DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

CUSTOMER SEGMENTATION USING A FUZZY AHP AND CLUSTERING BASED APPROACH: AN APPLICATION IN AN INTERNATIONAL TV MANUFACTURING COMPANY

by

Hülya GÜÇDEMİR

June, 2013

İZMİR

CUSTOMER SEGMENTATION USING A FUZZY AHP AND CLUSTERING BASED APPROACH: AN APPLICATION IN AN INTERNATIONAL TV MANUFACTURING COMPANY

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science
in Industrial Engineering, Industrial Engineering Program

by
Hülya GÜÇDEMİR

June, 2013
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"CUSTOMER SEGMENTATION USING A FUZZY AHP AND CLUSTERING BASED APPROACH: AN APPLICATION IN AN INTERNATIONAL TV MANUFACTURING COMPANY"** completed by **HÜLYA GÜÇDEMİR** under supervision of **ASSOC. PROF. HASAN SELİM** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Hasan SELİM

Supervisor

Yrd. Doç. Dr. Burcu Felekoğlu

(Jury Member)

Doç. Dr. Mehmet Aksarayl

(Jury Member)

Prof.Dr. Ayşe OKUR
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

Most of all I would like to thank to my research advisor, Assoc. Prof. Hasan SELİM for his valuable advice, encouragement, cooperation and guidance throughout this thesis. He guided me to make the right decisions by his scientific experience and creativity. I would also give special thanks again to him for his trust, patience and motivation. I am really glad to have worked with him.

I am grateful to all of the personnel of the TV manufacturing company who helped me to gather the data for the application of this study and share experience with me.

In addition, I would like to thank to Assist. Prof. Ceyhun ARAZ and Assist. Prof. Özlem UZUN ARAZ for their support and sharing their knowledge with me.

Very special thanks would be given to my family especially my mother for their endless support, love and patience throughout my whole life.

Hülya GÜÇDEMİR

CUSTOMER SEGMENTATION USING A FUZZY AHP AND CLUSTERING BASED APPROACH: AN APPLICATION IN AN INTERNATIONAL TV MANUFACTURING COMPANY

ABSTRACT

Today, the most valid way to achieve sustainable competitive advantage is shifting the focus from a product oriented view to a customer oriented view. However, due to more complex nature of customer behaviors, management of a customer base has become more difficult. Therefore, both business understanding and customer database analysis become vital. In this concern, customer segmentation plays an important role in marketing strategies and product development. This study aims to divide customer base in an international TV manufacturing company into discrete customer groups that share similar characteristics and also to find relative importance of these groups. Two different approaches are used for this purpose. First approach divides customer base using a characteristic called "*overall score"*. Overall score is a combined score of eight different characteristics namely, "recency", "loyalty", "average annual demand", "average annual sales revenue", "frequency", "long term relationship potential", "average percentage change in annual demand" and "average percentage change in annual sales revenue". This score computed by taking weighted average of the characteristics where weights are obtained by using Fuzzy Analytical Hierarchy Process (AHP). Second approach groups customers according to their similarities with respect to eight characteristics that mentioned above. Agglomerative hierarchical clustering algorithms (Ward's method, single linkage, complete linkage) and k-means algorithm are employed to segment the customers. Five customer segments are named as best, valuable, average, potential valuable and potential invaluable customers. The results reveal that the proposed approach can effectively be used in practice for proper customer segmentation.

Keywords: Customer segmentation, data clustering, fuzzy AHP

BULANIK AHP VE KÜMELEME TABANLI BİR MÜŞTERİ SEGMENTASYONU YAKLAŞIMI: ULUSLARARASI BİR TV İMALAT FİRMASINDA UYGULAMA

ÖZ

Günümüzde, sürdürülebilir rekabet avantajı elde etmek için en geçerli yol ürün odaklı bir anlayış yerine müşteri odaklı bir anlayışı benimsemektir. Ancak, müşteri davranışlarının karmaşık doğası müşteri tabanının yönetimini zorlaştırmaktadır. Bu nedenle hem işletmeyi anlamak hem de müşteri veri tabanlarının analizi önemli hale gelmiştir. Bu anlamda müşteri segmentasyonu, pazarlama stratejileri ve ürün gelişimi konularında önemli bir rol oynamaktadır. Bu çalışmada uluslar arası bir TV üreticisi firmanın müşteri tabanının benzer özellikler gösteren müşteri gruplarına bölünmesi ve aynı zamanda bu grupların göreli önemlerinin bulunması amaçlanmıştır. Bu amaç doğrultusunda iki farklı yaklaşım kullanılmıştır. İlk yaklaşım, müşteri tabanını "*genel skor*" olarak isimlendirilen tek bir karakteristiğe göre bölmektedir. Genel skor ise, "*güncellik", "sadakat", "yıllık ortalama talep", "yıllık ortalama satış geliri", "sıklık", "uzun vadeli ilişki potansiyeli", "yıllık talepteki ortalama değişim", "yıllık satış gelirindeki ortalama değişim"* olarak isimlendirilen sekiz farklı karakteristiğin birleşimidir. Burada, genel skor bulanık Analitik Hiyerarşi Süreci (AHS) kullanılarak ağırlıkları elde edilen karakteristiklerin ağırlıklı ortalaması alınarak hesaplanmaktadır. İkinci yaklaşım ise yukarıda belirtilen sekiz karakteristik açısından benzerliklerine göre müşterileri gruplamaktadır. Müşterileri gruplamada, yığılmalı hiyerarşik kümeleme yöntemleri (tek bağlantı, tam bağlantı, Ward's yöntemi) ve k-ortalamalar yöntemi kullanılmıştır. Oluşturulan beş segment, en iyi, değerli, ortalama, potansiyel değerli, potansiyel değersiz müşteriler olarak isimlendirilmiştir. Sonuçlar, önerilen yaklaşımın müşteri segmentasyonu uygulamalarında etkin bir şekilde kullanılabileceğini ortaya koymuştur.

Anahtar sözcükler: Müşteri segmentasyonu, veri kümeleme, bulanık AHP

CONTENTS

Page

THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZ v
LIST OF FIGURES ix
LIST OF TABLES xi

CHAPTER ONE - INTRODUCTION 1

CHAPTER TWO - CUSTOMER SEGMENTATION 3

2.1 Basic Concepts of Customer Segmentation 3
2.2 Literature Survey 10

CHAPTER THREE - DATA MINING 17

3.1 Definitions and Basic Concepts 17
3.2 Data Mining Process 17
3.2.1 Business Understanding Phase 18
3.2.2 Data Understanding Phase 19
3.2.3 Data Preparation Phase 20
3.2.4 Modeling Phase 21
3.2.5 Evaluation Phase 22
3.2.6 Deployment Phase 23
3.3 Data Mining Models 24
3.3.1 Predictive Models 25
3.3.1.1 Classification 25
3.3.1.2 Regression 25
3.3.2 Descriptive Models 26
3.3.2.1 Clustering 26

3.3.2.2 Association 27
3.4 Data Clustering 27
3.4.1 Data Clustering Steps 28
3.4.1.1 Pattern Representation 28
3.4.1.2 Pattern Proximity 29
3.4.1.2.1 Euclidean Distance 29
3.4.1.2.2 Squared Euclidean Distance 29
3.4.1.2.3 Pearson Distance 29
3.4.1.2.4 Manhattan (City-Block) Distance 30
3.4.1.2.5 Minkowski Distance 30
3.4.1.3 Grouping 30
3.4.1.3.1 Hierarchical Clustering Algorithms 31
3.4.1.3.1.1 Agglomerative Hierarchical Clustering Algorithms 32
3.4.1.3.1.1.1 Single Linkage Algorithm 32
3.4.1.3.1.1.2 Complete Linkage Algorithm 35
3.4.1.3.1.1.3 Average Linkage Algorithm 37
3.4.1.3.1.1.4 Ward's Algorithm 38
3.4.1.3.1.2 Divisive Hierarchical Clustering Algorithms 38
3.4.1.3.2 Partitional Clustering Algorithms 38
3.4.1.3.2.1 k-means Algorithm 40
3.4.1.4 Data Abstraction 41
3.4.1.5 Assessment of Output 41
3.4.1.5.1 Dunn Index 41
3.4.1.5.2 Davies-Bouldin Index 42
3.4.1.5.3 Silhouette Index 42
3.4.1.5.4 Sum of Squares 43
3.4.1.5.5 C Index 43
3.4.1.5.6 Calinski-Harabasz Index 44

CAHPTER FOUR - CUSTOMER SEGMENTATION: AN APPLICATION IN AN INTERNATIONAL TV MANUFACTURING COMPANY 45

4.1 Problem Definition and Business Understanding 45
4.2 Proposed Customer Segmentation Approach 46
4.3 Order-Selling Process of the Company 49
4.4 Data Understanding and Preparation 51
4.5 Determining Customer Evaluation Characteristics 52
4.5.1 Recency 53
4.5.2 Loyalty 54
4.5.3 Average Annual Demand 55
4.5.4 Average Annual Sales Revenue 55
4.5.5 Frequency 56
4.5.6 Long Term Relationship Potential 57
4.5.7 Average Percentage Change in Annual Demand 58
4.5.8 Average Percentage Change in Annual Sales Revenue 59
4.6 Determining Importance Weights of the Characteristics Using Fuzzy AHP .. 61
4.7 Segmenting Customers via Data Clustering Algorithms 68
4.7.1 Single Dimension (SD) – Based Customer Segmentation 68
4.7.1.1 SD – Based Customer Segmentation using AHC Algorithms……70
4.7.1.2 SD – Based Customer Segmentation using k-means Algorithm. 75
4.7.2 Multiple Dimensions (MD) – Based Customer Segmentation 81
4.7.2.1 MD – Based Customer Segmentation using AHC Algorithms... 81
4.7.2.2 MD – Based Customer Segmentation using k-means Algorithm 86
4.8 Evaluation of the Results 90
4.9 Final Customer Segments 93

CHAPTER FIVE - CONCLUSIONS 97

REFERENCES 101

APPENDIX A **-** THE RANK OF SEGMENTS TO WHICH CUSTOMERS WERE ASSIGNED 110
APPENDIX B **-** FUZZY EVALUATION MATRIX 116

LIST OF FIGURES

Page

Figure 2.1 Customer focused marketing 3
Figure 2.2 CRM management cycle 5
Figure 2.3 Analytical CRM 6
Figure 2.4 Analytical CRM tasks and tools 7
Figure 2.5 Classification framework on data mining techniques in CRM 7
Figure 2.6 Types of customer data 8
Figure 2.7 Common customer segmentation approaches 9
Figure 3.1 Interdisciplinary nature of data mining 17
Figure 3.2 CRISP data mining process 18
Figure 3.3 Business understanding phase 19
Figure 3.4 Data understanding phase 20
Figure 3.5 Data preparation phase 21
Figure 3.6 Modeling phase 22
Figure 3.7 Evaluation phase 23
Figure 3.8 Deployment phase 24
Figure 3.9 Classification of data mining models 25
Figure 3.10 Clustering example 28
Figure 3.11 Clustering methods 31
Figure 3.12 Agglomerative and divisive clustering 32
Figure 3.13 Single linkage agglomerative clustering on the sample data set 33
Figure 3.14 Iterative procedure of single linkage algorithm 34
Figure 3.15 Complete linkage agglomerative clustering on the sample data set 35
Figure 3.16 Iterative procedure of complete linkage algorithm 36
Figure 3.17 Iterative procedure of average linkage algorithm 37
Figure 3.18 Iterative procedure of k-means algorithm 41
Figure 4.1 The proposed approach 48
Figure 4.2 Order-selling process 50
Figure 4.3 Number of customers over the years 52
Figure 4.4 A triangular fuzzy number 62
Figure 4.5 Membership functions for linguistic variables 63

Figure 4.6 Importance weights of the characteristics 68
Figure 4.7 SSW values of AHC algorithms (SD) 71
Figure 4.8 Cluster centroids (Ward's - SD) 72
Figure 4.9 Scatter plot of groups (Ward's - SD) 74
Figure 4.10 SSW values of k-means (SD) 76
Figure 4.11 Cluster centroids (k-means - SD) 77
Figure 4.12 Determinant (W) over the iterations (k-means - SD) 77
Figure 4.13 Scatter plot of groups (k-means - SD) 80
Figure 4.14 Group averages in terms of each characteristic (k-means – SD) 80
Figure 4.15 SSW values of AHC algorithms (MD) 82
Figure 4.16 Profile plot of clusters (Ward's - MD) 85
Figure 4.17 SSW values of k-means (MD) 86
Figure 4.18 Determinant (W) over the iterations (k-means - MD) 88
Figure 4.19 Profile plot of clusters (k-means - MD) 90

LIST OF TABLES

Page

Table 2.1 Literature review on customer segmentation problem 16
Table 3.1 Proximity matrix of the sample data set 33
Table 4.1Order transaction data 51
Table 4.2 Geographical dispersion of customer portfolio 52
Table 4.3 Recency values 54
Table 4.4 Loyalty values 55
Table 4.5 AAD values 55
Table 4.6 AASR values 56
Table 4.7 Order schedule of C165 57
Table 4.8 Order schedule of C299 57
Table 4.9 LTRP values 58
Table 4.10 Demand values 59
Table 4.11 Sales revenues 60
Table 4.12 Initial data set 61
Table 4.13 Standardized data 61
Table 4.14 Linguistic variables for the importance of customer evaluation characteristics 63
Table 4.15 Notations for the characteristics 65
Table 4.16 The pair-wise comparison matrix 65
Table 4.17 Importance weights of the characteristics 67
Table 4.18 Overall scores of the customers 69
Table 4.19 Sum of squares of AHC algorithms (SD) 71
Table 4.20 Summary statistics of the dataset (SD) 71
Table 4.21 Cluster centroids and ranks (Ward's - SD) 72
Table 4.22 Distances between cluster centroids (Ward's - SD) 73
Table 4.23 Central objects (Ward's - SD) 73
Table 4.24 Distances between central objects (Ward's - SD) 73
Table 4.25 Results of Ward's AHC for five clusters (SD) 74
Table 4.26 Characteristics of the groups (Ward's - SD) 75
Table 4.27 Sum of squares of k-means algorithm (SD) 75

Table 4.28 Initial cluster centroids (k-means - SD) 76
Table 4.29 Final cluster centroids and ranks (k-means - SD) 76
Table 4.30 Statistics for the iterations (k-means - SD) 77
Table 4.31 Optimization summary for k-means (SD) 78
Table 4.32 Results of k-means for five clusters (SD) 78
Table 4.33 Distances between the cluster centroids (k-means - SD) 79
Table 4.34 Central objects (k-means - SD) 79
Table 4.35 Distances between the central objects (k-means - SD) 79
Table 4.36 Characteristics of groups (k-means - SD) 81
Table 4.37 Sum of squares of AHC algorithms (MD) 82
Table 4.38 Summary statistics of the characteristics (MD) 83
Table 4.39 Cluster centroids (Ward's - MD) 83
Table 4.40 Distances between centroids (Ward's - MD) 83
Table 4.41 Central objects (Ward's - MD) 84
Table 4.42 Distances between central objects (Ward's - MD) 84
Table 4.43 Results of Ward's AHC for five clusters (MD) 84
Table 4.44 Rank of clusters (Ward's - MD) 86
Table 4.45 Sum of squares for k-means (MD) 86
Table 4.46 Statistics for the iterations (k-means - MD) 87
Table 4.47 Optimization summary for k-means (MD) 87
Table 4.48 Initial cluster centroids (k-means - MD) 87
Table 4.49 Cluster centroids (k-means - MD) 88
Table 4.50 Distances between cluster centroids (k-means - MD) 88
Table 4.51 Central objects (k-means - MD) 89
Table 4.52 Distances between central objects (k-means - MD) 89
Table 4.53 Results of k-means for five clusters (MD) 89
Table 4.54 Rank of clusters (k-means - MD) 90
Table 4.55 The results of paired-t tests 92
Table 4.56 Number of customers assigned to the segments 92
Table 4.57 Comparison of the methods in terms of assignment similarities 93
Table 4.58 SSW/TSS ratio of different approaches for five clusters 93
Table 4.59 Best customers 95

Table 4.60 Valuable customers 95
Table 4.61 Average customers 95
Table 4.62 Potential valuable customers 95
Table 4.63 Potential invaluable customers 96

CHAPTER ONE
INTRODUCTION

Globalization and increased competition have changed customer buying behaviors and expectations dramatically. Managing changing customer behaviors according to business goals and objectives necessitates understanding and interpreting the customer relationship management (CRM) concepts correctly. In today's markets, creating value for customers and increasing loyalty are vital for companies for their survival. With a successful CRM, companies can understand their customers, their needs, expectations and they can quickly adapt to changing conditions.

With the help of innovations in computer technology, companies can keep large amounts of data about their customers and they can process and convert the data into meaningful information for their business decisions. Today, companies can keep many details extending from demographic characteristics to buying behaviors about customers on their databases. It is possible to extract hidden patterns, associations and relationships from these large databases using data mining techniques. Data mining helps the companies on issues that classify and identify the customers, and predict their behaviors. In addition, data mining provide strategic information for many customer-centric applications.

One of the most common application areas of data mining in CRM is customer segmentation. Customer segmentation is the division of the market into small groups of customers with similar characteristics. It groups customers based on different aspects such as their geographic, behavioral and demographic characteristics. It allows an organization to understand which customers are most valuable and also helps companies to manage their large customer base. As a data mining technique, data clustering can be employed for customer segmentation. Data clustering algorithms group customers based on their predefined characteristics. Then, companies can develop sales and marketing activities for their customer groups.

This study was carried out in an international TV manufacturing company and aims to divide the customers into manageable groups using clustering algorithms and also to find relative importance of these groups using multi criteria decision making technique.

This study is organized as follows. Chapter 1 presents brief descriptions of CRM, data mining and customer segmentation. Chapter 2 introduces customer segmentation and presents the survey of the related studies. Data mining techniques are presented in Chapter 3, while the proposed customer segmentation approach is explained in Chapter 4. Finally, Chapter 5 concludes the study.

CHAPTER TWO
CUSTOMER SEGMENTATION

2.1 Basic Concepts of Customer Segmentation

Customers are the most important assets of an organization and they differ from each other on issues such as buying behaviors, geography, education, expectations, preferences, profitability, loyalty etc. In today's competitive market, there is an extensive diversification of both products and services. The most valid way to achieve sustainable competitive advantage is shifting the focus from a product oriented view to a customer oriented view. However, companies have limited resources to serve their customers. Therefore, companies should use their limited resources in an effective manner by selecting the valuable customers and making efforts to keep them.

CRM is one of the most important topics in marketing. CRM has various definitions depending on different perspectives. For instance, Parvatiyar & Sheth (2001) defined CRM as ''a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and the supply chain functions of the organization to achieve greater efficiencies and effectiveness in delivering customer value". Moreover, CRM suggests that organizational thinking must be changed from the current focus on products to include both customers and products, as illustrated in Figure 2.1 (Srivastava et al., 2002).

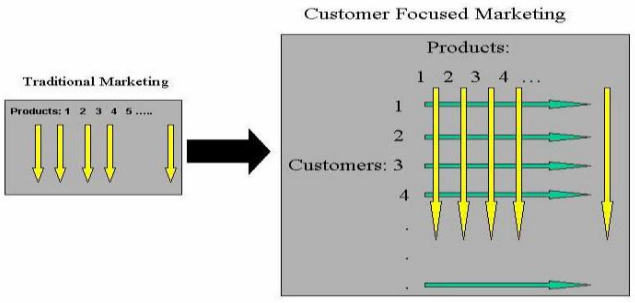

Figure 2.1 Customer focused marketing (Srivastava et al., 2002)

CRM is a comprehensive process of acquiring and retaining customers, understanding and satisfying their needs with the help of business intelligence to maximize the customer value and loyalty to the organization furthermore to gain sustainable competitive advantage. The most notable benefits of CRM are the following (Bergeron, 2002):

- Improved customer satisfaction levels
- Increased customer retention and loyalty
- Improved customer lifetime value
- Transfer of better strategic information to relevant departments
- Attraction of new customers
- Lower costs
- Customization of products and services
- Improving and extending customer relationships, generating new business opportunities
- Knowing how to segment customers, differentiating profitable customers from those who are not, and establishing appropriate business plans for each case
- Increasing the effectiveness of providing customer service by having complete, homogeneous information
- Sales and marketing information about customer requirements, expectations and perceptions in real time
- Improvement in the quality of business processes
- Competitive advantage
- Increase in customer demands

According to Swift (2001), Parvatiyar & Sheth (2001), Kracklauer et al. (2004), and Ngai et al. (2009), CRM consists of four dimensions: *customer identification, customer attraction, customer retention* and *customer development.* These dimensions can be considered as a closed cycle of a customer management system as illustrated in Figure 2.2 (Kracklauer et al., 2004; Ling & Yen, 2001).

All these dimensions share the common goal of deeper understanding of customers to maximize customer value to the organization in the long term.

Figure 2.2 CRM management cycle (Kraclauer et al., 2004)

CRM can be evaluated in two categories such as operational and analytical. Operational CRM comprises the business processes and technologies that can help improve the efficiency and accuracy of day-today customer-facing operations. This includes sales, marketing, and service automation (Iriana & Buttle, 2006).

The general objective of operational CRM is to improve the efficiency and effectiveness of customer management processes, by personalizing the relationship with customers, by improving organizational response to customers' needs (Xu & Walton, 2005) and by increasing the speed and quality of information flows in the organization, and between the organization and its external employees and partners (Speier & Venkatesh, 2002).

In the past, companies focused on operational tools, but this tendency seems to be changing (Reynolds, 2002). Decision-makers have realized that analytical tools are necessary to drive strategy and tactical decisions, related to customer identification, attraction, retention and development (Oliveira, 2012). Buttle (2004) defined analytical CRM as "a bottom-up perspective, which focuses on the intelligent mining of customer data for strategic or tactical purposes."

Analytical CRM mainly focuses on analyzing the data collected and stored, in order to make more meaningful and profitable business decisions (see Figure 2.3). This includes the underlying data warehouse architecture, reporting, and analysis (Iriana & Buttle, 2006). It is also consistent suite of analytical applications that help measure, predict, and optimize customer relationships (SAP, 2001).

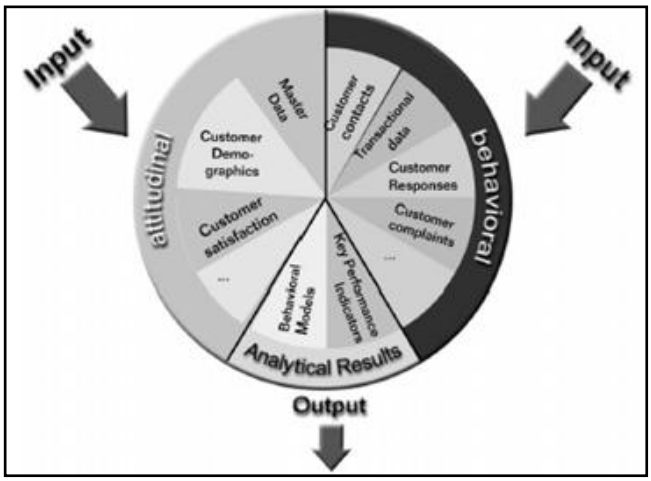

Figure 2.3 Analytical CRM (SAP, 2001)

Herschel (2002) identified several applications within analytical CRM, including customer segmentation analysis, customer profitability analysis, "what if" analysis, real-time event monitoring and triggering, campaign management, and personalization. Doyle (2002) also suggested other analytical tools such as, analysis of the characteristics and behavior of customers, modeling to predict customer behavior, communications management with customers, personalized communications with customers, interactive management and optimization to determine the best combination of customers, products, and communication channels. Figure 2.4 represents the dimensions of CRM and the tactical tools for achieving the core tasks.

Figure 2.4 Analytical CRM tasks and tools (Kraclauer et al., 2004)

Data mining plays a critical role in the analytical CRM applications. There exist various data mining techniques that are used in CRM applications such as decision trees, neural networks, genetic algorithms, clustering, classification and regression trees, logistic regression, association rules. The reader may refer to Ngai et al. (2009) for a review of the studies on use of data mining techniques in CRM. Figure 2.5 illustrates the classification framework that depicts the relationship between data mining techniques and analytical CRM.

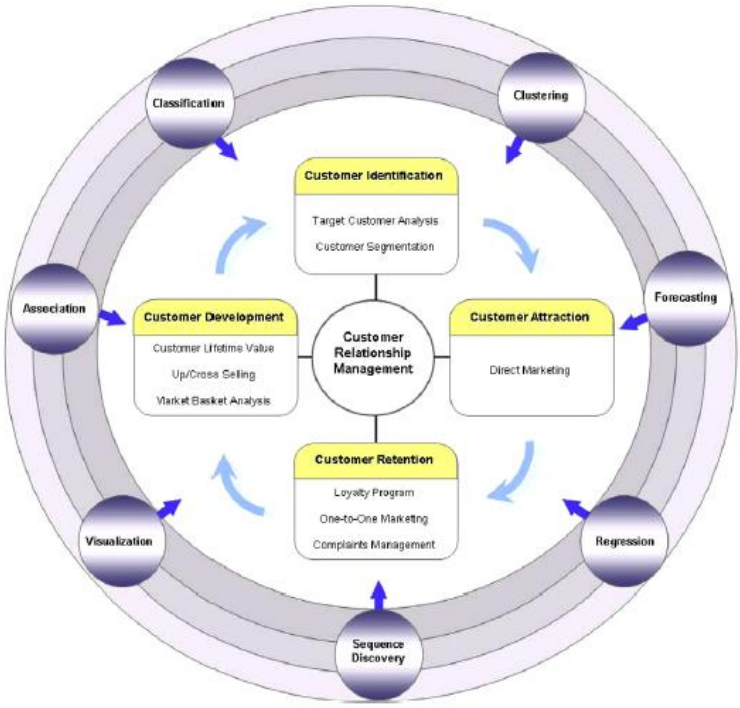

Figure 2.5 Classification framework on data mining techniques in CRM (Ngai et al., 2009)

Customer segmentation is one of the core functions of analytical CRM and it can be defined as dividing market into customer groups that share similar characteristics (Chen et al., 2006). The goal of customer segmentation is division of market into customer groups in accordance with their value for the company (Dannenberg & Zupancic, 2009). Segmentation allows companies to understand which customers are most profitable, how to develop marketing campaigns and pricing strategies to the customer segments and provide more personalized, more attractive product and service offerings to individual customer groups (Xu & Walton, 2005). A company can use customer segmentation for general understanding of a market, product positioning studies, new product concepts, pricing decisions, advertising decisions and distribution decisions (Wind, 1987).

Customer information helps the organization to understand customer behavior better, to conduct the right transaction at the right time, and to be able to segment its market effectively (Plakoyiannaki & Tzokas, 2002; Xu & Walton, 2005). So, the key enabler of any segmentation strategy is customer data (Kelly, 2003). Customer segmentation begins with depth analysis of customer data base that includes characteristics of a specific customer including customer demographics, purchasing behavior, channel preferences, profitability, loyalty, past and expected future spending, satisfaction etc. Figure 2.6 shows commonly used customer characteristics in segmentation studies.

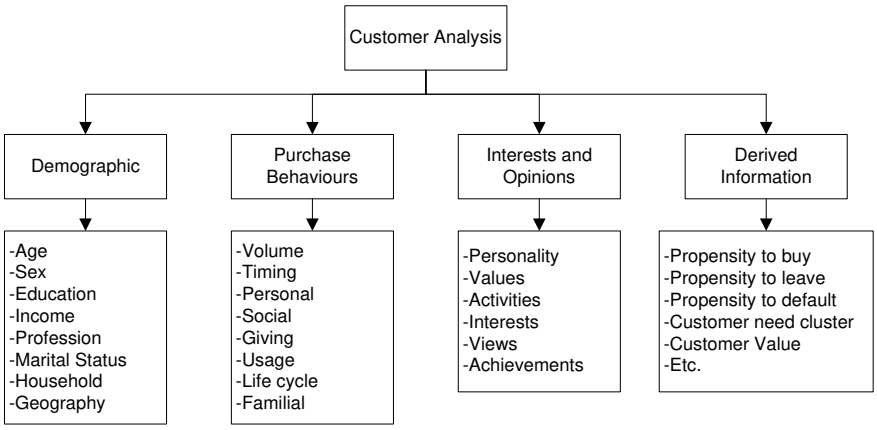

Figure 2.6 Types of customer data (Meltzer, 2005)

After defining customer characteristics, data mining techniques that extract or detect hidden customer characteristics and behaviors from large databases can be used for the customer segmentation (Carrier & Povel, 2003). Data clustering is a powerful data mining technique for customer segmentation (Punj & Stewart, 1983; Pham & Afify, 2007). The logic behind cluster analysis includes analyzing customer data and dividing the customers into smaller manageable groups according to the similarities between them with respect to predefined characteristics. On the other hand, customers can be segmented based on different perspectives. Figure 2.7 presents the common approaches to customer segmentation.

