THE REPUBLIC OF TURKEY
BAHÇEŞEHİR UNIVERSITY

COMPARING CLUSTERING METHODS BASED ON LIFESTYLE TRENDS

Master Thesis

DATEV YAZDANOGLU

İSTANBUL, 2011

THE REPUBLIC OF TURKEY

BAHÇEŞEHİR UNIVERSITY

THE GRADUATE SCHOOL OF SOCIAL SCIENCES

MARKETING

COMPARING CLUSTERING METHODS BASED ON LIFESTYLE TRENDS

Master Thesis

DATEV YAZDANOGLU

Thesis Supervisor: ASSOCIATE PROF. DR. HUSEYIN INCE

İSTANBUL, 2011

FOREWORD

The completion of this dissertation was made possible through support and encouragement of many individuals, and I would like to take this opportunity to acknowledge them.

I would like to express my deepest appreciation and thanks for my advisor, Associate Prof. Dr. Huseyin INCE, for his guidance, support, helpful suggestions and endless enthusiasm for this study.

Very special thanks to BAREM research for significant support and helpful comments during the analysis stage of this study and gathering the life style survey data.

I owe my special thanks to my parents, my brother and my grandmother for them extensive help throughout this study.

Finally, I would like to thank beloved Natali BAPUCOGLU for her endless encouragement and support. This dissertation is dedicated to her.

ISTANBUL, 2011 **Datev YAZDANOĞLU**

ABSTRACT

COMPARING CLUSTERING METHODS
BASED ON LIFESTYLE TRENDS

Yazdanoglu, Datev

Thesis Supervisor : Associate Prof. Dr. Huseyin Ince

January 2011, 73 Pages

Cluster analysis is one of the most widely used methods to market segmentation which means that dividing a market into distinct subsets of consumers with common needs or characteristics. However, variety of factors are affecting the consumers' decision and purchasing behaviour directly. The bases of these factors are respect of demographical, geographical, socio cultural and psychological causes.

In this research three clustering approaches, k-means, two-stage and self organizing map, are analyzed on the life style trend data set. The two main objectives of this study is identifying the possible consumer segments based on their life style trends and comparing the performance of three clustering approaches. Trend Statements were measured to cluster the consumers and three potential consumer segments, which were named as consistent, rebellious and traditional, were determined. After selecting the segments, the demographical characteristics of each segment are examined.

It can be mentioned that the consumers' life style related concerns are significantly more important during the segmentation of market. While consistent segment are more open to eco living however not interested in living on the net and being master of the youniverse. The rebellious segment are more open to accelerated society trend and has an aim to living on the net and being master of the youniverse. On the other hand conventional group are close to erosion of trust, self reflection and culturel diversity trends.

Keywords: Cluster Analysis, Market Segmentation, Life Style Trends, Kohonen Networks

ÖZET

KÜMELEME ANALİZİ YAKLAŞIMLARININ YAŞAM BİÇİMİ TRENDLERİ BAZINDA KARŞILAŞTIRILMASI

Yazdanoglu, Datev

Tez Danışmanı : Doç. Dr. Huseyin Ince

Ocak 2011, 73 Sayfa

Kümeleme analizi, pazar segmentleri için sıkça kullanılan metodlardan biri olup, pazarın tüketicilerin genel ihtiyaç veya özellikleri açısından farklı gruplara ayrılması anlamına gelmektedir. Ancak faktör çeşitliliği, tüketicinin karar ve satın alma davranışını etkilemektedir. Bu faktörlerin temelleri demografik, coğrafi, sosyo kültürel ve psikolojik etmenlere dayanmaktadır.

Bu araştırmada, üç çeşit kümeleme yaklaşımı, k-means, two stage ve self organizing map, yaşam biçimleri veri seti üzerinde analiz edilmiştir. Yapılan çalışmanın amaçlarından ilki, muhtemel tüketici segmentini yaşam biçimlerine göre uygulamak, diğeri ise üç kümeleme yaklaşımlarının performansını karşılaştırmaktır. Çalışmanın uygulamada aşamasında tüketici davranışlarını kümelemek için yaşam biçimleri ifadeleri kullanılmıştır. Uygulama sonucunda oluşan üç küme asi, uyumlu ve geleneksel olarak adlandırılmıştır. Segmentler seçildikten sonra her bir segmentin demografik özellikleri incelenmiştir.

Pazar araştırmasında tüketicilerin yaşam biçimleri ile ilgili durumların çok önemli olduğu açıkça belirtilebilir. Uyumlu olarak nitelendirilen grup, çevreye karşı daha duyarlı ancak internette yaşamayı tercih etmeyip risk almaktan kaçınmaktadırlar. Asi grup ise, hayatı ve zamanı yakalamaya çalıştıkları için, internette yaşamak kaçınılmaz bir sonuç olmaktadır ve kendilerine olan güvenleri yüksektir. Diğer yönden, geleneksel grup kurumlara karşı güven problemi yaşamalarına rağmen kültürel farklılıklara açık ve başkaların karşı saygılı olmaktadırlar.

Anahtar Kelimeler: Kümeleme Analizi, Pazar Segmentasyonu, Yaşam Biçimi Trendleri, Kohonen Networks

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The most challenging concept in marketing deals with understanding consumer behaviour. The term of consumer behaviour is defined as the behaviour that consumers display in *searching for, purchasing, using, evaluating and disposing of products and services that they expect will satisfy their needs* (Schiffman & Kanuk, 2000). In today's dynamic marketplace, in order to succeed in any business marketers need to know everything about consumers - what they want, what they think, how they work, how they spend their time. Having this knowledge and strong understanding of consumer behaviour would create successful marketing activities and the result of this rapidly evolving in the marketplace.

The main objective of this study is comparing three widely applied clustering techniques, k-means clustering, two stage clustering and self organizing map (SOM) clustering, to market segmentation which is one of the fundamental concepts of marketing to understand the consumer behaviour and determine the proper market strategy. These three approaches were performed on the real life-style research data set. The experimental results based on a life-style survey data set compared in respect of segments, determined by k-means, two stage and self organizing map clustering methods.

Although there are many hierarchical clustering approaches, in this study we applied k-means, two stage non-hierarchical clustering and self organizing map approaches. K-means clustering is the most frequently used and dominant technique for segmenting large datasets in the marketing area among the clustering areas (Gehrt & Shim, 1998). The method defines a fixed number of clusters, iteratively assigns records to clusters, and adjusts the cluster centers until further refinement can no longer improve the model. K-means clustering approach search for the optimum center locations to minimize the total distance between the data and the centers based on the Euclidean distance. The main reason to prefer this approach is due to the faster computational algorithm, respect to the presence of outliers and producing tighter clusters than other hierarchical clustering methods (Mashor, 1998).

The other clustering method performed to classify the data is a non-hierarchical two stage clustering. If the number of clusters is unpredictable the user can choose two-stage clustering procedure. Two-stage is a clustering method that involves preclustering the records into a large number of subclusters and then applying a hierarchical clustering technique to those subclusters to define the final clusters. The first step is formation of preclusters which will be input of the second stage of hierarchical clustering. Preclusters are just clusters of the original cases that are used in place of the raw data to reduce the size of the matrix that contains distances between all possible pairs of cases. In the second step standard hierarchical clustering algorithm is used on the preclusters in order to merge the subclusters into larger and larger clusters (Huang, 1998). Two stage clustering approach has the advantage of automatically estimating the optimal number of clusters for the training data and it can handle large datasets or datasets that have a mixture of continuous and categorical variables.

Self organizing map, which was introduced by Kohonen, is a special type of neural network to define clusters and transform into visually decipherable maps in large data bases. Basically, the self organizing map network model consists of input and output layers, which is visually examining the relationship between input data and output data for identifying important patterns and clusters. When the network is fully trained, records that are similar should appear close together on the output map, while records that are different will appear far apart. Because of outperforming the traditional data reduction and clustering techniques and because of operating on very large samples and no need a priori assumptions about the distribution, self organizing map method has attracted a wide range of application especially in market segmentation (Kohonen, 1989).

Because of life-style characteristics provide a rich view of the market and depth understanding to the consumer behaviour, life-style information has become a popular tool in marketing application since the 1960's. The main objective of life-style studies classifying consumers into segments with specific and identifiable patterns. On the bases of measured activities, interest and opinions (AIO) life-style research constructing consumers' psychographic profiles (Plummer, 1971). Values, Attitudes and Lifestyles (VALS) is another concept of classifying American population into segments based on attitudes, lifestyles and decision making styles, which developed by SRI consulting in the late 1970's and then revised in 1989 to focus more explicitly on explaining consumer purchase behaviour (Linda, 1999). Yankelovich's segmentation methodology, Monitor Mindbase (1971), is based on consumers' actual purchasing behaviour and their likely behaviour. The Yankelovich Monitor study has gathered and trended the values, motivations and attitudes that redefine segments as market conditions change.

The rest of the thesis is organized as follows: Section 2 reviews the market segmentation and cluster analysis approaches. Section 3 describes the methodology of survey and the characteristics of data that we have used. In section 4 experimental results of two stage clustering, k-means clustering and self organizing map approaches are compared with the base on cluster outputs. The characteristics of segments are searched according to fit approach. The conclusion of our study is presented in section 5 with a summary of our findings.

2. LITERATURE REVIEW

This section explores the literature review of hierarchical, non-hierarchical and self organizing map clustering methods. The scope of this literature review is expanded to include market segmentation regardless of the using specific clustering methods.

2.1. CLUSTER ANALYSIS

Classification in the widest sense is one of the oldest scientific pursuits undertaken by humanity. In the general terms of classification can be defined as the process of giving names to a collection of different types of events, objects and people which are thought to be similar to each other in some respect. Classification has played an important role in the development of many areas of science including psychology, artificial intelligence, biology and zoology, chemical, astronomy, pattern recognition and marketing research area for segmentation (Jain, Murty, & Flynn, 1999). The property of clustering makes it become a popular tool for market segmentation. Segmentation theory proposes designed to identify groups of entities that share certain common characteristics such as needs and purchasing behaviours. Segmentation has consequently been regarded as one of the most critical element in marketing area for the achievement of successful marketing activities and customer relationship management by the companies (Berson, Smith, & Thearling, 2000).

Clustering is an unsupervised process of dividing patterns into groups such that each group is homogeneous with respect to predefined attributes. Cluster analysis is widely applied to many areas, such as customer and market segmentation, pattern recognition and image processing, bioinformatics or biomedicine application. These inductive techniques have been employed as a classification tool of market segmentation (Jain, Murty, & Flynn, 1999). Applications are distinct from the use of cluster analysis for classification and represent an alternative to representing similarity of data.

Classification is concerned with the identification of numerical taxonomies, where on the contrary structural representation is concerned with the development of a faithful representation of relationships. Both uses of cluster analysis are legitimate, but the objectives of these applications different from them. The best clustering algorithm regarding of these objectives is not necessarily the best for the other objective (Wedel & Kamakura, 2000).

Cluster analysis is a generic term for numerous methodologies which attempt to find similarity measures into homogeneous groups with respect to predefined attributes. Cluster analyses have been used in a variety of fields including both natural and social sciences. Some of these fields are data mining, identification of different consumer's profiles, building up the stratified sampling and identification of the variables that have an inevitable important to describe a phenomenon (Mingoti & Lima, 2006). Cluster analysis, which takes a sample of variables and groups them such that the statistical variance among elements grouped together, is minimized while between-group variance is maximized. A critical issue to perform successful cluster analysis is the selection of the variables. Segmentation variables can be broadly classified into general variables and product specific variables. The general variables include the customer demographics and lifestyles. The product specific variables involve customer purchasing behaviours and intentions (Wedel & Kamakura, 1997).

Cluster analysis consists of grouping similar objects into distinct, mutually exclusive subsets known as clusters. Cluster analysis can be defined also as a statistical technique that sorts observations into similar sets or groups. Cluster analysis' sorting ability is powerful enough that it will provide clusters even if no meaningful groups are embedded in a sample. Thus, cluster analysis has the potential not only to offer inaccurate depictions of the groupings in a sample but also to impose groupings where none exist (Barney & Hoskisson, 1990).

The purpose of cluster analysis is to response to the general question facing researchers in many areas of inquiry is how to organize observed data into meaningful structure. Cluster analysis amalgamates data objects into constituent groups (natural groupings) such that objects belonging to the same cluster are similar, while those belonging to different ones are dissimilar. Natural groupings are also named as clusters having properties such as internal cohesion and external isolation. Until the 80's the discussion concentrated mainly on techniques that encompass a number of different classification algorithms. At the end of the 80's the whole process of clustering – starting with the selection of cases and variables then ending with the validation of cluster became dominant (Everitt, 1993).

