SKAP JOUNRALS
SLOVAKIA ACADEMIC PUBLISHING
Impact factor 2015 0, 76 ISBN: 10 12-3869

**Arsim Begunca**

# *USING CLUSTER K-MEANS METHOD FOR SOFT DRINKS CONSUMERS SEGMENTATION BY THE BENEFITS SOUGHT-CASE STUDY CITY OF PRISTINA*

Arsim Begunca, phd candidate, European University of Tirana, Albanian-Kosovo,

**Abstract**

Segmentation by benefits sought is a market segmentation approach with which it is possible to identify more market segments through causal factors than description factors. The basic belief of this strategy is that the benefits that people are looking at the consumption of any product are the fundamental reason for the existence of real segments and are better determinants of behavior than other approaches (Haley, 1968). Benefits better predict behavior rather than personality and lifestyle variables, demographic and geographic measurements, which simply describe the behavior without explaining them (Haley, 1985; Young, Ott and Feagin, 1980). To achieve the goal of customer segmentation by benefits required we have used the K-Means clustering method , as a basis for the division of customers in segments. Based on the differentiating factors in the selection of soft drinks, social demographic variables and soft drink consumer behavior throughout these segments, managers can modify their marketing strategies to satisfy the needs of these segments and at the same time maximize the profit.

Key words: benefit sought, soft drink, K-means cluster analysis

Impact factor 2015 0, 76 ISBN: 10 12-3869

## 1.0 **Introduction**

The segmentation based on benefits sought is the consumer segmentation based in what is the product special benefit appealed to them. Different customers require various benefits and marketers need to understand every segment and based on that to develop their communication for each group.

According to Haley (1968) benefits people look for in a particular product are the basic reasons for the existence of real market segmentation. Segmentation by benefits is the technique that makes the segmentation of consumers based on the benefit or requested wishes.

These benefits when present as the attribute of the product, service or market supply, make the consumers buy these products rather than simply a description of who are the consumers in terms of social-economics, demographics or psychografics. Segmentation based on benefits sought offers more benefits than traditional methods because it explains the reasons why customers choose to buy a particular product or to hold to a particular supplier or a service provider.

Important issue in this study is soft drinks consumer segmentation under the required benefits

Researchers often face the need for grouping or splitting similar objects from a greater whole in small groups or communities. These groups include elements with similar ingredients, but in general the composition of groups can be defined with the help of the components which constitute this community. One of the methods with whose help is carried out the classification of similar objects is called cluster analysis. With the application of cluster analysis, clusters are gained, which are gained that way that the objects similar among themselves are grouped in the same cluster, while many objects that differ among themselves are placed in different clusters. Cluster analysis is the name for a group of multivariate techniques whose main purpose is to group objects based on key features that they possess (Hair et al., 2010). Cluster analysis method which is applicable in various research disciplines, that is a big advantage of this method.

Impact factor 2015 0, 76 ISBN: 10 12-3869

Researchers often face a number of observations that are meaningless or impossible to interpret, if not grouped in clusters.

Therefore, the main advantage of cluster analysis is objective reduction of data based on information reduction in total population and reduction of population characteristics in representative features of groups, with minimal loss of information. Considering the basic characteristics of cluster analysis it can be concluded that this method is useful for the development of hypotheses/theories and verification of existing ones.

### 1.1. Problem definition and formulation

Soft drinks market is a very large and important market with very fierce competition. In recent years this market in global terms has not increased greatly. Thus for all the actors of this market without understanding of the customer/consumer, they may lose a market share from competitors given the fierce competition. The problem of this study is to find the relationship between the required benefits and purchasing behavior, in the soft drinks market. Nowadays most soft drink products mark (target) all classes of people, such as children, adults, middle age and old age. The aim of this study is to develop a better understanding of the benefits required by certain consumer and certain purchasing behavior, to find which factors or benefits have more impact on the purchasing behavior.

### 1.2 Purpose of the study

The main purpose of this study is the soft drink consumer segmentation in Prishtina based on benefits sought by using K-Means Cluster method.