Segmentation type	Segment definition
Buyer-readiness segmentation	The division of prospects and customers into groups reflecting the different stages which consumers normally pass through during the purchase process. These usually comprise ignorance, awareness, knowledge, preference and conviction.
Benefits segmentation	Dividing the market into groups according to the different benefits that consumers seek from the product.
Behaviour segmentation	The division of customers into groups based on attitude, usage or response to a product or promotion.
Occasion segmentation	The division of customers into groups which consume a product or service at particular times, in certain situations, in response to particular events or according to seasonal or cyclical times.
Psychographic/lifestyle segmentation	The division of customers into groups based on lifestyle, social behaviour, values, sensitivities and personality characteristics.
Demographic segmentation	The division of customers into different groups based on demographic variables such as age, gender, family size, income, occupation, education, language, religion, race and nationality.
Life-cycle segmentation	The division of customers into different groups that recognise the different needs of consumers at different stages in their life.
Geographic segmentation	The division of customers into different groups based on countries, regions, climate and population density.
Loyalty segmentation	The division of customers into different groups based on different degrees of loyalty to supplier or brand.
Product segmentation	The division of customers into different groups based on levels and type of usage of the product or service.
Profitability segmentation	The division of customers into different groups based on the different levels of value or profitability of the customers.
Interaction segmentation	The division of customers into different groups based on their preferences regarding channels, payment method, promotions and communications.
Satisfaction segmentation	The division of customers into different groups based on their recorded satisfaction levels, complaint history, fault history and upgrade history.

Figure 2.7 Common customer segmentation approaches (Kelly, 2003)

2.2 Literature Survey

In the literature, many studies handle customer segmentation problem in variety of sectors and various methods are proposed for this purpose. Wind (1987), provided a detailed research on customer segmentation approaches in his study. In this section, methodologies and studies about the customer segmentation problem will briefly be introduced.

The most common segmentation approach is grouping customers based on customer lifetime value (CLV) or the components of the recency-frequency-monetary (RFM) model.

CLV represents the economic value of a customer to the firm and defined as the "net present value of the profit streams a customer generates over the average customer lifetime" (Reichheld & Sasser, 1990). It is computed by the following formula;

$$CLV = -AC + \sum_{t=0}^{n} \frac{C_t}{(1+d)^t} \tag{2.1}$$

where,
t=time index
n=lifetime of the customer
AC=acquisition cost
C_t=contribution margin at time t (revenues-cost)
d=discount rate

The RFM model, which is proposed by Hughes in 1994, is very effective for customer segmentation (Newell, 1997). RFM model is the translation of customer behavior into numbers and it distinguishes important customers from large data bases by three attributes. On the other hand, definition and computation of these attributes can change depending on the problem (Miglautsch, 2000).

For instance, Buckinx & Poel (2005) described recency as the number of days that passed between the last transaction and the end of observation period in their study. They defined monetary value as the total amount of spending that a customer made during its lifetime. In addition, Hosseini et al. (2010) described frequency as the total number of purchases that customer made in a particular period. The reader may refer to Wei et al. (2010) for the details on the application of RFM model.

There are many researchers who use RFM model in their segmentation studies. Chan (2008), performed a segmentation study for the customers of Nissan automobile retailer. He used generic algorithm (GA) to segment customers based on RFM model. Customer LTV was taken as the fitness value of GA and customers were segmented into eight groups. Additionally, correlation between customer values and campaign strategies was considered in the study. The results of the study reveal that the proposed approach can increase potential value, customer loyalty and customer lifetime value.

Chiu et al. (2009) used RFM variables and proposed a conventional statistic analysis and intelligent clustering methods (artificial neural network and particle swarm optimization) integrated decision support system for the market segmentation.

Cheng & Chen (2009) joined the quantitative value of RFM attributes and k-means algorithm into rough set (RS) theory in their study. They segmented customers of a company that operates in Taiwan's electronic industry. They used RFM model and portioned 401 customers into 3, 5 and 7 clusters. Then, decision rules were generated by RS Learning from Examples Module, version 2 (LEM2) method. The number of segments was defined based on subjective view of top management for the company.

Dhandayudam & Krishnamurthi (2012) suggested a clustering algorithm to overcome the difficulties of traditional clustering algorithms. They used R, F, M attributes and clustered customers of a fertilizer manufacturing company into two, three and four clusters.

Then, they compared the performance of this improved algorithm against single link, complete link and k-means algorithms using mean square error, intra cluster distance, inter cluster distance and intra/inter cluster distance ratio indicators. As a result, their algorithm produced better results than other clustering algorithms.

There are also other researchers who propose to extend the standard RFM model by including additional variables into analysis. In 2005, Buckinx & Poel classified customers of a fast moving consumer goods (FMCG) retailer. Logistic regression, automatic relevance determination, neural networks and random forests were used to predict partial defection. They used additional variables to RFM such as "the length of customer relationship", "mode of payment", "buying behavior across categories", "usage of promotions" and "brand purchase behavior". Classification accuracy and area under the receiver operating characteristic curve are used to evaluate the classifier performance.

Li et al. (2011) extended traditional RFM model by adding variable relation length and segmented customers of a textile manufacturing business using two step clustering method (Ward with k-means).

In many applications companies weight the R, F, M scores in favor of importance of the attributes (Reinartz & Kumar, 2002). Liu & Shih (2005) combined group decision-making and data mining techniques in their study. Customers were segmented based on RFM variables. The analytic hierarchy process (AHP) was applied to determine the relative weights of RFM variables in evaluating customer lifetime value. They used k-means method and clustered customers into eight groups according to the weighted RFM values. Finally, an association rule mining approach was implemented to provide product recommendations to each customer group. The results of the study reveal that their methodology is more effective for more loyal customers.

Hosseini et al. (2010) segmented customers of SAPCO Co., one of the leading car manufacturing supplying companies in Iran, based on expanded RFM model by including additional loyalty parameter. They used k-means algorithm and portioned customers into 34 clusters according to Davies-Bouldin index. They used both weighted and unweighted parameters to compute the values of clusters. Also they assessed customer loyalty using decision trees and artificial neural network methods.

On the other hand problem specific variables can be used instead of RFM attributes. Kim et al. (2006) carried out a segmentation study in a wireless telecommunication company. The researchers evaluated the customer value from three viewpoints and displayed customers of a wireless telecommunication company with 3D space with axes denoting current value, potential value and customer loyalty and segmented customers. Lifetime value (LTV) model was used for the analysis, and they analyzed characteristics of each segment and built strategies for them.

Another segmentation study was carried out for the airline passengers by Teichert et al. in 2008. Data were collected choice-based conjoint survey that consists of seven attributes (flight schedule, total fare, flexibility, frequent-flyer program, punctuality, catering and ground services). They used class flown (economy/business) as a priori segmentation criterion and estimated separate logit models for the business and economy-class segments. Then they built two sub-groups within the business and economy segments according to the a priori criterion of travel reason such as business reason and leisure reason. Furthermore, they applied latent class modeling and segmented airline passengers into five segments based on behavioral and socio-demographic variables.

Ahn & Sohn (2009) carried out a study in order to identify customer groups and provide suitable after sales services to these groups. 376 customers were divided into three groups by using fuzzy c-means clustering according to indicators of customer satisfaction index. Furthermore, they used association rules to find out which after sales operations are important for each customer group.

Wu & Chou (2011) segmented online consumers of an electronic commerce market. Data was obtained from an online questionnaire. The questions were organized into four categories such as “satisfaction with service”, “shopping behavior”, “internet usage” and “demographics”. They developed a soft clustering method that uses a latent mixed class membership clustering approach to group customers based on their purchasing data. 2329 customers were portioned into five segments then these segments were portioned into nine micro segments.

Hosseni &Tarokh (2011) implemented a case study on the insurance database. They segmented customers based on their current value and churn rate. Logistic regression is used to predict the churn probability of a specific customer. They classified churn probability and current value as “high” and “low”. Then, four segments were composed as “high current value-high churn rate”, “high current value-low churn rate”, “low current value-high churn rate” and “low current value-low churn rate”. Moreover, cross/up-selling strategies were proposed for the segments.

Rajagopal (2011) clustered customers of a retail store into four clusters as “high value”, “medium value”, “low value” and “negative value” using IBM Intelligent Miner tool. Recency, total customer profit, total customer revenue and top revenue department parameters were used for the clustering. Additionally, clusters were profiled for the assessment of the potential business value of each cluster and some possible marketing strategies were proposed for the clusters.

Montinaro & Sciascia (2011) aimed to define new types of customer loyalty by using market segmentation strategies and customer satisfaction in their study. They measured customer satisfaction combining two items as satisfaction of purchase and satisfaction of brand of cellular phone purchased. Respondents were divided into three clusters using k-means algorithm based on age and school-leaving examination mark variables. Customer loyalty was calculated as a function of market segmentation and customer satisfaction for each cluster.

Genetic algorithm based k-means clustering algorithm is proposed by Ho et al. (2012), to segment customers of a window curtain manufacturer using volume, revenue and profit margin per order attributes.

Gilboa (2009) segmented Israeli mall customers in his study. Data was obtained from a questionnaire that consists of four main categories such as motivation for mall visits, activities performed during the visit, visiting patterns and personal details. 636 mall customers attended the questionnaire and then two step cluster analysis (Ward with k-means) was performed. Customers were divided into four clusters such as disloyal, family bonders, minimalists and mall enthusiasts.

Tarokh & Sekhavat (2006) segmented the customers of mental health clinic of the University of Tehran. “Customer loyalty”, “current value” and “expected future value” variables were used for the segmentation. Customer future value and churn rate were computed by using logistic regression models. Three customer segments were defined according to the 3D diagram and different marketing strategies were suggested for these segments.

Bayer (2010) considered four different segmentation schemes for the telecommunications industry as customer value segmentation, customer behavior segmentation, customer life cycle segmentation and customer migration segmentation.

Table 1 presents the techniques and variables adopted by the articles considered in the literature review. In many segmentation studies end users are grouped but, in this study customers of a contract manufacturer are evaluated from the perspective of the manufacturer and customer base is divided into groups using data clustering algorithms.

Since each of the customers is a company, customer evaluation characteristics are determined according to this and traditional RFM model is extended by including additional characteristics.

Because of the long inter-purchase time, computations were made based on an annual basis. Furthermore, weights are assigned to those characteristics using Fuzzy AHP method that is summarized in section 4.6 in order to provide a more realistic structure.

Cluster analysis achieves only the grouping of similar observations in the same cluster. So, additionally relative importance of the clusters was found in this study. In this way, we can see which segment is more important / valuable for the company.

Table 2.1 Literature review on customer segmentation problem

	RFM	Other Variables	Weight	Data Clustering	Other Techniques
Buckinx & Poel (2005)	X	X			X
Liu & Shih (2005)	X		X	X	
Kim et al. (2006)		X			X
Tarokh & Sekhavat (2006)		X			X
Teichert et al. (2008)		X			X
Chan (2008)	X				X
Ahn & Sohn (2009)		X		X	
Chiu et al. (2009)	X			X	X
Gilboa (2009)		X		X	
Cheng & Chen (2009)	X			X	
Hosseini et al. (2010)	X	X	X	X	
Bayer (2010)		X			X
Wu & Chou (2011)		X		X	
Hosseni &Tarokh (2011)		X			X
Rajagopal (2011)		X		X	
Montinaro & Sciascia (2011)		X		X	
Li et al. (2011)	X	X		X	
Dhandayudam & Krishnamurthi (2012)	X			X	
Ho et al. (2012)		X		X	X
This study	X	X	X	X	

CHAPTER THREE
DATA MINING

3.1 Definitions and Basic Concepts

"Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques" (Larose, 2005). Data mining is an interdisciplinary domain that gets together artificial intelligence, database management, machine learning, data visualization, fuzzy logic, mathematical algorithms, and statistics (see Figure 3.1).

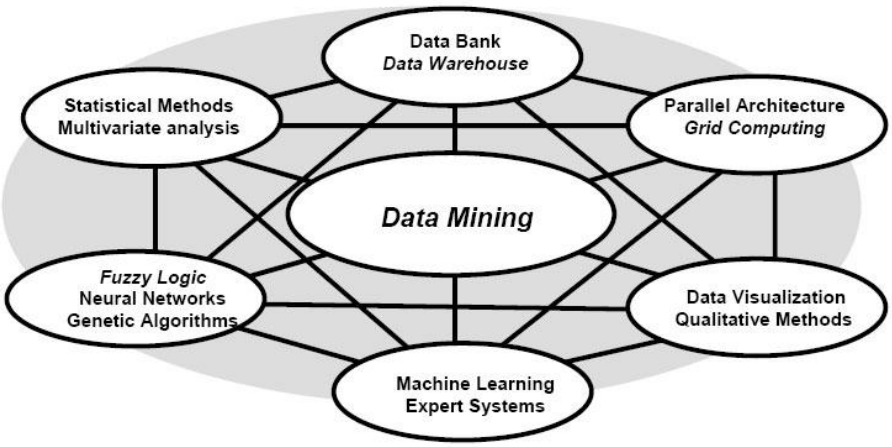

Figure 3.1 Interdisciplinary nature of data mining (Pereira et al., 2008)

3.2 Data Mining Process

The Cross Industry Standard Process for Data Mining (CRISP–DM) (Chapman et al., 2000) was developed in 1996. CRISP considers the data mining process as the general problem solving strategy. According to CRISP–DM, a given data mining project has a life cycle consisting of six phases, as illustrated in Figure 3.2.

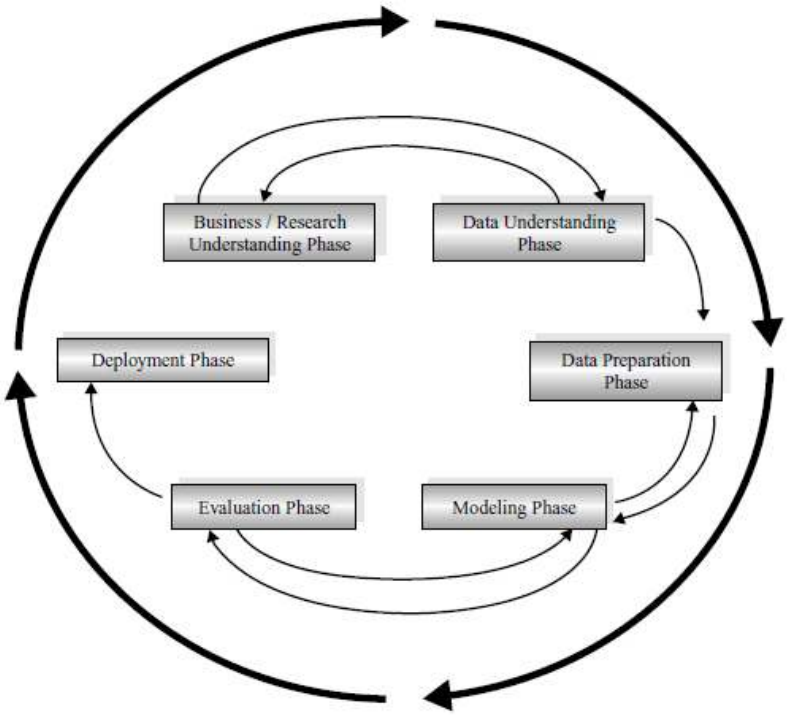

Figure 3.2 CRISP data mining process (Larose, 2005)

3.2.1 Business Understanding Phase

Business understanding is the first phase in CRISP-DM. This phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives. Tasks of this phase are summarized in Figure 3.3.

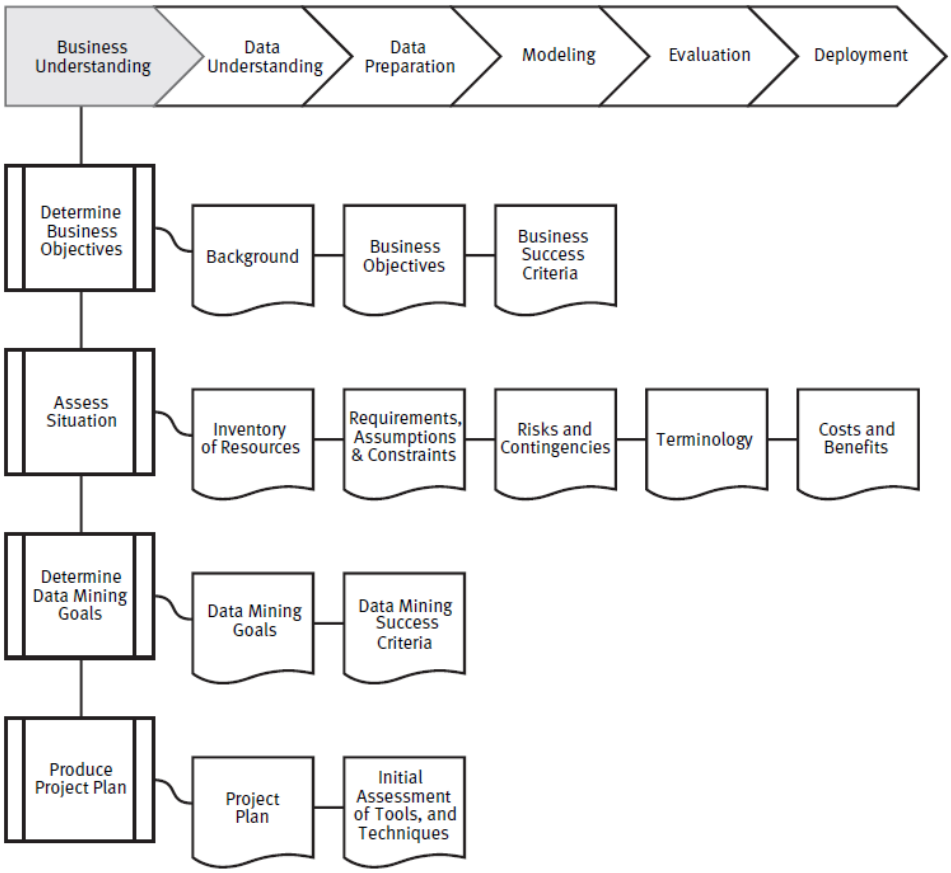

Figure 3.3 Business understanding phase (Chapman et al., 2000)

3.2.2 Data Understanding Phase

Data understanding phase starts by collecting data, then get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses about hidden information (Jackson, 2002). This phase is illustrated in Figure 3.4.

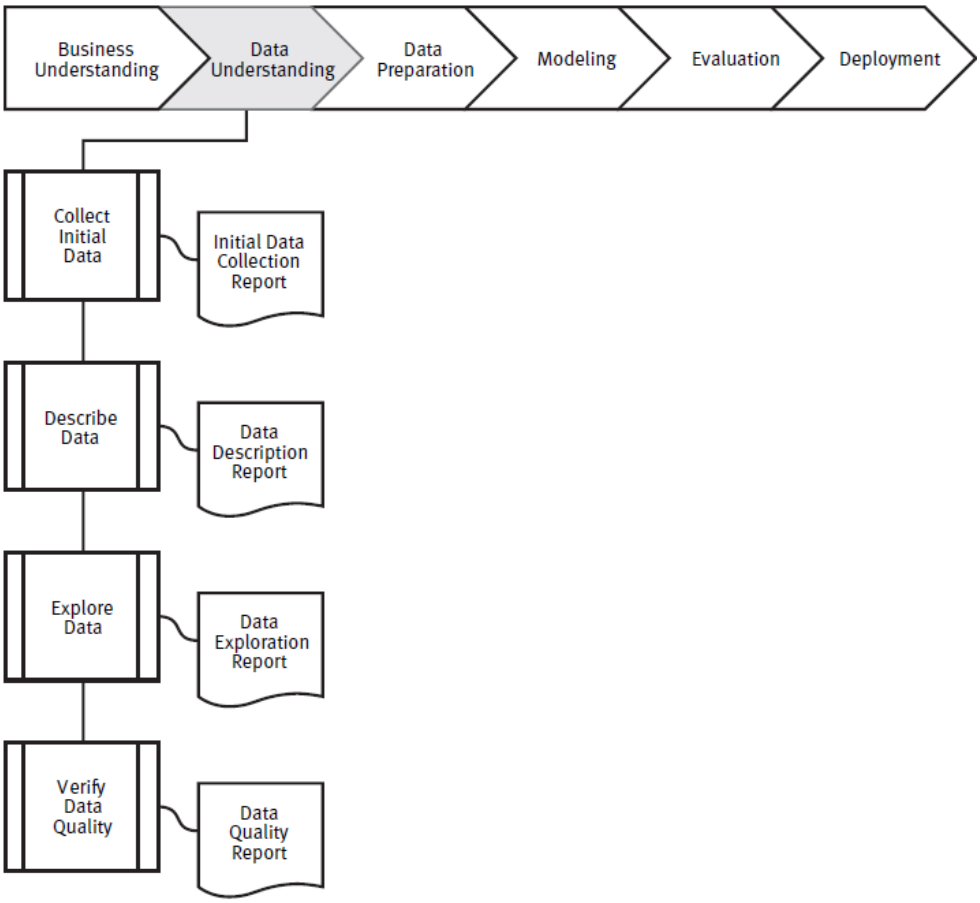

Figure 3.4 Data understanding phase (Chapman et al., 2000)

3.2.3 Data Preparation Phase

This phase includes all activities required to construct the final data set (data that will be fed into the modeling tool) from the initial raw data. As illustrated in Figure 3.5, data cleaning, data transformations, case and attribute selection, data reduction are the main tasks of this phase.

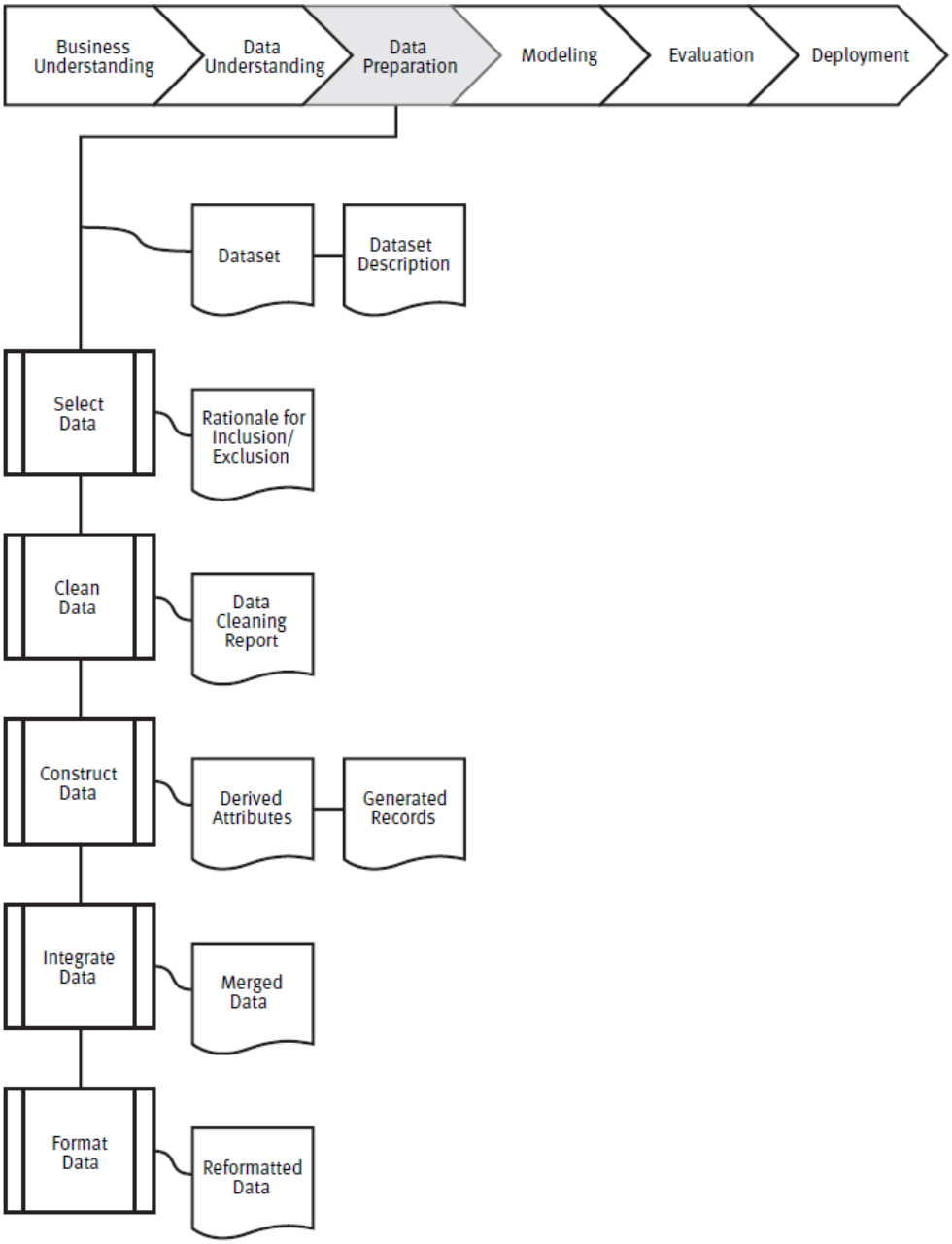

Figure 3.5 Data preparation phase (Chapman et al., 2000)

3.2.4 Modeling Phase

In this phase, appropriate modeling techniques are selected and applied for the predefined problem (see Figure 3.6). There may exist several techniques for the same data mining problem. Therefore, various modeling techniques are established and model settings are calibrated to optimize the results. Optimal values are obtained by comparing the results of modeling techniques.

A well-established model will also affect the quality of the results. Therefore, data preparation and modeling phases repeat until the model is considered to be the best.

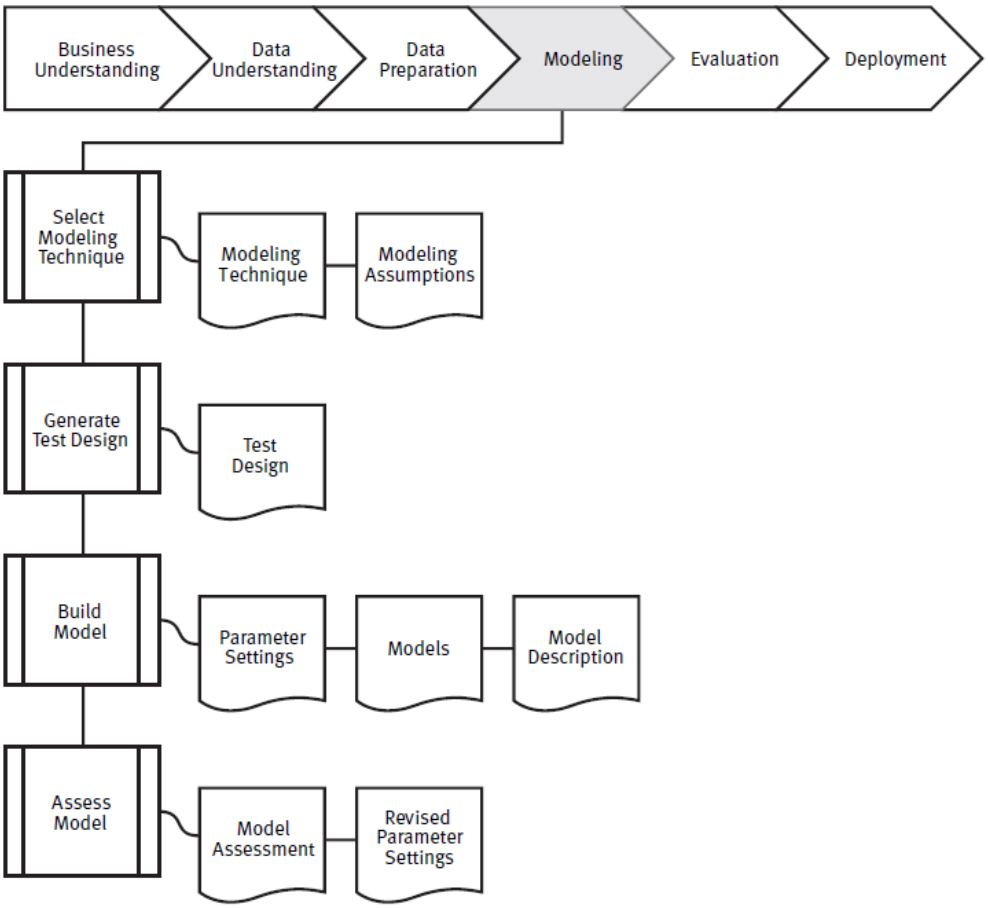

Figure 3.6 Modeling phase (Chapman et al., 2000)

3.2.5 Evaluation Phase

In this phase, the model is evaluated in order to be certain it properly achieves the business objectives. Analysts determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results is reached. Main tasks of evaluation phase are presented in Figure 3.7.