Clustering methods can be divided into classification within each of the classification exists a wealth of subtypes and different algorithms for finding the clusters. The first classification of clustering procedures is between sequential and simultaneous algorithms. The most clustering algorithm is sequential method which searches for the equivalence relation repeated at all levels of similarity in the association matrix. The simultaneous algorithms, searches the solution which is obtained in a single step. The second classification is between agglomerative and divisive. Agglomerative procedures the objects are considered as being separate from one another. They are successively grouped into larger and larger clusters until a single cluster is obtained. On the other hand, if a single group containing all objects, divisive algorithms subdivides the group into sub-clusters, and so on until the discontinuous partition is reached. The most known common classification is hierarchical and non-hierarchical methods. In hierarchical methods, the members of inferior ranking clusters become members of larger, higher-ranking clusters. Non-hierarchical methods produce a single partition which optimizes within group homogeneity, instead of a hierarchical series of partitions optimizing the hierarchical attribution of objects to clusters (Sneath & Skol, 1973). In a such kind of various clustering algorithms affords a basis for establishing some general guidelines for the appropriate use of cluster analysis and The selection of the clustering algorithm appears to be critical to the successful use of cluster analysis.

2.1.1. Hierarchical Cluster Analysis

Hierarchical clustering algorithms are heuristic algorithms which begin by placing each unit in an individual group and proceed by combining these groups in hierarchical fashion until all units are grouped together and instead of producing a single clustering they produce a hierarchy of clustering (Theodoridis & Koutroumbas, 2006). Hierarchical approach help users on the way of identifying not only distinct clusters but also the subgroups they may contain. Additionally, agglomerative hierarchical clustering methods are considered to be the most popular cluster analysis technique. Agglomerative algorithms begin by grouping each unit individually and proceed by grouping the pair of units which has the greatest similarity, in another word the smallest distance (Gong & Richman, 1995).

The increasing number of cluster analysis methods available has led the selection of appropriate clustering algorithms is critical to the effective use of cluster analysis. Hierarchical algorithms progress through a series of steps that structure by either adding individual elements to (agglomerative) or deleting them from (divisive) clusters. The most popular agglomerative algorithms are single linkage, complete linkage, average linkage, centroid method, and Ward's method (Hair, 2000). The differences among them lie in the mathematical procedures used to calculate the distance between clusters. Each has different systematic tendencies or biases in the way it groups observations.

2.1.1.1. Single Linkage

Single linkage clustering method’s logic is natural, so that it is defined as the principles of clustering and a simple to understand method. In single linkage clustering, distance between two clusters is defined as the minimum distance from any member of one cluster to any member of the other cluster. The method of forming clusters is observation joined to as a cluster if it has a certain level of similarity with at least one of the members of that cluster. Connections between clusters are based on links between single entities. In this method, the distance between two clusters is determined by the distance of the two closest cases (neighbours) in the different clusters (Small, 1998). The equation of distance between clusters is,

$$D\ (C_I, C_{II}) = \min X_I \in C_I,\ X_{II} \in C_{II}\ d(X_I, X_{II}) \tag{2.1}$$

The measure of inter clusters is illustrated in following figure.

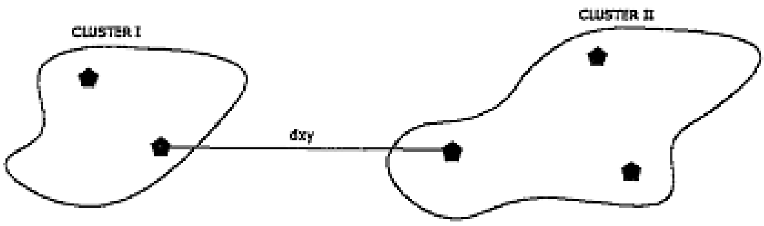

Figure 2.1 : Single linkage distance

Source : Mulvey & Gingold, 2009

The algorithm for single linkage clustering is sequential, agglomerative, and hierarchical. Its starting point is any association matrix including similarity or distance among the objects or descriptors to be clustered. The method involves in two steps. First, the association matrix is rewritten in order of decreasing similarities or increasing distances, heading the list with the two most similar objects of the association matrix, followed by the second most similar pair, and proceeding until all the measures comprised in the association matrix have been listed. Second, the clusters are formed hierarchically, starting with the two most similar objects, and then letting the objects clump into groups, and the groups aggregate to one another, as the similarity criterion is relaxed (Demirel, 2004).

$$
\mathbf{D_1} = \begin{array}{c|ccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & & & & \\
2 & 3 & 0 & & & \\
3 & 4 & 2 & 0 & & \\
4 & 8 & 7 & 6 & 0 & \\
5 & 4 & 10 & 9 & 7 & 0
\end{array}
$$

The smallest entry is that for individuals 2 and 3, consequently these are joined to form a two-member cluster. Distances between this cluster and the other three individuals are below;

$$d_{(23)1} = \min [d_{12}, d_{13}] = d_{12} = 3 \qquad (2.2)$$

$$d_{(23)4} = \min [d_{24}, d_{34}] = d_{34} = 6 \qquad (2.3)$$

$$d_{(23)5} = \min [d_{25}, d_{35}] = d_{35} = 9 \qquad (2.4)$$

A new matrix is now constructed whose entries are inter-individual distances and cluster individual values.

$$
\mathbf{D_2} = \begin{array}{c|cccc}
 & 23 & 1 & 4 & 5 \\
\hline
23 & 0 & & & \\
1 & 3 & 0 & & \\
4 & 6 & 8 & 0 & \\
5 & 9 & 4 & 7 & 0
\end{array}
$$

The smallest entry is in D_2 is that for individual (23) and 1, so these now form a three member cluster, and a new distance matrix is found,

$$d_{(123)4} = \min [d_{231}, d_{234}] = d_{231} = 3 \quad (2.5)$$

$$d_{(123)5} = \min [d_{231}, d_{235}] = d_{231} = 3 \quad (2.6)$$

$$\mathbf{D_3} = \begin{array}{c c} & \begin{array}{ccc} 123 & 4 & 5 \end{array} \\ \begin{array}{c} 123 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & & \\ 3 & 0 & \\ 3 & 7 & 0 \end{pmatrix} \end{array}$$

The smallest entry is in D3 is that for individual (123) and 4 or 5. The entry 4 is added to the cluster containing individuals 1,2,3. Finally the groups containing individuals 1, 2, 3, 4 and 5 are combined into one single. The partitions produced at each level are;

Level	**Groups**
5	[1], [2], [3], [4], [5]
4	[1], [2 3], [4], [5]
3	[1 2 3], [4], [5]
2	[1 2 3 4], [5]
1	[1 2 3 4 5]

As a consequence of proceeding, results of single linkage clustering are sensitive to noise in the data, because noise changes the similarity values and may thus easily modify the order in which objects cluster.

2.1.1.2. Complete Linkage

A variation on the simple linkage method is often known as complete linkage or furthest neighbour or maximum method. The distance between two clusters is defined as the greatest distance between objects in them. In complete linkage clustering methods, an object joins a cluster only when it is linked to all the objects already members of that cluster. Two clusters can fuse only when all objects of the first are linked to all objects of the second, and vice versa. Therefore in the complete linkage strategy, it becomes more and more difficult for new objects to join to it because the new objects should bear links with all the objects already in the cluster before being incorporated (Gong & Richman, 1995). In the agglomerative complete linkage method, the most similar clusters are combined; on the other hand in the divisive methods the most dissimilar clusters are splited. The equation of distance between clusters is,

$$D\ (C_I, C_{II}) = \max X_I \in C_I,\ X_{II} \in C_{II}\ d(X_I, X_{II}) \qquad (2.7)$$

The measure of inter clusters is illustrated in following figure.

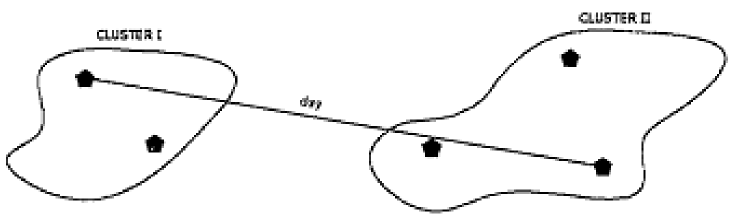

Figure 2.2 : Complete linkage distance

Source : Mulvey & Gingold, 2009

If we compare complete linkage with the single linkage, the former is opposite approach of the latter in the sense that the distance definition. The complete linkage clustering model has one drawback when compared to single linkage. In all cases where two incompatible candidates present themselves at the same time to be included in a cluster, algorithms use a pre established arbitrary rule to choose one and exclude the other. The recommendation of this problem is chosen the fusion leading to the largest cluster. If equality persists, the recommendation is chosen the fusion that most reduces the number of clusters or as a last criterion, choose the fusion that maximizes the average similarity within the cluster. This problem does not exist in single linkage. Beside of these differences, the results depend very much on which two cases are taken as starting point in the process. Briefly, complete linkage tends to produce small, compact clusters in which the observations or cases are very similar to each other, while single linkage tends to produce long and stringy clusters (Gong & Richman, 1995).

2.1.1.3. Average Linkage

There are two kinds of average clustering that unweighted and weighted arithmetic average clustering. This technique treats the distance between two clusters as the average distance between all pairs of objects in the two different clusters. This method using arithmetic averages (Sneath & Sokal, 1973). The distance between cluster a, merged by cluster i and j, and another cluster b is determined by the following formula;

$$d_{ab} = \frac{N_i}{N_i + N_j} d_{ia} + \frac{N_i}{N_i + N_j} d_{jb} \qquad (2.8)$$

The measure of inter clusters is illustrated in following figure

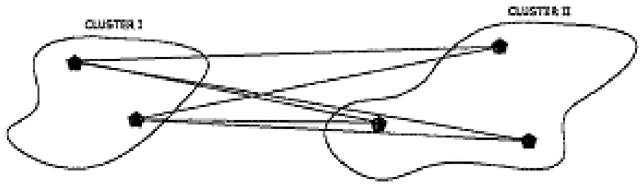

Figure 2.3 : Average linkage distance

Source : Mulvey & Gingold, 2009

The method computes an arithmetic average of the similarities or distances between candidates object all members of the two clusters. The highest similarity or smallest distance identifies the next cluster to be formed. All objects receive equal weights in the computation. The unweighted arithmetic average method assumes that the objects in each group form a representative sample of the corresponding larger groups of objects in the reference population under study. For that reason, unweighted arithmetic average clustering should only be used in connection with simple random or systematic sampling designs if the results are to be extrapolated to some larger reference population. On the other hand, weighted arithmetic average clustering occurs that groups of objects unweighted arithmetic average or representing different regions of a territory. Weighted arithmetic average, method consists in giving equal weights, when computing fusion similarities. Weighted arithmetic average clustering increase the separation of the two main clusters, compared to unweighted arithmetic average. This gives sharper contrast to the classification (Demirel, 2004).

2.1.1.4. Ward's Method

All the linkage techniques such as single linkage, complete linkage and average linkage are based on similar principle, but the rules of produce cluster differ from one linkage technique to another. Ward's method is another approach which is more complex than other linkage algorithms. The main objective of this method is to join cases into cluster such that the variance within a cluster is minimized. Ward's minimum variance method is forming clusters, with minimizes an objective function which is the same "squared error" criterion as that used in multivariate analysis of variance. Distances are computed as squared Euclidean distances in Ward's method. To measure the distance between two objects i and j, the Euclidean distance function, d_{ij} is used (Gong & Richman, 1995), which is determined by the following formula;

$$d_{ij} = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]^{1/2} \qquad (2.9)$$

At each clustering step, two objects or clusters h and i are merged into a new cluster hi, as in previous sections. Since changes occurred only in the groups h, i, and hi, the change in the overall sum of squared errors, ΔE_{hi}, may be computed from the changes that occurred in these groups only with the following formula;

$$\Delta E_{hi}{}^2 = e_{hi}{}^2 - e_h{}^2 - e_i{}^2 \qquad (2.10)$$

Ward's method tends to produce clusters with roughly the same number of observations and the solutions it provides tend to be heavily distorted by outliers. At each step in Ward’s method procedure, union of every possible pair of clusters is considered and the two clusters whose fusion results in the minimum increase in information loss in term of an error sum of squares criterion are combined (Everitt, 1993). Using squared Euclidean distance as dissimilarity measurement, together with Ward’s method for linkage, produced the most distinctive groups because it uses an analysis of variance approach to evaluate the distances between clusters. Ward’s method has been limited to

Euclidean metric space while getting the most accurate solutions and that the large majority of investigators applied this metric for their study (Güler, 2002).

2.1.2. Partitional (Non-Hierarchical) Cluster Analysis

Non-hierarchical algorithms, which are also referred as iterative methods, are based on partitioning a data set of observations into a pre specified number of clusters. Observations are then reassigned to clusters until some decision rule terminates the process. The decision rule is used for terminating clustering, and the frequency with which cluster centroids are updated during the reassignment process. The objects inside a cluster show a certain degree of closeness or similarity (Gerstengarbe & Werner, 1999).

Non-hierarchical methods have two potential advantages over hierarchical methods. The first advantage is non-hierarchical methods are less impacted by outlier elements. Although outliers can initially distort clusters, this is often corrected in subsequent passes as the observations switch cluster membership. Second, the final solution optimizes within cluster homogeneity and between cluster heterogeneity. Obtaining this improvement requires that the number of clusters be specified a priori (Hair, 2000).

