Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi • Cilt: 41 • Sayı: 2 • Aralık 2019, ISSN: 2149-1844, ss/pp. 338-363
DOI: 10.14780/muiibd.665054

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

SOSYOEKONOMİK YAKLAŞIMLA ZİNCİR PERAKENDE MAĞAZALARININ SEGMENTASYONU

STORE SEGMENTATION OF RETAIL CHAINS VIA SOCIO-ECONOMIC APPROACH

Emrah BİLGİÇ[*]
Özgür ÇAKIR[**]

Özet

Günümüzde perakende zincir mağazaların sayısı giderek artmaktadır. Sayıları binlerle ifade edilen bu zincirler, sayıları milyonlarla ifade edilebilecek müşterilere sosyoekonomik açıdan farklı özelliklerdeki şehirlerde hatta aynı şehirlerin farklı özelliklerdeki semtlerinde hizmet vermektedirler. Bu bakımdan, zincir firmalar tarafından özellikle pazarlama süreçleri için kararlar alınırken genel olarak mağazaların tamamına değil de bölgelere veya belirli mağaza gruplarına yönelik stratejiler geliştirmek gerekmektedir. Bu çalışmada, zincir mağazalara sahip perakende firmalarının kümeleme analizini kullanarak sosyoekonomik faktörlere göre mağazalarını nasıl segmentlere ayırabileceği araştırılmaktadır. Bu amaçla, bir perakendecinin İstanbul'daki 175 mağazasına ait farklı veriler çeşitli kaynaklardan bir araya getirilmiş ve Ward'ın kümeleme tekniği kullanılarak mağazalar altı segmente ayrılmıştır. Mağaza segmentasyonu sonucunda elde edilen segmentler incelendiğinde segmentlerin gerçekten farklı özelliklerde olduğu konum olarak birbirine yakın olan mağazaların bile farklı segmentlerde yer alabilecekleri tespit edilmiştir.

Anahtar Kelimeler: Mağaza Segmentasyonu, Kümeleme, Ward Algoritması, Zincir Mağazalar

JEL Sınıflandırması: C38, M31, C88

Abstract

The number of the retail chain stores is increasing very fast nowadays. These chains which may have thousands of stores and millions of customers are serving at the cities with different socio-economic characteristics and even at the same city of different areas with different characteristics. So in most of the decision processes, especially in marketing decisions, the chains should focus to the areas or store groups rather than focusing to the entire. In this study, our goal is to develop a methodology for retailers on how to segment their stores with cluster analysis based on multiple data sources. A

[*] Dr. Öğr. Üyesi Kayseri Üniversitesi Sosyal ve Beşeri Bilimler Fakültesi Sağlık Yönetimi Bölümü, emrahbilgic@kayseri.edu.tr

[**] Doç. Dr. Marmara Üniversitesi İşletme Fakültesi İşletme Bölümü , ocakir@marmara.edu.tr

Makale Gönderim Tarihi: 15.01.2019
Yayına Kabul Tarihi: 29.07.2019

retailer's 175 stores in Istanbul have been segmented into six segments with Ward's clustering algorithm using the data of socio-economic factors which are gathered from different sources. After clustering the stores and analyzing the characteristics of segments one would notice that the stores even closer to each other are segmented into different groups.

Keywords: Store Segmentation, Clustering, Ward's Algorithm, Retail Chain Stores

JEL Classification: C38, M31, C8

1. Giriş

Bu çalışmada, çok sayıda ve farklı kategorilerdeki müşterilere ulaşabilen zincir perakende işletmelerinin dış kaynaklardan elde edebilecekleri verilerle ve Veri Madenciliği tekniklerinden olan Kümeleme Analizini kullanarak mağazalarını nasıl gruplara ayırabileceği konusu ele alınmaktadır.

Zincir mağazalara sahip olan bir perakende firmasının farklı yerleşim yerlerinde yer alan birçok satış mağazası ve milyonlarla ifade edilebilecek müşterisi olabileceğinden, hem bu müşterilerin hem de mağazaların bulunduğu bölgelerin farklı özelliklere sahip olabileceği de kaçınılmaz bir unsurdur. Bu bakımdan firmanın, pazarlama süreçleri için kararlar verirken genel olarak mağazaların bütününe değil de bölgelere veya belirli mağaza gruplarına yönelik stratejiler geliştirmesi gerekebilmektedir.

Firma için bu yaklaşım sayesinde, tespit edilen farklı müşteri profillerine farklı hizmet anlayışı geliştirilerek müşteri memnuniyetinin yükseltilmesini sağlamak ve dolaylı olarak hem mevcut müşterilerden daha fazla gelir elde etmek hem de potansiyel müşteri grubunu gerçek müşteri haline getirerek pazar payını ve kârlılığı arttırmak mümkün olacaktır.

Daha önce, perakende firmaları ile ilgili yürütülen çalışmalarının çoğunda, firmaların sahip olduğu müşterilere ait bazı veriler (demografik veriler, satış verileri vb.) yardımıyla müşteri segmentasyonu uygulamaları yapılmıştır. Hem müşteri ile ilgili hem de mağaza ve mağaza bölgesi ile ilgili verilerin birlikte kullanımı yoluyla yapılmış olan segmentasyon çalışmaları nadirdir. Farklı kaynaklardan elde edilebilecek bu gibi verilerin kullanımı, zincir mağazalara sahip firmalar için yapılmış az sayıdaki "mağaza segmentasyonu" çalışmalarında karşımıza çıkmaktadır.

Bu çalışma, zincir mağazalara sahip perakende firmalarının farklı kaynaklardan bir araya getirdikleri sosyoekonomik faktörleri kullanarak mağazalarını nasıl segmentlere ayırabileceği hususunda bir fikir sunmaktadır.

2. Pazar Segmentasyonu

Segment kelimesi, büyük bir topluluk içindeki farklı ve tanımlanabilir karakteristik davranışlar sergileyen küçük seçkin bir grubu ifade etmektedir. Pazar segmentleri ise; bir pazarlama stratejisine gruplar halinde benzer tepki veren insanlardan veya organizasyonlardan oluşmaktadır[1].

1 Myers, J. H. (1996). Segmentation and positioning for strategic marketing decisions, s.16.

Segmentasyon, Wendell Smith'in[2] çığır açan çalışmasında, ekonomistlerin diliyle şöyle tanımlanmıştır: "Segmentasyon, yaptığı etkiyle bütünü parçalarına ayırma yeteneğine sahiptir ve daha önce tek bir olarak algılanan talep planları (demand schedule) ortaya çıkarmaya çalışmaktadır. Smith'in çalışmasından sonra pazar segmentasyonu, pazarlama alanında, hem araştırmacılar hem de uygulayıcılar için en önemli ve en çok araştırılan konulardan biri haline gelmiştir. Çünkü Smith artık "Kitle Pazarlama" çağının sonunun geldiğine işaret etmiş, basit bir şekilde bir grup ürün çeşidi sunan firmalarla (hedef homojen pazarlar), özenle seçilmiş ürün çeşitleri sunan firmaların (hedef çoklu heterojen pazar segmentleri) ayrı tutulması gerektiğine kanaat getirmiştir[3].

Pazar segmentasyonu için Smith'den sonra da birçok tanımlar yapılmış, ayrıca segmentasyon işinin nasıl yapılacağı hususunda da farklı teknikler ve yöntembilimler geliştirilmiştir.

Bu bağlamda, Peter Bennett'in[4] yaptığı basit ve anlaşılır tanımla ayrıca segmentasyon sürecinin nasıl işleyeceğinin ayrıntılı açıklanması önem arz etmektedir. Bennett pazar segmentasyonunu, bir pazarı, aynı şekilde davranan veya benzer ihtiyaçlar sergileyen müşteri alt gruplarına bölme süreci olarak açıklayıp, muhtemelen her bir alt grubun birer hedef pazar olarak seçilebileceğini ve bu gruplara farklı pazarlama stratejileri uygulanacağını vurgulamaktadır.

Myers[5] bir segmentasyon sürecini şu şekilde anlatmaktadır.

- Segmentasyon değişkenlerine karar verilir (temel değişkenler adını alır)
- Veri analizi yöntemine karar verilir.
- Segmentleri elde etmek için metodoloji uygulanır.
- Temel değişkenleri ve diğer değişkenleri kullanarak segmentler tanımlanır.
- Hedef segmentler seçilir ve
- Her bir hedef segment için bir pazarlama karması geliştirilir.

Amacı, kısaca pazarları analiz etmek, niş fırsatlar yakalamak ve üst düzey rekabet edebilme pozisyonunu alabilmek olan segmentasyonun, firmalara ne denli önemli avantajlar sağladığı hususu günümüzde bu iş için yapılan büyük yatırımlardan da anlaşılmaktadır[6]. Segmentasyon sayesinde artık, pazarlamacılar üründen ziyade müşteriye odaklanmaya başlamış, müşteri ile diyaloğun artması sonucunda da müşterilerin ihtiyaçlarının tespiti kolaylaşmış, böylece daha önce kitle pazarlama sebebiyle bilinmeyen yeni ve kârlı müşteri grupları keşfedilmiştir. Segmentasyon ayrıca, pazarın bundan sonra sadece büyük üreticiler tarafından domine edilemeyeceği müjdesini

2 Smith, W. R. (1956). Product Differentiation And Market Segmentation As Alternative Marketing Strategies. Journal of Marketing 21(1), s.5.

3 Doyle, C. (2011). *A dictionary of marketing*: Oxford University Press, s.341.

4 Bennett, P. D. (1995). Dictionary of Marketing Terms, NTC Business Books, s.165.

5 Myers, 1996, 165.

6 Weinstein, A. (2004). Handbook of market segmentation: Strategic targeting for business and technology firms, Psychology Press, s.3.

de beraberinde getirmiştir, çünkü küçük üreticilerin en azından bir tanesinin, bazı segmentlerde baskın olabileceği düşünülmektedir.[7]

2.1 Mağaza Segmentasyonu

Mağaza segmentasyonu ile ilgili akademik çalışmaların sayısının, yaptığımız araştırmalara göre çok az olduğu saptanmıştır. Yapılan araştırmaya göre erişilebilen çalışmalara bu bölümde kısaca yer verilecektir. Mağaza segmentasyonu Bermingham vd.nin[8] tanımıyla, birbiri ile bağlantılı olan mağazaları anlamlı gruplara ayırmaktır. Örneğin, 350 mağazaya sahip olan bir perakendeci, daha önce belirlenen bazı değişkenler ışığında, bütün mağazalarından sadece altı farklı mağaza grubu oluşturabilir, her bir grupta bazı özelliklere göre birbirine benzeyen mağazalar yer almaktadır.