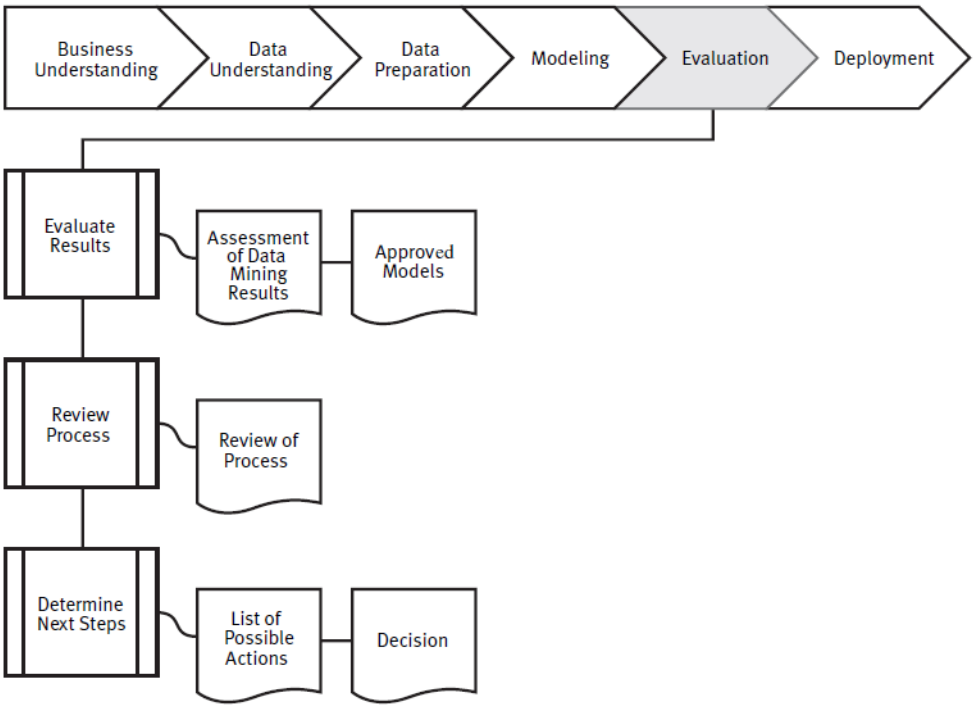

Figure 3.7 Evaluation phase (Chapman et al., 2000)

3.2.6 Deployment Phase

In general, model creation and evaluation is not enough. Data mining results should be organized and presented in a way that a company can easily understand. Deployment can be as simple as generating a report or as complex as implementing a repeatable data mining process (see Figure 3.8).

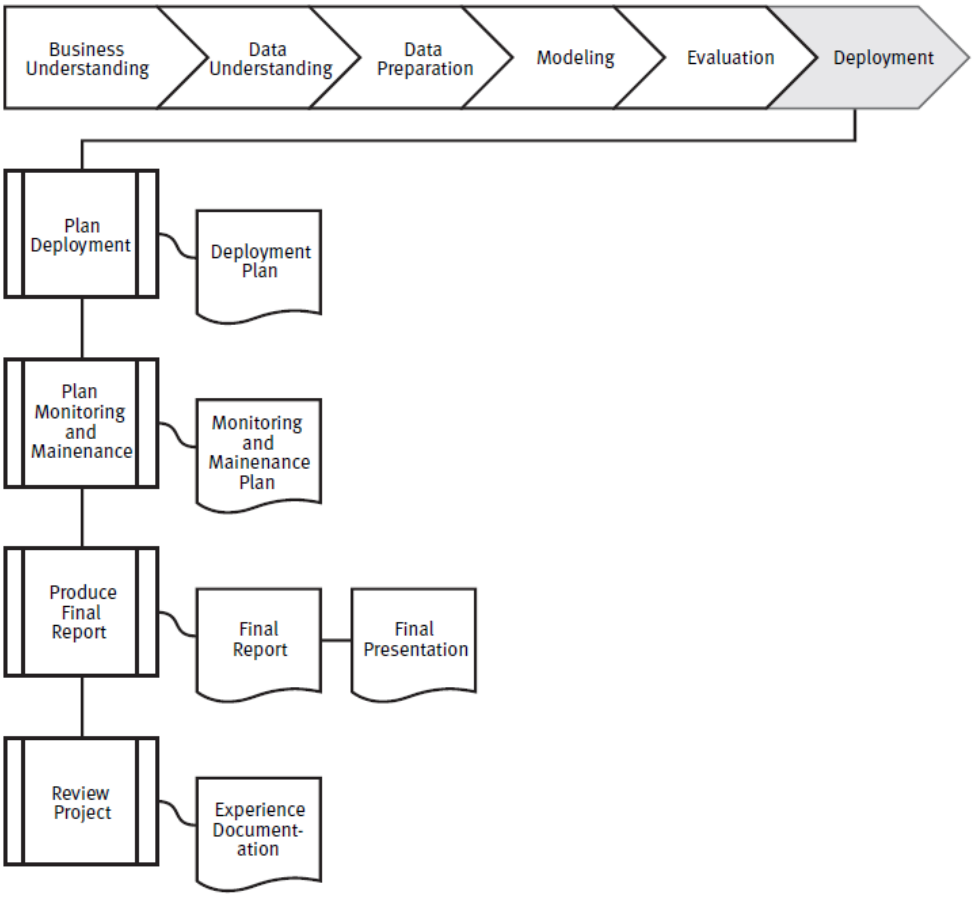

Figure 3.8 Deployment phase (Chapman et al., 2000)

3.3 Data Mining Models

Data mining tools take data and construct a representation of reality in the form of a model (Rygielski et al., 2002). Data mining models can be categorized in predictive models (supervised learning) and descriptive models (unsupervised learning). Predictive models predict a target value. So, these models require that the data set contains predefined targets. Descriptive models extract hidden information from the dataset. Therefore, they do not require the dataset to contain the target variables. General classification of data mining models is illustrated in Figure 3.9.

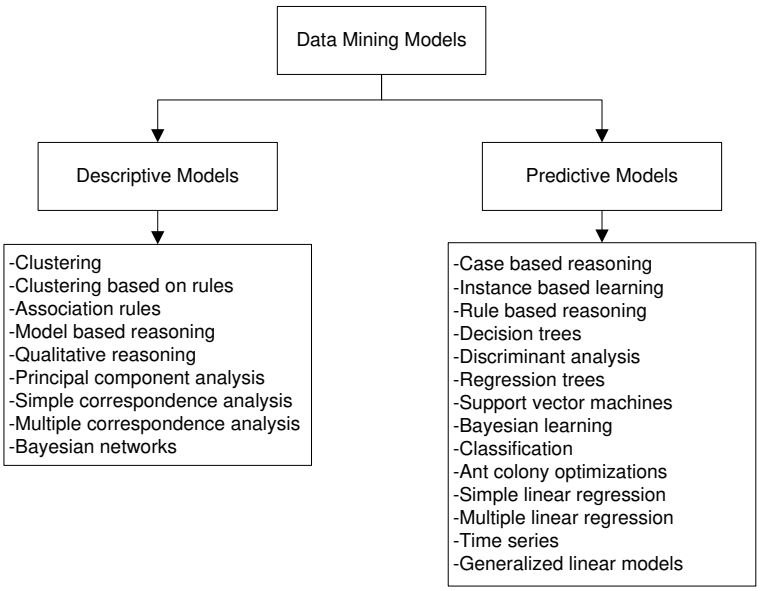

Figure 3.9 Classification of data mining models (Gilbert et al., 2012)

3.3.1 Predictive Models

Predictive models are generated using data with known targets and it is aimed to predict the results of data with unknown targets by using these models. Classification and regression are the most commonly used predictive models.

3.3.1.1 Classification

Classification is a data mining technique used to predict group membership for data instances (Phyu, 2009). In classification, there is a target categorical variable. The data mining model examines a large set of records. The record contains information on the target variable as well as a set of input or predictor variables.

3.3.1.2 Regression

Regression models establish a relationship between a dependent or outcome variable and a set of predictor(s). If there is only one predictor in the model, this model is named as linear regression model.

On the other hand, there can be more than one predictor in the model. In this case, the model is named as multiple regression model. The formula for simple linear regression is as follows:

$$\hat{y} = a + \beta x \tag{3.1}$$

where, $\hat{y}$ is the outcome variable, and x is the predictor. a and β are unknown parameters or in other words regression coefficients

The multiple regression formula is:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{3.2}$$

where, $\hat{y}$ is the outcome variable and x_k's are the predictors. β_0 and β_k's are regression coefficients.

3.3.2 Descriptive Models

Descriptive models are used to describe all of the data in a given dataset. Specifically, these models synthesize all of the data to provide information regarding trends, segments and clusters that are present in the information searched. Descriptive models try to find models for the data to help the decision maker. Most commonly used descriptive models are clustering and association.

3.3.2.1 Clustering

Clustering is the division of a heterogeneous population into more homogenous groups. Clustering differs from classification in that there is no target variable for clustering in other words clustering is an unsupervised learning technique where there are no predefined classes. The clustering task does not try to classify, estimate, or predict the value of a target variable.

One example of clustering application is market segmentation in which marketers take larger customer groups and segment them by homogeneous characteristics.

3.3.2.2 Association

The association models try to find correlations between different attributes in a dataset. The most common application of this kind of algorithm is for creating association rules, which can be used in a market basket analysis. Association rules are of the form "if antecedent, then consequent," together with a measure of the support and confidence associated with the rule.

In this study data clustering is used to segment the customers of an international TV manufacturing company. Accordingly, data clustering is introduced in the next section.

3.4 Data Clustering

Clustering refers to the grouping of records, observations or cases into classes of similar objects. A cluster is a collection of records that are similar to one another and dissimilar to records in other clusters. In general, to be useful in an engineering application, a clustering algorithm should have the following abilities (Pham & Afify, 2007):

- Dealing with different types of data (numerical, categorical, text, and images)
- Handling noise, outliers, and fuzzy data
- Discovering clusters of irregular shapes
- Dealing with large data sets and data of high dimensions
- Producing results that are easy to understand
- Being insensitive to the order of the input data
- Being simple to implement

An example of grouping data points into two, four and six clusters is can be seen in Figure 3.10.

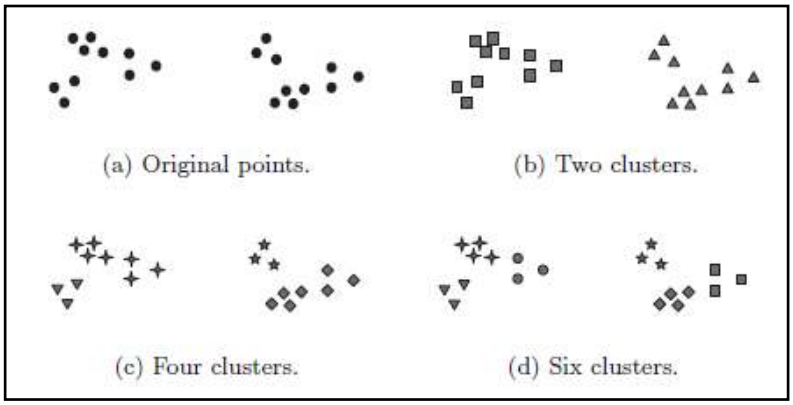

Figure 3.10 Clustering example (Tan et al., 2006)

3.4.1 Data Clustering Steps

Typical pattern clustering activity involves the following steps (Jain & Dubes, 1988):

- Pattern representation
- Definition of a pattern proximity measure appropriate to the data domain
- Clustering or grouping
- Data abstraction
- Assessment of output

3.4.1.1 Pattern Representation

Pattern representation refers to the determination of number, type, scale of variables or features for the determination of the similarity that exists between the observations. This is the stage of creation of the data matrix. $N * p$ dimensional data matrix where N is number of observations and p is number of attributes can be shown as follows:

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{N1} & \cdots & X_{Np} \end{bmatrix} \quad (3.3)$$

3.4.1.2 Pattern Proximity

Pattern proximity includes the calculation of similarities or dissimilarities of pair of objects with an appropriate distance measure. It can also refer to the determination of proximity matrix. A variety of distance measures are in use in the literature for this purpose. The most commonly used distance measures are explained in the following for *n* dimensional two points, $x = (x_1, x_2, \dots x_n)$ and $y = (y_1, y_2, \dots y_n)$.

3.4.1.2.1 Euclidean Distance. The Euclidean distance between two points like *x* and *y* is the length of the line segment connecting them. It is computed by using the following equation;

$$d(x,y) = d(y,x) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.4}$$

3.4.1.2.2 Squared Euclidean Distance. Standard Euclidean distance can be squared in order to place greater weight on objects (Patel & Mehta, 2011). Squared Euclidean distance between two points is computed as follows:

$$d(x,y) = d(y,x) = \sum_{i=1}^{n}(x_i - y_i)^2 \tag{3.5}$$

3.4.1.2.3 Pearson Distance. Pearson distance is computed by using Eq. 3.6. Where, S_i denotes variance of the variable.

$$d(x,y) = d(y,x) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2/S_i^2} \tag{3.6}$$

3.4.1.2.4 Manhattan (City-Block) Distance. Manhattan distance computes the absolute differences between coordinates of a pair of objects by using the following equation (Grabusts, 2011):

$$d(x,y) = d(y,x) = \sum_{i=1}^{n} |x_i - y_i| \tag{3.7}$$

3.4.1.2.5 Minkowski Distance. Minkowski distance is a generalized metric distance. This formula derives new formulas based on different p values. For instance, when p=2, the distance becomes the Euclidean distance (Grabusts, 2011). It is computed as follows.

$$d(x,y) = d(y,x) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} \tag{3.8}$$

3.4.1.3 Grouping

Clustering methods are classified in different ways. The most common distinction is the categorization of the methods as hierarchical and non-hierarchical (partitional) methods. Hierarchical clustering algorithms (HCA) recursively find nested clusters either in agglomerative mode or in divisive mode.

Compared to hierarchical clustering algorithms, partitional clustering algorithms find all the clusters simultaneously as a partition of the data and do not impose a hierarchical structure (Jain, 2010). Figure 3.11 shows the general classification of clustering methods.

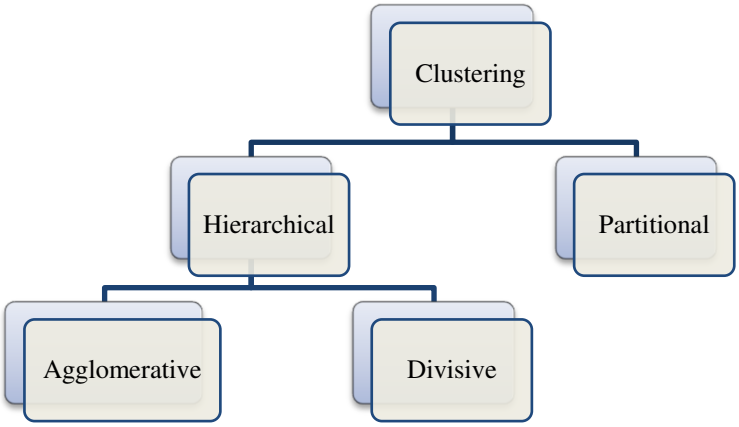

Figure 3.11 Clustering methods

3.4.1.3.1 Hierarchical Clustering Algorithms. Hierarchical clustering algorithms consist of steps that based on adding an object to a cluster or deleting an object from a cluster and these steps show a tree-like structure.

Based on a bottom-up or to down decomposition, the hierarchical algorithms can be classified as agglomerative and divisive. Agglomerative clustering treats each data point as a single cluster and then successively merges clusters until all points have been merged into single cluster. Divisive clustering treats all data points in a single cluster and successively breaks the clusters till one data point remains in each cluster (Manu, 2012). Figure 3.12 illustrates the agglomerative and divisive hierarchical clustering approaches.

One of the major problems in clustering analysis is termination of the algorithm, in other words determination of the number of clusters. The ideal number of clusters is the level of minimum variation within clusters and the maximum variation between clusters. However, the final decision on the number of clusters is left to decision maker.

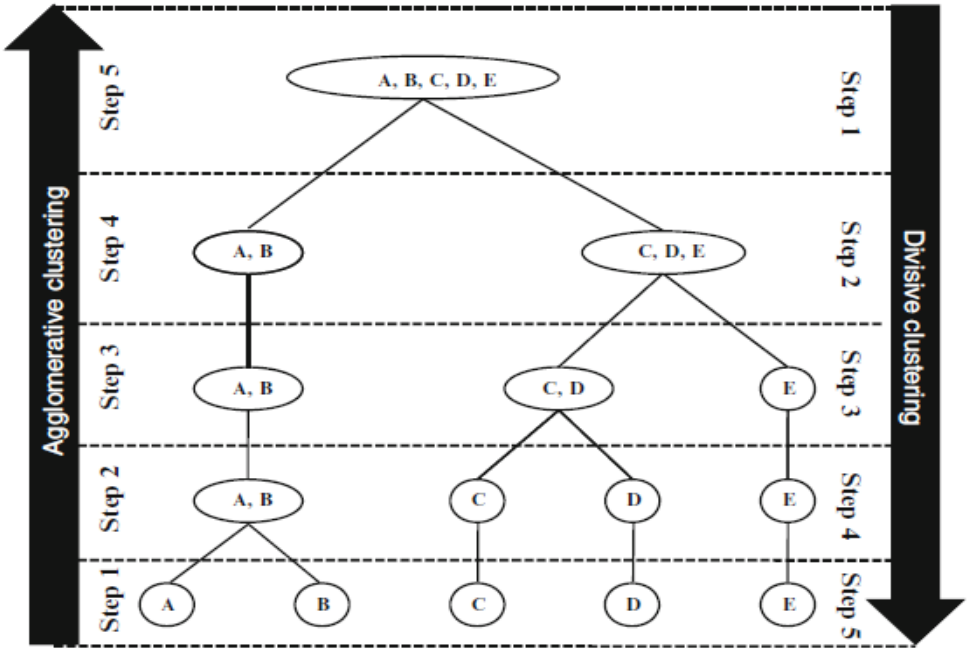

Figure 3.12 Agglomerative and divisive clustering (Mooi & Sarstedt, 2011)

3.4.1.3.1.1 Agglomerative Hierarchical Clustering (AHC) Algorithms. AHC starts with each data point in its own cluster and merges the most similar pair of clusters successively to form a cluster. The most commonly used AHC algorithms are "single linkage", "complete linkage", "average linkage" and "Ward's algorithms" (Punj & Stewart, 1983). These algorithms are explained in the following.

3.4.1.3.1.1.1 Single Linkage (Nearest Neighbor) Algorithm. In this method, all distances between items are computed and then two items that have the minimum distance are selected and combined into a new cluster. Then, an item that have the smallest distance to this cluster is added to the cluster or another two items that have the minimum distance are combined into a new cluster. This process continues until all clusters merge into a single cluster. The steps of single linkage algorithm are as follows (Dhandayudam & Krishnamurthi, 2012):

(1) Assign each object to its own cluster,

(2) Calculate the distance from each object to all other objects using a distance measure and store it in a distance matrix,

(3) Identify the two clusters with the shortest distance in the matrix and merge them together,

(4) The distance of an object to the new cluster is the minimum distance of the object to the objects in the new cluster,

(5) Update the distance of each object to the new cluster in the distance matrix,

(6) Repeat steps 3 to 5 until the required number of clusters are obtained.

If we consider a one-dimensional data set {2 5 9 15 16 18 25 33 33 45}, single linkage algorithm works as follows;

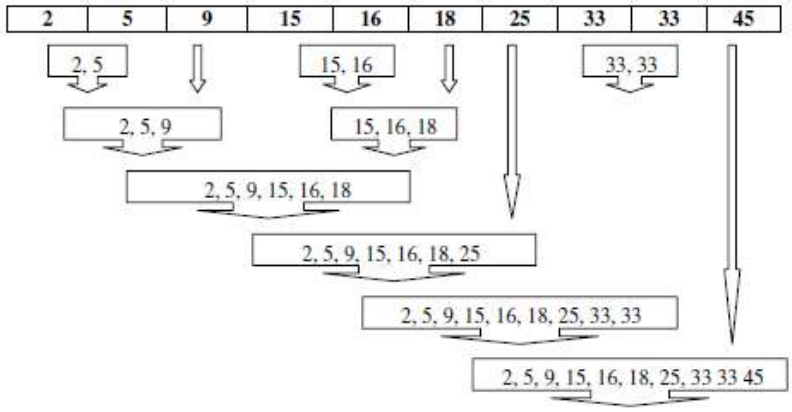

Figure 3.13 Single linkage agglomerative clustering on the sample data set

Distance between 33 and 33 has the smallest value with 0 (see Table 3.1). So, these two items should be combined at first step. Then, distance matrix should be updated. Iterative procedure of this algorithm can be seen in Figure 3.14.

Table 3.1 Proximity matrix of the sample data set

	2	5	9	15	16	18	25	33	33	45
2		3	7	13	14	16	23	31	31	43
5			4	10	11	13	20	28	28	40
9				6	7	9	16	24	24	36
15					1	3	10	18	18	30
16						2	9	17	17	29
18							7	15	15	27
25								8	8	20
33									**0**	12
33										12
45										

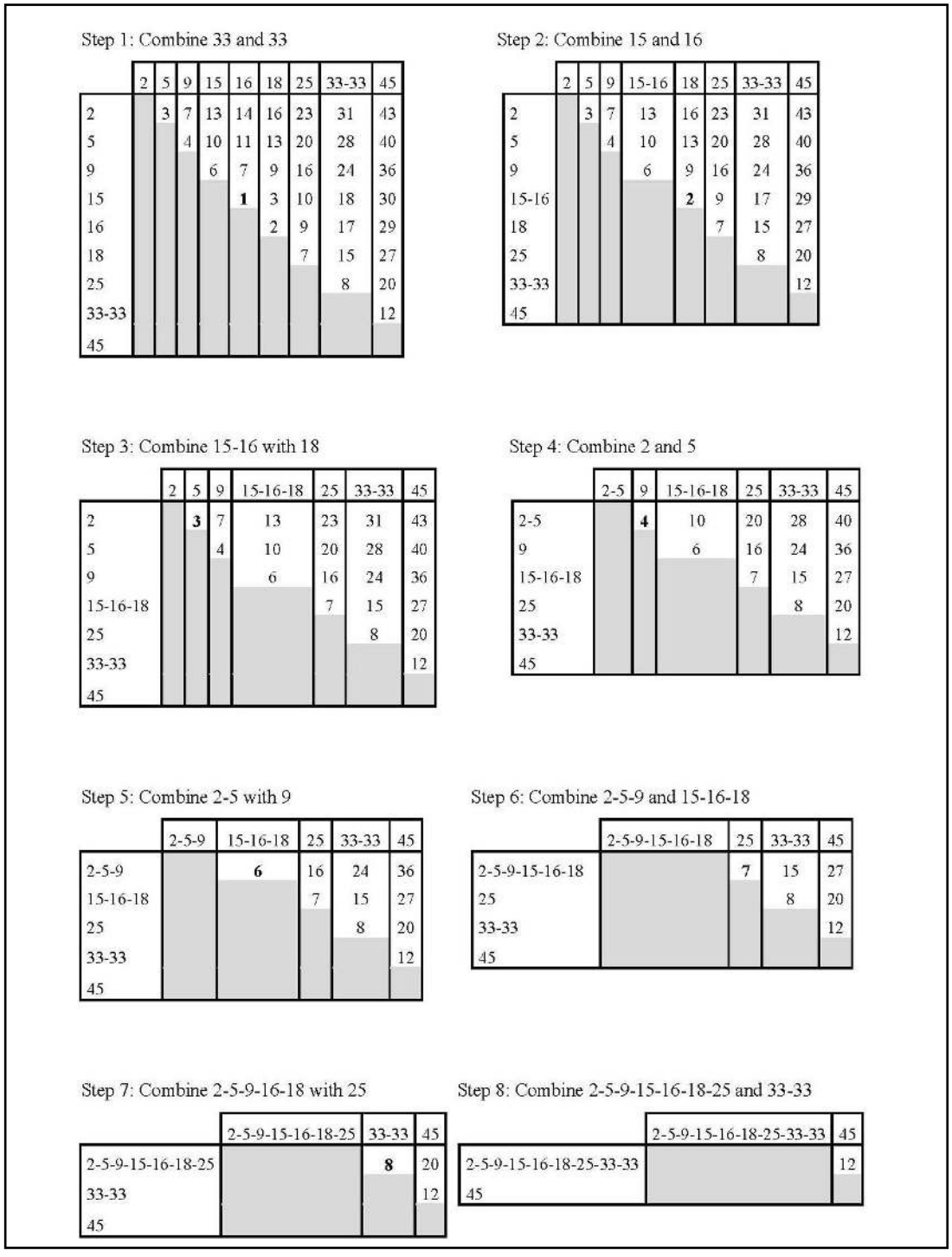

Figure 3.14 Iterative procedure of single linkage algorithm

Finally in step 9; 2, 5, 9, 15, 16, 18, 25, 33, 33 are combined with 45.

3.4.1.3.1.1.2 Complete Linkage (Farthest Neighbor) Algorithm. The complete linkage algorithm computes the distances between all items in two clusters and selects the highest as the measure of similarity. The steps in complete linkage algorithm are as follows (Dhandayudam & Krishnamurthi, 2012):

(1) Assign each object to its own cluster,

(2) Calculate the distance from each object to all other objects using a distance measure and store it in a distance matrix,

(3) Identify the two clusters with the shortest distance in the matrix and merge them together,

(4) The distance of an object to the new cluster is the maximum distance of the object to the objects in the new cluster,

(5) Update the distance of each object to the new cluster in the distance matrix,

(6) Repeat steps 3 to 5 until the required number of clusters are obtained.

If we consider the same one-dimensional data set {2 5 9 15 16 18 25 33 33 45}, complete linkage algorithm works as follows;

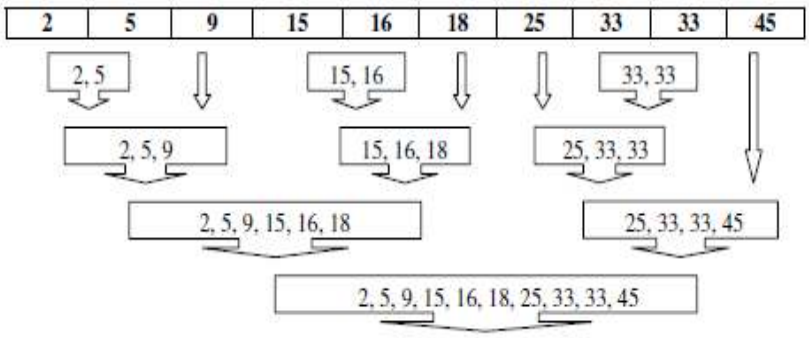

Figure 3.15 Complete linkage agglomerative clustering on the sample data set

Details of the iterations are illustrated in Figure 3.16.