Clustering methods can be considered as either hard or fuzzy depending on whether a pattern data belongs exclusively to a single cluster or to several clusters with different degrees. In hard clustering, a membership value of zero or one is assigned to each pattern data whereas in fuzzy clustering, a value between zero and one is assigned to each pattern by a membership function. Two types of non-hierarchical methods are generally used; K-means and Fuzzy C-means. Clustering algorithms such as K-Means as known hard clustering and C-Means, as known fuzzy clustering are based on the sum of intra cluster distances criterion. Additionally, the two stage clustering algorithm is used when some specifications about the clustering procedures are absent and not available to priori descriptions of expected clusters (Kaya, 2005).

2.1.2.1. K-means

K-Means clustering is one of the most widely used market segmentation technique among the other clustering techniques to classify the data (Gehrt & Shim, 1998). The algorithms aim at finding a K-partition of the sample, with within cluster sum of squares which cannot be reduced by moving points from one cluster to the other. The method defines a fixed number of clusters and adjusts the cluster centers until further refinement can no longer improve the model. There are two versions of K-Means clustering, a non-adaptive version and an adaptive version. The most commonly used K-Means clustering is the adaptive K-means clustering based on the Euclidean distance (Everitt, 2001).

K-Means clustering algorithm search for the optimum center locations and it is assumed that the initial centers are provided. The search for the final clusters or centers starts from these initial centers (Mashor, 1998). The centers should be selected to minimize the total distance between the data and the centers. A simple and widely used square error cost function is used to measure the distance, which is defined as:

$$E = \sum_{j=1}^{n_c} \sum_{i=1}^{N} (\| x_i - c_j \|)^2 \qquad (2.11)$$

where *N*, and n_c are the number of data and the number of centers respectively; *xi* is the data sample belonging to center *cj*. K-Means clustering tries to minimize function by searching for the center on line as the data are presented and with the Euclidean distances between the data sample and all the centers are calculated and the nearest center is updated.

Even if K-means clustering perform when non-random starting point is specified, the procedure appears to be more forceful than any of the hierarchical methods with respect to the presence of outliers, error perturbations of the distance measures, and the choice of a distance metric. If we compare the K-means algorithm with hierarchical clustering, K-Means may be computationally faster than hierarchical clustering respect to with a large number of variables. Additionally, K-Means may produce tighter clusters than hierarchical clustering. However, difficulty in comparing quality of the clusters produced and difficult to predict what K should be are the main disadvantages of K-Means clustering.

2.1.2.2. Fuzzy C-means

Fuzzy C-Means is generally used in pattern recognition. The aim of Fuzzy C-means is to find cluster centers that minimize the dissimilarity function (Albayrak & Amasyalı, 2003). Fuzzy C-means clustering is separated from K-means clustering with the using of fuzzy partitioning method which means a data point can belong to all groups with different degrees (Berks et al., 2000). In general, fuzzy clustering methods can be considered to be superior to that can represent the relationship between the input pattern data and clusters more naturally.

Fuzzy C-means algorithm determines the following steps (Jang, Sun, & Mizutani, 1997).

Steps	Formula
Step1. Randomly initialize the membership matrix (U).	$\sum_{i=1}^{c} u_{ij} = 1 \quad , \quad \forall_j = 1, \ldots\ldots, \quad n$
Step2. Calculating centers (c_i).	$c_i = \frac{\sum_{j=!}^{n} u_{ij}{}^m x_j}{\sum_{j=!}^{n} u_{ij}{}^m}$
Step3. Computing dissimilarity between centers and data points. (We have to stop if its improvement over previous iteration is below a threshold)	$J(U, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^m d_{ij}$
Step4. Computing a new "U". (Then we have to go Step 2)	$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{kj}})^{2/(m-1)}}$

Figure 2.4 : Algorithm steps of fuzzy c-mean

Source : Jang, Sun, & Mizutani, 1997

Because of cluster centers are initializing using *U,* Fuzzy C-means doesn't always good solution. Therefore two approaches are proposed that using an algorithm to determine all of the centers or run algorithm several times each starting with different initial centers.

2.1.2.3. Two Stage Clustering

Although, partitioning methods require some more specifications comparing with hierarchical methods, studies of clustering algorithms suggest that iterative partitioning methods are preferable to the hierarchical methods. Partitioning methods can be performed only when a non random starting point can be specified or the number of clusters can be desired. Under the absence of this information, the user can choose two-stage clustering procedure, in order to determine specifications and demonstrated superior performance of cluster method. Research has shown that in this situation the best solutions may be those obtained by using hierarchical and non-hierarchical methods with two-stage clustering. Additionally, the two-step cluster is appropriate for large datasets or datasets that have a mixture of continuous and categorical variables (Huang, 1998).

Empirical studies of the clustering algorithm performance suggest that the integration of hierarchical and non-hierarchical methods can provide a better solution than hierarchical methods. In addition, iterative partitioning methods require prior specification of the number of clusters desired, while hierarchical methods do not need such specification. Thus, the researcher is confronted with determining both an initial starting point and the number of clusters in order to use the non-hierarchical methods. Therefore, first, the hierarchical methods, such as Ward's minimum variance method, can be applied to obtain a rough solution. The main reason for such integration is that Ward's minimum variance method can provide the number of clusters and also provide the starting point which the *K*-means method requires. Thus the number of clusters need not be assumed by the researcher and starting point is not randomly selected. Then the non-hierarchical methods, like the *K*-means method, can use the information to obtain the final clustering results (Kuo, Ho, & Hu, 2002).

In the first step one of the hierarchical methods such as average linkage or Ward's minimum variance method, may be used to obtain a first approximation of a solution. By examining the results of this preliminary analysis, one can determine both a candidate number of clusters and a starting point for the iterative partitioning analysis. Additionally, this analysis can be used for examining the order of clustering of various observations and the distances between individual observations and clusters to provide an opportunity for the identification of outliers. The remaining cases may then be submitted to an iterative partitioning analysis for refinement of the clusters (Hair, 2000). Two step clustering has the advantage of automatically estimating the optimal number of clusters for the training data and it can handle mixed field types and large data sets efficiently. The Figure 2.4 is a schematic representation of the procedure by Punj and Stewart in 1983.

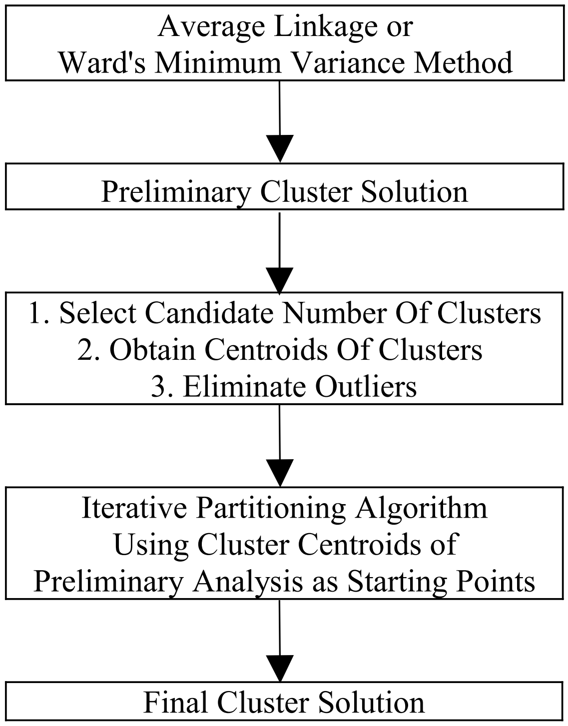

Figure 2.5 : Two-stage clustering procedure

Source : Punj & Stewart, 1983

2.1.3. Self Organizing Map (Kohonen Network)

Although the first model in artificial neural networks dated from the 1940s, which was explored by Hebb, it started been more used in the 1980's (Mingoti & Lima, 2006). The bases artificial neural network consists of a set of simple processing units, neurons that are connected to each other to form a network topology. In clustering problems, the artificial neural networks cluster observations in two main stages. The first step is learning stage which is used to train the network for a specific data set. The second step is observations are classified, which is called a recall stage. There are two ways to train a network, supervised and unsupervised. In supervised learning the network is presented with examples of known input-output data pairs and the network is then tested to see whether it is able to produce correct output, when only input is presented to it. In unsupervised learning, the output data is not available and usually not even known beforehand. Similar samples form clusters that constitute the output of the network. The user is responsible for giving an interpretation to each cluster (Back, Sere, & Vanharanta, 1998).

In another word the artificial neural networks are working into layers. Basically, the input layer contains the nodes through which data are input. The output layer generated the output interpreted by the user. Between these two layers there can be more layers called hidden layers (Mingoti & Lima, 2006). The output of each layer is an input of the next layer until the signal reaches the output layers as shown in following figure.

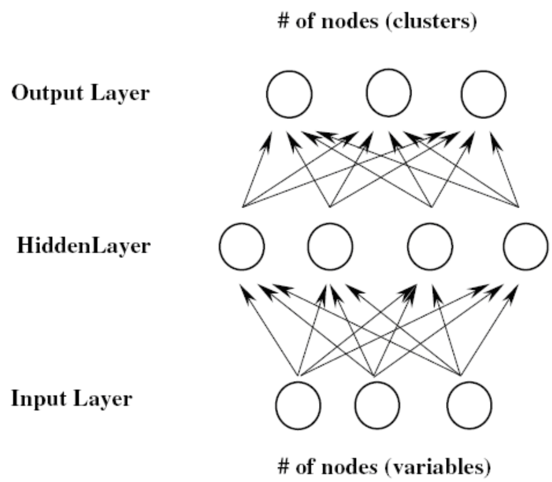

Figure 2.6 : Neural network for clustering

Source : Mingoti & Lima, 2006

Self organizing map, which proposed by Kohonen, is one of the most common neural network model. It was originally designed for solving problems that involve tasks such as clustering, visualization, abstraction and the theory is motivated by the observation of the operation of the brain. Various human sensory impressions are neurologically mapped into the brain such that spatial or other relations among stimuli correspond to spatial relations among the neurons organized into a two-dimensional map (Kohonen, 1989). SOM belong to a general class of neural network methods, which are non linear regression techniques that can be applied to find relationship between inputs and outputs or to organise data so as to disclose previously unknown patterns or structure. Kohonen's Self organizing map approach has been successfully applied because of its good performance as a classification tool to various problem domains including speech recognition, image or character recognition, robot control, medical diagnosis and market segmentation (Kiang, Hu, & Fisher, 2006).

SOM network is a special type of neural network that can learn from complex, multi-dimensional data and transform them into visually decipherable clusters. In another word SOM is a dimensionality reduction visualization technique such as one or two-dimensional map, to generate compact but distorted map visualizations for an expertise data set. Dimensional map provides an easy to use graphical user interface to help the decision maker visualize the similarities between consumer preferences. These dimensional maps not only help the companies to see fully visualized clusters of market but also reveal mutual non-linear correlations between different customers' characteristic variables. Briefly, the main function of SOM networks are discovering of structure in large high-dimensional data sets and mapping the input data from an n-dimensional space to a one or two-dimensional plot while maintaining the original topological relations (Kiang & Fisher, 2007).

Self Organizing Maps have some advantages which make them appropriate in much area of studies. The first advantage is they are able to outperform the traditional data reduction and clustering techniques, in both speed and quality of solution (Smith, 1999). Second, they have the capacity to operate on very large samples and need no a priori assumptions about the distribution of the sample. The other advantage is using SOM helps overcome structuring task problems associated with finding the appropriate underlying distribution and the functional form of the underlying data. The fourth advantage is ability of SOM is to reduce the input space to a one or two dimensional output map and also projecting them non-linearly in a two-dimensional map while maintaining the original topological relations (Vesanto, 1999). Furthermore, it detects clusters existing in the original data while avoiding creating artificial ones, thus providing a true representation of the original data's characteristics. Another advantage of Self Organizing Map is that the user can determine the number of clusters needed by examining the merge process visually on the map. Self Organizing Maps are useful in visually examining the relationship between input data and identifying important patterns and clusters.

SOM belongs to unsupervised neural networks and the basic idea behind them is competitive learning that, clusters objects having multi-dimension attributes into a lower-dimension space, in which the distance between every pair of objects captures the multi-attribute similarity between them. Although more common approach to neural networks required supervised training of the network, the SOM network performs unsupervised training which provides a visual representation of the relationships that exist in the original data, while avoiding creating artificial clusters (Deichmann, 2003). The SOM network model consists of two layers, an input layer and an output layer (Kohonen layer). An input layer represents the object features and output layer in the shape of a two-dimensional grid that determines the positions of the objects. In another word, the input layer neurons present an input pattern to each of the output neurons and it is fully connected to a two-dimensional Kohonen layer. The neurons in the output layer are usually arranged in a grid, and are influenced by their neighbours in this grid. The output layer acts as a distribution layer to summarize general feature patterns in the collection of objects and the output from SOM networks is a dimensional map (Kiang & Fisher, 2007).

