



## Using Clustering To Boost Sales In Chain Stores:A Real Case In Iran

Peyman Gholami\*

Department of Industrial Engineering,  
Arak Branch, Islamic Azad University,  
Arak, Iran  
Peyman711@gmail.com

Meysam Eftekhary

Department of Industrial Engineering,  
Arak Branch, Islamic Azad University,  
Arak, Iran  
maysameftekhary@yahoo.com

Nezamedin Faghieh

Department of management, Shiraz University  
Iran  
NezamFaghieh@gmail.com

**Abstract:** the volumes of data are doubling every four years and the price of data strongly is reducing. For using of these data, so many of instruments and techniques are being applied. One of these kinds of instruments is data mining technique. That this technique can extract a new and understandable pattern through a high volume of data and it has been used in business. Discovering the rules of selling with considering the appropriate time for it causes a high increase in selling of products. So that with considering the importance of the time of special record, profitability will be increased. In this paper, with using clustering techniques, we have tried to analyze data that have been recorded in hyperstar store in 2010. Our main objective is extracting the rules and analyzing the volume of different products

**Keywords:** Data Mining, Clustering, K-means Algorithm, Chain Stores

### I. INTRODUCTION

Data mining is a technique that can help the new models, understandable, reliable and potentially useful in the search and found a huge collection of data. The main reason that data mining has been attracted by information industry in recent years is that to deal with high volume of data in wide range and there is absolute need of transferring data to information. Data mining tools, predict future trends and, behaviors and will help managers in appropriate decision making[1,3]. Data mining tools can be answer to the questions which in the past, it needs too much time for them to be answered. Data Mining is the combination of data management, statistics, machine learning and data illustrated, is. to examine and analyze the Customers shopping cart. they use common techniques in the field of data mining techniques such as the dependency rules, clustering, etc[5].

The main objective of dependency rules is to discover the relationship among different items. It also can be used to discover Customer buying patterns. The culture, needs and affordability of customers in different areas are various so that these differences influence on the amount and the way of cash purchases, using credit cards and bank account the customers who has a high ability of purchasing and has a high income comparing with other people, They have a different behavior of purchasing. with using of clustering techniques and analyzing dependency rules and considering the volume of purchasing by customers in different areas, the Customer buying patterns, can be extracted and analyzed.[2,4] This paper is arranged in six sections that the K-means, clustering techniques and algorithms has been described in Section 2, the question raised in Section 3, and research method in

Section 4 and in Section 5 data has been analyzed and in Section 6 we came to Conclusion.

### II. MATERIALS AND METHODS

#### THE BACKGROUND OF CLUSTERING AND K-MEANS ALGORITHM

In today world, data are considered as one of the most valuable assets. With the current dramatic increase in magnitude of available data and also their low cost storage, it became interesting to discover knowledge in these data. Therefore, the importance of how to efficiently structures from large amounts of data stored in databases, data warehouses, or other information repositories (Liao, Chen, & Wu, 2008). In the literature, there are many data mining models such as classification, estimation, predictive modeling, clustering, affinity grouping or association rules, description and visualization, as well as sequential modeling. Clustering is a widely used technique, whose objective is to provide insight into the data by partitioning the data (objects) into disjoint and homogeneous groups (clusters) of objects, such that objects in a cluster are more similar to each other than to objects in other clusters. According to Boutsinas and Gnardellis (2002), clustering algorithms have been frequently studied in various fields including machine learning, neural networks and statistics, among others (Fensel, 2001; Corcho, Lopez, & Perez, 2003; Davies & Fensel, 2003).

The k-means algorithm, first proposed by MacQueen (1967), is the most popular partition-clustering method that has attracted great interest in the literature. The goal of the k-means algorithm is to partition the objects into k clusters so

that the within-group similarity is maximized. The procedure of k-means methods could be described as follows.