Birçok mağazası ve milyonlarca müşterisi olan bir perakende firmasının, bu kadar yüksek sayıdaki müşterilerini segmentlere ayırmak yerine, sahip olduğu mağazalarını segmentlere ayırması daha kolay bir iş olarak görünmektedir[9]. Örneğin, müşterilerinin satın alma davranışlarına göre mağazalarını gruplara ayıracak olan bir firma bu sayede sadece müşterileri aynı davranışı sergileyen mağazalarını tespit etmekle kalmayacak aynı zamanda aynı grupta yer alıp da grubundaki mağazalara göre daha düşük veya daha yüksek performans sergileyen mağazalarını da tespit etmiş olacaktır[10].

Ayrıca kümeleme yoluyla belki yüzlerce mağaza belirli özelliklerdeki birkaç gruba ayrılacağı için firmaların her türlü planlama, satış, hatta satın alma süreçleri daha da kolaylaşacaktır. Mağazaların satış performanslarına göre, büyüklüklerine veya konseptlerine göre, bulundukları bölgelere veya hedef müşterilere göre (emekliler, öğrenciler vb.) kümelere ayrılması ile beraber firma mağazalarını mikro pazarlama için hazırlamış olacaktır (Donofrio, 2009).[11]

Ayrıca Infoys, Oracle, Category Management Knowledge Group, The Parker Avery Group, Wilson Perumal & Company, Symphony EYC, Revionics, Pitney Bowes, Data Ventures vb. kurumlar da mağaza kümeleme konusunda danışmanı olduğu şirketlere hizmet vermektedir.

Mağaza segmentasyonunda kullanılacak temel değişkenler şu şekilde sınıflandırılmıştır.[12],[13]

- Mağaza ile ilgili: Ciro, kâr, satış alanı, konsept vb.

7 Doyle, (2011), 342.

8 Bermingham, P., Hernandez, T. and Clarke, I. (2013). "Network Planning and Retail Store Segmentation: A Spatial Clustering Approach." International Journal of Applied Geospatial Research (IJAGR) 4(1): s.68.

9 Vohra, G. (2011). "Store Clustering." http://analyticstraining.com/2011/store-clustering/ (26.05.2016).

10 Bornac, G. (2015). "The Power of Store Clustering." http://www.manh.com/resources/articles/2015/08/27/power-store-clustering (30.05.2016).

11 Donofrio, T. J. (2009). Advanced Planning and Optimization Part 3: Store Clustering. Retail Systems and Services, www.risnews.edgl.com/retail-news/Advanced-Planning-and-Optimization-Part-3—Store-Clustering38904 26.05.2016.

12 Wilson Company (2013). A Simple Approach to Retail Clustering. www.wilsonperumal.com/media/publications/PDFs/Vantage_Point_2013_Issue3.pdf (30.05.2016).

13 Bermingham vd., (2013).

- Müşteri ile ilgili: Demografik özellikler, satın alma verileri vb.
- Bölge ile ilgili: Bölgede yaşayan insanların demografik özellikleri, iklim koşulları, bölgenin sosyoekonomik özellikleri, rakip mağazaların sayısı vb.

Hawkes ve McLaughlin[14] çalışmalarında mağazaların geo-demografik kümelenmesi ismini verdiği yöntem ile 78 adet süpermarketi, satış performanslarına göre kümelemiştir. Bu çalışmada amaç ürün bazlı (margarin) satışların arttırılmasıdır.

Koehn[15] Starbucks firmasının mağazalarını kümeleyerek, markasının farkındalığını arttırdığını vurgulamış, fakat firmanın bu uygulamayı nasıl yaptığı bilgisine çalışmasında yer verilmemiştir. Yine Lippmann[16] da çalışmasında birçok farklı bölgede mağazası bulunan market perakendecilerine mağaza segmentasyonunu çok önemli bir strateji olarak tavsiye etmiş fakat makalede herhangi bir uygulamaya yer verilmemiştir.

Clarke vd.[17], ayrıca Bermingham vd.[18] mağaza kümeleme tekniğini, perakende firmalarının yer tespiti için karar verme süreçlerinde kullanmışlardır. Her iki çalışmada da MIRSA ismi verilen, perakende firmaları için hazırlanılmış erişime kapalı özel bir paket program ile mağazalara ait değişkenlere ait veriler programa girildikten sonra, mağazalar kümelere ayrılmaktadır. Bu programın temelinde bilişsel haritalar (cognitive maps) kullanılmaktadır. Program, şirketlere satılan özel bir program olduğu için içeriği hakkında fazla bilgi verilmemiş, ayrıca çalışmanın uygulama bölümünde analiz edilen firma ve kullanılan değişkenler gizli tutulmuştur. Fakat sadece değişkenler genel olarak; ticaret yapılan bölgenin demografik özellikleri, rekabet, ekonomik durum gibi, açıklanmışlardır. Bu gibi değişkenler bizim çalışmamızda da kullanılacak olan değişkenlerdir. MIRSA programı segmentasyon için kümeleme tekniklerinden k-ortalamalar tekniği kullanmaktadır.

Mendes ve Cardoso[19] da performans değerlendirmesi yapmak ve yer tespit çalışmalarını desteklemek amacıyla outlet mağazalarını kümelere ayırmışlardır. Bu çalışmada ise kullanılan değişkenler açıkça anlatılmıştır. Değişkenler mekân ve mağaza ile ilgili değişkenler (örneğin satış alanı, demografik değişkenler), müşteri tercihleri, tutum ve davranışları ile ilgili değişkenler olarak sınıflara ayrılmıştır. Bu çalışmada kümeleme tekniklerinden olan Ward'ın tekniği kullanılmıştır. Çalışmada az sayıda mağaza (25 mağaza) çok sayıdaki değişkenlere göre (250

14 Hawkes, G. F. and McLaughlin, E. W. (1994). STARS: Segment Targeting at Retail Stores, Department of Agricultural, Resource, and Managerial Economics, Cornell University.

15 Koehn, N. F. (2001). "Howard Schultz and Starbucks Coffee Company." Harvard Business School Cases.

16 Lippman, B. W. (2003). "Retail revenue management—Competitive strategy for grocery retailers." Journal of revenue and pricing management 2(3): 229-233.

17 Clarke, I., Mackaness, W. ve Ball, B. (2003). Modelling Intuition in Retail Site Assessment (MIRSA): making sense of retail location using retailers' intuitive judgements as a support for decision-making.The International Review of Retail, Distribution and Consumer Research 13(2): 175-193.

18 Bermingham vd., (2013).

19 Mendes, A. B. and Cardoso, M. G. M. S. (2006). "Clustering supermarkets: the role of experts." Journal of Retailing and Consumer Services 13(4): 231-247.

değişken) kümelendiği için, sonuçlar uzman görüşlerine başvurularak değerlendirilmiştir. Çalışmada mağaza müşterilerine anket yapılarak veri toplanmıştır.

Kargari ve Sepehri[20], otomotiv parçaları üreten bir firmanın sahip olduğu mağazalarını, ulaştırma maliyetlerini azaltmak amacıyla, kümelere ayırmıştır. Bu çalışmada kümeleme tekniklerinden k-ortalamalar tekniği kullanılmıştır. 75 ürün firma için çok kritik kabul edilmiş ve bu ürünlerin satışına göre mağazalar gruplara ayrılmıştır. Son üç yılın verisi 815 mağaza için incelenmiştir.

Yapılan bu ve benzeri çalışmaların genelinde şu sonuç çıkarılabilmektedir: Segmentasyon için hangi değişkenlerin kullanıldığına çalışmaların çoğunda açıkça yer verilmemiş, uygulama ve sonuç kısımları ise detaylandırılmamıştır. Bu çalışmada kullanılacak olan değişkenler açık bir şekilde tanımlanacak ve segmentasyon sonuçları da detaylı bir şekilde tartışılacaktır.

2.2. Segmentasyon Analizinde Kullanılan İstatistiksel Teknikler

Segmentasyon analizini yapabilmek için bazı tekniklere ihtiyaç vardır. Çalışmalarda genel olarak aşağıda sıralanan istatistiksel teknikler, ya direk segmentasyon işini yapmada ya da segmentasyon sürecine destek olarak kullanılmaktadırlar[21],[22].

- Ki-kare Otomatik Etkileşim Detektörü (CHAID)
- Konjoint Analizi
- Sınıflandırma ve Regresyon Ağacı (C&RT)
- Diskriminant Analizi
- Kümeleme Analizi

İlk dört tekniğe kısaca değinildikten sonra bu çalışmada da kullanılacak olan Kümeleme Analizi konusu detaylı bir şekilde ayrı bir başlık altında incelenecektir.

2.2.1 Ki-kare Otomatik Etkileşim Dedektörü (CHAID)

CHAID, bir karar ağacı tekniğidir. Literatürde Otomatik Etkileşim Dedektörü (AID), Sınıflandırma ve Regresyon Analizi, Genetik Algoritma gibi farklı tekniklerle benzerlik gösteren bu teknik Gordon Kass tarafından geliştirilmiştir[23]. CHAID sınıflandırma ve tahmin amacıyla, ayrıca değişkenler arasındaki etkileşimi keşfetmek amacıyla da kullanılmaktadır. Pazarlama alanındaki en sık kullanım amacı ise, müşteri gruplarının belirlenmesi ve müşterilerin bazı değişkenlere verdikleri tepkilerin diğer değişkenleri nasıl etkileyeceğinin tahminidir. CHAID

20 Kargari, M. and Sepehri, M. M. (2012). "Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs." Expert Systems with Applications 39(5): 4740-4748.

21 Lilien, G. L. and Kotler, P. (1983). Marketing decision making: A model-building approach, Harper & Row New York, NY.

22 Myers, (1996), 31.

23 Díaz-Pérez, F. M. and Bethencourt-Cejas, M. (2016). CHAID algorithm as an appropriate analytical method for tourism market segmentation, Journal of Destination Marketing & Management.

uygulanmadan önce bağımlı değişken ve bağımsız değişkenlerin belirlenmesi gerekir, çünkü segmentasyon işi bu değişkenlerin etkileşimine göre yapılır[24].

2.2.2 Konjoint Analizi

Pazarlama Araştırmasında sıkça kullanılan konjoint analizi, bir ürün/hizmetin özelliklerine tüketicilerin ne kadar değer atfettiğini ölçebilen, tüketici tercihlerine ürün özelliklerinin ve düzeylerinin katkılarını belirleyebilen, optimum özellikli ürün tasarlamaya destek olan ve tüketici kararları için bir model oluşturabilen bir tekniktir[25],[26]. Green ve Rao[27] pazarlama çalışmalarında konjoint analizini kullanmaya başlamışlardır. Bu tekniği kullanarak yapılan çalışmalardan bazıları şu şekildedir: Djokic vd.[28], Guo vd.[29], Ceylan[30].

2.2.3 Regresyon Ağacı

Regresyon Analizi, Karar Ağaçları ile birlikte kullanılarak müşterileri farklı özelliklerine göre gruplara ayırma işine destek olur. Sınıflandırma ve Regresyon tekniği (C&RT) çok yaygın bir şekilde pazarlama alanındaki uygulamalarda kullanılmaktadır. Bu tekniği kullanarak yapılan çalışmalar şu şekilde sıralanabilir: Ekinci vd.[31], De Ona ve De Ona[32], Mehrotra ve Agarwal[33], Timor ve Şimşek[34].

