

CLUSTERING OF DAIRY PRODUCTS CONSUMERS BASED ON BRAND CONSIDERATION: A CASE STUDY OF IRAN

Abbas Keramati^{1*}, Mehdi Keshavarz², & Abolghasem Yousefi-Babadi³

Received.....03/01/2016
Accepted.....06/15/2016
Published Online.....07/24/2016

^{1*} Department of Industrial Engineering, Faculty of Engineering, University of Tehran, keramati@ut.ac.ir

² Department of Industrial Engineering, Faculty of Engineering, University of Tehran, mehdikeshavarz@ut.ac.ir

³ Department of Industrial Engineering, Faculty of Engineering, University of Tehran, a.yousefibabadi@ut.ac.ir

ABSTRACT

An increasing demand of dairy product worldwide gives a good opportunity for dairy industries; and under the blows of globalization and pierce of technology to various corners, the competition amongst dairy industries are getting tougher and harder day-by-day. Dairy companies in order to sustain their competitiveness should segment and analyze customer attitudes and behaviors. This paper segments the customers in the Urmia city of Iran based on their attention and important degree of dairy products brand. The various customer characteristics such as age, income, gender and ethnicity are used for segmenting and clustering. Several dairy products such as cheese, butter, milk, chocolate milk, yogurt, Doogh and ice cream are utilized in this paper and for each of them a separate clustering has been performed. In this paper, the hierarchical approach is used to find the initial number of clustering, then by utilizing the K-mean algorithm the best cluster is calculated.

KEY WORDS: Dairy Product, Brand, Clustering, K-Mean Algorithm, Urmia

1. INTRODUCTION

Nowadays the consumers' demands in the field of food production have changed. They need the food not only to cure their hunger and to supply their body nutrients, but also to safeguard them against nutrition shortage disease and promote their physical and mental strength (Roberfroid 2000; Menrad 2003). Therefore, Consumers are increasingly believing that foods directly influence their healthcare (Young 2000; Mollet and Rowland 2002).

Dairy products are the foods which play important role in building healthy bones, maintaining a healthy weight and protecting your teeth and gums by reducing the risks of gum disease. Dairy products in this paper are defined as the branches of the food industry that involve processing milk for producing various dairy.

An increasing demand of dairy product worldwide gives a good opportunity for dairy industries; and under the blows of globalization and pierce of technology to various corners, the competition amongst dairy industries are getting tougher and harder day-by-day. Therefore, there is an opportunity and threat in dairy industries and they must perform the best effort for attracting customers, whereas this function without having a true cognition of customer's behavior and requirements will not be possible.

One of the best ways for understanding customer behavior is segmentation of markets and consumers. Smith (1956) technique of segmentation has become a dominating method of

modern marketing management and marketing research (Wedel and Kamakura 2002). Smith (1956) defined market segmentation as “involves viewing a heterogeneous market as a number of smaller homogeneous markets in response to differing product preferences among important market segments. It is attributable to the desires of consumers or users for more precise satisfaction of their varying wants”. Conceptually, market segmentation is meant to subdivide a market into distinct group of consumers who have similar conducts or have comparable requirements. Where every subset may be chosen as a market objective to be achieved with a defined marketing strategy (Bennett 1995). Market segmentation helps the managers for recognizing suitable segments in target marketing, better realizing the market, finding out opportunities in current trade and obtaining competitive advantage (Kotler and Turner 1979). As seen one of the important goals of segmenting is clustering costumers based on similar and common characteristic for preparing and developing best method for marketing.

Before starting marketing it must be clear that if marketing and advertising is really effective in selling a product, whether people pay attention to a brand of a product and for which group of consumers the brand of product is most important. Answering such questions help manager to focus on the group of consumers and the brand which is important for them.

The aim of this paper is segmenting the customers in the Urmia city of Iran based on their attention to brands and degree of its importance for them. The various customer characteristics such as age, income, gender and ethnicity are taken into account for segmenting and clustering. Several dairy products such as cheese, butter, milk, chocolate milk, yogurt, Doogh and ice cream are studied in this paper and for each of them a separate clustering has been performed.