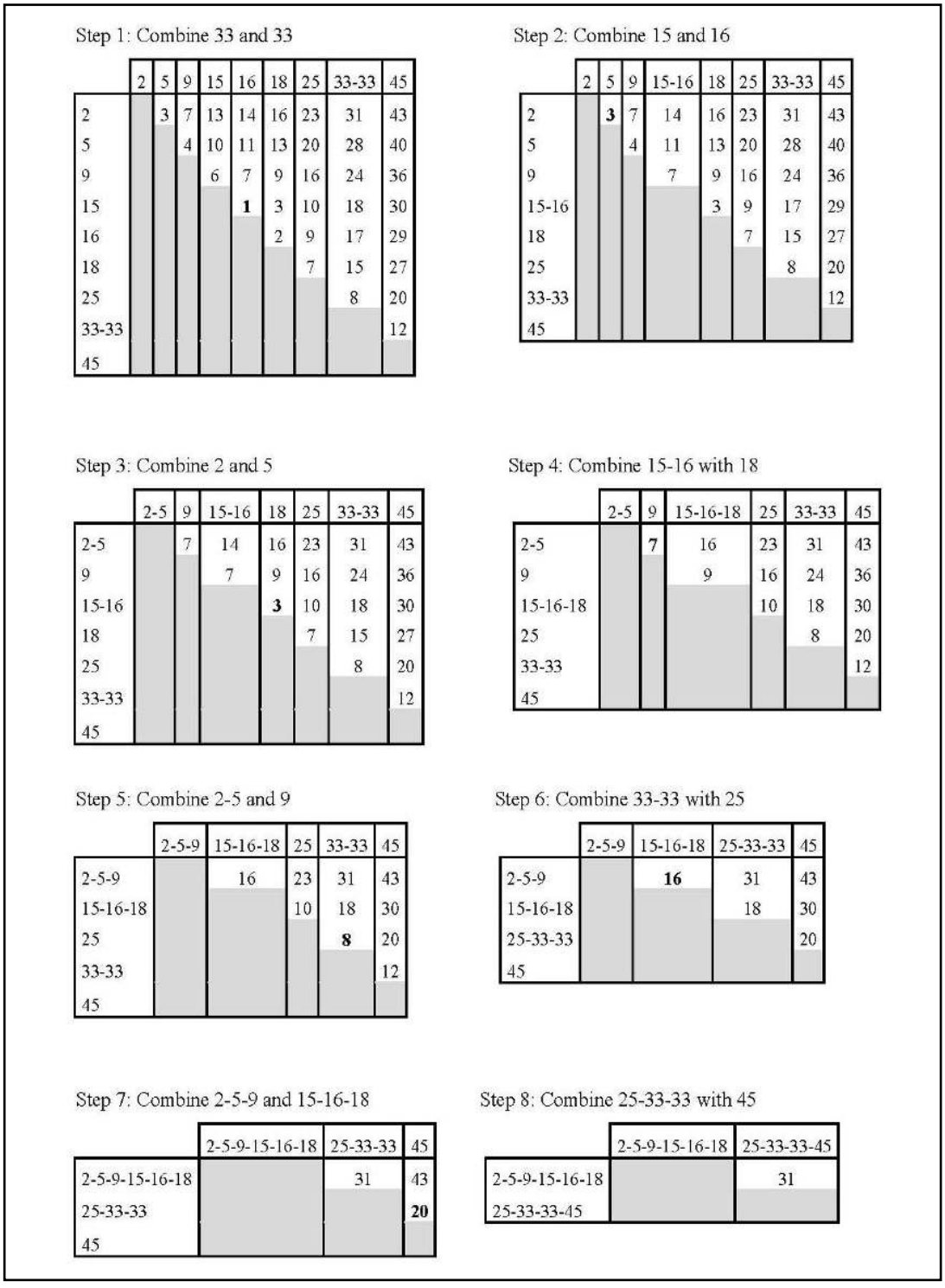

Figure 3.16 Iterative procedure of complete linkage algorithm

Finally combine 2-5-9-15-16-18 with 25-33-33-45 in step 9.

3.4.1.3.1.1.3 Average Linkage Algorithm. This algorithm defines distance between groups as the average of the distances between all pairs of individuals in the two groups. For the previously used sample data set, average-linkage method works as follows:

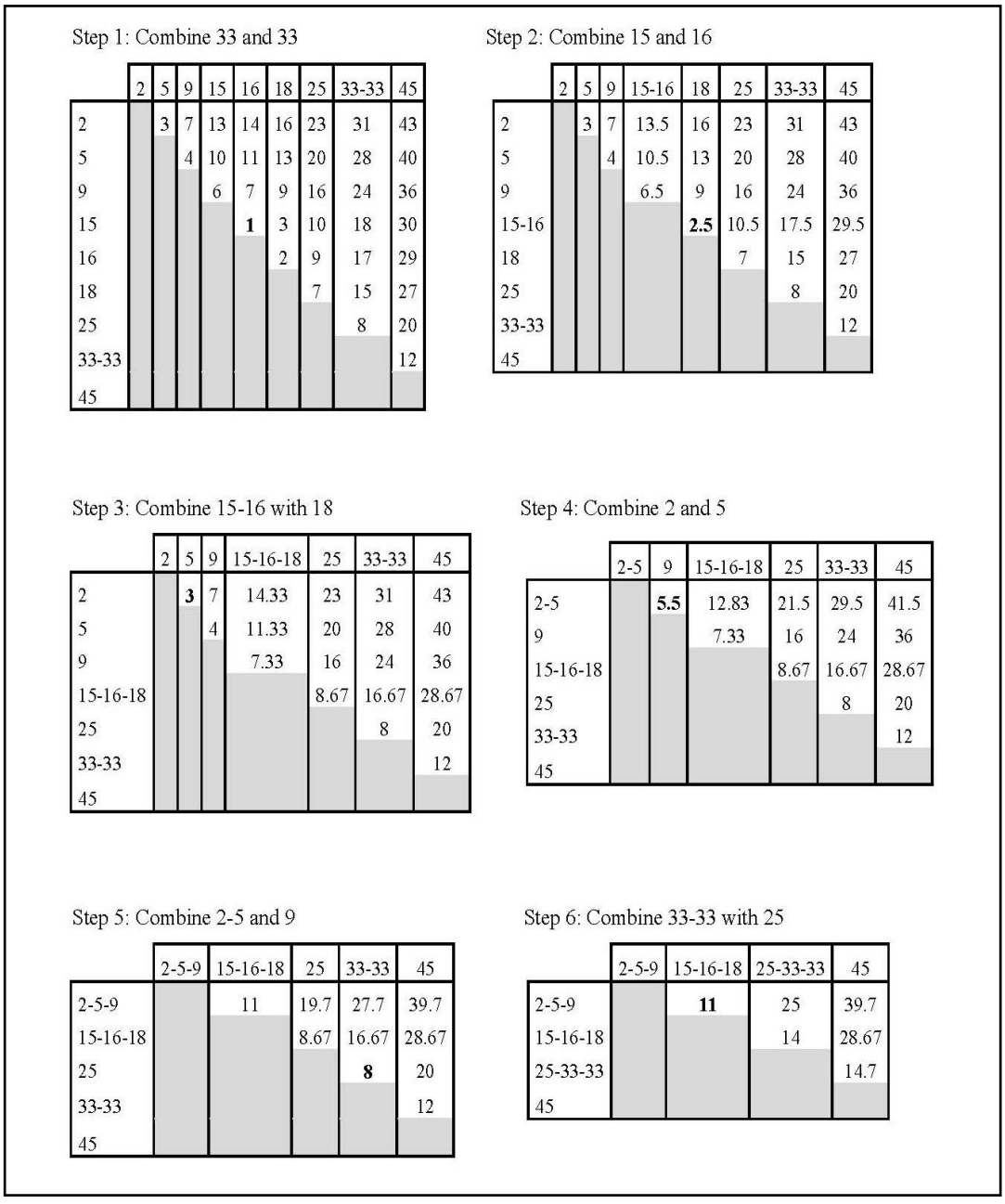

Figure 3.17 Iterative procedure of average linkage algorithm

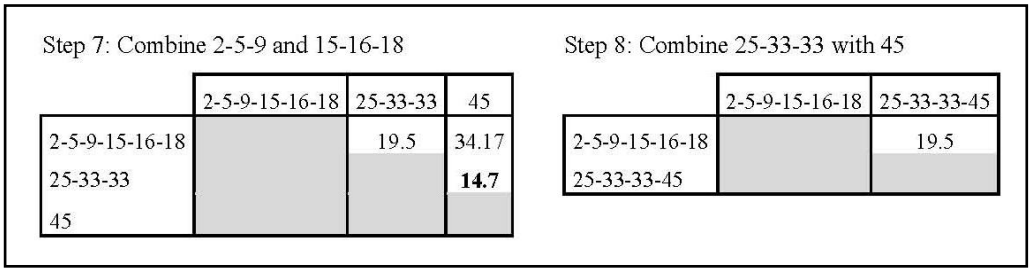

Figure 3.17 continued

Finally combine 2-5-9-15-16-18 with 25-33-33-45 in step 9.

3.3.1.3.1.4 Ward's Algorithm. Ward's method doesn't work over distances. This method groups objects in order to maximize homogeneity of clusters (Ward, 1963). In this method, within cluster sum of squares are considered instead of group linkages. At each generation, within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation. With hierarchical clustering, the sum of squares starts out at zero (because every point is in its own cluster) and then grows as we merge clusters. Ward's method keeps this growth as small as possible.

3.4.1.3.1.2 Divisive Hierarchical Clustering Algorithms. Divisive algorithms work contrary to agglomerative algorithms. In divisive approach, at first, all data points are in the same cluster. Then, items in this cluster divided into two sub-groups. Then, these groups again divided into dissimilar sub-groups. This process continues until the number of clusters equals to the number of observations.

3.4.1.3.2 Partitional Clustering Algorithms. These algorithms partition the database into a set of *k* clusters so that it optimizes the chosen partition criterion. Each object is placed in exactly one of the *k* non-overlapping clusters. Generally, number of clusters assumed to be known for non hierarchical clustering algorithms. There exist several partitioning criteria in the literature. Trace (W) and Determinant (W) are the most commonly used criteria (İyigün, 2008). These criteria are explained in the following.

The portioning of the data points x_i (which are the rows of the $N \times p$ data matrix) gives rise to the $p \times p$ total dispersion matrix,

$$T = \sum_{k=1}^{K} \sum_{x_i \in C_k} (x_{ik} - \bar{x})(x_{ik} - \bar{x})' \tag{3.9}$$

where, p-dimensional vector $\bar{x}$ is the mean of all data points and K is the number of clusters. Total dispersion matrix T can be portioned into within group dispersion matrix,

$$W = \sum_{k=1}^{K} \sum_{x_i \in C_k} (x_{ik} - \bar{x}_k)(x_{ik} - \bar{x}_k) \tag{3.10}$$

Herein, $\bar{x}_k$ is the mean of the data points in cluster C_k. Between-cluster dispersion matrix can be computed as follows:

$$B = \sum_{k=1}^{K} N_k(\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})' \tag{3.11}$$

where, N_k is the number of data points in C_k. So that,

$$T = W + B \tag{3.12}$$

For univariate data (p=1), Eq. 3.12 represents the division of total sum of squares of a variable into the within and between clusters sum of squares.

Minimization of Trace (W)

Minimization of Trace (W) means the minimization of the sum of within cluster sum of squares. Minimizing trace works to make the clusters more homogeneous, thus the problem, min {trace W} is equivalent to max {trace B}.

Minimization of Determinant (W)

The differences in cluster mean vectors are based on the ratio of the determinants of the total and within-cluster dispersion matrices. Large values of $\frac{\det(T)}{\det(W)}$ ratio indicate that the cluster mean vectors differ. Thus, a clustering criterion can be constructed as the maximization of this ratio. Since *T* is the same for all partitions of *N* data points into *K* clusters, this problem is equivalent to min det (W).

k-means clustering, metoid clustering, fuzzy clustering, hill climbing clustering are some of the non-hierarchical clustering techniques. However, k-means is the most commonly used algorithm in the literature.

3.4.1.3.2.1 k-means Algorithm. This algorithm is a well known algorithm and finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized (Jain, 2010). The steps in k-means algorithm are as follows (Dhandayudam & Krishnamurthi, 2012):

(1) Initialize centers for *k* clusters randomly
(2) Calculate distance between each object to *k*-cluster centers using the a distance measure
(3) Assign objects to one of the nearest cluster center
(4) Calculate the center for each cluster as the mean value of the objects assigned to it
(5) Repeat steps 2 to 4 until the objects assigned to the clusters do not change.

The assignment of objects to *k* clusters depends on the initial centers of the clusters. The output differs if the initial centers of the clusters are varied. Typical k-means clustering is illustrated in Figure 3.18.

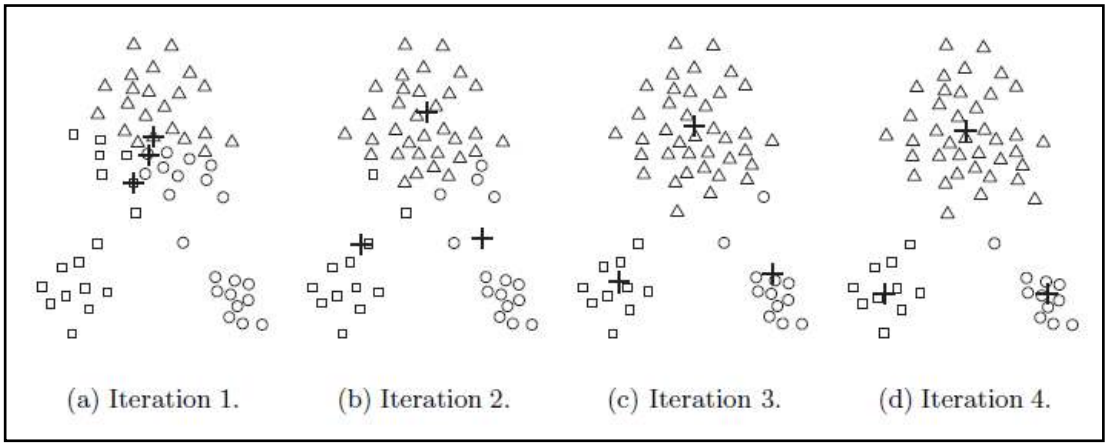

Figure 3.18 Iterative procedure of k-means algorithm (Tan et al., 2006)

3.4.1.4 Data Abstraction

Data abstraction is the process of extracting a simple and compact representation of a data set. In the clustering context, a typical data abstraction is a compact description of each cluster, usually in terms of cluster prototypes or representative patterns such as the centroid (Diday & Simon, 1976).

3.4.1.5 Assessment of Output

Different clustering algorithms often result in entirely different partitions even on the same data. Validity assessments are usually objective (Dubes, 1993) and are performed to determine whether the output is meaningful. Validation is a technique to find a set of clusters that best fits natural partitions (number of clusters) without any class information (Rendon et al., 2011). Statistical approaches that use optimality of a specific criterion are often used for the validation. The most commonly used cluster validity indices to evaluate the quality of the discovered clusters are described in the following.

3.4.1.5.1 Dunn Index. Dunn's validity index (Dunn, 1974), attempts to define the separation of clusters.

If a data set contains compact clusters, the distances among the clusters are large and the diameters of clusters are expected to be small (Halkidi et al., 2002). So, larger value of this index means better clustering. This index can be computed as follows:

$$D = min_{i=1\dots n_c}\left\{min_{j=i+1\dots n_c}\left(\frac{d(c_i, c_j)}{max_{k=1\dots n_c}(diam(c_k))}\right)\right\} \tag{3.13}$$

where, n_c denotes the number of clusters; i, j are cluster labels; then $d(c_i, c_j) = min_{x\in c_i, y\in c_j}\{d(x,y)\}$ and $diam(c_i) = max_{x,y\in c_i}\{d(x,y)\}$.

3.4.1.5.2 Davies Bouldin Index. This index is based on similarity measure of clusters and defined as (Davies & Bouldin, 1979):

$$DB = \frac{1}{n_c}\sum_{i=1}^{n_c} max_{1\le j\le n_c, i\ne j}\left\{\frac{d(c_i) + d(c_j)}{d(c_i, c_j)}\right\} \tag{3.14}$$

where, n_c denotes the number of clusters; i, j are cluster labels, then, $d(c_i)$ and $d(c_j)$ are the average distances of all samples in clusters i and j to their respective cluster centroids. $d(c_i, c_j)$ is the distance between these centroids. Smaller value of DB indicates a "better" clustering solution.

3.4.1.5.3 Silhouette Index. This index computes the silhouette width for each cluster and overall average silhouette width for the entire data set (Rousseeuw, 1987). To compute the silhouette width of i^{th} data point, following equation is used:

$$S_i = \frac{b_i - a_i}{\max(a_i\ ,\ b_i)} \tag{3.15}$$

where a_i is the average distance between the i^{th} data point to all other points in the same cluster. b_i is the minimum average distance between the i^{th} data point to all other points in other cluster. This index takes values between -1 and 1. A value of S_i close to 1 indicates better clustering. The overall average silhouette width for the data set is the average S_i for all data points.

Therefore, the number of cluster with maximum average overall silhouette width can be defined as optimal number of clusters.

3.4.1.5.4 Sum of Squares. A good clustering clusters objects such that similarity within a cluster is high (small sum of squares within cluster) while similarity between clusters is very low (high sum of squares between cluster). It is possible to show that the sum of the total sum of squares within cluster (*SSW*) and the total sum of squares between clusters (*SSB*) is a constant that is equal to the total sum of squares (*TSS*) which is the sum of squares of the distance of each point to the overall mean of the data (Eq. 3.16). The importance of this result is that minimizing *SSW* is equivalent to maximizing *SSB*.

$$TSS=SSW+SSB \tag{3.16}$$

$$TSS = \sum_{k=1}^{K} \sum_{\forall x_i \in C_k} (x_i - \mu)^2 \tag{3.17}$$

where, K denotes the number of clusters; C_k is the set of instances in cluster k; μ is the vector mean of data set.

SSW is the most widely used criterion to evaluate the validity of clustering results and determine the number of clusters. *SSW* is defined as follows:

$$SSW = \sum_{k=1}^{K} \sum_{\forall x_i \in C_k} (x_i - \mu_k)^2 \tag{3.18}$$

where, K denotes the number of clusters; C_k is the set of instances in cluster k; μ_k is the vector mean of cluster k. Smaller value of *SSW* indicates a "better" clustering.

3.3.1.5.5 C Index. C index is formulated as follows:

$$C = \frac{S - S_{min}}{S_{max} - S_{min}} \tag{3.19}$$

Herein, *S* is the sum of distances over all pairs of objects forms the same cluster. Let *m* be the number of those pairs and S_{min} is the sum of the *m* smallest distances if all pairs of objects are considered. Likewise, S_{max} is sum of the *m* largest distances out of all pairs. *C* index is limited to the interval [0,1] and should be minimized (Ansari et al., 2011).

3.3.1.5.6 Calinski - Harabasz Index. This index is computed by the following formula,

$$CH = \frac{SSB/(k-1)}{SSW/(n-k)} \tag{3.20}$$

where, *n* is number of data points, *SSB* is sum of squares between clusters, *SSW* is sum of squares within cluster and *k* is the number of clusters. Larger value of this index indicates a better clustering (Rendon et al., 2011).

CHAPTER FOUR
CUSTOMER SEGMENTATION: AN APPLICATION IN AN INTERNATIONAL TV MANUFACTURING COMPANY

4.1 Problem Definition and Business Understanding

Today, companies may have thousands of customers. It is complex and difficult to manage such large customer bases. In addition, developing customer-specific marketing strategy is time consuming and difficult to follow. Therefore, dividing customers into small groups according to their similarities will be more meaningful with respect to develop strategies and management.

It is possible to extract hidden patterns, associations and relationships from large customer related databases using data mining techniques. Data mining helps the companies on issues that classify and identify the customers, and predict their behaviors. One of the most common application areas of data mining in CRM is customer segmentation. As a data mining technique, data clustering can be employed for customer segmentation.

Cluster analysis achieves only the grouping of similar observations in the same cluster. Whereas, finding the relative importance of clusters by using specific evaluation characteristics is also important in customer segmentation. In this way, we can see which segment is more important / valuable for the company.

The aim of this study is to divide customers into small manageable groups using clustering algorithms and also to find relative importance of these groups using multi criteria decision making technique. In this regard, a customer segmentation approach is proposed and implemented in an international TV manufacturing company that is located in Turkey.

The company is the leader in its segment in Turkey and one of the most successful TV manufacturing companies over the world. It has an integrated TV production process extending from the production of electronic cards to the final assembly.

It has a wide export network that covers 127 countries and this network keeps expanding. The company has been in electronics industry for thirteen years and it performs 82 percent of Turkey's total exports of LCD TV. It manufactures industry-leading products by the adoption of innovation. With its high manufacturing technologies, it produces TVs not only for its own brand but also for the electronic leaders of the world.

4.2 The Proposed Customer Segmentation Approach

The proposed customer segmentation approach uses two different approaches as illustrated in Figure 4.1. It starts with understanding the business environment and preparation of the data set. Then, customer evaluation characteristics are defined and computed for each customer. In this study, eight different characteristics were defined namely *"recency", "loyalty", "average annual demand", "average annual sales revenue", "frequency", "long term relationship potential", "average percentage change in annual demand" and "average percentage change in annual sales revenue"*. In the next step, data set is normalized using a normalization method. Then, importance weights of the characteristics are determined using a multi-criteria decision making (MCDM) technique. In this study, fuzzy AHP is employed in this stage.

Customer segmentation involves two different approaches; single dimension (SD)-based segmentation and multiple dimensions (MD)-based segmentation. First approach segments customers according to a combined characteristic called "*overall score"* that is computed for each customer by taking the weighted average of the predetermined characteristics. On the other hand, second approach groups customers according to their similarities with respect to the characteristics under concern. As stated before, customer base of the company is segmented by using eight characteristics in this study.

Customers of the company under concern are segmented using both hierarchical and partitional clustering algorithms and then the results are validated using cluster validity indexes. More specifically, Ward's method, single linkage and complete linkage methods are used as the agglomerative hierarchical clustering algorithms, and k-means is employed as the partitional clustering algorithm in this study. Additionally, to validate the clustering results *SSW* is used.

Finally, importance levels of the segments were determined. In case of SD-based segmentation, the clusters were ranked according to the cluster centroids. In MD-based segmentation, they were ranked by using weighted average of the cluster centroids. Consequently, we determined the final customer segments and profiled these segments.

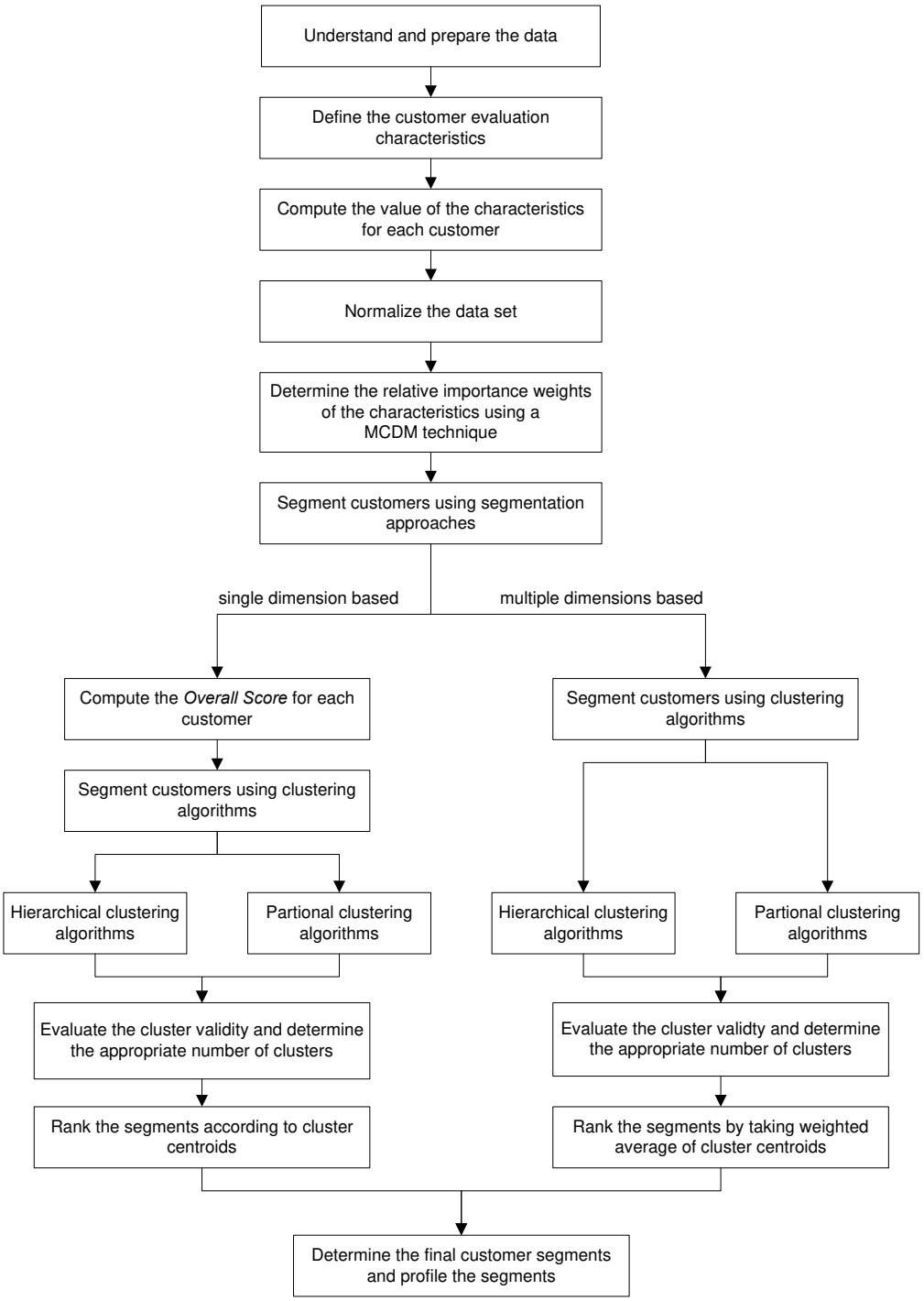

Figure 4.1 The proposed approach

4.3 Order-Selling Process of the Company

Foreign trade is an important department of the company and carries out sales and marketing activities. It performs international sales and marketing activities with the sales offices that located in France, Germany, Spain, England, Holland, Italy, Finland, Russia and Romania.

Order selling process of the company is summarized in Figure 4.2. The process starts with customer visits that are carried out by sales specialists to find out customer expectations. If both parties agree with the terms and conditions at the end of negotiations, sales agreement is signed. Then, the customer orders products through web-order channel and after that foreign trade enter the order to the ERP (Enterprise Resource Planning) system of the company. An order can be for a new product or an existing product. If the order is for a new product, BOM (Bill of Materials) specialists create new BOM list for this product in coordination with R&D (Research and Development) department. Both R&D and production planners determine the quantity and qualifications of pilot production of the new product. If pilot production is successful, production order is entered to the system by production planning department. Otherwise, qualifications of this new product are investigated by R&D team and they make the required changes.

In case of ordering for an existing product, BOM specialists check the existing BOM list and they revise it if needed. Then, production planners control the production capacity and material stocks, and afterward confirm the order. The confirmation includes production and completion dates of the order. Then, they open supply requests to the procurement department for the materials by using MRP. When all materials completed, the production starts. At the end of the production, if products pass the quality tests they are shipped to the customer. Shipment and payment are made simultaneously for the security of the process.

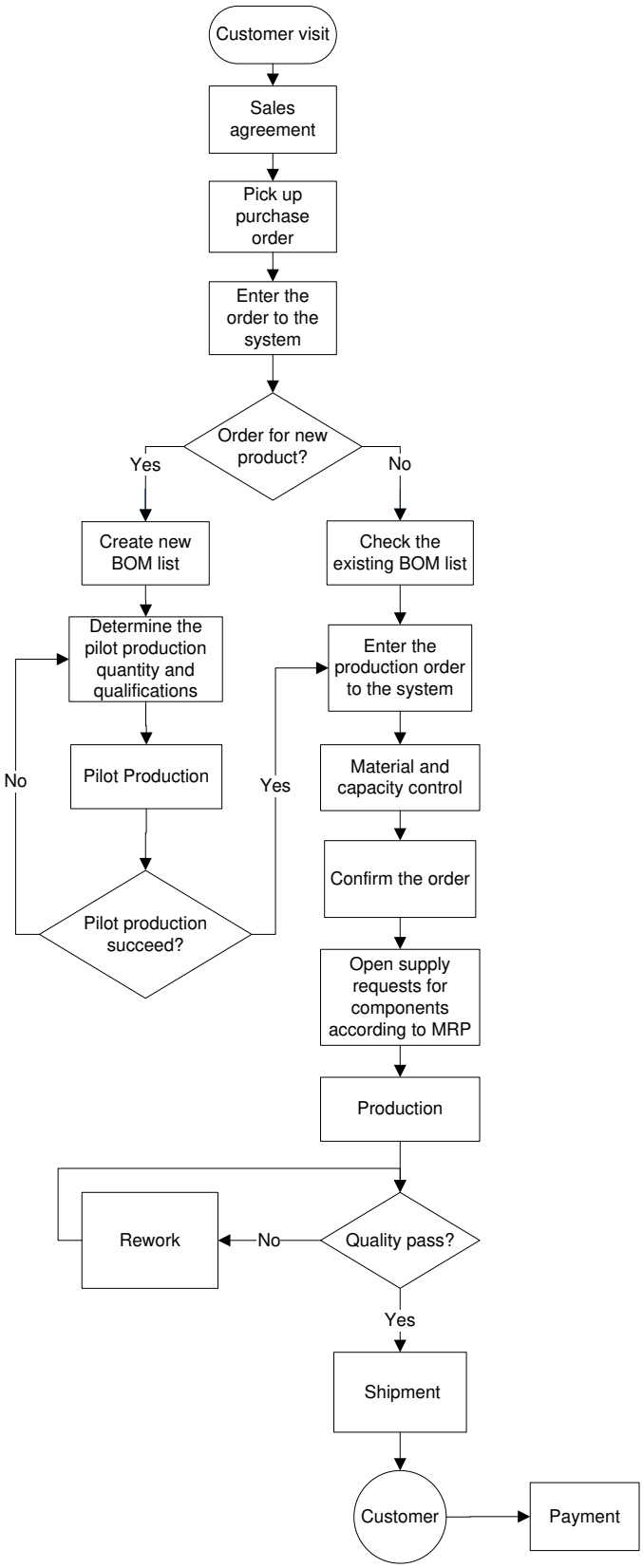

Figure 4.2 Order-selling process

4.4 Data Understanding and Preparation

In this study, customer transaction data was extracted from the ERP system of the company. 28880 order transactions of 329 customers were considered. 40 of the transactions associated with sample orders were found unusable and ignored. As a result, 28840 records of 317 customers between January 2002 and December 2011 were analyzed. Raw data consists of product order (PO) number, customer ID, customer name, country, product ID, product name, product group, year, month, quantity, unit price and total price information as reported in Table 4.1.

