

T.C.
MANİSA CELAL BAYAR ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI

TÜRKİYE’ DE İLLER BAZINDA ULAŞIM FAALİYETLERİNİN
GELİŞİM DURUMUNUN KÜMELEME ANALİZİ İLE
BELİRLENMESİ

Erkan ALTIN

Danışman
Dr. Öğr. Üyesi Aynur İNCEKIRIK

MANİSA–2021

T.C.
MANİSA CELAL BAYAR ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ PROGRAMI

TÜRKİYE’ DE İLLER BAZINDA ULAŞIM FAALİYETLERİNİN
GELİŞİM DURUMUNUN KÜMELEME ANALİZİ İLE
BELİRLENMESİ

Erkan ALTIN

Danışman
Dr. Öğr. Üyesi Aynur İNCEKIRIK

MANİSA–2021

YEMİN METNİ

Yüksek lisans tezi olarak sunduğum “Türkiye’de İller Bazında Ulaşım Faaliyetlerinin Gelişim Durumunun Kümeleme Analizi ile Belirlenmesi” adlı çalışmanın, tarafımdan bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin bibliyografyada gösterilen eserlerden oluştuğunu, bunlara atıf yapılarak yararlanmış olduğumu belirtir ve bunu onurumla doğrularım.

.../.../2021

Adı Soyadı

Erkan ALTIN

İmza

ÖZET

TÜRKİYE’ DE İLLER BAZINDA ULAŞIM FAALİYETLERİNİN GELİŞİM DURUMUNUN KÜMELEME ANALİZİ İLE BELİRLENMESİ

Ulaştırma kavramı geçmişten günümüze insanlar arasında bağlantıyı ve yer değiştirmeyi sağlayarak eğitim, sağlık, ekonomi, savunma gibi bir uygarlıktaki olmazsa olmaz nitelikteki hayati konuları doğrudan etkileyen en önemli kavramlardan birisidir. Ulaştırma sistemleri gelişmiş devletler, tarih boyunca hep daha avantajlı konumda yer almıştır. En müreffeh toplumların ulaşım imkânları ve altyapılarının güçlü olması tesadüf değildir. İnsanlık tarihinde karayolu ulaşımı ile başlayan ulaşımındaki bu gelişim süreci günümüze kadar denizyolu, demiryolu, boru hatları ve havayolu hatlarıyla giderek farklı dallarda çeşitlenmiştir.

2004-2018 yıllarında ulaşım faaliyetlerini oluşturan 40 adet göstergenin göz önüne alınarak yapıldığı bu çalışmada Türkiye’nin iller bazında ulaşım konusunda içinde bulunduğu durumu analiz etmek ve izleyeceği ulaşım politikasında bir ışık tutabilmesi için kümeleme analizi yöntemleri kullanılmıştır. Burada ulaşım konusunda illerin birbirlerine göre benzerliği ve illerin gelişim durumları hakkında bir çıkarım yapılabilmesi amaçlanmıştır. Kümeleme analizi bir grup gözlemi içlerinde benzer (homojen) ve aralarında benzer olmayan (heterojen) belirlenen sayıda kümelere ayrılıp çeşitli yöntemlerle incelenmesini amaçlayan çok değişkenli istatistik tekniğidir. Veri setinde min-maks standartlaştırması ve öklid uzaklık ölçüsü kullanılarak altı adet hiyerarşik kümeleme yöntemi (tek bağlantı, tam bağlantı, ortalama bağlantı, ward, medyan, ağırlık merkezi) ve iki adet hiyerarşik olmayan kümeleme yöntemi (k-ortalamlar, k-medoid) belirlendikten sonra R 3.5.1 ve SPSS 16.0 programları kullanılmıştır. Küme sayısını belirlemek amacıyla alt limiti 5 ve üst limiti 12 küme olmak üzere 14 adet küme geçerlilik indeksi kullanılmış ve en uygun küme sayısının 5 olmasına karar verilmiştir.

Uygulanan analizler sonucunda tüm kümeleme analizi yöntemlerinde İstanbul ili tek bir aykırı gözlem şeklinde büyük bir farkla öne çıkmıştır. Hiyerarşik yöntemler arasında kümeleme analizi yöntemlerinin başarısını ölçmek amacıyla kullanılan kofenetik korelasyon katsayısının en iyi değeri gösteren yöntemi tek bağlantı yöntemi olmuştur. Bu yöntemin sonucuna göre de İstanbul, İzmir, Ankara

ve Antalya diğ er illerden ayrılmıřtır. Hiyerarřik olmayan y ntemlerde de genel itibariyle Marmara, Ege b lgesi ve Akdeniz b lgesi illeri birbirine yakın karakterde il grupları olarak  ne  ıkmıř, Dođu Karadeniz, bir ok Dođu Anadolu ve G neydođu Anadolu B lgesi illeri de birbirine yakın benzerlikte  ıkmıřlardır. Karadeniz b lgesinde Samsun ve Trabzon illeri bir nebze farklı g r nmektedir. Analiz sonu ları, en b y k n fusa sahip b y křehir illerinin kalan illere g re ulařım bakımından farklı durumda olduklarına iřaret etmektedir.

Anahtar Kelimeler: K meleme Analizi, Ulařtırma, R Programlama, Kofenetik Korelasyon Katsayısı, k-Ortalamlar K meleme Y ntemi, K me Ge erlilik İndeksleri.

ABSTRACT

DETERMINATION OF DEVELOPMENT STATUS OF TRANSPORTATION FACILITIES IN TURKEY ON THE BASIS OF CITIES BY USING CLUSTERING ANALYSIS

The concept of transportation is one of the most important concepts that directly affects vital issues that are indispensable in a civilization such as education, health, economy, defense by providing connection and relocation between people from past to present. For this reason, developed states have always been in a more advantageous position throughout history. It is no coincidence that the transport facilities and infrastructure of the most prosperous societies are strong. This development process in transportation, which started with road transportation in human history, has increasingly diversified in different branches with seaways, railways, pipelines and airlines until present times.

Forming the transportation activities between 2004-2018 year, 40 transportation indicator of Turkey in this study was made considering 81 provinces to analyze the situation in terms of transportation basis and will follow clustering analysis methods to help the transport policy are used. Here, it is aimed to draw an inference about the similarity of provinces relative to each other in terms of transport and the development status of provinces. Clustering analysis is a multivariate statistical technique that aims to divide a group of observations into a set number of clusters that are similar (homogeneous) and not similar (heterogeneous) between them and study them by various methods. Min-max standardization and Euclidean distance measure were used with R 3.5.1 and SPSS 16.0 programs in the data set. Among the cluster analysis methods, six hierarchical clustering methods (single linkage, complete linkage, average linkage, ward method, median linkage, centroid method) and two non-hierarchical clustering methods (k-means, k-medoid) were used. In addition, 14 cluster validity indexes were used, with the lower limit 5 and the upper limit 12 sets, to determine the number of clusters and the most appropriate cluster number was decided to be 5.

As a result of the applied analyzes, Istanbul came to the fore with a big difference in the form of a single outlier in all cluster analysis methods. Among the

hierarchical methods, the method that shows the best value of the cophenetic correlation coefficient used to measure the success of cluster analysis methods was the single connection method. According to the result of this method, Istanbul, Izmir, Ankara and Antalya were separated from other provinces. In non-hierarchical methods, Marmara, Aegean Region and Mediterranean Region came to the fore as provincial groups, and Eastern Black Sea, many Eastern Anatolia and Southeast Anatolian provinces emerged closely. Samsun and Trabzon provinces looks some what different in the Black Sea Region. The results of the analysis indicate that metropolitan provinces (provinces with a large population) are different in terms of transportation compared to the remaining provinces.

Keywords: Cluster Analysis, Transportation, R Programming, Cophenetic Correlation Coefficient, k-Means Clustering Method, Cluster Validation Indices.

TEŞEKKÜR

Çalışmamın bütün aşamalarında bana destek olan, bilgi ve deneyimleri ile yol göstermiş; lisansüstü öğrenim hayatımda etrafı kasıp kavuran oldukça zorlu 2020 yılı başta olmak üzere hemen tüm zorlu aşamalarında maddi manevi her yönden yardımcı olan, tecrübeleri ile beni aydınlatan ve desteğini hiç eksik etmeyen, kendisini tanımaktan büyük onur duyduğum sevgili hocam Sayın Dr. Öğretim Üyesi Aynur İNCEKIRIK'a, ayrıca eğitim hayatımda uzaktan veya yakından farketmeksizin katkısı bulunan tüm hocalarıma, öğrenim hayatım boyunca beni maddi ve manevi olarak destekleyen ve hep yanımda olan aileme ve yakınlarıma yürekten teşekkür ederim.

Erkan ALTIN
Manisa, 2021

İÇİNDEKİLER

ÖZET.....	v
ABSTRACT	vi
TEŞEKKÜR.....	vii
İÇİNDEKİLER	viii
SİMGELER VE KISALTMALAR DİZİNİ	xi
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ	xiv
GİRİŞ	1

BİRİNCİ BÖLÜM

ULAŞTIRMA KAVRAMI

1.1. ULAŞTIRMA KAVRAMI	4
1.1.1. Ulaştırmanın Tarihi Önemi	5
1.1.2. Ulaştırmanın Sosyal Önemi.....	6
1.1.3. Ulaştırmanın Ekonomik Önemi.....	6
1.1.4. Ulaştırmanın Politik Önemi.....	6
1.1.5. Ulaştırmanın Çevresel Önemi	7
1.2. ULAŞTIRMA TÜRLERİ	7
1.2.1. Karayolu Ulaşımı	9
1.2.2. Demiryolu Ulaşımı	11
1.2.3. Deniz ve Suyolu Ulaşımı.....	12
1.2.4. Havayolu Ulaşımı.....	14
1.2.5. Boru Hatları Ulaşımı	18
1.3. TÜRKİYE’DE ULAŞTIRMA HAKKINDA BİLGİLER	19

İKİNCİ BÖLÜM

KÜMELEME ANALİZİ

2.1. KÜMELEME ANALİZİ.....	24
2.2. VERİLERİN STANDARTLAŞTIRILMASI	31
2.3. UZAKLIK ÖLÇÜLERİ	32
2.3.1. Öklid Uzaklığı	33
2.3.2. Karesi Alınmış Öklid Uzaklığı.....	35
2.3.3. Manhattan (City-Block) Uzaklığı.....	35
2.3.4. Mahalanobis Uzaklığı.....	36
2.3.5. Minkowski Uzaklığı.....	36

2.3.6. Canberra Uzaklığı.....	36
2.3.7. Chebychev Uzaklığı	37
2.3.8. Hotelling T^2 Uzaklığı	37
2.4. HİYERARŞİK KÜMELEME ANALİZİ YÖNTEMLERİ.....	37
2.4.1. Tek Bağlantı Yöntemi	40
2.4.2. Tam Bağlantı Yöntemi	42
2.4.3. Ortalama Bağlantı Yöntemi.....	43
2.4.4. Ward Yöntemi	44
2.4.5. Ağırlık Merkezi (Centroid) Yöntemi	45
2.4.6. Medyan Yöntemi	46
2.5. HİYERARŞİK OLMAYAN KÜMELEME ANALİZİ YÖNTEMLERİ	46
2.5.1. k-Ortalamalar Yöntemi.....	48
2.5.2. k-Medoid Yöntemi	49
2.6. KÜME GEÇERLİLİK İNDEKSLERİ VE KÜME SAYISININ BELİRLENMESİ.....	50
2.6.1. Küme Sayısının Belirlenmesi	51
2.6.2. Küme Geçerlilik İndeksleri	52
2.6.2.1. Siluet (Silhouette) İndeksi.....	53
2.6.2.2. Dunn İndeksi	54
2.6.2.3. Calinski ve Harabasz İndeksi	55
2.6.2.4. Kübik Kümeleme Kriteri (CCC).....	55
2.6.2.5. C İndeksi	56
2.6.2.6. DB İndeksi	57
2.6.2.7. Duda İndeksi	57
2.6.2.8. Pseudo T^2 İndeksi	58
2.6.2.9. Beale İndeksi.....	59
2.6.2.10. Ratkowsky İndeksi.....	60
2.6.2.11. Point-Biserial İndeksi.....	60
2.6.2.12. Frey İndeksi	61
2.6.2.13. McClain İndeksi.....	61
2.6.2.14. SD İndeksi.....	62
2.7. KÜMELEME YAPISININ ARAŞTIRILMASINDA GRAFİKSEL YAKLAŞIMLAR.....	63
2.7.1. Saçılım Grafiği	63

2.7.2. Ana Bileşen Skorları Grafiği.....	63
2.7.3. Diğer Gösterimler.....	64
2.8. KÜMELEME ALGORİTMASININ SEÇİMİ.....	65
2.9. KÜMELER İÇİN İSTATİSTİKSEL ÖNEM TESTLERİ	66

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. ÇALIŞMANIN AMACI	68
3.2. LİTERATÜR TARAMASI.....	70
3.3. KULLANILAN DEĞİŞKENLER VE YÖNTEMLER	81
3.4. KÜMELENME YAPISININ GÖRÜLMESİNDE ANA BİLEŞEN SKOR GRAFİKLERİ.....	85
3.5. KÜMELEME ANALİZİ UYGULAMALARI	88
3.5.1. Hiyerarşik Kümeleme Analizi Yöntemleri Uygulamaları.....	94
3.5.2. Hiyerarşik Olmayan Kümeleme Analizi Yöntemleri Uygulamaları	112
SONUÇ VE DEĞERLENDİRME	119
KAYNAKÇA.....	125
EKLER.....	137

SİMGELER VE KISALTMALAR DİZİNİ

%	Yüzde
\$	Dolar
-	Bilgi Mevcut Değil veya Bulunamamıştır
AB	Avrupa Birliği
ABD	Amerika Birleşik Devletleri
AB-27	Avrupa Birliğine Üye Olan 27 Ülke
BM	Birleşmiş Milletler
d	Uzaklık
DHM	Devlet Hava Meydanları İşletmesi
DWT	Bir Geminin Taşıyabileceği En Çok Ağırlık (Dead Weight Tonnage)
ESS	Hata Kareler Toplamı
EİT	Ekonomik İşbirliği Teşkilatı
GRT	Sicile Kayıtlı Gros Tonaj (Gross Register Tonnage)
GSYİH	Gayri Safi Yurt İçi Hâsıla
KEİ	Karadeniz Ekonomik İşbirliği
KGM	Karayolları Genel Müdürlüğü
KİT	Kamu İktisadi Teşebbüsü
km	Kilometre
km²	Kilometre Kare
max	En Büyük
min	En Küçük
sa	Saat
OECD	Ekonomik İşbirliği ve Kalkınma Örgütü
Ort.	Ortalama
PAM	Medoid Yöntemi (Partition Around Medoids)
TCDD	Türkiye Cumhuriyeti Devlet Demiryolları İşletmesi Genel Müdürlüğü
TEU	20 Feet’lik Konteyner (Twenty-foot Equivalent Unit)
TL	Türk Lirası
TRACECA	Avrupa Kafkasya Ulaşım Koridoru
UDHB	Ulaştırma, Denizcilik ve Haberleşme Bakanlığı
SPSS	Sosyal Bilimler için İstatistiksel Paket Programı (Statistical Package for the Social Sciences)

TABLolar LİSTESİ

Tablo 1: Ulaştırma Türleri Değerlendirme Tablosu	8
Tablo 2: Türkiye’de Karayolu Ulaşımıyla İlgili Bazı Tanımlar	20
Tablo 3: Bazı Ülkelerin 2001 Yılı için Yolcu Taşımacılığı Tercihleri (%).....	20
Tablo 4: Denizyolu Ulaşımı ile İlgili Bazı Tanımlar	21
Tablo 5: İncelenen Çalışmalarda Kullanılan Yazılımlar	71
Tablo 6: Kullanılan Değişkenler	82
Tablo 7: Veri Setindeki Değişkenler için Bazı Tanımlayıcı İstatistikler.....	83
Tablo 8: Veri Setindeki Değişkenlerin Çarpıklık ve Basıklık Değerleri	84
Tablo 9: Veri Setine Ait Bileşenler için Özdeğer ve Önem Değerleri.....	86
Tablo 10: Uygulamada Kullanılan Hiyerarşik Kümeleme Yöntemlerindeki Kofenetik Korelasyon Katsayıları.....	89
Tablo 11: Tek Bağlantı Yöntemi için Küme Geçerlilik İndeksleri Sonuçları	90
Tablo 12: Tam Bağlantı Yöntemi için Küme Geçerlilik İndeksleri Sonuçları	90
Tablo 13: Ortalama Bağlantı Yöntemi Küme Geçerlilik İndeksleri Sonuçları.....	91
Tablo 14: Ward Yöntemi için Küme Geçerlilik İndeksleri Sonuçları	91
Tablo 15: Ağırlık Merkezi Yöntemi Küme Geçerlilik İndeksleri Sonuçları	92
Tablo 16: Medyan Yöntemine Göre Küme Geçerlilik İndeksleri Sonuçları	92
Tablo 17: k-Ortalamlar Yöntemi için Küme Geçerlilik İndeksleri Sonuçları.....	93
Tablo 18: Tek Bağlantı Yöntemine Göre Kümeler.....	94
Tablo 19: Tek Bağlantı Yöntemi için Uygulanan Aglomerasyon Tablosu	95
Tablo 19.1: 77 Adet İl için Tek Bağlantı Yöntemine Göre Kümeler	97
Tablo 20: Tam Bağlantı Yöntemine Göre Kümeler.....	99
Tablo 21: Tam Bağlantı Yöntemi için Uygulanan Aglomerasyon Tablosu	100
Tablo 22: Ortalama Bağlantı Yöntemine Göre Kümeler.....	102
Tablo 23: Ortalama Bağlantı Yöntemi için Uygulanan Aglomerasyon Tablosu.....	103
Tablo 24: Ağırlık Merkezi Yöntemi için Uygulanan Aglomerasyon Tablosu	105
Tablo 25: Medyan Yöntemine Göre Kümeler	107
Tablo 26: Ward Yöntemine Göre Kümeler.....	109
Tablo 27: Ward Yöntemi için Uygulanan Aglomerasyon Tablosu	110
Tablo 28: k-Ortalamlar Yöntemine Göre Kümeler.....	113
Tablo 29: k-Ortalamlar Yöntemi için Küme Merkezleri	114
Tablo 30: k-Ortalamlar Küme Merkezleri Arasındaki Öklid Uzaklık Değerleri...	115
Tablo 31: k-Ortalamlar Yönteminde Kümelerde Bulunan İllerin Sayısı	115

TABLÖLAR LİSTESİ (Devamı)

Tablo 32: k-Medoid Yöntemine Göre Kümeler.....	116
Tablo 33: Kruskal Wallis Testi Sonuç Tablosu -2 (k-Ortalamalar Yöntemi).....	118



ŞEKİLLER LİSTESİ

Şekil 1: Türkiye Demiryolları Haritası	22
Şekil 2: Türkiye’de Mevcut Havaalanları Haritası (2018)	23
Şekil 3: Üç Boyutlu Koordinat Düzleminde Bir Kümeleme Örneği	25
Şekil 4: John Snow’un 1854 Londra’da Ölümlü Kolera Vakaları İşaretlemeleri.....	27
Şekil 5: Kümeleme Analizi Karar Diyagramı.....	28
Şekil 6: Kümeleme Tekniklerinin Genel Yapısı.....	29
Şekil 7: Ölçülen İki Nesne Arasındaki Öklid Uzaklık Örneği.....	34
Şekil 8: Aglomeratif ve Ayırıcı Kümelerin Çalışma Yönleri	38
Şekil 9: Dendrogramda Bir Ters Çevirme (Inversion) Görüntüsü.....	40
Şekil 10: Kümeleme Uzaklığı (En Kısa)	41
Şekil 11: Tek Bağlantı Yönteminde Bir Birleşim Örneği.....	42
Şekil 12: Kümeleme Uzaklığı (En Uzun)	42
Şekil 13: Ortalama Bağlantı Yöntemi.....	43
Şekil 14: Ağırlık Merkezi Yöntemi	46
Şekil 15: İllerin İlk İki Ana Bileşen Skorlarının Kullanımıyla Oluşan Grafiği.	63
Şekil 16: Ana Bileşen Skorlarının Kullanımıyla Oluşan Üç Boyutlu Grafik Örneği	64
Şekil 17: Veri Seti Karakteristiklerine Ait Yığın Grafiği	87
Şekil 18: Veri Setinde İllerin İki Boyutlu Ana Bileşen Saçılım Grafiği.....	87
Şekil 19: Tek Bağlantı Yöntemi İçin Kümeleme Analizi Dendrogramı.....	96
Şekil 19.1: 77 İlden Oluşan Veri Setiyle Tek Bağlantı Yöntemi Dendrogramı	98
Şekil 20: Tam Bağlantı Yöntemi İçin Kümeleme Analizi Dendrogramı	101
Şekil 21: Ortalama Bağlantı Yöntemi İçin Kümeleme Analizi Dendrogramı	104
Şekil 22: Ağırlık Merkezi Yöntemi İçin Kümeleme Analizi Dendrogramı	106
Şekil 23: Medyan Yöntemi İçin Kümeleme Analizi Dendrogramı	108
Şekil 24: Ward Yöntemi İçin Kümeleme Analizi Dendrogramı	111
Şekil 25: İl Ulaştırması Faaliyet Durumu Türkiye İl Haritası	117

EKLER LİSTESİ

EK: 1 (R kodları)

EK: 2 (Türkiye Boru Hatları Haritası)

EK: 3 (İller Arası Sıcaklık Benzerlik Matrisi (Öklid Uzaklık))

EK: 4 (Veri Seti İçin Üç Boyutlu Ana Bileşen Skor Grafiği)

EK: 5 (Türkiye Karayolları Haritası)

EK: 6 (k-Ortalamalar Yöntemi Küme Grafiği (k=5))

EK: 7 (İki Boyutlu Ana Bileşen Skor Grafiği)

EK: 8 (Kruskal-Wallis Test Sonuç Tablosu – 1)

EK: 9 (İl Ulaştırması Faaliyet Durum Tablosu)

EK: 10 (Tek Bağlantı Yöntemine Göre Türkiye İl Haritası)

EK: 11 (Tam Bağlantı Yöntemine Göre Türkiye İl Haritası)

EK: 12 (Ortalama Bağlantı Yöntemine Göre 5 Küme İçin Türkiye İl Haritası)

EK: 13 (Ward Yöntemine Göre Türkiye İl Haritası)

EK: 14 (k-Ortalamalar Yöntemi (k=5) Sonucu Ortaya Çıkan İller Haritası)

EK: 15 (Medoid Yöntemine göre İl Haritası)

EK: 16 (Türkiye Limanları Haritası)

EK: 17 (Türkiye Konteyner Faaliyetleri Haritası)

GİRİŞ

Geçmişten günümüze, bir uygarlığın meydana gelmesinde ulaştırma konusu vazgeçilmez bir konumda yer almaktadır ve ulaştırma olmadan bir uygarlığın gelişimi düşünülemez. Roma, Mısır, Çin, gibi çeşitli coğrafyalardaki büyük uygarlıkların temelleri ulaştırma sistemlerinin üzerinde yükselmiştir. Günümüzde de farklı bir durum söz konusu değildir. Uygun ulaştırma olanaklarının bölgedeki durumu o bölgenin gelişmişliğini doğrudan etkilemektedir. Bu da ulaştırma kavramını medeni yaşam gereksiniminin temel unsurlarından birisi haline getirmiştir. Ulaştırma kavramını oluşturan temel bileşenler, bölgenin dolaşım sistemleri olarak nitelendirilebilir. Bir canlıda dolaşım sisteminin önemi neyse bölgedeki ulaştırma sistemlerinin öneminin de bundan farkı bulunmamaktadır.

Bir objenin yer değiştirme işlemi için ulaştırma faaliyetini gerçekleştirecek unsurlardan faydalanılması şarttır. Ulaştırma büyüklüğü, etki alanı ve ulaşımın ivedilik durumuna göre ulaştırma altyapısı, araç ve hız talebi değişebilir olup İngilizcede “transport”, “transportation” şeklinde ifade edilen ulaştırma, aynı zamanda nakliye, taşımacılık şeklinde de kullanılabilmektedir (Kaya, 2012: 4).

Ulaştırma faaliyetleri çeşitli sistemler kullanılarak hava, kara ve deniz unsurlarıyla yapılmakta olup hepsinin kendine özgü çeşitli avantaj ve dezavantajları bulunmaktadır. Çalışmamızda ise günümüzde ulaştırma faaliyetlerinde kullanılan ulaştırma türleri arasındaki karayolu, demiryolu, boru hattı, denizyolu ve havayolu ulaştırması olarak ele alınacaktır. Bazı araştırmacılar genel olarak telekomünikasyon konusunu da ulaştırma kavramı içinde değerlendirse de gerek bazı araştırmacıların ulaşımında etkisinin henüz yeterli olmadığını belirtmesi, gerekte ülkemiz verileri olarak incelediğimizde birçok verinin henüz yeni toplanmaya başlanmış olması ve birçok eksik bilgi olması göz önünde bulundurularak araştırmamızda bu konuyla ilgili değişkenler analizde ele alınmamıştır. Aynı şekilde iç su yolu ve nehir taşımacılığı konusunda da Türkiye sınırları içerisinde bu taşımacılıkta en uygun nehir Sakarya nehridir ve özellikle kereste taşımacılığı konusunda bazı planlama ve uygulamalar yapılmasına rağmen karayolu daha tercih edilebilir bulunduğu için su yolu taşıması yapılmamaktadır. Ek olarak Van Gölü’nde yapılan feribot seferleri de esas olarak lokomotif taşımacılığında kullanıldığı için ayrı bir değerlendirmeye tabi tutulmamıştır.

Çalışmamızda ulaştırma kavramını gerek uluslararası boyutta tanımlamak, gerekse ulusal boyutta ulaşım sistemlerini ele alarak Türkiye’de iller bazında elde edilmiş verilerle bu 81 ilin benzerliklerine göre uygun biçimde gruplara ayrılması amacı bulunmaktadır. Bu gruplara göre illerin mevcut ulaşım durumları temel alınarak yeni politika ve hedefler belirlenebilmesinin mümkün olması da bu amaca dâhildir. Literatürde buna benzer çalışmalar genelde illerin sosyoekonomik durumu, eğitim, sağlık ve kültür gibi belli alanlarda yapılmış olup daha farklı bir açıdan ele alınmak istenmiştir. İnceleme için birden fazla sayıda değişkeni bir arada inceleyebilme yeteneğine sahip çok değişkenli istatistiksel yöntemlerden biri olan kümeleme analizi tercih edilmiştir. Bu çalışma kümeleme analizi ve ulaştırma kavramlarıyla ilgili temel unsurlar tanıtılarak gerçekleştirilecek olup üç ana bölümde incelenecektir.

Kümeleme analizi, istatistik biliminin temel araştırma alanlarından birisi olan çok değişkenli istatistikte önemli ve geliştirilmeye açık bir yöntemler bütünüdür. Bu analizin kullanımındaki amaç ise araştırma hedefindeki çok sayıda değişken unsurlara sahip gözlemleri, birbirine benzer homojen yapıda gruplar elde etmek, ayrıca bu gruplar arasında ise yüksek heterojen yapılar elde etmektir. Bir obje veya bir yapının gruplanması uygarlık tarihinin başlangıcından beri bir ihtiyaç olup, günlük yaşamın hemen her alanında birçok önemli amaçla kullanılabilmektedir.