- a. Place k points into the space represented by the objects that are being clustered. These points represent initial group centroids.
- b. Assign each object to the group that has the closest centroids.
- c. When all objects have been assigned, recalculate the positions of the k centroids.
- d. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

### III. THE EXPRESSION OF PROBLEM

In respect to this fact that in the modern world, most of the organizations face with vast range of data and unprocessed records and using efficiently from these data with ancient analysis methods is so hard and time consuming. Using of new data mining can be effective for solving this problem. With the help of data mining techniques such as clustering, data existing in database , will process and change usable and useful and Because data mining is the process of extracting knowledge from data it is performed through discovering patterns that is existed in data. by using clustering algorithms, customers in respect to their shopping cart, proposed clustering indicators , their buying patterns and their clusters, are being discovered ,. Here there is the challenge, and today, most of the chain stores have vast datasets that is included recorded data. And there is not a appropriate usage of these datasets and all data are just being saved, on the other hand, most of the world wide chain stores in order to increasing quality and quantity of their services and increasing their benefits ,They are using scientific data analysis in the new knowledge of data mining and have been achieved considerable success .in this paper, in order to increase satisfaction and the quality of served services in the store, by considering the time situation and the place of customers, the volume of sales in the hyperstar store is being analyzed and clustered .then in any cluster, the dependency rules has been discovered and analyzed.

### IV. RESEARCH METHOD

In this paper , k –mean clustering algorithm ,applied on the data of recording sales that has been done in shiraz hyperstar store over last year(2010). The output of k- mean algorithm includes 6 clusters that the average of sales in any of clusters in the form of percentage of selling in table 1 has been presented. Also in table2, each of clusters has been named according to this table, a cluster 1containing dairy products, cluster 2 containing beverage products ,cluster 3 containing food products, cluster 4 containing nut products, cluster 5

containing cleaning products and cluster 6 containing cosmetic products . in the next section , percentage of selling each of the clusters with related charts will be analyzed and then the factors that affect the profitability of most stores, to be discovered.

Table 1 : The percentage of sales in clusters in 12 months of a year

Cluster						Month
6	5	4	3	2	1	1
7	10	21	20	15	21	1
8	10	17	16	19	17	2
11	13	16	19	16	20	3
10	14	14	14	16	15	4
14	16	17	17	18	17	5
14	16	17	17	18	17	6
13	15	18	18	19	18	7
13	14	18	18	19	19	8
13	12	19	19	18	18	9
12	10	17	20	19	21	10
10	10	20	20	19	21	11
10	10	20	19	20	20	12

Table 2 : The names of clusters

Cluster1	dairy products
Cluster2	beverage products
Cluster3	food products
Cluster4	nut products
Cluster5	cleaning products
Cluster6	cosmetic products

### V. ANALYSIS OF DATA

In this section, the percentage of selling of each of clusters in 12 months of year with the related chart has been analyzed .and then the factors that cause more sales and more profitability in hyper star store, to be discovered. In respect to chart 1, as It is being seen, selling of dairy products in winter and spring in comparison with summer and fall has been increased, also it has been shown, that the best selling of dairy products is related to may and June and most selling is related to February. From this analysis we can come to conclusion that if the store increase the volume of its sales of dairy products in summer and fall and reduce the volume of its sales of this product in winter and spring, the hyperstar store can make more profit from dairy product.

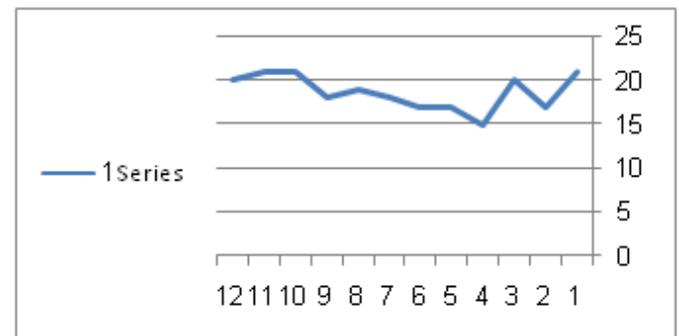


Chart 1: sales volume of dairy products in 12 months of a year

In chart 2, the sales of beverage product in 12 months of a year has been shown. According to this chart the selling of

beverage products in spring and summer has a maximum percent and in winter has a minimum percent. In result if the hyper star store increase the volume of its sales in spring and summer and reduce in winter and fall, they can make more profit from beverage product.