Other study objectives are:

1. To identify differences between the behavior of consumers in the market segmentation based on benefits sought/required.

2. To identify behavior features/ characteristics based on benefit sought approach.

Impact factor 2015 0, 76 ISBN: 10 12-3869

3. To identify personal features based on benefit sought.

4. Providing a guide for the development of effective segmentation strategies.

### 1.3 **Research hypothesis**

The main hypothesis of the study is: there are important differences between consumer behavior in different market segments based on the benefits sought.

### 2.1 . Research methodology

This study intends to segmentize soft drinks consumers based on benefits sought testing proposed hypothesis based on current theories.Quantitative approach is used more in this study than the qualitative approach. Quantitative approach is more suitable to test and rate/evaluate the assumptions/hypothesis and to follow standard procedures which are strict in the used instruments and allow the statistical data analysis to identify relevant information related to soft drinks consumer lifestyles in Prishtina. The face to face survey with self-administered questionnaire was adopted in this study as a data collecting method. Alreck and Settle (1995) pointed out that self-administered questionnaires are an excellent way to ask responders and provide information, provided by an expert and when the appropriate instrument has been used.

### 2.2 Data collection

Primary data on this topic are provided through various survey respondents in city of Prishtina. Through a self administered questionnaire we have surveyed a total of 399 responders, where the information obtained from these surveys is used as primary data and presents the essence/core of the research. The questionnaire consists of 27 questions; the series of questions are designed to have a range of responses focusing on the same general topic.

### 2.3 Sample selection

Impact factor 2015 0, 76 ISBN: 10 12-3869

A self-administered questionnaire is used in this study that is directly distributed by the researcher in randomly selected sample. The study uses one scale sampling method. Therefore, Prishtina residents of ages 18 and up have been randomly selected in conducting this study.

### 2.4 Methods of analysis

Programm called Statistical Package for Social Science (SPSS version 24) is used for data analysis. This study used factorial analysis and K-means cluster analysis

## 3.0 Data analysis

### 3.1 Factorial Analysis

Factor analysis is applied in this study to reduce the number of variables, in order to make the study interpreted easier.

A set of questions is in the study instrument including 16 issues related to the benefit sought when consuming soft drinks.Issues related to the sought benefit are the subject of our research and division into segments will be done based on these benefits, therefore the subject of factorial analysis are only the issues related to the benefit sought/required. Other issues related to the consumption of soft drinks and demographics, used in the instrument of our study will be investigated to see their connections with the segments identified, with the purpose that information about consumption of refreshments would be more complete.

Kaiser-Meyer-Olkin sample adequacy measuring unit that provides the measuring unit in quantification of the inter-correlation scale between variables and the factorial analysis adaptability/ adequacy, is 0.805 exceeding recommended value of 0.6 (Coakes & Steed,1999; Kaiser.1970, 1974). Bartlett's Test of Sphericity reached the reached statistical significance (p = 0.000) supporting the suitability of the correlation matrix for the factor analysis.

Impact factor 2015 0, 76 ISBN: 10 12-3869

Table.3.1 Kaiser-Meyer-Olkin measure for Sampling Adequacy

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.805 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 6530.618 |
| | df | 1020 |
| | Sig. | 0.000 |

Preliminary assumptions of factorial analysis request an existence of the correlation between variables, which will be included in the analysis. Kaiser-Meyer-Olkin Measurement for the sample adequacy, or the choice to conduct a factor analysis of 0805 shows that our choice is estimated between a choice "with merit" and a "very good" choice (a coefficient of 0.9 is very good and a coefficient below 0.5 is not acceptable), and there is no need to examine the Anti-Image correlation matrix.

Bartlett's test of Sphericity tests the hypothesis that the correlation matrix is the matrix unit (all diagonal elements are 1 and off diagonal elements are 0, which means that all variables are uncorrelated with each other). The value of statistical significance in this table shows that the zero hypothesis (that this matrix is a matrix unit) can almost certainly be rejected outright. So, we can say that our data have correlations that are adequate for performing factor analysis.