Bu çalışma ulaştırma kavramı, kümeleme analizi ve uygulama olmak üzere üç bölümden oluşmaktadır. Çalışmanın birinci bölümünde kümeleme analizi uygulamalarında kullanacağımız ulaştırma kavramı hakkında temel bilgiler verilip tarihi, sosyal, ekonomik, politik, çevresel açılardan önemine vurgu yapılmıştır. Temel olarak bu bilgiler verildikten sonra ulaşım yöntemleri olan karayolu, denizyolu, demiryolu, havayolu, boru hatları ve bu yöntemleri kullanan temel ulaşım araç gereçleri hakkında açıklamalar, uluslararası alandaki örnekleri, avantaj ve dezavantajları verilmiştir. Bölümün sonunda da bu anlatılanlara uygun, çalışmamızda kullanacağımız verilere dönük olacak biçimde Türkiye’de ulaştırma konusunda açıklamalar yer almıştır. Çalışmanın ikinci bölümünde tezin temel konusu olan çok değişkenli istatistik alanında yer alan kümeleme analizine değinilmiştir. Söz konusu uygulamalar ile hedef alınan gözlem veya değişkenleri aralarında mümkün olduğunca heterojen kümelere ayırarak bu küme içlerinde de homojen objeleri bir araya getirme amacı güden; böylece incelenen veriyle ilgili inceleme yapılarak hakkında yorumlanmasını kolaylaştıran bu analiz yöntemlerinde hiyerarşik ve

hiyerarşik olmayan yöntemler olarak iki şekilde yaklaşılmıştır. Hiyerarşik yöntemlerde ağırlık merkezi yöntemi, tek, tam, ortalama, ward ve medyan bağlantı yöntemleri olmak üzere altı yöntem ele alınmış, hiyerarşik olmayan yöntemlerde ise k-ortalamlar yöntemi ile daha yeni bir yöntem olan k-medoid yöntemleri ele alınmış ve açıklanmıştır. Bununla birlikte bu yöntemleri doğru ve etkili biçimde kullanabilmek için verilerin inceleme ve standartlaştırma işlemleriyle doğru küme sayısının elde edilebilmesi için başta küme geçerlilik indeksleri ile iki ve üç boyutlu ana bileşen skor grafiklerinin oluşturularak incelenmesi olmak üzere bazı yöntemler üzerinde de durulmuştur. Çalışmanın son bölümünde ise daha önce yapılan literatürdeki benzer çalışmalardan bahsedildikten sonra önceki bölümlerde anlatılan kavramlar uygulamalı bir şekilde Türkiye'nin iller bazında 2004-2018 yılları arasındaki ulaştırma göstergeleri bir araya getirilerek yukarıda anlatılan yöntemlerle kümeleme analizi yapılmıştır. Bu analizler ışığında illerin durumuyla ilgili sonuçlar incelenmiş; daha sonra farklı çalışmalarda kullanılabilmesi adına yorumlanarak Tablo ve haritalar eşliğinde durumları ele alınmıştır.

Analizlerin yapılması için açık kaynak kodlu ve istatistiksel açıdan kendine yaygın bir kullanım alanı oluşturabilmiş R 3.5.1 programı ve yine bu konuda kullanılan en yaygın istatistik programından birisi olan SPSS 16.0 istatistik paket programından faydalanılmıştır.

BİRİNCİ BÖLÜM

ULAŞTIRMA KAVRAMI

1.1. ULAŞTIRMA KAVRAMI

Ulaştırma, herhangi bir canlı veya cansız varlığın bir noktadan başka bir noktaya belli araçlar kullanılarak belli amaçlar doğrultusunda ekonomik bir kavram olarak ele alınabilir nitelikte yer değiştirmesini sağlayan bir hizmettir. Burada hedef, faydalı olacak biçimde herhangi bir canlı veya eşyanın belli bir amaca dönük olacak biçimde taşınmasıdır. Bu ulaşımı sağlayan herhangi bir unsur ise taşıma sistemi olarak tanımlanmaktadır. Ulaştırma sistemleri birkaç temel unsurun bir araya gelmesiyle oluşmaktadır. Bunlar genel olarak aşağıdaki gibi ifade edilebilir (Kaya, 2012: 6):

- * Ulaşım yolu
- * Ulaşım aracı/taşıt
- * Taşınan (canlı ve cansız varlıklar)
- * Taşıyan (işletmeci)
- * Taşıma sözleşmesi

Tüm modern ekonomik faaliyetlerin merkezinde ticaret vardır. İnsanlar nakit para karşılığında emeğini, fikirlerini, malını ve hizmetlerini ticaret yoluyla kullanıma sunarak geçimlerini sağlarlar; bu amaçla firmalar birbiriyle, bireylerle veya hükümetlerle teknoloji, uzmanlık, mali kapasite, ara mallar, idari işlevler ve diğer birçok şeyin ticaretini yaparlar. Bütün bu işlemler çeşitli bağlantılar gerektirir ve çoğunlukla da malların veya insanların işe, alışverişe, turistik yerlere, toplantı yerlerine taşınmasını gerektirir. Bu nedenle, ulaşımın tüm bu ekonomik faaliyetin merkezinde olduğunu söylemek doğru bir ifade olacaktır (Small ve diğerleri, 2007: 1).

Ulaştırma belli bir zaman periyodu içinde kullanılan araçlardan ziyade bu araçların bir araya toplanıp oluşturduğu sistemler ve ulaşımın yapıldığı yerler bakımından oldukça büyük bir genişlik ve çeşitlilik kazanmıştır. Ulaşımında yaşanan belli değişimler bir dönüm noktası olarak kabul edilmekte olup bu dönüm noktalarından birisi de motorun icat edilerek çok sayıda ve çeşitte araçların ortaya çıkması olmuştur (Bakırcı, 2012: 342).

Ürün ve hizmet dağıtımı, yurt içi ve yurt dışı ticaretin geliştirilmesi, doğal kaynakların verimli dağıtımı, taşıma sektörünün düzgün planlama ve

uygulamalarına bağlıdır. Ulaştırma sektörü üretim yapma üzerine bir kavram değildir, fakat diğer sektörlerde verimli çalışma ve üretkenliklerinde en önemli konumda bulunan bir hizmet sektörüdür (Akgüngör ve Demirel, 2004: 424).

Ulaşım tarihi, sosyal, ekonomik, politik, çevresel öneme sahip çok boyutlu bir faaliyettir. Ulaştırma sistemleri de bu faaliyetin sürdürülmesini sağlamaktadır.

1.1.1. Ulaştırmanın Tarihi Önemi

Ulaştırma faaliyet ve sistemleri, tarih boyunca medeniyetlerin yükselişinde (Mısır, Roma ve Çin medeniyetleri gibi), toplumların gelişmesinde (sosyal yapıların yaratılması) ve ulusal savunmada (Roma İmparatorluğu karayolu ağı, Amerikan karayolu ağı gibi) birçok farklı rol oynamıştır (Rodrigue ve diğerleri, 2013: 5).

Ulaştırmanın önemi tarihsel gelişimiyle doğrudan ilişkilidir. İlk çağlardaki çeşitli yer değiştirme hareketleri beslenme, ticaret, spor ve inanç gibi nedenlerle meydana gelmiştir. İnsanlar daima av hayvanları, tarım ürünleri, endüstri ürünleri gibi temel ihtiyaçlarını bir yerden başka bir yere taşıma ihtiyacı içinde kalmıştır. Ulaştırma ihtiyacını eski çağlarda hayvanlar, rüzgâr, akıntı, bizzat kendi gücü, yerçekimi gibi doğal enerjiyle karşılayan insanoğlu daha sonra medeniyetin gelişiminde en büyük kırılma noktalarından biri olan tekerleğin icadı ile basit ulaşım araçları kullanmaya başlamıştır. Bu şekilde daha az kas gücüyle daha fazla yükün, daha uzun mesafelere ve daha hızlı şekilde ulaştırılma olanağı elde edilmiştir. Bu gelişmeyi daha sonra karayolları ve motorsuz hayvan gücüyle ilerleyen ulaşım araçları izlemiştir. Suyun kaldırma kuvvetinin ve ulaşım yüzeyi olarak kullanılabileceğinin keşfedilmesiyle ulaşım farklı bir boyut kazanmıştır. İlerleyen zamanlarda ana ulaşım yolu denizler ve büyük suyolları üzerinde yoğunlaşmıştır. Bunun sonucunda denizyollarına bağlı kara parçalarının önemi artarak büyük yerleşim yerleri genellikle denizler ve büyük suyolları üzerinde kurulmaya başlanmıştır (Kaya, 2012: 8-9).

Günümüz ulaşım sistemlerinin temelleri ise buhar motoru ve içten yanmalı motorun icadıyla atılmış, hava ve demiryollarının oluşması ve motorlu ulaşım sistemlerinin yaygınlaşması bunu takip etmiştir.

1.1.2. Ulaştırmanın Sosyal Önemi

Ulaştırma türleri sağlık hizmetlerine, sosyal hizmetlere, kültürel ve sanatsal etkinliklere erişim kolaylığı sunarak sosyal hizmetlere katkı sağlamaktadır. Ayrıca sosyal hayata ticari kolaylıklar ve canlılık getiren bir tür enerji işlevi görmektedir. Başlangıçta hayatı idame etmeye yardımcı olması gibi nedenlerle yapılan ulaşım faaliyetleri, 20. yüzyılın ikinci yarısından itibaren zorunlu ihtiyaçları karşılamının yanında, zevk ve kültürel gelişim, küreselleşme amaçlı da yapılmaya başlamıştır. Ulaşım faaliyetleri, zararlı maddelerin, salgın hastalıkların yayılımı, kaza ve yaralanmalar gibi olumsuzlukları beraberinde getirebilmektedir (Kaya, 2012: 12).

1.1.3. Ulaştırmanın Ekonomik Önemi

Taşımacılığın gelişimi her zaman ekonomik kalkınma ile de doğrudan ilişkili olmuştur. Tek başına bir endüstri olarak da (otomobil imalatı, hava taşımacılığı şirketleri vb.) tanımlanabilir. Taşımacılık sektörü de mal ve hizmet üretiminde ekonomik bir faktördür. Ekonomik faaliyetlerin katma değerine katkıda bulunur, ölçek ekonomilerini olumlu etkiler, arazi (emlak) değerini ve bölgesel kalkınmayı artırır. Ulaştırmanın ekonomik faaliyetleri hem şekillendirme hem de bu faaliyetler tarafından şekillendirilme durumu her zaman mevcuttur (Rodrigue ve diğerleri, 2013: 5).

1.1.4. Ulaştırmanın Politik Önemi

Devletler, ulaşırmada yatırım kaynakları ve düzenleyiciler olarak kritik bir rol oynamaktadır. Çoğu ulaşım talebi ekonomik zorunluluklarla ilgili olmakla birlikte, ulusal erişilebilirlik veya iş yaratma gibi politik nedenlerle birçok iletişim koridoru inşa edilmiştir. Dolayısıyla ulaşım, ulus inşası ve ulusal birlik üzerinde doğrudan bir etkiye sahiptir (Rodrigue ve diğerleri, 2013:5). Bunun aynı zamanda savunma anlamında da mutlak bir önemi bulunmaktadır. Örneğin Türk Kurtuluş Savaşı'nda demiryolları ve karayollarının eksikliği her noktada zorluklara sebep olmuş olup birçok belgede bu zorlukların kayıtları mevcuttur. Ayrıca tarih boyunca deniz kuvvetlerine sahip birçok devletin savaşlarda daha avantajlı olduğu bilinmektedir. Günümüzde hava ve uzay gücü kuvvetli devletler aynı avantajlara sahip durumdadır (Kaya, 2012: 13).

1.1.5. Ulaştırmanın Çevresel Önemi

Ulaştırma faaliyetlerinin getirdiği belirgin avantajlara rağmen, günümüz çevreye olan etkileri bakımından önemli sonuçlar meydana getiren faktörlerden birisidir. Hava ve su kalitesi, gürültü seviyesi, halk sağlığı, incelenmesi gereken sonuçlardan bazılarıdır. Taşımacılık ile ilgili tüm kararlarda, söz konusu çevreye karşı olumsuz etkileri göz önünde bulundurulmalıdır (Rodrigue ve diğerleri, 2013: 5).

Bu konuya ilişkin olarak karayolu ulaşımının, özellikle otomobil kullanımının ulaştırma sektöründe ağırlıkta olduğu günümüzde sebep olduğu hava kirliliği, trafik sıkışıklığı ve kazalar gibi çeşitli olumsuz çevre etkileri görülmektedir. Karayolu ulaşımı alt sektörünün bu şekilde bir ulaşım modelini temel almaya devam etmesi, çevresel sorunlarla birlikte hem zaman hem maliyetlerde meydana gelen olumsuzlukların devam etmesine sebep olmaktadır (Nalçakan ve diğerleri, 2012: 69).

1.2. ULAŞTIRMA TÜRLERİ

Her bir ulaşım çeşidi bir dizi teknik, operasyonel ve ticari özellik ile tanımlanır (Rodrigue ve diğerleri, 2013: 89). Ulaştırma türleri yeryüzünde hava, kara ve deniz olmak üzere üç tür mekânda faaliyet göstermekte olup bu türlerin her biri kendine has avantaj ve dezavantajlara sahip beş ulaştırma şekliyle incelenmektedir:

- * Karayolu ulaştırması,
- * Demiryolu, ulaştırması,
- * Denizyolu ulaştırması,
- * Havayolu ulaştırması,
- * Boru hatları ulaştırması,

Bu sistemlerin aynı zamanda birden fazla ulaştırma türü bir araya gelerek multimodal yani çok modlu taşımacılık hizmeti verebilme gibi seçenekleri de bulunmaktadır. Bu ulaştırma türlerinden boru hatları ulaştırması sadece yük taşıma amacıyla kullanılmakta olup diğer türlerde hem yük hem yolcu taşımacılığı yapılabilir.

Her ulaştırma türü kendini özgü farklı karakteristikler barındırır ve hangi koşulda en iyi seçim olacağı ise taşınacak malın türüne, konuma, mesafeye, miktar ve bir dizi başka faktörlere bağlıdır. Mevcut ulaşım seçenekleri çoğu zaman oldukça dardır. Brezilya'dan Amsterdam'a kahve taşınmak istenirse deniz nakliyesi seçeneği;

Meksika Körfezi'nden Dallas'a doğal gaz taşınmak istenirse, boru hattı seçeneği; Atlantik'te hızlı bir paket servisi istenirse, hava taşımacılığı seçeneği kullanılmak durumunda kalınacaktır. Yine de genel olarak çoğu taşıma işlemi karayoluyla yapılmaktadır. Örneğin İngiltere'de de tüm yüklerin %65'i ve 240 milyar ton km' den 160 milyar ton km değeri karayoluyla taşınmıştır. Bu konuyla ilgili Tablo 1'de ulaştırma türleriyle ilgili bir değerlendirilme tablosu hazırlanmıştır. Bu tabloya göre karayolu, demiryolu, denizyolu, havayolu, boru hatları ulaşım türleri maliyet, hız, esneklik, hacim/ağırlık limiti ve erişilebilirlik durumlarına göre 1 en iyi, 5 en kötü puan olmak üzere bir değerlendirme yapılmıştır (Waters, 2003: 310-317).

Tablo 1: Ulaştırma Türleri Değerlendirme Tablosu

	Demiryolu	Karayolu	Denizyolu	Havayolu	Boru Hattı
Maliyet	3	4	1	5	2
Hız	3	2	4	1	5
Esneklik	2	1	4	3	5
Hacim\Ağırlık Limiti	3	4	1	5	2
Erişilebilirlik	2	1	4	3	5

Kaynak: (Waters, 2003: 317)

Tablo 1'deki ulaştırma türleri, hareketliliğin sağlanmasında bir aracı olduğu için taşımacılık sistemlerinin temel birer bileşenidir. Her bir türde kendine özgü gereksinimler, özellikler vardır. Buna bağlı olarak yük ve yolcu trafiğinin özel taleplerini karşılamak için tasarlanmışlardır. Bu türlerde yeryüzünün farklı yerlerinde kullanılma ve kullanım şekillerinde belirgin farklılıklar ortaya çıkmaktadır. Daha yakın zamanlarda, ulaşım türlerini intermodalite yoluyla bütünleştirme ve bu türlerin üretim ve dağıtım faaliyetlerine daha yakından bağlantı sağlanması yönünde bir eğilim vardır. Bununla beraber, yolcu ve yük taşıma faaliyeti de pek çok türe giderek daha fazla ayrılmaktadır (Rodrigue ve diğerleri, 2013: 89).

1.2.1. Karayolu Ulaşımı

Özellikle yirminci yüzyılın ilk yarısından itibaren günümüz yol ağı gelişimi hızlanmaya başlamıştır. 1920'lerde, ilk her türlü hava koşullarına uygun kıtalararası karayolu olan Lincoln otoyolu, New York ve San Francisco arasında 5.300 km'den fazla uzunlukta inşa edilmiştir. Bununla birlikte Almanlar, 1932'de modern karayolunu (otoban) ilk kez kullanmışlardı ve sonuçta karayolu sistemlerinin ortak özellikleri olacak kısıtlı erişimli noktalar, üst geçitler ve yol ayrımı gibi birçok özelliklere sahiptiler. İkinci Dünya Savaşı sonrası dönem ise küresel çapta karayolu ağlarının hızlı bir şekilde gelişim çağını temsil ediyordu. Bu konuda en dikkat çekici başarılarından birisi, şüphesiz 1956'da başlatılan Amerikan eyaletler arası otoyol sisteminin kurulmasıydı. Bu sistemin stratejik hedefi, Amerikan ekonomisine hizmet veren, aynı zamanda askeri hareketleri destekleyebilen ve acil durumlarda hava şeridi gibi davranabilen ulusal bir yol sistemini sağlamaktı. 1950'lerden 1970'lere kadar yaklaşık 56.000 km inşa edildi, ancak 1975 ile 2006 yılları arasında sisteme yalnızca 15.000 km ilave edilmiştir. Bu duruma artan inşaat maliyetleri ve azalan gelirlerin payı bulunmaktaydı. Genel olarak, tüm büyük Amerikan şehirlerini bir uçtan diğerine bağlamak için yaklaşık 70.000 km dört şeritli ve altı şeritli otoyollar yapılmıştır. 1962'de tamamlanan Trans-Canada Otoyolu ile Kanada'da da benzer bir proje gerçekleştirilmiştir. 1970'lerde her modern ulus, Batı Avrupa örneğinde Pan-Avrupa Ulaştırma Koridorları örneğinde olduğu gibi ulusal bir otoyol sistemi kurmuştur. Bu eğilim şimdi birçok sanayileşen ülkede gerçekleşmeye devam etmektedir. Örneğin 2011 yılına kadar Çin 80.000 km uzunluğa çıkan ulusal bir otoyol sistemi inşa etmekte olup bu inşaat yılda yaklaşık ortalama 2.000 km hızla gerçekleşmeye devam etmektedir (Rodrigue ve diğerleri, 2013: 90).

Nakliye maliyetlerinin yüksek olmasına ve diğer ulaştırma türlerine göre düşük kapasiteli taşıma oranlarına rağmen ilk yatırım maliyetleri diğer ulaşım türlerine göre düşüktür. Ancak buna rağmen güvenilirlik konusunda diğerlerine göre bir farkının bulunmaması, kısa mesafeli nakliyyede en uygun yöntem olması gibi avantajlar onu diğer taşıma yöntemlerinden ayrı bir şekilde değerlendirilmesine yol açmıştır (Xukuo ve Qiong, 2013: 414).

Karayolu taşımacılığı, demiryolunun aksine geniş karayolu ağlarını kullanabilme avantajına sahip olması nedeniyle kullanıcıların kendi yollarını oluşturmaları ve sürdürmeleri gerekmez. Ayrıca, araçların böyle esnek olmayan

sistem rotasına uyması gerekmediğinden kısa sürede ve az planlama ile seyahatlere devam edebilirler (Waters, 2003: 313).

Tarihsel olarak karayolu taşımacılığı, motorsuz taşımacılık biçimlerini (yürüyüş, hayvanların evcilleştirilmesi ve on dokuzuncu yüzyılın sonunda bisiklet kullanımı) desteklemek için geliştirilirken, yirminci yüzyılın başından bu yana gelişimini en çok şekillendiren ise motorlu taşıtlarla ulaşım şeklidir. Kullanışlılığı, kolay satın alınabilirliği ve rahatlığına bağlı olarak kullanıcının seçebileceği kısa mesafeli hareketler için çeşitli motorlu ve motorsuz seçenekler sunar. Otomobil, esneklik ve kolaylık sağladığı için tercih edilen bir yolcu taşımacılığı şekli olarak ortaya çıkmıştır. Karayolu ulaşımının diğer türlere göre önemli avantajları bulunmaktadır. Araçların sermaye maliyeti nispeten küçüktür ve bu da yeni kullanıcıların bu sektöre katılmasını nispeten kolay hale getirir. Bu da örneğin kamyon taşımacılığı endüstrisinin oldukça rekabetçi olmasını sağlar. Düşük sermaye maliyetleri, yeniliklerin ve yeni teknolojilerin endüstride hızla yayılmasını sağlar. Göreceli olarak yüksek ulaşım hızlarını engelleyen tek kısıt devletlerin koyduğu yasal sınırlamalardır. Karayolu taşımacılığı, rota seçimindeki yüksek esneklik sebebiyle hem yolcular hem de taşınan yük için kapıdan kapıya servis sağlama konusunda eşsiz bir fırsata sahiptir. Bu gibi avantajlar, otomobilleri ve kamyonları oldukça fazla miktarda tercih edilen bir tür haline getirerek kısa mesafeli yolculuklar için pazarın domine edilmesine neden olmuştur. Fakat arabaların ve kamyonların bu başarısı, ciddi sorunları da beraberinde getirmiştir. Trafik sıkışıklığı, Dünya'daki çoğu kentsel alanın doğal bir parçası haline gelmiştir. Buna ek olarak, taşımayla bağlantılı birçok temel çevresel sorunların arkasındaki sebebidir. Bu meselelerin ele alınması, yerelden Dünya'ya, gün geçtikçe her hukuk düzeyinde katı bir şekilde belirlenip uygulanması gereken bir politik zorunluluk haline gelmektedir (Rodrigue ve diğerleri, 2013: 90-93).

1.2.2. Demiryolu Ulaşımı

Demiryolu ulaşımı esasen sanayi devrimi ve buhar makinesinin makineleşmede kullanılabilesinin meyvelerinden birisidir. Diğer ulaştırma türlerine göre ve özellikle kara taşımacılığına göre daha az maliyetli olma, daha çok yük taşıma ve daha çok güvenlik sağlama gibi bazı avantajlara sahip bir ulaşım türüdür.

Demiryolu ulaştırması, araçların bağlı olduğu özel tasarlanmış bir yoldan oluşmaktadır. Lokomotif türleri ile bağlantılı ortalama bir fiziksel kısıtlama seviyesine sahiptirler ve özellikle navlun için düşük bir eğim gerekmektedir (Rodrigue ve diğerleri, 2013: 93). Demiryolunun ana dezavantajı, esnek olmamasıdır. Tüm tren hizmetleri yapım aşamasında önceden programlanmalıdır, böylece bu ulaşım ile bağlantılı her şeyin aynı raylara uyum sağlanacak şekilde düzenlenmesi kolaylaşır. Son dakika veya acil değişiklikler için yapılabilecek uygun esneklikler bu tür için çok sınırlıdır. Yine de tren operatörleri bir dizi farklılıklar sağlayabilir. Bazı özel hizmetler sunan servisler (bir trenin liman ve fabrika gibi iki yer arasında sürekli hareket ettiği yerlerde), tam tren hizmetleri (müşterilerin bir trenin tamamını tuttuğu yerler), tarifeli servislere bağlı tam yük vagonu, konteyner tarifeli servislerde nakliye veya paylaşımlı vagonlar gibi bazı ek seçenekler gündeme gelebilir (Waters, 2003: 311).

Demiryolu ulaşımı yüksek altyapı maliyetlerine karşılık, kurulu sistemlerde taşıma maliyetleri çok daha düşüktür. Nitekim yaygınlaşmaya başladığı 19. yüzyıldan itibaren yük taşımacılığında en çok kullanılan ulaştırma biçimi olmasına karşılık, zamanla motorlu kara taşıtları teknolojisinde yaşanan gelişmeler sonucu 20. yüzyılda karayolu taşımacılığı karşısında daha geri plana düşmüştür. Fakat ilerleyen zamanda daha az kaynak tüketimi ve daha çevreci bir sistem olması nedeniyle tekrar ön plana çıkarak Avrupa'daki gelişmiş ülkeler ve A.B.D. gibi yerlerde demiryolları lehine yasal düzenlemeler ve önlemler alınmıştır (Şahin ve diğerleri, 2012: 12).

Demiryoluyla ilgili bir diğer belirgin sıkıntılı durum ise, trenlerin sadece belirli terminaller arasında seyahat edebilmesi ve ara noktalarda duramamasıdır. Çoğu müşteri bu demiryolu terminallerinin hemen yakınında olmadığından, seyahatin devamında bu malları karayoluyla nakletmeleri gerekmektedir. Bu transferler fazladan zamana sebep olan ve demiryoluna göre oldukça yavaş bir alternatif olarak ayrılmaktadır. Bu yöntem Avustralya'daki Perth ve Sydney gibi uzun mesafeler için daha kullanışlıdır, ancak Hollanda'daki Rotterdam ve Arnhem

gibi küçük seyahatler için elverişsizdir. Bu tip erişim sınırı sorunları, çeşitli ulaştırma türlerinde de yaygın bir problemdir, ancak aşmanın çeşitli yolları vardır. Bunlardan en belirgin olanı, demiryolu terminallerine (veya limanlar, havaalanları, konteyner limanları veya uygun terminaller) yakın tesislerin belirlenmesidir. Fakat talep yeterince büyükse, bu durum için özel tesisler inşa etme seçeneği de göz önüne alınabilir. Örneğin bir elektrik santrali için, kömür madenine doğru özel bir demiryolu hattı yapmanın, kamyon kullanmaya göre uzun vadede daha düşük bir maliyet yükü getireceği unutulmamalıdır. Bu konuda bir diğer seçenekte birden fazla taşıma sisteminin bir arada kullanıldığı intermodal taşımacılık sistemidir (Waters, 2003: 311).

1.2.3. Deniz ve Suyolu Ulaşımı

Dünya'nın dörtte üçünün suyla kaplı olması ve ticari taşımanın büyük çoğunluğunun denizyolları üzerinden sağlandığı göz önüne alınırsa denizyolu ulaştırmasının önemi ortaya çıkmaktadır. Eski çağlardan beri deniz ticaretinin taşıdığı önemin günümüzde de değişmediğini söylemek yanlış olmaz. Bunda denizyolu ulaşım ve taşıma maliyetlerinin diğer ulaşımlara göre daha az maliyetli olmasının büyük payı bulunmaktadır. Denizyolu Taşımacılığı en ekonomik taşıma sistemi olması münasebetiyle bölgelerin gelişmesinde önemli bir paya sahiptir. Denizyolu taşımacılığı uluslararası taşımacılıkta en çok kullanılan taşıma şeklidir.

Temel olarak üç tür su taşımacılığı vardır; nehirler ve kanallar (genellikle iç suyolları olarak adlandırılır), kıyı taşımacılığı (kıyı boyunca denizden bir gemiden diğerine taşınması) ve deniz taşımacılığıdır (büyük denizlerde) (Waters, 2003: 314).

Kara ve hava ulaştırması türlerine benzeyen deniz taşımacılığı aynı zamanda fiziki özelliklerine göre coğrafi, stratejik ve ticari saha kontrolünde kullanıma sahip bir faaliyet alanıdır. Coğrafi düşünceler zaman içinde sabit olma eğilimindeyken, stratejik ve özellikle ticari konular çok daha dinamiktir. Deniz taşımacılığının fizyografisi nehirler ve denizler olmak üzere iki ana unsurdan oluşur. Bağlantılı olmalarına rağmen, her biri belirli bir deniz dolaşımı alanını temsil eder. Deniz taşımacılığı kavramı, daha iyi deniz yolu rotaları olarak bilinen düzenli güzergâhların varlığına dayanmaktadır. Deniz taşımacılığı güvenlik ve güvenilirliğinde kayda değer gelişmeler sağlamış olsa da deniz rotaları hâkim rüzgârlar, akıntılar ve genel hava durumları tarafından sık sık engellenmektedir. Kuzey Atlantik ve Kuzey Pasifik (50

ila 60 derece kuzeyde) kış aylarında bazen denizciliği engelleyen ağır dalga faaliyetlerine maruz kalır ve gemilerin daha düşük enlemlerde rota izlemelerine ve böylece rota uzunluklarını artırmalarına neden olabilir. Yaz muson mevsiminde (nisan-ekim), Hint Okyanusu ve Güney Çin Denizi'nde denizcilik daha tehlikeli hale gelebilir (Rodrigue ve diğerleri, 2013: 96).

Denizcilik sektörünün bir ticaret ve hizmet kolu olmasının yanında bir endüstri kolu olarak gerek liman hizmetleri, gerek gemi inşa sanayi olarak, gerekse canlı veya cansız doğal kaynakların çıkartılması olarak çeşitli faaliyet imkânları sağlamaktadır. Denizyolu ulaştırmasının ana amacı ekonomik, güvenli, çevreye zarar vermeyen ve hukuk kurallarına uygun çerçevede ülke gelişiminin ekonomik ve sosyal hedeflerinin gerektirdiği şekilde hizmet sağlamaktır. Petrol, doğalgaz ve birçok değerli madenin denizlerin altında olduğu ve Dünya'nın çoğunun sularla kaplı olduğu göz önüne alındığında denizyolu ulaşımının öneminin artması kaçınılmaz bir durumdur (Bakanlığı, 2005: 7.3.2, 7.3.3).