Kohonen network is trained using unsupervised learning. During the training process the network has no knowledge of the desired outputs and it's accomplished by iterative application. The training process is characterized by a competition between the output neurons. The input patterns are presented to the network one by one in the input space with a random order. As the training process proceeds, the nodes adjust their weight values according to the topological relations in the input data. Each node on the map may represent zero to many input data. The nodes that are closely located on the representational grid should have similar cluster center. Each node in the input layer corresponds to one of the features of an object and in the mapping layer is connected to all input layer nodes with certain link weights (Kiang & Fisher, 2007). Thus, a mapping layer node can be also viewed as a feature vector with link weights as the feature values and each node on the output map as one group and cluster the input data accordingly. The SOM exhibits the important property of topology preservation. In other words, if two input vectors are close in the input space, the corresponding closest neurons will also be close in the neural network. The network topology can be described by the number of output neurons present in the network and by describing which neurons in the output array are mutual neighbours (Huang et al., 2006)

Neurons on the output layer are arranged in either a rectangular or a hexagonal grid as the following figure.

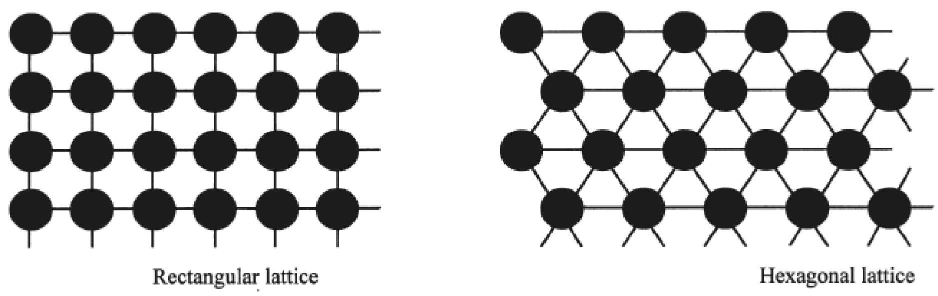

Figure 2.7 : Network topologies

Source : Back, Sere, & Vanaranta, 1998

In a rectangular grid each neuron is connected to four neighbours, except for the ones at the edge of the grid. The output neurons are arranged in a hexagonal lattice structure which means that every neuron is connected to exactly six neighbours, except for the ones at the edge of the grid. The basic SOM consists of a set of neurons that there are neighbourhood relations among them. The neighbourhood structure of the output layer will cause neighbouring neurons in the output layer to have similar weight vectors. Each neuron has an associated weight vector of the same dimension as the input space. The output neurons compete for each and every pattern. The output neuron with a weight vector that is closest to the input vector is called the winner. In another word the node with the minimum distance becomes the “winning” node and the link weights between output layer node and the input nodes are updated according to its distance to this “winning” node. For expressing the distance between two vectors, we use the Euclidean distance between the two vectors (Back, Sere, & Vanaranta, 1998).

The weight vector of the winner adjusts its weights to be closer to the value of the input pattern. The size of adjustment in the weight vectors of the neighbouring neurons is dependent on the distance of that neuron from the winner in the output array. There are widely used three adjustment criteria. The first is learning rate which influences the size of the weight vector adjustments after each training step, whereas the neighbourhood width parameter determines to what extent the surrounding neurons, the neighbours, are affected by the winner. The second criterion is training length, which measures the number of iterations through the training data. Another criterion is the average quantization error, which is an average of the Euclidean distances of each input vector and its best matching reference vector in the SOM. When the adjustment proceeds the clusters are formed by identifying neurons on the output layer that are close to each other using the weight vectors as a starting point which defined as a U-matrix (Kohonen, 1997). The matrix can be used to visualize the distances between neighbouring neurons. Thus, a set of neurons form a cluster, if they are sufficiently close to each other.

Before examining the steps of self organizing map, we should focus on the methodology which has important principles. The first is choosing the data material which is often advisable to pre process the input data so that the learning task of the network becomes easier (Kohonen, 1997). Then choose the network topology, learning rate, and the neighbourhood width. The third principle is constructing the network which takes place by showing the input data to the network iteratively using the same input vector many times. Final principle is choosing the best map for further analysis; identify the clusters and interpreting the clusters. The process of self organizing map has following 4 steps;

(1) Initialize the weights as random small numbers.

(2) Put an input sample, X_i, into the SOM network, and the distances between weight vectors, $W_j = (w_{j1}, w_{j2}, \ldots, w_{jm})$, and the input sample, X_i, are calculated. Then, select the neuron whose distance with X_i is the shortest, as in equation (2.16). The selected neuron would be called the "winner"

$$\min d(j) = \sum_i (w_{ji} - x_i)^2 \qquad (2.12)$$

(3) Update the weights of the winner as well as its neighbours by the following equation when α is assumed to be the learning rate

$$W_j^{NEW} = W_j + \alpha \| X_i - W_j \| \qquad (2.13)$$

(4) Iterate Steps (2) and (3) until the weights have stabilized and stop criterion is satisfied

If we compare cluster analysis with the self organizing map method, cluster analysis is a technique for grouping subjects into clusters of similar elements and tries to identify similar elements by their attributes. The technique is forming clusters that are homogeneous but are different from other groups. On the other hand self organizing map networks combine competitive learning with dimensionality reduction by smoothing the clusters with respect to an a priori grid and provide a powerful tool for data visualization (Kiang & Fisher, 2007).

Self organizing maps have an important limitation which has no mechanism to determine the number of clusters, initial weights and stopping conditions where as the other neural network based algorithms hasn't. If we compare the number of clusters between cluster analysis and self organizing map, in clustering algorithm the number of clusters should be chosen according to the number of clusters there are in the data, but in the self organizing map algorithm the number of neurons and related weight vectors can be chosen to be much larger, irrespective of the number of clusters. Another difference is a self organizing map is more sophisticated than clustering methods in terms of presentation the relationship between the clusters in a two-dimensional space (Urso & Giovanni, 2007).

2.2. MARKET SEGMENTATION

In 20th century the concept of market segmentation entered in the literature and after this time it becomes nearly impossible to see the situation where mass marketing approach is feasible (Wedel & Kamakura, 2000). The reason, which has obviously long been accepted, is that consumer's shows differentiation between each other among their individual choices. Here appears a new terms heterogeneity which could be a core point of the segmentation (Hunt & Arnett, 2004). Customers are heterogeneous, which means that their purchasing behaviour over time varies, their willingness to pay varies from customer to customer, and they are attracted by different benefits offered by the same type of products. Market segmentation involves the identification of segmentation variables followed by segmentation of the market. Segmentation is a grouping task for, which a large variety of methods are available, leads to market targeting and evaluation of the attractiveness of the obtained segments and a selection of the target segments. For achieving target segments, positioning concepts are developed, selected and communicated. The segments are distinguishing from each other because of some factors like product, distribution, pricing, and communication strategy (Wedel & Kamakura, 2000).

The concept of market segmentation arises from viewing a heterogeneous market composed of a couple of smaller homogeneous markets. Market segmentation desires that groups of customers with similar needs and purchasing patterns are likely to demonstrate that specific customer groups (Tsai & Chiu, 2004). As our world getting industrialised, the importance of the market segmentation is dramatically increasing because unless considering customer needs and recognising the heterogeneity of those needs, goods and services can no longer be produced and sold. In other words able to balance diverse customer needs with the capabilities and resources of competing organizations in the marketplace getting the core point of the company's market strategies. Most markets the breadth of customer requirements is too extreme to allow single organizations to satisfy all customer products and service needs all of the time. Therefore, in order to satisfy various customer requirements within company's capacity, companies need to split consumer market into several segmentations and find out appropriate marketing strategies for them. As a result companies achieve a full understanding of a market; the ability to predict behaviour accurately; and an increased likelihood of detecting and exploiting new market opportunities (Kotler, 1997).

Segmentation is enable understanding market through collecting and then analysing several variables using sophisticated multivariate techniques. The companies use these techniques to divide subgroups because they would like to reach this segment by developing different offerings with aiming increasing their profit. Companies could be successful if segmentation allows them to determine which actions would attract or retain customers, or decide which additional products or services they could introduce profitably. With proper market segmentation, companies can deploy the right resource to target customer groups and cultivate the closer relationships with their customers more efficiently and effectively (Huang, Tzeng, & Ong, 2007). Based on this proper market segmentation perspective, companies need to understand their customers, differentiate between various customer groups and identify the most or the least valuable customers and increase customer loyalty through providing customized products and services. In another word segmentation must help identify different response groups which have uniform and stable responses to a particular set of marketing variables by analyzing customer information and the result of this marketers aim to change, reinforce or initiate behaviour patterns (Ha, 2007).

2.2.1. Bases for Segmentation

In the marketing strategy there are three phases; in sequence segmenting the market, selecting one or more segments to target and positioning the product or service. Market segmentation, which is the first phase of marketing strategy, can be defined as the process of dividing a market into distinct subsets of consumers with common needs or characteristics and selecting one or more segments to target. Segmentation studies have two major components: the information used as input, called the 'bases' of segmentation, and the methods used to identify segments/subpopulations based on the input data (Wedel & Kamakura, 2000). The first step in developing a segmentation strategy is to select the most appropriate bases on which to segment the market. Four major categories of consumer characteristics provide the most popular bases for market segmentation; which are geographical, demographical, socio cultural and psychological-psychographic. In general, demographical and geographical segmentation help to locate a target market, whereas psychological and socio cultural characteristics help to describe how its members think and how they feel. Therefore in some situation, marketers can segment the markets by combining several bases rather than relying on a single segmentation base (Schiffman & Kanuk, 2000).

In geographic segmentation, the market is divided by location using with region, city size, climate or density of area variables. The theory behind this strategy is that people who live in the same are share some similar needs and wants and these needs and wants differ from those of people living in other areas. Geographic segmentation is a useful strategy that can be easily reached through the local media, including newspapers, TV and radio, regional editions of magazines. Demographic segmentation is most wide used as the basis for segmentation research. Demography explains the main characteristics of a population with using the measurable statistics such as age, sex, marital status, income, occupation and education variables. Demographic segmentation is often the most accessible and cost effective way to segmentation. Additionally demographical variables are easier to measure than the other segmentation variables (Baker & Hart, 2008).

Socio cultural variables provide further bases for market segmentation. This approach divides the segments on the basis of the stage in the family life cycle, social class, core cultural values, sub cultural membership and cross-cultural affiliation. Family life cycle based on the premise that many families pass through similar phases in their formation, growth and dissolution. Social class is measured by a weighted index of several demographic variables in which individuals in the same class generally have the same degree of status Consumer segmentation strategies are often based on specific psychological variables such as motivations, personality, perceptions, learning and attitudes. On the other hand, psychographic segmentation can be thought of as a composite of consumers' measured activities, interests and opinions with the larger number of statements (Schiffman & Kanuk, 2000).

2.2.2. Criteria for segmentation

Segmentation is a very critical process that companies have to give attention to it therefore Morrison (1996) listed eight criteria for effective segmentation to help companies. According to these effective markets segmentation has the following characteristics (Morrison, 1996).

The first criterion is achieving homogeneity inside of the group which defined as segmentability. People within a segment should be similar to each other but at the same time they should be as different from each other as possible. In another word, segments should satisfy homogeneity within and heterogeneity between them. This is the base on the segmentation (Kotler, 1997). Second criterion is segments should be measurable in other words segments should be identified with a reasonable degree of accuracy. This criterion is used mostly in demographic and geographic segmentation which use commonly age, sex, income, education, occupation, gender, region, city size, density of area, climate variables (Raaij & Verhallen, 1994). The next criterion is segments should be substantial, it means that segments should be large enough in size to warrant separate attention. It is important because as large is the segment as better as the segmentation. If the targeted segments represent a large enough portion of the market the substantiality criterion is satisfied which is closely connected to the firms' goals and costs. Another criterion is segments should be accessible; if an organization needs to be able to easily reach or access the identified segments. That means for example that there are target group specifically advertising media, as magazines or websites the target audience likes to use because they usually have to contact with their segment to ale to accessible (Baker & Hart, 2008).

Another important criterion is each segment must need different marketing approaches (Dibb, 1998). This suggests that the segments must differ on those characteristics which will be most relevant to the organization's services or products. Addition to this until competitors copy or segment your segmentation, you have a competitive edge, even if you serve the segment with a standard product or service. The next criterion segments must have meaningful relation with o the products or services offered by the organization. Another criterion is identified segments need to be compatible with existing markets. The last criterion is there must be some stability in the segments. The identified segments must be durable, in another words need to remain relevant over an extended period of time (Morrison, 1996).

2.2.3. Market Segmentation Approaches

The segmentation approach can be divided into two groups. The first subgroup consists of priori and posteriori approaches. The other subgroup consists of forward, backward and simultaneous approaches. In priori approach the segments are chosen before the data are analyzed according to pre-existing demographic criteria such as age, sex, working status, education level, social economic status. A priori segmentations are easy to define and also the simplest segmentation to apply and use. The posteriori approach is one where the segments are determined by the data rather than by the researcher. The aim of these studies is not only to understand commonalities in opinion, but also what makes one group of users different from another. To understand how attitudes affect purchase statistical techniques are used where people with similar attitudes are combined together and so this approach is usually associated with cluster analysis (Mazanec, 1999).