2. LITERATURE REVIEW

Up to now, many researchers studied about clustering methods and application of clustering in the real life especially some researchers discuss about clustering method in the foods and dairy foods. Some literature on the clustering method and its application in the foods and dairy foods are explained as follows.

Windham et al. (1985) applied a mathematical clustering algorithm to classify foods within dairy, grain, and fat commodity groups. On the their paper, foods within dairy, grain, and fat commodity groups were clustered on the basis of similar content of vitamin B-6, calcium, iron, magnesium, folacin, zinc, and added sugar, fat, cholesterol, and sodium.

Austin et al. (2005) analyzed the concentration of fast food restaurants in region proximal to schools to characterize school neighborhood food environments. Geocoded databases of restaurant and school were used to perform the mentioned analyzed. Finally they could find that Fast-food restaurants were statistically significantly clustered in areas within a short walking distance from schools, with an estimated 3 to 4 times as many fast-food restaurants within 1.5 km from schools than would be expected if the restaurants were distributed throughout the city in a way unrelated to school locations.

Li et al. (2007) proposed a food clustering method for personalizing food alternative. Their presented approach would be helpful to nutritionists in creating new food groups.

Lingling (2010) carried out the grey relation clustering and PCA to evaluate the performance of vendors in fresh milk supply chain. He selected the three major fresh milk supply vendor and typical evaluation indices of fresh milk supply chain to illustrate the evaluation.

Rodenburg et al. (2013) presented the clustering of food and activity preferences in Dutch primary school children. To this study they used from 1480 parent-child dyads participating in the IVO Nutrition and Physical Activity Child cohort. Finally they found (1) a clustering of preferences for unhealthy foods and unhealthy drinks, (2) a clustering of preferences for various physical activity behaviors, and (3) a clustering of preferences for unhealthy drinks and

sedentary behavior. Also they shown the boys had a higher cluster score than girls on all three preference clusters.

In a similar research Buck et al. (2013) presented the clustering of unhealthy food around German schools and its influence on dietary behavior in school children. They examined this clustering in a pilot place and analyzed its results; they understood that food stores and fast food restaurants do not significantly cluster around schools in the 95% confidence level.

Chifu et al. (2015) introduced the clustering of documents containing food recipes as an optimization problem solved using a Flower Pollination-based method.

Joo et al. (2015) studied the distribution fast food density outlets around schools, and the relationship between dietary health of children and adolescents and the fast food density. They found that the number of schools with a fast food outlet within 200 meters or in the Green Food Zone around its location was 48 of 189 (25.4%) in Suwon and 14 of 153 (9.2%) in Hwaseong and Osan. Students in the low-density group visited fast food outlets less often than those in the high-density group ($p < 0.01$). Also they understood that the distribution map of fast food outlets within 200 meters of schools was useful for identifying the effectiveness of the Green Food Zone Act and nutrition education programs.

3. METHODOLOGY

3.1. Method of analysis

Clustering is one of the most powerful techniques for discovering of the groups and natural relationship among data in such groups and also discovering the structural pattern of them without having any background of data characteristic (Michael 2007).

Clustering (Everitt et al. 2001; Xu and Wunsch 2005; Chang et al. 2009) is widely used technique where a number of patterns, usually vectors in a multidimensional space, are employed for recognizing clusters of similar characteristics.

A traditional classification of clustering algorithms is done by hierarchical and partitioning methods (Everitt et al. 2001; Xu-Wei and Min 2013). Each of these methods utilizes different algorithms. One of the main and evident differences between hierarchical and partitioning methods is in definite or indefinite number of clusters. In hierarchical clustering method the number of clusters is indefinite but in Partitioning method the number of clusters is definite. Because the number of the clusters is usually indefinite, therefore in this paper first the hierarchical method is utilized to find out the best number of clusters, then, K-means algorithm is apply for clustering.