2.2.4 Diskriminant Analizi

Çok değişkenli istatistik tekniklerinden biri olan Diskriminant Analizi, önceden sınıflandırılmış iki ya da daha fazla grubu birbirinden ayıran faktörleri tespit eden ve grup dışından alınan bir

24 Bijak, K. and Thomas, L. C. (2012). Does segmentation always improve model performance in credit scoring?, Expert Systems with Applications, 39(3): 2433-2442.

25 Nakip, M. (2006). Pazarlama araştırmaları teknikler ve (SPSS destekli) uygulamalar, Ankara, Seçkin Yayıncılık, s.553.

26 Yaman, T. T., Çakır, Ö (2017). Üniversite Tercihlerinin Seçime Dayalı Konjoint Analizi ile Belirlenmesi, Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi 1(1): 65-84.

27 Green, P. E. and Rao, V. R. (1971). "Conjoint measurement for quantifying judgmental data." Journal of Marketing research: 355-363.

28 Djokic, N., Salai, S., Kovac-Znidersic, R., Djokic, I. and Tomic, G. (2013). "The use of conjoint and cluster analysis for preference-based market segmentation." Engineering Economics 24(4): 343-355.

29 Guo, Y., Denizci Guillet, B., Kucukusta, D. and Law, R. (2015). "Segmenting Spa Customers Based on Rate Fences Using Conjoint and Cluster Analyses." Asia Pacific Journal of Tourism Research: 1-19.

30 Ceylan, H. H. (2013). "Perakende Sektöründe Konjoint ve Kümeleme Analizi ile Fayda Temelli Pazar Bölümlendirme." Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 20(1): 141-154.

31 Ekinci, Y., Ulengin, F. and Uray, N. (2014). "Using customer lifetime value to plan optimal promotions." The Service Industries Journal 34(2): 103-122.

32 De Oña, R. and De Oña, J. (2015). "Analysis of transit quality of service through segmentation and classification tree techniques." Transportmetrica A: Transport Science 11(5): 365-387.

33 Mehrotra, A. and Agarwal, R. (2009). "Classifying customers on the basis of their attitudes towards telemarketing." Journal of Targeting, Measurement and analysis for Marketing 17(3): 171-193.

34 Timor, M. and Şimşek, U. (2008). "Veri Madenciliğinde Sepet Analizi ile Tüketici Davranışı Modellemesi." Yönetim, 19 (59): 3-10.

gözlemin hangi grup ile ilişkilendirilebileceğini ortaya çıkaran bir tekniktir[35]. Diskriminant analizi ile sınıflandırma işi, farklı gruplardaki nesnelerin birbirleri arasındaki varyansın maksimum, aynı sınıftaki nesnelerin birbirleri arasındaki varyansı da minimum yapan bir anlayışa dayanmaktadır[36]. Bu tekniği kullanarak bazı yapılan segmentasyon çalışmaları şu şekildedir: Zhiyu ve Congdong[37], Nakahara ve Yada[38].

3. Kümeleme Analizi

Everitt[39] bir kümenin tanımını basitçe şöyle yapar: Bir küme, birbirine benzeyen bir takım birimden oluşur. Kümeleme ise veri noktalarını, küme adı verilen doğal gruplara ayırma işidir, öyle ki bir grup içerisindeki veri noktaları birbirine çok benzemekte, farklı kümelerdeki veri noktaları ise mümkün olduğunca birbirine benzememektedir[40].

Daha teknik bir tanım ise şu şekildedir: Kümeleme, sonlu bir grup nesneye yapılan bir sınıflandırma çeşididir. Nesneler arasındaki ilişki bir yakınlık matrisinde temsil edilirler. Eğer bahsi geçen nesneler d-boyutlu ölçü uzayındaki örnekler veya noktalar ise, yakınlıklar nokta çiftleri arasındaki uzaklıklardır (örneğin Öklidyen uzaklık gibi). Eğer, nokta çiftleri arasında bir uzaklık veya benzerlik olmazsa kümeleme analizi de olmaz. Bunun sebebi benzerlik (yakınlık) matrisinin kümeleme analizini yapabilmek için tek girdi olmasıdır[41]. Bu tanımdan anlaşılacağı üzere, kümeleme analizi ham veri kümesi ile değil, kümelemesi yapılacak gözlem çiftlerinin benzerlik (yakınlık) veya benzeşmezlik (uzaklık) değerlerinin oluşturduğu matris ile yapılmaktadır.

Aggarwal[42] kümelemenin bazı önemli uygulamalarını şu şekilde anlatmaktadır.

- Verinin özetlemesi: Daha önce belirtildiği gibi, kümeleme problemi, bir çeşit veri özetlemesi olarak düşünülebilir. Zaten birçok veri madenciliği uygulamasında da verinin daha iyi anlaşılması açısından kümeleme analizi sürecin ilk aşamasıdır. Örneğin kümeleme, sınıflandırma modelleri için, birliktelik analizleri için ve aykırı değer analizleri için ilk aşamadır.
- Müşteri Segmentasyonu: Benzer müşterilerinin ortak satın alma davranışlarını inceleyebilmek, firmalar tarafından daima arzu edilen bir uygulamadır. Bu uygulama müşteri segmentasyonu ile yapılır. Kümeleme analizinin belki de en çok kullanıldığı alan budur.

35 Nakip, (2006), 477.

36 Badea, L. M. (2014). "Predicting consumer behavior with artificial neural networks." Procedia Economics and Finance 15: 238-246.

37 Zhiyu, Z. ve Congdong, L. (2009). Research on Application to Customer Classification in Management Decision-Making Based on Multivariate Statistics. Information Technology and Applications, 2009. IFITA'09. International Forum on, IEEE.

38 Nakahara, T. ve Yada, K. (2012). "Analyzing consumers' shopping behavior using RFID data and pattern mining." Advances in Data Analysis and Classification 6(4): 355-365.

39 Everitt, B. (1974). Cluster analysis 122, Heinemann, London.

40 Zaki, M. J. ve Meira Jr, W. (2014). Data mining and analysis: fundamental concepts and algorithms, Cambridge University Press, s.28.

41 Jain, A. K. ve Dubes, R. C. (1988). Algorithms for clustering data, Prentice-Hall, Inc., s.55.

42 Aggarwal, C. C. (2015). Data mining: The textbook. Switzerland, Springer, s.153.

- Sosyal ağlar analizi: Ağ verisinde, bağlantı ilişkileri ile kümelenen düğüm noktaları, çoğu zaman benzer arkadaş gruplarını veya cemiyetleri işaret etmektedir. Sosyal ağlar analizinde bu tip cemiyetleri veya arkadaş gruplarını saptamak en çok çalışılan konulardan biridir.

3.1 Kümeleme Analizi Süreci

Kümeleme analizinin nasıl yapılacağı en genel anlamda; değişkenlerin seçimi, uzaklık fonksiyonu ile kümeleme algoritmasının seçimi, kümelemenin geçerlemesi ve sonuçların yorumlanması şeklinde sıralanabilir. Dört adımdan oluşan aşamalarının her biri birbiriyle bağlantılı olup birbirlerinden beslenirler ve elde edilen kümeleri belirlerler[43]. Çalışmamızın uygulama bölümünde, kümeleme analizinde kullanılacak olan değişkenler Türkiye İstatistik Kurumu, Google Haritalar ve bazı emlak şirketlerinden, yani farklı kaynaklardan bir araya getirilmiştir. Uzaklık ölçüsü olarak Gower'in[44] uzaklık ölçüsü kullanılacak ve algoritma olarak da hiyerarşik kümeleme tekniklerinden Ward'ın tekniği kullanılacaktır.

3.2. Uzaklık Fonksiyonları

Uzaklık fonksiyonunu belirleme veri madenciliği uygulamalarında oldukça önem taşıyan bir kavramdır. Çünkü çok yaygın olarak kullanılan kümeleme yöntemleri bir yakınlık (proximity), benzerlik veya ilişki göstergesine ihtiyaç duyarlar. Bu gösterge ham veriden elde edilmektedir[45].

Yakınlıktan kastedilen literatürde sıkça karşılaşılan benzeşmezlik (dissimilarity) veya benzerliğe (similarity) karşılık gelmektedir[46]. Benzerlik, iki nesnenin birbirine ne kadar yakın olduğunun ölçüsü iken, benzeşmezlik ise aksine ne kadar uzak olduğunun ölçüsüdür. Benzeşmezlik genel olarak bir uzaklık ölçüsü ile hesaplanmaktadır, Öklidyen uzaklık en bilinen uzaklık ölçüsüdür.

Uzaklık ölçüsü için literatürde Öklidyenden başka farklı uzaklık fonksiyonları da mevcuttur ve uzaklık fonksiyonunun seçiminin kritik olma sebebi, elde edilecek sonuçların kalitesini direkt olarak etkilemesindendir. Uzaklık fonksiyonları için birçok yöntemin karşımıza çıkma nedeni ise, fonksiyonun veri tipine, veri boyutuna ve veri dağılımına oldukça hassas olmasındandır[47].

Uzaklık ölçülerinin aldığı değerler negatif olamazlar. Eğer birbirine yakın iki nokta ile ilgileniliyorsa bunlar arasındaki uzaklık sıfıra yaklaşırken birbirine uzak iki nokta söz konusu ise uzaklıkları bire yaklaşırlar. Farklı ölçekle ölçülmüş değişkenler için farklı yöntemler kullanarak, gözlem değerleri (veri noktaları) arasındaki uzaklıkları bulabiliriz.

43 Xu, R. and Wunsch, D. C. (2009). "Clustering." from http://site.ebrary.com/id/10257659.

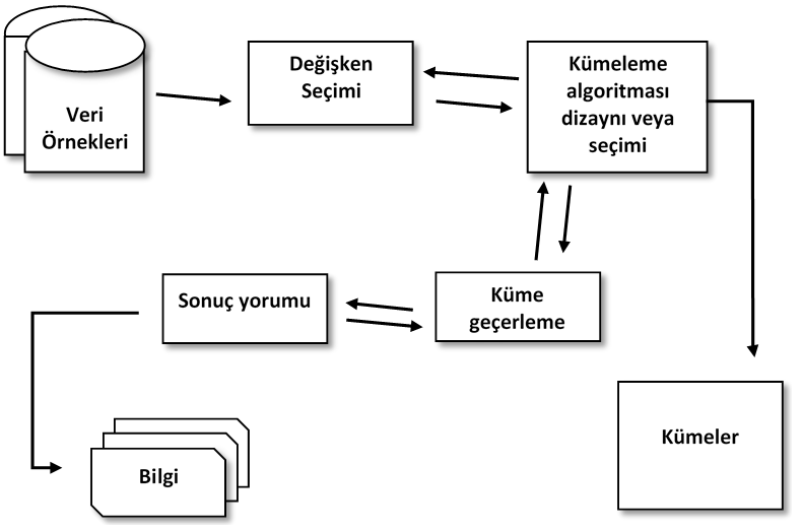
44 Gower, J. C. (1971). "A general coefficient of similarity and some of its properties." Biometrics: 857-871.