19. yüzyılın sonlarından bu yana küresel bir deniz taşımacılığı sistemi kurulmuş ve konteynerizasyon ile genişletilmiştir. Bu sistem doğu-batı, kuzey-güney rotalarını ve bu rotalarda yer alan artelleri de kapsayan devasa bir alanı ve içindeki deniz limanlarını kapsamaktadır (Rodrigue ve diğerleri, 2013: 99).

Deniz Limanları: Limanlar, kara ve deniz alanları olmak üzere iki yük (bazen yolcular) dolaşım alanı arasındaki yakınsama noktalarıdır. Limanın İngilizce karşılığı olan “port” kelimesi Latince geçit veya kapı anlamındaki portus kelimesinden gelmiştir. Limanların görev şekli, sadece gemilere hizmet verilmesi şartında sınırlıdır ve bu nedenle deniz taşıtlarının seyrine elverişli olan suya erişim, tarihsel olarak en önemli saha konusu olmuştur. Alışlagelmiş dökme terminallerden 1960'ların başından itibaren konteyner terminallerine kademeli olarak geçiş, terminal seçiminin yanı sıra saha seçiminde de temel bir değişiklik getirmiştir. Konteynerle taşıma, özel konteyner limanlarının ortaya çıkması için liman dinamiklerini önemli ölçüde değiştirmiştir. Geleneksel dökme yük kargo gemileri ile karşılaştırıldığında, konteyner gemileri güverte vinçlerine sahip değildir ve bu nedenle konteyner terminali tesisleri, konteyner taşıyarak istiflemek için büyük miktarda yoğun vinçler ve bol miktarda depolama alanı sağlamalıdır. İlerleyen zamanlarda parmak iskeleleri artık yeterli gelmediğinden geminin daha hızlı geri dönmesi ve vinç ile konteyner depolama alanları arasında daha etkili liman operasyonlarını gerçekleştirmek için rıhtımlar yeniden tasarlanmıştır. Sonuç olarak, konteyner taşımacılığı küresel liman

operasyonlarının temel bir fonksiyonu haline gelerek daha fazla alan kaplama eğilimindeki liman terminallerinin yapısını ve yapılandırmasını değiştirmiştir. İç liman bölgeleri (örneğin bir körfezin sonunda veya bir nehir boyunca) genellikle nihai pazara daha yakın olma avantajına sahipken, deniz taşımacılığı rotalarından daha uzun sapmalar gerektirmektedir. Bu nedenle, en başarılı iç liman alanları geçit işlevi gören alanlardır (örneğin Anvers, Montreal, Constanza). Bir liman çıkışı, Dünya'nın en büyük limanlarının büyük sanayi bölgelerine açılan kapısı olması nedeniyle çeşitli yerel ve bölgesel endüstriyel faaliyetlerle bağlantılıdır (Rodrigue ve diğerleri, 2013: 139-142).

Denizyolu taşımacılığının belki de tek ve en önemli avantajı taşıma maliyetidir. Gerek ölçek ekonomisi, gerekse olası uzun rotaları karşılayabilme ve kapasite potansiyeli göz önüne alındığında en belirgin taşıma türü olarak göze çarpmaktadır. Bu faktörlerin hepsi de maliyetle ilişkilidir. Bunun dışında sefer düzenlerinin olmaması, yavaşlık, varış noktalarına erişim kısıtlılığı gibi çok belirgin dezavantajları da beraberinde getirmekte, bu sebeple üretim endüstrilerine ait tesisler denizdeki liman tesislerine çok yakın, hatta entegre edilmek zorunda kalınmaktadır. Örneğin çeşitli petrol ve petrol ürünleri tesislerinin genellikle limanlara yakın noktalarda kurulması da bu duruma bağlıdır (Şahin ve diğerleri, 2012: 14). Buna bağlı olarak deniz taşımacılığının diğer bir ana dezavantajı elbette uygun limanlarla sınırlı olma konusundaki kısıtlı esnekliğidir. Tedarikçi ve müşterilere yapılan yolculuklar, limanlara yakın olsalar bile, kaçınılmaz olarak bir ulaşım türü değişikliğine ihtiyaç duyarlar. Deniz taşımacılığındaki diğer bir sorun ise, nispeten yavaş olması, yükleri birleştirmek ve limanlara aktarmak için zamana ihtiyacı olmasıdır. Bununla birlikte, bu tür transferler daha verimli hale getirilebilir ve kıyı taşımacılığı, nispeten kısa mesafelerde bile karayolu taşımacılığı ile rekabet edebilir (Waters, 2003: 314).

1.2.4. Havayolu Ulaşımı

Ulaşım tarihinde yaşanan bazı gelişmeler bu noktada bazı kırılma noktaları oluşturmuştur. İçten yanmalı motorun icadı da bu kırılma noktalarından birisidir. İçten yanmalı motorun icat edilmesi birçok sayı ve çeşitte ulaşım aracının geliştirilip kullanımına zemin hazırlamıştır. Wright kardeşlerin ilk uçuşlarını gerçekleştirdiği uçakta bunlardan birisidir. Bu girişimler sonucu insanların uçarak bir yere ulaşması

imkân dâhiline girmiş ve I. Dünya Savaşı'nda askeri amaçlarla kullanılabilecek seviyeye ulaşmıştır. Kullanıcılarına savaşlarda büyük avantajlar veren bu araçlar II. Dünya Savaşı'nda da çok daha aktif bir biçimde rol alarak oldukça gelişmiş bir platforma dönüşmüştür. Savaştan sonra da hem sivil hem askeri amaçlarla kullanımları için gelişmeye devam etmiştir. Gerek gelişen hızları gerekse konforuyla ulaştırmada hızla öne çıkan bir ulaşım şekli haline gelmiştir. Gün geçtikçe artan talepler ve gelişen olanaklar, mevcut ülkelerin havayolu ulaştırmasına giderek daha fazla yatırım yapmasını bir zorunluluk haline getirmiştir. Uçakların grünüş, kapasite ve hızlarında her geçen gün yeni imkânlar devreye sokulmakta, aynı şekilde havaalanlarının donanım ve kapasiteleri de gün geçtikçe daha üst seviyelere çıkarılmaktadır (Bakırcı, 2012: 342).

Günümüz koşullarında düşük birim maliyetleri nedeniyle deniz taşımacılığı, uluslararası taşımacılıkta en yaygın kullanılan tür olsa da hızındaki düşüklüğü çoğu zaman kabul edilebilir değildir. Örneğin, Arjantin'de bir fabrika işletiyorsanız ve kritik bir makine arızalanırsa, Japonya'dan dört hafta içinde gelecek olan bir sonraki tarifeli gemiye yedek parçanın konması mantıklı olmayacaktır. Bu gibi durumlar için alternatif bir ulaşım şekli, hızlı bir ulaşım yöntemi olan hava taşımacılığıdır (Waters, 2003: 315).

Hava taşımacılığı, teorik olarak mükemmel bir rota seçim özgürlüğü sağlamaktadır. Bu ulaştırma şekli karayolu taşımacılığına göre bariz olarak daha az sınırlı olsa da olması gerekenden daha fazla kısıtları bulunmaktadır. Buna sebep olarak kısmen, hava taşıtının hızını arttırmak ve yakıt tüketimini azaltmak için üst atmosferik rüzgârlardan, özellikle de jet akımından yararlanmaya (veya kaçınmaya) yönelik fiziksel koşullar gösterilebilir. Buna ek olarak, navigasyon ve güvenliği kolaylaştırmak için özel koridorlar kurulmuştur. Stratejik ve politik faktörler de rota seçimini etkileyebilir. Örneğin, Güney Afrika Hava Yolları'nın uçuşlarının ırkçı ayırım döneminde birçok Afrika ülkesinin üzerinden olmasına izin verilmemiştir. Aynı şekilde Küba Havayollarının düzenli olarak Amerika Birleşik Devletleri'nden geçmesi yasaklanmıştır. Bu konuda çok önemli bir diğer önemli örnek ise Sibirya hava sahasının Soğuk Savaş sonrası Batı havayollarına açılması olmuştur. Hava ulaşımında bir diğer önemli unsur olan havaalanları, küresel ekonominin de dolaşım sistemindeki düğüm noktalarıdır. İnsan ve eşya trafiğine aracılık ederler. En önemli havaalanları, Dünya'nın en önde gelen şehir bölgelerinden birinin merkezinde olmasıyla, kilit pazarlar arasında aracılık rolünden veya her ikisinden de yararlanan

bir konumda bulunmalarıyla yakından ilişkilidir. Üretim bağlantılarının Dünya genelinde yayılması, bu aracılığın küresel ölçekte önemini artırmaktadır. Örneğin, Dubai'nin bir hava taşımacılığı merkezi olarak büyümesini teşvik eden önemli bir faktör, A340–500 gibi ultra uzun menzilli uçaklarla Dünya'daki herhangi iki konumun Dubai'deki bir duraktan bağlanabilmesidir (Rodrigue ve diğerleri, 2013: 104-146).

Deniz taşımacılığındaki gibi benzer şekilde, havayolları seyahatlerinde de malzeme yüklenip boşaltmak bir sorundur. Teslimatların doğru uçaklara yüklenmesi, sonra uçakların dışına ve müşterilere taşınması için büyük havaalanları etrafında yer alan her türlü tesis genellikle bulunmaktadır. Ne yazık ki, bu transferler burada da yine zaman alabilmekte ve hava yolculuğunun faydalarını azaltabilmektedir (Waters, 2003: 315).

Dünya'daki hava trafiğinin ifade edildiği üç ana havaalanı yoğunluğu vardır: Kuzey Doğu Amerika, Batı Avrupa ve Japonya'dır. Bu alanlarda kilit havalimanları veya daha doğrusu birden fazla havalimanına sahip olduğu için başlıca havalimanı şehirleri New York, Londra ve Tokyo'dur. Bu şehirler aynı zamanda Dünya'nın en seçkin şehirlerine ve en önemli finans merkezlerine karşılık gelmektedir. Yine de, bu üstünlük Pekin ve Dubai gibi yeni faaliyet merkezleri tarafından zorlanmaktadır. Dolayısıyla, hava yolcu trafiği seviyesi ile uluslararası kentsel sisteminde bir şehrin önceliği arasında doğrudan bir ilişki vardır. Büyük havaalanı terminalleri aynı zamanda dağıtım merkezleri, tam zamanında üreticileri, ofis parkları, oteller, restoranlar ve kongre merkezleri gibi önemli etkinlik yapılanmalarını da beraberinde getirmektedir. Havalimanlarındaki navlun faaliyeti düzeyi, özellikle Amerika Birleşik Devletleri gibi bazı ülkelerde, yolcu trafiğinden farklı olma eğilimindedir. Örneğin Midwest, Amerika Birleşik Devletleri'nin demografik ve ekonomik merkezi olma yolunda ilerleyen bir bölgedir ve burada birçok havayolu nakliye şirketi, merkezlerini Memphis (Federal Express) ve Louisville (UPS) gibi havaalanlarında çok az yolcu trafiğine sahip yerlere yerleştirmişlerdir. Pasifik Asya havaalanlarının önemi, bölgenin küresel ekonomideki spesifik rolü ile, özellikle elektronik ürünler konusunda bağlantılıdır. Bu ürünler değer / ağırlık oranı yüksek olma eğiliminde olduklarından, hava taşımacılığıyla Kuzey Amerika ve Batı Avrupa pazarlarına gönderilmeleri için özellikle uygundur. Bu önemin sebebi ise uzun mesafeli kargo uçakları yolcu uçaklarından daha az menzile sahip olmalarında yatmaktadır ve burada iki havaalanı önemli bir ara rol oynamaktadır, bunlar Anchorage (Asya

Pasifik – Kuzey Amerika trafiği) ve Dubai (Asya Pasifik-Batı Avrupa trafiği) havaalanlarıdır (Rodrigue ve diğerleri, 2013: 147).

Ülkemizde ise havayolu ulaştırması 80’li yıllardan itibaren önemli atılımlar gerçekleşmiştir. Çeşitli sanayi sektörünün gittikçe gelişmesi havayolu altyapısında, sistemlerinde ve yük taşımacılığında talep artışları gittikçe hissedilmiştir. Ülke hava sahasının Avrupa, Ortadoğu ve Asya ortasında geçiş güzergâhı konumunda olması önemli bir potansiyeli bünyesinde barındırmasına sebep olmuştur (Ulaştırma, D., ve Bakanlığı, H., 2013: 146).

Havaalanları büyük alan ihtiyacı olan yapılardır; pistler, terminal binaları, bakım hangarları ve otopark için geniş alanlar gerekmektedir. Pist, sistemin kapasitesini belirlediği için hava taşımacılığının en hayati unsurlarından biridir. Farklı havaalanlarında ölçek bazında önemli farklılıklar olsa da 500 hektarı aşan asgari boyutlar, kentsel alanda muazzam bir yükü beraberinde getirmektedir. Bu sebeple de havaalanları yeterli miktarda alanların bulunduğu kentsel alanların çevresine konumlandırılmaktadır. Fakat eski zamanlarda inşa edilen birçok havalimanı kentlerin büyümesi sebebiyle zaman içinde etrafı çevrelenmiştir. Yeni alanlar geliştirmek günümüzde, en azından Kuzey Amerika ve Avrupa’da gittikçe zorlaşmakta, planlama izni alınmış olsa bile, kent merkezlerinde genellikle uzakta gerçekleşmektedir. Kentlerdeki iş dünyasına ve demografik merkezlerine artan uzaklıkların yüksekliği önemli olumsuzlukları beraberinde getirmektedir. Kuzey Amerika’da son 30 yılda az sayıda yeni büyük ölçekli havalimanı inşası bu duruma örnektir. Sonuçta, çoğu havaalanının mevcut konumlarında uyum sağlaması gerektiği ortaya çıkmıştır. Bunun da yolu pistleri yeniden yapılandırarak ve mevcut terminal tesislerini yenilemekten geçmektedir (Rodrigue ve diğerleri, 2013: 148).

Havayolları için bir diğer sorun, üzerinde çok az kontrol sahibi oldukları maliyetleridir. Havayolu taşımacılığında yüksek sabit maliyetlerin (uçakların mali değeri çok yüksektir) ve yüksek değişken maliyetlerin (yakıt, iniş ücretleri, personel vb.) birleşimi vardır. Uçakların uçmasını sağlamak pahalıdır ve bu maliyetleri düşürmenin de gerçek bir yolu yoktur. Yine de havayollarında rekabet şiddetli olabilir ve talep edebileceği tutarı sınırlandırır, bu da sıkça iflasa uğrayan yeni havayolu şirketleri ortaya çıkartır (Waters, 2003: 316).

1.2.5. Boru Hatları Ulaşımı

Kara ve denizyolu hatlarına göre yatırım maliyetleri pahalı olmasına karşın diğer taşımalara göre hızlı, çevreci, güvenli olma ve atmosfer koşullarından etkilenmemesi sebebiyle yatırımlarının karşılığını kısa sürede verme avantajına sahiptir. Bu sebeple petrol ve doğalgazın üretim merkezlerinden tüketim merkezlerine kadar taşınması konusunda revaçta olan bir ulaşım şeklidir (Bakanlığı, 2005: 8).

Boru hatlarının temel kullanımları su ve atık su ile birlikte petrol ve doğal gazdır. Ayrıca yağda toz haline getirilmiş kömür gibi birkaç başka üründe de kullanılabilir. Boru hatları, uzun mesafelerde büyük miktarlarda hareket etme avantajına sahip olsa da ne yazık ki, bu taşımanın yavaş olması (taşınan ürün saatte 10 km' den daha az hareket eder), esnek olmaması (sadece sabit noktalar arasında taşınır) ve sadece belirli miktarlarda sıvıyı büyük miktarda taşıma gibi dezavantajları vardır. Bunlara ek olarak, özel boru hatları inşa etmeye yönelik ilk yatırım maliyetleri de çok büyük olacaktır. Bu yükseklikteki ilk yatırım maliyeti durumuna rağmen, boru hatları sıvıların (özellikle petrol ve gazın) uzun mesafelerde taşınmasında kullanılabilecek en ucuz yoldur. İlave yerel ağlarla, daha geniş konumlara (evlere su ve gaz temini gibi) iletim sağlayarak belirli bir esneklik kazandırılabilir (Waters, 2003: 316).

Boru hatları, neredeyse her yerde, yalnızca bir hammaddeyi bir yerden diğerine taşımak için belirli bir amaç üzere tasarlanmıştır. Büyük ölçüde özel sermaye ile inşa edilmişlerdir ve yerleşim yerlerinin gelişmesinde etkili olabilir. İnşası sırasında çevresel koşullara dayalı ağır denetimlere rağmen, boru hatlarının yönlendirilmesinde çoğu zaman coğrafi ortam koşullarına karşı duyarsız kalınmaktadır. Hassas alanlarda, örneğin kutuplarda görülen permafrost nedeniyle boruların gömülemeyeceği kutup / yarı arktik bölgelerde, Göçmen yaban hayatı üzerindeki etkileri bu sebeple şiddetli olabilir ve 1970'lerde Kanada'da önerilen McKenzie Vadisi boru hattında olduğu gibi yapımının onaylanmaması için yeterli olabilir. Jeopolitik faktörler de uluslararası sınırları geçen boru hatlarının yönlendirilmesinde çok önemli bir rol oynamaktadır. İsrail'den kaçınacak şekilde Orta Doğu'dan Akdeniz'e boru hatları döşenmiş ve Orta Asya'yı Akdeniz'e bağlayan yeni boru hatları, Kafkasya'daki cumhuriyetlerin etnik ve dinsel mozaigine bağlı olarak yönlendirilmektedir. Boru hattı inşaat maliyetleri değişken, işletme

maliyetleri çok düşüktür, ancak yukarıda belirtildiği gibi, boru hatları sıvı ve gaz ürünlerin taşınması için çok önemli bir ulaştırma şeklini temsil eder. Boru hatlarının en büyük dezavantajı, bu ulaşım şeklinin yapısal esneksizliğidir. Bir kez oluşturulduktan sonra (genellikle büyük bir masrafla), talebin genişlemesi kolayca karşılanamaz. Taşıma kapasitesi için belirli sınırlamalar vardır. Tersine bir arz veya talebin azaltılması, sistemin uygulanabilirliğini etkileyebilecek gelirlerin düşürülmesine neden olacaktır. Üretim veya tüketimdeki coğrafi kaymalardan kaynaklanan olası başka sınırlamalar da söz konusudur; burada bir yerden diğerine inşa edilmiş bir boru hattı, gerekli değişikliklere kolayca uyum sağlayamayabilir (Rodrigue ve diğerleri, 2013: 95-96).

Boru hattı taşımacılığı esasen demiryolu ve denizyolu taşımacılığı ile rekabet halinde olsa da işleme tesisleriyle dağıtımının yapılacağı tesisler arasında yapılması sebebiyle yüksek olan ilk kuruluş maliyeti karşılandıktan sonra oldukça düşük maliyetli bir ulaştırma alternatifine dönüşür (Şahin ve diğerleri, 2012: 15).

1.3. TÜRKİYE'DE ULAŞTIRMA HAKKINDA BİLGİLER

Karayolu Ulaşımı: Türkiye Cumhuriyeti, Osmanlı Devleti zamanından kalma 18.000 km uzunluğunda bakıma muhtaç durumda bir yol ağı devralmıştır. O yıllarda demiryollarına ağırlık verilmesinin yanında bu konuda 1929 yılında çıkarılan yol kanunu ile karayolu çalışmalarına da ağırlık verilmiştir. Yollar önceleri kazma, kürek gibi basit el aletleriyle ve insan gücüne dayanarak yapılırken, 1945 yılı sonrası iş makineleri teknolojisindeki yeniliklere paralel olarak makineli yol yapımına geçilmiştir. Yol yapım ve motorlu taşıt teknolojilerindeki muazzam gelişmeler sonucunda karayolu ulaşımı Türkiye'de giderek aynı paralelde gelişmelere sahne olmuş ve 1960'lı yıllarda karayolu ağı 60.000 km'ye ulaşmıştır (Korkmaz ve Alacahan, 2013: 6).

İkinci Dünya Savaşı'nın ardından Türkiye, ABD'den aldığı yardım ve direktifler doğrultusunda ulaşımında özellikle karayollarında büyük bir ilerleme sürecine girilmiştir. Bu dönem içinde demiryolu ulaşım sisteminde daha önceden uygulanan politika terk edilerek karayolu ulaşımı birinci plana çıkmıştır (Aytaç ve diğerleri, 2007: 138).

Günümüz 2017 yılı itibarıyla ise Türkiye karayolu ağı 67.620 km uzunluğunda olup 31.066 km devlet yolu, 33.897 km il yolu ve 2657 km otoyol

olarak bulunmaktadır. Bu yollarda kayıtlı bulunan araç sayısı 22.218.945 adettir. Bu konu için kullanılan bazı tanımlar Tablo 2’de verilmiştir (UDHB, 2017: 14-89):

Tablo 2: Türkiye’de Karayolu Ulaşımıyla İlgili Bazı Tanımlar

Ton-Km	Bir ton ağırlığındaki nesnenin bir kilometre uzaklığa taşınmasıyla elde edilen trafik ölçü birimidir.
Taşıt-Km	Herhangi bir motorlu kara taşıtının bir kilometre mesafede hareket etmesiyle elde edilen trafik ölçü birimidir.
Yolcu-Km	Bir yolcunun bir kilometre mesafeye taşınmasıyla elde edilen trafik ölçü birimidir.
Asfalt Beton (BSK)	Belli bir gradasyona sahip agrega ve filler ile bitümlü bağlayıcının belirli şartlarda karışımından meydana gelen yüksek niteliğe sahip bir sıcak karışımdır.
Sathi Kaplama	Temel Tabakası üzerine daha önceden belirlenmiş özellikteki bitümlü bağlayıcı ve agreganın birbiri peşi sıra serilip sıkıştırıldığı bir kaplama türüdür.
Stabilize Yol	Granüler alt temel tabakası kalınlığı 20 cm’ den az olmayacak şekilde ve granüler tabakası kalınlığı da 10 cm’ den az olmayacak şekilde inşa edilen yoldur.
Devlet Yolu	Belediye ve diğer kurumların sorumluluğunda bulunan yollar dışında kalan ve transit trafiği illerle, limanlara, tersanelere, havaalanlarına, demiryolu istasyonlarına ve sınır kapılarına kesintisiz olarak ulaştıran ana karayolu hattıdır.
İl Yolu	Belediyeler ve diğer kurumların sorumluluk sahasındaki yollar dışında kalan, illeri birbirine, devlet yollarını limanlara, tersanelere, havaalanlarına, demiryolu istasyonlarına ve kamu ihtiyacının gerektirdiği diğer yerlere bağlayan ve bir il sınırı içinde başlayıp bu sınırlar içinde biten karayoludur.
Otoyol	Yüksek standartlarda inşa edilmiş, trafik seyrinde asgari hız sınırlaması uygulanan, seyahat hızı yüksek ve üzerinde erişim kontrolünün uygulandığı karayoludur.
Bölünmüş Yol	Farklı yönlerde trafiğe sahip taşıt yollarının bir ayırıcıyla birbirlerinden ayrılmasıyla meydana gelen karayoludur.
Viadük	Açıklığı ve yüksekliği çok büyük olan vadiden, demiryolu ve karayolunun aynı hatta olacak şekilde geçişini sağlayan, yol ve kavşak kesişmelerini ortadan kaldırmak için yapılan köprüdür.

Kaynak: UDHB, 2017: 89

Denizyolu Ulaşımı: Denizlerde yolcu taşımacılığı ise özellikle ülkemiz bazında oldukça düşük bir paya sahiptir. Tablo 3’te görüleceği üzere yolcu taşımacılığında deniz ulaştırmasının payı sadece % 0,1 değerindedir. Diğer oranlara bakıldığında ülkemizin denizden yolcu taşımacılığının yok denecek kadar az olduğu görülmektedir.

Tablo 3: Bazı Ülkelerin 2001 Yılı için Yolcu Taşımacılığı Tercihleri (%)

Ülke	Karayolu	Demiryolu	Denizyolu	Havayolu
ABD	27,2	38,3	24	10,5
Almanya	58.2	22	12	7,3
Türkiye	96,2	2,0	0,1	1,7

Kaynak: Aras, 2012: 107

Ülkemizde nehir ve su yolu taşımacılığı akarsularımızın bu işe uygun olmaması sebebiyle yok denebilecek düzeydedir. Sadece Sakarya Nehri’nin belirli

bir bölümü üzerinde kereste taşımacılığı yapılmakta, fakat karayolu alternatifi nedeniyle bu faaliyet de minimum seviyede bulunmaktadır (Aras, 2012: 119).

Türkiye'nin İthalat ve İhracat için temel ulaşım türü olan denizyolu ulaşımında limanlarımızda 2017 yılı içinde toplam 471,15 milyon ton yük ve 10 milyon TEU biriminden fazla konteyner elleçlenmiştir. Bu rakamlar 2004 yılı için 189,9 milyon ton yük ve 2,49 milyon TEU büyüklüğünde konteynerde bulunmaktaydı. Denizyolu ulaşımı ile ilgili bazı tanımlar ise aşağıda belirtilmiştir (UDHB, 2017: 56-91):

Tablo 4: Denizyolu Ulaşımı ile İlgili Bazı Tanımlar

Liman Başkanlığı	Sorumluluk sahası sınırları çerçevesinde deniz ulaşımındaki idareden sorumlu birimdir.
Elleçleme	Bir yükün boşaltılma, depolama veya yüklenmesi işlemidir.
Kabotaj	Bir ülkenin kendi karasularında kendi bayrağını taşıyan araçlarla deniz taşımacılığı yapma hakkıdır.
Kurvaziyer Gemi	Bir veya daha fazla liman noktasında tarifeli geçici turistik ziyaret amaçlı bir grup programına katılıp yolcu taşıyan, yolcu veya yük alıp vermeyen veya tahliye etmeyen uluslar arası sefer yapan gemi türüdür.
Ro-Ro Gemi	Tekerlekli, kendi kendine hareket edebilen araçları taşıyan gemilerdir.
Yolcu-Mil	Bir yolcunun bir mil mesafeye taşınmasını ifade eden ölçü birimidir.
Ton-Mil	Bir ton yükün bir mil mesafeye taşınmasını ifade eden ölçü birimidir.
TEU	20'lik (feet) veya 34 m ³ lük hacme sahip konteynerdir.
GT	Bir gemide ölçü altı güverte ve yaşam ortamı hacimleri toplamının metrik sistemde bulunarak 2,83 m ³ bölünmesiyle elde edilen hacimsel birimdir.
DWT	Geminin yükleme, kaldırma taşıma kapasitesi ya da tuzlu suda yaz yükleme hattına kadar geminin 2.240 librelik ton olarak kaldırabileceği ağırlıkların tümüne denir.
Paris Memorandum	Gemilerin çevre, can-mal güvenliği, uluslar arası kural ve standartlara uyup uymadığını denetleyen organizasyondur.
Ölçek Ekonomisi	Bir malın üretiminin artırılarak maliyetlerinin düşürülmesine dayalı ekonomi kavramıdır. Üretimin artırılması da işletmelerin büyümesiyle doğru orantılıdır.

Kaynak: UDHB, 2017: 91

Demiryolu Ulaşımı: Ülkemizde demiryolları konusunda en büyük gelişme zamanları 1950 yılına kadar olan zaman dilimidir. Bu yıl itibariyle Türkiye'deki demiryolu uzunluğu toplamı cumhuriyet dönemi öncesi olan 4136 km, 1923-1950 arası dönem de 3764 km olmak üzere 7900 km'ye ulaşmıştır. Ülkemizde 1950 yılında T.C. Karayolları Genel Müdürlüğü'nün kurulmasıyla başlayan, izlenen politikalar sonucu elde edilen Marshall yardımları ve ilerleyen yıllarda ivme kazanan karayollarındaki gelişmeler, zaten savaş sebebiyle yavaşlamış durumda olan demiryolu hat inşasını daha da yavaşlatmış ve son 50 yılda demiryoluyla yük taşımacılığını %60'lara kadar geriletmiştir (Müdürlüğü, T. İ. G., 2016: 3). Günümüz demiryolu ulaşım hattı genel görünüm itibarıyla Şekil 1'de görüldüğü gibi tüm illerimizde olmasa da belli başlı bölgelerden geçmektedir.