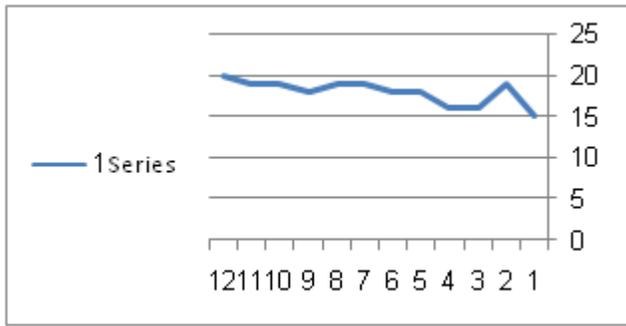


Chart 2: sales volume of beverage products in 12 months of a year

In chart 3 the sales of food products in 12 months of a year can be seen. According to this chart, the selling of food products in all of the season is between 17 to 20 this means, the volume of sales of this product in 12 months of a year is uniformly in the peak. Also, in can be seen, selling of this product in August, October and November is more than other months. All things that can be received from this analysis is that, it is better to increase the volume of sales of food products in all over the year especially in August, October and November.

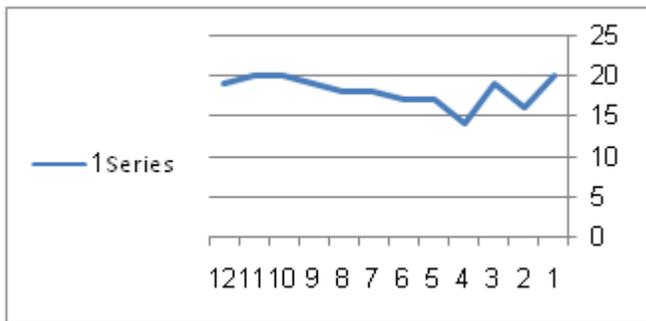


Chart 3: sales volume of food product in 12 months of a year

In chart 4, the sale of nut products in all over the year has been shown. According to this chart, the selling of nut product in all of the season are approximately the same. It means the selling of this product uniformly is in its peak .and it being observed, that the selling of this product in November, December and January has been reached the maximum percent and in march has been reached the minimum percent. All things that can be received from this analysis is that it is better to increase the volume of sales of nut product in 12 months of a year especially in November December and January and moderately, reduce in march.

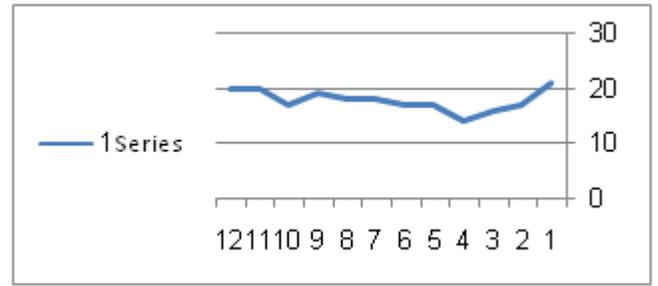


Chart 4: sales volume of nut products in 12 months of a year.

In chart 5, the sales of cleaning products in all months of a year have been shown. According to this chart ,the selling of cleaning products in fall and winter have been reached the maximum percent and in spring and summer have been reached the minimum percent but in total the percentage of selling of this product in comparison with previous products has a least percent so it can be concluded, if hyperstar store ,in comparison with previous products, reduce the volume of sales of cleaning product in all over the year and also, increase the volume of sales that it has been assigned to this product in spring and summer in comparison with winter and fall the profits from cleaning product will be much more .

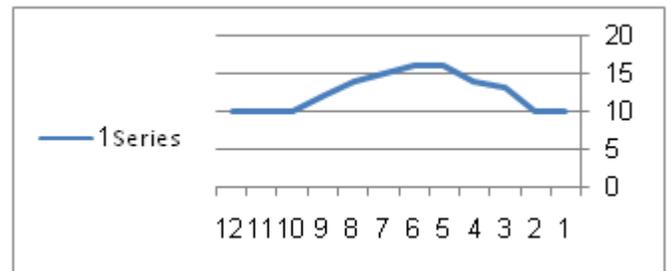


Chart 5: sales volume of cleaning products in 12 months of a year

In chart 6 the sales of cosmetic products in 12 months of a year has been shown in respect to this chart, the selling of cosmetic chart has a highest percent in spring and in other seasons, uniformly, has a least percent, but in total, the selling of this product in comparison with the products that have been considerate in cluster 1 to 4, is so low and trivial so it can be concluded, if the hyperstar store, in comparison with the products that have been investigated in cluster1 to 4,reduce the volume of sales of cosmetic products during a year ,and also, in comparison with other seasons, increase the assigned volume of this product in spring . The hyperstar store can make more profit from cosmetic product.