45 Aggarwal, 2015, s.88.

46 Jain, 1988, s.11.

47 Aggarwal, 2015, 88.

Şekil 1: Kümeleme Prosedürü



Kaynak: Xu ve Wunsch 2009, s.6

3.2.1 Karma Tipteki Veriler için Uzaklık Ölçüsü

Gerçek hayat uygulamalarında yaygın olarak farklı ölçeklerle ölçülmüş değişkenlerin aynı veri kümesi içerisinde yer aldığını görürüz. Bu tip verilerle kümeleme analizi yapmanın farklı yolları mevcuttur. Örneğin her bir farklı ölçekle ölçülmüş verilere ayrı ayrı kümeleme yapılabilir. Fakat kümelemelerin sonuçları farklı çıkarsa, bu sonuçları bağdaştırmak zor olabilir. Bu sebeple en iyi yol veri kümesini bir bütün olarak alıp analizi yapmaktır. Bütün değişken tiplerini tek bir yakınlık/uzaklık matrisinde birleştirmek en ideal yol olarak görülmektedir. Log-Olabilirlilik Uzaklığı ve Gower'in uzaklığı akla ilk gelen kullanılabilecek ölçülerdir. Çalışmamızda Gower'in önerdiği uzaklık ölçüsü temel olarak alınacaktır.

Gower[48] farklı ölçekle ölçülmüş değişkenleri barındıran veri kümeleri için (aralıklı, nominal ve ikili değişkenler) uzaklık ölçüsü tasarlanmıştır. Daha sonraki 1990 yılında ise Gower'in formülünü genelleştirilerek sıra ve oran ölçekli değişkenler de hesaplamaya dâhil edilmiştir[49]. Her ne kadar Gower'in tanımı benzerlik katsayısı hesaplasa da daha önce belirtildiği gibi bu katsayı benzeşmezlik katsayısına $d(i,j) = 1 - s(i,j)$ yardımıyla dönüştürülebilir. Gower'in uzaklık ölçüsü, birçok alanda kümeleme çalışmaları için kullanılan bir ölçüdür[50].

48 Gower, (1971).

49 Kaufman, L. ve Roussew, P. (1990). Finding Groups in Data-An Introduction to Cluster Analysis. A Wiley-Science Publication John Wiley & Sons, Inc., s.32.

50 Pavoine, S., Vallet, J., Dufour, A. B., Gachet, S. ve Daniel, H. (2009). On the challenge of treating various types of variables: application for improving the measurement of functional diversity, Oikos 118(3): 391-402.

Kauffman ve Rousseeuw[51] benzeşmezlik matrisini hesaplamak için karma ölçeklerle ölçülmüş değişkenleri nasıl kullanacaklarını şöyle anlatmışlardır: Diyelim ki veri kümemiz *p* sayısı kadar farklı ölçekle ölçülmüş değişkeni barındırsın. *i* ve *j* nesneleri arasındaki benzeşmezlik şöyle bulunur;

$$d(i,j) = \frac{\sum_{f=1}^{p} \delta_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^{p} \delta_{ij}^{(f)}}$$

Buradaki $\delta_{ij}^{(f)}$ *f* inci değişken için eğer x_{if} ve x_{jf} veri kümemizde yer alıyorlarsa 1'e, herhangi biri kayıpsa 0'a eşittir. Ayrıca, f değişkeni asimetrik ikili değişkense ve *i* ile *j* nesneleri 0-0 eşleşmesine sahip ise yine 0'a, diğer durumlarda ise 1'e eşittir. Formüldeki diğer terim ise, f değişkeninin *i* ve *j* arasındaki benzeşmezliğe katkısıdır.

Tablo 1: Bazı Uzaklık Ölçüleri ve Formülleri

Uzaklık ölçüsü	Formül
Manhattan Uzaklığı	$d(x,y) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$
Öklit Uzaklığı	$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
Tchebyschev Uzaklığı	$d(x,y) = max_{i=1,2,\ldots,n} \lvert x_i - y_i \rvert$
Minkowski Uzaklığı	$d(x,y) = \sqrt[p]{\sum_{i=1}^{n} (x_i - y_i)^p}, p > 0$
Canberra Uzaklığı	$d(x,y) = \sum_{i=1}^{n} \frac{\lvert x_i - y_i \rvert}{x_i + y_i}$

Örneğin, f değişkeni ikili veya nominal değişken ise o halde $d_{ij}^{(f)}$ şu şekilde tanımlanır.

$$d_{ij}^{(f)} = 1 \text{ eğer } x_{if} \neq x_{jf}$$

$$= 0 \text{ eğer } x_{if} = x_{jf}$$

Diyelim ki, f değişkeni aralıklı ölçekle ölçülmüşse o zaman $d_{ij}^{(f)}$:

$$d_{ij}^{(f)} = \frac{\lvert x_{if} - x_{jf} \rvert}{R_f}$$

51 Kaufman, (1990), 35.

Buradaki *R*, *f* değişkeni için yayılma bandıdır. Eğer *f* değişkeni asimetrik ikili olarak ölçülmüşse, Jaccard'ın[52] katsayısı benzeşmezlik matrisini hesaplamak için uygun olur. Eğer bütün değişkenler aralıklı ölçülmüşse, Gower'in formülü Manhattan uzaklık formülü halini almaktadır. İlk olarak değişkenler yayılma bantlarına bölünür. Gower'in bu formülü $d_{ij}^{(f)}$ değerlerini [0 – 1] aralığına sınırlamaktadır, yani her bir değişken benzeşmezliğe 0 ile 1 arasında katkı yapmaktadır ve *d(i,j)* sonucu da 0 ile 1 arasında olmaktadır.

3.3. Kümeleme Algoritmaları

Kümeleme algoritmasının seçimi de uzaklık fonksiyonu ile bağlantılı bir durumdur. Çünkü algoritmaların çoğu bir benzerlik ölçüsüne bağlı olarak çalışmaktadır. Bu adımda benzerlik ölçüsü seçilir ve bir amaç fonksiyonu ışığında kriterler belirlenir. Benzerlik ölçüsü kümeleme işini doğrudan etkileyecektir.

Daha önce belirtildiği gibi, çok boyutlu veri kümelerinin yapısındaki çeşitliliği kümeleyebilecek veya bu veri kümesinden doğru veya en iyi kümeleri ortaya çıkarabilecek araştırmacılar tarafından ortak kabul görmüş bir kümeleme tekniği yoktur[53],[54]. Literatürde birçok kümeleme tekniği olduğu gibi, bu teknikler gün geçtikçe geliştirilmekte ve hatta yeni teknikler de bu teknikleri uygulayabilecek yeni yazılımlar ve paket programlarla birlikte ortaya çıkmaktadır. Genel olarak kümeleme teknikleri iki yaklaşım altında toplanmaktadır.

- Hiyerarşik olmayan kümeleme
- Hiyerarşik kümeleme

3.3.1. Hiyerarşik Olmayan Kümeleme Algoritmaları

Hiyerarşik olmayan kümeleme tekniklerinde, küme sayısı (k) daha önce araştırmacı tarafından bilinmektedir. Bu tekniklerin en bilinenleri, temsilciye dayalı algoritmalar başlığı altında incelenebilir.

Temsilciye dayalı algoritmaların kümeleme algoritmaları arasında en basit ve yaygın olmasının temel sebebi, bu algoritmaların veri noktalarını kümelemek için direkt olarak uzaklıkları (veya benzerlikleri) temel almalarıdır. Bu algoritmalarda kümeler arasında herhangi bir hiyerarşi söz konusu olmayıp, kümeler tek hamlede oluşturulmaktadır. Bu oluşum, bir grup temsilcinin birleşmesiyle ortaya çıktığı için bu algoritmalar temsilciye dayalı algoritmalar ismini almıştır. Temsilciler belirlendiğinde, veri noktalarını en yakın temsilcilere atamak için, bir uzaklık fonksiyonu da kullanılabilir[55].

52 Jaccard, (1908).

53 Xu, (2009), 7.

54 Hennig, C. ve Liao, T. F. (2013). "How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification." Journal of the Royal Statistical Society: Series C (Applied Statistics) 62(3): 309-369.

55 Aggarwal, (2015), 159.

D veri kümesinin, *d* boyutlu uzayda, *n* kadar veri noktası $\overline{X_1}\ldots\overline{X_n}$ içerdiğini düşünelim. Amaç, aşağıdaki amaç fonksiyonu *O*'yu minimize edecek, *k* kadar temsilci $\overline{Y_1}\ldots\overline{Y_k}$ belirlemektir.

$$O=\sum_{i=0}^{n}\left[min_j Dist(\bar{X}_i,\bar{Y}_j)\right]$$

Bu şu anlama gelmektedir: Farklı veri noktalarının kendilerine en yakın temsilcilere olan uzaklıkları toplamı minimize edilmelidir. Veri noktalarının temsilcilere atanması, temsilcilerin $\overline{Y_1}\ldots\overline{Y_k}$ seçimine bağlıdır. Temsilciye dayalı algoritmaların genel olarak şu adımları takip ettiği söylenebilir[56].

- (Atama adımı) Her bir veri noktasını, uzaklık fonksiyonunu *Dist(.,.)* kullanarak, kendilerine en yakın olan temsilciye ata ve kümeleri $C_1\ldots C_k$ olarak adlandır.
- (Optimizasyon adımı) Her bir küme C_j için, yerel amaç fonksiyonunu $\sum_{\overline{X_i}\in C_j}\left[Dist(\bar{X}_i,\bar{Y}_j)\right]$ minimize edecek en uygun temsilciyi $\bar{Y}_j$ belirle.

