

# PREPAID TELECOM CUSTOMERS SEGMENTATION USING THE K-MEAN ALGORITHM

**Băcilă Mihai-Florin**

*Babeş-Bolyai University of Cluj-Napoca, Faculty of Economics and Business Administration*

**Rădulescu Adrian**

*Business Logic Systems Ltd*

**Mărar Ioan Liviu**

*Business Logic Systems Ltd*

*The scope of relationship marketing is to retain customers and win their loyalty. This can be achieved if the companies' products and services are developed and sold considering customers' demands. Fulfilling customers' demands, taken as the starting point of relationship marketing, can be obtained by acknowledging that the customers' needs and wishes are heterogeneous. The segmentation of the customers' base allows operators to overcome this because it illustrates the whole heterogeneous market as the sum of smaller homogeneous markets. The concept of segmentation relies on the high probability of persons grouped into segments based on common demands and behaviours to have a similar response to marketing strategies. This article focuses on the segmentation of a telecom customer base according to specific and noticeable criteria of a certain service. Although the segmentation concept is widely approached in professional literature, articles on the segmentation of a telecom customer base are very scarce, due to the strategic nature of this information. Market segmentation is carried out based on how customers spent their money on credit recharging, on making calls, on sending SMS and on Internet navigation. The method used for customer segmentation is the K-mean cluster analysis. To assess the internal cohesion of the clusters we employed the average sum of squares error indicator, and to determine the differences among the clusters we used the ANOVA and the post-hoc Tukey tests. The analyses revealed seven customer segments with different features and behaviours. The results enable the telecom company to conceive marketing strategies and planning which lead to better understanding of its customers' needs and ultimately to a more efficient relationship with the subscribers and enhanced customer satisfaction. At the same time, the results enable the description and characterization of expenditure patterns for services that are continuously growing. Also, the study demonstrates this analysis model is efficient for a large customer base.*

*Keywords: market segmentation, profiling segments, telecommunication services, k-mean cluster, relationship marketing.*

*Cod JEL: M31*

## **1. Introduction**

Companies now face severe competition in a highly dynamic and unstable marketing environment. In order to be successful and hold a leading place on the market, they have to provide quality service and respond to changes in their customers' needs, wishes, characteristics and behaviours. Thus, instead of looking at the customer base as homogenous and engage all customers in the same campaigns or marketing incentives, companies should approach customers differently, depending on their needs, characteristics and behaviours (Bose& Chen, 2010:197).

Given the severe competition and the highly dynamic environment, companies can no longer afford to simply attract new customers. They must place the same importance on strategies focused on retaining customers and bringing back former subscribers, rather than focusing on increasing their market share. Most of the times, retention and increasing the clients' profitability rates are equally important and easier to achieve than attracting new customers from the competition (Tsipstis& Chorianopoulos, 2009: 291). To develop a proper relationship with the

subscribers, telecom operators need to implement and apply the principles of customer relationship management.

Introducing CRM enables operators to customize products, services and communication, to create and offer higher customer value, increase customer loyalty, improve business stability with higher customer retention rates and create value, boost sales and reduce the number of dissatisfied clients (Diller, 2000: 32; Bruhn, 2001: 117).

Relationship marketing focuses on identifying the customers most suitable and profitable for developing long term, mutually beneficial relationships (Niarn and Bottomley, 2003:243). Segmentation is an extremely important marketing concept because it is needed, together with better understanding of customers' needs, in order to improve the relationship with existing customers (Storbacka, 1997:480).

## **2. Literature review**

Mobile operators' activity consists of gathering and managing a large amount of information and data. Thus, millions of people, in millions of places can perform tens or hundreds of transactions in a short period resulting in billions of events to be recorded. In order to handle such an enormous quantity of data, special analyses methods need to be involved. These have appeared and grown at the same pace with the information technology (Foss and Stone, 2001: 67).

Market segmentation is a fundamental component in the companies' strategic marketing planning in industrialized countries because goods and services can no longer be produced and retailed without taking into consideration the customers' needs and wishes and the fact that they differ (Wedel and Kamamura, 2000:3). The segmentation concept relies on the fact that it is more likely for persons grouped according to common behaviours and needs to have a homogenous response to marketing actions (Dibb 1999:108).