K-mean algorithm is the most popular partitioning algorithm among various clustering algorithms because it is simple, easy to use and very efficient in dealing with large amount of data with linear time complexity (Chen and Ye 2004). Although K-mean is very popular algorithm but it is very sensitive to the initial cluster centers (González and Tou 1974).

In this paper, the hierarchical approach is used to find the initial number of clustering, then by utilizing the K-mean algorithm the best cluster is calculated.

3.2. Research data

We prepared a questionnaire consisting of two sections. The first section contained of 7 questions about the degree of brand importance for the cheese, butter, milk, chocolate milk, yogurt, Doogh and ice cream dairy product, respectively. The second section contained demographic questions such as age, gender, income and ethnicity.

Data were collected by questionnaires in the Urmia city of Iran. First about 60 questionnaires have been distributed to the residents of Urmia city for pre-test and Pilot Study. After reviewing the results and feedback obtained through the questionnaire, initial questionnaire was completed and necessary reformation was applied to them. The final version of the questionnaire was distributed to 300 residents and was used for further analysis.

3.3. The Sample

In order the questionnaire to represent the entire community, it has been tried to distribute the questionnaire among people with respect to population pyramid and ethnic composition of the city of Urmia. Table 1 presents the detailed demographic characteristics of the sample. As seen in the Table 1 the proportion of male and female is roughly equal. The age 21-30 have higher proportion than other ages in the sample due to Iran's population being roughly young and also young people usually buy the products. The Urmia city's inhabitants are predominantly Iranian Azerbaijanis who speak the Azerbaijani language and also minorities of Kurds and Fars. Therefore, it will be rational that the Azerbaijani language have high ratio in the sample.

Table 1- Demographic characteristics of the sample

Demographics	Frequently	Percentage s
Gender		
✓ Male	166	55.3
✓ Female	134	44.7
Age		
✓ 15-20	11	3.7
✓ 21-30	145	48.3
✓ 31-40	85	28.3
✓ 41-55	47	15.7
✓ Over 55	12	4
Income		
✓ Less than 500	171	57
✓ 500-1000	96	32
✓ 1000-1500	22	7.3
✓ Over 1500	8	2.7
ethnic		
✓ Azary	223	74.3
✓ Kord	47	15.7
✓ Fars	24	8
✓ Else	6	2

3.4. Reliability

Cronbach's alpha is the most common method of reliability measurement. It is most commonly used when you have multiple questions in questionnaire. Alpha coefficient ranges in value from 0 to 1 and Nunnally (1978) has indicated 0.7 to be an acceptable reliability coefficient. Reliability coefficients (Cronbach's alphas) for the questionnaires are 0.762. The Cronbach's Alpha of the questionnaires is above 0.7 therefore, the questionnaires are reliable.

4. RESULT

In this section the result for clustering of dairy product is presented.

4.1. Ice cream

Table 2 shows the result for the clustering of ice cream dairy product. As shown in Table 2, there are 3 clusters for ice cream products as follows:

- ✓ Cluster 1: It is comprised of 31 to 55 years-old Azary female and male with medium income for whom the significance of brand is important.
- ✓ Cluster 2: It is comprised of 21 to 30 years-old Azary female with low income which the brand of ice cream product is not important for them. Therefore, it is better for dairy company to have no advertisement for customers with such characteristics. Because this kind of customer chooses ice cream products without looking at their brand.
- ✓ Cluster 3: It is comprised of the 21 to 30 years-old Azary female with low income which the significance of brand is very important for them. Therefore, it is better for dairy company to focus their advertisement to this kind of customers.

From this 3 cluster it can be deduced that the brand of ice cream for males is more significant than females. Another important point is that there is no significance of brand for young female with low income.