3.3.1.1 k-ortalamalar Algoritması

Kümeleme analizinde belki de en çok kullanılan tekniktir. Bu teknikte genellikle Öklidyen uzaklık girdi olarak kullanılmaktadır.

k-ortalamalar algoritmasında; veri noktalarının, kendilerine en yakın temsilcilere olan Öklidyen uzaklıklarının karesi, kümelemenin amaç fonksiyonunun değerini belirler. Bu yüzden aşağıdaki ifadeyi yazmak mümkündür[57].

$$Dist(\bar{X}_i,\bar{Y}_j)=\left\|\bar{X}_\iota-\bar{Y}_j)\right\|_2^2.$$

Burada $\|.\|_p$, L_p normunu temsil etmektedir. $Dist(\bar{X}_\iota,\bar{Y}_j)$ ifadesi, bir veri noktasının kendisine en yakın olan temsilciye yaklaştırılmasıdır. Böylece, farklı noktalar üzerinden hesaplanan Hata Kareleri Toplamı (HKT) minimize edilir. Daha geniş bir anlatımla, k-ortalamalar ilk etapta, veri uzayında rastgele k kadar nokta oluşturarak bu noktaları küme ortalamaları olarak alır. Her bir yineleme iki adımdan oluşmaktadır. Birinci adımda küme ataması yapılır, ikinci adımda ise merkez güncellenir. Küme atama aşamasında; örneğin elimizde k küme ortalaması olsun; ilk etapta her bir nokta (gözlem) kendisine en yakın olan ortalamaya atanır ve böylece kümeler oluşmaya başlar. Her bir kümeyi (C_i) oluşturan noktalar, μ_i'ye diğer küme ortalamalarına göre daha yakın olan noktalardır. Merkez güncelleme aşamasında ise, her bir kümedeki (C_i; (i=1,2…k)) noktalar ile yeni ortalamalar hesaplanır. Küme atama ve merkez güncelleme aşamaları, sabit bir noktaya ulaşıncaya veya yerel minimuma ulaşıncaya kadar yinelenir durur.

$$HKT(C)=\sum_{i=1}^{k}\sum_{x_j\in C_i}\left\|x_j-\mu_i\right\|^2$$

56 Aggarwal, (2015), 160.

57 Aggarwal, (2015), 162.

3.3.1.2 k-Medoids Algoritması

k-medoids algoritmasında temsilci her zaman veri tabanından seçilmektedir. k-medoids algoritmasının diğer algoritmalara tercih edilmesinin nedenleri şöyle sıralanabilir. Birincisi k-ortalamalar algoritmasındaki temsilcinin o kümede yer alan aykırı bir değer sebebiyle sapmış olabileceği ihtimali ve ikinci sebep ise, karma veri tipinde yer alan veri noktalarının, ortalama veya medyan ile temsilci seçiminin bazen zor olmasıdır. Bu yüzden, k-medoids algoritmasının en önemli özelliği, veri kümesine uygun benzerlik veya uzaklık fonksiyonu tanımlanabildiği sürece, bu algoritmanın her veri tipi için kullanılabiliyor olmasıdır[58],[59]

k-medyanlar algoritması, ile k-medoids algoritması birbirine karıştırılmamalıdır. k-medoids temsilciyi orijinal veri tabanından seçerken, her bir boyutta bağımsız olarak seçilecek medyanların, *d* boyuttaki temsilcisi orijinal veri kümesine ait olmayabilir.

3.3.2. Hiyerarşik Kümeleme Algoritmaları

Hiyerarşik algoritmalar veri kümesine uzaklıkların oluşturduğu matrisi kullanarak kümeleme analizi uygular. Hiyerarşik kümeleme yönteminin en önemli faydası, oluşacak olan hiyerarşinin farklı seviyelerinin uygulama açısından farklı bakış açıları sunmasıdır[60],[61]

Hiyerarşik teknikler birleştirici (alttan üste) ve bölücü (üstten alta) yöntemler olarak ikiye ayrılırlar. Birleştirici yaklaşım alttan üste doğru mantığını kullanır, ilk başta n veri noktasının her biri ayrı ayrı küme olarak düşünülür ve yineleme esasına dayanarak, birbirine en benzer parçalar, tek bir küme oluşuncaya kadar birleşir. Bölücü yaklaşım ise birleştirici yaklaşımın aksine üstten alta doğru mantığını kullanır. İlk etapta bütün veri noktalarının oluşturduğu tek bir küme yineleme esasına dayanarak, her bir nokta başlı başına birer küme oluncaya kadar bölünür.

Hiyerarşik kümelemenin amacı, d boyutlu bir uzayda n veri noktası için, bir dizi bölümler oluşturmaktır ve bu bölümler bir ağaç yardımı ile veya dendrogram denilen kümelerin hiyerarşisini gösteren kullanışlı bir şekille görselleştirilir. Ağacın tabanında her bir nokta ayrı ayrı birer küme oluşturmakta, en üst noktasında ise bütün noktalar birleşerek tek bir küme oluşturmaktadır. Dolayısıyla dendrogram kümeleme analizinde, her bir adımda gerçekleşen birleşmeleri veya bölünmeleri gözler önüne sermektedir[62]. Aşağıdaki şekilde *a, b, c, d, e, f, g* ve *h* gözlemlerinin nasıl birleşip kümeler oluşturduğu dendrogramla gösterilmiştir.

58 Aggarwal, (2015), 165.

59 Zaki, (2014), 333.

60 Everitt, B. ve Hothorn, T. (2011). Cluster analysis. An Introduction to Applied Multivariate Analysis with R, Springer**,** s.55.

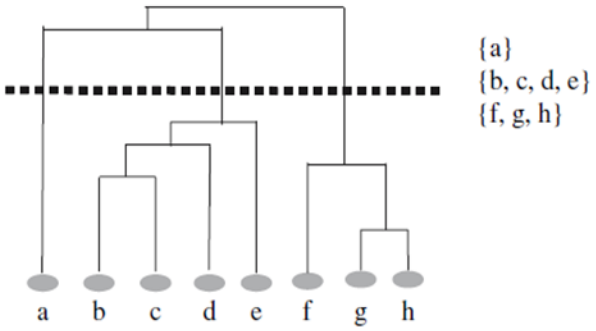
61 Aggarwal, (2015), 166.

62 Zaki, (2014), 364.

3.3.2.1 Minimum Varyans: Ward'ın Yöntemi

Bu ölçüt birleşme sonucunda daha önceden belirlenmiş bir amaç fonksiyonunda oluşacak olan değişmeyi minimum yapmayı hedefler. Birleşmeler belirlenen amaç fonksiyonunda kötüleşmeye

Şekil 2: Dendrogram Örneği



neden olmaktadır. Bu sebeple amaç fonksiyonunda en az değişim olmasını sağlayacak birleşme seçilir. Varyanstaki değişimi kullanmak yerine, birleşme kriteri olarak Hata Kareleri Toplamı da kullanılabilir. Bu yöntem aslında merkezi (centroid) yöntemin diğer bir çeşididir. Ward her bir adımda birleşebilecek her bir küme çiftini göz önüne almış ve birleşmeleri halinde "malumat kaybı" artışının minimum olacağı iki küme seçmiştir. Ward bu kaybı, Hata Kareleri Toplamı olarak tanımlamıştır[63]. Dolayısıyla bu ölçüde iki küme arasındaki uzaklık iki küme birleştiğinde Hata Kareleri Toplamındaki artış olarak tanımlanacaktır. C_i kümesinin HKT şu formülle hesaplanır[64].

$$HKT_i = \sum_{x \in C_i} \|x - \mu_i\|^2$$

$C=\{C_1, C_2, \ldots, C_m\}$ kümelemesindeki HKT şöyle hesaplanır.

$$\text{HKT} = \sum_{i=1}^{m} HKT_i = \sum_{i=1}^{m} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Belirtildiği gibi örneğin, C_i ve C_j arasındaki uzaklık, C_i ve C_j kümeleri birleşip C_{ij} kümesini oluşturduğu zaman, HKT'deki net değişim değeri olarak açıklanır.

$$\delta(C_i, C_j) = \Delta HKT_{ij} = HKT_{ij} - HKT_i - HKT_j$$

Uzaklık matrisinin güncellenmesi için ise Lance ve Williams'ın[65] formülü kullanılarak yineleme yapılır. Bu formül her bir adımda (küme birleşmeleri yapılırken) küme uzaklıklarını güncellemek için kullanılır.

63 Aggarwal (2015), 170.

64 Zaki, (2014), 368.

65 Lance, G. N. and Williams, W. T. (1967). "A general theory of classificatory sorting strategies II. Clustering systems."

C_i ve C_j sembollü iki küme C_{ii} kümesi olarak birleşeceği zaman, yeni oluşmuş olan C_{ij} kümesinin diğer bütün kümelere C_k (k$\neq i \; ve \; k \neq j$) olan uzaklığının yeniden hesaplanması gerektiği için, uzaklık matrisi güncellenmelidir. Bu noktaya kadar bütün uzaklıklar daha önceden belirlenmiş ve dolayısıyla elimizde uzaklık matrisi mevcuttur[66].

d_{ij}, d_{ik} ve d_{jk}, C_i, C_j ve C_k kümeleri arasındaki uzaklıklar olsun.

$d_{(ij)k}$ ise, $C_i \cup C_j$ (yeni küme) ve C_k kümesi ile arasındaki uzaklık olsun.

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma \left| d_{ik} - d_{jk} \right|$$

Burada $\alpha_i, \alpha_j, \beta$ ve γ katsayıları, küme boyutuna bağlı olabilirler ve küme uzaklık fonksiyonu d_{ii} ile birlikte kümeleme algoritmasını belirlerler. Örneğin $\alpha_i = \alpha_j = 1/2, \beta = 0$ ve $\gamma = -1/2$ olduğu zaman formül aşağıdaki hali alır ve algoritma bir diğer hiyerarşik tekniklerden olan tek bağlantı yöntemine karşılık gelir;

$$d_{(ij)k} = \min(d_{ik}, d_{jk}) .$$

$\alpha_i = \alpha_j = \gamma = 1/2$ ve $\beta = 0$ olduğunda ise, formül tam bağlantı yöntemine karşılık gelir;

$$d_{ijk} = max(d_{ik}, d_{jk})$$

Tek bağlantı, çok bağlantı ve ortalama bağlantı hiyerarşik kümeleme yöntemleri, uzaklık hesaplanırken, bir çift kümedeki bütün noktaları göz önünde bulundurur ve bu yüzden bunlara grafik yöntemler de denir. Diğer yöntemler ise küme temsilcisi için geometrik merkezleri kullandıkları için, geometrik yöntemler olarak adlandırılırlar.

4. Uygulama

Bu bölümde, Türkiye'de faaliyet gösteren bir süpermarket zinciri firmasına ait, İstanbul ili sınırları içerisinde bulunan 175 mağaza pazarlama stratejileri geliştirme amacı doğrultusunda segmentlere ayrılacaktır. Segmentasyon işi kümeleme analizi kullanılarak yapılacak olup kümeleme analizine girdi olacak sosyoekonomik değişkenler firmaya ait 175 mağazanın adresleri kullanılarak farklı kaynaklardan elde edilmiştir.

4.1 Veri Seti

Veri setini oluşturabilmek için, ilk etapta bu mağazaların ticari çevresi; mağazanın yakınlarında bir veya birkaç fabrika, ticaret merkezi, üniversite ve turistik merkez bulunup bulunmadığı tespit edilerek incelenmiştir. Bu inceleme Google Haritalar kullanılarak yapılmış ve sonunda ikili

The computer journal 10(3): 271-277.

66 Zaki, (2014), 370.

değişkenler (0: Hayır, 1: Evet) oluşturulmuştur. Bu değişkenlerin aldığı değerler, örnek teşkil etmesi amacıyla beş mağaza için Tablo 2'de sunulmuştur.