Market segmentation refers to the process of splitting current or potential customers into several groups both as homogenous as possible and as heterogeneous among themselves (McDonalds & Dunbar, 2004: 34; Paina & Pop, 1998: 103; Wedel and Kamamura 2000:3). The perspective of segmentation being just an organizational tool is shallow because it bears both strategic and tactical marketing implications. Strategically speaking, segmentation allows the identification of profitable customers, the stabilization of market decisions and market segments and placing the product or service on the market. At operational level it drives the companies to lay more stress on enhanced customer understanding to develop more efficient relations with them (Dibb, 2001:195; Storbacka, 1997:485).

The benefits of market segmentation for a company include: better customer understanding, identification of the most attractive consumer segments, efficient prioritisation of resources, proper positioning of product and market services both for the customer and against the competition, conception of personalised campaigns and incentives, selection of the best performing distribution channels, winning competitive edges after customization of products and services and creating value offers and making the most of market opportunities (Dibb & Simkin, 2010: 113; Kiang et al., 2006: 36; Feldman, 2006: 24; Tsiptsis & Chorianopoulos, 2009: 190).

The studies (Young, Oti and Feigin 1978:411, Dibb 2001:196; McDonalds and Dunbar 2004:431) prove that although segmentation brings benefits, marketers can also face difficulties when putting theory into practice: increased costs for marketing actions as compared to the previous market approach, obtaining reduced segments or mobile segments, segments that against their common features have different consumption behaviours, difficult implementation of marketing actions.

CRM segmentation has become more efficient in the recent years due to the development of database marketing techniques. Profiling techniques provide marketers with superior tools for customer segmentation and adaptation of marketing strategies to the specific needs of each consumer segment (McCarty & Hastak, 2007: 656).

In general, in order to perform customer segmentation, companies use criteria that relate to geographical, demographical, psycho-graphical, socio-economic, behavioural characteristics and psychological attitudes toward the respective product or service (McDonalds and Dunbar 2004:35-37, Kotler and Keller 2006:247; Tsiptsis and Corianopoulos 2009:191-193). Most companies use segmentation based on demographics. In the case of markets that feature high competition, such as the telecom market, this approach is not enough. These companies also need to consider the information related to the customers' needs, consumer behaviour, service or payment preferences, perception of product, probability of leaving the network, growth potential and customer migration. To become subject to segmentation, a market needs to be heterogeneous (Dibb and Simkin 2010:115). Segmentation criteria need to fulfil the following conditions (Paina and Pop, 1998:106; Dibb 1999:108; Kotler and Keller 2006:248; Tonks 2009:345):

- Segments will be measurable so that the size, buying power and other characteristics can be quantified and determined;
- Segments will be substantial and have a potential for profit so that they justify the creation of special later created marketing programs;
- Segments are homogenous, each being distinct in terms of clients' profiles and needs;
- Segments are accessible and differentiable so that they can be differently approached marketing-wise;
- Segments enable action, namely they allow the formulation of efficient programs to attract and serve customers;
- Market segments are stable; they preserve their characteristics for a longer period;
- Market segments are compatible with the organisation's mission, culture and structure, with the marketing strategy, with the existing resources.

According to Bayer (2010: 248) mobile operators use basically the following segmentation types: subscriber value-based segmentation, subscriber behaviour-based segmentation, subscriber lifecycle-based segmentation and subscriber (possible) migration-based segmentation. They are used for different situations and focus on different aspects.

### **3. Consumption-based segmentation**

This study was carried out on natural persons who are prepaid subscribers. These people do not have a contractual relation with the telecom carrier and they buy credit in advance. They receive no bill and pay for services before actually using them. This survey was carried out on 76,753 prepaid subscribers. Usage data refers to last year's second semester. The analysis excluded the people who failed to recharge within the past three months of 2011 and did not spend anything on calls, SMS or Internet in three months.

This analysis will identify the subscribers' profiles in the overall population and determine the efficiency of the K-mean cluster analysis in the case of high data volumes. Since extracting the profiles within a population presupposes learning unknown information, the definition of hypotheses is not possible (Foss and Stone 2001:84).