Table 4.1 Order transaction data

PO Number	Customer ID	Customer Name	Country	Product ID	Product Name	Product Group	Year	Month	Quantity	Unit Price	Total Price
***	1	***	Tunisia	***	***	LCD	2011	11	1200	319.21	38052
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

The data that was obtained from the ERP system was transferred to Excel and then pivot tables were generated for the calculations. Customer IDs, customer names, product IDs and product names were hidden as the privacy policy of the company. Herein, customer IDs were sorted in ascending order and the original IDs were converted to numeric values from 1 to 317. Unit prices are quoted in TL. Actually, sales are conducted in dollars but the ERP system converts the monetary values from dollar to TL using daily exchange rate.

If we evaluate the basic statistics derived from the database, Figure 4.3 indicates that the number of customers who conducted a transaction show an increasing trend. In addition, customers were grouped according to their countries. 14 different zones were defined and geographical dispersion of the customer portfolio was summarized in Table 4.2. We should underline that demand and sales revenue related statistics are not presented in this study as the privacy policy. As reported in Table 4.2, customers spread a wide range of geographic area. Most of the customers are from Balkans, Middle East and Middle Europe.

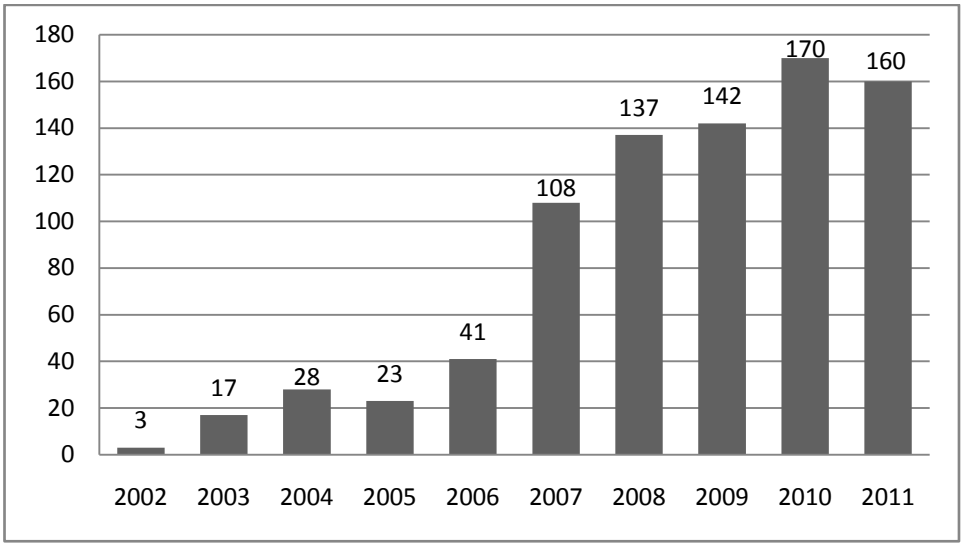

Figure 4.3 Number of customers over the years

Table 4.2 Geographical dispersion of customer portfolio

Zone	# of customers	%
BALKANS	64	20.19
MIDDLE EAST	58	18.30
MIDDLE EUROPE	56	17.67
NORTH AFRICA	35	11.04
SCANDINAVIA	30	9.46
MEDITERRANEAN	18	5.68
EAST EUROPE	11	5.47
CENTRAL ASIA	11	5.47
BALTICS	9	2.84
SOUTH AFRICA	9	2.84
SOUTH ASIA	7	2.21
FAR EAST	5	1.58
OCEANIA	3	0.95
SOUTH AMERICA	1	0.32

4.5 Determining Customer Evaluation Characteristics

Customers are evaluated according to their different features with respect to the sector of the company. Types of data held about customers vary across sectors and even across companies. For example, sectors like FMCG, retail, financial services, telecommunication etc. have a rich customer database and they care about different features of customers according to their CRM policies.

On the other hand, sectors that have a long inter-purchase time like durable consumer goods have limited information about customers. So, their assessments are different from other sectors. At this point, both business understanding and sector analysis become vital for the customer or market segmentation. In other words, to develop company specific solutions come into prominence.

As stated previously, eight different characteristics were defined in this study to evaluate and segment the customers. These characteristics were determined by a comprehensive literature review and a group of expert consisting of sales specialists and supervisor working in the company. Brain storming technique was used in deciding on the characteristics. Brainstorming should address a specific question (Osborn, 1953). In this study, the question is "What do you consider while characterizing a customer as important or valuable"? Group members answered this question and then using their answers and a literature survey, customer evaluation characteristics were defined. Definition and calculation of these characteristics are explained in the following. It is aimed to maximize the values of all characteristics. To show the computations, two customers were selected randomly. These two customers were named briefly as C165 and C299 according to their customer IDs. Before the calculations, it would be appropriate to explain "Length of relationship (LoR)". This is the number of years between the first transaction of a specific customer and the end of the observation period (2011).

4.5.1 Recency

Recency can be described as the date of last transaction of a specific customer within the observation period. It is important for companies since recent order indicates that the relationship is live. The value of recency is scaled from 1 to 7 and it equals to 1 for 2005 and 7 for 2011. For example, in Table 4.3, it can be seen that C165's last order was on 2011. Then, the recency value of this customer equals to 7. In the same way, C299's last order was on 2008 corresponding to recency value of 4.

Table 4.3 Recency values

	C165	C299
Year	Order	Order
2008	√	√
2009	-	-
2010	√	-
2011	√	-
Recency	*7*	*4*

4.5.2 Loyalty

The term customer loyalty is the behavior of repeat customers. Customers can be said "loyal" when they consistently purchase a certain product / product type or brand over a long period of time. The higher the value of LoR means the higher the probability that a customer stays loyal.

In this study, loyalty value is calculated by Equation 4.1. In this equation; active years mean the total number of years that transactions were conducted by a specific customer during its LoR. The main reason to multiply the ratio (active years / LoR) with active years is to distinguish between loyalty values of customers whose active years equal to their LoR. For example, a customer whose active years and LoR are 2 cannot be considered as loyal as a customer whose active years and LoR are 8.

As shown in Table 4.4, LoR of C165 equals to 4. Since the customer gave order in three of these four years, the active years equal to 3. Based on this information, this customer's loyalty value equals to 2.25. In the same manner, LoR of C299 equals to 4 and this customer gave an order only 1 of these 4 years. Then, the loyalty value is computed as 0.25 for this customer.

$$\text{Loyalty} = \frac{(\text{Active Years})^2}{LoR} \tag{4.1}$$

Table 4.4 Loyalty values

	C165	C299
Year	Order	Order
2008	√	√
2009	-	-
2010	√	-
2011	√	-
Loyalty	*2.25*	*0.25*

4.5.3 Average Annual Demand (AAD)

The value of this characteristic is determined by averaging demand of a customer during its LoR with the company. In other words, it is the ratio of total demand to LoR. Total demand of C165 is 4346 TVs. To obtain the average annual demand of this customer, we divide 4346 by 4 and obtain 1086.5 TV (see Table 4.5). In the same way, the average annual demand of C299 is computed as 541. The main reason to use this characteristic is to calculate comparable values for the customers with respect to total demand.

Table 4.5 AAD values

	C165	C299
Year	Demand	Demand
2008	150	2164
2009	0	0
2010	1931	0
2011	2265	0
AAD	*1086.5*	*541*

4.5.4 Average Annual Sales Revenue (AASR)

This characteristic can be defined as the average expenditure of the customer made during its LoR. It is the ratio of total sales revenue to LoR with the company. The reason to use this characteristic is to calculate comparable values for the customers with respect to total sales revenue.

Total sales revenue of C165 is 1960367.42 TL and average annual sales revenue is 490091.9 TL (see Table 4.6). Average annual sales revenue of C299 is 156861 TL.

Table 4.6 AASR values

	C165	C299
Year	Revenue	Revenue
2008	108117.5	627444
2009	0	0
2010	913414.8	0
2011	938835.1	0
AASR	*490091.9*	*156861*

4.5.5 Frequency

Frequency indicates the average number of transactions conducted per year. It is computed by multiplying the ratio of total number of months in which at least a transaction was conducted to total number of months between the first transaction and the end of observation period by 12. As reported in Table 4.7, C165's first order is in June 2008 and the customer gave 13 orders in 43 months. By using Eq. 4.2, the frequency values of C165 and C299 are computed as 3.63 and 0.32, respectively (see Tables 4.7 and 4.8 for the order schedules).

On the other hand according to, C299 has an order only 1 of 37 months, so the frequency value of this customer equals to 0.32 (Equation 4.3).

$$Frequency_{C165} = \left(\frac{13}{43}\right) \times 12 = 3.63 \tag{4.2}$$

$$Frequency_{C299} = \left(\frac{1}{37}\right) \times 12 = 0.32 \tag{4.3}$$

Table 4.7 Order schedule of C165

C 165				
Month	2008	2009	2010	2011
1	NA	-	-	√
2	NA	-	-	√
3	NA	-	√	√
4	NA	-	-	√
5	NA	-	-	-
6	√	-	√	-
7	-	-	√	√
8	-	-	-	-
9	-	-	√	-
10	-	-	-	-
11	-	-	√	-
12	-	-	√	√

*NA: Not available

Table 4.8 Order schedule of C299

C 299				
Month	2008	2009	2010	2011
1	NA	-	-	-
2	NA	-	-	-
3	NA	-	-	-
4	NA	-	-	-
5	NA	-	-	-
6	NA	-	-	-
7	NA	-	-	-
8	NA	-	-	-
9	NA	-	-	-
10	NA	-	-	-
11	NA	-	-	-
12	√	-	-	-

*NA: Not available

4.5.6 Long Term Relationship Potential (LTRP)

Long term relationship simply means building customer loyalty for the company. It is thought that if a customer works with the company for a long time or in other words its loyalty is high and at the same time has a recent order we may assume that

this customer will keep working with this company in the future. Long term relationship potential is calculated as a score by the following equation;

$$LTRP = \text{Loyalty} * \text{Recency} \quad (4.4)$$

For C165, loyalty value equals to 2.25 and recency value is 7. If we multiply these two values, LTRP of C165 is computed as 15.75 (see Table 4.9). In addition, LTRP of C299 is 1.

Table 4.9 LTRP values

	C165	C299
Year	Order	Order
2008	√	√
2009	-	-
2010	√	-
2011	√	-
LTRP	15.75	1

4.5.7 Average Percentage Change in Annual Demand (APCIAD)

Not only the amount of annual demand but also the variation in annual demand of a specific customer is important for companies. Therefore, a characteristic that indicates the average percentage change in annual demand is necessary. With this characteristic, we can see whether the annual customer demand increases or decreases. If we consider demand values in two consecutive years, respectively *a* and *b*, APCIAD is calculated as *(b-a) / a*. This characteristic equals to zero for customers whose first transaction was conducted in 2011. For instance, using the information given in Table 4.10, APCIAD can be computed for C165 by the following steps.

- *Step 1:* Compute the percentage change in annual demand between 2008 and 2009. (0-150) / (150) = -1.

- *Step 2:* Compute the percentage change in annual demand between 2009 and 2010. (0-1931) / (0) = error of division by zero.

The operations defined in step three should be carried out to avoid this error.

-*Step 3:* Compute the percentage change in annual demand between 2008 and 2010. Since demand value in 2009 is zero, the percentage change in annual demand between 2008 and 2010 should be calculated as (1931-150) / 150 = 11.87. This value should be divided by two (11.87 / 2 = 5.94) since this change occurs in two years.

-*Step 4:* Compute the percentage change in annual demand between 2010 and 2011. (2265-1931) / 1931 = 0.17.

-*Step 5:* Compute the average percentage change in annual demand.
APCIAD is computed as 3.05 (% 305) by taking the average of 5.94 and 0.17. As a result, we can say that the customer demand increases.

In the same way, APCIAD is computed as -0.33 for C299.

Table 4.10 Demand values

	C165	C299
Year	Demand	Demand
2008	150	2164
2009	0	0
2010	1931	0
2011	2265	0

4.5.8 Average Percentage Change in Annual Sales Revenue (APCIASR)

This characteristic is the measure of change in annual sales revenue. By evaluating the value of APCIASR we can see whether the annual sales revenue obtained from a particular customer increases or decreases. If we consider sales revenues of two consecutive years respectively, *c* and *d*, APCIASR is calculated by *(d-c)/c*. APCIASR equals to zero for customers whose first transaction was conducted in 2011. For instance, using the data reported in Table 4.11, APCIASR value for C165 can be computed by the following steps.

- *Step 1:* Compute the percentage change in annual sales revenue between 2008 and 2009.

(0-108117.5) / (108117.5) = -1.

- *Step 2:* Compute the percentage change in annual sales revenue between 2009 and 2010.

(0-913414.8) / (0) = error of division by zero.

The operations defined in step three should be carried out to avoid this error.

- *Step 3:* Compute the percentage change in annual sales revenue between 2008 and 2010.

Since sales revenue in 2009 is zero, the percentage change in annual sales revenue between 2008 and 2010 should be calculated as (913414.8-108117.5) / 108117.5 = 7.45. This value should be divided by two (7.45 / 2 = 3.72) since this change actually occurs in two years.

- *Step 4:* Compute the percentage change in annual sales revenue between 2010 and 2011.

(938835.1-913414.8) / 913414.8 = 0.03.

- *Step 5:* Compute the average percentage change in annual demand

APCIASR is computed as 1.876 (% 187.6) by taking the average of 3.72 and 0.03. As a result, it can be stated that C165's sales revenue increases.

In the same way, APCIASR is computed as -0.33 for C299.

Table 4.11 Sales revenues

	C165	C299
Year	Revenue	Revenue
2008	108117.5	627444
2009	0	0
2010	913414.8	0
2011	938835.1	0

After computation of the aforementioned characteristics for each customer, initial data set is obtained as shown in Table 4.12.

Table 4.12 Initial data set

Customer ID	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
165	7	2.25	1086.5	490091.86	3.63	15.75	3.055	1.876
299	4	0.25	541	156861	0.32	1	-0.33	-0.33

Initial data set is normalized to a range between 0 and 1 by using the min-max normalization on MINITAB 14.0. Min–max normalization works by seeing how much greater the field value is than the minimum value and scaling this difference by the range (Larose, 2005). The purpose of the normalization is to create a common scale for the characteristics and make them commensurate. Equation 4.5 was used to compute the standardized values of the characteristics. Let X_{ij} refer to original field value and r_{ij} refer to the normalized field value. Where, X_j^* is the best and X_j^- is the worst values of characteristic j.

$$r_{ij} = (X_{ij} - X_j^-)/(X_j^* - X_j^-) \tag{4.5}$$

The standardized data is reported in Table 4.13.

Table 4.13 Standardized data

Customer ID	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
165	1.000	0.236	0.002	0.003	0.291	0.246	0.119	0.019
299	0.500	0.009	0.001	0.001	0.010	0.011	0.023	0.006

4.6 Determining Importance Weights of the Characteristics Using Fuzzy AHP

Many real world decision problems have complexity and usually based on human judgments. In addition, human judgments are based on unclear linguistic assessments. The linguistic term is a variable whose values are words or phrase in natural or artificial language (Jamalnia & Soukhakian, 2009). Linguistic assessments are often characterized by fuzzy numbers (Li & Lai, 2001).

To deal with uncertainty of human thought, Zadeh (1965) first introduced the fuzzy set theory, which was oriented to the rationality of uncertainty due to imprecision or vagueness.

A fuzzy set is a class of object s with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function, which assigns to each object a grade of membership ranging between zero and one. Triangular fuzzy number (*TFN*), $\widetilde{M}$, is illustrated in Figure 4.4, and it is denoted simply as *M= (l, m, u).* The parameters *l*, *m*, and *u*, respectively denote the smallest possible value, the most promising value, and the largest possible value that describe a fuzzy event (Kahraman et al., 2004).

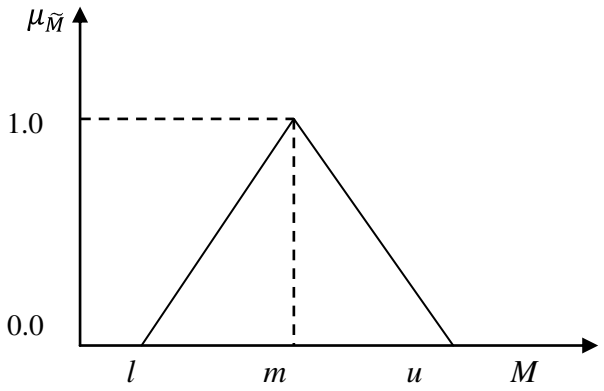

Figure 4.4 A triangular fuzzy number

Since the fuzzy approaches allow more accurate description of human judgments, fuzzy AHP method is a well known and effective tool for MCDM problems. In this study, fuzzy AHP approach is used to determine the importance weights of the customer characteristics.

Each characteristic is evaluated by the sales supervisor by stating the importance level of the characteristics using linguistic variables and filling the pair wise comparison matrix. L= {VHI, HI, SHI, M, SLI, LI, VLI} is defined as a set of linguistic values where VHI=Very High Important, HI=High Important, SHI=Somewhat High Important, M=Medium, SLI=Somewhat Low Important, LI=Low Important, VLI=Very Low Important. Linguistic values and corresponding triangular fuzzy numbers are given in Table 4.14. Membership functions for linguistic values are presented in Figure 4.5.

Table 4.14 Linguistic variables for the importance of customer evaluation characteristics

Linguistic Variables	Triangular Fuzzy Numbers
Very low important (VLI)	(0, 0, 0.10)
Low important (LI)	(0.05, 0.15, 0.25)
Somewhat low important (SLI)	(0.20, 0.325, 0.45)
Medium (M)	(0.40, 0.50, 0.60)
Somewhat high important (SHI)	(0.55, 0.675, 0.80)
High important (HI)	(0.75, 0.85, 0.95)
Very high important (VHI)	(0.90, 1, 1)

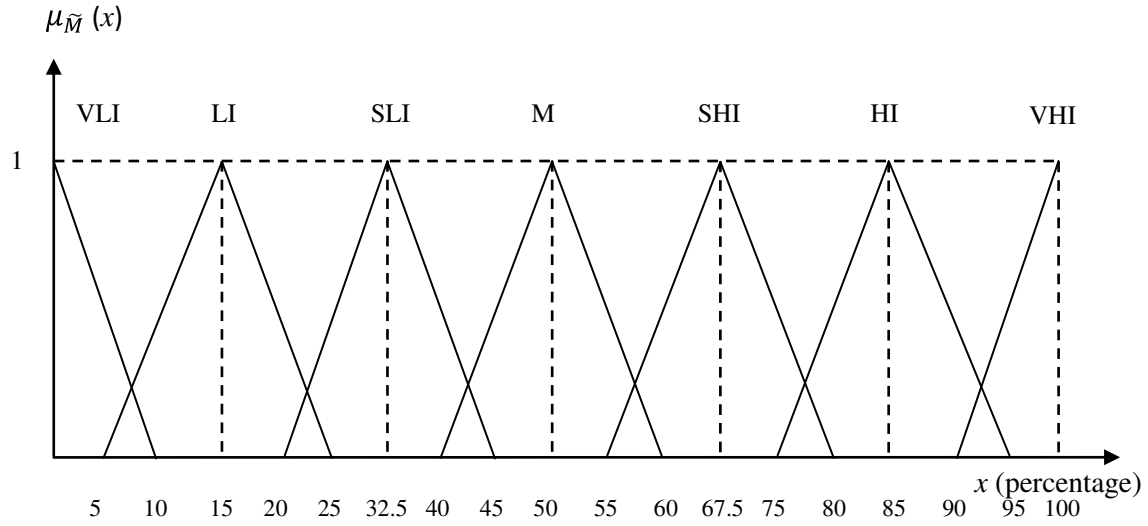

Figure 4.5 Membership functions for linguistic variables (Belmokaddem et al., 2009)

The computational procedure of Chang's (1992) extended fuzzy AHP is described as follows:

- *Step 1*: The value of fuzzy synthetic extent with respect to the *i*th object is defined by:

$$S_i = \sum_{j=1}^{m} M_{gi}^{j} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{gi}^{j} \right]^{-1} \tag{4.6}$$

where, all M_{gi}^{j} $(j = 1,2, \ldots . m)$ are triangular fuzzy numbers.

- *Step 2*: As $M_1 = (l_1, m_1, u_1)$ and $M_2 = (l_2, m_2, u_2)$ are two triangular fuzzy numbers, the degree of possibility of M1 ≥ M2 is defined as:

$$V(M_1 \geq M_2) = \begin{cases} 1, & if \quad m_1 \geq m_2 \\ 0, & if \quad l_2 \geq u_1 \\ \dfrac{l_2 - u_1}{(m_1 - u_1) - (m_2 - l_2)}, & otherwise \end{cases} \tag{4.7}$$

- *Step 3*: To compare M_1 and M_2, we need both the values of $V(M_1 \geq M_2)$ and $V(M_2 \geq M_1)$. The degree possibility for a convex fuzzy number to be greater than k convex fuzzy numbers M_i $(i = 1,2. \ldots . k)$ can be defined by;

$$V(M \geq M_1 M_2 M_3 \ldots M_k) = V\left[(M \geq M_1) \; and \; \left(M \geq M_2 \; and \ldots and \; (M \geq M_k)\right)\right] \tag{4.8}$$
$$= \min V(M \geq M_i), \; i = 1,2,3 \ldots . k.$$

Assume that $d'(A_i) = \min V(S_i \geq S_k)$ for $k = 1,2. \ldots n; k \neq i$, then the weight vector is given by:

$$W' = (d'(A_1), d'(A_2). \ldots . . d'(A_n))^T \tag{4.9}$$

where $A_i (i = 1,2. \ldots . . n)$ are n elements.

- *Step 4*: Via normalization, the normalized weight vectors are:

$$W = (d(A_1), d(A_2), \ldots . . d(A_n))^T \tag{4.10}$$

where, *W* is a crisp number.

As reported in Table 4.15, customer evaluation characteristics are referred as F_i (i = 1, 2 …8). The pair-wise comparison matrix is presented in Table 4.16. In addition, fuzzy evaluation matrix is given in Appendix B.

Table 4.15 Notations for the characteristics

Characteristic	Definition
F1	Recency
F2	Loyalty
F3	Average Annual Demand
F4	Average Annual Sales Revenue
F5	Frequency
F6	Long Term Relationship Potential
F7	Average Percentage Change in Annual Demand
F8	Average Percentage Change in Annual Sales Revenue

Table 4.16 The pair-wise comparison matrix

	F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*
F1	M	SLI	LI	LI	SHI	SLI	SHI	M
F2	SHI	M	SLI	LI	SHI	SLI	SHI	M
F3	HI	SHI	M	SLI	HI	SHI	SHI	SHI
F4	HI	HI	SHI	M	HI	SHI	HI	SHI
F5	SLI	SLI	LI	LI	M	M	SLI	M
F6	SHI	SHI	SLI	SLI	M	M	HI	SHI
F7	SLI	SLI	SLI	LI	SHI	LI	M	SLI
F8	M	M	SLI	SLI	M	SLI	SHI	M

The value of fuzzy synthetic extent is calculated as follows:

$$S_{Z_1} = (2.4, 3.3, 4.2) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.061, 0.103, 0.162)$$

$$S_{Z_2} = (2.9, 3.825, 4.75) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.074, 0.119, 0.184)$$

$$S_{Z_3} = (4.3, 5.225, 6.15) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.109, 0.163, 0.238)$$

$$S_{Z_4} = (5.05, 5.925, 6.8) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.129, 0.185, 0.263)$$

$$S_{Z_5} = (1.9, 2.775, 3.65) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.048, 0.087, 0.141)$$

$$S_{Z_6} = (3.6, 4.525, 5.45) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.09, 0.141, 0.211)$$

$$S_{Z_7} = (1.85, 2.775, 3.7) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.047, 0.087, 0.143)$$

$$S_{Z_8} = (2.75, 3.65, 4.55) \otimes \left(\frac{1}{39.25}, \frac{1}{32}, \frac{1}{25.85}\right) = (0.07, 0.114, 0.176)$$

Then the fuzzy values are compared;

$V(S_{Z_1} \geq S_{Z_2}) = 0.844$ $V(S_{Z_2} \geq S_{Z_1}) = 1$ $V(S_{Z_3} \geq S_{Z_1}) = 1$

$V(S_{Z_1} \geq S_{Z_3}) = 0.468$ $V(S_{Z_2} \geq S_{Z_3}) = 0.629$ $V(S_{Z_3} \geq S_{Z_2}) = 1$

$V(S_{Z_1} \geq S_{Z_4}) = 0.292$ $V(S_{Z_2} \geq S_{Z_4}) = 0.456$ $V(S_{Z_3} \geq S_{Z_4}) = 0.833$

$V(S_{Z_1} \geq S_{Z_5}) = 1$ $V(S_{Z_2} \geq S_{Z_5}) = 1$ $V(S_{Z_3} \geq S_{Z_5}) = 1$

$V(S_{Z_1} \geq S_{Z_6}) = 0.649$ $V(S_{Z_2} \geq S_{Z_6}) = 0.808$ $V(S_{Z_3} \geq S_{Z_6}) = 1$

$V(S_{Z_1} \geq S_{Z_7}) = 1$ $V(S_{Z_2} \geq S_{Z_7}) = 1$ $V(S_{Z_3} \geq S_{Z_7}) = 1$

$V(S_{Z_1} \geq S_{Z_8}) = 0.894$ $V(S_{Z_2} \geq S_{Z_8}) = 1$ $V(S_{Z_3} \geq S_{Z_8}) = 1$

$V(S_{Z_4} \geq S_{Z_1}) = 1$ $V(S_{Z_5} \geq S_{Z_1}) = 0.830$ $V(S_{Z_6} \geq S_{Z_1}) = 1$

$V(S_{Z_4} \geq S_{Z_2}) = 1$ $V(S_{Z_5} \geq S_{Z_2}) = 0.672$ $V(S_{Z_6} \geq S_{Z_2}) = 1$

$V(S_{Z_4} \geq S_{Z_3}) = 1$ $V(S_{Z_5} \geq S_{Z_3}) = 0.292$ $V(S_{Z_6} \geq S_{Z_3}) = 0.822$

$V(S_{Z_4} \geq S_{Z_5}) = 1$ $V(S_{Z_5} \geq S_{Z_4}) = 0.113$ $V(S_{Z_6} \geq S_{Z_4}) = 0.653$

$V(S_{Z_4} \geq S_{Z_6}) = 1$ $V(S_{Z_5} \geq S_{Z_6}) = 0.475$ $V(S_{Z_6} \geq S_{Z_5}) = 1$

$V(S_{Z_4} \geq S_{Z_7}) = 1$ $V(S_{Z_5} \geq S_{Z_7}) = 1$ $V(S_{Z_6} \geq S_{Z_7}) = 1$

$V(S_{Z_4} \geq S_{Z_8}) = 1$ $V(S_{Z_5} \geq S_{Z_8}) = 0.722$ $V(S_{Z_6} \geq S_{Z_8}) = 1$

$V(S_{Z_7} \geq S_{Z_1}) = 0.833$ $V(S_{Z_8} \geq S_{Z_1}) = 1$

$V(S_{Z_7} \geq S_{Z_2}) = 0.679$ $V(S_{Z_8} \geq S_{Z_2}) = 0.949$

$V(S_{Z_7} \geq S_{Z_3}) = 0.305$ $V(S_{Z_8} \geq S_{Z_3}) = 0.575$

$V(S_{Z_7} \geq S_{Z_4}) = 0.128$ $V(S_{Z_8} \geq S_{Z_4}) = 0.400$

$V(S_{Z_7} \geq S_{Z_5}) = 1$ $V(S_{Z_8} \geq S_{Z_5}) = 1$

$V(S_{Z_7} \geq S_{Z_6}) = 0.485$ $V(S_{Z_8} \geq S_{Z_6}) = 755$

$V(S_{Z_7} \geq S_{Z_8}) = 0.728$ $V(S_{Z_8} \geq S_{Z_7}) = 1$

The priority weights are calculated as follows:

$d'(Z_1) = \min(0.844, 0.468, 0.292, 1, 0.649, 1, 0.894) = 0.292$

$d'(Z_2) = \min(1, 0.629, 0.456, 1\ ,0.808, 1, 1) = 0.456$

$d'(Z_3) = \min(1, 1, 0.833, 1, 1, 1, 1) = 0.833$

$d'(Z_4) = \min(1, 1, 1,1, 1, 1, 1) = 1$

$d'(Z_5) = \min(0.830, 0.672, 0.292, 0.113, 0.475, 1, 0.722) = 0.113$

$d'(Z_6) = \min(1, 1, 0.822, 0.653, 1, 1, 1) = 0.653$

$d'(Z_7) = \min(0.833, 0.679, 0.305, 0.128, 1, 0.485, 0.728) = 0.128$

$d'(Z_8) = \min(1, 0.949, 0.575, 0.4, 1, 0.755, 1) = 0.4$

Priority weight vector can be stated as follows: W'=(= (0.292, 0.456, 0.833, 1, 0.113, 0.653, 0.128, 0.4). The normalized weight vector is computed as W= (0.075, 0.118, 0.215, 0.258, 0.03, 0.168, 0.033, 0.103).