If we examine the second subgroup we consider that, there are three approaches that mentioned in the segmentation literature are forward, backward and simultaneous. In the forward approach or the analysis of consumer response, consumers are divided into group according to the similarity in their behavioural response to the supply of goods and services. Here, consumers are assigned to groups by their similarity in one or more consumer characteristics. Additionally, the differences between the groups are related to behavioural differences for specific consumer characteristics. A forward segmentation approach includes grouping consumer characteristics based on their similarity in demographics, personality, attitude and benefits sought followed by discriminating groups by consumer response for a specific chosen product or service (Kuylen & Verhallen, 1981).

The second approach which is backward segmentation approach, the analysis start point is the consumer characteristics. The segment group is formed by their similarity in one or more consumer characteristics. Consumer characteristics are distinguished as general characteristics and situation-specific consumer characteristics. General characteristics consist of sex, age, stage in life-cycle, lifestyle or personality. On the other hand consumer characteristics consist of attitudes, opinions, perceptions and preferences. The differences between the groups are related to behavioural differences. Backward segmentation approach involves grouping consumer response based on their similarity in choice of products and services followed by discriminating groups by consumer characteristics (Raaij & Verhallen, 1994).

The third approach is the simultaneous segmentation that analysis of consumer characteristics and responses and consumers are assigned to groups on the basis of the relationship between consumer characteristics and behavioural responses to the supply of goods and services. Several researchers in marketing and consumer behaviour have emphasized that consumers may belong to multiple segments rather than one and only one segment. Large number of respondents fell into more than one segment and revealed both overlapping and unique (mutually exclusive) product-benefit segments. Therefore, as an alternative method simultaneous approach was developed on lodging market segmentation to analyzed both consumer response and behavioural responses simultaneously (Mazanec, 1999).

There is no certainty that which approach is being utilized because each approach has weaknesses whereas consumers have diverse demand for products and services. Consumer characteristics are more likely to show a relationship with broad patterns of consumer response than a specific response. Additionally, the relationship will often be weak and unstable over time when consumers have diverse demands and choices. Briefly, in each approach, consumer characteristics are assumed to be relevant to the explanation of consumer responses and the successive approaches were used to specify segments.

2.2.4. Market Segmentation by Using Cluster Analysis

The primary use of cluster analysis in marketing has been for market segmentation. Since 1950's market segmentation has become an important tool in marketing and all segmentation research is designed to identify groups of entities that share certain common characteristics. Researchers tend to select grouping methods largely on the basis of familiarity, availability, and cost rather than on the basis of the methods' characteristics and appropriateness. Using cluster analysis in segmentation research has been important in seeking a better understanding of buyer behaviours by identifying homogeneous groups of buyers whereas heterogeneity is probably the most important reason for segmentation (Hunt & Arnett, 2004).

Cluster analysis has been employed in the development of potential new product opportunities by clustering brands, products or competitive sets within the larger market structure. This method also has been employed by several researchers in the problem of test market selection. The identification of such homogeneous sets of test markets allows generalization of the results obtained in one test market to other test markets in the same cluster, thereby reducing the number of test markets required. Despite of these benefits of using cluster analysis in marketing segmentation research, it has short computation time and easy accommodation owing to increasing power of soft computing (Wang, 2009).

In most of market segmentation priori or backward type of analysis, also known as the traditional approach is used. They are particularly well suited to situations where it is known, from either prior research or experience, which many variables can be used to divide consumers into homogeneous sub-groups in terms of their responses or prefer. But on the other hand a response-based, post hoc or a posterior approach can be used to construct homogeneous response sub-groups. In another word the segments are formed according to continuous or post not past indication (Dolnicar & Leisch, 2004).

Cluster analysis has become a common tool for the marketing researcher. Both the academic researcher and the marketing applications researcher rely on the technique for developing empirical groupings of persons, products, or occasions which may serve as the basis for further analysis. Despite its frequent use, little is known about the characteristics of available clustering methods or how clustering methods should be employed, because of the numerous authors' failure in the marketing literature to specify what clustering method is being used. On the other hand although cluster analysis is the most common technique in market segmentation, as disadvantages it has some limitations an exploratory technique which relies considerably on the analyst's judgement and it has a result to have inaccurate strategy formulation (Bottomley & Nairn, 2004).

K-means clustering is the most widely used and fundamental technique of market segmentation among the clustering techniques. One study that applying K-means method, focuses on the current automotive maintenance industry in Taiwan to analyze and promote customer value. By the results of the study, customers are divided into high, middle and low value groups (Liang, 2009). K-means clustering techniques are also applied to determine the pricing strategies and models. In the study a limited number of clusters are obtained and each of them is representing a market segment. That model provides practical insights into pricing mechanisms (Dolgui & Proth, 2010).

In many studies k-means clustering is used as a reference method for comparison with the other method performances and outputs. In this respect, the study of k-means algorithm and artificial neural network is compared and classification of respondents on the basis of external criteria is determined among these methods. The data set is analyzed consists of usages of household cleaners brands category in different usage situations (Hruschka & Natter, 1999). The Internet is emerging as a new marketing channel, so understanding the characteristics of online customers' needs and expectations is considered a prerequisite for activating the consumer oriented electronic commerce market. Therefore proposed study is effectively segmented the real world online shopping market case based on k-means clustering and then the results of k-means compared with self-organizing maps (Kim & Ahn, 2008).

Two stage clustering is also widely applied technique in the literature with the respects of using hierarchical and non-hierarchical methods. The study is performed on the empirical data from one of the largest credit card issuing banks in China is collected that including customer satisfaction attributes and credit card transaction history. A two-stage clustering approach proposed to first, grouping similar regions together and then finding customer segmentation for each region group (Mo et al., 2010). In another study the conventional two-stage method and proposed two-stage method compared via both simulated and real-world data. The proposed two-stage method is a combination of the self-organizing feature maps and the k-means method (Kuo, Ho, & Hu, 2002). Clustering method has become an important issue with the development of information technology and has been applied to many fields as an analytical tool of data mining. A modified two-staged method is proposed which first uses the self organizing map to determine the number of clusters and the starting points and then employs the k-means method to find the final solution in information technology area (Chiu et al., 2009). In telecommunication area, the two-step procedure, that combines factor analysis and k-means, and SOM network cluster analysis performed in uncovering market segments. This research has been realized throughout a consumer data set from American Telephone and Telegraph Company (Kiang, Hu, & Fisher, 2006).

During the last decades neural network clustering models are proposed to market segmentation widely in the literature. A method for clustering time-varying data by using self-organizing maps is suggested allowing clustering of temporal data which is applied to telecommunications market segmentation on real data (Urso & Giovanni, 2007). Evaluating the feasibility of using a self-organizing map to mine web log data and providing a visual tool to assist user navigation are applied in internet area. Besides providing a meaningful navigation tool for web users, it also serves as a visual analysis tool for webmasters to understand better the characteristics and navigation behaviours of web users visiting their pages (Smith & Ng, 2003). In the education area, the study is applied the clustering and visualization capabilities of self organizing map to group and plot the top 79 MBA schools as ranked by US News and World Report into a two-dimensional map with four segments. The map should assist prospective students in searching for the MBA programs that best meet their personal requirements (Kiang & Fisher, 2007). Self organizing map is also applied to focus on the connection between human resource management as a source of competitive advantage and perceived organizational performance in the European Union’s private and public sectors (Stavrou, Charalambous, & Spiliotis, 2007).

Beyond these frequently used techniques, k-means, two-stage and self organizing map, other clustering methods are applied in many studies in the literature. The study attempts to examine the potential of fuzzy clustering in enriching methods for identifying housing submarkets. A fuzzy c-means algorithm is applied to obtain fuzzy set membership degree of census tracts to housing submarkets defined within an 85 metropolitan area in U.S. (Hwang & Thill, 2007). A maximum likelihood method is used as a cluster analysis technique in the study, which allows varying size and orientation and assumes constant volume. Models are estimated on the basis of household scanner data and determined market segments by clustering households on the basis of their purchases, brands price, sales promotion and brand loyalty (Hruschka, Fettes, & Probst, 2004). A latent class clustering approach is proposed to identify the appropriate number of health-related segments based on their socio-demographic characteristics and attitudes towards healthy eating. The study is also explored differences across segments in types of associations with food and health, as well as perceptions of food healthfulness on the 316 Danish consumers participated survey data (Chrysochou et al., 2010).

3. METHODOLOGY

In this section, we reviewed the framework of research covering research design, the scope of sample and population, measurement instruments, and data collection method. Then we described the three adopted clustering methods, k-mean, two stage clustering and self organizing map techniques under the subject of data analysis.

3.1. RESEARCH DESIGN

Research design for a segmentation study depends primarily on the segmentation model used. Parallel to this, segmentation model requires the selection of a basis both dependent and the descriptor of independent variables for the various segments. These variables can be divided into two types; general consumer characteristics and specific consumer characteristics. General consumer characteristics, including demographic and socioeconomic characteristics, personality and life style characteristics. Specific consumer characteristics are including product usage and purchase patterns, attitudes toward the product and its consumption which depends on the situation (Wind, 2004).

The clustering based segmentation design is one of the prototypical research patterns in segmentation studies, in which segments are determined on the basis of a clustering of respondents on a set of "relevant" variables (Schaffer & Green, 1998). In another word in the clustering based approach the number and type of segments are not known in advance and are determined from the clustering of respondents on their similarities on some selected set of variables. Most commonly the variables used in the clustering-based models are needs, attitudes, life style and other psychographic characteristics, or benefits sought. In our research we preferred to use clustering based approach design, using set of lifestyle trend variables to cluster respondents into similar clusters.

3.2. POPULATION and SAMPLING

The population framework of research is people who are older than 15 years and the education level is at least secondary school. According to population census 2009 of Address Based Population Registration system statistics, which was obtained from Turkish Statistical Institute, the total number of people is 20.833.552 regarding this criterion and including population of province and towns in Turkey.

Because of the population is spread on the wide geographic regions, we have to use required convenience sampling technique of cluster random (multi-stage) sampling. Cluster-area random sampling is a probability sampling technique that the population is divided into mutually exclusive groups such as blocks, and the researcher draws a sample of the groups to interview within sampled clusters. The first stage of sampling procedure was selecting a random sample of cities. Then, a sample of smaller areas for each selected cities randomly selected. Further stage was adding blocks or streets within each district. Then, systematic randomly selected of households within each block and finally randomly selection of respondent amongst residents of the household selected. This sampling procedure was done using a Kish Grid (Kish, 2004). Kish Grid is a probability based approach that requires all eligible respondents within a household be listed by age from either the oldest to the youngest or the youngest to the oldest. After all eligible respondents are identified; the respondent to be interviewed is selected by using a random number table. In this research the clusters are covered by 9 provinces of Turkey including both rural and urban population. The sample had representative characteristics of Turkey’s population. The cities and number of conducted interviews in each city are presented below.

City	Frequency	Percent
Adana	145	7,8
Ankara	274	14,7
Antalya	94	5,0
Bursa	142	7,6
Denizli	85	4,6
İstanbul	590	31,6
İzmir	307	16,4
Samsun	135	7,2
Trabzon	96	5,1
Total	**1868**	**100**

Table 3.1 : Distribution of city

Regarding to the cluster-area random sampling technique total sample number is 1868.

3.3. MEASUREMENT INSTRUMENTS

The information necessary to carry out the empirical study was collected through face-to face interviews accompanied by survey questionnaire administration to avoid unnecessary misunderstanding. Respondents whom had agreed to participate, spent approximately 2,5 - 3 hours to fill out the long questionnaire. Selecting the variables to include in an analysis is always crucial, because must be made before the analyse process and so design of questionnaire must be made carefully by researcher. The questionnaire has two main parts; the first is including demographical characteristics and respondents' purchase behaviour, the second part is the rating of lifestyle trend items.

The first part was designed by the researcher respect to the purpose of research and this part collected background and descriptive information on the respondents. Questions about participants' gender, age, education level, occupation size of household and place of living were included. Additionally, respondents were asked to indicate their thoughts, beliefs, values, attitudes, behaviours, emotions, perceptions and interests based on a wide spread of specific sections. These sections are health, travel, sports, music, eating, saving habits, house furnishing, purchasing, media habits, personal cleanup and care products, food, drinks, cleaning products, durable goods, automobile, furniture and clothing. Each part has many questions including agreement statements and brand usage to get depth information about their attitudes and purchase behaviour.

The second part consists of life style trend items that was very important to this study and had a great impact on clustering results. Literature review and qualitative research techniques become invaluable to identify the proper lifestyle statements before design the trends and items on the questionnaire. Through the review of the relevant literature, the primary trends for consumer behavior were identified. Earlier studies of life style segmentation by Lastovicka (1982), Yankelovich (1971) and Plummer (1971) helped to determine the trend statements. On the other hand, a qualitative investigation allowed exploring the richness and depth of individual perceptions, experiences, aspirations and values. Additionally, conducting focus groups and depth interviews provide the insight, the conceptual knowledge, and the consumer's exact language necessary to design the segmentation questionnaire. Typically, verbatim comments from consumers are used to build batteries of lifestyle statements. Greater understanding of these relations could lead to new theoretical insights about market segmentation.