We continue by presenting the classification of subscribers into several categories using the following variables: the sum of the amounts recharged in 6 months, the value of the SMS sent within the 6 months, the Internet traffic value in the 6 months and the value of calls made in the 6 months.

To group subscribers into segments we used SPSS, the K-Mean Cluster non-hierarchical method. This algorithm follows the segmentation of the populations so that the variation inside the clusters will be down to a minimum. This analysis pursues the grouping of subscribers into various segments based on their behavioural values (recharge values, call values, SMS expenditure and Internet expenditure). The ANOVA analysis revealed the following order in the case of the factors' contribution to the splitting of the population into groups: recharge value, call

value, Internet expenditure and sent SMS value. The value of Sig. is smaller than 0.05 thereby results are significant.

The data was submitted to the logarithm to reduce the differences among values, and the values were standardised.

To examine and assess the solution of the segmentation analysis we took the following aspects into account: the number of segments and the size of each segment, the clusters' cohesion and the differences among clusters (Tsiptsis and Chorianopoulos, 2009). After the analysis (the internal homogeneity of the groups resulted and the distances among the groups) and considering the practical aspects (increasing the size of the segments, so they can be substantial, and the number of consumer groups that can be managed by the mobile operator) we concluded that the analysed population should be split into seven segments. The analyses revealed the absence of a dominant segment which would require the analysis of another segmentation of the population. In addition, no segment is too small, so no separate analysis is required.

To assess the quality of the cohesion of the clusters we analysed, for each cluster, the maximum (Euclidean) distance to the central values begot the first reduced values. Moreover, to determine the distance between the members of every cluster and the cluster's central value, we calculated the Average sum of squares error indicator, obtaining a value of 0.9820 which indicates high cohesion.

To determine the differences between the clusters we used the ANOVA test, which indicated the significant differences (at a 0.000 significance level) among groups in the case of each variable taken into consideration. The Tukey post-hoc test confirmed (at a 0.000 significance level) that the averages of the four variables differ significantly in the case of the seven clusters.

The following segments were identified inside the population:

Cluster 1. This cluster consists of 12,366 people and includes subscribers with a small recharge value. On average, these subscribers recharged by 92.06 monetary units (m.u.) in the past 6 months. These subscribers feature credit used for making calls, then to connect to the Internet and then to send SMS. In absolute value, they used 38.76 m.u. for calls, 26.87 m.u. for Internet and 4.31 m.u. for SMS. In percentage, of the spent (recharged) value, they spent 55.42 % (42.10%) on calls, 38.42% (29.19%) on data and 6.16 % (4.68%) on SMS.

Compared to the entire population's average values, subscribers in this cluster spent 25.62% less on calls, 43.68% less on SMS and 54.48% less on their Internet connection. These consumers are comfortable surfing the web and communicating at ease. Against the population's average, they consume (recharge) by 43.67% (50.26%) less. These subscribers used 75.98% of the recharge on the three services.

Cluster 2. This cluster consists of 8,209 people and includes subscribers with small to average recharge values. On average, these subscribers recharged 123.14 monetary units in the past 6 months. They use the credit to make calls and send SMS, and have no interest in Internet. In absolute value, they spent 67.32 m.u. on calls, 17.22 m.u. on SMS and 0.59 m.u. on Internet. Percentage-wise, of the recharge value they spent 79.08% (67.32%) on calls, 20.23% (13.99%) on SMS and 0.69% (0.47%) on the Internet.

Compared to the average values for the entire population, subscribers in this cluster spend 29.19% more on calls, 41.55% less on SMS and 99% less on Internet surfing. These clusters indicate the subscribers' preference for traditional phone services and reluctance to use the handset to surf the web. Compared to the population's average, their recharge value is 39.46% (24.65%) smaller. These subscribers used 69.14% of the recharge on the three services.

Cluster 3. This cluster consists of 11,389 people and includes subscribers with the greatest recharge value. On average, these subscribers' recharge value in the past six months is 403.24 m.u. They use the credit on all mobile phone services. In absolute value, they consumed 165.86 m.u. on calls, 83.89 m.u. on SMS and 107.31 m.u. on Internet. Namely, of the entire recharge the calls took 46.45% (41.13%), the SMS 23.49% (20.80%) and the Internet 30.06% (26.61%).