Şekil 1: Türkiye Demiryolları Haritası (2018)



Kaynak: <https://rayhaber.com/2019/07/tcdd-demiryollari-haritasi-2019-yuksek-cozunurluk/>
(9.11.2020)

Günümüzde ise ülkemiz demiryolu ulaşımı bazında yatırımlarda belirgin bir artış olsa da uluslararası standartlar açısından geride bulunmaktadır. 2017 yılı itibariyle ülkemiz demiryolu uzunluğu toplamı 12.608 km'dir ve karayolundan sonra 71.7 milyar TL yatırımla ikinci en çok yatırım yapılan ulaştırma türü konumundadır (UDHB, 2017: 8).

Bu çalışmada, Türkiye Cumhuriyeti Devlet Demiryolları (T.C.D.D.) raporlarında belli zamana kadar hatlar bazında değerlendirme göz önüne alındığı için yolcu ve yük taşıma işlemlerine ait bilgiler genel olarak iller arasında ve bölge çaplı gösterilmiştir. Bu sebepten dolayı veriler iller bazında değerlendirilmeden sadece demiryolu hat uzunlukları bilgileri iller bazındaki değerlendirmelerde dikkate alınmıştır.

Havayolu Ulaşımı: Ülkemizde ilk havaalanı 1912 yılında Yeşilköy'de açılan Atatürk Havaalanı'nın da temelini oluşturan Yeşilköy havaalanıdır. Daha sonra bu havaalanının ardından yeni havaalanları inşa edilerek 1947 yılına kadar 17 tane inşa edilmiştir. 2018 yılı itibariyle bu sayı 56'ya ulaşmış durumdadır. Ayrıca Türkiye geneli havayolu ulaşımında 1.256.224 tonluk kargo taşıması, 193.576.844 kişilik yolcu taşıması ve 1.914.017 adetlik uçak trafiği gerçekleşmiştir (UDHB, 2017: 41-53). Günümüz Türkiye'de 2018 yılındaki mevcut havaalanlarının konumları kabaca Şekil 2'de gösterilmiştir. Şekle nicelik olarak bakıldığında havaalanlarının İstanbul ilinde ve Güneydoğu Anadolu'da yoğunlaştığı fark edilebilir. Şekilde görüldüğü üzere Doğu Karadeniz'de sadece Trabzon'da havaalanı bulunmaktadır.

Şekil 2:Türkiye’de Mevcut Havaalanları Haritası (2018)



Kaynak: <https://cozumpedia.com/turkiyedeki-havalimanlari-listesi> (9.11.2020)

Boru Hattı Ulaşımı: Türkiye Dünya’nın en büyük ham petrol ve doğal gaz rezervlerinin bulunduğu Ortadoğu ve Orta Asya ülkeleri ile Avrupa’nın ham petrol ve doğal gaz ithal eden ülkeleri arasında köprü konumunda olması dolayısıyla önemli bir jeopolitik konumda bulunmaktadır. Bu konumun yarattığı avantaj ile kolayca elde edebildiği bu ürünleri en büyük sanayi kuruluşu olan TÜPRAŞ ile petrolün tedariki, rafine etmesi, ithalat ve ihracat amaçlı çeşitli sistemlerin inşa edilmesi ve işletimi gibi işlemler gerçekleştirilmektedir. Doğalgaz tedariki ise yerli üretimin düşüklüğü sebebiyle büyük oranda ithalat ile sağlanmaktadır. Türkiye Petrolleri Anonim Ortaklığı (TPAO) tarafından üretilen doğal gaz genelde Boru Hatları ile Petrol Taşıma A.Ş.’nin (BOTAŞ) ana iletim hattına verilerek Rusya’dan ithal edilen doğal gaz ile birlikte kullanıma arz edilmektedir. Ayrıca kısmen de TPAO tarafından kurumun doğrudan satış yaptığı sanayi müşterilerine verilmektedir (Bakanlığı, 2005: 7.5.2).

Türkiye’de 2017 yılı itibariyle doğal gaz hatları ulaşmayan Artvin, Şırnak ve Hakkâri illeri bulunmaktadır. Aynı yıl ayrıca TÜİK 2017 verilerine göre toplam sekiz adet 3053 km uzunluğundaki petrol boru hattından 52.095.254 ton petrol aktarımı yapılmıştır.

İKİNCİ BÖLÜM

KÜMELEME ANALİZİ

2.1. KÜMELEME ANALİZİ

Tarih boyunca insanlar karşılaştıkları gözlem ve objeleri grupta ve bu sayısı, biçimi ve özellikleri gibi kavramlarının belirlenmesine her dönemde ve hemen her alanda ihtiyaç duymuştur. Yakın dönemde kullanılmaya başlanan kümeleme analizi kavramı da bu ihtiyaç dolayısıyla ortaya çıkmıştır.

Kümeleme Analizi, araştırmacıların objeleri (gözlemleri) bir dizi özelliklere göre sınıflayarak birtakım kümeler elde edilmesini sağlamaktadır. Bu sınıflama işlemi sonunda elde edilen kümelerde, kümelerin sahip olduğu gözlemlerde yüksek homojenlik, bu kümelerin arasında ise yüksek heterojenlikte grupların elde edilmesi beklenmektedir (Hair ve diğerleri: 2014: 418).

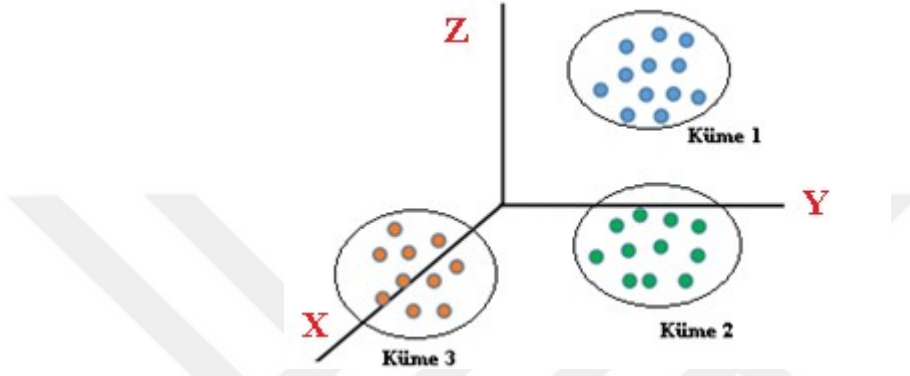
Bir veri setinin en doğru şekilde gruplanması için çok sayıda seçenektan en uygun olanını bulmak oldukça zordur. Nitekim 16 birimlik bir gözlem grubunu bile iki farklı biçimde gruplamanın 32.767 adet farklı yolu bulunmaktadır ve bu sayı 3 farklı gruplamada 7.141.686 olası farklı yola dönüşmektedir. Zaman ve işlem kısıtları göz önüne alındığında bu durum, günümüzün güçlü bilgisayarlarının bile kolayca başa çıkabileceği bir problem değildir. Bu sebeple mümkün olan en kısa zamanda imkânlar dâhilinde uygun çözüme ulaştıracak algoritmalara ihtiyaç duyulmaktadır (Johnson ve Wichern, 2007: 672).

Aynı zamanda büyük miktarda verinin işlenerek işe yarayacak bilgilerin ortaya çıkarılmasını amaçlayan veri madenciliği alanında bir uygulama alanıdır (Berkhin ve diğerleri, 2006: 3). Çok değişkenli istatistik yöntemlerinden birisi olan kümeleme analizi kavramı, çok büyük ve karmaşık yapıdaki bir veri setinde birbirine mümkün olduğunca benzeyen gözlemlerin bir arada gruplandırılarak uygun sayıda kümelerde karşılaştırılması amacıyla kullanılan yöntemlerdir. Bu yöntemler, verilerin uzaklık ve benzerliklerine göre kümelere ayrılarak, bu ayrılmaların sebeplerinin yorumlanmasına dayanır (Ayrıca bu kümelerin aralarında heterojen olması istenir).

Kümeleme analizindeki temel amaç, objelerin doğru şekildeki gruplarını keşfetmektir. Buna karşılık, önce objeler arasındaki ilişkiyi (uzaklık veya benzerliği) ölçmek için nicel bir ölçek geliştirilmelidir. Doğru bir gruplandırmanın kesin bir ölçütü olmasa bile, objeler iki veya üç boyutlu düzlemlerde göz kararı gruplanabilir

(Şekil 3). Radar(yıldız) grafikleri ve Chernoff yüzleri bu amaç için kullanılabilecek yöntemlerdendir. En büyük ve en hızlı bilgisayarlarda bile, tüm olası gruplama seçeneklerine nadiren ulaşabiliriz. Bu probleme karşı, tüm olası gruplara bakmak zorunda kalmadan “makul” kümeleri bulan çok çeşitli kümeleme algoritmaları ortaya çıkmıştır (Johnson ve Wichern, 2007: 672-680). Kümeleme analizi yöntemlerinin amacının şekilsel olarak ifadesi Şekil 3’te görüldüğü gibi ifade edilebilir.

Şekil 3: Üç Boyutlu Koordinat Düzleminde Bir Kümeleme Örneği



Kaynak: Özeri, 2019: 41

Kümeleme analizi, bir örneklemden elde edilen parametrelerin bir kitleyi temsil ettiği şeklinde değerlendirildiği istatistiksel bir çıkarsama tekniği değildir. Bunun yerine, bir dizi gözlemin yapısal özelliklerini ölçmek için kullanılan bir yöntemdir. Bu sebeple güçlü matematiksel özellikleri olmasına rağmen istatistiksel temelleri yoktur. Diğer tekniklerde çok önemli olan normallik, lineerlik ve eş varyanslılık konusundaki gereklilikler kümeleme analizinde çok az etkiye sahiptir (Hair ve diğerleri, 2014: 436).

Kümeleme analizinin kullanıldığı alanlar yakından incelendiğinde, neredeyse tüm alanlarda yer alabildiği görülmektedir. Fizik, kimya, biyoloji, tıp, psikoloji, sosyoloji, arkeoloji, tarih, çeşitli mühendislik alanları gibi karmaşık yapıların ve belirsizlik koşullarının yer aldığı bilim alanlarında oldukça yaygın bir yöntem olduğu görülebilir. Örneğin (Özdamar, 2004: 281),

-Tıp alanında; hastalık ya da bulgulara göre tanı ve tedavi yöntemlerinin gruplandırılması.

-Psikiyatride; şizofreni, paranoya gibi belirtilerin sınıflandırılması, koyulacak teşhisler için görüntü oluşturulması, laboratuvar bulguları ile klinik bulgulardan ortaya

ıkan veri matrislerinden hastalık alt sınıflandırmalarının veya yeni hastalık tablolarının tanımlanabilmesi.

-Arkeoloji alanında; kazılardan toplanan bulguların ve tarih ncesi canlıların trlerinin sınıflandırılması.

-Tarih biliminde eşitli uygarlıkların kullandıkları dil ve yazıların incelenerek birbirleriyle etkileşime girmiş toplulukların sınıflandırılması amacıyla kullanılmaktadır.

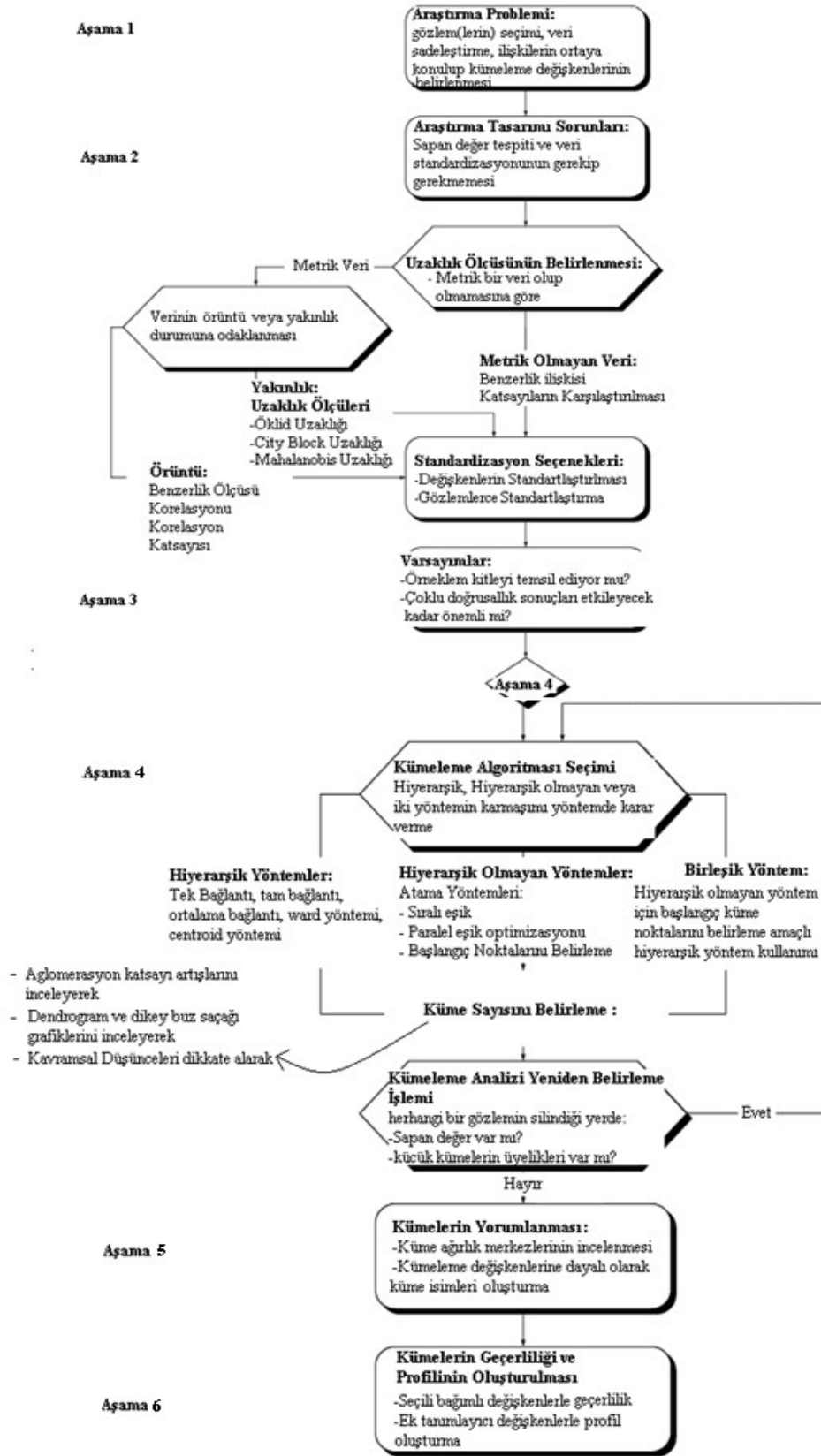
ok sayıda olgunun söz konusu olduėu her araştırmada kmeleme analizi yer alabilmektedir. 1753 yılında Carl Linnaeus'un hayvan ve bitkileri sınıflandırmasında kmeleme analizi alışmasına rastlanabilir. Bu konuda rnek olabilecek eski uygulamalardan birisi de İngiltere'de 1854 yılında zamanın nl doktoru John Snow Şekil 4'teki Londra haritası zerinde kolera kaynaklı lmlerin meydana geldiėi noktaları işaretlemeye başlamış, belli bir blgede bu vakaların oėalarak yoėun bir kme oluştunuunu fark etmiştir. Bu blgeyi incelediėinde bir tulumba sisteminin bulunduėu sokak etrafında hastalığın yoėunlaştığını gözlemlemiştir. Söz konusu tulumba kaldırıldığında her ne kadar tulumbada hastalıkla ilgili bir ize rastlanmasa da bu işlemden sonra salgının n kesilmiştir (Brody ve diėerleri, 2000: 68). İlk kez 1939 yılında Robert Tryon tarafından kullanılan kmeleme analizi daha sonra 1960'larda yaygınlaşarak yeni yntemler geliştirilmiştir. Sayısal olarak sınıflandırmada konu olarak yaşıyan organizmaların gruplanması, personel ve kişilik zellik tabanlı sınıflandırma, piyasa araştırması segmentasyonu arasında deėişebilen kmeleme analizi uygulamaları byk lde bireyleri gruplama zerine odaklanmaktadır. Bununla birlikte kmeleme analizinin birok farklı şekilde kullanım alanları vardır. Bu alanlar veri "gzetleme", tabakalı bir rnekleme şemasının ilk aşamasını oluşturmak iin benzer objelerin "doėal" gruplarını arama, hipotez oluşturma, uygun bir sınıflandırma arama ve veri indirgeme gibi ok farklı etkinlikleri kapsar. Farklı hedefler farklı tekniklerin kullanımını gerektirir ve sıradan bir gzlemci iin ok farklı eşitte uygulamalar olduėu kadar da bu konuda ok farklı kmeleme tekniėinin var olduėu sylenebilir (Seber, 2009: 347).

Kmeleme analizi altı aşamalı bir model yaklaşıımı zerinden gsterilebilir (Şekil 5). Bu model Anderberg (1973), Everitt (1980), Cormack (1971) gibi bu yntemler stnde alışan birok araştırmacının belirttiėi kmeleme analizi srelerinden ok farklı deėildir (Milligan ve Cooper, 1987: 331). Keşfedici ya da

- Her kümenin özelliklerini açıklamanın yanı sıra, demografik bilgiler gibi belirlenen amaca uygun boyutlarda nasıl farklı olabileceklerinin açıklanması.

Kümeleme analizinin amacı daha önce açıklandığı üzere gözlemleri kümeler halinde her değişkenlerine göre olabildiğince homojen olacak gruplamaktır. Kümeleme analizinde ilk adım, uygulamaya hazır veri seti için bir benzerlik veya korelasyon ölçüsü seçmektir. Daha sonra, kullanılacak kümeleme tekniğinin türüne karar verilir (Örneğin, hiyerarşik veya hiyerarşik olmayan). Üçüncüsü, seçilen teknik için kümeleme yönteminin türü seçilmesidir (Örneğin, hiyerarşik kümeleme tekniğinde ağırlık merkezi yöntemi). Dördüncü olarak, küme sayısına ilişkin bir karar verilmesidir. Son olarak, küme çözümünün yorumlanmasıdır (Sharma, 1996: 187). Şekil 5’te Hair, Black, Babin ve Anderson’a göre ise bu durum altı aşamada tanımlanmış ilk aşamaya araştırma probleminin belirlenmesi ve son aşamaya küme geçerliliğinin belirlenmesi eklenmiştir.

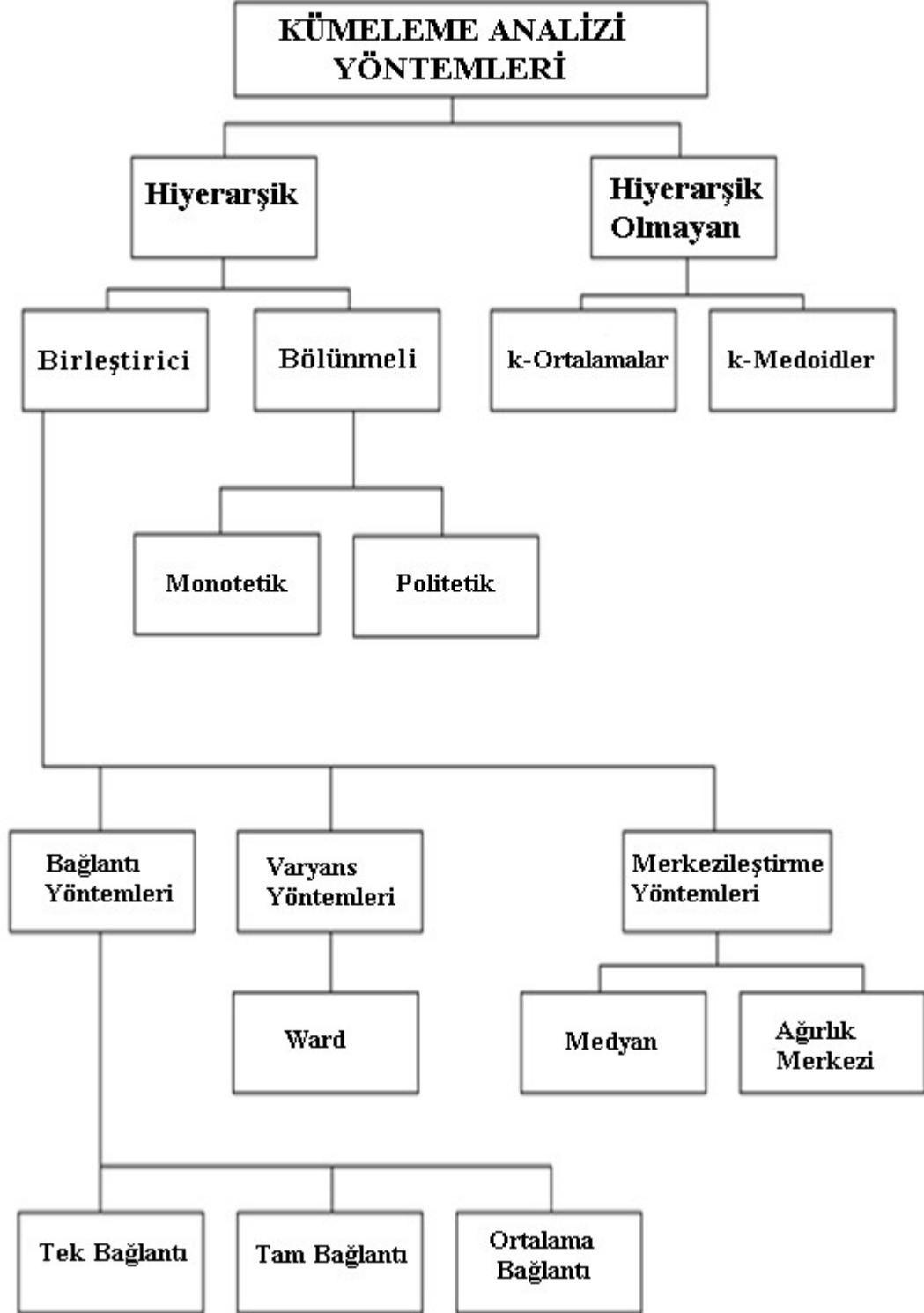
Şekil 5: Kümeleme Analizi Karar Diyagramı



Kaynak: Hair ve diğerleri, 2009: 426-438

Klasik kümeleme analizi yöntemleri genel olarak hiyerarşik (aşamalı) ve hiyerarşik (aşamalı) olmayan yöntemler olarak iki temel gruba ayrılmaktadır. Bu kümeleme yöntemlerin genel yapısı Şekil 6’da verilmiştir.

Şekil 6: Kümeleme Analizi Yöntemlerinin Genel Yapısı



Kaynak: Grabmeier ve Rudolph, 2002: 340

Klasik anlayışa dayalı bu yöntemler dışında günümüz şartlarına daha uygun analizler üstünde çalışılmaktadır. Kümeleme literatüründe bu yöntemler veri yapılarına göre tanımlanabilir. Bulanık kümeleme, soft set kümeleme, gri kümeleme, rough kümeleme gibi farklı yaklaşımlar bunlardan bazılarıdır (Erilli ve Gündoğdu, 2019: 6).

Klasik anlayışa dayalı kümeleme analizi yöntemleri, herhangi bir obje veya gözlemin bir kümede varlığı veya yokluğuna göre şekil alır. Bir başka deyişle bir obje veya gözlemin üyelik gücü sıfır veya bir olmak üzere iki değerlidir. Bir bulanık kümeleme analizi yönteminde, alt kümelerin sayısının bilindiği varsayılır ve her bir kümedeki nesnenin üyelik işlevi, genellikle bir sezgisel hedef fonksiyonuna dayalı standart bir optimizasyon tekniği olan yinelemeli bir yöntem kullanılarak tahmin edilir. Genel olarak, üyelik fonksiyonları olasılık teorisinin kurallarına uymaz, ancak bir kez bulunduklarında üyelikler sıfır ile bir arasında olacak şekilde ölçeklenebilir ve ardından olasılıklar olarak yorumlanabilir. Klasik anlayışta olduğu gibi sadece iki değerden ibaret olmaz. Bu esneklik sayesinde bulanık kümeleme yöntemleri klasik yöntemlere göre iki büyük avantaja sahiptir. Öncelikle üyelikler diğer bilgilerle birleştirilebilir. Özellikle üyeliklerin olasılıklar şeklinde olduğu özel durumda sonuçlar bayes teoremi kullanılarak farklı kaynaklardan birleştirilebilir. İkinci avantajı ise, herhangi bir obje veya gözlem üyelikleri, diğer kümeleme teknikleri kullanıldığında genellikle saklı kalan bir fenomen olan, "en iyi" küme kadar iyi olan "ikinci bir en iyi" kümenin olup olmadığını ortaya çıkarmasıdır (Everitt ve diğerleri, 2011: 242).

Bulanık kümeleme analizi kavramı 1960'lı yılların ortalarında Loutfi A. Zadeh tarafından ortaya atılmıştır. Bu kavramın temeli gerçek hayat problemlerinin içerdiği kesin olmayan ve belirsiz verilerdir (İncekırık, 2014: 38). İlerleyen yıllarda bulanık c-ortalamlar yöntemiyle 1973 yılında Dunn, 1974 yılında Bezdek gibi araştırmacıların katkısıyla bu yöntem giderek çeşitlenmiştir. c-ortalamlar yönteminin yanı sıra Gath-Geva ve Gustafson-Kessel gibi yöntemler bulanık kümeleme konusunda bilinen ve sık kullanılan yöntemler arasındadır (Erilli, 2019: 157-165). Bulanık kümeleme analizi kavramına bağlı yöntemlerin sunduğu imkânlar günümüz koşullarında hiyerarşik ve hiyerarşik olmayan kümeleme analizi yöntemlerine göre daha ön planda olmasını sağlamıştır.

2.2. VERİLERİN STANDARTLAŞTIRILMASI

Farklı ölçme düzeyleri (aralıklı, oran, sınıflama, sıralama), uzunluk, ağırlık ölçüleri gibi farklı değişkenlerin bir arada bulunduğu ve büyük varyans farklılıklarının söz konusu olduğu veri setlerinde standartlaştırma gündeme gelmektedir. Bir değişkenin küme çözümü üzerindeki etkisinin olması gerekenden daha büyük olduğu durumlardan kaçınmak için gerektiğinde aynı ölçekte olmayan kümeleme değişkenleri standartlaştırılmalıdır. Standartlaştırma yapılmasındaki temel amaç analizde kullanılan çeşitli özelliklerin veya değişkenlerin ölçeklerindeki farklılıkların oluşturduğu yanlılığı ortadan kaldırarak kümeleme analizini daha doğru sonuçlara ulaştırmaktır (Grimm ve Yarnold: 1995: 172).

Kümeleme analizinde kullanılacak veri matrisindeki sütunların ölçme düzeyleri ve varyans değerlerinde farklılıklar gözlemlendiğinde uzaklık değerleri hesaplanmadan önce verinin standartlaştırılması önerilmektedir (Bulut, 2018: 368).

Standartlaştırılmış değişkenleri kullanılması, yalnızca değişkenler arasında değil, aynı değişken için de ölçek farklılıklarından kaynaklanan etkileri ortadan kaldırmaktır. Değişkenlerin tümü aynı yanıt ölçeğinde (örneğin, bir dizi davranışsal yapı sorusu gibi) ölçüldüğünde standartlaştırma ihtiyacı asgariye indirilir. Ancak, oldukça farklı ölçme düzeyi kullanan değişkenler küme değişkenine dâhil edildiğinde bu süreç oldukça önemli hale gelir. Standartlaştırma işlemi, bazı veri setlerinde tüm değişkenlerin aynı ölçüm düzeyinde ölçülebilmesi sebebiyle gerekli olmayabilir (Hair ve diğerleri: 2014: 435).

Literatürde en çok kullanılan z-standartlaştırmasının yanı sıra min-maks yöntemi gibi çeşitli standartlaştırma biçimleri mevcuttur. Özellikle IBM'in yapımını gerçekleştirdiği SPSS programlarındaki mevcudiyeti sebebiyle pozitif ve negatif verilerin bir arada bulunduğu heterojen yapıda aykırı değerlerin bulunduğu veri setlerinde kullanılan bir $-1 < 0 < 1$ aralığında standartlaştırma, aritmetik ortalaması veya standart sapması her değişkende 1 olacak şekilde standartlaştırma biçimleri de yaygın şekilde kullanılmaktadır (George ve Mallery, 2019: 278). Ancak burada standartlaştırma konusunda yaygın olarak kullanılan iki yöntem anlatılmıştır.