The lifestyle trend instrument consisted of totally 50 items regarding to 7 trends (See Appendix 1.1). Each item was measured on a 4 point Likert type agreement scale (1 = Certainly disagree to 4 = Certainly agree). All Cronbach's alphas (α) were above 0.60 indicating a satisfactory internal consistency of the instrument (Nunnally & Bernstein, 1994). The functionality, accelerated pace of life and time dimensions was measured with the accelerated society trend scale (3 items; α = 0,73). Social acceptance, respect to others and intercultural competence dimensions was measured with the cultural diversity trend scale (9 items; α = 0,60). Preferring natural things and being consciousness about the re-cycling and eco dimensions was measured with the eco-living trend scale (6 items; α = 0,66). Conspiracy, hypocrisy, self-security and loss of corporate trust dimension was measured with the erosion of trust trend scale (10 items; α = 0,64). Virtual living, transparency, sharing privacy and the other activities on the net such as blogging wiking and virtual communities dimensions was measured with the living on the net trend scale (6 items; α = 0,72). Risk-taking, self- reliance, denial of authority, refusal of taboos dimensions was measured with the masters of the youniverse trend scale (5 items; α = 0,63). Physical attention, form-consciousness, personal style, health consciousness dimensions was measured with the self reflection trend scale (11 items; α = 0,63).

3.4. DATA COLLECTION METHOD

For this study the data were obtained from a syndicated-tracking survey that the DAP Research company undertook since 1994, for the purpose of market segmentation and get a depth understanding about consumer characteristics. Syndicated research is a single research study conducted by a research company with its results available, for sale, to multiple companies. On the other hand tracking research technique is a periodic market research that involves repeated observations of the same items over time and used to pick up important base-line information about the market. Additionally, tracking research helps to ensure a proper representation of the population and all data is collected via face-to-face interviews. The research data have both two aspects, one according to customer which is syndicated, and the other tracker research according to periodical and continues research methodology. The database represents data for a two year period. It is common that due to life style trends in the society, some changes in the data-mining model are expected after the model becomes stable. Therefore, the base assumption for this data set is that significant changes would be observed over time.

The raw data was collected from 1868 households by cluster-area random sampling. The interviewers were given cluster cards and record sheets which outlined the procedures for the selection of the households. Respondent selection was done randomly to choose eligible respondents for an interview among the household members. The interviewers told the respondents that they were from a research company which was conducting a survey to obtain information for the consumers' purchase behaviour and their attitudes toward the life style trends. Before the fieldwork process, a pilot study had involved on a small sample of 30 respondents. Only the minor changes on the wording were deemed necessary by the authors. Due to the length of the interview, if required 2 visits were organized for some respondents.

During the fieldwork process the questionnaire was administered by interviewers using computer aided personal interview (CAPI) technique through face to face interviews. The CAPI is a method of the research which is combining the properties of traditional methods with new information technologies. Instead of collecting data on paper questionnaires, interviewers would key in responses from their respondent or interviewee directly into a purpose built computer program on a small device or laptop to enter data directly via a keyboard. The main reasons to prefer this method are decreasing interviewer source of error, eliminate a separate process of data entry and supply of a clean data set with obtained predefined logic checks automatically via scripts, for inconsistent or contradictory responses during the interview. Due to these advantages, computer aided personal interview method provides increasing the cost of effectiveness and the saving of time especially when the survey population is large and the questionnaire design is complex.

Other than the single and multiple choice categorical questions, respondents were asked to answer lifestyle trend measurement items on the basis of a 4-point scale. Scale points require the individuals to make a decision on their level of agreement. In order to understand positive or negative way of agreement for each trend statement, the

“Neither/Undecided” point cancelled from 5 point Likert type scale. The scale ranged from 1 ‘Certainly disagree’, 2 ’Disagree’, 3 ’Agree’ and 4 ‘Certainly agree’.

3.5. DATA PROCESSING

After the standard procedures, such as verify, check, delete and coding the data, the mean score of each customer on each trend is calculated, and becomes the basis of cluster analysis. Then, before feeding data into the clustering analysis, the mean score of trends were normalized to eliminate scale effects. Scaling variables is believed to be essential especially when clustering algorithm uses Euclidean metric to measure distance between vectors. Standardization of the data was not only necessary to achieve physically meaningful classifications but it is also a useful precursor to the application of cluster analysis. Although there are many different standardization measures, in this study we applied traditional standardization approach that function is as following.

$$z = \frac{x - \mu}{\sigma} \qquad (3.1)$$

where μ is the mean, σ is the variance of all measurements of trends. To this end, we normalized dataset and scaled all variables linearly so that their variances were equal to one and means equal to zero.

3.6. DATA ANALYSIS

After the fieldwork process, data was obtained from the laptops with transferred into the SQL server. Then by the help of software, the raw data exported to the Statistical

Package for Social Sciences (SPSS) programme. Version 15.0 for Windows of the SPSS statistical package was used to perform statistical analysis. Additionally, SPSS Clementine was used to perform cluster analysis.

Primarily the characteristics of sample and demographical information about respondents will be shown. Afterwards cluster analysis was performed on the standardized mean scores which has become the basis of cluster analysis. In order to determine the most appropriate clusters and focused on finding the best results for this data set, we applied k-means, two stage clustering and self organizing map algorithms. We propose a cluster analysis for market segmentation into two phases. First, we perform the k-means and two stage cluster analysis as a standard analytical tool of market segmentation. Second we perform the self organizing map which has attracted a wide range of application and used in market segmentation in recent years. The results of self organizing map, traditional k-means and two stage clustering methodologies compared base on the characteristics of segments.

Finally, after segmentation of the market, we used Anova to test for the significant differences among k-means, two stage clustering and self organizing map approached to find more accurate classification algorithm. Additionally, Anova and cross tabulation analysis was used to recognize the characteristics of sub divided clusters.

4. RESULTS AND DISCUSSION

This section includes the demographic characteristics of participants, and experimental results of clustering analysis. We compare the performance of the three methods, k-means, two stage and self organizing map, based on three clusters output.

4.1. DEMOGRAPHIC CHARACTERISTICS

The data for the main survey was collected in Turkey. The characteristics of the respondents such as age, gender, marital status, education, income level, employment status and social economical status were asked in questionnaire. 1868 surveys were carried out with participants in this study.

Table 4.1 summarizes the profile of the respondents. Most of the respondents' age is between 15 and 24 years (39.9%) and the mean of the respondents' age is 30,7 years old. 48,6% of the respondents are male and 51,6% are female. The statistical results show that 33,7% of the respondents education level is secondary school, while 45,6% of the respondents education level is from high school to university and 20,8% of the participants education level is university and above. If we look at the employment status, we consider that totally 43,7% of the participants consists of housewives and students. On the other hand 9,4% are unemployed and 4,7% are retired.

Table 4.1 : Characteristics of the Sample of Study

		Frequency	Percent%
Age	15-24	751	39,9
	25-34	509	27
	35-44	342	18,2
	45 +	283	15
	Mean	30,7	
Gender	Male	916	48,6
	Female	969	51,4
Education	Secondary school	636	33,7
	High school	577	30,6
	Vocational high school	254	13,5
	Private high school	28	1,5
	University	371	19,7
	Master's degree / graduate	20	1,1
Employment	Business owner	58	3,1
	Professionals / Self employment	51	2,7
	Trader	11	0,6
	Small Trader	52	2,8
	Manager	31	1,6
	Servant in private sector	141	7,5
	Servant in public sector	142	7,6
	Skilled / Unskilled Worker	232	12,3
	Student	398	21,1
	Housewife	427	22,6
	Unemployed	178	9,4
	Retired	89	4,7
	Other	74	3,9
Family Class	Upper Class	35	1,9
	Upper Middle Class	217	11,5
	Middle Class	1169	62
	Lower Middle Class	345	18,3
	Lower	120	6,4
Socio-Economic Status	AB	350	18,5
	C1	892	47,3
	C2	556	29,5
	DE	88	4,7

The Table 4.1 results shows that, the percentage of the given answers to the question that, which class is more appropriate for your family by you, is as follow; 13,4% of the respondents belong to upper and upper middle class and %24,7 of the respondents belong to lower and lower middle class. The most part of the respondents belong to a middle class (62%). Socio-Economic Status was labelled as AB, C1, C2 and DE from high to low according to head of the households' education and employment cross results (See Appendix 4.1). 18,5% of the households statue is A and B, 29,5% of the households statue is C2 and 4,7% of the households statue is DE. Approximately, the half of the respondents' socio economic statue is C1 (%47,3).

4.2. CLUSTER ANALYSIS RESULTS

K-means and SOM clustering approaches don't automatically determine how many clusters are represented in the data and so deciding where to cut the steams of a dendogram is a subjective evaluation. Therefore, Ward's hierarchical method is used to determine the number of clusters and then k-means, two stage, self organizing map clustering method is performed to yield clusters. According to Ward's method the input of seven trends can be represented by three outputs, which form the segments.

4.2.1. Characteristics of Segments based on Lifestyle Trends

Cluster Analysis, which refers to a group of techniques for attribute based classification, was used to derive the consumption segments. This clustering technique identifies members of existing groups by looking at the responses to questions of each respondent in the sample to see if that respondent is similar to any existing group and, simultaneously, different from the respondents in any other group. The following table shows the distrubition of final cluster sample for three clustering approaches.

Table 4.2 : Final cluster sample distribution

	Frequency			Percentage (%)		
	K-mean	Two Stage	SOM	K-mean	Two Stage	SOM
Cluster 1	556	839	363	29,8	44,9	19,4
Cluster 2	301	358	612	16,1	19,2	32,8
Cluster 3	1011	671	893	54,1	35,9	47,8

As Table 4.2 shows that 54,1% of the sample is in Cluster 3 by K-mean clustering approach; 44,9% of the sample is in Cluster 1 by Two Stage clustering approach; and 47,8% of the sample is in Cluster 3 by SOM clustering approach.

The trends mean score of each cluster have been identified in the final cluster center as shown in Table 4.3 All trends descriptions have been examined and appropriate name is given to the clusters. Based on the average values, segments were labelled as consistent, rebellious and traditional.

Table 4.3 : Final cluster centers with trends

	K-Means			Two Step			SOM		
	1	2	3	1	2	3	1	2	3
Accelerated Society	0,12	0,34	-1,95	-0,62	0,22	-2,03	-0,89	0,60	-2,07
Cultural Diversity	0,26	-0,17	0,36	0,29	-0,16	0,40	0,28	0,03	0,38
Eco Living	0,85	-0,02	0,67	0,97	0,04	0,47	0,72	0,41	0,71
Erosion of Trust	0,28	-0,07	0,45	0,25	-0,10	0,62	0,29	0,10	0,48
Living on the Net	-1,56	0,12	-0,97	-1,37	-0,11	-0,94	-1,05	-0,89	-1,00
Masters of the Youniverse	-0,44	0,18	-0,03	-0,53	0,22	0,21	-0,17	-0,19	-0,06
Self Reflection	0,28	-0,05	0,42	0,38	-0,06	0,40	0,31	0,11	0,44

Cluster 1, having mostly the characteristics of eco living trend and was named as consistent. Because they are consciousness about eco and they try to prefer natural foods. On the other hand, living on the net and master of the youniverse trends have negative affects on this segment, which means that they don’t prefer to live on the net, and they believe to others to live without taking risks.

Cluster 2, having mostly the characteristics of accelerated society trend and was named as rebellious. Because they are accelerating the life and they save the time with functionality. Living on the net is an inevitable necessity for them and so they are being master of oneself as far as risk taking and being powerful at the same time.

Cluster 3, having mostly the characteristics of erosion of trust trend and was named as conventional. This group is loss of their self-security and corporate trust. On the other hand cultural diversity and self reflection trends have also affects on this segment, which means that they open to everyone, have respect to others and reclaim a contemporary self-identity. On the other hand, negative affect of accelerated society trend indicates that, accelerating the life and time is not the fundamental purpose of these people.

In the same way, the final segments for two stage and SOM clustering approaches are broadly overlapping with the result of k-means' segments based on trends. However, there is a noticeable difference that living on the net trend has become a negative affect on the rebellious segments for two stage and SOM approaches.

One way analysis of variance (ANOVA) was conducted to capture the attribute differences in each of the 7 trends across the three segments formed through k-means, two stage and self organizing map approaches. Statistically significant differences are detected for all trends among all three clusters at $p < 0,05$. The results for the k-means clustering are presented in Table 4.4.

Table 4.4 : Significance test for k-means cluster analysis

ANOVA				
K-Means	**Mean Square**	**df**	**F**	**Sig.**
Accelerated Society	1081,46	2	2021,00	0,000
Living on the Net	277,51	2	545,47	0,000
Eco Living	77,19	2	202,62	0,000
Masters of the Youniverse	46,69	2	74,50	0,000
Erosion of Trust	32,57	2	107,89	0,000
Cultural Diversity	32,19	2	121,99	0,000
Self Reflection	25,57	2	120,16	0,000

As shown in the Table 4.4 the K-means cluster analysis is significant at the 0.05 significance level. The top three effective variables are Accelerated Society, Living on the Net and Eco Living in the cluster analysis while segmenting the consumers in terms of lifestyle trends.