Table 2- Result for the clustering of ice cream dairy product

Cluster NO.	Brand significant	Gender	Age	Ethnicity	Income
1	Important	Female/male	31-55	Azary	Under 1500
2	No significant	Female	21-30	Azary	Under 500
3	Very important	Male	21-30	Azary	Under 500

4.2. Milk

Table 3 shows the result for the clustering of milk product dairy product. As shown in Table 3, there are 4 clusters for milk products as follows:

- ✓ Cluster 1: It is comprised of the 21 to 30 years-old Azary male with low income which for whom the significance of brand is important for them.
- ✓ Cluster 2: It is comprised of the 31 to 40 years-old Azary male with medium income which the brand of milk product is not important for them. Therefore, it is better for dairy company to have no advertisement for customers with such characteristics. Because this kind of customer choose milk product without looking at their brand.
- ✓ Cluster 3: It is comprised of the 21 to 55 years-old Azary male with low income and which of brand is very important for them. Therefore, it is better for dairy company to have no advertisement for customers with such characteristics.
- ✓ Cluster 4: It is comprised of the 21 to 40 years-old Azary female with medium income which the significance of brand is very important for them. Therefore, it is better for dairy company to focus their advertisement to this kind of customers.

Table 3- Result for the clustering of milk dairy product

Cluster NO.	Brand significant	Gender	Age	Ethnicity	Income
1	Important	Male	21-30	Azary	Under 500
2	No significant 1	Male	31-40	Azary	500-1000
3	No significant 2	Female	21-30/41-55	Azary	Under 500
4	Very Important	Female	21-40	Azary	1000-1500

4.3. Cheese

Table 4 shows the result for the clustering of cheese product. As shown in Table 4, there are 3 clusters for cheese products as follows:

- ✓ Cluster 1: It is comprised of the 21 to 30 years-old Azary female with low income for whom the significance of brand is important.
- ✓ Cluster 2: It is comprised of the 31 to 55 years-old Azary male with medium income for whom the significance of brand is important.
- ✓ Cluster 3: It is comprised of the 31 to 40 years-old Azary female with medium income for whom the significance of brand is not important for them. Therefore, it is better for dairy company to have no advertisement for customers with such characteristics. Because this kind of Customer choose the cheese product without looking at the brand of it.

From this 3 cluster it can be concluded that the significant of cheese is not very important in none of clusters. Also it can be said that the brand of cheese is more important from male rather female. Another important point is that the significant of brand cheese for young female with low income is important but for female with higher income and age, there is no significant of cheese brand. One could say that by increasing the age and income of female, their attention and significant for brand of cheese product will be decrease.

Table 4- Result for the clustering of cheese dairy product

Cluster NO.	Brand significant	Gender	Age	Ethnicity	Income
1	Important 1	Female	21-30	Azary	Under 500
2	Important 2	Male	31-55	Azary	1000-1500
3	No significant	Female	31-40	Azary	Under 1500

4.4. Butter

Table 5 shows the result for the clustering of butter product. As shown in Table 5, there are 2 clusters for butter products as follows:

- ✓ Cluster 1: It is comprised of the 21 to 30 and 41-50 years-old Azary female and male with medium income for whom the significance of brand are not important Therefore, it is better for dairy company to have no advertisement for customers with such characteristics. Because this kind of Customer choose the butter product without looking at the brand of it.
- ✓ Cluster 2: It is comprised of the 21 to 30 years-old Azary male with low income characteristics.

From this 2 cluster it can be result that the brand of butter is less important from male rather female. Another important point is that the significant of brand butter for young female with low income is important but for female with higher income and age, there is no significant of butter brand. One could say that by increasing the age and income of female, their attention and significant for brand of butter product will be decrease. This result is as same as the result of cheese.