Z-Standardlaştırması: Literatürde en sık karşılaşılan veri standartlaştırma yöntemidir. Normal dağılım gösterdiği varsayılan oransal ya da aralıklı ölççeklerle ölçülmüş verilere uygulanır. Elde edilen Z skorlarının beklenen değeri 0, standart sapması 1'dir (Alpar, 2017: 84). Z aralığı negatif ile pozitif arasında değerler alabilmekte olup negatif değerler ortalamanın altında, pozitif değerler ise ortalamanın üstündedir. Dönüştürmede aşağıdaki formülden yararlanılmaktadır (Seber, 2009: 164):

$$Z_i = \frac{x_i - \bar{x}}{S} \quad (2.1)$$

0<x<1 Aralığında Min-Maks Normalizasyon Yöntemi: Mevcut verileri en düşük değeri 0, en yüksek değeri 1 olmak üzere pozitif değerlere dönüştürmek amacıyla kullanılmaktadır. Formülü aşağıdaki biçimde gösterilir. Formülden de anlaşılacağı üzere uygulanması son derece kolaydır. Bir gözlem değeriyle o değişkendeki en küçük gözlem değerinin farkının yine o değişkendeki maksimum ve minimum gözlem değerleri arasındaki farka bölümüyle elde edilir. 0-1 aralığı dışında da farklı varyasyonlar kullanılarak bu formüle bağlı olarak veriler üzerinde standartlaştırma yapmak da mümkündür. Diğer birçok yöntemle göre daha dayanıklı bir standartlaştırma yöntemi olduğuyla ilgili bazı çalışma sonuçları elde edilmiştir (Markov ve diğerleri, 2012: 1162). Formülü şu şekilde gösterilebilir:

$$Z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (2.2)$$

2.3. UZAKLIK ÖLÇÜLERİ

Kümeleme analizinde gözlemler arasında kullanılabilecek benzerlik ölçüleri korelasyon, uzaklık ve ortaklık ölçüleri olmak üzere üç ana biçimde incelenebilir. Ölçülerin kullanımı, veri setindeki değişkenlerin tiplerine bağlı olarak değişmektedir. Uzaklık ölçüleri ve korelasyon ölçüleri metrik (aralıklı, oransal ölçeğe sahip değişkenler) verilerde kullanılırken, ortaklık ölçüleri de kategorik veriler için uygundur. Korelasyon ölçüsünde ortaya çıkacak yüksek bir değer, gözlem çifti arasındaki benzerliğin yüksekliğine işaret eder (Kalaycı, 2014: 354-355). Uzaklık ölçülerinde ise durum farklıdır. Yüksek bir değer iki gözlem arasında düşük benzerliğe işaret etmektedir. Burada kullanılan ölçülerin alınmasındaki temel amaç

ise gözlemler arasındaki ilişkilerin sayısal anlamda incelenerek birbirlerine ne kadar benzediğinin uzaklık ölçüleri yardımıyla ortaya çıkarılmasıdır.

Karma yapıda bir veri setinden daha basit bir grup yapısı oluşturma işlemlerinin çoğu, “yakınlık” veya “benzerlik” ölçüsünü gerektirir. Bir benzerlik ölçüsü seçiminde çoğu zaman araştırmacı temelli bir öznellik söz konusudur. Bu durumda göz önünde tutulması gerekenler kısmında önemli değişkenler (kesikli, sürekli, ikili), ölçüm çeşitleri (sınıflama, sıralama, aralık, oran) ve konu bilgisi bulunmaktadır (Johnson ve Wichern, 2007: 673).

Kümeleme analizinde en sık kullanılan ölçü biçimi uzaklık ölçüsüdür. Değişkenler arasında gözlemlerin birbirine yakınlığı ile benzerliğini temsil ederler. Üç farklı benzerlik ölçütü formu mevcut olmasına rağmen, bu durumun bazı sebepleri bulunmaktadır. Bunun sebeplerinden ilki, uzaklık ölçüsünün kümeleme analizinde temel olan yakınlık kavramını en iyi şekilde temsil etmesidir. Korelasyon ölçüleri, diğer tekniklerde yaygın olarak kullanılmasına rağmen, yakınlıktan ziyade bir veri setinden elde edilen belirli model kalıplarını temsil eder. İkinci sebep, kümeleme analizinin tipik olarak metrik değişkenlerle ölçülen özelliklerle ilişkilendirilmesidir. Bazı uygulamalarda metrik olmayan karakteristikler hâkimdir, ancak çoğu zaman bu karakteristikler sık tercih edilen ölçü uzaklığını oluşturan metrik ölçülerle temsil edilirler. Bu nedenle, herhangi bir durumda araştırmacıya, bir dizi metrik veya metrik olmayan değişken arasındaki objelerin yakınlığını temsil edebilecek benzerlik ölçüleri geliştirilmiştir (Hair ve diğerleri, 2014: 431-434).

Sonuç olarak, hangi ölçünün bir anlamda “en uygun” olduğunu bilmek son derece yararlı olacaktır. Ancak ne yazık ki bu sorun 1966 yılında Williams, Lambert, Lance ile 1969 yılında Cheetham ve Hazel gibi araştırmacıların ilgisini çekmiş olsa da hala mutlak anlamda cevaplanamaz ve ölçüm seçimi büyük ölçüde kullanılan değişkenlerin türü ve araştırmacının inisiyatifine bağlı kalır (Everitt ve diğerleri, 2011: 69). Literatürde sıkça bilinen ve kullanılan uzaklıklardan bazıları aşağıda tanıtılmıştır.

2.3.1. Öklid Uzaklığı

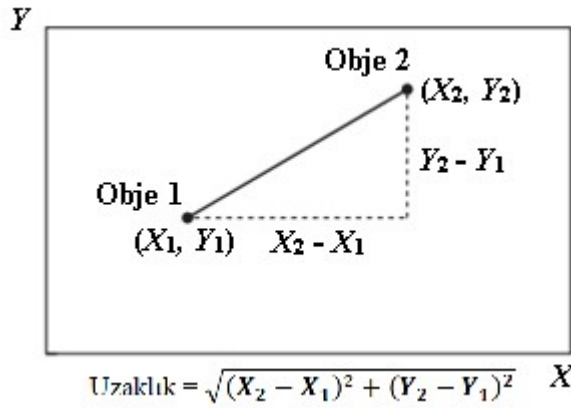
Çoğu zaman vektör uzaklığı olarak da adlandırılan, en bilinen uzaklık ölçüsüdür. Öklid uzaklığının nasıl elde edildiğinin bir örneği, Şekil 7’de geometrik olarak gösterilmektedir. İki boyuttaki iki noktanın sırasıyla (X_1, Y_1) ve (X_2, Y_2)

koordinatlarına sahip olduğunu varsayalım. Noktalar arasındaki Öklid uzaklığı, şeklin altındaki formülle hesaplandığı dik üçgenin hipotenüsünün uzunluğudur. Bu kavram ikiden fazla değişken olması durumunda da kolayca genelleştirilebilir (Hair ve diğerleri, 2014: 431). Genel formülü aşağıdaki gibi ifade edilir:

$$d(x_i, x_k) = \sqrt{\sum_{j=0}^n w_j (x_{ij} - x_{kj})^2} \quad (2.3)$$

Burada kullanılan w_j ifadesi öklid uzaklığın standartlaştırılmamış şekliyle 1 değerini, S^2 değeri j . değişkenin varyansı olmak üzere $w_j = 1/S_j^2$ olarak Karl-Pearson uzaklığına dönüşümünü veya standartlaştırılmış $1/R_j^2$ aralığını ifade edecek şekilde üç formda bulunabilir (Mardia ve diğerleri, 1995: 381).

Şekil 7: Ölçülen İki Obje Arasındaki Öklid Uzaklık Örneği



Kaynak: Grimm ve Yarnold, 1995: 166

Literatürde en çok kullanılan ve birçok çalışmada standart olarak kullanılıp hemen tüm çok değişkenli analiz yazılımlarında bulunan bir ölçüdür. Öklid uzaklıkta en yakın değer 0 olarak ifade edilir. Bu hesaplamada verilerin birbirinden bağımsız olduğu varsayılır.

İki adet p boyutlu gözlem (obje) arasındaki Öklid (doğrusal) Uzaklığı formülü şu şekilde hesaplanır (Johnson ve Wichern, 2007: 673):

$$\begin{aligned} x' &= [X_1, X_2, \dots, X_p] \text{ ve } y' = [Y_1, Y_2, \dots, Y_p] \text{ iken} \\ d(x, y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \\ &= \sqrt{(x - y)'(x - y)} \text{ olur.} \end{aligned}$$

Aynı iki gözlem arasındaki istatistiksel uzaklık (mahalanobis) ise şu şekilde gösterilir:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{A} (\mathbf{x} - \mathbf{y})} \quad (2.4)$$

Formüldeki $\mathbf{A}$ ifadesi $\mathbf{A} = \mathbf{S}^{-1}$ 'e eşit olup bu eşitlikte bulunan $\mathbf{S}$, örnek varyans ve kovaryanslarını içemektedir. Bununla birlikte, farklı gruplar hakkında ön bilgi sahibi olmaksızın bu ifadedeki miktarlar hesaplanamaz (Johnson ve Wichern, 2007: 673). Ayrıca gözlemler arası ilişkinin benzerlik ve dağılımları en fazla öklid uzaklıkta göze çarpar. Bu sebeple kümeleme için genellikle öklid uzaklığı tercih edilir. Öklid uzaklığı hesaplanırken değişkenlerin ölçek farklılıkları önemli ölçüde etkilediğinden bu uzaklığın hesaplanmasında birbirleriyle karşılaştırılabilir aynı ölçekte ölçülmüş olması önemlidir (Sharma, 1996: 168).

2.3.2. Karesi Alınmış Öklid Uzaklığı

Karekökünü almadan karesel farklılıklarının toplamıdır. Karesi alınmış öklid uzaklığı, hesaplamaları belirgin bir şekilde hızlandırmayı sağlayan kareköksüz hesaplama avantajına sahiptir. Ağırlık merkezi ve Ward kümeleme yöntemleri için önerilen uzaklık ölçüsüdür (Hair ve diğerleri, 2014: 431). Karekökü alınmadığı için de özellikle normal öklid uzaklığına göre aykırı gözlemlere karşı daha hassastır (Çilingirtürk, 2011: 168). Formülü şu şekilde gösterilebilir.

$$d(x_i, x_k) = \sum_{j=0}^n (x_{ij} - x_{kj})^2 \quad (2.5)$$

2.3.3. Manhattan (City-Block) Uzaklığı

Öklid uzaklığına dayalı değildir. Bunun yerine, değişkenlerin mutlak farklılıklarının toplamını kullanır (yani, hipotenüs yerine sağ üçgenin iki tarafı). Bu prosedür hesaplaması basittir ancak kümeleme değişkenleri arasında yüksek bir korelasyon varsa geçersiz kümeler neden olabilir (Hair ve diğerleri, 2014: 432). Daha çok kesikli nicel verilere sahip değişkenler için kullanılması önerilmektedir. Hesaplanması şu şekildedir (Alpar, 2017: 172):

$$d(x_i, x_k) = \sum_{j=0}^n |x_{ij} - x_{kj}| \quad (2.6)$$

2.3.4. Mahalanobis Uzaklığı

Değişkenler arasındaki korelasyonu dikkate almak üzere tasarlanmıştır ve aynı zamanda ölçek sabittir. İlişkisiz değişkenler için Mahalanobis uzaklığı, standartlaştırılmamış veriler için öklid uzaklığına dönüşür. Yani standartlaştırılmış veriler için öklid uzaklığı, Mahalanobis uzaklığının özel bir formudur. Değişkenler birbiriyle ilişkilendirildiğinde (pozitif veya negatif), Mahalanobis uzaklık ölçüsü en uygun olanıdır. Standartlaştırılmış değişkenlere dayanır. SPSS ve SAS gibi önemli programlarda bu ölçü mevcut değildir. Formülü şu şekilde gösterilir (Sharma, 1996: 220):

$$d(x_i, x_k) = (x_i - x_k)' S^{-1} (x_i - x_k) \quad (2.7)$$

2.3.5. Minkowski Uzaklığı

$m = 1$ durumu için burada $d(x_i, x_k)$ p boyutlarında iki nokta arasındaki "Manhattan" uzaklığını ölçer. $m = 2$ olması durumunda ise öklid uzaklığına dönüşür. Genel olarak burada m 'nin değerinin değişimine göre daha küçük ve daha büyük farklar meydana getirir. Formülü şu şekildedir (Everitt ve diğerleri, 2011: 50):

$$d(x_i, x_k) = \left[\sum_{j=1}^p |x_{ij} - x_{kj}|^m \right]^{1/m} \quad m \geq 1 \quad (2.8)$$

2.3.6. Canberra Uzaklığı

Pozitif tanımlı gözlemler için uygundur. Formülü şu şekildedir:

$$d(x_i, x_k) = \sum_{i=1}^p \frac{|x_i - x_k|}{|x_i + x_k|} \quad (2.9)$$

p burada boyut sayısını, x_i ve x_k ise iki farklı gözlem değerini ifade eder. Gürültülü ve karmaşık yapıdaki verilerde uygun değildir (Emran ve Ye, 2001, 84).

2.3.7. Chebychev Uzaklığı

Maximum değer uzaklığı olarak da adlandırılabilir. Aşağıdaki formülden de anlaşılacağı üzere bir çift gözlem arasındaki farkın mutlak değerinin maksimumu olarak tanımlanan bir uzaklık ölçüsüdür. Formülü de aşağıdaki gibi ifade edilebilir (Irani ve diğerleri, 2016: 10).

$$d(x_i, x_k) = \max |x_{ij} - x_{kj}| \quad (2.10)$$

2.3.8. Hotelling T² Uzaklığı

Normal dağılımlı iki grup ya da kümenin ortalama vektörlerinin karşılaştırılmasında kullanılabilen Hotelling T² değeri aynı zamanda bir uzaklık ölçüsü olarak kullanılmaktadır (Johnson, 1998: 329). Hotelling T² uzaklığının hesaplanmasında aşağıdaki formülden yararlanılır (Sapp ve diğerleri, 2007: 68):

$$d(x_i, x_k) = \frac{n_1 n_2}{n_1 + n_2} (x_i - x_k)' S^{-1} (x_i - x_k) \quad (2.11)$$

2.4. HİYERARŞİK KÜMELEME ANALİZİ YÖNTEMLERİ

Hiyerarşik kümeleme teknikleri, birbirini takip eden bir seri ile devam eden bir dizi ardışık birleşme veya bölünme ile ilerler. Birleştirici (aglomeratif) hiyerarşik yöntemler bireysel objelerle başlar. Böylece, başlangıçta objeler kadar çok küme vardır. Çoğu benzer obje ilk önce gruplandırılır ve bu ilk gruplar daha sonra benzerliklerine göre birleştirilir. En sonunda, benzerlik azaldıkça, tüm alt gruplar tek bir kümeye kaynaşır. Birleştirici hiyerarşik yöntemler ters yönde çalışır. Başlangıçta bir tek obje grubu, bir alt gruptaki objelerin diğerindeki objelerden "uzak" olacak şekilde iki alt gruba ayrılır. Bu alt gruplar daha sonra farklı alt gruplara ayrılır; işlem, objeler alt grup olana kadar, yani bütün objeler bir grup oluşturana kadar devam eder (Johnson ve Wichern, 2007: 680). Bölünmeli (ayırıcı) yöntemler ise birleştirici (aglomeratif) yöntemlerin tam tersidir. Tüm gözlemlerden oluşan tek bir küme ile başlanarak her gözlemin tek bir küme oluşturana kadar işleme devam edilir. Bölünmeli yöntemler ise genel olarak monotetik ve politetik yöntemler olmak üzere ikiye ayrılır. Monotetik yöntemlerde her adımda tek bir ideal belirleyici kullanılırken politetik yöntemlerde ise her adımda iki gruba ayrılarak benzer objeler ortaya

çıkartılır (Seber, 2009: 377-378). Monotetik yöntemler her adımda tek bir değişkenle çalışabilirken politetik yöntemler ise her bir adımda tüm değişkenlerle çalışabilirler (Billard ve Diday, 2019: 197; Grabmeier ve Rudolph, 2002: 343).

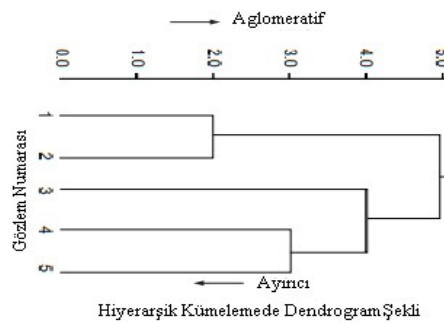
Hem aglomeratif hem de ayırıcı metotların sonuçları, dendrogram olarak bilinen iki boyutlu bir diyagram şeklinde gösterilebilir.

Dendrogram: Sol taraftaki y yükseklik değerlerinin uzaklıkları temsil ettiği, x ekseninin ise "objeleri" temsil ettiği bir ağaç diyagramıdır. Ağacın dallanması n-1 bağlantılarının sırasını gösterir; ilk çatal ilk bağlantıyı, ikinci çatal ikinci bağlantıyı temsil eder ve hepsi bir araya gelinceye kadar bu devam eder (Mardia ve diğerleri, 1995: 372).

Hiyerarşik yöntemler uygulamalarda kümeleme analizinin bel kemiğini oluştururlar. Yazılım paketlerinde yaygın olarak bulunurlar ve kullanımı kolaydır; ancak yapıları gereği büyük verilerde sonuçlandırılması zaman alıcıdır. Araştırmacının bu analizde yapması gerekenler uzaklık ve benzerlik ölçüsü seçimi, kümeleme yöntemi seçimi ve küme sayısının belirlenmesidir. Uygulamadaki ana problem, uygun matematiksel özelliklere (tek bağlantı gibi) sahip yöntemler genellikle deneysel olarak yorumlanabilir sonuçlar üretmediği için belirli bir kümeleme yönteminin önerilememesidir (Everitt ve diğerleri, 2011:110).

Şekil 8’de görüldüğü gibi, dendrogramlar art arda yapılmış olan birleşme veya bölünmeleri göstermektedir:

Şekil 8: Aglomeratif ve Ayırıcı Kümelerin Çalışma Yönleri



Kaynak: Everitt ve diğerleri, 2011: 75

Birleştirici(aglomeratif) yöntemleri kullanan algoritmaların işleniş yapısı genel olarak aşağıdaki gibi dört adımlıdır (Johnson, 1998: 323-324):

1. Tüm gözlemler birer küme olarak başlanır (yani, her gözlem tek üyeli bir küme oluşturur), böylece küme sayısı gözlem sayısına eşittir.

2. Bir benzerlik ölçüsünü kullanarak, en benzer iki kümeyi yeni bir kümede birleştirilir (iki gözlem içeren durum), böylece küme sayısını bir azaltır.

3. En benzer iki kümeyi yeni bir kümeleme işleminde birleştirmek için benzerlik ölçüsünü kullanarak kümeleme işlemi tekrarlanır.

4. Her bir adımda en benzer iki kümeyi yeni bir kümede birleştirerek bu işleme devam edilir. Tüm gözlemler bir kümede tutulana kadar işlem toplamda $n - 1$ kez tekrarlanır.

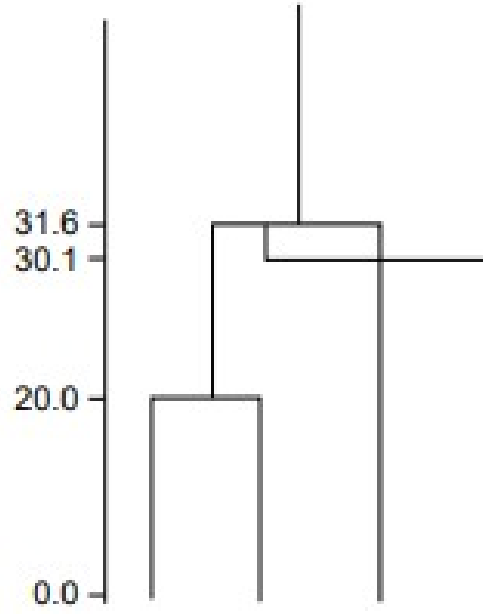
Hiyerarşik kümeleme yöntemleri veri seti üzerinde olduğu gibi uygulanabilen yöntemlerdir. İlave bir ön parametre girdisine ihtiyaç duymaz. Elde edilen dendrogram grafikleriyle herhangi bir küme sayısında gözlemlerin gruplanması incelenebilmektedir. Bununla birlikte algoritma yapıları gereği işlem yükünün fazla olması nedeniyle çok büyük miktarda veride uygulanmasının pratik olmaması ve aykırı gözlemlere karşı hiyerarşik olmayan yöntemlere göre daha duyarlı olması gibi dezavantajları da beraberinde getirmektedirler.

Hiyerarşik kümeleme yöntemleri, küme sayısı veya başlangıç bölümü hakkında önceden bilgi sahibi olmayı gerektirmez. Bu, hiyerarşik olmayan yöntemlere göre mutlak bir avantajdır. Fakat hiyerarşik yöntemlerin dezavantajlarından biri, kümeye bir gözlem atandığında başka kümeye atanamamasıdır. Bu nedenle, hiyerarşik yöntemler sıklıkla sadece bir keşif amaçlı kullanılırlar ve ortaya çıkan çözüm, küme çözümünü daha da hassaslaştırmak için hiyerarşik olmayan bir yöntemle uygulanır. Yani, hiyerarşik ve hiyerarşik olmayan yöntemler, rakip yöntemler yerine tamamlayıcı kümeleme yöntemleri olarak kullanılabilir (Sharma, 1996: 211).

Bu konudaki çok sayıdaki algoritma yaklaşımları arasında en sık kullanılan beş birleştirici algoritma yöntemlerine gelirsek (1) tek bağlantı, (2) tam bağlantı, (3) ortalama bağlantı, (4) ağırlık merkezi yöntemi ve (5) ward yöntemidir. Çalışmada da bu yöntemler incelenip kullanılacaktır.

Hiyerarşik yöntemlerden ağırlık merkezi ve medyan yönteminde kullanılan veri setine ve algoritmaya bağlı olarak tersine dönme (inversion) durumu meydana gelebilmektedir. Tersine dönme, mevcut bir kümeye önceki birleşimden daha küçük bir uzaklıkta (daha büyük benzerlikle) bağlandığında meydana gelir. Bir tersine dönme durumu Şekil 9'daki dendrogramda gösterilmiştir (Johnson ve Wichern, 2007: 695):

Şekil 9: Bir Tersine Dönme Görüntüsü Örneği



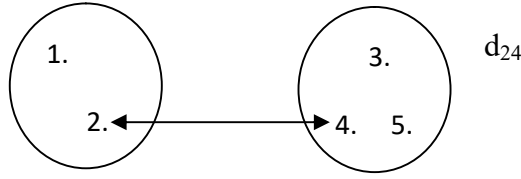
Kaynak: Everitt ve Diğerleri, 2011: 93

Ağırlık merkezi veya medyan yöntemlerinde tersine dönme durumuyla karşılaşılması halinde söz konusu yöntemlerin kullanılmaması gerektiği, olumsuz görüş bildiren ve başka yöntemlere geçiş yapılmasını öneren Murtagh, Morgan ve Ray gibi araştırmacılar mevcuttur (Morgan ve Ray, 1995: 132).

2.4.1. Tek Bağlantı Yöntemi

1951 yılından itibaren tanımlanıp uygulanmaya başlanan tek bağlantı yönteminin ilk uygulayıcıları olarak Florek ve arkadaşları (1951), Sneath (1957) söylenebilir (Hartigan, 1981: 388). Hartigan, 1973 yılında bu yöntemi evrimsel tarihin yeniden yapılanma modelindeki mutasyon sayısını en aza indirmek amacıyla kullanmıştır. Gower ve Ross, 1969 yılında bu yöntemin minimum yayılan ağaç kavramıyla ilişkisini göstermiştir ve Ross tarafından aynı yılda bir bilgisayar algoritması olarak verilmiştir (Mardia ve diğerleri: 1995: 372). En yakın komşuluk yöntemi olarak da bilinir. Kümeler arasındaki benzerliği, bir kümedeki herhangi bir objeden diğerindeki herhangi bir objeye en kısa uzaklık olarak tanımlar. Bu kural gözlemler arasındaki orijinal uzaklık matrisini yeni uzaklık ölçümleri hesaplamadan kullanmamızı sağlar. İki kümede gözlemler arasındaki tüm uzaklıklar bulunarak küme benzerliğinin ölçüsü olarak en küçüğü Şekil 10'daki gibi seçilir (Grimm ve Yarnold, 1995: 178).

Şekil 10: Kümeleme Uzaklığı (En Kısa)



Kaynak: Johnson ve Wichern, 2007: 681

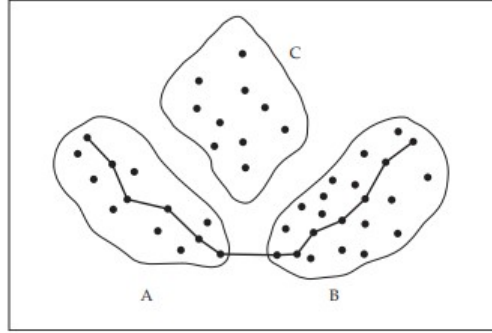
Tek bağlantı kümeleme analizi sonuçları, bir dendrogram veya ağaç diyagramı şeklinde grafiksel olarak gösterilebilir. Ağaçtaki dallar kümeleri temsil eder. Bu dallar, bir uzaklık (veya benzerlik) eksenini boyunca belirli konumlarda oluşan düğümlerle birleşir. Tek bağlantı yöntemini uygularken ilk olarak, elde edilen uzaklık matrisindeki en küçük $D = \{d_{ik}\}$ bulunmalı ve (UV) kümesini elde etmek için U ve V 'ye karşılık gelen objeler birleştirilmelidir. Daha önce yukarıda belirtilen hiyerarşik kümeleme analizindeki genel algoritma adımlarından üçüncü adımında, (UV) ifadesi ile diğer herhangi bir W kümesi arasındaki uzaklıklar aşağıdaki şekilde hesaplanır (Johnson ve Wichern, 2007: 682):

$$d_{(uv)w} = \min\{d_{uw}, d_{vw}\} \quad (2.12)$$

Burada d_{uw} ve d_{vw} değerleri sırasıyla U , W kümeleriyle V , W arasındaki en yakın komşu uzaklıklarıdır.

Bu yöntem kullanılan veri setine göre avantajlı olduğu kadar dezavantajlı da olabilmektedir. Şekil 11'de bunun bir örneği görülebilir. Şekilde üç küme (A, B ve C) birleştirilecektir. Her kümede yalnızca en yakın noktalara odaklanan tek bağlantı yöntemi, kümelerin en uç noktalarındaki kısa mesafeleri nedeniyle A ve B kümelerini birbirine bağlayacaktır. A ve B kümelerinin birleştirilmesi, C kümesini çevreleyen yeni bir küme oluşturur. Bununla birlikte, kümelerin homojenliği için çabalanacaksa, C kümesinin A veya B kümesi ile birleşmesi daha makuldür. Bu şekil, tek bağlantı yönteminin temel dezavantajını ortaya çıkarmaktadır (Grimm ve Yarnold, 1995: 178).

