have smaller values. Therefore, a further study on standardized segmentation variables (with a mean of 0 and a standard deviation of 1) is carried out. A comparative analysis of the results of both methods examined (with initial and standardized segmentation variables) and of a two-step clustering (obtained in [19]) is made within each of the strategies.

The analysis is done using SPSS (see [15]).

K-MEANS CLUSTER ANALYSIS WITH INITIAL VARIABLES

Table 1 shows the final cluster centers obtained by the *K*-means cluster analysis, using the initial segmentation variables. It could be seen form the table that Cluster 1 includes the clients with the highest average loan amount (92 667 EUR), average of 0 missed payments and average time with bank 92 months. The clients from Cluster 2 rank second on the average loan amount and on the average number of missed payments for the last 12 months. The borrowers from Cluster 3 have average time with bank close to those of Cluster 1 but are far behind them on the other two benchmarks.

TABLE 1. K-means cluster analysis – final cluster centers

Cluster No / Variable	**1**	**2**	**3**
Loan Amount	92667	50150	3165
Time with Bank	92	69	90
Worst Status Last 12 Months	0.00	0.25	0.44
Number of cases	**3**	**4**	**93**

A good strategy is to encourage the clients from Cluster 1 as "platinum clients," Cluster 2 borrowers to be chosen as "gold clients," and Cluster 3 to remain the weakest group of loyal customers – "silver." Thus, very few borrowers (3% and 4%, respectively, of the total number of cases) will fall in the first two groups, and will be more precisely awarded that type of clients the bank seeks (without missed payments and with large loan amount).

According to the results from the analysis of variance (Table 2), the loan amount has the biggest influence in the clusters formation (*F*-value = 959.9), while the time with bank has the least influence (*F*-value = 0.216).

TABLE 2. ANOVA table

Variable	**Cluster**		**Error**		***F*-value**	**Significance level**
	Mean square	**df**	**Mean square**	**df**		
Loan Amount	1.539×10^{10}	2	1.603×10^{7}	97	959.900	0.000
Time with Bank	850.590	2	3928.970	97	0.216	0.806
Worst Status Last 12 M	0.343	2	0.574	97	0.597	0.552

The great influence of the loan amount in clusters formation is determined by the statistical procedure for calculating the distances to the individual cluster centers. In this procedure, input variables with larger variances have commensurately greater influence on the results.

K-means clustering is also very sensitive to outliers, as they are usually chosen as initial cluster centers, forming their own tiny groups. This is often cited as a reason to exclude the outliers from the analysis. In this type of customer segmentation, however, the outliers may be the most important customers.

The most appropriate clustering method should be determined depending on creditor goals. If the loan amount is crucial for determining the individual segments, it is appropriate to use *K*-means clustering as it assigns the biggest weight to this variable. If the large impact of loan amount on cluster formation is not a sought-after effect, then two-step clustering could be used or variables used for segmentation with *K*-means clustering could be standardized.

K-MEANS CLUSTER ANALYSIS WITH STANDARDIZED VARIABLES

Table 3 represents the results obtained by the *K*-means clustering using standardized variables*, i.e.*, variables with a mean of 0 and a standard deviation of 1.

TABLE 3. K-Means Cluster Analysis with standardized variables – final cluster centers

Cluster No / Variable	1	2	3
Loan Amount (Z)	-0.215	3.356	-0.280
Time with Bank (Z)	-0.955	-0.165	0.711
Worst Status Last 12 M (Z)	0.565	-0.367	-0.360
Number of cases	**39**	**7**	**54**

Table 4 represents the mean values of the initial segmentation variables for the respective clusters. It can be seen from the table that Cluster 1 includes loyal borrowers with a small average loan amount (3 845 EUR), the smallest value of average time with bank (30 months) and the highest average number of missed payments for the last 12 months (0.85). Customers in this cluster form the weakest group of loyal borrowers – the "silver" one.