As reported in Table 4.17 and Figure 4.6, the most important characteristic is "*average annual sales revenue"* with importance weight of 0.258. On the other hand, the least important characteristic is "*frequency"* with the weight of 0.03. The reason why the frequency is the least important characteristic is related with the sector in which the company operates. The company produces TV that is a kind of durable consumer goods. Both production and distribution processes are time consuming and costly. In addition, most of the customers located in geographically dispersed countries over the world and they are far from the production facility. Therefore, the company does not prefer frequent orders. Instead of frequent, low volume orders they preferred less frequent, high volume orders to avoid high setup, production and distribution costs.

Table 4.17 Importance weights of the characteristics

Characteristic	Definition	Importance Weight
F1	Recency	0.075
F2	Loyalty	0.118
F3	Average Annual Demand	0.215
F4	Average Annual Sales Revenue	0.258
F5	Frequency	0.030
F6	Long Term Relationship Potential	0.168
F7	Average Percentage Change in Annual Demand	0.033
F8	Average Percentage Change in Annual Sales Revenue	0.103

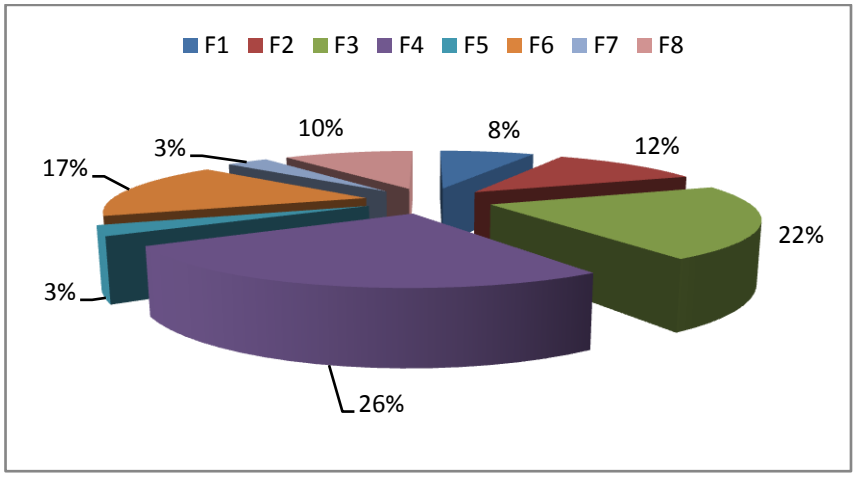

Figure 4.6 Importance weights of the characteristics

4.7 Segmenting Customers via Data Clustering Algorithms

In the application part of this study, customers are grouped using two different approaches. First approach clusters customers according to a combined characteristic that merges all the characteristics in one dimension. The second one considers all of the mentioned characteristics and then clusters customers according to these predefined characteristics.

Recall that the aim of this study is to divide customers into small manageable groups. Determining the number of clusters is important for the management of these groups and planning of the sales & marketing activities. Increasing the number of clusters make the management more difficult. Therefore, the numbers of clusters are selected as 3, 5 and 7 in implementation of the clustering algorithms.

4.7.1 Single Dimension (SD) - Based Customer Segmentation

In this section, customers were grouped in terms of their overall scores and it is looked for natural breakpoints in the data set. Herein, the data set is one dimensional, in other words it is a vector with 317 rows and 1 column. The rows represent the customers and the column represents the overall scores of the customers.

Such a composite score is computed for the following purposes:

- To reduce the size of the dataset without losing the essential characteristic of the dataset,
- To reflect the priorities of the companies for the characteristics,
- To partition the customers into comparable groups according to their importance levels.

Overall score takes values between 0 and 1, and it is computed for each customer by using Equation 4.11, and presented in Table 4.18.

$$Overall\ Score_i = \sum_{j=1}^{8} w_j F_{ij} \tag{4.11}$$

where *i* is the customer ID *(i= 1, 2…317)* and *j* is the number of characteristics *(j=1, 2...8)*, w_j is the importance weight of characteristic *j* and F_{ij} is the value of characteristic *j* of customer *i*.

Table 4.18 Overall scores of the customers

Customer ID	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*Overall Score*
165	1	0.236	0.002	0.003	0.291	0.246	0.119	0.019	*0.160*
299	0.5	0.009	0.001	0.001	0.010	0.011	0.023	0.006	*0.043*

After computing overall score for each customer, different data clustering techniques that involved in XLSTAT 2012 statistical package were used to segment the customers. XLSTAT is a Microsoft Excel statistical add-in that has been developed since 1993 to enhance the analytical capabilities of Excel. XLSTAT relies on Excel for the input of data and the display of results, but the computations are done using autonomous software components. The use of Excel as an interface makes XLSTAT a user-friendly and highly efficient statistical and multivariate data analysis package (XLSTAT).

4.7.1.1 SD - Based Customer Segmentation using AHC Algorithms

Among the SD-based customer segmentation algorithms, at first, AHC algorithms were used in this study. Herein, Euclidean distance was chosen as the dissimilarity metric. Since the data set has single dimension, distance between two points equals to their numerical difference. Thus, if *x* and *y* are two points on the real line, then the distance between them is given by;

$$d(x,y) = \sqrt{(x-y)^2} = |x-y| \tag{4.13}$$

In the application phase, the data was clustered according to Ward's, single linkage and complete linkage methods. It is known that clustering algorithms aim to obtain clusters of objects such that similarity within a cluster is high while similarity between clusters is very low. In this study, *SSW* is used to evaluate the validity of the clustering results.

Smaller value of *SSW* indicates "better" clustering. Results of AHC algorithms are presented in Table 4.19. It is clear that Ward's agglomeration method provides superior results than the others. After determining the best method, we should determine the number of clusters. It is obvious that when the number of clusters increases, *SSW* will decrease. However, there must be a balance between the number of clusters and *SSW*. A clustering with small number of clusters and low value of *SSW* is treated as adequate. Figure 4.7 visualize *SSW* values for AHC methods for different number of clusters. The results reveal that *SSW* value of Ward's method is subject to small changes for the number of clusters bigger than 5. Therefore, we can conclude that segmenting customers into five clusters is proper.

Table 4.19 Sum of squares of AHC algorithms (SD)

Agglomeration Method	*SSW*	*SSB*	*TSS*	*SSW/TSS*	# of clusters
Ward's method	1.082	3.190	4.272	0.253	3
Ward's method	0.316	3.956	4.272	0.074	5
Ward's method	0.149	4.122	4.272	0.035	7
Single Linkage	2.303	1.968	4.272	0.539	3
Single Linkage	1.679	2.593	4.272	0.393	5
Single Linkage	1.672	2.600	4.272	0.391	7
Complete Linkage	0.920	3.352	4.272	0.215	3
Complete Linkage	0.692	3.579	4.272	0.162	5
Complete Linkage	0.220	4.051	4.272	0.052	7

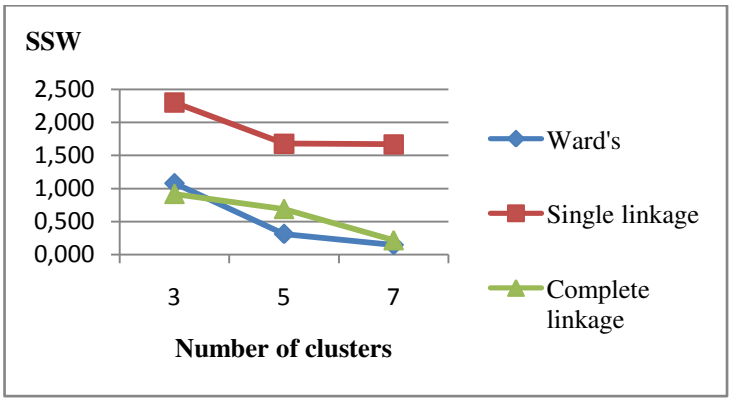

Figure 4.7 *SSW* values of AHC algorithms (SD)

Results of Ward's AHC method are presented in the following. The number of observations, minimum, maximum and mean values of the overall score, and standard deviation of the overall score are reported in Table 4.20. Herein, number of observations is equal to 317, which is the number of customers. Minimum and maximum values of the overall score are 0.015 and 0.837, respectively. The mean and standard deviation of the overall score are 0.148 and 0.116, respectively.

Table 4.20 Summary statistics of the dataset (SD)

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Overall Score	317	0.015	0.837	0.148	0.116

Table 4.21 shows the centroids of the clusters. Centroid is the mean value of overall scores of the customers that assigned to a specific cluster. For the SD data set, importance of the segments was computed according to their centroids (mean value of the overall score) and the higher the value of centroid indicates the segment is of greater importance. The clusters were ranked from 1 to 5 where, 1 indicates the most important and 5 indicates the least important clusters. As illustrated in Figure 4.8, cluster three has the highest mean value in terms of the overall score. So, we can conclude that the most important cluster is cluster three. On the other hand, cluster five has the lowest mean value in terms of the overall score. So, it is the least important cluster.

Table 4.21 Cluster centroids and ranks (Ward's - SD)

Cluster	Overall Score	Rank
1	0.121	4
2	0.188	3
3	0.601	1
4	0.299	2
5	0.055	5

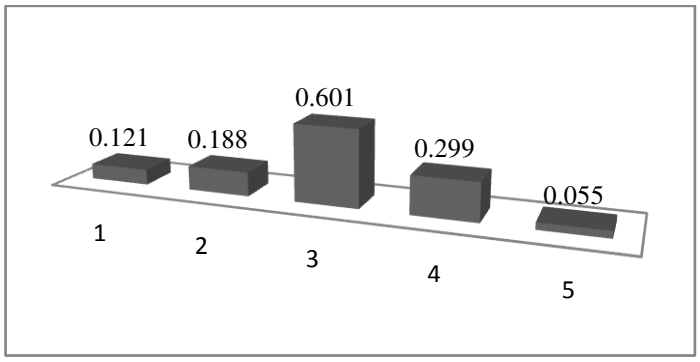

Figure 4.8 Cluster centroids (Ward's - SD)

Distances between cluster centroids are presented in Table 4.22. Max distance is occurred between cluster three and five with the value of 0.547. This value is computed as the difference between the centroid values of these two clusters *(0.601-0.055=0.547)*. It means that these two clusters are the most dissimilar clusters with respect to overall score.

Table 4.22 Distances between cluster centroids (Ward's - SD)

	1	2	3	4	5
1	0	0.067	0.480	0.178	0.067
2	0.067	0	0.413	0.111	0.134
3	0.480	0.413	0	0.302	0.547
4	0.178	0.111	0.302	0	0.244
5	0.067	0.134	0.547	0.244	0

Central object is defined as the nearest object to the centroid of a specific cluster. As reported in Table 4.23, C38, which is a member of cluster one is the nearest object to the centroid of cluster 1. Table 4.24 indicates the distances between central objects.

Table 4.23 Central objects (Ward's - SD)

Cluster	Overall Score
1 (38)	0.121
2 (97)	0.189
3 (290)	0.620
4 (209)	0.297
5 (240)	0.056

Table 4.24 Distances between the central objects (Ward's - SD)

	1 (38)	2 (97)	3 (290)	4 (209)	5 (240)
1 (38)	0	0.067	0.499	0.176	0.065
2 (97)	0.067	0	0.432	0.108	0.132
3 (290)	0.499	0.432	0	0.323	0.564
4 (209)	0.176	0.108	0.323	0	0.241
5 (240)	0.065	0.132	0.564	0.241	0

Table 4.25 defines the clusters by using the number of objects, minimum distance to the centroid, maximum distance to the centroid and mean distance to the centroid. 86 of 317 customers are assigned to Cluster 1, 66 of them are assigned to Cluster 2, 9 of them are assigned to Cluster 3, 41 of them are assigned to Cluster 4 and finally, 115 of them are assigned to Cluster 5 by the clustering process.

Table 4.25 Results of Ward's AHC for five clusters (SD)

Cluster	1	2	3	4	5
Number of objects	86	66	9	41	115
Minimum distance to centroid	0.000	0.000	0.019	0.002	0.001
Average distance to centroid	0.014	0.019	0.074	0.045	0.016
Maximum distance to centroid	0.031	0.043	0.236	0.155	0.040

Scatter plot of the customer groups for the 5-cluster segmentation is illustrated in Figure 4.9.

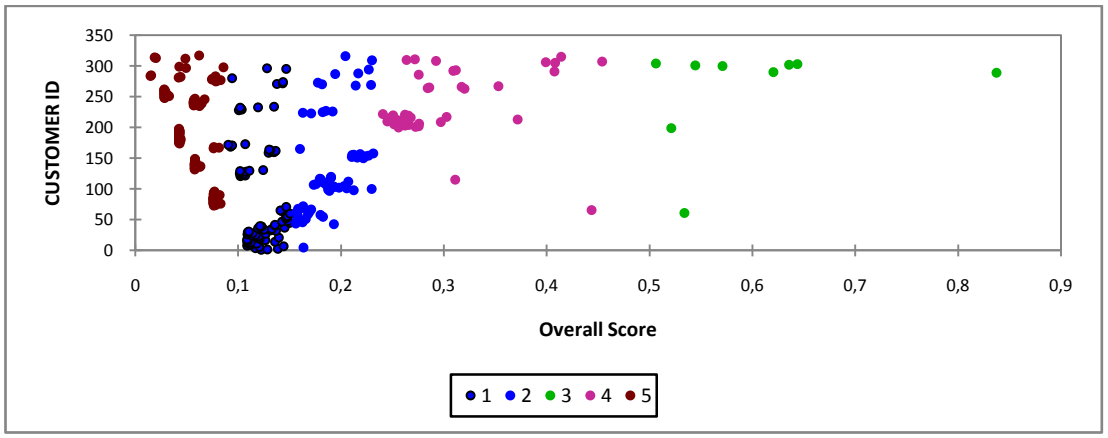

Figure 4.9 Scatter plot of groups (Ward's - SD)

In Table 4.26, the maximum, minimum and average values of the clusters with respect to "recency", "loyalty", "AAD", "AASR", "frequency", "LTRP", "APCIAD" and "APCIASR" are presented. Considering the average values of the characteristics for each cluster, we can state that there is big gap between cluster three and other clusters in terms of AAD (*F3*) and AASR (*F4*). In addition, cluster three has dominance on the other clusters with respect to all characteristics. On the other hand, clusters three and five are the farthest clusters, in other words, the most different clusters.

Table 4.26 Characteristics of the groups (Ward's - SD)

		F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*
Cluster 1	avg	0.922	0.141	0.003	0.003	0.287	0.139	0.038	0.009
	max	1.000	0.335	0.023	0.023	1.000	0.250	0.134	0.043
	min	0.667	0.094	0.000	0.000	0.028	0.090	0.000	0.000
Cluster 2	avg	0.972	0.317	0.010	0.011	0.467	0.316	0.110	0.021
	max	1.000	0.597	0.079	0.094	1.000	0.497	0.712	0.094
	min	0.667	0.094	0.000	0.001	0.100	0.106	0.009	0.003
Cluster 3	avg	0.918	0.673	0.522	0.713	0.865	0.523	0.133	0.060
	max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.366
	min	1.000	0.207	0.176	0.278	0.790	0.218	0.047	0.009
Cluster 4	avg	0.992	0.607	0.045	0.053	0.616	0.606	0.128	0.053
	max	1.000	1.000	0.491	0.652	1.000	1.000	0.563	1.000
	min	0.833	0.207	0.000	0.001	0.166	0.218	0.031	0.007
Cluster 5	avg	0.559	0.038	0.000	0.001	0.051	0.030	0.025	0.006
	max	0.833	0.182	0.004	0.007	0.312	0.090	0.162	0.033
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.002

4.7.1.2 SD - Based Customer Segmentation using k-means Algorithm

k-means algorithm requires the user to specify the number of clusters to be formed (*k* clusters). The numbers of clusters were determined as 3, 5 and 7 respectively, in the applications performed in this study. Herein, determinant (W) is selected as the clustering criterion. Table 4.27 and Figure 4.10 show the *SSW* values for different number of clusters. The results reveal that there is a substantial difference between the *SSW* values of the solutions with 3 and 5 clusters. However, *SSW* value of the k-means algorithm is subject to small changes for the number of clusters bigger than 5. Therefore, we can conclude that five clusters are proper to define the segments of customers in terms of k-means algorithm.

Table 4.27 Sum of squares of k-means algorithm (SD)

Clustering Criterion	*SSW*	*SSB*	*TSS*	*SSW/TSS*	# of clusters
Determinant (W)	0.732	3.539	4.272	0.171	3
Determinant (W)	0.307	3.965	4.272	0.072	5
Determinant (W)	0.173	4.099	4.272	0.040	7

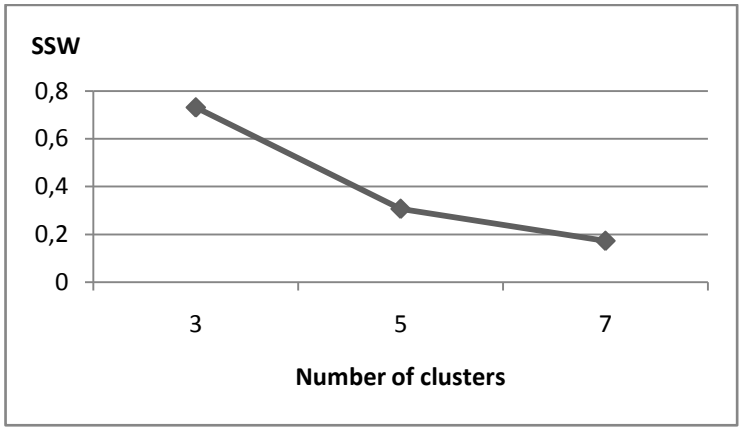

Figure 4.10 *SSW* values of k-means (SD)

k-means starts by selecting *k* initial records randomly as cluster centroids (initial centers) and assigns each record to its ‘‘nearest’’ cluster. As new records are added to the clusters, the cluster centroids are recalculated to reflect their new members. Then, cases are reassigned to the adjusted clusters. Then, final cluster centroids obtained after an iterative process. Table 4.28 reports the randomly selected initial cluster centroids, and Table 4.29 indicates the final cluster centroids.

Table 4.28 Initial cluster centroids (k-means - SD)

Cluster	Overall Score
1	0.160
2	0.141
3	0.159
4	0.126
5	0.160

Table 4.29 Final cluster centroids and ranks (k-means - SD)

Cluster	Overall Score	Rank
1	0.121	4
2	0.188	3
3	0.574	1
4	0.055	5
5	0.292	2

If we consider the final cluster centroids, it can be stated that cluster three has the highest overall score (0.574). So, it is the most important cluster among the clusters with the rank of 1. On the other hand, the least important cluster is cluster four with the rank of 5.

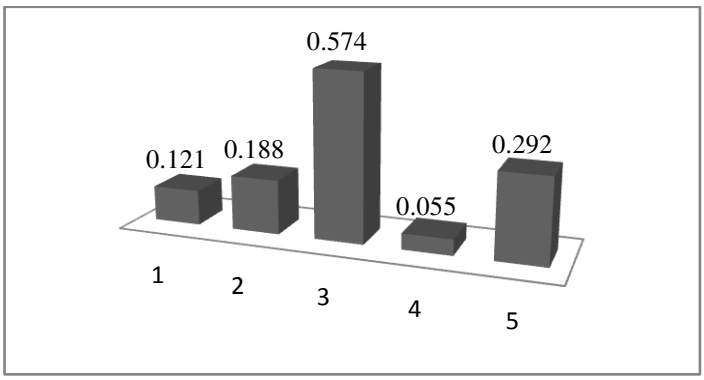

Figure 4.11 Cluster centroids (k-means - SD)

It is known that k-means consists of an iterative procedure. Iteration summary of the algorithm is given in Table 4.30 and Figure 4.12. Determinant (W) reached the minimum value (0.307) in the fourth iteration and then the algorithm is terminated.

Table 4.30 Statistics for the iterations (k-means - SD)

Iteration	Within-cluster variance	Trace(W)	Determinant(W)
0	0.013	4.209	4.209
1	0.004	1.276	1.276
2	0.001	0.326	0.326
3	0.001	0.307	0.307
4	0.001	0.307	0.307

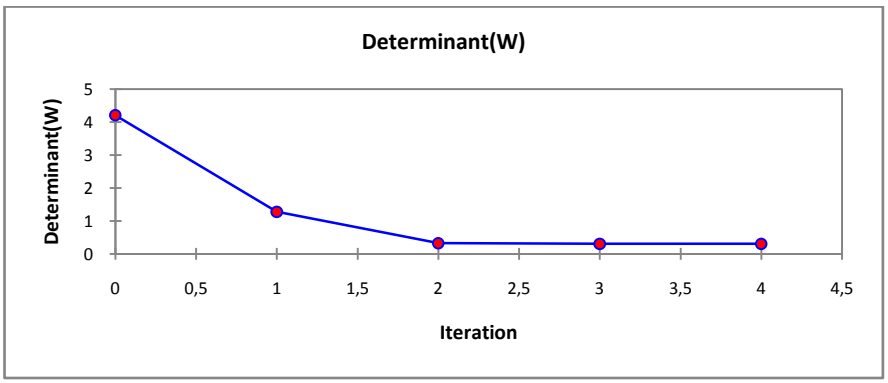

Figure 4.12 Determinant (W) over the iterations (k-means - SD)

In the options, tab number of repetitions is chosen as 10 in order to increase the quality and the stability of the results. Optimization summary can be found in Table 4.31. The first repetition gives the best result with minimum Determinant (W) value of 0.307.

Table 4.31 Optimization summary for k-means (SD)

Repetition	Iteration	Initial within-cluster variance	Final within-cluster variance	Determinant(W)
1	**4**	**0.013**	**0.001**	**0.307**
2	4	0.014	0.001	0.340
3	4	0.014	0.001	0.340
4	4	0.013	0.001	0.329
5	4	0.013	0.001	0.336
6	4	0.013	0.001	0.336
7	4	0.014	0.001	0.327
8	4	0.013	0.001	0.328
9	4	0.014	0.001	0.330
10	4	0.014	0.001	0.336

Table 4.32 defines the clusters. As reported in the Table 4.32; 86 of 317 customers are assigned to Cluster1, 66 of them are assigned to Cluster 2, 11 of them assigned to Cluster 3, 115 of them are assigned to Cluster 4 and 39 of them are assigned to Cluster 5.

Table 4.32 Results of k-means for five clusters (SD)

Cluster	1	2	3	4	5
Number of objects	86	66	11	115	39
Minimum distance to centroid	0.000	0.000	0.003	0.001	0.000
Average distance to centroid	0.014	0.019	0.080	0.016	0.037
Maximum distance to centroid	0.031	0.043	0.264	0.040	0.122

Table 4.33 shows the distances between the cluster centroids. Maximum distance is occurred between clusters three and four with the value of 0.519. This value is computed as the difference between the centroids of these clusters *(0.574 - 0.055=0.519)*. It means that these two clusters are the most dissimilar clusters with respect to the overall score.

Table 4.33 Distances between the cluster centroids (k-means - SD)

	1	2	3	4	5
1	0	0.067	0.452	0.067	0.170
2	0.067	0	0.385	0.134	0.103
3	0.452	0.385	0	0.519	0.282
4	0.067	0.134	0.519	0	0.237
5	0.170	0.103	0.282	0.237	0

Central objects are presented in Table 4.34. In addition, Table 4.35 shows the distances between the central objects.

Table 4.34 Central objects (k-means - SD)

Cluster	*Overall Score*
1 (38)	0.121
2 (97)	0.189
3 (300)	0.571
4 (240)	0.056
5 (308)	0.292

Table 4.35 Distances between the central objects (k-means - SD)

	1 (38)	2 (97)	3 (300)	4 (240)	5 (308)
1 (38)	0	0.067	0.449	0.065	0.171
2 (97)	0.067	0	0.382	0.132	0.104
3 (300)	0.449	0.382	0	0.514	0.279
4 (240)	0.065	0.132	0.514	0	0.236
5 (308)	0.171	0.104	0.279	0.236	0

Scatter plot of the customer groups for the 5-cluster segmentation is illustrated in Figure 4.13.

Figure 4.14 and Table 4.36 denotes the valuation of groups under each characteristic. Figure 4.14 reveals that there is big gap between cluster three and the other clusters in terms of AAD (*F3*) and AASR (*F4*). In addition, cluster three dominates other clusters with respect to all of the characteristics. Clusters three and four are the farthest, in other words, most different clusters. All clusters except cluster three has similar values in terms of AAD (*F3*) and AASR (*F4*).

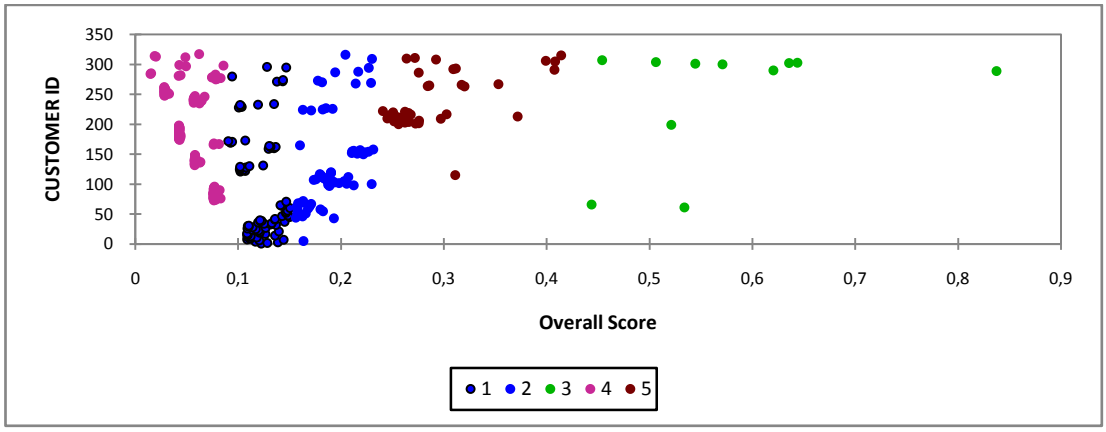

Figure 4.13 Scatter plot of the groups (k-means - SD)

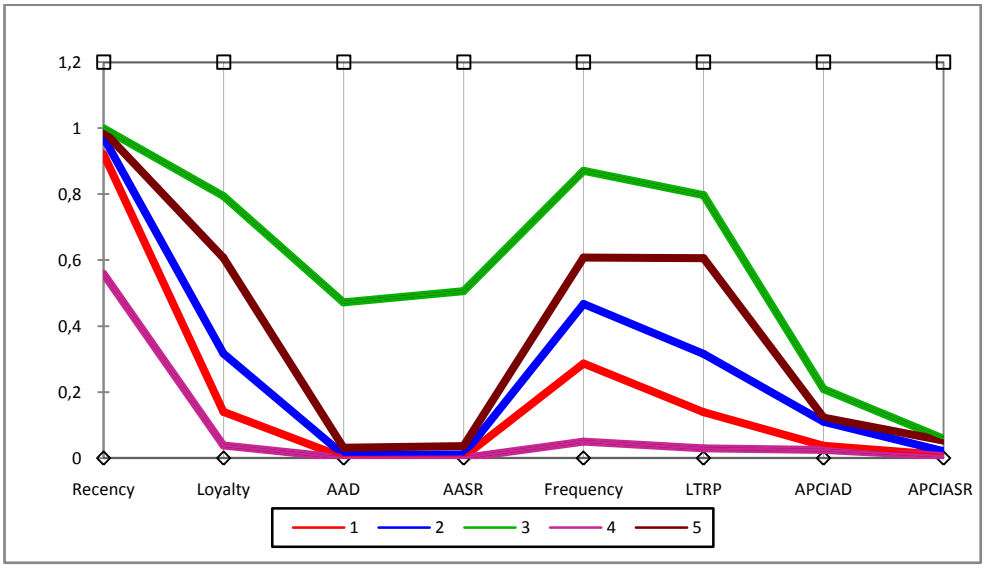

Figure 4.14 Group averages in terms of each characteristic (k-means - SD)

Table 4.36 Characteristics of the groups (k-means - SD)

		F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*
Cluster 1	avg	0.922	0.141	0.003	0.003	0.287	0.139	0.038	0.009
	max	1.000	0.335	0.023	0.023	1.000	0.250	0.134	0.043
	min	0.667	0.094	0.000	0.000	0.028	0.090	0.000	0.000
Cluster 2	avg	0.972	0.317	0.010	0.011	0.467	0.316	0.110	0.021
	max	1.000	0.597	0.079	0.094	1.000	0.497	0.712	0.094
	min	0.667	0.094	0.000	0.001	0.100	0.106	0.009	0.003
Cluster 3	avg	1.000	0.794	0.472	0.506	0.871	0.797	0.209	0.059
	max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.366
	min	1.000	0.207	0.117	0.114	0.748	0.218	0.047	0.009
Cluster 4	avg	0.559	0.038	0.000	0.001	0.051	0.030	0.025	0.006
	max	0.833	0.182	0.004	0.007	0.312	0.090	0.162	0.033
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.002
Cluster 5	avg	0.991	0.607	0.032	0.036	0.607	0.606	0.123	0.053
	max	1.000	1.000	0.244	0.261	1.000	1.000	0.563	1.000
	min	0.833	0.321	0.000	0.001	0.166	0.330	0.031	0.007

4.7.2 *Multiple Dimensions (MD) - Based Customer Segmentation*

The analyses conducted in the previous sections are consider single dimensional (n by 1) dataset. In this section, different approach was used to group customers and compare the importance of groups. First, customers were clustered according to their predefined characteristics (317×8 data set). Then, importance levels of the segments are determined by taking weighted average of the cluster centroids.