Şekil 11: Tek Bağlantı Yöntemi



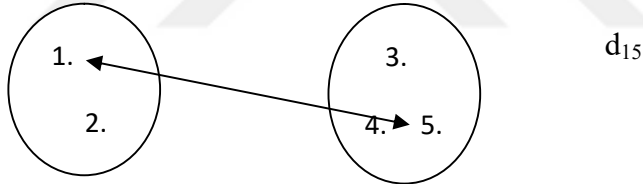
Benzer Olmayan A ve B Kümelerinin Tek Bağlantıda Birleşiminin Örneği

Kaynak: Grimm ve Yernold, 1995: 179

2.4.2. Tam Bağlantı Yöntemi

Sorensen tarafından 1948 yılında önerilen bu yöntemin tek bağlantı yönteminden farkı gözlem çiftleri arasında en kısa değer yerine en uzak mesafe değerini almasıdır (Hartigan, 1981: 388). Bu durum Şekil 12'deki gibi gösterilebilir:

Şekil 12: Kümeleme Uzaklığı (En Uzun)



Kaynak: Johnson ve Wichern, 2007: 681

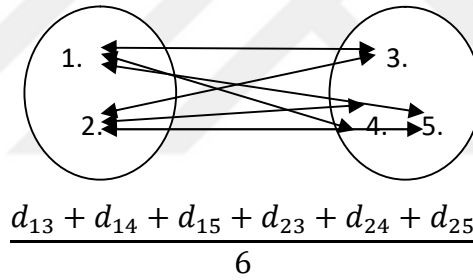
Tek bağlantı yöntemi için yapılan yorumların çoğu, bu yöntem için geçerlidir. Tek bağlantı gibi, bu yöntem de d_{ij} uzaklığının monoton dönüşümleri altında değişmezlik özelliğine sahiptir. Bununla birlikte, algoritmanın bazı aşamalarında en küçük uzaklık için bir düğüm noktası varsa (tek bağlantının aksine) tam bağlantı yöntemi tam olarak tanımlanamayabilir (Mardia ve diğerleri, 1995: 374). Verilerin sadece bir görünüşünü temsil etmesine rağmen (yani üyeler arasındaki en uzak mesafeyi), bu yöntem de benzer şekilde tıpkı tek bağlantı yönteminde olduğu gibi sapan gözlemlere karşı hassastır (Everitt ve diğerleri, 2011: 102). Formülü (UV) ve W kümeleri arasında şu şekilde kullanılır (Johnson ve Wichern, 2007: 686):

$$d_{(uv)w} = \max \{d_{uw}, d_{vw}\} \quad (2.13)$$

2.4.3. Ortalama Bağlantı Yöntemi

Sokal ve Michener tarafından 1958 yılında geliştirilmiştir. Herhangi iki kümenin benzerliğinin, bir kümedeki tüm bireylerin diğerindeki tüm bireylerle olan ortalama benzerliğine dayalı olması nedeniyle bu yöntem, tek bağlantı veya tam bağlantı yöntemlerinden ayrılır. Bu yöntem, tek veya tam bağlantı gibi uç değerlere (en yakın veya en uzak çiftler) bağlı değildir. Bunun yerine, benzerlik kavramı bir çift aykırı değere sahip üyeden ziyade kümelerin tüm üyelerine dayanır ve bu nedenle aykırı değerler tarafından daha az etkilenir (Seber, 2009: 363-375). Ortalama bağlantı yöntemi, Tek ve tam bağlantı yöntemleri arasındaki bir arabulucu tür olarak, homojen içerikli kümeler oluşturma eğilimindedir. Ayrıca mümkün olduğunca eşit değerde grup içi varyansa sahip kümelerin de oluşturulmasına meyillidir (Hair ve diğerleri, 2014: 441).

Şekil 13: Ortalama Bağlantı Yöntemi



Kaynak: Johnson ve Wichern, 2007: 681

Ortalama bağlantı yönteminin şekilsel görünümü Şekil 13'teki gibi ifade edilebilir. Genel tanımlanmış aglomeratif algoritmanın 3. Adımında, (UV) ve diğer W kümeleri arasındaki uzaklıklar bu yöntemde aşağıdaki gibi belirlenir (Johnson ve Wichern, 2007: 690):

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)} N_w} \quad (2.14)$$

d_{ik} burada (UV) kümesindeki i objesi ile W kümesindeki k objesi arasındaki uzaklıktır, $N_{(uv)}$ ve N_w kümeleri, sırasıyla (UV) ve W kümelerindeki gözlem sayısıdır.

2.4.4. Ward Yöntemi

1963 yılında Ward tarafından oluşturulan bu yöntem, küme (grup) içi kareler toplamını en aza indirgeyerek mümkün olduğunca en yüksek homojenlikte kümeler elde etmeye çalışır (Sharma, 1996: 193). Yöntemin her adımında birleştirilecek iki küme, tüm kümelerdeki değişkenlerin tamamının toplam kareler toplamındaki artışı en aza indiren kümelerdir. Bu yöntem, kümeleri az sayıda gözlemle birleştirme eğilimindedir, çünkü karelerin toplamı doğrudan ilgili gözlem sayısı ile ilişkilidir. Bir toplam kareler ölçüsünün kullanılması, bu yöntemin aykırı değerler tarafından kolayca bozulmasına neden olur (Grimm ve Yarnold, 1995: 180).

Ward yönteminde, gözlemler ile bu gözlemlerin ait olduğu kümenin merkezi arasındaki uzaklık karelerinin toplamı en aza indirilir. Bu yöntem, çok etkili olduğu düşünülse de, benzer kardinalitelere sahip kümeler oluşturma eğilimindedir. Bu sebeple eşit görüntüde kümeler elde etmek isteyen araştırmacılar için uygun bir yöntemdir. Fakat küçük oranlarda temsil yeteneğine sahip olacak kümelerin ortaya çıkarılması bu yöntemde zorlaşmaktadır (Wierzchoń ve Kłopotek, 2018: 30).

Ward yöntemi yapısı gereği küme (grup) içi karesi alınmış uzaklıkları en küçük, kümeler arası karesel uzaklıkları da en büyük yapmayı amaçlar. A ve B kümeleri için kare toplamaları (KT) şu formüllerle ifade edilir (Bulut, 2018: 378):

$$KT_A = \sum_{j=1}^{n_A} (x_j - \bar{x}_A)' (x_j - \bar{x}_A) \quad (2.15)$$

$$KT_B = \sum_{j=1}^{n_B} (x_j - \bar{x}_B)' (x_j - \bar{x}_B) \quad (2.16)$$

A ve B kümeleri birleştirildiğinde elde edilen yeni eşitlik ise aşağıdaki formüldeki gibi hesaplanır:

$$KT_{AB} = \sum_{k=1}^{n_{AB}} (x_k - \bar{x}_{AB})' (x_k - \bar{x}_{AB}) \quad (2.17)$$

Bu formüldeki $\bar{x}_{AB}$ ifadesi aşağıdaki şekilde hesaplanır:

$$\bar{x}_{AB} = \frac{n_A \bar{x}_A + n_B \bar{x}_B}{n_A + n_B} \quad (2.18)$$

Ward yönteminde A ve B kümesi birleştirildiğinde kareler toplamındaki artışın düşük olması ve aşağıdaki formülle bu durumun ölçülerek küçük bir değer alması istenir. Tüm mümkün durumlarda en küçük I_{AB} değerine sahip kümeler bir araya getirilir.

$$I_{AB} = KT_{AB} - (KT_A + KT_B) \quad (2.19)$$

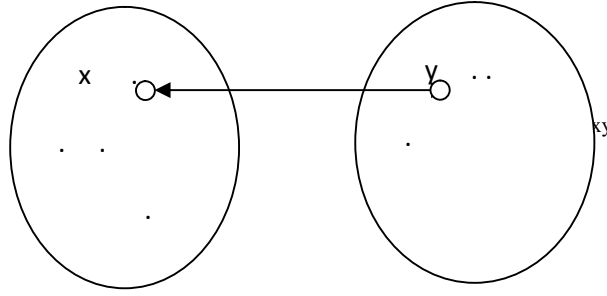
2.4.5. Ağırlık Merkezi (Centroid) Yöntemi

İlk olarak Sokal ve Michener tarafından 1958 yılında geliştirilip uygulanmıştır. Bu yöntemde, iki küme arasındaki benzerlik küme ağırlık merkezleri arasındaki uzaklıktır. Küme ağırlık merkezleri, küme değişkenindeki değişkenler üzerindeki gözlemlerin ortalama değerleridir. Bu yöntemde, bireyler her gruptandırıldığında yeni bir ağırlık merkezi hesaplanır. Fakat daha önce de belirtildiği gibi bu yöntemle çizilen dendrogramda tersine dönme durumu oldukça ihtimal dâhilindedir. Bu yöntemde $d_{AB} = d(\bar{x}_A, \bar{x}_B)$ uzaklık formülasyonu kullanılmaktadır. Burada $\bar{x}_A$ A'nci kümeye ait gözlem değerlerinin ortalamasını, $\bar{x}_B$ ise B'inci kümeye ait gözlem değerlerinin ortalamasını göstermektedir. Yeni oluşturulan küme merkezleri ise aşağıdaki formül kullanılarak hesaplanır (Seber, 2009: 362):

$$\bar{x}_{AB} = \frac{n_A \bar{x}_A + n_B \bar{x}_B}{n_A + n_B} \quad (2.20)$$

Küme ağırlık merkezleri küme birleşimleri gerçekleştikçe güncellenir. Başka bir deyişle, bir küme ağırlık merkezi, var olan bir kümeye her yeni bir birey veya birey grubu eklendiğinde değişir. Bu yöntemler, yaşam bilimlerinde (örneğin biyoloji) oldukça popülerdir, ancak dağınık ve çoğu zaman kafa karıştırıcı sonuçlar verebilir. Bu karışıklık tersine dönme nedeniyle, yani, bir çiftin ağırlık merkezleri arasındaki uzaklığın, daha önceki bir kombinasyonda birleştirilen başka bir çiftin ağırlık merkezleri değerleri arasındaki mesafeden daha az olduğu durumlar sonucu meydana gelir. Bu yöntemin avantajı, ortalama bağlantı yöntemi gibi, diğer hiyerarşik yöntemlere göre aykırı değerlerden daha az etkilenmesidir (Hair ve diğerleri, 2014: 442). Bu yöntem şekil 14'te x ve y'nin ağırlık merkezleri kabul edildiği kümeler ile aşağıdaki gibi tasvir edilebilir:

Şekil 14: Ağırlık Merkezi Yöntemi



2.4.5. Medyan Yöntemi

Bu yöntem ilk olarak Gower tarafından 1967 yılında ağırlık merkezinin çeşitli olumsuzluklarına karşı alternatif olarak geliştirilerek uygulanmıştır. Ağırlık merkezi yöntemi ile arasındaki tek fark birleştirilmiş iki kümenin yeni merkezinin hesaplamasında meydana gelen görüş farklılığıdır. Bu fark da iki küme merkezi arasındaki noktanın merkez olarak alınarak kümelerdeki birim sayısının etkisinin minimuma indirilmesidir (Seber, 2009: 363). Buradaki medyan ifadesi istatistikte sıklıkla kullanılan medyanla birebir aynı değildir ve şöyle tanımlanır (Bulut, 2018: 378):

$$m_{AB} = \frac{1}{2}(\bar{x}_A + \bar{x}_B) \quad (2.21)$$

Medyan yöntemi tıpkı ağırlık merkezinde olduğu gibi algoritma yapısı nedeniyle tersine dönme durumuyla karşılaşılabilir bir yöntemdir. Bu nedenle çizilen dendrogramlar ayrıntılı incelenmelidir.

2.5. HİYERARŞİK OLMAYAN KÜMELEME ANALİZİ YÖNTEMLERİ

Hiyerarşik olmayan yöntemlerin yapıları itibarıyla hiyerarşik yöntemlere göre oldukça belirgin farkları bulunmaktadır. Bunlardan temel olarak bahsedecek olursak ayrılacak küme sayısının önceden belirlenmiş olması (k sayısının belirlenmesi), bir dendrogram Tablosu oluşturulmaması ve bir uzaklık veya benzerlik ölçüsü kullanılmamasıdır. Bu yöntemler bölümlenmeli (partitional) veya aşamalı olmayan kümeleme yöntemleri olarakta isimlendirilmektedir. Hiyerarşik olmayan kümeleme yöntemlerine ait her ne kadar çok sayıda yöntem mevcut olsa da bu yöntem, literatürde çok sık kullanıldığından dolayı genellikle k-ortalamlar (k-means) kümeleme yöntemi olarak anılmaktadır.

Hiyerarşik olmayan kümeleme yöntemleri, daha sonra iteratif olarak iyileştirilen bir dizi başlangıç kaynak noktasına (veya kümelere) dayanarak çalışmaktadır. Hiyerarşik yöntemler ise tek kümelerde tek bir veri noktalarıyla başlayabilir ve kümelemeyi oluşturabilir. Uzaklık ölçülerinin rolü bu yöntemlerin her ikisinde de farklıdır. Hiyerarşik kümelemede, uzaklık ölçüsü başlangıçta temel düzeydeki veri noktalarına uygulanır ve daha sonra alt kümeler için mutlak temsili noktalar seçilerek kademeli olarak alt kümelere uygulanır. Hiyerarşik olmayan kümeleme yöntemleri söz konusu olduğunda ise, farklı iterasyonlarda seçilen temsili noktalar genel olarak kümenin ağırlık merkezi veya (verilerde mevcut olmayan) bir grup sanal nokta olabilmektedir (Aggarwal ve Reddy, 2014: 88).

Hiyerarşik yöntemlerde $n \times n$ boyutlu uzaklık gözlemler matrisi kullanılırken, hiyerarşik olmayan yöntemlerde $n \times p$ boyutlu değişken ve gözlemler boyutunda örüntü matrisi ya da yine $n \times n$ boyutlu uzaklık matrisi kullanılmaktadır (Çilingirtürk, 2011: 174).

Hiyerarşik olmayan yöntemler daha fazla kabul edilebilirlik ve kullanım kazanmıştır, ancak bu yöntemdeki herhangi bir uygulamanın başarısı, araştırmacının kaynak noktalarını bazı pratik, objektif veya teorik temellere göre seçebilme yeteneğine bağlıdır. Bu durumlarda hiyerarşik olmayan yöntemler, hiyerarşik yöntemlere göre çeşitli avantajlar sunar. Bunlar (Hair ve diğerleri, 2014: 445):

1. Sonuçların duyarlılığı verilerdeki aykırı değerlere, kullanılan uzaklık ölçüsüne, alakasız veya uygunsuz değişkenlerin dâhil edilmesine karşı daha azdır.

2. Hiyerarşik olmayan yöntemler son derece büyük veri kümelerini algoritma yapıları gereği daha hızlı ve rahat bir biçimde analiz edebilir, çünkü tüm gözlemler arasında benzerlik matrislerinin hesaplanmasını gerektirmez. Bunun yerine, her gözlemin sadece küme ağırlık merkezlerine olan benzerliğini dikkate alır. Kümeler arasında gözlemlerin yeniden atanmasına izin veren optimizasyon algoritmaları bile her boyuttaki veri kümesine kolayca uygulanabilir.

Hiyerarşik olmayan kümelemede veriler k adet bölüm veya gruplara ayrılır ve her bölüm bir kümeyi temsil eder. Bu nedenle, hiyerarşik kümelemenin aksine küme sayısının önceden bilinmesi gerekir. Hiyerarşik olmayan kümeleme analizi yöntemleri temel olarak şu adımları izler (Sharma, 1996: 202):

1. k başlangıç küme ağırlık merkezleri veya kaynak noktaları seçilir, burada k , istenen küme sayısıdır.

2. Her bir gözlem en yakın olduğu kümeye atanır.
3. Her bir gözlemi önceden belirlenmiş bir durdurma kuralına göre k adet kümeden birine yeniden atanır veya tahsis edilir.
4. Veri noktalarının yeniden tahsisi veya yeniden atama yoksa durdurma kuralı tarafından belirlenen ölçütleri karşılaması durumunda işlem durdurulur. Aksi takdirde ikinci adıma geri dönülür.

Objelerin tamamının kümelere atanması, büyük oranda ilk bölünme veya ilk kaynak noktası seçimine bağlı olacaktır. Bu konudaki çalışmalarda tecrübeler, bir çalışmadaki en büyük kırılma noktasının bu ilk atamaların belirlenme aşamasında meydana geldiğini göstermektedir (Johnson ve Wichern, 2007: 696).

Birçok araştırmacı her iki yöntemi de kullanan bir birleşim önermektedir. Bu şekilde bir yaklaşımın zayıflıkları diğeriyle birleştirilerek telafi edebilir. Bu işlem iki adımda gerçekleştirilir (Hair ve diğerleri, 2014: 446):

1. İlk olarak, eksiksiz bir küme çözümleri grubu üretmek, uygulanabilir bir küme çözümü oluşturmak için hiyerarşik bir yöntem kullanılır ve uygun sayıda küme belirlenir.
2. Aykırı gözlemler ayıklandıktan sonra, kalan gözlemler hiyerarşik olmayan bir yöntemle toplanabilir.

Bu adımlar sonucunda her iki yöntemin avantajlarından yararlanılıp dezavantajlarının etkisini minimuma indirerek analizi gerçekleştirebiliriz. Bu çalışmada k-ortalamlar yöntemi ve daha sonra ortaya çıkan k-medoid yöntemleri üzerinde durulacaktır.

2.5.1. k-Ortalamlar Yöntemi

k-ortalamlar terimi ilk defa Macquenn tarafından 1967 yılında tanımlanıp kullanılmasına rağmen bu yöntemin uygulanma fikri 1956 yılına kadar uzanmaktadır. Zaman içinde en popüler kümeleme analizi algoritmalarından birisi haline gelmiştir. Hiyerarşik olmayan kümeleme analizi kavramında ilk akla gelen yöntemlerden birisi haline gelmiş ve neredeyse tüm kümeleme analizi paket yazılımlarında da yerini almıştır. Fonksiyonu ise aşağıdaki gibidir (Macqueen, 1967: 283):

$$\arg \min x_1, \dots, x_k = \sum_{i=0}^n \min_{j=1,2,\dots,k} \|z_i - x_j\|^2 \quad (2.22)$$

k-ortalamalar yönteminin amacı diğer kümeleme analizi yöntemlerindeki gibi grup içi homojenliği ve gruplar arası heterojenliği en büyüklemektir (Bulut, 2018: 389). Verileri kullanıcı tarafından belirlenen sayıda kümeye bölerek çalışır ve ardından bazı sayısal ölçütler yerine getirilinceye kadar gözlemleri yinelemeli olarak kümelere yeniden atar. k-ortalamalar ölçütü, kümelerdeki gözlem uzaklığını en aza indirmeye ve kümeler arasındaki uzaklığı en yükseğe çıkarma ile ilgili bir amacı belirler. k-ortalamalar yöntemi o kadar yaygın kullanılır ki, kimi araştırmacılar tarafından sadece hiyerarşik olmayan kümeleme analizine atıfta bulunmak için bile kullanılır (Hair ve diğerleri, 2014: 444).

En basit uygulamasında süreç şu üç adımdan oluşur (Johnson ve Wichern, 2007: 696):

1. Objeler k adet başlangıç noktasına rastgele ayrılır.
2. Ağırlık Merkezine (ortalama) en yakın olan kümeye bir obje atanarak, obje sırasınca ilerlenir (Uzaklık genellikle standartlaştırılmış veya standartlaştırılmamış gözlemlerle Öklid uzaklık kullanılarak hesaplanır). Yeni gözlem eklenen küme ve gözlemi kaybeden küme için ağırlık merkezi değeri yeniden hesaplanır.
3. Atamalar durana kadar bir önceki adım tekrarlanır.

k-ortalamalar yöntemi hiyerarşik yöntemlere göre daha iyi sonuçlar üretebilir fakat kümelerin ayrılmasında kullanılacak başlangıç noktalarının çok iyi seçilmesi gerekmektedir (Tinsley ve diğerleri, 2000: 314).

2.5.2. k-Medoid Yöntemi

Bu yöntemin uygulanışında veri seti içinde yer alan bazı gözlemlerin temsili k adet gözlem birimleri olabileceği düşünülebilir. Bu durum k-ortalamalar yöntemindeki rastgele belirlenen kaynak noktası belirleme yöntemine göre farklı bir seçim biçimidir. Bu temsili k adet gözlem birimleri medoid (üyelik) olarak adlandırılırlar (Sönmez ve Er, 2007: 20). k-medoid yöntemi k-ortalamalar yöntemine kıyasla aykırı değerlere karşı daha dayanıklı olan bir kümeleme algoritmasıdır. k-ortalamalara benzer şekilde, k-medoid yönteminin amacı da en az işlemle bir kümeleme çözümü bulmaktır. Bu yöntemde, gerçek veri noktaları ilk örnek olarak seçilerek kümeleme işlemi yapılır. (Aggarwal ve Reddy, 2014: 93).

Diğer klasik yöntemlere göre nispeten yeni olan k-medoid yöntemi ilk olarak 1987 yılında Rousseeuw ve Kaufman tarafından önerilmiştir. Fonksiyonu ise aşağıdaki gibidir:

$$S = \sum_{k=1}^K \sum_{i \in C_k, j \in C_k} d(x_i, x_j) \quad (2.23)$$

Bu fonksiyonda S küme içi gözlemlerin arasındaki toplam uzaklıktır ve bu değerin en küçük olması durumunda en iyi kümeleme çözümüne ulaşıldığı kabul edilir (Bulut, 2018: 393).

k-medoid yöntemi k-ortalamlar yöntemine göre daha dayanıklı bir yöntemdir, ancak k-medoid yönteminin hesaplama karmaşıklığı daha yüksektir ve bu nedenle büyük veri kümeleri için uygun değildir. Aykırı değerlere karşı duyarlılığı düşüktür. Bu yöntemin adımları aşağıdaki şekilde uygulanır (Aggarwal ve Reddy, 2014: 94):

- 1: İlk temsili k gözlemleri için noktalar seçilir.
- 2: Verideki her nokta seçilmiş en yakın temsili obje ile kümeye atanır.
- 3: Temsili olmayan bir x_i gözlemi rastgele seçilir.
- 4: Temsili obje olan m ile x_i değerini değiştirmeyeyle oluşan S toplamı hesaplanır.
- 5: $S < 0$ ise, yeni k temsili obje kümesini oluşturmak için m değeri x_i ile değiştirilir.
- 6: Yakınsama ölçütü karşılanıyorsa işlem durdurulur. Karşılanmıyorsa ikinci adıma dönlür.

2.6. KÜME GEÇERLİLİK İNDEKSLERİ VE KÜME SAYISININ BELİRLENMESİ

Kümeleme analizi yöntemlerinin temel amacı olan kendi içlerinde homojen, düşük varyansa sahip ve aralarında da yüksek varyansa sahip heterojen gözlemlerin gruplanabilmesi için söz konusu veri setinin uygun sayıda kümeye ayrılması gerekmektedir. Söz konusu veri setinde bulunan gözlemlerin gruplandığı uygun küme sayısının gerektiğinden daha az veya fazla olması gözlemlerin açıklanabilirliğinde hatalara sebep olacaktır. Ayrıca hiyerarşik olmayan kümeleme analizi yöntemlerinde algoritmaların çalışabilmesi için uygun bir küme sayısının da

analizden önce belirlenmesi gerekmektedir. Uygun küme sayısının belirlenmesinde genel bir ortak yöntem literatürde bulunmamaktadır. Bunun temel nedeni çalışmalarda araştırılacak verinin değişik yapılarda olabilmesidir. Bu sebeple küme sayısının belirlenmesinde çoğunlukla söz konusu araştırmayı yapacak kişilerin önceliğine bırakılmaktadır. Fakat yine de bu araştırmanın yapılabileceği çeşitli yöntemler mevcuttur.

Kümeleme analizi yöntemleri sezgisel tekniklerdir; Uygulanan verilerde herhangi bir doğal grup veya küme olmasa bile bir kümeleme çözümü veya gruplama ortaya çıkacaktır. Bu nedenle, bir küme çözümünün güvenilirliğini ve dışsal geçerliliğini belirlemek daha da önemlidir (Sharma, 1996: 221). Kümeleme analizinin optimizasyon yöntemlerinin çoğu uygulamasında araştırmacının verilerdeki küme sayısını 'tahmin etmesi' gerekecektir. Belirli durumlarda yardımcı olabilecek çeşitli yöntemler önerilmiştir. Çoğu nispeten literatürde yaygın yöntemler değildir ve esasen, kümeleme kriteri değerinin grup sayısına göre belirlenmesini içeren, kimi zaman subjektif içerikli değerlendirmelerdir (Everitt ve diğerleri, 2011:126).

2.6.1. Küme Sayısının Belirlenmesi

Araştırmacıların küme sayısını belirlemesi için herkesin kabul ettiği belli bir yöntem olmamakla birlikte çok sayıda yöntem mevcuttur. Bu yöntemler arasında küme geçerlilik indekslerinin kullanımı, ana bileşen skorları ve grafiksel yöntemleri, dendrogramları inceleme ve bazı matematiksel yöntemler bulunmaktadır.

Uygulamada, küme sayısının (k 'nın) değişkenlik göstemesinin mümkün olduğu durumlarda maksimum olabilirlik yöntemleri (MLE), verileri daima mümkün olan en yüksek sayıda kümeye ayırır. Bu sebeple k 'nın başka yöntemlerle seçilmesi gerekir. Eşit kovaryans matrisleri için 1971 yılında Marriott, $|W|$ ölçütünü araştırmış ve $k^2|W|$ ifadesinin en küçük yapacak şekilde belirlenen k 'nın alınmasını önermiştir. Bu eşitlikte W , grup içi kareler ve çarpımlar toplamı matrisini ifade eder. Bunun yanında 1974 yılında Everitt, farklı ölçümlerle ilgili yaptığı araştırmalarda $k^2|W|$ ifadesinin en faydalı yöntem olduğunu ortaya koymuştur. Bu algoritmalarda temel kural aşağıdaki formülü başarıyla uygulamak olup k burada elde edilecek küme sayısı, n de gözlem sayısıdır. Örneklem hacminin büyük olduğu veri setlerinde bu

yöntemin kullanımı hatalı bir küme sayısı belirlemeye neden olabilmektedir (İncekırık, 2005: 36).

$$k \cong (n/2)^{1/2} \quad (2.24)$$

Marriott'un 1971 yılında belirlemiş olduğu $k^2|W|$ ölçütü küme sayısı belirlemede kullanılabilen örnek yöntemlerden birisidir. Yukarıdaki formülde olduğu gibi çoğu zaman sağlıklı sonuçlar üretse de bu da her zaman doğru bir küme sayısı veremeyebilir. W burada grup içi kareler toplamı olan $\sum_{i=1} \sum_{j=1} (y_{ij} - \bar{y}_i)'(y_{ij} - \bar{y}_i)$ işlemini, k ise küme sayısını temsil etmektedir (Seber, 2009: 388).

Araştırmacının küme sayısına karar vermesinde mesleki deneyimi ve mümkün olduğunca fazla yöntemi göz önünde bulundurarak faydalanması doğru karar vermesini kolaylaştıracaktır.

2.6.2. Küme Geçerlilik İndeksleri

Kümeleme analizi sonuçlarını değerlendiren küme geçerliliği ölçütü kümeleme analizi uygulamalarının başarısı için önemli konulardan birisidir. Bu duruma eğilip çaba harcayan çok sayıda uzmana rağmen küme geçerliliğinde tutarlı ve kesin bir çözüm yoktur. Uygulamalarda kullanılacak en uygun ölçü de bilinmemektedir. Kümeleme analizi literatüründe tam olarak ele alınmayan birçok geçerlilik ölçüsü kavramı bulunmaktadır. Ayrıca, farklı geçerlilik ölçüleri arasında net bir ilişki yoktur. Küme geçerlilik indeksleri içsel (internal) indeksler ve dışsal (external) indeksler olmak üzere iki ana kategoride sınıflandırılabilir. Bu indeksler arasındaki temel fark, küme geçerliliği için harici bilgilerin kullanılıp kullanılmadığıdır (Aggarwal ve Reddy, 2014: 572).

Dışsal İndeksler: Dışsal ölçü, bir kümeleme algoritması tarafından keşfedilen kümeleme yapısını, örneğin verilen sınıf etiketlerinde belirtilen dış yapı ile ne ölçüde eşleştirdiğini değerlendirmek için verilerde bulunmayan harici bilgileri kullanan indekslerdir (Aggarwal ve Reddy, 2014: 572). Bunun anlamı ise dışsal indekslerin kümeleme sonucuna ilişkin önsel bir bilgi olduğunda kullanılabilir olmasıdır. Bu indekslerin altında yatan temel düşünce, bilinen bir kümeleme sonucuyla elde edilen kümeleme yönteminin ulaştığı sonuçların benzerliğine dayanmasıdır. İki sonuç birbirine yakınlaştıkça elde edilen kümeleme sonuçlarının başarılı olduğuna karar verilir (Bulut, 2018:414).

Dışsal geçerlilik, kümeleme analizinden elde edilen sonuçlar ile bir dış ölçüt karşılaştırılarak elde edilir. Örneğin, firmaları belirli finansal oranlara göre kümelendirdiğimizi ve böylece finansal olarak sağlıklı firmalar ve finansal olarak sağlıklı olmayan firmalar olmak üzere iki küme elde ettiğimizi varsayalım: Dış geçerlik daha sonradan kümeleme analizi sonuçları bağımsız değerlendiriciler tarafından (denetçiler, finansal analistler, borsacılar, endüstri analistleri vb.) elde edilen sınıflandırma ile ilişkilendirilerek oluşturulabilir (Sharma, 1996: 221).

İçsel İndeksler: İçsel indeksler dışarıdan bir bilgi gerekmezken kullanılabilir ve ek bilgiye ihtiyaç duymadan en iyi kümeleme algoritmasını ve en uygun küme sayısını seçmek için kullanılabilirler. Uygulamada, sınıf etiketleri gibi harici bilgiler bazı gerçek hayat uygulamalarında mevcut değildir. Bu durumda, harici doğrulama da olmadığından küme doğrulaması için tek seçenek içsel geçerlilik ölçüleridir. Bu çalışmada da içsel indeksler kullanılmıştır (Aggarwal ve Reddy, 2014: 572). Bu indekslerin çalışma dayanağı küme içi homojenlik ve heterojenlik üzerinedir (Bulut, 2018:414).