Sixteen issues of the section "benefits sought" have been subjected to factor analysis to reduce the number of variables in a smaller community of fundamental factors.

Before applying the Principal Components Analysis (PCA), the adequacy of the data to factor analysis is assessed or reviewed. Any issues in correlation matrix having a coefficient of less than 0:36 on any other matter is removed. Also commonality table is examined and any issue with less than 0.50 has also been removed as low commonality value, since it shows that the variables do not have relations with other variables in community (Field, 2005; Tabachnick & Fidell, 2000).

Factor analysis was performed using the Principal Component Analysis - PCA with orthogonal rotation(varimax).Orthogonal rotation was chosen because the resulting factors will be used in

Impact factor 2015 0, 76 ISBN: 10 12-3869

subsequent statistical analysis and this approach minimizes collinearity (Hair,Anderson,Totham & Black,1998). Two sources of information were examined in order to determine the number of factors that will be elected, Eigenvalue size and Eigenvalue Screeplot. Only factors with an Eigenvalue of 1 or more are taken into account and Screeplot diagram is checked for the information on the number of factors. The final solution resulted in 6 common factors, which significantly reduce the original number of variables introduced in the analysis and explain in total 77,013 percent of the total variance, as shown in Table nr.3.2:

Interpretation of Common factors consists in summarizing of information that contains each of the factors. The meaning of each of the factors derives after analyzing the variables that have higher correlations with each of these factors. The inferior limit of weight factors is set at 0.360. Table 4.17 shows the weights for each variable factors on components or factors after the spinning/rotation. Each number represents the partial correlations between the original variable and spin factor. These correlations also help in the formulation of the factors interpretation.

Impact factor 2015 0, 76 ISBN: 10 12-3869

Table.3.3 Rotated Component Matrix

| Benefits | Factors | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Taste | .631 | .146 | -.144 | -.438 | .241 | .273 |
| Low Calories | .079 | .553 | .494 | -.206 | -.184 | .390 |
| Low Prices | .113 | .025 | -.071 | -.134 | .160 | .917 |
| (Popular) Brands | .060 | .815 | .100 | -.030 | .222 | -.152 |
| Energy | -.041 | .722 | -.312 | .187 | .245 | .162 |
| Quality | .709 | .079 | .501 | -.074 | -.056 | .073 |
| Being Natural | .187 | -.254 | .721 | .039 | -.449 | .019 |
| Quenching thirst | .005 | .119 | .826 | -.005 | .216 | -.107 |
| Image (of the drink) | -.219 | .053 | .194 | .778 | .356 | .013 |
| Gassed (Sparkling) | -.081 | .162 | -.010 | .241 | .862 | .126 |
| Packing/Wrapping | -.095 | -.039 | -.105 | .830 | -.015 | -.110 |
| Refreshment/Freshness | .774 | -.180 | .075 | .046 | -.015 | .038 |
| Entertainment | .342 | .281 | -.441 | .437 | .100 | .346 |
| Vitamins | .802 | .228 | -.098 | -.080 | -.233 | .028 |
| Health Benefits | .105 | .660 | .048 | .113 | -.328 | .509 |
| Cafeinne | .317 | .309 | -.384 | .552 | .179 | -.058 |

Impact factor 2015 0, 76 ISBN: 10 12-3869

The first factor relates to these controlled factors explicitly in our analysis, "taste", "quality", "freshness" and "vitamins".All these factors during the ranking had the highest degree of compliance and can be considered the most required benefits and also this factor may be labeled as "key benefits/major benefits".

The second factor is closely correlated with other factors such as "low calorie", "popular brands", "energy" and "health benefits". These factors are mainly related to health issues, and may be labeled as the "health"factor.

The third factor has the highest correlation with the factors "natural" and "quenching of thirst ', but also is correlated to " quality "and" low calorie" factors. But given the high correlation of this factor it can be labeled as "natural".