4.7.2.1 MD - Based Customer Segmentation using AHC Algorithms

In this section, customers are first clustered using Ward's, single linkage and complete linkage methods. Herein, Euclidean distance is used as dissimilarity metric. Results of AHC algorithms are presented in Table 4.37. The results reveal that Ward's agglomeration method gave better results. Figure 4.15 shows the *SSW* values for AHC methods for different number of clusters.

The results reveal that *SSW* value of Ward's method is subject to small changes for the number of clusters bigger than 5. Therefore, we can conclude that segmenting customers into five clusters is proper.

Table 4.37 Sum of squares of AHC algorithms (MD)

Agglomeration Method	*SSWC*	*SSBC*	*TSS*	*SSWC/TSS*	# of clusters
Ward's method	37.111	54.295	91.406	0.406	3
Ward's method	23.777	67.628	91.406	0.260	5
Ward's method	17.389	74.017	91.406	0.190	7
Single Linkage	87.226	4.180	91.406	0.954	3
Single Linkage	81.470	9.936	91.406	0.891	5
Single Linkage	80.076	11.329	91.406	0.876	7
Complete Linkage	61.218	30.188	91.406	0.670	3
Complete Linkage	32.055	59.351	91.406	0.351	5
Complete Linkage	29.783	61.623	91.406	0.326	7

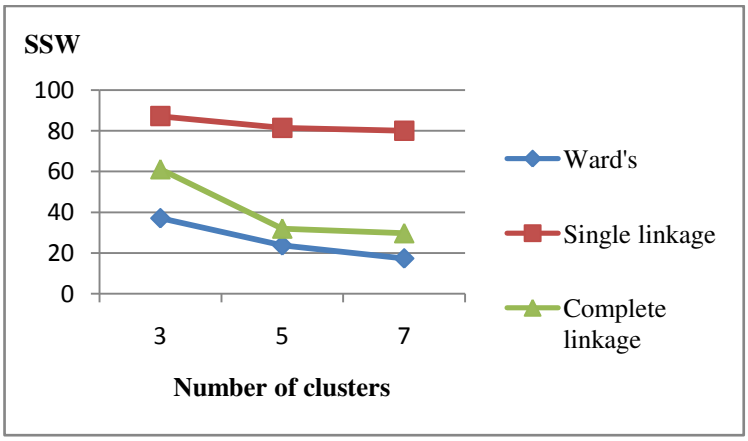

Figure 4.15 *SSW* values of AHC algorithms (MD)

The summary statistics presented in Table 4.38 show the number of observations, minimum, maximum, and mean values of the characteristics. Because of the standardization, minimum and maximum values are 0 and 1, respectively, for each characteristic.

Table 4.38 Summary statistics of the characteristics (MD)

Characteristic	Observations	Minimum	Maximum	Mean	Std. deviation
Recency	317	0.000	1.000	0.812	0.236
Loyalty	317	0.000	1.000	0.220	0.233
AAD	317	0.000	1.000	0.023	0.100
AASR	317	0.000	1.000	0.026	0.103
Frequency	317	0.000	1.000	0.298	0.302
LTRP	317	0.000	1.000	0.216	0.232
APCIAD	317	0.000	1.000	0.064	0.099
APCIASR	317	0.000	1.000	0.018	0.061

Table 4.39 indicates the cluster centroids that are obtained by Ward's method.

Table 4.39 Cluster centroids (Ward's - MD)

Cluster	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
1	0.972	0.216	0.010	0.011	0.551	0.217	0.088	0.017
2	0.896	0.130	0.001	0.001	0.103	0.131	0.031	0.007
3	0.990	0.775	0.346	0.372	0.822	0.770	0.198	0.102
4	0.495	0.047	0.001	0.001	0.051	0.033	0.030	0.007
5	0.993	0.542	0.014	0.017	0.531	0.544	0.104	0.026

The distances between the cluster centroids are given in Table 4.40. The results reveal that clusters three and four are the farthest clusters (1.486). It means that these two clusters are the most different clusters from each other with respect to predefined characteristics. On the other hand, clusters two and four are the most similar clusters (0.424).

Table 4.40 Distances between centroids (Ward's - MD)

	1	2	3	4	5
1	0	0.474	0.978	0.737	0.463
2	0.474	0	1.283	0.424	0.735
3	0.978	1.283	0	1.486	0.665
4	0.737	0.424	1.486	0	0.995
5	0.463	0.735	0.665	0.995	0

Customers that are the closest object to the centroid of a specific cluster are given in Table 4.41. For instance, C57 is the closest object to the centroid of Cluster 1.

Table 4.41 Central objects (Ward's - MD)

Cluster	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
1 (57)	1.000	0.207	0.002	0.002	0.542	0.218	0.081	0.017
2 (124)	0.833	0.131	0.002	0.003	0.099	0.122	0.026	0.007
3 (301)	1.000	1.000	0.298	0.345	0.893	1.000	0.069	0.013
4 (247)	0.500	0.071	0.002	0.004	0.045	0.046	0.024	0.006
5 (200)	1.000	0.547	0.005	0.007	0.511	0.553	0.087	0.021

Distances between the central objects are presented in Table 4.42.

Table 4.42 Distances between the central objects (Ward's - MD)

	1 (57)	2 (124)	3 (301)	4 (247)	5 (200)
1 (57)	0	0.492	1.252	0.741	0.478
2 (124)	0.492	0	1.546	0.351	0.748
3 (301)	1.252	1.546	0	1.717	0.867
4 (247)	0.741	0.351	1.717	0	0.977
5 (200)	0.478	0.748	0.867	0.977	0

Table 4.43 defines the clusters by presenting their number of objects, minimum distance to the centroid, maximum distance to the centroid and mean distance to the centroid. The results reveal that 78 of 317 customers are assigned to Cluster 1, 80 of them are assigned to Cluster 2, 17 of them are assigned to Cluster 3, 96 of them are assigned to Cluster 4 and 46 of them are assigned to Cluster 5.

Table 4.43 Results of Ward's AHC for five clusters (MD)

Cluster	1	2	3	4	5
Number of Objects	78	80	17	96	46
Minimum distance to centroid	0.033	0.063	0.369	0.030	0.033
Average distance to centroid	0.282	0.154	0.616	0.149	0.250
Maximum distance to centroid	0.755	0.332	1.146	0.502	0.582

In Figure 4.16, we can see the evaluation of groups for each characteristic.

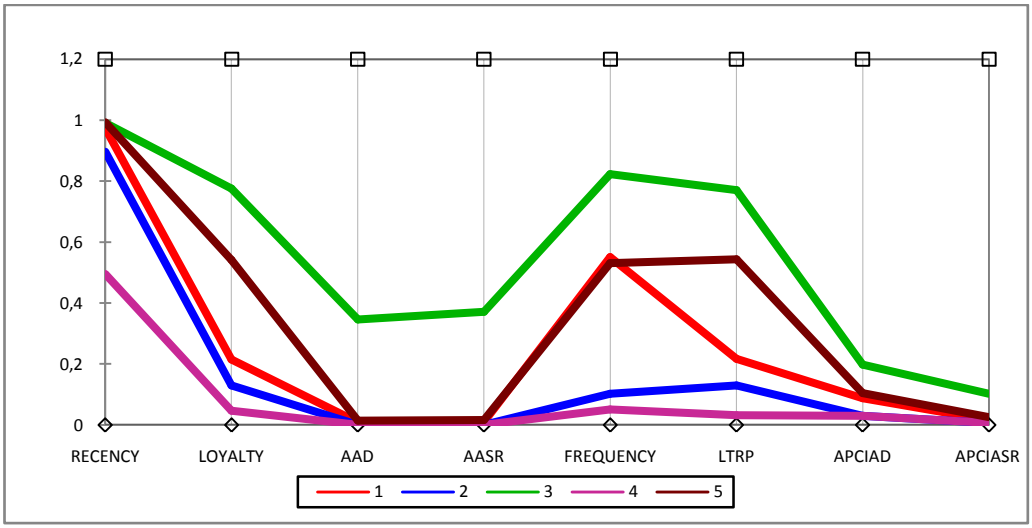

Figure 4.16 Profile plot of clusters (Ward's - MD)

Figure 4.16 reveal that there is big gap between cluster three and the other clusters in terms of AAD and AASR. In addition, cluster three has the greatest values of the characteristics. On the other hand cluster four has the lowest values with respect to the characteristics. Considering the results, it can be concluded that cluster three corresponds to the most important segment while cluster four corresponds to the least important one.

Importance of a cluster for a multi-dimensional data set is computed as follows:

$$C_i = \sum_{j=1}^{8} w_j F_{ij} \tag{4.14}$$

where i is the cluster number ($i=0, 1...5$), j is the number of characteristics, w_j is the importance weight of characteristic j and F_{ij} is the mean value of characteristic j of cluster i.

In Table 4.44, importance level of the clusters are computed by using importance weights of the characteristics, which are obtained by fuzzy AHP and ranked through 1 to 5, where 1 is the most important and 5 is the least important clusters. The results reveal that cluster three is the most important cluster and cluster four is the least important cluster.

Table 4.44 Rank of the clusters (Ward's - MD)

weight	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*		
cluster	0.075	0.118	0.215	0.258	0.029	0.168	0.033	0.103	C_i	*Rank*
1	0.972	0.216	0.010	0.011	0.551	0.217	0.088	0.017	0.161	3
2	0.896	0.130	0.001	0.001	0.103	0.131	0.031	0.007	0.110	4
3	0.990	0.775	0.346	0.372	0.822	0.770	0.198	0.102	0.507	1
4	0.495	0.047	0.001	0.001	0.051	0.033	0.030	0.007	0.052	5
5	0.993	0.542	0.014	0.017	0.531	0.544	0.104	0.026	0.259	2

4.7.2.2 MD - Based Customer Segmentation using k-means Algorithm

In k-means clustering, clustering criterion is selected as Determinant (W) and the number of repetitions is determined as 10. Table 4.45 and Figure 4.17 summarize the *SSW* for different number of clusters. The results reveal that the proper number of cluster is five.

Table 4.45 Sum of squares of k-means (MD)

Clustering Criterion	*SSW*	*SSB*	*TSS*	*SSW/TSS*	# of clusters
Determinant (W)	35.303	56.103	91.406	0.386	3
Determinant (W)	22.996	68.410	91.406	0.252	5
Determinant (W)	17.212	74.194	91.406	0.188	7

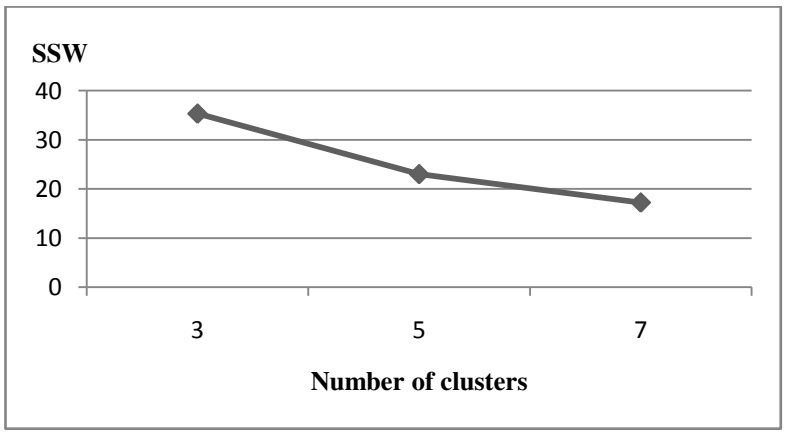

Figure 4.17 *SSW* values of k-means (MD)

If *k* takes the value of 5, k-means algorithm was truncated itself after four iterations. Statistics for the iterations are presented in Table 4.46. In addition, Figure 4.18 illustrates the change in Determinant (W) over the iterations.

Table 4.46 Statistics for the iterations (k-means - MD)

Iteration	Within-cluster variance	Trace(W)	Determinant(W)
0	0.290	90.595	206.089
1	0.099	30.966	6.746
2	0.093	28.993	4.787
3	0.082	25.500	3.297
4	0.074	22.996	1.767

Table 4.47 summarizes the results of each repetition. The results reveal that repetition 10 has the minimum value of Determinant (W).

Table 4.47 Optimization summary for k-means (MD)

Repetition	Iteration	Initial within-cluster variance	Final within-cluster variance	Determinant(W)
1	6	0.289	0.073	2.149
2	6	0.290	0.072	2.220
3	6	0.289	0.072	2.059
4	6	0.288	0.071	1.925
5	6	0.291	0.074	2.677
6	6	0.289	0.073	2.149
7	6	0.288	0.072	2.066
8	6	0.286	0.074	2.077
9	4	0.284	0.072	2.061
10	**4**	**0.290**	**0.074**	**1.767**

Coordinates of the initial cluster centroids are presented in Table 4.48.

Table 4.48 Initial cluster centroids (k-means - MD)

Cluster	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
1	0.833	0.214	0.032	0.036	0.329	0.213	0.078	0.019
2	0.812	0.262	0.020	0.024	0.295	0.255	0.076	0.032
3	0.836	0.206	0.008	0.008	0.319	0.209	0.063	0.015
4	0.772	0.233	0.033	0.035	0.287	0.221	0.049	0.012
5	0.804	0.192	0.025	0.026	0.256	0.190	0.056	0.011

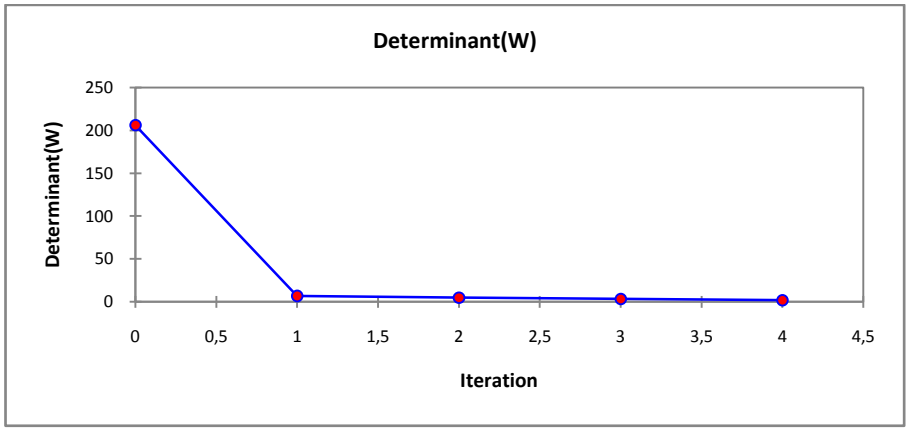

Figure 4.18 Determinant (W) over the iterations (k-means - MD)

Table 4.49 represents the final cluster centroids that are obtained by k-means algorithm.

Table 4.49 Cluster centroids (k-means - MD)

Cluster	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
1	0.995	0.174	0.036	0.037	0.627	0.184	0.083	0.017
2	0.830	0.114	0.001	0.001	0.097	0.110	0.035	0.008
3	0.980	0.490	0.014	0.018	0.492	0.488	0.101	0.037
4	0.424	0.040	0.001	0.001	0.045	0.024	0.033	0.008
5	0.986	0.935	0.342	0.372	0.811	0.925	0.210	0.056

Distances between cluster centroids can be seen in Table 4.50. As reported in the table, clusters four and five are the farthest clusters (1.674). On the other hand, clusters four and two are the most similar clusters (0.425).

Table 4.50 Distances between the cluster centroids (k-means - MD)

	1	2	3	4	5
1	0	0.568	0.461	0.844	1.178
2	0.568	0	0.684	0.425	1.470
3	0.461	0.684	0	0.965	0.858
4	0.844	0.425	0.965	0	1.674
5	1.178	1.470	0.858	1.674	0

Customers that are the closest object to the centroid of a specific cluster are given in Table 4.51. For instance, C53 is the closest object to the centroid of Cluster 1.

Table 4.51 Central objects (k-means - MD)

Cluster	*Recency*	*Loyalty*	*AAD*	*AASR*	*Frequency*	*LTRP*	*APCIAD*	*APCIASR*
1 (53)	1.000	0.207	0.002	0.004	0.619	0.218	0.021	0.006
2 (124)	0.833	0.131	0.002	0.003	0.099	0.122	0.026	0.007
3 (153)	1.000	0.434	0.004	0.004	0.503	0.442	0.099	0.023
4 (182)	0.500	0.009	0.000	0.000	0.058	0.011	0.023	0.006
5 (301)	1.000	1.000	0.298	0.345	0.893	1.000	0.069	0.013

Distances between the central objects are presented in Table 4.52.

Table 4.52 Distances between the central objects (k-means - MD)

	1 (53)	2 (124)	3 (153)	4 (182)	5 (301)
1 (53)	0	0.559	0.348	0.804	1.233
2 (124)	0.559	0	0.624	0.375	1.546
3 (153)	0.348	0.624	0	0.906	0.994
4 (182)	0.804	0.375	0.906	0	1.766
5 (301)	1.233	1.546	0.994	1.766	0

Table 4.53 defines the clusters. The results reveal that 61 of 317 customers are assigned to Cluster1, 109 of them are assigned to Cluster 2, 67 of them are assigned to Cluster 3, 68 of them are assigned to Cluster 4 and 12 of them are assigned to Cluster 5.

Table 4.53 Results of k-means for five clusters (MD)

Cluster	1	2	3	4	5
Number of objects	61	109	67	68	12
Minimum distance to centroid	0.093	0.024	0.079	0.085	0.203
Average distance to centroid	0.294	0.177	0.290	0.127	0.451
Maximum distance to centroid	1.122	0.720	1.094	0.433	0.915

Figure 4.19 shows the group averages for each characteristic. It is clear that there is a big gap between clusters five and the other clusters especially in terms of loyalty, AAD, AASR and LTRP. The results conclude that cluster five corresponds to the most important segment while cluster four corresponds to the least important segment.

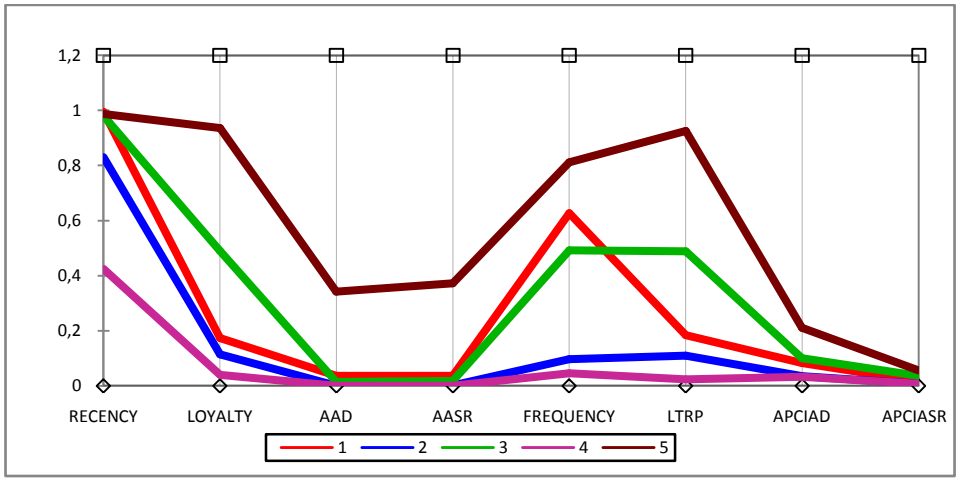

Figure 4.19 Profile plot of the clusters (k-means - MD)

In Table 4.54, importance of clusters are computed and ranked through 1 to 5. Herein, cluster five is the most important cluster and cluster four is the least important cluster.

Table 4.54 Rank of the clusters (k-means - MD)

weight	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*		
cluster	0.075	0.118	0.215	0.258	0.029	0.168	0.033	0.103	C_i	*Rank*
1	0.995	0.174	0.036	0.037	0.627	0.184	0.083	0.017	0.166	3
2	0.83	0.114	0.001	0.001	0.097	0.11	0.035	0.008	0.100	4
3	0.98	0.49	0.014	0.018	0.492	0.488	0.101	0.037	0.243	2
4	0.424	0.04	0.001	0.001	0.045	0.024	0.033	0.008	0.044	5
5	0.986	0.935	0.342	0.372	0.811	0.925	0.21	0.056	0.546	1

4.8 Evaluation of the Results

Importance level of a customer segment varies depending on the clustering method used. In the application performed in this study, importance levels of the segments are compared in terms of the best results of two different approaches by using paired-t test. As known, paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample.

To test the null hypothesis that the true mean difference is zero, or in other words, these two approaches are indifferent and alternative hypothesis is that these approaches are significantly different from each other, the procedure is as follows (Shier, 2004):

- Calculate the difference $(d_i = y_i - x_i)$ between the two observations on each pair,
- Calculate the mean difference, $\bar{d}$,
- Calculate the standard deviation of the differences, s_d, and use this to calculate the standard error of the mean difference, $SE(\bar{d}) = \frac{S_d}{\sqrt{n}}$
- Calculate the t-statistic, which is given by $T = \frac{\bar{d}}{SE\ (\bar{d})}$. Under the null hypothesis, this statistic follows a t-distribution with $n - 1$ degrees of freedom,
- Use tables of the t-distribution to compare your value for T to the t_{n-1}distribution. This will give the *p*-value for the paired t-test. If the *p* value is greater than the threshold chosen for statistical significance then the null hypothesis is accepted.

The results are obtained by SPSS 13.0, and summarized in Table 4.55. If the results of paired-t test are considered for a %95 confidence interval, it can be concluded that all of the approaches are different except Ward's SD and k-means SD. In addition, second approach that uses multidimensional data set assigns more customers to the upper segments compared with the first approach. On the other hand, for multidimensional data set, k-means algorithm assigns more customers to the upper segments than Ward's method does.

Table 4.55 The results of paired-t tests

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 k-means-SD Ward's-SD	-0.006	0.079	0.004	-0.015	0.002	-1.416	316	0.158
Pair 2 k-means-MD Ward's -MD	-0.120	0.455	0.026	-0.170	-0.070	-4.690	316	0.000
Pair 3 k-means -SD kmeans-MD	0.319	0.542	0.030	0.259	0.379	10.470	316	0.000
Pair 4 Ward's -SD Ward's -MD	0.205	0.489	0.027	0.151	0.259	7.460	316	0.000
Pair 5 Ward's -SD k-means -MD	0.325	0.532	0.030	0.266	0.384	10.870	316	0.000
Pair 6 k-means -SD Ward's -MD	0.199	0.485	0.027	0.145	0.252	7.290	316	0.000

Numbers of customers that assigned to these segments are presented in Table 4.56.

Table 4.56 Number of customers assigned to the segments

	1	2	3	4	5
k-means-SD	11	39	66	86	115
Ward's-SD	9	41	66	86	115
k-means-MD	12	67	61	109	68
Ward's-MD	17	46	78	80	96

Assignment similarities of the approaches are compared in Table 4.57. More detailed information on the assignments is presented in Appendix A. If the results of Ward's-SD and Ward's-MD are compared, it can be seen that 71.9 percent of the customers are assigned to the same segment, 24.3 percent of them assigned to one upper segment and 3.8 percent of them assigned to one lower segment in Ward's-MD method. In addition, the most similar results are obtained by k-means-SD and Ward's-SD.

Table 4.57 Comparison of the methods in terms of assignment similarities

	k-means-SD	Ward's-MD	k-means-MD
Ward's-SD			
one upper segment	0.006	0.243	0.353
same segment	0.994	0.719	0.621
one lower segment	0.000	0.038	0.022
two lower segment	0.000	0.000	0.003
k-means-SD			
one upper segment		0.237	0.350
same segment		0.726	0.625
one lower segment		0.038	0.019
two lower segment		0.000	0.006
Ward's-MD			
two upper segment			0.006
one upper segment			0.142
same segment			0.826
one lower segment			0.016
two lower segment			0.009

4.9 The Final Customer Segments

In this study, we aimed to divide customers into manageable groups and compose segments according to their importance to the company. With the integration of cluster analysis and different approaches the customers of the company under concern were grouped into five segments. Final customer segments are determined using *SSW/TSS* ratios of the different approaches. It is aimed to minimize this ratio since the smaller value of it means that the clusters are compact or in other words variation in clusters is low. As reported in Table 4.58, k-means-SD has the smallest value of this ratio. Therefore, the customer segments are determined by the results of k-means-SD algorithm where *k* equals to five.

Table 4.58 *SSW/TSS* ratio of different approaches for five clusters

Approach	*SSW/TSS*	# of clusters
Ward's -SD	0.074	5
k- means-SD	0.072	5
Ward's -MD	0.260	5
k- means-MD	0.252	5

The segments are named as “best”, “valuable”, “average”, “potential valuable” and “potential invaluable” customers. Their features are given in Tables 4.59 to 4.63.

The results reveal that cluster three is the most important segment. With respect to “AAD”, “AASR” and “LTRP” characteristics, this cluster has the greatest values among the clusters. Customers in this segment are totally loyal customers. They order products with high volumes and they are the most important source of income for the company. Therefore, customers in this segment should be retained.

Cluster five is the second important segment. Customers in this segment have a long relationship with the company. They have great volume of orders and they are important source of income for the company. Generally, their orders are recent, so they have a higher potential for a long term relationship. In addition, they show an increasing trend with respect to AAD and AASR.

Cluster two is the third important segment for the company. Customers in this segment have an average length of relationship with the company. They are partially loyal to the company and they have average amount of orders. Their LTRP value is also moderate. These customers should be focused to increase their order volumes.

Cluster one is the fourth important segment. Most of the customers in this segment are new customers. They have relatively shorter relationship with the company, but they have great amount of annual demand relatively to their LoR.

Cluster four is the fifth important segment. Most of customers in this segment don’t have recent orders and their relationships with the company are inactive. They can be qualified as disloyal according to their degree of loyalty. They have low volume orders and also sales revenue obtained from them is low. In addition, some of them are relatively new customers but they have lower amount of annual demand and sales revenue than the new customers in cluster one. Relationship with these customers should be reconsidered.