Table 5- Result for the clustering of butter dairy product

Cluster NO.	Brand significant	Gender	Age	Ethnicity	Income
1	No Significant	Female/Male	21-30/ 41-50	Azary	Under 1500
2	Important	Female	21-30	Azary	Under 500

4.5. Yoghurt Error! Reference source not found. shows Result for the clustering of yoghurt dairy product. As shown in Error! Reference source not found., there are 2 clusters for yoghurt products as follows:

- ✓ Cluster 1: It is comprised of the 41 to 50 years-old Azary female/male with medium income which the significant of yoghurt product are not significant for them. Therefore, it is better for dairy company to have no advertisement for customers with such characteristics. Because this kind of customer choose the yoghurt product without looking at the brand of it.
- ✓ Cluster 2: It is comprised of the 21 to 30 years old Azary female/male with low income which the significant of brand is important for them.

As the result of this 2 cluster, it is obvious that the brand of yoghurt for young customer is more significant than middle-age.

Table 6- Result for the clustering of yoghurt dairy product

Cluster NO.	Brand significant	Gender	Age	Ethnicity	Income
1	No Significant	Female/Male	41-50	Azary	Under 1000
2	Important	Female/Male	21-30	Azary	Under 500

4.6. Milk Chocolate **Error! Reference source not found.** shows the result for the clustering of milk chocolate dairy product. As shown in **Error! Reference source not found.**, there are 2 clusters for milk chocolate products as follows:

- ✓ Cluster 1: It is comprised of the 41 to 50 years-old Azary female/male with medium income which the significant of ice cream product are not significant for them. Therefore, it is better for dairy company to have no advertisement for customers with such property. Because this kind of customer choose the milk chocolate product without looking at the brand of it.
- ✓ Cluster 2: It is comprised of the 21 to 30 years old Azary female/male with low income which the significant of brand is important for them.

As the result of this 2 cluster, it is obvious that the brand of milk chocolate for young customer is more significant than middle-age.

Table 7- Result for the clustering of Milk Chocolate dairy product

Cluster NO.	Brand significant	Gender	Age	Ethnicity	Income
1	No Significant	Female/Male	41-50	Azary	Under 1000
2	Important	Female	21-30	Azary	Under 500

4.7. Doogh

Table 8 shows the result of the clustering of Doogh product. As shown in Table 8, there are 4 clusters for Doogh products as follows:

- ✓ Cluster 1: It is comprised of the 31 to 40 Azary female with low income which the significant of brand is very important for them. Therefore, it is better for dairy company to focus their advertisement to this kind of customers.
- ✓ Cluster 2: It is comprised of the 21 to 40 Azary female with medium income which the significant of brand is important for them.
- ✓ Cluster 3: It is comprised of the 21 to 40 Azary female with medium income which the significant of Doogh product are not significant for them. Therefore, it is better for dairy company to have no advertisement for customers with such characteristics. Because this kind of customer choose the Doogh product without looking at the brand of it.
- ✓ Cluster 4: It is comprised of the 41 to 55 Azary male with low income which the significant of brand is medium for them

Table 8- Result for the clustering of Doogh dairy product

Cluster No.	Brand significant	Gender	Age	Ethnicity	Income
1	Very important	Female	31-40	Azary	Under 500
2	Important	Female	21-30/31-40	Azary	Under 1500

3	No significant	Male/female	21-30/31-40	Azary	Under 1500
4	medium	Male	41-55	Kurd	Under 500

5. CONCLUSION

This paper clusters the dairy product according to criteria such as age, income, gender and ethnicity based on their attention to brands and degree of its importance for consumers. For clustering hierarchical approach is utilized to find out the initial number and then K-mean algorithm is used. Base on the results, it can be concluded that brand are more significant for young people with low incomes than those with higher age and income.