İçsel ve dışsal geçerlilik indeksleri birbiriyle karşılaştırıldığında içsel indekslerin daha başarılı oldukları gözlemlenen araştırmalar mevcuttur. Bu konuda verilebilecek örnek araştırmalardan birisi olan Abundez, Arizmendi, Rendon ve Quiroz'un yapmış olduğu bir çalışmada 13 adet veri seti alınarak dört adet içsel ve dört adet dışsal geçerlilik indeksi seçilmiştir. Burada k-ortalamlar yöntemi kullanılarak her bir veri seti için içsel ve dışsal geçerlilik indekslerinin verdiği sonuçlar karşılaştırılmıştır. Elde edilen indeks sonuçlarının karşılaştırılması sonucu içsel indekslerin dışsal indekslere göre daha doğru (içsel indekslerde yaklaşık %77, dışsal indekslerde % 61,5) sonuçlar verdiği ortaya çıkmıştır (Rendon ve diğ., 2011: 33).

2.6.2.1. Siluet (Silhouette) İndeksi

Bu yöntem Rousseeuw tarafından önerilmiştir. Siluet indeksi (S), küme içi ve kümeler arası mesafelerin ikili karşılaştırmalarındaki farklara dayanarak kümeleme performansını analiz eder (Rousseeuw, 1987a: 55). Ek olarak, uygun küme sayısını, bu indeksin o anki küme sayısındaki değerlerine bakılarak maksimuma ulaşan değer seçilerek belirlenmektedir (Aggarwal ve Reddy, 2014: 572). S değeri aşağıdaki formülle hesaplanır (Bulut, 2018:416):

$$S = \frac{\sum_{i=1}^N \frac{b_i - a_i}{\max(a_i, b_i)}}{N} \quad (2.25)$$

Bu indeks $[-1,1]$ aralığında değer alır ve 1 değerine yakın olması elde edilen kümelemenin çok iyi olduğuna, 0.5 üstünün makul olduğuna, fakat 0.2'nin altında bir genişlikte ise küme yapısında önemli bir eksikliğe işaret eder (Everitt ve diğerleri, 2011: 129). Bu indeks hem küme içi hem de kümeler arası mesafeler göz önüne alınarak formüle edilir. Belirli bir x_i değeri için, ilk önce aynı kümedeki tüm noktalara olan uzaklıkların ortalaması hesaplanır. Bu değer a_i değeri olarak belirlenmiştir. Daha sonra x_i içermeyen her küme için, x_i 'nin her kümedeki tüm veri noktalarına ortalama mesafesi hesaplanır. Bu değer b_i olarak ayarlanmıştır. Bu iki değeri kullanarak bir noktanın siluet katsayısı tahmin edilir. Veri kümesindeki tüm siluetlerin ortalaması, veri kümesindeki tüm noktalar için ortalama siluet genişliği olarak adlandırılır. Bir kümelemenin kalitesini değerlendirmek için tüm noktaların ortalama siluet katsayısı hesaplanabilir (Aggarwal ve Reddy, 2014: 93).

2.6.2.2. Dunn İndeksi

Dunn tarafından tasarlanıp geliştirilmiştir. Kümeler arası ayırma ölçütü olarak farklı kümelerdeki objeler arasındaki minimum ikili mesafeyi, küme içi kompaktlığı ve tüm kümeler arasındaki maksimum çapı kullanır. Bu üç indeks, a ve b 'nin ağırlık katsayıları olduğu bir $İndeks = (a \cdot \text{Ayırma}) / (b \cdot \text{Kompaktlık})$ biçimini alır. Optimal küme sayısı, bu indekslerin değeri maksimize edilerek belirlenir (Dunn, 1974: 95; Aggarwal ve Reddy, 2014: 591). Fonksiyonu:

$$D(C) = \min_i \left\{ \min_j \left(\frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \{ \max_{x, y \in C_k} d(x, y) \}} \right) \right\} \quad (2.26)$$

Şeklinde gösterilebilir. Bu ifade 0'dan sonsuza kadar değerler alabilir ve iyi bir kümeleme sonucu için bu ifadenin mümkün olduğunca büyük olması istenir (Bulut, 2018:416).

2.6.2.3. Calinski ve Harabasz (CH) İndeksi

Calinski ve Harabasz tarafından geliştirilmiştir. Formülü aşağıdaki şekilde ifade edilebilir (Calinski ve Harbarz, 1974: 10):

$$C_h(q) = \frac{\frac{(B_q)}{(q-1)}}{\frac{(W_q)}{(n-q)}} \quad (2.27)$$

Burada;

$$B_q = \sum_{k=1}^q n_k * (c_k - c)(c_k - c)^T \quad (2.28)$$

$$W_q = \sum_{k=1}^q \sum_{i \in C_k} (x_i - c_k)(x_i - c_k)^T \quad (2.29)$$

İşlemlerine karşılık gelir. Bu indekste q küme sayısıdır. B_q , q kümeleri halinde gruplanan veriler için gruplar arası dağılım matrisi, B_w de q kümeleri halinde kümelenmiş veriler için grup içi dağılım matrisidir. x_i , küme k 'deki i . objenin gözlemlerinin p -boyutlu vektörü, C_k küme k 'nin ağırlık merkezi, c veri matrisinin ağırlık merkezi n_k , C_k kümesindeki obje sayısıdır. C_h değerini maksimum yapan küme sayısı en uygun olarak kabul edilir (Charrad ve diğerleri, 2013:4-5).

2.6.2.4. Kübik Kümeleme Kriteri (CCC)

1983 yılında Sarle tarafından önerilmiştir. Bu indekse göre q küme sayısının bu indekste maksimuma ulaşıldığı durum en uygun küme sayısı olarak kabul edilir. Formül gösterimi aşağıdaki gibidir (Sarle, 1983:1-8):

$$CCC = \ln \left[\frac{1 - E(R)^2}{1 - R^2} \right] \frac{\sqrt{\frac{np^*}{2}}}{(0.001 + E(R^2))^{1.2}} \quad (2.30)$$

$$R^2 = 1 - \frac{iz(W)}{iz(T)} \quad (2.31)$$

$$E(R^2) = 1 - \left[\frac{\sum_{j=1}^{p^*} \frac{1}{n+u_j} + \sum_{j=p^*+1}^p \frac{u_j^2}{n+u_j}}{\sum_{j=1}^p u_j^2} \right] \left[\frac{(n-q)^2}{n} \right] \left[1 + \frac{4}{n} \right] \quad (2.32)$$

T burada $X'X$ toplam örneklemin kareler toplamı ve $p \times p$ kadar çapraz çarpım matrisidir.

W değeri $T-B$ grup içi toplam örneklemin kareler toplamı ve $p \times p$ kadar çapraz çarpım matrisidir.

$B = \bar{X}'Z'Z\bar{X}'$ gruplar arası toplam örneklemin kareler toplamı ve $p \times p$ kadar çapraz çarpım matrisidir.

$$\bar{X} = (Z'Z)^{-1}Z'X \text{ dir.}$$

Z değeri, eğer i 'inci gözlem k 'ıncı kümeye aitse $z_{ik} = 1$ elementli bir küme gösterge matrisidir ($n \times q$), aksi takdirde 0'dır.

Bu gösterimde $u_j = \frac{s_j}{c}$ değerine eşittir. s_j ise $\frac{T}{(n-1)}$ 'in j 'inci özdeğerinin kareköküdür. c değeri de $(\frac{v^*}{q})^{\frac{1}{p^*}}$, v^* ise $\prod_{j=1}^{p^*} s_j$ değeridir. p^* , q 'dan küçük en büyük tamsayı olarak seçilmiştir, böylece u_p^* birden az olmayacaktır (Charrad ve diğerleri, 2013:7-8).

2.6.2.5. C İndeksi

Hubert ve Levin tarafından önerilmiştir. q küme sayısı değerinin minimize ettiği indeks değeri en uygun küme sayısı olarak kabul edilir. Formülü aşağıdaki biçimde ifade edilir (Hubert ve Levin, 1976: 20):

$$cindex = \frac{Du - (r \times D_{min})}{r \times D_{max} - (r \times D_{min})} \quad (2.33)$$

$$D_{min} \neq D_{max}, cindex \in (0,1)$$

Du , küme içi farklılıkların toplamı

r = küme içi farklılıkların sayısı

D_{min} = en küçük küme içi farklılığı

D_{max} = en büyük küme içi farklılığı

Olarak bu fonksiyondaki yerlerini almaktadırlar (Charrad ve diğerleri, 2013: 6).

2.6.2.6. DB İndeksi

Davies ve Bouldin tarafından önerilmiştir. Formülü aşağıdaki biçimde gösterilebilir (Davies ve Bouldin, 1979: 224-225):

$$DB = \frac{1}{NC} \sum_i \max_{j \neq i} \left\{ \left[\frac{1}{n_i} \sum_{x \in C_i} d(x, c_i) + \frac{1}{n_j} \sum_{x \in C_j} d(x, c_j) \right] / d(c_i, c_j) \right\} \quad (2.34)$$

Bu fonksiyonda NC küme sayısı, n C 'deki obje sayısı, c C 'nin ağırlık merkez değeri, d de uzaklık ifadesidir. Her C kümesi için, C ile diğer tüm kümeler arasındaki benzerlikler hesaplanır ve en yüksek değer küme geçerlilik indeks değeri olarak C 'ye atanır. Daha sonra DB dizini, tüm küme benzerliklerinin ortalaması alınarak elde edilebilir. Dizin ne kadar küçük olursa, kümeleme sonucu o kadar iyi olur. Bu indeks en aza indirgeyerek kümeler birbirinden en belirgin olanıdır ve bu nedenle en iyi biçimi elde eder (Aggarwal ve Reddy, 2014: 590-591).

2.6.2.7. Duda İndeksi

Duda ve Hart tarafından 1973 yılında önerilmiştir. Formülü şu şekilde gösterilebilir (Duda ve diğerleri, 1973: 46-47):

$$duda = \frac{je(2)}{je(1)} = \frac{W_k + W_l}{W_m} \quad (2.35)$$

$je(2)$, veriler iki kümeye ayrıldığında küme içindeki hata kareleri toplamıdır ve $Je(1)$ yalnızca bir küme olduğunda hata kareleri verir.

$$W_k, W_l, W_m \text{ CH İndeksinde bulunan } W_q = \sum_{k=1}^q \sum_{i \in C_k} (x_i - c_k)(x_i - c_k)^T$$

olarak tanımlanır.

c_k ve c_l kümelerinin c_m oluşturmak üzere birleştirildiği varsayılmaktadır.

$$B_{kl} = W_m - W_k - W_l, c_m = c_k \cup c_l \text{ tanımlanır.}$$

n_i = küme c_i 'deki gözlem sayısı, $i = k, l, m$ 'ye eşittir.

Bu indekste en uygun küme sayısı aşağıdaki şartı sağlayacak en küçük q küme sayısı değeridir (Charrad ve diğerleri, 2013: 5):

$$duda \geq 1 - \frac{2}{\pi p} - z \sqrt{\frac{2(1 - \frac{8}{\pi^2 p})}{n_{mp}}} = \text{Duda İndeksi Kritik Değeri} \quad (2.36)$$

p, veri kümesindeki değişkenlerin sayısıdır.

z standart normal bir skor değeridir. 3.2 değerinde en iyi sonuçları vermektedir.

Bu indeks, dendrogramda doğru kesme seviyesini seçerek hiyerarşik kümeleme sürecini durdurmaya içeren bir tahmin yöntemidir. Tipik bir dendrogramı kesmek için ise kullanılan yöntemler sırasıyla şunlardır: Dendrogramın kullanıcı tarafından, önceden belirli bir eşik değerinde ve yine önceden belirlenmiş bir benzerlik düzeyinde kesilmesi, dendrogramın iki ardışık birleşme arasındaki boşluğun en büyük olduğu kısmın kesilmesi ve birleşmeden kaynaklanan kümenin yoğunluğu belirli bir eşiğin altında olduğunda işlemin durdurulmasıdır (Aggarwal ve Reddy, 2014: 92).

2.6.2.8. Pseudo T² İndeksi

Bu geçerlilik indeksi de Duda indeksine bağlı olarak çalışan bir indekstir. Bunun bir varyasyonu, iki kümenin tek bir kümede birleştirilip birleştirilmemesi gerektiğine karar vermeye çalışırken kullanılabilir. Bu indeks iki kümenin ortalamalarını karşılaştırmak için de kullanılır. Temel olarak, iki kümenin araçları önemli ölçüde farklı değilse, kümeler birleştirilebilir (Johnson, 1998: 329).

Formülü aşağıdaki şekilde gösterilebilir:

$$Pseudo-T^2 = \frac{B_{kl}}{\frac{W_k + W_l}{n_k + n_l - 2}} \quad (2.37)$$

W_k, W_l, W_m Duda indekslerinde bulunan $W_q = \sum_{k=1}^q \sum_{i \in C_k} (x_i - c_k)(x_i - c_k)^T$ olarak tanımlanır.

$B_{kl} = W_m - W_k - W_l, c_m = c_k \cup c_l$ 'ye karşılık gelir.

n_k ve n_l sırasıyla C_k ve C_l kümelerindeki objelerin sayısıdır. En uygun küme sayısı da aşağıdaki en küçük değer şartını sağlayan q küme sayısıdır (Charrad ve diğerleri, 2013:6):

$$Pseudo-T^2 \leq \left(\frac{1 - Duda \text{ İndeksi Kritik Değeri}}{Duda \text{ İndeksi Kritik Değeri}} \right) \times (n_k + n_l - 2) \quad (2.38)$$

2.6.2.9. Beale İndeksi

Beale tarafından geliştirilmiş bir yöntemdir. Bu indekste örneğin birincisi c_1 kümelerinden ve ikincisi c_2 kümelerinden oluşan iki olası kümeleme olduğunu ve $c_2 < c_1$ olduğunu varsayılır. Burada bir diğer ifadeyle ikinci oluşan küme birinciye göre daha az sayıda küme içermektedir. W_1 ve W_2 , küme ortalamalarından hesaplanmış küme içi uzaklıklarının kareler toplamlarıdır. Burada $r=1,2,\dots,c_1$ iken r . kümede n_r adet veri noktası olduğu varsayılır. Eğer x_{rq} r . kümede bulunan q . gözlem vektörünü temsil ettiğinde ve $\bar{x}_r$ de r . kümedeki gözlemleri temsil ettiğinde W_1 için aşağıdaki forma ulaşılır (Beale, 1969: 90-92; Johnson, 1998: 328):

$$W_1 = \sum_{r=1}^{c_1} \sum_{q=1}^{n_r} (x_{rq} - \bar{x}_r)' (x_{rq} - \bar{x}_r) \quad (2.39)$$

W_1 'e benzer şekilde W_2 de yukarıdaki gibi hesaplanabilir. Bu ifadelerin büyüklükleri arasında ne kadar az fark olursa, daha az küme sayısına sahip olan ifade de çok küme sayısına sahip ifade kadar başarılı bir sonuç verir. Bu sebeple daha az küme sayısına sahip kümelerin seçilmesi daha mantıklıdır. Bununla birlikte W_1 , W_2 'den daha küçük olursa ilk kümelenmenin ikincinin üzerinde arttığı ve daha büyük kümelenme sayısına sahip kümeyi seçeceğimiz anlamına gelir. Bu indekste ilk kümelenmenin ikinciden daha iyi kümelenip kümelenmediğini tespit etmek için aşağıdaki pseudo F istatistiği önerilmiştir (Johnson, 1998: 328):

$$F^* = \frac{W_2 - W_1}{W_1} \cdot \frac{(N - C_1)k_1}{(N - C_2)k_2 - (N - C_1)k_1}, \quad k_1 = c_1^{-2/p}, \quad k_2 = c_2^{-2/p} \quad (2.40)$$

Eşitlikte yer alan payda değeri $(N - C_2)k_2 - (N - C_1)k_1$ serbestlik derecesi ve pay için $(N - C_1)k_1$ değeriyle hesaplanan F kritik değerinden F^* istatistiği daha büyük olduğunda ilk kümeleme (daha fazla kümeye sahip) ikinciye (daha az kümeye sahip) tercih edilir (Johnson, 1998: 329).

Beale indeksi gerekli düşünülenden daha büyük bir k değeri ile başlatılır ve azaltmak için aşama aşama işlemler gerçekleştirilir, İstatistiğin türetilmesi sezgiseldir, bu test sadece göreceli bir kılavuz olarak kullanılmalıdır (Seber, 2009: 389).

2.6.2.10. Ratkowsky İndeksi

Ratkowsky ve Lance tarafından önerilmiştir. Formülü aşağıdaki gibi gösterilebilir (Ratkowsky ve Lance, 1978: 115-117):

$$Ratkowsky = \frac{\bar{S}}{q^{1/2}} \quad (2.41)$$

$\bar{S}$ değeri $\frac{B}{T}$ değerine eşittir. Burada B her değişken için kümeler arası kareler toplamı ve T her değişken için toplam kareler toplamıdır. En uygun küme sayısı, *Ratkowsky* 'nin maksimum değerine sahip olduğu q değeridir (Charrad ve diğerleri, 2013: 10).

2.6.2.11. Point-Biserial İndeksi

İlk kez Milligan tarafından 1980 ve 1981 yıllarında önerilmiştir. Bu indekste ulaşılan maksimum değer en uygun küme sayısı olarak kabul edilir. Formülü şu şekildedir (Milligan, 1981: 196):

$$ptbiserial = \frac{(\bar{d}_b - \bar{d}_w)(f_w f_b / n_d^2)^{1/2}}{S_d} \quad (2.42)$$

d_w = Küme içi uzaklıkları toplamı

d_b = Küme arası uzaklıklar toplamı

$\bar{d}_w, \bar{d}_b$ = İlgili değerlerin ortalamaları

S_d = Tüm uzaklıkların standart sapması

n_d = Toplam uzaklıkların sayısı

f_w = Küme içi uzaklıkların sayısı

f_b = Küme arası uzaklıkların sayısı

2.6.2.12. Frey İndeksi

Frey ve Van Groenewoud tarafından yılında önerilmiştir. Formülü aşağıdaki gibi gösterilmektedir (Frey ve Groenewoud, 1972: 875):

$$K = \frac{\bar{d}_{vj+1} - \bar{d}_{vj}}{\bar{d}_{sj+1} - \bar{d}_{sj}} \quad (2.43)$$

$\bar{d}_v$ = ortalama dışsal küme uzaklığı

$\bar{d}_s$ = ortalama küme içi uzaklığı

Frey indeksi sadece hiyerarşik yöntemlere uygulanabilir. Bu indeks, hiyerarşide birbirini izleyen iki seviyedeki fark puanlarının oranıdır. Pay, iki hiyerarşi seviyesinin (seviye j ve seviye $j+1$) her biri arasındaki ortalama dış küme uzaklıkları olan $\bar{d}_v$ 'ler arasındaki farktır. Payda, iki seviyeden (seviye j ve seviye $j+1$) küme mesafelerindeki ortalama arasındaki farktır. Yazarlar doğru küme düzeyini belirlemek için 1.00 oranını kullanmayı önermişlerdir. Oranlar genellikle 1.00'ın üstünde ve altında değişmektedir. En iyi sonuçlar, kümeleme süreci devam ettiğinde, son seriler için oran 1.00'ın altına düşene kadar devam eder. Bu noktada, bu seriden önceki küme düzeyi en uygun bölüm olarak alınmıştır. Oran asla 1.00'ın altına düşmezse, tek küme çözümü varsayılmıştır (Charrad ve diğerleri, 2013: 9).

2.6.2.13. McClain İndeksi

McClain ve Rao tarafından geliştirilmiştir. Formülü aşağıdaki gibi gösterilebilir (McClain ve Rao, 1975: 457):

$$mcclain = \frac{ort\left(\sum_{q=1}^k \sum_i^{n_k} \sum_{j=i+1}^{n_k} d_{ij}\right)}{ort\left(\sum_{k=1}^q \sum_{i \in C_k} \sum_{l=k+1}^q \sum_{j \in C_l} d_{ij}\right)} \quad (2.44)$$

q = küme sayısı

n_k = k 'inci kümedeki obje sayısı $k \in [1 \dots q]$

d_{ij} = i ve j 'inci objeler arasındaki uzaklık

Mcclain indeksinde en düşük değer üreten q küme sayısının en uygun sonuca ulaştığı kabul edilir (Charrad ve diğerleri, 2013: 13).

2.6.2.14. SD İndeksi

Halkidi tarafından önerilmiştir. Formülü aşağıdaki şekilde ifade edilebilir (Halkidi, 2000: 269):

$$Dis(NC_{max})Scat(NC) + Dis(NC) \quad (2.45)$$

$$Scatter(NC) = \frac{1}{NC} \sum_i \|\sigma(C_i)\| / \|\sigma(D)\|, \quad \sigma(C_i): C_i \text{'ye ait varyans vektörü.}$$

NC : Küme Sayısı, D : Veri seti,

$$Dis(NC) = \frac{\max_{i,j} d(c_i, c_j)}{\min_{i,j} d(c_i, c_j)} \sum_i (\sum_j d(c_i, c_j))^{-1}, \text{e karşılık gelmektedir.}$$

$Dis()$ uzaklığı, $Scat()$ ortalama küme içi yoğunluğu ifade etmektedirler. SD indeksi fikri ortalama küme içi yoğunluğu ve kümelerin toplam ayrılması kavramlarına dayanmaktadır. İlk terim, kümelenen objelerin varyanslarına göre kompaktlığı değerlendirir ve ikinci terim, küme merkezleri arasındaki uzaklıklara göre ayırma farkını değerlendirir. SD indeksi bu iki terimin toplamıdır ve SD'nin değeri en aza indirilerek en uygun küme sayısı elde edilebilir (Aggarwal ve Reddy, 2014: 591).

2.7. KÜMELEME YAPISININ ARAŞTIRILMASINDA GRAFİKSEL YAKLAŞIMLAR

Kümeleme analizinin başarısı büyük oranda ilgili araştırmacının yeteneğine bağlıdır ve genel kabul görececek bir istatistiksel teoriye tam anlamıyla sahip bir yöntem değildir. Bu sebeple daha önce bahsedilen geçerlilik indekslerinin yanı sıra bu konuda araştırmacının analizini kolaylaştıracak grafiksel yöntemler bu konuda en sık kullanılan genel yöntemlerin başında gelmektedir. Bu grafikler verilerin dağılımını görsel olarak ifade etmekte ve ilgili verilerin diğer verilere karşı konumlarını tespit ederek analizler için doğru kararlar alınmasına yardımcı olmaktadır. Saçılım grafikleri, ana bileşen skor grafikleri, Andrew grafiği gibi gösterimler bu grafiklere örnek olarak verilebilir.

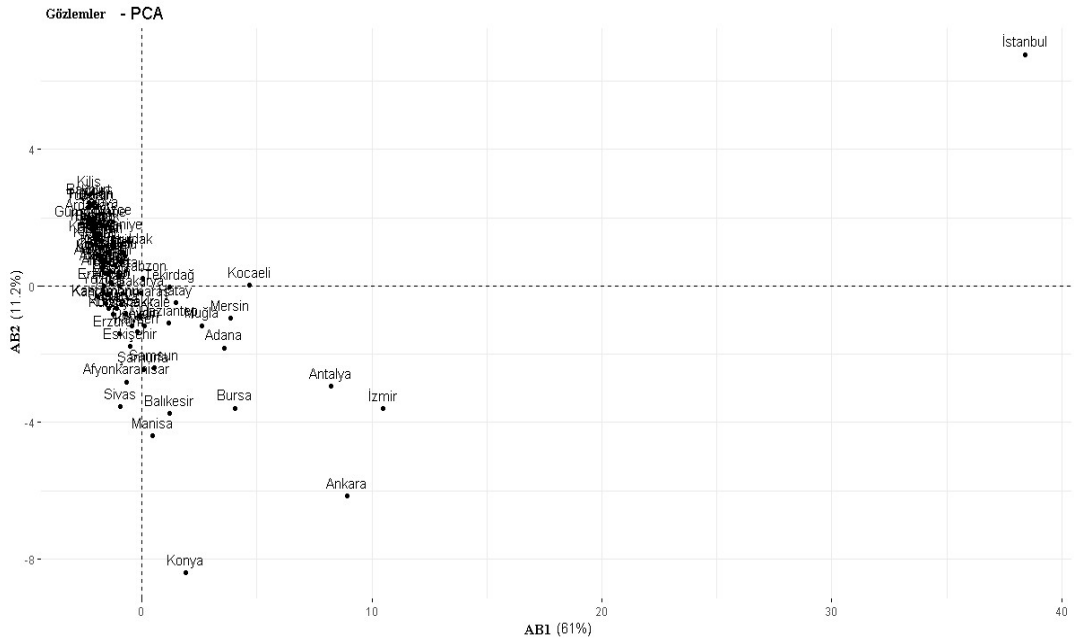
2.7.1. Saçılım Grafiği

En uygun şekilde iki değişkenli veri setleri için kullanılan grafik şeklidir. Görsel açıdan verilerin dağılımları, oluşturdukları gruplar ve yoğunlaşmanın yaşandığı bölgeleri gösterebilmeleri açısından eşsiz bilgiler sağlarlar. Olası grupların tespiti, aykırı gözlemler gibi önemli bilgiler rahatlıkla tespit edilebilir. En yaygın ve etkin kullanıldığı alanlardan birisi paralel ve meridyen konumlarının işaretlenerek konum bilgisinin kullanılabildiği coğrafi araştırmalardır.

2.7.2. Ana Bileşen Skorları Grafiği

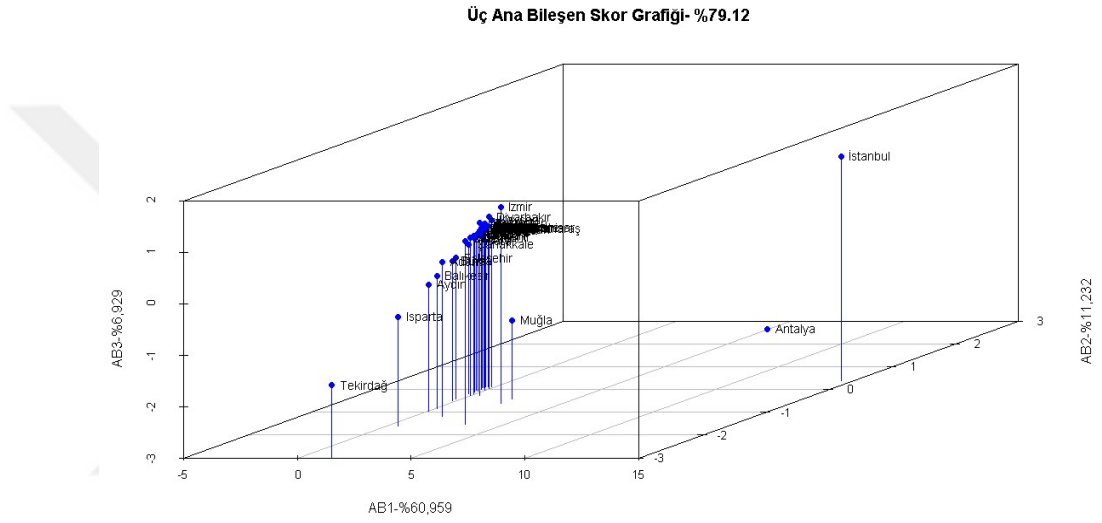
Boyut indirgeme amacıyla kullanılan ana bileşenler analizi kümeleme analizi için veri setlerinin grafiksel gösteriminde de kullanılabilir. Burada önemli olan toplam değişkenliğin yeterince iki veya üç ana bileşen skorunda açıklanabilir olmasıdır. Bu oran ne kadar yüksek olursa doğru sonuçlar gösterebilme ihtimali o denli yüksek olacaktır (Alpar, 2017: 331). Aşağıda yer alan Şekil 15'te olduğu gibi ilk iki ana bileşen yardımıyla iki boyutlu saçılım grafiği elde edilebilirken üç ana bileşen skorları kullanarak bu grafiği üçüncü boyuta taşımakta seçenekler arasındadır.

Şekil 15: İllerin İlk İki Ana Bileşen Skorlarının Kullanımıyla Oluşan Grafiği



Üç boyutlu grafikler artık çoğu istatistiksel yazılımdan doğrudan elde edilebilir ve bunlar bazen küme yapısının kanıtı aramaya yardımcı olabilir. Örnek olarak, Şekil 16’da ilk üç ana bileşen skorları kullanılarak üç boyutlu uzayda Türkiye’deki 81 il verilerinin bir grafiği gösterilmektedir. Şekilde İstanbul, Antalya gibi birçok büyükşehir statüsündeki illerin sapan gözlemler olduğu net bir şekilde görüntülenebilmektedir.