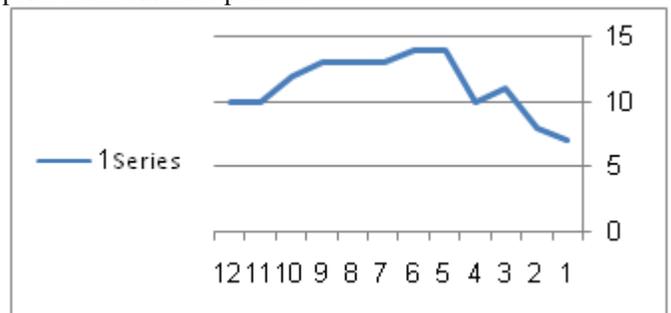


Chart 6: sales volume of cosmetic products in 12 months of a year.

## VI. CONCLUSION

The main objective of this paper is optimization the process of selling in chain stores to make more profit by using data mining techniques that it has been done by exploiting k-mean clustering algorithm on existing data in hyperstar store which dataset administrator had saved the selling of each product in data set

The output of this algorithm contains 6 clusters, that the volume of sales in a year in every cluster has been expressed in percentage. after analyzing the clusters, we concluded that if the hyperstar store decide to increase Sales volume of the first four clusters which include dairy products, beverage products , food products, and nut products, the average of selling will be increased and finally he can make more profit. Also, if hyperstar store, in comparison with winter and fall, increase the volume of sales in spring and summer and assign a less volume of its sales to cosmetic products and instead of this product, increase the volume of sales of clusters number 1 to 4 it would be more profitable.

## VII. REFERENCES

- [1] Yen-Liang Chena, Kwei Tangb, Ren-Jie Shena, Ya-Han Hua, “Market basket analysis in a multiple store environment”, Available online at [www.sciencedirect.com](http://www.sciencedirect.com) , Decision Support Systems, No 40 , 339–354, 2004.
- [2] Yasemin Boztug, Thomas Reutterer, “A combined approach for segment-specific market basket analysis”, European Journal of Operational Research, 2007.
- [3] R.J. Kuo , S.Y. Lin a, C.W. Shih ,” Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan”, Expert Systems with Applications , No 33 , 794–808, 2007.
- [4] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur. “Dynamic itemset counting and implication rules for market basket analysis.” In Proc. 1997 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD’97), pages 255–264, Tucson, Arizona, May 1997.
- [5] Lu ´ s Cavique ,” A scalable algorithm for the market basket analysis”, Journal of Retailing and Consumer Services, 2007.
- [6] Liao, S. H., & Chen, Y. J. (2004). Mining customer knowledge for elec-tronic catalo marketing. Expert Systems with Applications, 27, 521–532.
- [7] Liao, S. H., Chen, C. M., & Wu, C. H. (2008). Mining customer know-ledge for product line and brand extension in retailing. Expert Systems with Applications, 35(3), 1763–1776.
- [8] Boutsinas, B., & Gnardellis, T. (2002). On distributing the clustering process. Pattern Recognition Letters, 23, 999–1008
- [9] Fensel, D. (2001). Ontologies: A silver bullet for knowledge manage-ment and electronic commerce. New York: Springer.
- [10] Corcho, O., Lopez, M. F., & Perez, A. G. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? Data and Knowledge Engineering, 46, 41–64.
- [11] Davies, J., & Fensel, D. (2003). Toward the Semantic Web: Ontology driven knowledge management. John Wiley & Sons Ltd
- [12] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of 5-th Berkeley symposium on mathematical statistics and probability (vol. 1, pp. 281–297). Berkeley: University of California Press.