TABLE 4. Mean values of the initial segmentation variables

Cluster No / Variable	1	2	3
Loan Amount	3845	68371	2674
Time with Bank	30	79	133
Worst Status Last 12 Months	0.85	0.14	0.15
Number of cases	**39**	**7**	**54**

Clients from clusters 2 and 3 have close to one another average numbers of missed payments. The borrowers with the highest average loan amount (68 371 EUR) fall in Cluster 2, and those with the highest average time with bank (133 months) fall in Cluster 3. If the bank wants to encourage the long-term clients with a good credit history, the borrowers who fall in Cluster 3 would be chosen for "platinum clients." If the goal is to encourage good customers with large loans, they would choose for "platinum clients" those from Cluster 2, which are the smallest group (7% of the observations). The largest percentage of observations (54%) falls in Cluster 3.

TABLE 5. ANOVA table

Variable	Cluster		Error		*F*-value	Significance level
	Mean square	df	Mean square	df		
Loan Amount (Z)	42.423	2	0.146	97	290.721	0.000
Time with Bank (Z)	31.509	2	0.371	97	84.941	0.000
Worst Status Last 12 Months (Z)	10.198	2	0.810	97	12.584	0.000

According to the results from the analysis of variance (Table 5), the loan amount again has the biggest influence in the clusters formation (*F*-value = 290.721), and the worst status for the last 12 months has the least influence (*F*-value = 12584).

However, the use of standardized segmentation variables in the *K*-means clustering allows for a choice between strategies to encourage good payers with a large loan amount or long-standing customers with a good credit history.

It is important to note that the *F*-test results are only descriptive, as the clusters are formed to maximize the distance (in the multidimensional space) between the observations in the individual segments. The differences between the *F*-values allow the formulation of conclusions about the role of individual segmentation variables in cluster formation.

COMPARATIVE ANALYSIS OF THE RESULTS

Let us look at the possible strategies depending on which variable plays the most important role in defining the platinum group:

- *Strategy 1:* The smallest number of missed payments in the last 12 months and a large loan amount;
- *Strategy 2:* The smallest number of missed payments in the last 12 months and long-standing bank customers;
- *Strategy 3:* The largest loan amount and long-standing bank customers;
- *Strategy 4:* The largest loan amount and a small number of missed payments in the last 12 months;
- *Strategy 5:* The longest-standing bank customers with a large loan amount;
- *Strategy 6:* The longest-standing bank customers with a small number of missed payments in the last 12 months.

Table 6 contains a comparison between the results of both methods examined (*K*-means clustering with initial and standardized segmentation variables) and of a two-step clustering (obtained in [19]) within Strategy 1 (the smallest number of missed payments in the last 12 months and a large loan amount). The table shows that the most appropriate segmentation method for this strategy is *K*-means clustering with initial segmentation variables.

TABLE 6. Strategy 1: Comparative analysis of the clustering methods

Clustering Method	Variable	Platinum Clients	Gold Clients	Silver Clients
K-means, initial variables	Worst Status Last 12 Months	0	0.25	0.44
	Loan Amount	92667	50150	3165
	Time with Bank	92	69	90
	Number of cases	**3**	**4**	**93**
K-means, standardized variables	Worst Status Last 12 Months	0.14	0.15	0.85
	Loan Amount	68371	2674	3845
	Time with Bank	79	133	30
	Number of cases	**7**	**54**	**39**
Two-Step Clustering	Worst Status Last 12 Months	0	1	2
	Loan Amount	2963	31028	4689
	Time with Bank	100	93	41
	Number of cases	**68**	**16**	**16**

It is obtained in Table 7 that the two-step clustering and *K*-means clustering with standardized variables are appropriate to use if Strategy 2 (smallest number of missed payments and long-standing customers) is selected.

TABLE 7. Strategy 2: Comparative analysis of the clustering methods

Clustering Method	Variable	Platinum Clients	Gold Clients	Silver Clients
K-means, initial variables	Worst Status Last 12 Months	0	0.25	0.44
	Time with Bank	92	69	90
	Loan Amount	92667	50150	3165
	Number of cases	**3**	**4**	**93**
K-means, standardized variables	Worst Status Last 12 Months	0.15	0.14	0.85
	Time with Bank	133	79	30
	Loan Amount	2674	68371	3845
	Number of cases	**54**	**7**	**39**
Two-Step Clustering	Worst Status Last 12 Months	0	1	2
	Time with Bank	100	93	41
	Loan Amount	2963	31028	4689
	Number of cases	**68**	**16**	**16**

It could be seen from Table 8 that the most appropriate method when Strategy 3 (the largest loan amount and long-standing bank customers) is selected is *K*-means cluster analysis with initial segmentation variables.