Compared to the average values for the entire population, subscribers in this cluster spend 218.29% more on calls, 184.74% more on SMS and 81.77% more on Internet. Referring to the structure of the population, although the subscribers in this cluster use all services, they tend to employ traditional services: voice and SMS. Compared to the population's average, they recharge 153.94% (146.75%) more. These subscribers spent 88.54% of the recharge value.

Cluster 4. This cluster consists of 20,816 people and includes subscribers with an average recharge value. On average, these subscribers' recharge value in the past 6 months was 166.14 monetary units. These subscribers use the credit to connect to the Internet, then to talk and then to send SMS. In absolute value, they spent 61.30 m.u. on Internet, 44.87 m.u. on calls and 38.28 m.u. on SMS. Percentage-wise, they allotted 42.44% (36.90%) of the recharge value on the internet, 31.07% (27.01%) on calls and 26.50% (23.04%) on SMS.

Compared to the average values for the entire population, subscribers in this cluster spend 13.88% less on calls, 29.90% more on SMS and 3.83% more on connecting to the Internet. This cluster indicates the subscribers' high preference for voice communication. They recharge 2.73% (1.67%) more than the average population. These subscribers only spent 86.94% of the recharge value.

Cluster 5. This cluster consists of 13,227 people and includes subscribers with small to medium recharge value. On average, these subscribers' recharge in the past 6 months reached 134.51 monetary units. They use their credit to surf the web, then to send SMS and only last to initiate calls. In absolute value, they used 97.62 m.u. on Internet, 22.30 m.u. on SMS and 8.50 m.u. on calls. In per cent, the Internet used 76.03% (72.57%) of their recharge value, SMS used 17.37% (16.58%) and calls 6.60% (6.30%).

Compared to the average values for the entire population, subscribers in this cluster spend 83.73% less on calls, 24.31% less on SMS and 65.34% more on the Internet. This cluster indicates higher preference on written rather than verbal communication. Compared to the population average, their recharge value is 8.69% (17.69%) smaller. These subscribers spent 95.46% of the recharge value.

Cluster 6. This cluster consists of 4,864 people and includes subscribers with the smallest recharge value. On average, these subscribers' recharge value was 18.38 monetary units in 6 months. These subscribers use their credit to connect to the Internet, then on voice services and ultimately for SMS. In absolute value, they used 4.72 m.u. on Internet, 4.33 m.u. on calls and 1.64 m.u. on SMS. Of the recharge value, they used 40.5% (25.68%) on internet, 44.13% (23.56%) on calls and 15.37% (8.94%) on SMS.

Compared to the average values for the entire population, subscribers in this cluster spend 91.69% less on calls, 94.42% less on SMS and 92.01% less on the Internet. This segment is the least attractive for the company since it includes people who spend little money on mobile services and are careful not to exceed their recharge value. Compared to the population average, their recharge value is 92.40% (88.75%) less. These subscribers only spent 58.18% of the recharge value.

Cluster 7. It consists of 5,882 people and includes subscribers with a small recharge value. On average, these subscribers recharged 80.61 monetary units in the past 6 months. These subscribers use the credit for Internet and show a complete lack of interest for other services. In absolute value, they used 64.93 m.u. on Internet, 1.98 m.u. on SMS and 1.92 m.u. on calls. In per cent, the Internet consumed 94.33% (80.55%) of their credit, SMS 2.87% (2.45%) and calls 2.80% (2.38%).

Compared to the average values for the entire population, subscribers in this cluster spend 96.31% less on calls, 93.3% less on SMS and 9.97% more on connecting to the Internet. This cluster indicates very high preference for the Internet. Compared to the population average, their recharge value is 51.05% (50.67%) less. These subscribers spent 85.38% of the recharge.

#### **4. Conclusions**

The study pursued the applicability of the K-mean cluster method to provide business intelligence. The method employed proved its efficiency in processing large data volumes leading to obtaining consumer segments featuring high internal homogeneity of the segments and high heterogeneity among segments.