Table 4.5 : Significance test for two stage cluster analysis

ANOVA				
Two Stage	**Mean Square**	**df**	**F**	**Sig.**
Accelerated Society	684,57	2	712,54	0,000
Living on the Net	198,14	2	333,64	0,000
Masters of the Youniverse	128,51	2	238,43	0,000
Eco Living	119,14	2	354,62	0,000
Erosion of Trust	63,37	2	235,74	0,000
Cultural Diversity	37,94	2	147,20	0,000
Self Reflection	29,17	2	139,64	0,000

As shown in the Table 4.5 the Two Stage cluster analysis is significant at the 0.05 significance level. the top three effective variables are Accelerated Society, Living on the Net and Masters of the Youniverse in the cluster analysis while segmenting the consumers in terms of lifestyle trends.

Table 4.6 : Significance test for SOM cluster analysis

ANOVA				
SOM	**Mean Square**	**df**	**F**	**Sig.**
Accelerated Society	1293,27	2	4199,42	0,000
Erosion of Trust	26,44	2	85,72	0,000
Cultural Diversity	21,45	2	77,88	0,000
Self Reflection	19,97	2	91,28	0,000
Eco Living	19,68	2	44,47	0,000
Living on the Net	3,72	2	4,64	0,010
Masters of the Youniverse	3,59	2	5,34	0,005

As shown in the Table 4.6 the SOM cluster analysis is significant at the 0.05 significance level. The top three effective variables are Accelerated Society, Erosion of Trust and Cultural Diversity in terms of lifestyle trends.

Table 4.7 gives the characteristics results of each cluster for k-means clustering approach. According to the age mean results, rebellious segment are younger than the others. The percentage of 15-24 years old participants are 46,4%, where as 40,7% for consistent segment and %37,5 for conventional segments. As the same way, conventional group is the oldest segment within them.

If we examine the gender results, there is no difference between male and female for consistent segment. In rebellious segment male percentage (52,4%) is higher than woman (47,6%). On the other hand, opposite characteristics of rebellious segment is seen for conventional segment, as female percentage (52,9%) is higher than man (47,1%).

Table 4.7 : Sample characteristics of clusters for k-means approach

		Consistent	Rebellious	Conventional
Age	15-24	40,7	46,4	37,5
	25-34	29,1	28,9	25,3
	35-44	16,1	16,3	19,8
	45 ve uzeri	14,1	8,5	17,3
	Mean	29,9	28,2	31,8
Gender	Male	49,5	52,4	47,1
	Female	50,5	47,6	52,9
Education	Secondary school	33,9	29,6	34,8
	High school	27,5	34,8	31,1
	Vocational high school	16	10,7	12,9
	Private high school	2,1	0,7	1,3
	University	19,3	23,6	18,8
	Master's degree / graduate	1,1	0,6	1,2
Employment	Business owner	3,5	5	2,3
	Professionals / Self employment	3,1	3	2,4
	Trader	0,4	0,5	0,7
	Small Trader	2,3	3,5	2,8
	Manager	1,2	3,1	1,5
	Servant in private sector	5,9	9	7,9
	Servant in public sector	6,6	8,5	7,8
	Skilled / Unskilled Worker	14,5	12,3	11,2
	Student	22,2	23,4	19,9
	Housewife	19,4	15,4	26,4
	Unemployed	10	11,1	8,6
	Retired	4,8	2,6	5,3
	Other	5,9	2,6	3,2
Family Class	Upper Class	0,8	2,1	2,4
	Upper Middle Class	14,2	11,4	10,1
	Middle Class	59,2	66,2	62,3
	Lower Middle Class	18,9	15,2	18,8
	Lower	6,9	5,1	6,4
Socio-Economic Status	AB	19,8	19,1	17,7
	C1	45,6	47	48,3
	C2	29,5	28,2	29,8
	DE	5,1	5,7	4,2

The results show that percentage of university and above (24,2%) education level for rebellious segment is higher than other segments. The other education levels percentages are very near regarding of three segments.

Employment status show that conventional segment contains of mainly housewives (26,4%). The student percentages are also high for consistent and rebellious segments. If we compare rebellious segment with the others, unemployed people percentage is higher and retired people percentage is lesser.

There is no significant difference between segments according to socio economic status. On the other hand, if we examine the family class results we consider that middle class percentage (66,2%) is higher for rebellious group than the other segments and lower class percentage (20,3%) is lesser.

4.2.2. Performance Comparison of Three Clustering Approaches

One way to evaluate the performance of three clustering results is to compare the total within cluster and between cluster variance. For the same number of clusters, the smaller the within cluster variance is, the more homogenous are the cluster members. Therefore, it is a good indication of the clustering performance. Table 4.8 compares the total within clusters variance of three approaches.

Table 4.8 : Total within cluster variance of the three approaches

Approach	Variance
K-Means	5278,04
Two Stage	5902,73
SOM	5648,12

The results show that the K-means analysis generate clustering results slightly better than that of the two stage and self organizing map approach based on total within cluster variance.

We also compared the similarity of the clustering results generated by the three different approaches based on the composition of the cluster members. The following tables show the cross-tabulation of the three cluster results for each pair of the methods. Cases in the diagonal entries of a table are grouped in the same clusters by the pair of clustering methods.

Table 4.9 : Cross-tabulation of the k-means' clusters and two stage's clusters

		Two Stage		
		Cluster 1	**Cluster 2**	**Cluster 3**
K-means	**Cluster 1**	497 (26,6%)	44 (2,4%)	15 (0,8%)
	Cluster 2	7 (0,4%)	287 (15,4%)	7 (0,4%)
	Cluster 3	335 (17,9%)	27 (1,4%)	649 (34,7%)
	Total	839 (44,9%)	358 (19,2%)	671 (35,9%)

Table 4.9 indicates that 76,7% (1433 out of 1868) of the cases are grouped in the same clusters by K-means and Two Stage clustering approaches.

Table 4.10 : Cross-tabulation of the k-means' clusters and SOM's clusters

		Self Organizing Map		
		Cluster 1	**Cluster 2**	**Cluster 3**
K-means	**Cluster 1**	181 (9,7%)	357 (19,1%)	18 (1%)
	Cluster 2	45 (2,4%)	255 (13,7%)	1 (0,1%)
	Cluster 3	137 (7,3%)	0 (0%)	874 (46,8%)
	Total	363 (19,4%)	612 (32,8%)	893 (47,8%)

Table 4.10 shows that 70,2% (1310 out of 1868) of the cases are grouped in the same clusters by K-means and SOM clustering approaches.

Table 4.11 : Cross-tabulation of the two stage's clusters and SOM's Clusters

		Self Organizing Map		
		Cluster 1	**Cluster 2**	**Cluster 3**
Two Stage	**Cluster 1**	236 (12,6%)	313 (16,8%)	290 (15,5%)
	Cluster 2	53 (2,8%)	290 (15,5%)	15 (0,8%)
	Cluster 3	74 (4%)	9 (0,5%)	588 (31,5%)
	Total	363 (19,4%)	612 (32,8%)	893 (47,8%)

Table 4.11 shows that 59,6% (1114 out of 1868) of the cases are grouped in the same clusters by Two Stage and SOM clustering approaches.

The results of three cross-tabulation tables shows that, the largest frequencies occur in the diagonal terms, indicating pairs of procedures yield a large percent to the same three segments. However, the most overlapped methods are k-means and two stage respect of the 76,7% of the respondents are grouped in the same clusters.

5. CONCLUSIONS

Studies on consumer lifestyles trends have long historical past in the related literature. On the other hand, there are several conceptual articles and few empirical studies associated with understanding consumer behaviour based on life style trends during the market segmentation. However, there is still a lack of empirical research which aims to explore determinants of market segmentation based on lifestyles trends.

In this study, we applied the three clustering approaches, k-means, two stage and self organizing map, to group the seven trends into three segments. We first used Ward's hierarchical method to determine the number of clusters. We then performed three clustering approaches and compared the result of analysis. The results show that k-means method generates distinct groups as good as, if not better than, that of the two-stage and self organizing map approaches. Both the cross tabulation for measuring the similarity between the cluster compositions revealed high degree of agreement among the outputs of the k-means, two stage and self organizing map.

We applied the three clustering approaches to group the trends into three segments: consistent, rebellious and conventional. Consistent segment is more close to eco living trend, rebellious segment is more open to accelerated society living on the net and being master of the youniverse, and conventional segment is more close to erosion of trust, self reflection and cultural diversity trends.

The characteristics of segments are differing according to demographical variables. Consistent segment has no significant gender discrimination, but the students' percentage is high in this segment than the others. Conventional segment's characteristics is older, female percentage higher than male percentage and housewives are generally belongs to this group. Rebellious segment's characteristic is younger, male percentage higher than female percentage, university and above education level is higher than the other segments.

The major contribution of this thesis is that it is to investigate the determinants of market segmentation based on lifestyles scale from literature and which is one of the biggest concern. Moreover in heterogeneous markets, segmenting consumers according to their values should be an important tool in the strategic kit of marketing segmentation.

REFERENCES

Books

Baker, M., & Hart, S. (2008). The Marketing Book. 5th ed, *Elsevier's Science & Technology*, pp. 222-242.

Berson, A., Smith, S., & Thearling, K. (2000). Building data mining applications for CRM. New York: McGraw-Hill.

Everitt, B. S., (1993). Cluster analysis.3rd ed., Halsted Press, Division of Wiley, New York.

Hair, J.F., Anderson, R.E., Tatham, R. L., & Black, W. C. (2000). *Multivariate Data Analysis*. 5th ed, London.

Kish, L., (2004). Statistical design for research. *John Wilney and Sons,* Hoboken, New Jersey, pp.41-44.

Kotler, P., (1997). *Marketing Management-Analysis, Planning, Implementation and Control.* 9th ed., Prentice-Hall, Upper Saddle Rive, NJ.

Morrison, A.M., (1996). *Hospitality and Travel Marketing*. 2nd ed., Albany, New York: Delmar.

Mulvey, L., & Gingold, J. (2007). Microarray Clustering Methods and Gene Ontology. *PBworks*.

Schiffman, L.G., & Kanuk, L.L. (2000). *Consumer Behaviour*. 7th ed, London. Prentice Hall.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco: W. H. Freeman.

Theodoridis, S., & Koutroumbas, K. (2006). 3rd ed, Pattern Recognition, University of Athens, Greece.

Wedel, M., & Kamakura, W.A. (2000). Market Segmentation: Conceptual and Methodological Foundations. 2nd ed, Boston, MA: Kluwer Academic.

Wind, P.E., (2004). Marketing Research and Modeling. *International Series in Quantitative Marketing*, Vol. 14

Periodicals

Back, B., Sere, K., & Vanharanta, H. (1998). Managing complexity in large data bases using self-organizing maps. *Accounting Management and Information Technologies*, pp. 191-210.

Barney, J. B., & Hoskisson, R. E. (1990). Strategic groups: Untested assertions and research proposals. *Managerial and Decision Economics*, 11, pp. 187- 198.

Bottomley, P., & Nairn, A. (2004). Blinded by science: the managerial consequences of inadequately validated cluster analysis solutions. *International Journal of Marketing Research*, 46(2), pp. 171–187.

Chiu, C.Y., Chen, Y.F., Kuo, T., & Ku, H.C. (2009). An intelligent market segmentation system using k-means and particle swarm optimization. *Expert Systems with Applications*, pp. 4558–4565.

Deichmann, J., (2003). Geography matters: Kohonen classification of determinants of foreign direct investment in transition economies, *Journal of Business Strategies,* 20 (1), pp. 23–25.

Dibb, S., (1998). Market segmentation: strategies for success, *Marketing Intelligence & Planning,* pp. 394–406.

Dolgui, A., & Proth, J.M. (2010). Pricing strategies and models. *Annual Reviews in Control*, pp. 101–110.

Dolnicar, S., & Leisch, F. (2004). Segmenting markets by bagged clustering. *Australasian Marketing Journal*, 12(1), pp. 51–65.

Everitt, B. S., (2001). Cluster analysis. *John Wiley & Sons*, New York.

Gehrt, K. C., & Shim, S. (1998). A shopping orientation segmentation of French consumers: Implications for catalog marketing. *Journal of Interactive Marketing*, 12(4), pp. 34–46.

Gerstengarbe, F.W. & Werner, P. C. (1999). Applying Non-Hierarchical Cluster Analysis Algorithms to Climate Classification: Some Problems and their Solution. *Theoretical Applied Climatol*ogy, 64, pp.143-150.

Gong, X., & Richman M.B. (1995). On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of climate*, 8., pp. 897 – 931.

Güler, C., Geoffrey, D.T., McCray J.E., & Turner A.K. (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*, 10. pp. 455-474.

Ha, S. H., (2007). Applying knowledge engineering techniques to customer analysis in the service industry, *Advanced Engineering Informatics*, 21, pp. 293–301.

Hatten, K. J., Schendel D. E., & Cooper A. C. (1978). A strategic model of the U.S. brewing industry. *Academy of Management Journal*, 21, pp. 592-610.

Hruschka, H., & Natter, M. (1999). Comparing performance of feedforward neural nets and K-means for cluster-based market segmentation. *European Journal of Operational Research,* pp. 346-353.

Hruschka, H., Fettes, W., & Probst, M. (2004). Market segmentation by maximum likelihood clustering using choice elasticities. *European Journal of Operational Research,* pp. 779-786.