TABLE 8. Strategy 3: Comparative analysis of the clustering methods

Clustering Method	Variable	Platinum Clients	Gold Clients	Silver Clients
K-means, initial variables	Loan Amount	92667	50150	3165
	Time with Bank	92	69	90
	Worst Status Last 12 Months	0	0.25	0.44
	Number of cases	**3**	**4**	**93**
K-means, standardized variables	Loan Amount	68371	3845	2674
	Time with Bank	79	30	133
	Worst Status Last 12 Months	0.14	0.85	0.15
	Number of cases	**7**	**39**	**54**
Two-Step Clustering	Loan Amount	31028	4689	2963
	Time with Bank	93	41	100
	Worst Status Last 12 Months	1	2	0
	Number of cases	**16**	**16**	**68**

Table 9 shows that the most appropriate segmentation method when selecting Strategy 4 (the largest loan amount and a small number of missed payments in the last 12 months) is *K*-means with initial variables.

TABLE 9. Strategy 4: Comparative analysis of the clustering methods

Clustering Method	Variable	Platinum Clients	Gold Clients	Silver Clients
K-means, initial variables	Loan Amount	92667	50150	3165
	Worst Status Last 12 Months	0	0.25	0.44
	Time with Bank	92	69	90
	Number of cases	**3**	**4**	**93**
K-means, standardized variables	Loan Amount	68371	3845	2674
	Worst Status Last 12 Months	0.14	0.85	0.15
	Time with Bank	79	30	133
	Number of cases	**7**	**39**	**54**
Two-Step Clustering	Loan Amount	31028	4689	2963
	Worst Status Last 12 Months	1	2	0
	Time with Bank	93	41	100
	Number of cases	**16**	**16**	**68**

The most appropriate segmentation method when selecting Strategy 5 (the longest-standing bank customers with a large loan amount) is *K*-means cluster analysis with initial segmentation variables (see Table 10).

TABLE 10. Strategy 5: Comparative analysis of the clustering methods

Clustering Method	Variable	Platinum Clients	Gold Clients	Silver Clients
K-means, initial variables	Time with Bank	92	90	69
	Loan Amount	92667	3165	50150
	Worst Status Last 12 Months	0	0.44	0.25
	Number of cases	**3**	**93**	**4**
K-means, standardized variables	Time with Bank	133	79	30
	Loan Amount	2674	68371	3845
	Worst Status Last 12 Months	0.15	0.14	0.85
	Number of cases	**54**	**7**	**39**
Two-Step Clustering	Time with Bank	100	93	41
	Loan Amount	2963	31028	4689
	Worst Status Last 12 Months	0	1	2
	Number of cases	**68**	**16**	**16**

Appropriate methods when Strategy 6 (longest-standing bank customers with a small number of missed payments in the last 12 months) is selected are two-step clustering and *K*-means cluster analysis with standardized segmentation variables (see Table 11). *K*-means clustering with initial segmentation variables does not provide a good combination between the time with bank and the number of missed payments.

TABLE 11. Strategy 6: Comparative analysis of the clustering methods

Clustering Method	Variable	Platinum Clients	Gold Clients	Silver Clients
K-means, initial variables	Time with Bank	92	90	69
	Worst Status Last 12 Months	0	0.44	0.25
	Loan Amount	92667	3165	50150
	Number of cases	**3**	**93**	**4**
K-means, standardized variables	Time with Bank	133	79	30
	Worst Status Last 12 Months	0.15	0.14	0.85
	Loan Amount	2674	6837	3845
	Number of cases	**54**	**7**	**39**
Two-Step Clustering	Time with Bank	100	93	41
	Worst Status Last 12 Months	0	1	2
	Loan Amount	2963	31028	4689
	Number of cases	**68**	**16**	**16**

The following recommendations could be formulated as a result of the comparative analysis:

- If the bank wants to encourage the long-term clients with a good credit history, it is appropriate to apply two-step clustering or *K*-means cluster analysis with standardized segmentation variables;

- If the goal is to encourage customers with large loan amount (which are long-term clients or have a good credit history), it is preferable to apply *K*-means cluster analysis with initial segmentation variables.