The fourth factor is in the highest correlation with factors such as "the image of the drink", "Packaging", "caffeine" and "entertainment". Given these data, this factor can be labeled as "external factor".

The fifth factor is in higher correlation with benefits such as "sparkling" and "drinking image". Because the highest correlation is with the variable of the benefit "gassed/sparkling" this factor can be labeled as "Gassed/sparkling".

The sixth factor is in high correlation with the variable benefit such as "low prices", "low calorie" and "health benefit". But it has the highest correlation with "low prices" variable and 2 other variables associated with other factors, this factor can be properly called/labeled as factor of "price".

Këta gjashtë faktorë të benefitëve të kërkuara të fituar përmes analizës faktoriale do të përdoren në analizën cluster me qëllim të segmentimit të konsumatorëve në bazë të variablave të benefitit të kërkuar.

Impact factor 2015 0, 76 ISBN: 10 12-3869

3.2 Cluster Anaylsis

The purpose of this study is soft drink consumer segmentation based on variables of benefit sought. In our cluster analysis the aim is to segmentize soft drinks customers based on the answers from the used questionnaire and to explore the features of the group of respondents, which is the first step in the marketing procedure that brings together potential customers in specific marked markets with common features.

Cluster analysis is the term used to describe a group of multivariate techniques to develop usable subgroups of individuals or objects based on similarity or association across entities (Coakes and Steed, 1999; Hair, Anderson, Tatham and Black, 1998). Spath (1980) emphasized that the main goal of cluster analysis is to divide a set of objects into basic groups based on their similarities.

Cluster analysis is one of statistical analysis that is used in many segmentation studies in defining groups. There are two main subdivisions of clustering procedures, when the number of clusters is default/determined then K-means Cluster method is used and hierarchy cluster analysis is used when the number of clusters is not predetermined.

K-Means cluster analysis was used to identify relatively homogeneous groups of cases based on selected characteristics for this study. K-Means cluster analysis procedure requires a specified number of clusters for further analysis.

K-Means procedure is applicable to the database community with a large number of cases while hierarchical procedure is preferred when the number of cases is limited. K-Means try to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can manipulate large numbers of cases.

399 cases are included in our cluster analysis since 1 other case had missing values for at least one of the marketing elements listed in the questionnaire.
This means that for these cases, the result is not taken at least for one of the factors derived from factor analysis (result/outcome is based on the weight factor and individual responses). Therefore cluster analysis will target segmentation of these 399 individuals based on the importance given

Impact factor 2015 0, 76 ISBN: 10 12-3869

to each of those 6 factors. It will then investigate links to other variables, especially those demographic with identified segments, to determine their profile.

The cluster segmentation procedure has a final goal in separating inviduals in clusters that maximize the distance between the clusters based on the 6 factors.

It begins with the building of the clusters centers and determines the placement of individuals into segments based on the distance from the initial centers. Through iterative procedures, the initial cluster centers change being replaced by the average of individuals in that cluster, until the restoration of individual in to segments does not influence or affect less (statistically insignificant difference) to change the center of the cluster.

### 4.3.1 Identification of segments

Individuals surveyed by K-Means method were divided into three segments\clusters. 137 individuals are included in the first segment, 192 individuals in the second and 70 individuals in the third. The comparison of averages between segment evaluations/ratings on 6 factors, the results of which are summarized in the table show that all factors have statistically significant contribution in dividing in clusters, thus the importance given to these factors in different segments is different. ANOVA table presents all cases used in the cluster analysis.