Google haritalar yardımı ile ikinci etapta mağazaların yakın çevresinde faaliyet gösteren rakip mağazaların sayısı da tespit edilmiştir. Örneğin incelenen bölgede firmanın büyük ölçekli bir mağazası yer alıyorsa, rakip olarak yine büyük ölçekli mağazaların yakın çevrede yer alıp almadığı, varsa kaç tane yer aldığı tespit edilmeye çalışılmıştır. Bu amaçla tek tek 175 mağazanın çevresinde yer alan rakip mağazalarının sayısı bulunmuş ve rakip mağaza sayısı değişkeni oluşturulmuştur. Yine Tablo 2'den bu değişkenlerin aldığı değerler örnek beş mağaza için incelenebilir.

Ayrıca yine mağazaların adresleri kullanılarak Türkiye İstatistik Kurumu (TÜİK) tarafından yayınlanan Adrese Dayalı Nüfus Sistemi ile mağazaların yer aldığı mahallelerdeki demografik yapı ile ilgili veriler elde edilmiştir. TÜİK'ten İstanbul ilinde bulunan bütün mahallelere ait istatistikler satın alınmış, mağazaların bulunduğu 175 mahallenin verileri ham veri kümesinden süzülmüştür.

Adrese Dayalı Nüfus Sisteminde, yaş dağılımı (0-4 yaşındaki insan sayısı, 5-9 yaşındaki insan sayısı, 10-14 yaşındaki insan sayısı…) medeni durum (bekâr insan sayısı, dul insan sayısı…) ve eğitim durumu (okuryazar olmayan insan sayısı, ilkokul mezunu olan insan sayısı, ilköğretim mezunu olan insan sayısı…) gibi mahalle sakinlerinin demografik özelliklerine ait istatistikler mevcuttur. TÜİK'ten elde edilen verilerde bulunan değişkenler, firmanın pazarlama faaliyetleri ve segmentasyon analizi sonrası elde edilecek sonuçların kolay yorumlanabilir olması açısından tekrar düzenlenmiştir. TÜİK verilerine göre, 14 gruba ayrılmış olan yaş değişkeni (0-4, 5-9, 10-14…65 ve üzeri), pazarlama literatüründe hakkında birçok çalışma yapılmış ve halen yapılmakta olan X, Y ve Z kuşakları göz önüne alınarak, beş grupta (0-4, 5-14, 15-34, 35-54, 55 ve üzeri) birleştirilmiştir. Aynı şekilde eğitim seviyesi değişkeni de TÜİK ham verilerinde on grup iken, yine sonuçların kolay yorumlanabilmesi açısından dört grupta (eğitim yok, eğitim seviyesi düşük, eğitim seviyesi orta, eğitim seviyesi yüksek) toplanmıştır. Her bir gruba ait değerler TÜİK verilerinde insan sayısı olarak hesaplanmasının aksine, oran şeklinde hesaplanmıştır. Böylece Tablo 2'de gösterildiği gibi demografiyle ilgili değişkenlere ait değerler elde edilmiştir. Mağazaların bulunduğu mahallelerde yaşayan insanların ekonomik durumları hakkında TÜİK vb. kurumlar aracılığıyla oluşturulmuş herhangi bir veri kümesi yoktur. Mahalle sakinlerinin ekonomik durumlarının bölgede bulunan konutların aylık kira bedelleri ile doğru orantılı olduğu düşünülmüş ve bu bedellerin tespiti ise bazı emlak şirketlerinin web siteleri aracılığıyla yapılmıştır. Tablo 2'de ev kirası değişkeni Türk Lirası cinsinden oluşturulmuştur.

Tablo 2: Uygulamada Kullanılacak Olan Değişkenler ve Beş Mağazaya Ait Gözlem Değerleri

Değişkenler	Mağaza 1	Mağaza 2	Mağaza 3	Mağaza 4	Mağaza 5
Ev Kirası	1200	2200	900	3000	600
Fabrika Bölgesi	0	0	0	1	0
Üniversite Bölgesi	1	0	0	0	1
Ticaret Merkezi	0	0	0	0	0
Turistik Bölge	0	0	0	1	0
Rakip Sayısı	5	1	7	1	1
Yaş 0-4	0.0854	0.0546	0.0668	0.0376	0.0928
Yaş 5-14	0.1714	0.1502	0.1268	0.0749	0.1702
Yaş 15-34	0.4498	0.2675	0.3363	0.2632	0.3647
Yaş 35-54	0.2272	0.3631	0.3069	0.3246	0.2921
Yaş 55+	0.0663	0.1646	0.1631	0.2996	0.0802
Bekâr	0.5062	0.4475	0.4156	0.6010	0.3724
Evli	0.4938	0.5525	0.5844	0.3990	0.6276
Eğitim Yok	0.1826	0.0932	0.0908	0.0579	0.1356
Eğitim Düşük	0.3832	0.2049	0.4821	0.2252	0.5622
Eğitim Orta	0.3224	0.2246	0.2629	0.2954	0.1757
Eğitim Yüksek	0.1118	0.5021	0.1642	0.4481	0.1283

4.2 Kümeleme Analizi ile Mağaza Segmentasyonu

Kümeleme Analizinde kullanılacak olan sosyoekonomik değişkenlerin belirlenip her bir mağaza için aldığı değerler de tespit edildikten sonra, uzaklık ölçüsünün belirlenmesi aşamasına geçilir. Değişkenlerin farklı ölçeklerle ölçülmüş olmasından ötürü Gower'in ortaya çıkarmış olduğu uzaklık ölçüsü kullanılarak hesaplanması daha uygun olan uzaklık matrisi *R Programlama Dilindeki* bazı işlevler sayesinde gerçekleştirilmiştir.

R'daki *cluster* paketinde yer alan *daisy* işlevi yardımıyla ihtiyaç duyulan uzaklık matrisi kolayca elde edilmiştir. Bu işlev sayesinde veri kümeleri için Öklidyen ve Manhattan uzaklık ölçüleri de hesaplanabilmektedir. İlk etapta 17 değişkenin 175 mağazanın her biri için tespit edilen değerlerinin oluşturduğu veri matrisi R programına aktarılmış, daha sonra da değişkenlerin tipleri programa tanıtılmıştır.

Değişkenler arasındaki korelasyonun tespiti için, literatürde birçok uygulamada kullanılmış olan R programlama dilinin *polycor* paketindeki *hetcor* işlevi sayesinde farklı değişken tiplerini barındıran veri kümeleri için korelasyon değerleri hesaplanmaktadır. *Hetcor* işlevi sayısal değişkenler arasında Pearson çarpım-moment korelasyonu, sayısal ile kategorik değişkenler arasında Polyserial korelasyonu, kategorik değişkenler arasında ise, Polychoric korelasyonu hesaplamaktadır. Bu hesaplamalar sonucunda bekar oranı değişkeni ile evli oranı değişkeninin beklendiği gibi – 1 korelasyona sahip olduğu, yaş 0-4 ile yaş 5-14 değişkenlerinin ayrıca düşük

eğitimli oranı ile yüksek eğitimli oranı değişkenlerinin ise 0,90 korelasyon değerinden daha yüksek değerlere sahip olduğu tespit edilmiştir. Bu sebeple yaş 5-14 değişkeni, bekâr oranı değişkeni ve düşük eğitimli oranı değişkeni, uzaklık matrisi hesaplanmadan önce veri kümesinden çıkartılmıştır. Daha sonra segmentlerin özelliklerinin yorumlanması aşamasında bu değişkenler veri kümesinde tekrar göz önünde bulundurulmuştur.

Hesaplanan 175×175 boyutundaki uzaklık matrisi Hiyerarşik Kümeleme Tekniklerinden biri olan Ward'ın Kümeleme Algoritması için girdi olarak kullanılmıştır. Ward'ın algoritması için yine R'daki *stats* paketinde yer alan *hclust* işlevi kullanılmıştır. Bu işlev sayesinde uzaklık matrislerine Hiyerarşik Kümeleme Yöntemleri uygulanabilmekte ve sonuçlar dendrogramda görselleştirilebilmektedir. Ward'ın algoritmasının oluşturduğu dendrogram Ek 1'de sunulmuştur. Yine R Programlama Dilindeki *cluster* paketinde yer alan *agnes* işlevi sayesinde Birleştirici Katsayı %97 olarak hesaplanmıştır. Bu yüksek değer bize kullanılmış olan Ward'ın algoritmasının veri kümemizde açık küme yapıları bulduğunu ispatlamaktadır.

Diğer taraftan programdaki *stats* paketinde yer alan *cophenetic* işlevi ile Kophenetic Korelasyon Katsayısı % 55 (Spearman Korelasyon ile) olarak tespit edilmiştir. Bu değer hiyerarşik algoritmanın ortaya çıkardığı sonucun orijinal veri kümesini orta derecede yansıtabildiğini göstermektedir. Fakat dendrogramın ortaya çıkardığı kümelerin özellikleri incelendiğinde ve çalışmamızın amacı kümeleme algoritmalarının performanslarını karşılaştırmanın aksine pazarlama alanında bir uygulama yapmak olduğu için bu sonuç iyi bir değer olarak değerlendirilebilir.

Gerçekten de dendrogramdan elde edilebilecek farklı kümeler incelendiğinde 175 mağazanın altı segmente ayrılması her bir grupta farklı değişkenlerin baskın olması ve dolayısıyla her bir grubun farklı özelliklerinin bulunması sebebiyle anlamlı görülmüştür. Ek 1'de yer alan Dendrogramda ağacın kökünde firmanın sahip olduğu 175 mağaza görülmektedir.

Tablo 3 Segmentlerin Özellikleri

Segment	Sembol	Mağaza Sayısı	Özellikler (BASKIN DEĞİŞKENLER)
1		40	Düşük düzeydeki ev kiraları, mağazaların hepsi üniversite bölgesinde, rekabet yüksek düzeyde, 5-14 yaş ve 15-34 yaş oranı yüksek, eğitim seviyesi düşük
2		16	Ev kiraları düşük, mağazalar fabrika ve ticaret bölgesinde, 0-4 yaş, 5-14 ve 15-34 yaş oranı çok yüksek, yüksek evli oranı, eğitim seviyesi çok düşük
3		38	Yüksek ev kiraları, rekabet orta düzeyde, çok yüksek 55+ yaş oranı, yüksek bekâr oranı, eğitim seviyesi çok yüksek
4		45	Orta ev kirası, yüksek evli oranı, rekabet yüksek, eğitim seviyesi düşük
5		18	Yüksek ev kirası, düşük rekabet düzeyi, mağazalar ticaret bölgesinde
6		18	Ev kiraları yüksek düzeyde, mağazalar turistik bölgelerde, rekabet düzeyi düşük, 55+ yaş oranı yüksek, yüksek bekâr oranı

Dendrogramdan alınan sonuçla mağazalar ait oldukları segmentlerde toplanmıştır. Segmentlerin özelliklerinin tespit edilmesi amacıyla her bir sosyoekonomik değişkenin segmentler için ortalama değerleri hesaplanmış ve Tablo 4'de sunulmuştur. Yukarıdaki Tablo 3'de ise segmentlerdeki baskın değişkenler tespit edilmiş ve segmentlerin özellikleri incelenmiştir.