- Loyalty program managers should have in mind that the marketing trends are constantly changing, as well as do the needs and the behavior of the clients, and the banking services themselves are changing too. Therefore, customer segmentation should not be seen as a one-time action but as a permanent process.

CONCLUSIONS

The conclusions from the provided analysis could be generalized as follows:

1. Three clusters (segments) of loyal customers are identified. Three variables are used for the segmentation. A *K*-means cluster analysis is applied, with the initial segmentation variables taken for input data. Then a further study is performed on standardized segmentation variables.

2. The potential segmentation strategies are formulated. A comparative analysis of the results of *K*-means clustering with initial and standardized segmentation variables and of a two-step clustering (obtained in a previous study from one of the authors) is made within each of the strategies. We have specified which type of cluster analysis is best suited to each of the strategies.

3. Customer segmentation involves not only statistical analysis of individual user groups but also demand for a practical outcome in the sale of financial services. An appropriate segmentation of commercial banks' customers enables these banks to increase their competitiveness.

ACKNOWLEDGEMENTS

This paper contains results of the work on project No 2018-FNSE-05, financed by Scientific Research Fund of Ruse University, Bulgaria.

REFERENCES

1. A team of Investor.bg, “Can we also see interest rates of 3.5% on some retail products?” Investor.bg. [in Bulgarian]
2. V. Mihova and V. Pavlov (2017) An approach of estimating the probability of default for new borrowers, *Economy & Business Journal* **11**.1, 200-208.

3. V. Mihova and V. Pavlov, “Comparative analysis on the probability of being a good payer,’ in *AMiTaNS’17*, edited by M.D. Todorov, AIP CP**1895** (American Institute of Physics, Melville, NY, 2017), paper 050006, 15p.
4. E. Raeva and V. Pavlov, “Planning outstanding reserves in general insurance,” in *AMiTaNS’17*, edited by M.D. Todorov, AIP CP**1895** (American Institute of Physics, Melville, NY, 2017), paper 050009, 10p.
5. E. Raeva and V. Pavlov (2015) Modeling issues of the claim process and insurance risk, *Pliska Studia Mathematica* **25**, 101–108.
6. V. Pavlov and E. Raeva (2015) On the risk modeling in general insurance, *Economy & Business Journal*, **9**.1, 563–574.
7. E. Raeva and V. Pavlov (2015) Some approaches for modeling claims process in general insurance, *Mathematics and Education in Mathematics* **1**, 233–239.
8. Future Factors. How Regulation, Client Expectations and Technology are Transforming Retail Banking, The Economist Intelligence Unit Ltd, 2014, pp. 6.
9. G. Mladenova, *Bank Customers’ Segmentation: Problems and Marketing solutions* (Burgas Free University, Burgas, 2015). [in Bulgarian]
10. P. Gamble, A. Tapp, A. Marsella, and M. Stone, *Marketing Revolution: The Radical New Approach to Transforming the Business, the Brand and the Bottom Line* (Cogan Page Publishers, 2005).
11. B. Sharp and A. Sharp (1997) Loyalty programs and their impact on repeat-purchase loyalty patterns, *International Journal of Research in Marketing* **14**.5, 473–486.
12. Understanding financial consumers in the digital era: A survey and perspective on emerging financial consumer trends (CGI, 2014).
13. The impact of incentives on consumer behavior (White paper, Virtual Incentives, 2017).
14. Elevate engagement to unlock your loyalty program (Report, Deluxe Corporation, 2015).
15. V. Pavlov and V. Mihova, *Applied Statistics with SPSS* (Avangard Print, Ruse, 2016). [in Bulgarian]
16. *Cluster Analysis, Chapter 16*, see http://www.norusis.com/pdf/SPC_v13.pdf. Accessed 13 June 2018.
17. R. Cattell (1944) A note on correlation clusters and cluster search methods, *Psychometrica* **9**.3, 169–184.
18. R. Tryon, *Cluster Analysis* (McGraw Hill, New York, 1939).
19. V. Mihova, “A classification of borrowers from commercial banks,” in *Scientific Works of Ruse University*, Vol. 55, series 6.1 (Ruse University, 2016), pp. 47–51. [in Bulgarian]