Huang, Z., Chen, H., Guo F., Xu J.J., Wu, S., & Chen, W.H. (2006). Expertise visualization: An implementation and study based on cognitive fit theory. *Decision Support Systems*, pp.1539-1557.

Huang, J. H., Tzeng, G. H., & Ong, C. S. (2007). Marketing segmentation using support vector clustering. *Expert Systems with Applications*, 32, 313–317.

Hunt, S.D., & Arnett, D.B. (2004). Market segmentation strategy, competitive advantage, and public policy: grounding segmentation strategy in resource-advantage theory. *Marketing Journal*, Australasian, 12(1), pp. 7–25.

Jang, J.S.R., Sun, C.T., & Mizutani, E. (1997). Neuro-Fuzzy and Soft Computing. *Pearson Education*, pp. 426-427.

Kaya, M., (2005). An Algorithm for Image Clustering and Compression Using an Annealed Fuzzy Hopfield Neural Network, *International Journal of Signal Processing*, Vol.1, pp.80-88.

Kiang, M.Y., Hu, M.Y., & Fisher, D.M. (2006). An extended self-organizing map network for market segmentation a telecommunication example. *Decision support system*, pp. 36-47.

Kiang, M.Y., & Fisher, D.M. (2007). Selecting the right MBA schools–An application of self-organizing map networks. *Expert Systems with applications*, pp. 36-47.

Kim, K.J., & Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications,* pp. 1200–1209.

Kohonen, T., (1997). Self-Organizing Maps. Berlin: Springer-Verlag.

Kohonen, T., (1989). Self-organization and associative memory, Springer-Verlag.

Kuo, R.J., Ho, L.M., & Hu, C.M. (2002). Integration of self-organizing feature map and *K*-means algorithm for market segmentation. *Computers & Operations Research*, pp. 1475-1493.

Kuylen, A.A.A., & Verhallen, T.M.M. (1981). The Use of Canonical Analysis. *Journal of Economic Psychology*, 1, pp. 217-237.

Linda, P. M., (1999). Segmenting Publics by Lifestyles. *Public Relations Quarterly*, Fall, pp. 46-47.

Mashor, M.Y., (1998). Improving the Performance of K-Means Clustering Algorithm to Position the Centres of RBF Network. *International Journal of the Computer*, The Internet and Management, vol.6.

Mazanec, J.A., (1999). Simultaneous positioning and segmentation analysis with topologically ordered feature maps. *European Journal of Marketing*, pp. 183–207.

Mingoti, S. A., & Lima, J. O. (2006). Comparing SOM neural network with Fuzzy C-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research.*

Mo, J., Kiang, M.Y., Zou, P., & Li, Y. (2010). A two-stage clustering approach for multi-region segmentation. *Expert Systems with Applications,* pp. 7120–7131.

Plummer, Joseph T., (1971). Life Style Patterns and Commercial Bank Credit Card Usage. *Journal of Marketing*, 35, No. 2.

Punj, G., & Stewart, D. W. (1983). Cluster Analysis in Marketing Research. *Journal of Marketing*, Vol. 20 pp.134-148.

Schaffer, C.M., & Green, P.E. (1998). Cluster-based market segmentation: some further comparisons of alternative approaches. *Journal of the Market Research Society*, 40(2), pp. 155 –161.

Small, H., (1998). A general framework for creating large-scale maps of science in two or three dimensions. *Scientometrics*, 41, pp.125-133.

Smith, K.A., (1999). Neural networks for combinatorial optimization: A review of more than a decade of research. *Informs Journal on Computing*, 11 (1), pp. 15–34.

Smith, K.A., & Ng, A. (2003). Web page clustering using a self-organizing map of user navigation patterns. *Decision Support Systems*, pp. 245-256.

Stavrou, E.T., Charalambous C., & Spiliotis, S. (2007). Human resource management and performance: A neural network analysis. *European Journal of Operational Research* pp. 453–467.

Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications*, 27, pp. 265–276.

Ultseh, A., (1995). Self organizing neural networks perform different from statistical k-means clustering. *Gesellschaft f. Klassifikation*, Basel.

Wang, C.H., (2009). Outlier identification and market segmentation using kernel-based clustering techniques, *Experts Systems with Applications*, pp. 3744–3750.

Ward, J.H., (1963). “Hierarchical Grouping to Optimize an Object Function,” *Journal of the American Statistical Association*, 58, pp. 236-244.

Other Sources

Albayrak, S., & Amasyalı, F. (2003). Fuzzy C-Means Clustering On Medical Diagnostic Systems. International XII. Turkish Symposium on Artificial Intelligence and Neural Networks-TAINN 2003, İstanbul.

Berks, G., Keyserlingk, D.G., Jantzen, J., Dotoli, M., & Axer, H. (2000). Fuzzy Clustering- A Versatile Mean to Explore Medical Database. ESIT2000, Aachen, Germany.

Chrysochou, P., Askegaard, S., Grunert, K.G., & Kristensen, D.B. (2010). Social discourses of healthy eating. A market segmentation approach, *Appetite*, pp. 288-297.

Demirel, M.C., (2004). Cluster Analysis of Streamflow Data over Turkey, Istanbul Technical University.

Huang, Z., (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values, *Data Mining and Knowledge Discovery*, pp. 283–304.

Hwang, S., & Thill, J. C. (2007). Using fuzzy clustering methods for delineating urban housing submarkets. *In Proceedings of the 15th international symposium on advances in geographic information systems*.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Survey*, pp. 264–323.

Liang, Y.H., (2009). Combining the K-means and decision tree methods to Promote Customer Value for the Automotive Maintenance Industry, *IEEE,* Taiwan, ROC.

Nunnally, J.C., & Bernstein, J.H. (1994). Psychometric theory. 3rd ed, New York: McGraw-Hill.

Raaij, W. F.V., & Verhallen, T.M.M. (1994). *Domain-specific Market Segmentation*. Erasmus University and Tilburg University, Netherland.

Urso, P., & Giovanni, L.D. (2007). Temporal self-organizing maps for telecommunications market segmentation, *Neurocomputing*.

Vesanto, J., (1999). SOM-based data visualization methods. Intelligence Data Analysis 3 (2), pp. 111–126.

Wedel, S., & Kamakura, W. (1997). Market segmentation: Conceptual and methodological foundations. *Boston*: Kluwer.

APPENDICES

APPENDIX 1- Life-style trend statements

Trend Name	Statement
Accelerated Society	There are so many things I have to do during the day that I look for solutions that will make things easier, that will gain me time
Accelerated Society	No one in our house waits for everyone else to gather around the dinner table anymore, whoever's hungry puts some food in a tray and sits in front of the tv
Accelerated Society	Quickly prepared and quickly eaten foods have started to comprise more and more of my meals
Cultural Diversity	There is no right or wrong way to live. People who judge the way other people live are only enjoying voice to their own prejudices
Cultural Diversity	People living in parts of the country relatively free of social and economic problems are expected not to "bail out" the less fortunate
Cultural Diversity	Everybody should be free to do their own thing
Cultural Diversity	I'd like my neighbours to be people like me.
Cultural Diversity	I make friends with people who have similar life styles and values as me.
Cultural Diversity	I'd like to reveal my ethnic traditions, symbols and pass these on to my children
Cultural Diversity	My interest in far away cultures like ındian, african and south american increase thanks to documentaries, the ınternet, tv etc.
Cultural Diversity	From clothing to food, products with local features are mor meaningful and valuable to me
Cultural Diversity	I find it natural that today conservative people also reveal a new life style with ostentatious clothes, cars and homes
Eco-living	I would be willing to pay 10 % more for products when, I am sure that they would not harm the environment
Eco-living	We have limited resources, I am for sharing some products with other people.
Eco-living	It eases my conscience to use environment friendly brands.
Eco-living	The reason some illnesses are so prevalent these days is unnatural foods
Eco-living	The government and corporations should pioneer and build the infrastructure for recycling (separating glass, plastic, paper wastes)
Eco-living	I think it is less polluting to the environment when we throw away recyclable materials such as paper, glass separately
Erosion of trust	Problems with business world are so big business that it's often difficult to know where to go if you have a complaint or request
Erosion of trust	Its best at all times to be honest with people
Erosion of trust	It's becoming absolutely necessary to protect your home against intruders
Erosion of trust	Because of the changes in the economic condition of the country and the world, young people can no longer take for granted that they will live a better than their parents
Erosion of trust	These people are hopeless
Erosion of trust	The quality of products made by the very big companies has been decending
Erosion of trust	In order to ensure my family's security, I'd like to know what goes on in my environment and have a say in the decisions made
Erosion of trust	I'd like to do something to prevent my personal information from passing into other people's hands
Erosion of trust	I cannot speak freely on the phone for fear of being bugged
Erosion of trust	I face such outrageous lies on the media and the ınternet that I find it hard to believe anyone

APPENDIX 1- (Continued): Life-style trend statements

Trend Name	Statement
Living on the Net	The fact that some youngsters and children are too interested in the fictitious environments created on the ınternet severs their ties to the real world
Living on the Net	I always enjoy sharing my knowledge, experience and opinions. Now this is very easy with ınternet forums and blogs
Living on the Net	Some friendships I have made on the ınternet are more fulfilling from those in real life
Living on the Net	I don’t feel the need to hide my emotions and opinions on the Internet, that is why I become more intimate than in real life
Living on the Net	I use the many opportunities the ınternet provides in order to improve myself and increase my creativity
Living on the Net	I don’t have any inhibitions about revealing my private life in the virtual or real world
Masters of the Youniverse	I like to make my own decisions, nobody tells me what to do
Masters of the Youniverse	I don’t need anyone to tell me what products are good for me. I like to make my own choices
Masters of the Youniverse	I don’t care as much about rules and taboos as I used to, and I act the way ı feel comfortable
Masters of the Youniverse	I can take risks in real life in order to achieve my dreams
Masters of the Youniverse	There aren’t a specific set of rules to follow in order to be successful in life
Self Reflection	I like to buy products that reflect my style personality
Self Reflection	I am concerned about trying to stay in shape
Self Reflection	Sometimes it is worth making sacrifices in order to look more.
Self Reflection	In our society, “fine feathers make fine birds” ideology is gaining in importance.
Self Reflection	I would like to learn and implement the aproaches that will allow me to be healthy but ı cannot really trust the ever changing approaches
Self Reflection	My most precious organ is my brain, but I don’t really know what to do in order to keep it healthy
Self Reflection	In order to be healthy when old, it is important to maintain an adequate and balanced diet when young
Self Reflection	Using “make it your self products”/”products that are co-created with the consumer” is more satisfactiory than obtaining economic benefits.
Self reflection	I would welcome more novelty and change in my life
Self Reflection	I’d like to be famous, even for a short while
Self reflection	Every home should have things that are beautiful even though they may not have any use

APPENDIX 2- Socio economic statue table

Education Level and Occupation Head of Family	Primary School		Secondory	High	University		Master's degree /
	Drop Out	Graduated	Drop Out	Graduated	2Years/ Drop Out	4 ve +	Graduate
Expert Work Owner / No employee	-	-	-	-	B	B	A
Expert Work Owner / with employee	-	-	-	-	B	A	A
Expert Salesperson	-	-	-	-	B	B	B
Work Owner / Production / No employee	C2	C2	C2	C1	C1	B	B
Work Owner / Production / 1-3 employee	C2	C1	C1	C1	B	B	B
Work Owner / Production / 4-9 employee	C2	C1	C1	C1	B	B	B
Work Owner / Production / 10-25 employee	C2	C1	B	B	B	B	B
Work Owner / Production / 25+ employee	C1	C1	B	A	A	A	A
Work Owner / Sales-Service / No employee	C2	C2	C2	C1	C1	B	B
Work Owner / Sales-Service / 1-3 employee	C2	C2	C2	C1	B	B	B
Work Owner / Sales-Service / 4-9 employee	C2	C1	C1	B	B	B	B
Work Owner / Sales-Service / 10-25 employee	C2	C1	B	B	B	B	B
Work Owner / Sales-Service / 25+ employee	C1	C1	B	A	A	A	A
General Manager	-	-	C1	B	A	A	A
Middle Level Manager	C2	C2	C1	B	B	B	B
Low Level Manager	-	C2	C2	C1	C1	C1	B
Salesperson	D	C2	C1	C1	C1	C1	B
Black coated worker	D	C2	C2	C1	C1	C1	C1
Expert/Head Worker	D	C2	C2	C2	C1	-	-
Qualified Worker	D	C2	C2	C2	C1	-	-
Unqualified Worker	D	D	D	D	-	-	-
Unemployed	E	E	E	E	E	E	E
Unearned Income	D	D	C2	C2	C1	B	B

CURRICULUM VITAE

Personal Information

Name Surname	: Datev YAZDANOGLU
Date of Birth	: 01.01.1987
Place of Birth	: Istanbul

Education

1997-2003	High School	Getronagan High School.
2003-2008	Undergraduate	Yildiz Technical University, Statistics.

Professional Life

2008-Still	Data Processing in Barem Research.
2006-2008	Project Manager in Frekans Research
2005-2006	Reserach & Development Department in AVEA