The hereby study also has several managerial implications. An obvious conclusion of the segmentation analysis is the fact that most subscribers use their credit for certain purposes. The fact that most prepaid cards owners focus on certain types of services only indicates that the respective market is not mature. Only subscribers in the 3rd and 4th clusters use their credit for all types of services. Analyses indicate that in six months 19.9% of the subscribers spent less than six monetary units on calls, 33.3% spent less than 6 monetary for SMS and 23.6% spent below 3 monetary units to connect to the Internet. Therefore, marketing activities must be conceived to encourage subscribers in the rest of the segments to use their handsets for all types of services provided by the mobile network operator.

#### **5. Bibliography**

##### **Books**

Bruhn, Manfred. *Orientarea spre client*, București:Editura Economica, 2001.

Diller H. "Customer Loyalty: Fata Morgana or Realistic Goal? Managing Relationships with Customers" In *Relationship Marketing: Gaining Competitive Advantage through Customer Satisfaction and Customer Retention* edited by Henning-Thurau T. and Hansen U., 29-48, Berlin:Springer-Verlag , 2000.

Foss Bryan & Stone Merlin. *Succesfull Customer Relationship Marketing*, London: Kogan Page, 2001.

Kotler, Philip and Keller, Kevin L. *Management Marketing*, New Jersey:Pearson Education Inc., 2006.

McDonalds, Malcom and Dunbar, Ian. *Marketing Segmentation: How to Do It, How to Profit from It*, Oxford :Elsevier Butterworth-Heinemann, 2004.

Paina, Nicolaie și Pop, Marius, D. *Cercetări de marketing*, Cluj-Napoca: Presa Universitară Clujeană, 1998.

Tsiptsis, Konstantinos and Corianopoulos, A.ntonios. *Data Mining Tehniques in CRM: Inside Customer Segmentation*, Chicester:John Wiley& Sons Ltd., 2009.

Wedel, Michel and Kamacura, Wagner, A. *Market Segmentation: Conceptual and Methodological Foundations*, Boston: KluverAcademic Publishers, 2000.

##### **Articles**

Bayer, Judy. "Customer Segmentation in the Telecommunication Industry". *Database Marketing& Customer Strategy* 17, 3/4(2010):247-256.

Bose, Indranil and Chen, Xi. "Exploring Business Opportunities from Mobile Services Data of Customers: An inter-cluster Analysis Approach", *Electronic Commerce and Applications* 9, 3(2010): 197-208.

Dibb, Sally. "Criteria Guiding Segmentation Implementation: Reviewing the Evidence", *Journal of Strategic Marketing* 7, 2(1999):107-129.

Dibb, Sally. "New Millenium, New Segments: Moving Towards the Segment of One?", *Journal of Strategic Marketing* 9, 3(2001):193:213.

Dibb, Sally and Simkin, Lyndon. "Judging the Quality of Customer Segments: Segmentation Effectiveness", *Journal of Strategic Marketing* 18, 2(2010): 113-131.

Feldman, David. "Segmentation Building Blocks", *Marketing Research* 18, 2(2006): 23-29.

Kiang, Melody, Y. and Hu, Michael, Y. and Fisher, Dorothy, M. "An Extended Self-organizing Map Network for Market Segmentation – A Telecommunication Example", *Decision Support Systems* 42, 1(2006): 36-47.

Liou, James J., H. and Tzeng Gwo-Hsiung. "A Dominance-based Rough Set Approach to customer behavior in the airline market", *Information Sciences* 180, 11(2010): 2230-2238.

McCarty, John, A. and Hastak, Manoj. "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and Logistic Regression", *Journal of Business Research* 60, 6:(2007) 656-662.

Niarn, Agnes and Bottomley, Paul. "Something Approaching Science? Cluster Analysis Procedures in the CRM Era", *Journal of Market Research* 45, 2(2003):241-261.

Storbacka, Kaj. "Segmentation Based on Customer Profitability – Retrospective Analysis of Retail Bank Customer Bases", *Journal of Marketing Management* 13, 5(1997): 479-492.

Tonks, David, G. "Validity and the Design of Market Segments", *Journal of Marketing Management* 25, 3/4(2009):341-356.

Young, Shirley and Oti, Leland and Feigin, Barbara. "Some Practical Considerations in Market Segmentation", *Journal of Marketing Research* 15, 3(1978):405-412.