Tablo 4: Değişkenlerin Segmentler Bazındaki Ortalama Değerleri

Değişkenler	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
Ev Kirası (TL)	*1096*	*1090*	*1833*	*1159*	*1775*	*1791*
Fabrika bölgesindeki mağaza sayısı	*0*	*15*	*0*	*2*	*0*	*1*
Üniversite bölgesindeki mağaza sayısı	*40*	*3*	*4*	*1*	*7*	*3*
Ticaret Merkezine yakın mağaza sayısı	*0*	*12*	*0*	*1*	*18*	*0*
Turistik Bölgedeki mağaza sayısı	*3*	*0*	*0*	*1*	*0*	*18*
Ortalama Rakip	*2,7*	*2,3*	*2,6*	*2,5*	*1,3*	*1,8*
Yaş 0-4	*0.07*	*0.10*	*0.04*	*0.07*	*0.05*	*0.05*
Yaş 5-14	*0.13*	*0.17*	*0.09*	*0.14*	*0.10*	*0.10*
Yaş 15-34	*0.34*	*0.37*	*0.28*	*0.33*	*0.33*	*0.33*
Yaş 35-54	*0.30*	*0.28*	*0.32*	*0.32*	*0.32*	*0.31*
Yaş 55+	*0.16*	*0.08*	*0.27*	*0.14*	*0.20*	*0.23*
Bekâr	*0.44*	*0.38*	*0.53*	*0.42*	*0.53*	*0.59*
Evli	*0.56*	*0.62*	*0.47*	*0.58*	*0.47*	*0.41*
Eğitim Yok	*0.11*	*0.15*	*0.07*	*0.11*	*0.08*	*0.10*
Eğitim Düşük	*0.44*	*0.56*	*0.26*	*0.42*	*0.33*	*0.37*
Eğitim Orta	*0.25*	*0.18*	*0.28*	*0.25*	*0.26*	*0.25*
Eğitim Yüksek	*0.20*	*0.12*	*0.41*	*0.23*	*0.34*	*0.29*

5. Sonuç ve Tartışmalar

Zincir firmaların tüm mağazalarına birden mağaza ve müşteri ihtiyaçlarını gözetmeksizin kitle pazarlama stratejileri geliştirmesi günümüz şartları açısından uygun değildir. Günümüze uygun olan strateji ise küresel rekabet ortamında avantaj sağlamaya ve firmanın pazarda tutunabilmesine destek olacak olan hedef pazarlama stratejisidir

Zincir firmaların mağaza sayısı arttıkça başlangıçta sunulan hizmet kalitesinin sağlanması zorlaşmaktadır. Hizmet kalitesi sorunları, müşteri memnuniyetinin azalmasına ve haliyle satış hacminin küçülmesine neden olmaktadır. Mağaza segmentasyonu ile farklı mağaza segmentlerine yönelik yaklaşım ve uygulamalar, geliştirilmesi, müşteri memnuniyetini arttıracak ve hatta satış hacmini standart hizmet düzeyinin üzerine yükseltecek fırsatlar sunar. Aynı şekilde, farklı müşteri gruplarının beklentileri ve tüketim davranışları arasındaki fark nedeniyle her bir mağazada doğru ürünleri sunarak depolama hatalarını da en aza indirmek mümkün olacaktır.

Başarılı mağaza segmentasyonu yapabilen zincir firmalar, imkanlarını en doğru şekilde kullanma ve maliyetlerini düşürme konusunda rekabet avantajı elde ederler. Mağaza segmentasyonu aynı zamanda başarılı CRM uygulamalarının geliştirilmesine de çok büyük katkı sağlayacaktır.

Başarılı mağaza segmentasyonu; müşteri alt gruplarının farkında olmak, bu alt grupları tanımak, ihtiyaç ve beklentilerini bilmek ve doğru operasyon ile bu ihtiyaç ve beklentileri en doğru şekilde karşılayarak daha yüksek müşteri memnuniyeti ve satış hacmini daha düşük maliyetlerle elde etmek için son derece etkin bir yöntemdir.

Çalışmamız sonucunda ortaya çıkan ilk segment incelendiğinde, bu segmentteki tüm mağazaların üniversite bölgesinde olduğu (sembol olarak kep seçilmiştir) ve mağazaların bulunduğu mahallelerde yaşayan insanlar arasında gençlerin oranının yüksek olduğu, ikinci segmentteki mağazaların çoğunun ya ticari bölge ya da fabrika bölgesinde olduğu (sembol olarak fabrika seçilmiştir), bu mağazaların bulunduğu mahallelerde evli olanların yüksek oranda olduğu görülmektedir. Eğitim seviyesinin ve emekli oranının yüksek olduğu üçüncü segmentte (sembol olarak yaşlı insan seçilmiştir) ayrıca insanların ekonomik durumlarının da görece daha iyi olduğu söylenebilir. Beşinci segmentte insanların ekonomik durumu iyi ve mağazaların çevresinde ticaret merkezleri (sembol olarak dolar işareti seçilmiştir) mevcuttur. Rekabetin yoğun olmadığı altıncı segmentteki mağazaların çoğu turistik bölgelerde (sembol olarak turizm ofisi seçilmiştir) yer almakta, bu bölgelerde demografisi bekâr ve emekli olanların oranı yüksek görülmektedir. Dördüncü segment incelendiğinde yüksek evli oranı ve rekabet ortamı ile orta seviyedeki kira seviyesine sahip olması dikkat çekmektedir.

Firmalar özellikle pazarlama ile ilgili karar verme süreçlerinde, segmentasyon analizi sonucunda ortaya çıkan segmentlerin farklı özelliklerini göz önünde bulundurmalıdır. Eğitim seviyesinin yüksek olduğu yerlerde veya üniversite bölgelerinde diğer segmentlerden farklı bir strateji olarak internet aracılığıyla pazarlama faaliyetleri yürütülebilir. Mağazalara özel web siteleri oluşturulabilir, mobil indirim kuponları hazırlanabilir, kısa mesaj yoluyla özel indirimler uygulanabilir, çevrimiçi sipariş ve ödeme yöntemi ile evlere servis hizmeti sunulabilir. Ekonomik gelirin yüksek olduğu bölgelerde fiyat seviyesi ve karlılığı yüksek ürünlerin, üniversite bölgesindeki mağazalarda ise ucuz ve pratik ürünlerin pazarlanması yoluna gidilebilir. Düşük ekonomik seviyedeki bölgelerde fiyat seviyesi ve karlılığı düşük ürünlerin daha çok sunulması yoluna gidilebilir. Evli oranı ve çocuk oranı yüksek bölgelerde ailelere ve çocuklara yönelik kampanyalar geliştirilebilir. Ancak, mağaza segmentasyonu ile hedeflere tatmin edici düzeyde ulaşılabilmesi için ilgili operasyonun en doğru şekilde planlanması ve uygulanması da kritik öneme sahiptir.

Kaynakça

AGGARWAL, C. C. (2015). Data mining: The textbook. Switzerland, Springer.

BADEA, L. M. (2014). "Predicting consumer behavior with artificial neural networks." Procedia Economics and Finance 15: 238-246.

BENNETT, P. D. (1995). Dictionary of Marketing Terms, NTC Business Books.

BERMINGHAM, P., Hernandez, T. ve Clarke, I. (2013). Network Planning and Retail Store Segmentation: A Spatial Clustering Approach. International Journal of Applied Geospatial Research, 4(1): 67-79.

BIJAK, K. ve Thomas, L. C. (2012). "Does segmentation always improve model performance in credit scoring?" Expert Systems with Applications 39(3): 2433-2442.

BORNAC, G. (2015). "The Power of Store Clustering." http://www.manh.com/resources/articles/2015/08/27/power-store-clustering (30.05.2016).

CEYLAN, H. H. (2013). "Perakende Sektöründe Konjoint ve Kümeleme Analizi ile Fayda Temelli Pazar Bölümlendirme." Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 20(1): 141-154.

CLARKE, I., Mackaness, W. ve Ball, B. (2003). "Modelling Intuition in Retail Site Assessment (MIRSA): making sense of retail location using retailers' intuitive judgements as a support for decision-making." The International Review of Retail, Distribution and Consumer Research 13(2): 175-193.

COMPANY, W. P. (2013). "A Simple Approach to Retail Clustering." http://www.wilsonperumal.com/media/publications/PDFs/Vantage_Point_2013_Issue3.pdf (30.05.2016).

DE OÑA, R. ve De Oña, J. (2015). "Analysis of transit quality of service through segmentation and classification tree techniques." Transportmetrica A: Transport Science 11(5): 365-387.

DÍAZ-PÉREZ, F. M. ve Bethencourt-Cejas, M. (2016). "CHAID algorithm as an appropriate analytical method for tourism market segmentation." Journal of Destination Marketing & Management.

DJOKIC, N., Salai, S., Kovac-Znidersic, R., Djokic, I. ve Tomic, G. (2013). "The use of conjoint and cluster analysis for preference-based market segmentation." Engineering Economics 24(4): 343-355.

DONOFRIO, T. J. (2009). "Advanced Planning and Optimization Part 3: Store Clustering." Retail Systems and Services http://risnews.edgl.com/retail-news/Advanced-Planning-and-Optimization-Part-3—Store-Clustering38904 (26.05.2016).

DOYLE, C. (2011). A dictionary of marketing, Oxford University Press.

EKİNCİ, Y., Ulengin, F. ve Uray, N. (2014). "Using customer lifetime value to plan optimal promotions." The Service Industries Journal 34(2): 103-122.

EVERITT, B. (1974). Cluster analysis 122, Heinemann, London.

EVERITT, B. ve Hothorn, T. (2011). Cluster analysis. An Introduction to Applied Multivariate Analysis with R, Springer: 163-200.

GOWER, J. C. (1971). "A general coefficient of similarity and some of its properties." Biometrics: 857-871.

GREEN, P. E. ve Rao, V. R. (1971). "Conjoint measurement for quantifying judgmental data." Journal of Marketing research: 355-363.

GUO, Y., Denizci Guillet, B., Kucukusta, D. ve Law, R. (2015). "Segmenting Spa Customers Based on Rate Fences Using Conjoint and Cluster Analyses." Asia Pacific Journal of Tourism Research: 1-19.

HAWKES, G. F. ve McLaughlin, E. W. (1994). STARS: Segment Targeting at Retail Stores, Department of Agricultural, Resource, and Managerial Economics, Cornell University.

HENNIG, C. ve Liao, T. F. (2013). "How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification." Journal of the Royal Statistical Society: Series C (Applied Statistics) 62(3): 309-369.