Table 3.4 ANOVA for three selected clusters

| | Cluster | | Error | | | |
|---|---|---|---|---|---|---|
| | Mean Square | df | Mean Square | df | F | Sig. |
| Factor 1: Key benefits | 71.343 | 2 | 0.645 | 396 | 110.656 | 0.000 |
| F actor 2: Health | 39.284 | 2 | 0.807 | 396 | 48.7 | 0.000 |
| Factor 3 : Natural | 21.803 | 2 | 0.895 | 396 | 24.363 | 0.000 |
| Factor 4 : External factor | 28.844 | 2 | 0.859 | 396 | 33.564 | 0.000 |
| Factor 5: Gassed(Sparkling) | 40.015 | 2 | 0.803 | 396 | 49.834 | 0.000 |
| Factor 6: Price factor | 73.585 | 2 | 0.633 | 396 | 116.174 | 0.000 |

Impact factor 2015 0, 76 ISBN: 10 12-3869

The following table shows the average rating of 6 factors (differences in the average which resulted as important). For the first segment it can be said that it is the segment that gives greater importance to factor 5 or gives importance to the drink in being sparkling/gassed. The second factor, more important for this segment is factor 5 or price factor.

Second segment gives more importance to the first factor (key benefits) where taste, quality, freshness and vitamins are included. Second important factor is the "health factor" where low calories and health benefits are included. The third factor (natural factor ) has a great importance in this segment, seeking the drink to be a natural one.

The third section gives greater importance to the fourth factor (external fator), while other important factors are the "price" , "major benefits" and the 'natural' factor.

Table 3.5 Rating factor averages by segments

| | Cluster | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Factor 1: Key benefits | -0.79783 | 0.34062 | 0.73885 |
| Factor 2: Health | -0.01938 | 0.29991 | -0.89634 |
| Factor 3: Being Natural | -0.44814 | 0.28306 | 0.10067 |
| Factor 4: External Factor | -0.15217 | -0.19163 | 0.87344 |
| Factor 5: Gassed/ Sparkling | 0.54299 | -0.12342 | -0.72419 |
| Factor 6: Price factor | 0.40881 | -0.60908 | 0.82052 |

These cluster analysis results presented in tables 3.4 and 3.5 support the main hypothesis assertion that "there are important differences between consumer behavior in different market segments based on the benefits sought.

**4.0 Conclusions, Implications of the study, study limitations and suggestions for future studies**.

**4.1 Conclusions**

Through a series of analysis we have achieved through cluster analysis to group customers into three benefit sought segments. By putting consumers in 3 benefits sought segmentized groups we

Impact factor 2015 0, 76 ISBN: 10 12-3869

have provided answers to our main goal which has been soft drinks customer segmentation based on K-means cluster analysis. With these practical results of the study we have concluded that soft drinks consumers can be segmented according to benefits sought/required by supporting the conclusion of the existing theory that every industry needs its own instruments to conduct researches for effecient market segmentation.

### 4.2 Study Implications

Results from this study can help the soft drinks industry in establishing marketing strategies by providing a clear picture of segmented groups.Having submitted our main goal that was segmentation of soft drinks consumers based on benefits sought/required, we provide a good analytical base for soft drink market segmentation, better positioning of the product on market and more effective advertising.

### 4.3 Study Limitations

Main limitations can be summarized in two issues; responders number or sample's size and study instrument construction.

In order to accomplish the study we surveyed 399 responders, soft drinks consumers. Therefore small number of responders represents the main limitation/constraint of the study and its applicability since it can not represent the entire population of Pristina city broadly, despite efforts for a right inclusion.

Another constraint of the study relates to the construction of the survey instrument, where the scales used are not uniform.These varying scales/degrees are used because of the different references used for study questions in questionnaire.

### 4.4 Sugestions for future studies

Forthcoming researches in this area should include various issues that are not included in this study.It is suggested that study instruments should include more issues/cases, and responders

Impact factor 2015 0, 76 ISBN: 10 12-3869

number should not be under 1000. This is because the more issues are addressed and as many respondents interviewed the more accurate results will be and their application in practice easier and more effective. Also geographical distribution of respondents must be conforming to the expansion of settlements and population number. Also, in future studies the ethnic composition of respondents should be considered, taking in consideration the cultural differences that exist between different ethnic communities, which may have a decisive influence on many issues related to overall consumption, and in particular the consumption of soft drinks.

It is also suggested to accomplish a general research using behaviorist variables for the food industry in general throughout Kosovo territory, of which the data for each industry separately will derive more easily later on.