Table 4.59 Best customers

CLUSTER THREE-BEST CUSTOMERS	
General Characteristics	Smallest cluster with 11 customers
	Attractive subgroup of data set
Behavioral characteristics	Generally have long and consistent relationship with the company
	Greatest average amount of annual demand
	Greatest average amount of annual sales revenue
	Highest potential for long term relationships
	Shows an increasing trend with respect to AD and ASR

Table 4.60 Valuable customers

CLUSTER FIVE-VALUABLE CUSTOMERS	
General Characteristics	Second smallest cluster with 39 customers
	Not outliers
Behavioral characteristics	Generally have long relationship with the company
	Greater average amount of annual demand than other clusters except Cluster 3
	Greater average amount of annual sales revenue than other clusters except Cluster 3
	Higher potential for long term relationships than other clusters except Cluster 3
	Shows an increasing trend with respect to AD and ASR

Table 4.61 Average customers

CLUSTER TWO-AVERAGE CUSTOMERS	
General Characteristics	Third smallest cluster with 66 customers
	Ordinary subgroup of data set
Behavioral characteristics	Have average length of relationship with the company
	Average amount of annual demand
	Average amount of annual sales revenue
	Average potential for long term relationships
	Shows a static structure with respect to AD and ASR

Table 4.62 Potential valuable customers

CLUSTER ONE-POTENTIAL VALUABLE CUSTOMERS	
General Characteristics	Second biggest cluster with 86 customers
	Not outliers
Behavioral characteristics	Have shorter relationship with the company
	Great amount of annual demand relatively to their LoR
	Great amount of annual sales revenue relatively to their LoR
	Lower potential for long term relationships
	Shows a static structure with respect to AD and ASR

Table 4.63 Potential invaluable customers

CLUSTER FOUR-POTENTIAL INVALUABLE CUSTOMERS	
General Characteristics	Biggest cluster with 115 customers
	Not outliers
Behavioral characteristics	Have average length of relationship with the company
	Low amount of annual demand
	Low amount of annual sales revenue
	Lowest potential for long term relationships
	Shows a decreasing trend with respect to AD and ASR

CHAPTER FIVE
CONCLUSIONS

Companies may have thousands of customers. It is complex and difficult to manage such a large customer base. To know the value of customers is very important for sales and marketing strategies of companies. On the other hand, developing customer-specific marketing strategy is time consuming and difficult to follow. For this reason, dividing customers into small groups according to their similarities will be more meaningful. In addition, determining the value of these groups to the company is also important as grouping customers.

Data mining helps the companies on issues that classify and identify the customers, and predict their behaviors. In addition, data mining provide strategic information for many customer-centric applications. One of the most common application areas of data mining in CRM is customer segmentation. Customer segmentation is the division of the market into small groups of customers with similar characteristics. As a data mining technique, data clustering can be employed for customer segmentation. Data clustering algorithms group customers based on their predefined characteristics.

In today's competitive market, most of leading brands take advantage of the economies of scale by collaborating with contract manufacturers. Especially in the electronics industry, products enter the market with relatively high initial price but these prices rapidly fall and quality improves over time. On the other hand, there are many competing contract manufacturers in this industry. To become a successful contract manufacturer, they should understand and care about customer's business, goals, needs and expectations. Also, they should adopt long term partnership instead of traditional contract manufacturer-customer relationship.

Contract manufacturers have multiple customers that they produce for. Moreover, existing customer portfolio is an important reference to gain new customers in the future. Contract manufacturers should maintain their market share against potential competitors by providing high quality products and services like every manufacturer.

Other important issues for contract manufacturers are capacity planning, production and distribution. Companies should use their limited resources in an effective manner by selecting the valuable or in other words strategic important customers and making efforts to keep them.

This study was carried out in an international TV manufacturing company and aims to divide the customers into small manageable groups using clustering algorithms and also to find relative importance of these groups using multi criteria decision making technique.

At first, eight different characteristics are defined as "*recency", "loyalty", "average annual demand", "average annual sales revenue", "frequency", "long term relationship potential", "average percentage change in annual demand" and "average percentage change in annual sales revenue"* for the evaluation of the customers. Fuzzy AHP is used to determine the importance weights of the characteristics and the most important characteristic was obtained as "*average annual sales revenue*".

In the next stage, customers were grouped according to their characteristics using two different approaches, single dimension-based and multiple dimensions-based approaches. The first approach clusters customers according to a combined characteristic called *overall score* that merges all characteristics into one dimension. The numbers of clusters were taken as 3, 5 and 7 in both AHC and k-means algorithms. The results were compared according to *SSW* values. The results reveal that k-means segments customers into five clusters better than the other methods.

Second approach considers all of the characteristics and clusters the customers according to them. The results reveal that k-means segments customers into five clusters better than Ward's, single linkage and complete linkage algorithms.

Final customer segments are determined using *SSWC/TSS* ratio of the different approaches. According to the results, k-means that uses single dimensional data set has the smallest value of *SSW/TSS*. Therefore, the customer segments were profiled considering the results of this approach. The segments were named as best, valuable, average, potential valuable and potential invaluable customers.

"The best customers" segment is the most important segment. This cluster has the greatest values among all clusters with respect to "average annual demand", "average annual sales revenue" and "long term relationship potential" characteristics. Customers in this segment are totally loyal customers, they order products with high volumes and they are the most important source of income for the company. "Valuable customers" segment is the second important segment. Customers in this segment have a long relationship with the company, they have great volume of orders and they are important source of income for the company. Their orders are generally recent. So, they have a higher potential for a long term relationship. In addition, they show an increasing trend with respect to annual demand and annual sales revenue. "Average customers" segment is the third important segment for the company. Customers in this segment have an average length of relationship with the company. They are partially loyal to the company and they have average amount of orders. Their LTRP is also moderate. "Potential valuable customers" segment is the fourth important segment. Most of the customers in this segment are new customers. They have shorter relationship with the company but they have great amount of annual demand relatively to their LoR. Potential invaluable customers segment is the fifth important segment. Most of customers in this segment don't have recent orders and their relationships with the company are inactive. They can be treated disloyal according to their degree of loyalty. They have low volume orders and also sales revenue obtained from them is low. Furthermore, they show a decreasing trend with respect to annual demand and annual sales revenue.

This study can be treated as a road map for customer segmentation for the company under concern. In practice, customer base, customer evaluation characteristics and their importance levels are subject to change over time. Therefore, customer segments should be dynamically updated. Additionally, the clustering algorithms can be applied for different number of clusters and similarity/dissimilarity metrics. By using the output of this study, different marketing strategies can be developed for the customer groups with respect to the firm's resources, constraints and views. Also, loyalty programs can be developed.

REFERENCES

Ahn, J.S., & Sohn, S.Y. (2009). Customer pattern search for after-sales service in manufacturing. *Expert Systems with Applications, 36,* 5371-5375.

Ansari, Z., Azeem, M.F., Ahmed, W., & Babu, A.V. (2011). Quantitative evaluation of performance and validity indices for clustering web navigational sessions. *World of Computer Science and Information Technology Journal (WCSIT)*, *1*, 217-226.

Bayer, J. (2010). Customer segmentation in the telecommunications industry. *Database Marketing & Customer Strategy Management*, *17*, 247-256.

Belmokaddem, M., Mekidiche, M., & Sahed, A. (2009). Application of fuzzy goal programming approach with different importance and priorities to aggregate production planning. *Journal of Applied Quantitative Methods*, *4*(3), 317-331.

Bergeron, B. (2002). *Essentials of CRM: Customer relationship management for executives* (1st ed.). New York: John Wiley & Sons Inc.

Buckinx, W., & Poel D.V. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*, 252**-**268.

Buttle, F. (2004). *Customer relationship management: Concepts and tools* (1st ed.). Oxford: Elsevier.

Carrier, C.G., & Povel, O. (2003). Characterizing data mining software. *Intelligent Data Analysis*, *7*, 181 **-** 192.

Chan, C.C.H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications*, *34*, 2754-2762.

Chang, D.Y. (1992). Extent analysis and synthetic decision. *Optimization Techniques and Applications*, *1*, 352-366.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinart, T., Shearer, C., & Wirth, R. (2000). *CRISP–DM Step-by-Step Data Mining Guide.* Retrieved March 24, 2013, from ftp://ftp.software.ibm.com/software/analytics/spss/support/ Modeler/ Documentation/14/UserManual/CRISP-DM.pdf.

Chen, Y., Zhang, G., Hu, D., & Wang, S. (2006). Customer segmentation in customer relationship management based on data mining. *International Federation for Information Processing (IFIP)*, *207*, 288-293.

Cheng, C.H., & Chen, Y.S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*, 4176-4184.

Chiu, C.Y., Chen, Y.F., Kuo, I.T., & Ku, H.C. (2009). An intelligent market segmentation system using k-means and particle swarm optimization. *Expert Systems with Applications*, *36*, 4558-4565.

Dannenberg, H. & Zupancic, D. (2009). *Excellence in sales: Optimising customer and sales management* (1st ed.). Wiesbaden: Gabler.

Davies, D.L., & Bouldin, D.W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*, 224-227.

Dhandayudam, P., & Krishnamurthi, I. (2012). An improved clustering algorithm for customer segmentation. *International Journal of Engineering Science and Technology (IJEST)*. *4*(2), 695-702.

Diday, E., & Simon, J.C. (1976). *Clustering analysis in digital pattern recognition*. K. S. Fu. (Ed.). Heidelberg: Springer, 47-94.

Doyle, S. (2002). Software review: Communication optimisation–The new mantra of database marketing. Fad or fact?. *Journal of Database Marketing, 9*(2), 185-191.

Dubes, R.C. (1993). *Cluster analysis and related issues*. In *Handbook* of Pattern Recognition & Computer Vision. New Jersey: World Scientific Publishing Co.

Dunn, J.C. (1974). Well separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, *4*(1), 95-104.

Gilbert, K., Marre, M.S., & Codina, V. (2012). Choosing the right data mining technique: classification of methods and intelligent recommendation. *International Congress on Environmental Modeling and Software Modeling for Environment's Sake, Fifth Biennial Meeting,* Ottawa, Canada.

Gilboa, S. (2009). A segmentation study of Israeli mall customers. *Journal of Retailing and Consumer Services*, *16*, 135-144.

Grabusts, P. (2011). The choice of metrics for clustering algorithms. *Proceedings of the 8th International Scientific and Practical Conference*, *11*, 70-76.

Halkidi, M., Batistakis, Y. & Vazirgiannis, M. (2002). Cluster validity methods: part II. *SIGMOD Record*, *31*(2), 40-45.

Herschel, G. (2002). Introduction to CRM analytics. *Gartner Symposium ITxpo*, Orlando, Florida.

Ho, G.T.S., Ip, W.H., Lee, C.K.M., & Mou, W.L. (2012). Customer grouping for better resource allocation using GA based clustering technique. *Expert Systems with Applications*, *39*, 1979-1987.

Hosseni, M.B., & Tarokh, M.J. (2011). Customer segmentation using CLV elements. *Journal of Service Science and Management*, *4*, 284-290.

Hosseini, S.M.S., Maleki, A., & Gholamian, M.R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*, 5259-5264.

Hughes, A.M. (1994). *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable Customer – Based Marketing Program* (2nd ed.). Chicago: McGraw-Hill.

Iriana, R., & Buttle F. (2006). Strategic, operational, and analytical customer relationship management. *Journal of Relationship Marketing*, *5* (4), 23-42.

İyigün, C. (2008). *Probabilistic distance clustering*. Doctor of Philosophy Thesis in Operations Research, New Brunswick University, New Jersey, 15-17.

Jackson, J. (2002). Data Mining: A Conceptual Overview. *Communications of the Association for Information Systems*, *8*, 267-296.

Jain, A.K., & Dubes, R.C. (1988). *Algorithms for clustering data* (1st ed.). New Jersey: Prentice-Hall Inc.

Jain, A.K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, *31*, 651-666.

Jamalnia, A., & Soukhakian, M.A. (2009). A hybrid fuzzy goal programming approach with different goal priorities to aggregate production planning. *Computers and Industrial Engineering*, *56*, 1474-1486.

Kahraman, C., Cebeci, U., & Ruan, D. (2004). Multi-attribute comparison of catering service companies using fuzzy AHP: The case of Turkey. *International Journal of Production Economics, 87*,171-184.

Kelly, S. (2003). Mining data to discover customer segments. *Interactive Marketing*, *4*(3), 235-242.

Kim, S.Y., Jung T.S., Suh, E.H., & Hwang, H.S. (2006) Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, *31*,101-107

Kracklauer, A. H., Mills, D.Q., & Seifert, D. (Eds.). (2004). *Collaborative customer relationship management: taking CRM to the next level* (1st ed.). Berlin: Springer.

Larose, D.T. (2005). *Discovering knowledge in data: An introduction to data mining* (1st ed.). New Jersey: John Wiley & Sons Inc.

Li, D.C., Dai, W.L., & Tseng, W.T. (2011). A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business. *Expert Systems with Applications*, *38*, 7186-7191.

Li, L., & Lai, K.K. (2001). Fuzzy dynamic programming approach to hybrid multiobjective multistage decision-making problems. *Fuzzy Sets and Systems*, *117*, 13-25.

Ling, R., & Yen, D.C. (2001). Customer relationship management: An analysis framework and implementation strategies. *The Journal of Computer Information Systems, 41*(3), 82-97.

Liu, D.R., & Shih, Y.Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information and Management*, *42*, 387-400.

Manu, C. (2012). Analysis of clustering technique for CRM. *International Journal of Engineering and Management Sciences(IJEMS)*, *3*, 402-408.

Meltzer, M. (2005). *Customer Centricity : Making CRM Magical : Using Outcomes.* Retrieved April 7, 2013, from http://www.leader-values.com/article.php?aid=614

Miglautsch, J.R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing*, *8*(1), 67-72.

Montinaro, M., & Sciascia, I. (2011). Market segmentation to obtain different kinds of customer loyalty. *Journal of Applied Sciences*, *11*(4), 655-662.

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research: The process, data, and methods using IBM SPSS Statistics* (1st ed.). Berlin: Springer.

Ngai, E.W.T., Xiu, L., & Chau, D.C.K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification, *Expert Systems with Applications*, *36*, 2592-2602.

Newell, F. (1997). The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry. New York: McGraw-Hill Companies Inc.

Oliveira, V.L.M. (2012). *Analytical customer relationship management in retailing supported by data mining techniques*. Doctor of Philosophy Thesis in Industrial Engineering and Management, Universidade do Porto, Portugal.

Osborn, A.F. (1953). *Applied imagination: Principals and procedures of creative thinking*. New York: Charles Scribner's Sons.

Parvatiyar, A., & Sheth, J. (2001). Customer relationship management: emerging practice, process, and discipline. *Journal of Economic and Social Research*, *3*(2), 1-34.

Patel, V.R., & Mehta, R.G. (2011). Data clustering: integrating different distance measures with modified k-means algorithm. *Proceedings of the International Conference on SocProS*, *131*, 691-700.

Pereira, G.C., Coutinho, R., & Ebecken, N.F.F. (2008). Data mining for environmental analysis and diagnostic: A case study of upwelling ecosystem of Arraial Do Cabo. *Brazilian Journal of Oceanography*, *56*(1), 1-12.

Pham, D.T., & Afify, A.A. (2007). Clustering techniques and their applications in engineering. *Mechanical Engineering Science*, *221*, 1445-1459.

Phyu, T.N. (2009). Survey of classification techniques in data mining. *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, Hong Kong.

Plakoyiannaki, E. & Tzokas, N. (2002). Customer relationship management: A capabilities portfolio perspective. *Journal of Database Marketing*, *9*(3), 228-37.

Punj, G., & Stewart, D.W. (1983). Cluster analysis in marketing research: review and suggestions for application. *Journal of Marketing Research*, *20*(2), 134-148.

Rajagopal S. (2011). Customer data clustering using data mining technique. *International Journal of Database Management Systems (IJDMS), 3*(4), 1-11.

Reichheld, F.F., & Sasser W.E. (1990). Zero defections: quality comes to services. *Harvard Business Review*, *68*(5), 105-111.

Reinartz, W., & Kumar V. (2002). The mismanagement of customer loyalty. *Harvard Business Review*, R0207F.

Rendon, E., Abundez, I., Arizmendi, A., & Quiroz, E.M. (2011). Internal versus external cluster validation indexes. *International Journal of Computers and Communications*, *5*, 27-34.

Reynolds, J. (2002). *A Practical Guide to CRM: Building more profitable customer relationships* (1st ed.). New York: CMP Books.

Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

Rygielski, C., Wang, J.C. & Yen, D.C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, *24*, 483-502.

SAP White Paper. (2001). *Analytical CRM*. Retrieved April 7, 2013, from http://www.sap.com/belgie/solutions/business-suite/crm/pdf/brochures/AnalyticalCRM_50046585.pdf

Shier, R. (2004). *Paired t-tests*. Retrieved April 3, 2013, from http://mlsc.lboro.ac.uk/resources/statistics/Pairedttest.pdf.

Speier, C. & Venkatesh, V. (2002). The hidden minefields in the adoption of sales force automation technologies. *Journal of Marketing*, *66*(3), 98-111.

Srivastava, J., Wang, J. H., Lim, E. P., & Hwang, S.Y. (2002). A case for analytical customer relationship management. *Lecture Notes in Artificial Intelligence*, Heidelberg: Springer, *2336*, 14-27.

Swift, R.S. (2001). *Accelerating customer relationships: Using CRM and relationship technologies*. New Jersey: Prentice Hall.

Tan, P.N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining* (1st ed.). USA: Addison-Wesley Pearson.

Tarokh, M.J., & Sekhavat, A.A. (2006). LTV model in consultant sector. Case study: Mental health clinic. *Behaviour & Information Technology*, *25*(5), 399-405.

Teichert, T., Shehu, E., & Wartburg, I. (2008). Customer segmentation revisited: The case of airline industry. *Transportation Research Part A*, *42*, 227-242.

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, *58*, 236-244.

Wei, J.T., Lin, S.Y., & Wu, H.H. (2010). A review of the application of RFM model. *African Journal of Business Management*, *4*(19), 4199-4206.

Wind, Y. (1987). Issues and advances in segmentation research. *Journal of Marketing Research*, *15*(3), 317-337.

Wu, R.S., & Chou, P.H. (2011). Customer segmentation of multiple category data in e-commerce using a soft clustering approach. *Electronic Commerce Research and Applications*, *10*, 331-341.

XLSTAT. *Products and Solutions*. Retrieved February 16, 2013, from http://www.xlstat.com/en/products-solutions.html.

Xu, M., & Walton, J. (2005). Gaining customer knowledge through analytical CRM. *Industrial Management and Data Systems*, *105*(7), 955-71.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

APPENDIX A - THE RANK OF SEGMENTS TO WHICH CUSTOMERS WERE ASSIGNED

Customer ID	k-means-SD	Ward's- SD	k-means-MD	Ward's-MD
1	4	4	3	3
2	4	4	3	3
3	4	4	3	3
4	4	4	3	3
5	3	3	3	3
6	4	4	3	3
7	4	4	3	3
8	4	4	4	4
9	4	4	4	4
10	4	4	4	4
11	4	4	3	3
12	4	4	4	4
13	4	4	4	4
14	4	4	3	3
15	4	4	4	4
16	4	4	4	4
17	4	4	3	3
18	4	4	4	4
19	4	4	3	3
20	4	4	3	3
21	4	4	3	3
22	4	4	3	3
23	4	4	3	3
24	4	4	4	4
25	4	4	3	3
26	4	4	4	4
27	4	4	3	3
28	4	4	4	4
29	4	4	4	4
30	4	4	4	4
31	4	4	4	4
32	4	4	3	3
33	4	4	3	3
34	4	4	3	3
35	4	4	3	3
36	4	4	3	3
37	4	4	3	3
38	4	4	3	3
39	4	4	3	3
40	4	4	3	3
41	4	4	3	3
42	4	4	3	3
43	3	3	3	3
44	3	3	3	3
45	4	4	3	3
46	3	3	3	3
47	4	4	4	4
48	4	4	3	3
49	3	3	3	3
50	3	3	3	3
51	3	3	3	3

APPENDIX A – CONTINUED

Customer ID	k-means-SD	Ward's- SD	k-means-MD	Ward's-MD
52	4	4	4	4
53	3	3	3	3
54	4	4	4	4
55	3	3	3	3
56	4	4	4	4
57	3	3	3	3
58	3	3	3	3
59	3	3	3	3
60	4	4	4	4
61	1	1	3	1
62	3	3	3	3
63	3	3	3	3
64	4	4	4	4
65	4	4	4	4
66	1	2	3	1
67	3	3	3	3
68	3	3	3	3
69	3	3	3	3
70	3	3	3	3
71	4	4	4	4
72	3	3	3	3
73	5	5	4	4
74	5	5	4	4
75	5	5	4	4
76	5	5	4	4
77	5	5	4	4
78	5	5	4	4
79	5	5	4	4
80	5	5	4	4
81	5	5	4	4
82	5	5	4	4
83	5	5	4	4
84	5	5	4	4
85	5	5	4	4
86	5	5	4	4
87	5	5	4	4
88	5	5	4	4
89	5	5	4	4
90	5	5	4	4
91	5	5	4	4
92	5	5	4	4
93	5	5	4	4
94	5	5	4	4
95	5	5	4	4
96	5	5	4	4
97	3	3	2	3
98	3	3	3	3
99	3	3	2	3
100	3	3	3	3
101	3	3	3	3
102	3	3	3	3
103	3	3	3	3
104	3	3	2	3
105	3	3	3	3
106	3	3	2	3

APPENDIX A – CONTINUED

Customer ID	k-means-SD	Ward's- SD	k-means-MD	Ward's-MD
107	3	3	4	4
108	3	3	2	4
109	3	3	2	3
110	3	3	2	3
111	3	3	2	3
112	3	3	3	3
113	3	3	2	4
114	3	3	2	3
115	2	2	3	1
116	3	3	2	3
117	3	3	2	3
118	3	3	3	3
119	3	3	2	3
120	3	3	2	3
121	4	4	4	4
122	4	4	4	4
123	4	4	4	4
124	4	4	4	4
125	4	4	4	4
126	4	4	4	4
127	4	4	4	4
128	4	4	4	4
129	4	4	4	4
130	4	4	4	4
131	4	4	4	4
132	5	5	4	5
133	5	5	4	5
134	5	5	4	5
135	5	5	4	5
136	5	5	4	5
137	5	5	4	5
138	5	5	4	5
139	5	5	4	5
140	5	5	4	5
141	5	5	4	5
142	5	5	4	5
143	5	5	4	5
144	5	5	4	5
145	5	5	4	5
146	5	5	4	5
147	5	5	4	5
148	5	5	4	5
149	5	5	4	5
150	3	3	2	2
151	3	3	2	2
152	3	3	2	2
153	3	3	2	2
154	3	3	2	2
155	3	3	2	2
156	3	3	2	2
157	3	3	2	2
158	3	3	2	2
159	4	4	4	4
160	4	4	4	4
161	4	4	4	4

APPENDIX A – CONTINUED

Customer ID	k-means-SD	Ward's- SD	k-means-MD	Ward's-MD
162	4	4	4	4
163	4	4	4	4
164	4	4	4	4
165	3	3	3	3
166	5	5	4	5
167	5	5	4	5
168	5	5	4	5
169	4	4	4	4
170	4	4	4	4
171	4	4	4	4
172	4	4	4	4
173	4	4	4	4
174	5	5	5	5
175	5	5	5	5
176	5	5	5	5
177	5	5	5	5
178	5	5	5	5
179	5	5	5	5
180	5	5	5	5
181	5	5	5	5
182	5	5	5	5
183	5	5	5	5
184	5	5	5	5
185	5	5	5	5
186	5	5	5	5
187	5	5	5	5
188	5	5	5	5
189	5	5	5	5
190	5	5	5	5
191	5	5	5	5
192	5	5	5	5
193	5	5	5	5
194	5	5	5	5
195	5	5	5	5
196	5	5	5	5
197	5	5	5	5
198	5	5	5	5
199	1	1	1	1
200	2	2	2	2
201	2	2	2	2
202	2	2	2	2
203	2	2	2	2
204	2	2	2	2
205	2	2	2	2
206	2	2	2	2
207	2	2	2	2
208	2	2	2	2
209	2	2	2	2
210	2	2	2	2
211	2	2	2	2
212	2	2	2	2
213	2	2	2	1
214	2	2	2	2
215	2	2	2	2
216	2	2	2	2

APPENDIX A – CONTINUED

Customer ID	k-means-SD	Ward's- SD	k-means-MD	Ward's-MD
217	2	2	2	2
218	2	2	2	2
219	2	2	2	2
220	2	2	2	2
221	2	2	2	2
222	2	2	2	2
223	3	3	2	3
224	3	3	4	4
225	3	3	4	4
226	3	3	2	3
227	3	3	4	4
228	4	4	4	5
229	4	4	4	5
230	4	4	4	5
231	4	4	4	5
232	4	4	4	5
233	4	4	4	4
234	4	4	4	4
235	5	5	5	5
236	5	5	5	5
237	5	5	5	5
238	5	5	5	5
239	5	5	5	5
240	5	5	5	5
241	5	5	5	5
242	5	5	5	5
243	5	5	5	5
244	5	5	5	5
245	5	5	5	5
246	5	5	5	5
247	5	5	5	5
248	5	5	5	5
249	5	5	5	5
250	5	5	5	5
251	5	5	5	5
252	5	5	5	5
253	5	5	5	5
254	5	5	5	5
255	5	5	5	5
256	5	5	5	5
257	5	5	5	5
258	5	5	5	5
259	5	5	5	5
260	5	5	5	5
261	5	5	5	5
262	5	5	5	5
263	2	2	2	2
264	2	2	2	2
265	2	2	2	2
266	2	2	2	2
267	2	2	2	2
268	3	3	2	2
269	3	3	2	2
270	3	3	3	3
271	4	4	4	3

APPENDIX A – CONTINUED

Customer ID	k-means-SD	Ward's- SD	k-means-MD	Ward's-MD
272	4	4	4	4
273	3	3	4	3
274	4	4	4	4
275	5	5	5	5
276	5	5	5	5
277	5	5	5	5
278	5	5	5	5
279	5	5	5	5
280	4	4	4	5
281	5	5	5	5
282	5	5	5	5
283	5	5	4	4
284	5	5	5	5
285	5	5	5	5
286	2	2	2	2
287	3	3	2	3
288	3	3	2	2
289	1	1	1	1
290	1	1	1	1
291	2	2	2	1
292	2	2	2	2
293	2	2	2	2
294	3	3	2	2
295	4	4	4	3
296	4	4	4	4
297	5	5	5	5
298	5	5	4	5
299	5	5	5	5
300	1	1	1	1
301	1	1	1	1
302	1	1	1	1
303	1	1	1	1
304	1	1	1	1
305	2	2	1	1
306	2	2	1	1
307	1	2	1	1
308	2	2	2	2
309	3	3	2	3
310	2	2	2	2
311	2	2	2	2
312	5	5	5	5
313	5	5	5	5
314	5	5	5	5
315	2	2	1	1
316	3	3	2	3
317	5	5	5	5

APPENDIX B – FUZZY EVALUATION MATRIX

	F1	F2	F3	F4	F5	F6	F7	F8
F1	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)	(0.05. 0.15. 0.25)	(0.05. 0.15. 0.25)	(0.55. 0.675. 0.80)	(0.20. 0.325. 0.45)	(0.55. 0.675. 0.80)	(0.40. 0.50. 0.60)
F2	(0.55. 0.675. 0.80)	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)	(0.05. 0.15. 0.25)	(0.55. 0.675. 0.80)	(0.20. 0.325. 0.45)	(0.55. 0.675. 0.80)	(0.40. 0.50. 0.60)
F3	(0.75. 0.85. 0.95)	(0.55. 0.675. 0.80)	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)	(0.75. 0.85. 0.95)	(0.55. 0.675. 0.80)	(0.55. 0.675. 0.80)	(0.55. 0.675. 0.80)
F4	(0.75. 0.85. 0.95)	(0.75. 0.85. 0.95)	(0.55. 0.675. 0.80)	(0.40. 0.50. 0.60)	(0.75. 0.85. 0.95)	(0.55. 0.675. 0.80)	(0.75. 0.85. 0.95)	(0.55. 0.675. 0.80)
F5	(0.20. 0.325. 0.45)	(0.20. 0.325. 0.45)	(0.05. 0.15. 0.25)	(0.05. 0.15. 0.25)	(0.40. 0.50. 0.60)	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)	(0.40. 0.50. 0.60)
F6	(0.55. 0.675. 0.80)	(0.55. 0.675. 0.80)	(0.20. 0.325. 0.45)	(0.20. 0.325. 0.45)	(0.40. 0.50. 0.60)	(0.40. 0.50. 0.60)	(0.75. 0.85. 0.95)	(0.55. 0.675. 0.80)
F7	(0.20. 0.325. 0.45)	(0.20. 0.325. 0.45)	(0.20. 0.325. 0.45)	(0.05. 0.15. 0.25)	(0.55. 0.675. 0.80)	(0.05. 0.15. 0.25)	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)
F8	(0.40. 0.50. 0.60)	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)	(0.20. 0.325. 0.45)	(0.40. 0.50. 0.60)	(0.20. 0.325. 0.45)	(0.55. 0.675. 0.80)	(0.40. 0.50. 0.60)