JACCARD, P. (1908). Nouvelles recherches sur la distribution florale.

JAIN, A. K. ve Dubes, R. C. (1988). Algorithms for clustering data, Prentice-Hall, Inc.

KARGARI, M. ve Sepehri, M. M. (2012). "Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs." Expert Systems with Applications 39(5): 4740-4748.

KAUFMAN, L. ve Roussew, P. (1990). Finding Groups in Data-An Introduction to Cluster Analysis. A Wiley-Science Publication John Wiley & Sons, Inc.

KOEHN, N. F. (2001). "Howard Schultz and Starbucks Coffee Company." Harvard Business School Cases.

LANCE, G. N. ve Williams, W. T. (1967). "A general theory of classificatory sorting strategies II. Clustering systems." The computer journal 10(3): 271-277.

LILIEN, G. L. ve Kotler, P. (1983). Marketing decision making: A model-building approach, Harper & Row New York, NY.

LIPPMAN, B. W. (2003). "Retail revenue management—Competitive strategy for grocery retailers." Journal of revenue and pricing management 2(3): 229-233.

MEHROTRA, A. ve Agarwal, R. (2009). "Classifying customers on the basis of their attitudes towards telemarketing." Journal of Targeting, Measurement and analysis for Marketing 17(3): 171-193.

MENDES, A. B. ve Cardoso, M. G. M. S. (2006). "Clustering supermarkets: the role of experts." Journal of Retailing and Consumer Services 13(4): 231-247.

MYERS, J. H. (1996). Segmentation and positioning for strategic marketing decisions.

NAKAHARA, T. ve Yada, K. (2012). "Analyzing consumers' shopping behavior using RFID data and pattern mining." Advances in Data Analysis and Classification 6(4): 355-365.

NAKİP, M. (2006). Pazarlama araştırmaları teknikler ve (SPSS destekli) uygulamalar. Ankara, Seçkin Yayıncılık.

PAVOINE, S., Vallet, J., Dufour, A. B., Gachet, S. ve Daniel, H. (2009). "On the challenge of treating various types of variables: application for improving the measurement of functional diversity." Oikos 118(3): 391-402.

SMITH, W. R. (1956). "Product Differentiation And Market Segmentation As Alternative Marketing Strategies." Journal of Marketing 21(1): 3-8.

TİMOR, M. ve Şimşek, U. (2008). "Veri Madenciliğinde Sepet Analizi ile Tüketici Davranışı Modellemesi." Yönetim, 19 (59): 3-10.

VOHRA, G. (2011). "Store Clustering." http://analyticstraining.com/2011/store-clustering/ (26.05.2016).

WEINSTEIN, A. (2004). Handbook of market segmentation: Strategic targeting for business and technology firms, Psychology Press.

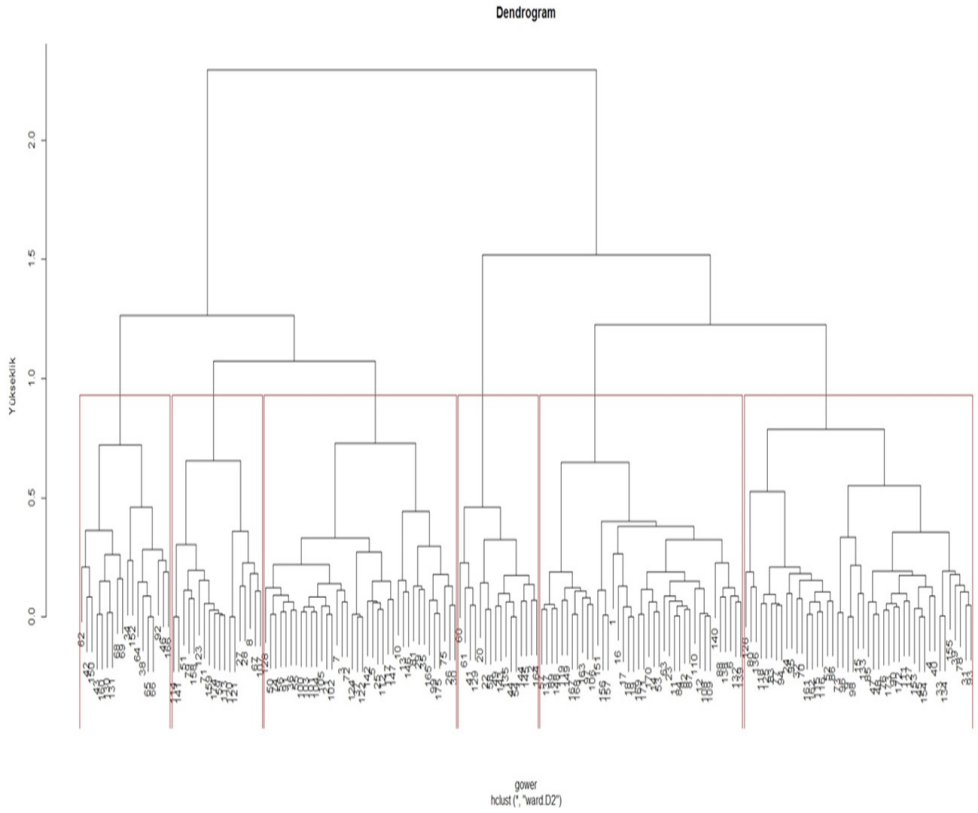
XU, R. ve Wunsch, D. C. (2009). "Clustering." from http://site.ebrary.com/id/10257659.

YAMAN, T. T. ve Çakır, Ö. "ÜNİVERSİTE TERCİHLERİNİN SEÇİME DAYALI KONJOİNT ANALİZİ İLE BELİRLENMESİ." Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi 1(1): 65-84.

ZAKI, M. J. ve Meira Jr, W. (2014). Data mining and analysis: fundamental concepts and algorithms, Cambridge University Press.

ZHIYU, Z. ve Congdong, L. (2009). Research on Application to Customer Classification in Management Decision-Making Based on Multivariate Statistics. Information Technology and Applications, 2009. IFITA'09. International Forum on, IEEE.

EK 1: DENDROGRAM



Extended Abstract

The value of analyzing databases is increasing day by day thus retailers and other companies started to invest more and more to coordinate their data analytic strategies for serving their customers better. In this study, we propose a Business Analytics approach that utilizes a Data Mining technique, Cluster Analysis for segmenting stores of a retail chain. The new methodology we offer sheds light on how to cluster the stores of a retailer effectively. Segmentation will be carried out using cluster analysis and socioeconomic variables will be used for clustering task. The data set of variables used in this study have been obtained from different sources by using mainly the addresses of 175 stores.

Although there have been many studies on customer/market segmentation over the last five decades, scant attention was devoted to store clustering, which is still a new approach in retailing. Store clustering aims to divide a network of stores into meaningful groups. For example, a retailer with a network of 350 stores may generate six different store types, with each comprised of a 'similar' set of stores. Since a retailer can have millions of customers and the number of stores is usually limited to from 100 to 1000, store clustering seems easier than customer/market segmentation.

Store clustering brings many benefits to retail chain companies. Boston Retail Partners' (BRP) survey (Boston Retail Partners, 2012, "1st Annual Merchandise Planning & Allocation Benchmark Survey") is a strong proof of the value of store clustering. BRP surveyed more than 500 top North American Retailers and found that half of the companies surveyed perform store clustering, which is generally used to support Assortment Planning and for Allocation to provide a localized product mix in an efficient and effective manner.

Store clustering studies done before can be summarized as follows: Koehn (2001) offers Starbucks as a success story, since the company has increased its brand awareness using store clustering. Clarke, Mackaness, and Ball (2003) used store clustering for the purpose of retail site assessment. They used a special package program called MIRSA, which is prepared for retailers for decision support. The authors used cognitive maps, based on the answers of experts from the largest retail chains in United Kingdom, to identify the main variables used in location decisions. Bermingham, Hernandez, and Clarke, (2013) also used MIRSA for store clustering to be able to support network planning and location decision making. In the study, a Canadian retailer's store operation data, sales data and trade area characteristics were used for the task of clustering. Mendes and Cardoso (2006) used store clustering to evaluate the performance of stores and to find new store site locations in Portugal. They used shop surveys of 6000 customers and national census data as clustering variables. Since a few stores (25 stores) with a huge number of variables (250 variables) are used for clustering, they ask experts their opinion on the interpretation of the results. Kargari and Sepehri (2012) clustered 815 stores of an automotive spare-parts distributor and after-sales services company for the purpose of reducing transportation costs. The three-year sales data with 40,750 records consist of variables such as store locations, type of order and order size.

To be able to create the data set used in this study we fulfill the following steps: the commercial environment of 175 stores such as, whether there are one or more factories, commercial centers, universities and tourist centers in the vicinity of the store. This review was conducted using Google Maps, and binary variables (0: No, 1: Yes) were generated. With the help of Google maps, the number of competing stores operating in the immediate vicinity of the stores was determined in the second stage. In addition, data on demographic structure in the neighborhood where the store with the Address Based Population System published by the Turkey Statistical Institute has been obtained. The statistics of all neighborhoods in Istanbul were purchased from TURKSTAT and the data of 175 neighborhoods where the stores were located were filtered from the raw data set. The economic conditions of the residents were thought to be directly proportional to the monthly rental prices of the residences in the region, and the prices were determined through the websites of some real estate companies. A sample of variables used for five stores is given in the table above.

The required distance matrix for cluster analysis can be easily obtained by the daisy function included in the cluster package in R. The calculated 175 × 175 distance matrix was used as an input for Ward's Clustering Algorithm, one of the Hierarchical Clustering Techniques. The hclust function in the stats package in R is used for Ward's algorithm. When the different clusters obtained from the dendrogram were examined, the partition of 175 stores into six segments was found to be significant because different variables were dominant in each group and therefore each group had different characteristics.

It is not appropriate for the company to develop mass marketing strategies for all 175 stores regardless of store and customer needs. The strategy that is suitable for today is the target marketing strategy which will provide advantage in global competitive environment and support the firm to keep in the market. As can be seen from Table 2, the six groups of stores have different characteristics.

The firm should take into account the different characteristics of the segments resulting from segmentation analysis, especially in the decision-making processes related to marketing. For example, if internet and smartphone usage is more common in places where education level is high or in university regions, marketing activities can be carried out via internet as a different strategy from other segments. Stores can be created by creating special web sites, campaigns can be announced here, mobile discount coupons can be prepared, special discounts can be applied to member customers via text message. Unlike the other segments, it is possible to market expensive products in the segments with high economic income and to market cheap and practical products in the stores in the university region. Taking into consideration the high married rate and high child age values in the segments, campaigns for families can be developed unlike other segments.

The most important outcome of this study is to show the benefits and ease of store clustering using socio-economic variables with appropriate data mining tools. The strengths of this study also make it different from the others, such as the fact that the company has 175 different stores located in different regions of Istanbul, a city populated with 20 million people.