REFERENCES

- Austin, S. B., Melly, S. J., Sanchez, B. N., Patel, A., Buka, S. and Gortmaker, S. L. (2005). Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments. *American Journal of Public Health*, Vol. 95, No. 9, 1575-1581.
- Bennett, P. D. (1995). *Dictionary of marketing terms*, Contemporary Books.
- Buck, C., Börnhorst, C., Pohlmann, H., Huybrechts, I., Pala, V., Reisch, L. and Pigeot, I. (2013). Clustering of unhealthy food around German schools and its influence on dietary behavior in school children: a pilot study. *International Journal of Behavioral Nutrition and Physical Activity*, Vol. 10, No. 1, 65.
- Chang, D.-X., Zhang, X.-D. and Zheng, C.-W. (2009). A genetic algorithm with gene rearrangement for K-means clustering. *Pattern Recognition*, Vol. 42, No. 7, 1210-1222.
- Chen, C.-Y. and Ye, F. (2004). Particle swarm optimization algorithm and its application to clustering analysis. *Networking, Sensing and Control*, 2004 IEEE International Conference on, IEEE.
- Chifu, V. R., Pop, C. B., Salomie, I. and Aczel, A. (2015). CLUSTERING FOOD RECIPES USING A FLOWER POLLINATION-BASED METHOD. *Acta Technica Napocensis*, Vol. 56, No. 4, 42.
- Everitt, B., Landau, S. and Leese, M. (2001). *Cluster Analysis* Arnold. A member of the Hodder Headline Group, London, No.
- González, R. and Tou, J. (1974). *Pattern recognition principles*. Applied Mathematics and Computation. Reading (MA): Addison-Wesley, No.
- Joo, S., Ju, S. and Chang, H. (2015). Comparison of fast food consumption and dietary guideline practices for children and adolescents by clustering of fast food outlets around schools in the Gyeonggi area of Korea. *Asia Pacific journal of clinical nutrition*, Vol. 24, No. 2, 299.
- Kotler, P. and Turner, R. E. (1979). *Marketing Management: analysis, planning, and control*, Prentice-Hall Englewood Cliffs, NJ.
- Li, H.-C., Ko, W.-M. and Tung, H.-W. (2007). Food Clustering Analysis for Personalized Food Replacement. *Fuzzy Information Processing Society*, 2007. NAFIPS'07. Annual Meeting of the North American, IEEE.
- Lingling, X. (2010). Applying grey relation clustering and PCA to performance evaluation of vendors in fresh milk supply chain. *Logistics Systems and Intelligent Management*, 2010 International Conference on, IEEE.

- Menrad, K. (2003). Market and marketing of functional food in Europe. *Journal of food engineering*, Vol. 56, No. 2, 181-188.
- Michael, W. B. (2007). *Survey of Text Mining: Clustering, Classification and Retrieval*, Published by Springer, USA.
- Mollet, B. and Rowland, I. (2002). Functional foods: at the frontier between food and pharma. *Current Opinion in Biotechnology*, Vol. 13, No. 5, 483-485.
- Nunnally, J. (1978). C.(1978). *Psychometric theory*, New York: McGraw-Hill.
- Roberfroid, M. B. (2000). A European consensus of scientific concepts of functional foods. *Nutrition*, Vol. 16, No. 7, 689-691.
- Rodenburg, G., Oenema, A., Pasma, M., Kremers, S. P. and van de Mheen, D. (2013). Clustering of food and activity preferences in primary school children. *Appetite*, Vol. 60, No. 123-132.
- Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *The Journal of Marketing*, No. 3-8.
- Wedel, M. and Kamakura, W. A. (2002). Introduction to the special issue on market segmentation. *Intern. J. of Research in Marketing*, Vol. 19, No. 181-183.
- Windham, C. T., Windham, M. P., Wyse, B. W. and Hansen, R. G. (1985). Cluster analysis to improve food classification within commodity groups. *Journal of the American Dietetic Association*, Vol. 85, No. 10, 1306-1314.
- Xu-Wei, P., and Min, J. (2013). Wnbac: A Weighted Network Based Adaptive Clustering Algorithm for Spatial Objects. *Information Technology Journal*, Vol. 12, No. 23, 7849.
- Xu, R., and Wunsch, D. (2005). Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*, Vol. 16, No. 3, 645-678.
- Young, J. (2000). Functional foods and the European consumer. *SPECIAL PUBLICATION-ROYAL SOCIETY OF CHEMISTRY*, Vol. 248, No. 75-81.