Şekil 16: Ana Bileşen Skorlarının Kullanımıyla Oluşan Üç Boyutlu Grafik Örneği



2.7.3. Diğer Gösterimler

Kümeleme analizinde aykırı gözlemlerin ortaya çıkarılmasında en etkili yöntemlerden biri olan Andrew grafiği yöntemi, üç boyutlu yöntemler kullanan bubble grafikleri, chernoff yüzleri grafiği, radar grafiği burada örnek olarak verilebilir. Bu grafikler, tüm uygulamada önemli bir yere sahip olan ana bileşenleri kullanarak kümelenme yapısının ortaya çıkarılmasına yardımcı olur (İncekırık, 2005: 38).

Chernoff Yüzleri: İlk kez Herman Chernoff tarafından tanıtılmış bir grafik yöntemidir. Her bir nesne üzerinde gözlemler bir insan yüzüyle temsil edilir (Mardia ve diğerleri, 1995: 20). Burun, göz boyutu, yüz şekli ile her bir değişkenin farklı biçimlerde temsil edilmesiyle oluşturulur. Bu yüzlerin şekli kullanılan değişkenlere göre farklı biçimler alabilir (Seber, 2009: 128).

Yıldız Grafiği: Pasta dilimlerine benzer. Radar grafiği olarak da adlandırılabilir. Her bir veri biriminin $p > 2$ değişkenleri üzerine negatif olmayan gözlemlerden oluştuğunu varsayıldığında iki boyutta, çemberin merkezinden yayılan p eşit aralıklı ışınlarla sabit (referans) yarıçaplı bir daire oluşturulabilir. Işınların uzunlukları değişkenlerin değerlerini temsil eder. Işınların uçları bir yıldız oluşturmak için düz çizgilerle birleştirilebilir. Her yıldız çok değişkenli bir gözlemi temsil eder ve yıldızlar (öznel) benzerliklerine göre gruplandırılabilir ve standartlaştırma işlemi yapılabilir (Johnson ve Wichern, 2007: 26).

Köpük (Bubble) Grafiği: Verilerin 3. boyutunun görüntülenmesini sağlayan bir grafik gösterimi biçimidir. 2 boyutu bir düzlem üzerinde gösterilen gözlemler, üçüncü boyutunun gösterilebilmesi için gözlemlerin grafikte gösterildiği daire yarıçapları çeşitli boyutlarda değişen şekilde çizilir (Bulut, 2018: 80).

2.8. KÜMELEME ALGORİTMASININ SEÇİMİ

Kümeleme analizinde algoritma seçimi oldukça esnek bir durumdur. Hiyerarşik ve hiyerarşik olmayan kümeleme analizi yöntemlerinden birisi kullanılabileceği gibi iki yöntemin de dezavantajlarının kapatabileceği şekilde birlikte kullanılabilmek seçeneği de mevcuttur. Bu konuda kesin biçimde çizilmiş kurallar yoktur. Burada kullanılan yöntemlerin yapısı, uzaklık ölçüsünün seçimi, verilerin standartlaştırılması konusunda karar verme veya veri setinin durumu gibi konularda araştırmacının önsel bilgi sahibi olması önem kazanmaktadır. Hiyerarşik yöntemler arasında bir karşılaştırma yapılmak istenmesi durumunda kofenetik korelasyon katsayısı değeri kullanımı da literatürde bilinen ve bu amaçla kullanılan yöntemlerden birisidir.

Grupların oluşturulması için seçilen yöntem uygulandıktan sonra grupta kalitesini doğrulamak için kofenetik korelasyon katsayısı (CCC) kullanılabilecek yöntemlerden birisidir. Sokal ve Rohlf tarafından 1962 yılında ortaya atılan CCC, söz konusu veri tabanının sınıflandırmasının uyum derecesini ve çeşitli kümeleme analizi yöntemlerinin verimliliğini değerlendirmek için bir ölçüt olarak çalışmalarda yaygın olarak kullanılmaktadır. Araştırmalar kofenetik korelasyon katsayısının genellikle ortalama bağlantı yönteminde en iyi sonuçları verdiğini göstermektedir (Carvalho ve diğerleri, 2019: 6). Bir Araştırmacı farklı dendrogramları kıyaslamak zorunda kaldığında, hangi yöntemin orijinal benzerlik matrisinde yer alan bilgilerin en az bozulmaya sebep olduğunu bilmek isteyebilir.

Sokal ve Rohlf, her bir kümeleme analizi yöntemi için, bu bozulmadan yola çıkarak orijinal benzerlik matrisi ile dendrogramdaki kofenetik değerler arasındaki korelasyonun hesaplanmasını önermiştir (Lessig, 1972: 83). Kofenetik değer bir dendrogramın sol kısmındaki yükseklik iki küme arasındaki benzemezlik veya uzaklık ölçüsünü ifade etmektedir. Bu değer azaldıkça kümelerdeki benzerlikte artmaktadır (Bulut, 2018: 384).

Hiyererşik kümeleme yöntemlerindeki dendrogram birleşim noktalarıyla uzaklık ölçü matrisi arasındaki ilişkiyi ölçen kofenetik korelasyon katsayısında Farris, Sneath ve Sokal gibi araştırmacılar 0.95 gibi yüksek bir değer için söz konusu yöntemde verilerin kabul edilebilir bir uyum içinde olduğunu, fakat 0.6 veya 0.7 gibi daha düşük değerlerde ise söz konusu yöntemde şüphe ile yaklaşılması gerektiğini belirtmişlerdir (Seber, 2009: 371). Bir kofenetik korelasyon katsayı değeri ne kadar yüksek çıkarsa o yöntemde ait kümeleme analizi sonuçlarının o kadar iyi olduğu kabul edilir.

Aşağıdaki korelasyon formülü uygulandığında elde edilecek değer kofenetik korelasyon katsayısıdır. Dendrogram birleşim değeri yüksekliklerinin t , uzaklık değerlerinin x olarak gösterildiği bir veri grubunda x ve t tekil simgeler $x(i, j)$ ve $t(i, j)$ değerlerinin genel ortalamalarını ifade etmektedir (Saraçlı ve diğerleri, 2013: 2).

$$c = \frac{\sum_{i < j} (x(i, j) - \bar{x})(t(i, j) - \bar{t})}{\sqrt{[\sum_{i < j} (x(i, j) - \bar{x})^2][\sum_{i < j} (t(i, j) - \bar{t})^2]}} \quad (2.45)$$

2.9. KÜMELER İÇİN İSTATİSTİKSEL ÖNEM TESTLERİ

Bir veri setinin içeriğinin incelenmesi kümeleme analizinin doğru yapılabilmesi için daha önce bahsedildiği gibi vazgeçilmez bir konumdadır. Bunu yaparken seçilen değişkenlerin üretilen kümelerdeki gözlemleri ne kadar başarılı ayırt ettiği mutlak olarak etkileyeceği unutulmamalıdır. Bu konuda yapılması gereken işlerden birisi de bu testlerin gidişatını etkileyecek değişkenlerin normal dağılımlı olup olmadığının tespit edilmesidir (Model tabanlı kümeleme analizi yöntemi uygulanacağı zaman verilerin normal dağılımdan geldiği kabul edilir). Normal dağılımlı olup olmama durumunun tespit edilmesinde kullanılabilecek en güçlü yöntemlerden birisi değişkene ait gözlemlerin çarpıklık ve basıklık

değerlerinin incelenmesidir. Eğer çarpıklık ve basıklık değerleri -1 ile +1 aralığındaysa o gözlemlerin normal dağılım gösterdiği söylenebilir (Hair ve diğerleri, 2014: 34). Değişkendeki gözlemlerin normal dağılımlı olup olmamasına göre varyans analizi veya Kruskal-Wallis testiyle değişkenlerin kümelemeleri ayırmadaki başarısı ölçülebilir ve standartlaştırma işlemi buna göre doğru olarak ayarlanabilir. Bu testlerde ölçülen ortak durum ise gruplar arasında en az bir farkın oluşup oluşmamasıdır. ANOVA veya Kruskal-Wallis testlerinden hangisinin uygulanacağı konusunda ise aşağıdaki 6 kıstas göz önüne alınır. Bu kıstaslardan herhangi birisi sağlanmıyor ise Kruskal-Wallis testi uygulanır (Höft ve diğerleri, 1999: 22-25):

1. Standartlaştırılmış veya dönüşmüş verilerin normal dağılıyor olması;
2. Varyans mesafelerinin eşit olması (varyansların homojenliği);
3. Değişkenler arasında önemli bir etkileşimin olmaması;
4. Grup ortalamaları ve standart sapmaların bağımsızlığı (yani grup ortalamalarının boyutunun, standart sapmalarının boyutuyla ilişkili olmaması);
5. Verilerde sapan değerlerin olmaması (sapan değerler, güvenilirlik kontrollerinin ardından analizden çıkarılabilir);
6. Farklı kategorilerdeki (hücreler) gözlem sayısının eşit olması (zorunlu değildir).

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. ÇALIŞMANIN AMACI

Kümeleme analizi yöntemleri, kendi içlerinde homojen aralarında heterojen küme yapıları oluşturarak araştırmacılara veri setini incelemeyi kolaylaştırıcı ve sağlıklı yorumlama yapılabilme imkânı sağlamayı amaçlamaktadır. Kümeleme analizi yöntemlerinin geçmişi 1940'lı yıllara kadar uzanmaktadır ve çeşitli veri setlerine uygulanışıyla ilgili başarılı çalışmalar literatürde mevcuttur. Ayrıca Türkiye'deki iller ile ilgili literatürde, ekonomi, sağlık, eğitim, kültür, ziraat gibi alanlarda kümeleme analizi yöntemleri kullanılarak yapılmış çeşitli çalışmalar bulunmaktadır.

Bu çalışmada, 2004-2018 yılları arasındaki ulaştırma faaliyetlerinden elde edilen 40 adet değişken dikkate alınarak Türkiye'deki mevcut 81 ilin kümelenme yapısının ortaya konulması amaçlanarak kümeleme analizi uygulanmıştır. Ayrıca bu çalışma ile tamamen açık kaynak kodlu ve ücretsiz bir yazılım olan R programının kümeleme analizi yöntemleri için etkin bir biçimde kullanılabileceğinin gösterilmesi hedeflenmiştir. R programının, birçok istatistiksel analizde kullanılan STATA, MATLAB, SPSS, SAS, MINITAB gibi açık kaynak kodlu olmayan ve lisanslı kullanım gerektiren programlar kadar güçlü bir istatistiksel bir program olduğu söylenebilir. Bu tip açık kaynak kodlu ve ücretsiz yazılımlar, eğitim ve araştırma alanlarında kullanımlarının yaygınlaşması ile ekonomik açıdan da oldukça büyük faydalar sağlayabilir.

Mevcut çalışmada, öncelikle ulaştırma türlerinin çerçevesi çizilerek; karayolu, denizyolu, demiryolu, havayolu ve boru hatları ana başlıkları içindeki mevcut göstergelerden Türkiye'deki 81 ile ait veriler elde edilerek 81x40 boyutlu bir veri seti oluşturulmuştur. Daha sonra oluşturulan bu veri setiyle ilgili bazı tanımlayıcı istatistikler elde edilmiştir. Ayrıca, veri setinin normal dağılıma uygunluğu çarpıklık ve basıklık değerlerine bakılarak incelenmiştir. Bu incelemeden sonra da veri setindeki farklı ölçü birimlerinin varlığı nedeniyle yapılacak standartlaştırma işleminin şekli ve kullanılacak istatistiksel önem testine karar verilmiştir. Bunun sonucunda, 0-1 aralığında min-maks standartlaştırma tekniği ve veri setindeki değişkenlerin tamamının normal dağılımlı olmadığının tespit edilmesi

sebebiyle parametrik olmayan bir istatistiksel yöntem olan Kruskal-Wallis testinin uygulanmasına karar verilmiştir. Ayrıca verilerdeki kümelenme yapısını görsel olarak inceleyebilmek amacıyla ana bileşenler analizi kullanılarak ana bileşen skorları elde edilmiş ve bu ana bileşen skorları ile iki ve üç boyutlu grafikler oluşturulmuştur.

Çalışmada, veri setinin ve grafiklerin oluşturulup standartlaştırma biçiminin belirlenmesinin ardından klasik kümeleme analizi anlayışına uygun olarak altı adet hiyerarşik kümeleme analizi yöntemi (tek bağlantı, tam bağlantı, ortalama bağlantı, ward, ağırlık merkezi, medyan), iki adet de hiyerarşik olmayan kümeleme analizi yöntemi (k-ortalamalar, k-medoid) kullanılmıştır. Küme sayısını belirlemek amacıyla da yaygın olarak kullanılan 14 adet küme geçerlilik indeksi uygulamaya dahil edilmiştir. Veri seti, illerin bilinen bazı özellikleri ve grafiklerde görülen sapan gözlem özellikleri gibi sebeplerle küme geçerlilik indekslerindeki uygulanacak limitleri minimum küme sayısının 5 ve maksimum küme sayısının da 12 olduğu kabul edilerek analizler yapıldığında veri seti için uygun küme sayısının 5 olması gerektiğine karar verilmiştir.

İstatistiksel yöntemlerle yapılan analizlerde açık kaynak kodlu ve istatistiksel olarak geniş bir kullanım alanına sahip önu açık bir yazılım olan R 3.5.1 programı kullanılmıştır. R 3.5.1 programında çalışma için gerekli analizlerin yapılabilmesi R 3.5.1 programı için kullanılabilen bazı paketlerden faydalanılmıştır. Kullanılan bu paketler ve bazı özellikleri şu şekildedir:

factoextra: 'PCA' (Ana Bileşen Analizi), 'CA' (Uygunluk Çözümlemesi), 'MCA' (Çok Boyutlu Uygunluk Çözümlemesi), 'FAMD' (Karmaşık Verilerde Faktör Analizi), farklı R paketlerinden 'MFA' (Çoklu Faktör Analizi) ve 'HMFA' (Hiyerarşik Çoklu Faktör Analizi) fonksiyonlarının işlenmesinde kullanılabilir (Kassambara ve Mumbi, 2017: 1).

readxl: Excel dosyalarının R programında işlem yapılmasını sağlar (Wickham ve diğerleri, 2019: 1).

scales: veri ölçeklendirilmesi için kullanılan pakettir (Wickham ve Seidel, 2016: 1).

NbClust: Tek bağlantı yöntemi, tam bağlantı yöntemi, ortalama bağlantı yöntemi, medyan yöntemi, ağırlık merkezi yöntemi, ward yöntemi ve k-ortalamalar yöntemi olmak üzere 7 kümeleme analizi yöntemi için 26 geçerlilik indeksi yöntemiyle uygun küme sayısı belirleme paketidir (Charrad ve diğerleri, 2013: 1).

moments: Çarpıklık ve basıklık değerlerini ölçmek için kullanılan pakettir (Komsta ve Novomestky, 2015: 1).

scatterplot3d: 3 boyutlu olarak veri görüntülemek için R programında kullanılan pakettir (Ligges ve diğerleri, 2018: 1).

psych: Uzaklık ölçülerini hesaplar. Çeşitli fonksiyonları yapısal eşitlik modellemesi için yararlı bir ön uç görevi görür. Bir veri seti için örnek sayısı, minimum, maksimum değerler, ortalama, varyans, standart hata gibi temel istatistikler Tablo şeklinde bu paketten hesaplanabilir (Revelle, 2015: 1).

cluster: Çalışmada hiyerarşik kümeleme analizi yöntemlerinin R programında ayrıntılı olarak yapılmasını sağlayan pakettir (Maechler ve diğerleri, 2019: 1).

stats: Birçok istatistiksel analiz ve grafik çıktıları sağlayan pakettir. Çalışmada k-ortalamalar yöntemiyle ilgili bazı fonksiyonları kullanılmıştır (Team, 2013: 193).

Sonuç olarak daha önce sağlık, eğitim, ekonomi gibi farklı göstergelerle iller bazında kümeleme analizi kullanılarak incelenen bazı çalışmalardan farklı olarak, bu çalışmada ulaştırma kavramına özgü faaliyetler ele alınarak, illerin birbirine göre benzerlik ve farklılıklarını ortaya çıkarıp gelişim durumlarının ele alınması amaçlanmıştır. Bu sonuçlar ışığında da oluşan benzerlik ve farklılıkların sebeplerinin yorumlanması, çalışmanın diğer amaçlarından birisidir.

3.2. LİTERATÜR TARAMASI

Bu bölümde kümeleme analizi konusunda daha önce yapılmış bazı çalışmalara değinilecektir. Bu çalışmalarda çok değişkenli istatistik alanında kullanılan kümeleme analizi yöntemleri esas alınmıştır. İncelenen çalışmalar, bölge (il, ülke vs.) sınıflaması veya ulaştırma kavramıyla ilgili bir kümeleme analizi çalışmasına yer vermiş olanlardan 2004-2020 yılları arasında yapılanlardan seçilmiştir. Bunlara ek olarak araştırmada kullanılan yazılımlar da ele alınmıştır. Tablo 5'te görüleceği üzere 14 adet yerli çalışma, 17 adet de yabancı çalışma ele alınmıştır. Ele alınan bu çalışmalarda yerli olanlarda 11 tanesinde SPSS, 3 tanesinde R programı kullanılmıştır. İncelenen yabancı çalışmalarda ise 8 adet SPSS, 3 adet R, 2 adet MATLAB, 2 adet Arcgis, birer tanesinde Stata, XI stat, C++ yazılımı kullanılmıştır. İncelenen çalışmaların tamamında klasik kümeleme analizi yöntemleri arasında bulunan hiyerarşik veya hiyerarşik olmayan kümeleme analizi yöntemlerinden birisi mutlaka kullanılmıştır.

Tablo 5: İncelenen Çalışmalarda Kullanılan Yazılımlar

YERLİ ÇALIŞMALAR		YABANCI ÇALIŞMALAR	
YAZAR	YAZILIMI	YAZAR	YAZILIMI
Atalay, Tortum-2013	SPSS	Allonso, Monzon, Cascajo-2015	SPSS
Bulut-2019	R	Baculakova-2018	XL Stat
Cengiz, Öztürk-2012	SPSS	Gitto, Mancuso-2019	R
Çakmak, Uzgören, Keçek-2005	SPSS	Haustein, Nielsen-2016	SPSS
Çelik, Şengül, İnci-2019	SPSS	Klinger, Kenworthy, Lanzerdorf -2013	SPSS
Er-2004	R	Lachapelle, Burke, Brotherton, Leung-2018	Stata, ArcGis
Karaoğlu-2018 (YL)	SPSS	Marta-2017	SPSS
Levent-2016 (YL)	SPSS	Mayer-2016	SPSS
Öcalır, Akunal, Erol-2016	SPSS	Nezerenko, Koppel, Tuisk-2017	SPSS
Öztürk, Çiftçi, Hırlak (2017)	SPSS	Park, Lee, Sohn-2019	C++
Sonmez, Er-2007	R	Ren, Li-2018	R
Yalçın, Ayyıldız (2018)	SPSS	Nashri ve Zang-2018	ArcGis
Yaşar-2009 (Doktora Tezi)	SPSS	Cabral, Raos-2014	R
Yılmaz, Uzgören-2013	Minitab, SPSS	Rocha vd.-2020	Matlab
		Wang ve Pham-2020	Matlab
		Wolf, Schröder-2019	SPSS
		Zemp, Stauffacher, Jang, Scholz-2011	SPSS

Alonso, Monzon, Cascajo (2015): 2010 yılı referans alınarak 23 Avrupa şehrinin kentsel ulaşımının durumuyla ilgili Ward yöntemi ve 4 adet küme alınacak şekilde k-ortalamlar yöntemi kullanılarak kümeleme analizi yapılmıştır. Bu analizlerde ulaşımında çevresel, ekonomik ve sosyal etkenlere dayalı olarak değişkenler seçilip özel dönüşümler uygulanmış 9 adet gösterge kullanılmıştır. Sonuç olarak çevresel etkinlik bakımından sürdürülebilir şehirler, sosyal açıdan sürdürülebilirliği en yüksek şehirler, ekonomik sürdürülebilirliği olan ve rekabetçi yapıdaki olan şehirler ve en az sürdürülebilir küçük nüfuslu şehir grupları tespit edilip şehirler bu dört farklı şekilde belirlenmiştir. Bu çalışmada analizlerde SPSS programı kullanılmıştır.

Atalay, Tortum (2013): Bu çalışmada 2012 yılında Türkiye’deki iller bazında trafik hareketliliğini analiz etmek amacıyla KGM’den elde edilen yolcu-km, taşıt-km ve ton-km verileri kullanılarak hiyerarşik olmayan kümeleme analizi yöntemlerinden k-ortalamalar yöntemi kullanılmıştır. k değeri için 6 küme belirlenmiş ve iller bu altı kümeye dağıtılmıştır. Bu çalışma sonucunda İstanbul ve Ankara en yüksek trafik hareketliliğine sahip iller olarak, Antalya, Bursa, Kocaeli, Konya, İzmir ise bu iki ili takip eden ikinci en yüksek trafik hareketliliğine sahip iller olarak belirlenmiştir. Bu çalışmanın analizinde SPSS 20 istatistik programından yararlanılmıştır.

Baculakova (2018): Slovakya’da bölgesel ve ulusal düzeyde destekleyici politikalar üretmek amacıyla kreatif endüstri bölgelerini tanımlamaya dönük kümeleme analizi yapılmıştır. Bu amaçla 2016 yılındaki 8 yöreye ait 79 endüstriyel bölge verisi, çeşitli ekonomik göstergeler ışığında belirlenmiştir. Bu veriler için öklid uzaklıkları hesaplanarak hiyerarşik kümeleme yöntemlerinden birisi olan ward yöntemi uygulanmıştır. Bu yöntemle elde edilen sonuçlar 4 ve 8 adet sınıflamada ele alınmıştır. Sekiz yöre için yapılan dörtlü sınıflamada ilk iki sırayı en iyi birer bölge almıştır. Bu iki bölgenin durumu diğerlerine göre oldukça farklı ve iyi durumda çıkmıştır. İkinci analizde ise 79 endüstriyel bölgenin tamamı 8 kümeli bir sınıflamaya tabii tutulmuş ve bu sınıflamalara göre bölgeler gruplara ayrılmıştır. Buradaki sonuçlarda da ilk analizdeki yöreye ait endüstriyel bölgeler yine ilk analize paralel durumda ön planda çıkmıştır. Bu çalışmada gerekli analizler için XL Stat yazılımından faydalanılmıştır.

Bulut (2019): 2015 yılı yaşam endeksine göre illerin sınıflandırılmasına yönelik yapılan kümeleme analizi çalışmasında TÜİK tarafından oluşturulan 11 adet gösterge ışığında hiyerarşik olmayan kümeleme yöntemlerinden EM ve k-Ortalamalar kümeleme algoritmaları kullanılmış, bu algoritmalarda sırasıyla 2 ve 5 adet küme oluşturulmuştur. Analiz sonucunda özellikle doğu ve güneydoğu illeri diğer illere göre aynı sınıfta ve birbirlerine benzer durumda olduğu gözlenmiştir. Çalışmada açık kaynak kodlu ve istatistik alanında yaygın olarak kullanılmaya başlanan R programından yararlanılmıştır.

Cabral, Raos (2014): Bu çalışmada 2009 yılında Brizalya’nın 17 liman bölgesinin performans durumunu ziyaret eden gemiler, liman ve konteyner özellikleriyle fiyatlamaları kapsayan 9 farklı değişken kullanılarak hiyerarşik kümeleme yöntemleri uygulanmış ve 3 seviyede sınıflama yapılmıştır. Yapılan bu analiz sonuçlarında en yüksek performansa sahip limanın Sao Pulo bölgesi Tecon

terminalindeki Santos liman bölgesi olduğu belirlenmiştir. Çalışmadaki analizlerin gerçekleşmesinde R programından faydalanılmıştır.

Cengiz, Öztürk (2012): Bu çalışmada illerin eğitim performansını belirlemek amacıyla TÜİK verileri esas alınarak 2010 yılı Türkiye'nin 81 ili için 8 farklı eğitim göstergesi belirlenmiştir. Bu veriler üzerinde hiyerarşik olmayan yöntemlerden k-ortalamalar yöntemi uygulanmış ve sınıflamak için uygun 6 küme belirlenmiştir. Yapılan analiz sonucunda doğu ve güneydoğu illerinin durumunun diğer illere göre oldukça kötü bir durumda olduğu tespit edilmiş ve diğer gruplara nazaran bir bütün olarak ayrı bir sınıflamada yer almışlardır. Diğer gruplarda ise eğitim performansı bakımından coğrafi ve sosyoekonomik sınırlar oluşmamıştır. Çalışmanın uygulama aşamasında SPSS 17 programı uygulanmıştır.

Çakmak, Uzgören, Keçek (2005): Bu makale çalışmasında 1990 ve 2000 yılında iki dönem şeklinde 73 adet ilin kültürel açıdan gelişmişliklerini belirlemek amacıyla hiyerarşik kümeleme (Ward ve tek bağlantı yöntemleriyle kareli öklid uzaklığı) ve k-ortalamalar yöntemiyle kümeleme analizi uygulanmıştır. Kümeleme analizinin yapılması için SPSS 10 istatistik programından faydalanılmıştır. Veri seti için 10 adet değişken kullanılmış ve bunlar sırasıyla illerdeki matbaa sayıları, kitap sayıları, 6 yaş üstü nüfus, müzelerdeki kitap sayıları, müzelere bağlı taşınmaz kültür varlıkları sayısı, tiyatrolardaki koltukların toplam sayısı, turneye çıkan tiyatrolara giden 6 yaşın üstündeki toplam seyirci sayısı, sinemalardaki toplam koltuk sayısı, sinemaya giden 6 yaş üstü seyirci sayısı ve müzelerde kayıt altındaki toplam kitap sayısı olarak belirlenmiştir. Analiz sonucunda ise İstanbul ilinin diğer illere göre oldukça farklı bir kültürel konuma sahip olduğu, İzmir ve Ankara'nın genel olarak benzer kültürel durumda oldukları, bölgesel ve iktisadi gelişmişlik yönünden illerin farklılıklarının kültürel açıdan da farklılığı etkilediği belirlenmiştir.

Çelik, Şengül, İnci (2019): Türkiye'de illere göre bitkisel ve hayvansal ürünlerin incelenmesi amacıyla 2017 yılı 11 adet değişken kullanılarak kümeleme analizi yapılmıştır. Bu analiz için hiyerarşik olmayan kümeleme yöntemlerinden birisi olan k-ortalamalar yöntemi ve hiyerarşik yöntemler kullanılmış ve k değeri 5,7 ve 10 alınarak yapılmıştır. Bu analiz sonucunda iller tarımsal değişkenlere göre bölgelerle sınıflandırılmıştır. İncelemede Konya ili hem hiyerarşik hem de hiyerarşik olmayan yöntemlerde hep ayrı bir bölge olarak ortaya çıkmıştır. Bu çalışmanın analizi yapılırken SPSS 23 istatistik programından faydalanılmıştır.

Er (2006): Bu makale çalışmasında 2004 yılı Türkiye'ye gelen turistlerin kullandıkları havayolu, denizyolu, karayolu ve denizyolu araçları değişkenleri esas alınarak 79 ülke ele alınmış ve kümeleme analizi yapılmıştır. Hiyerarşik olmayan yöntemlerden birisi olan k-medoid yöntemi kullanılarak belirlenen 3 grup ışığında benzer yapıdaki gruplar belirlenerek sınıflama yapılmıştır. Bu veriler ışığında yapılan iki adet kümeleme analizi işlemi sonuçlarına göre ilk analizde Türkiye'ye en çok turist gönderen ülkeler Almanya, Hollanda, Rusya, İngiltere öne çıkmıştır. Bu ülkelerin çıkarılmasıyla yapılan ikinci kümeleme analizi sonucunda ise Türkiye'ye yakın ve komşu olan ülkeler (Azerbaycan, İran, Bulgaristan, Gürcistan, Yunanistan, Suriye) ayrı bir sınıf oluşturarak öne çıkmışlardır. İlk analizde öne çıkan ülkeler ağırlıklı olarak havayolu kullanırken, ikinci analizdeki Türkiye'ye komşu ülkeler ağırlıklı olarak karayolunu kullanmışlardır. Çalışmada R programından faydalanılmıştır.

Gitto, Mancuso (2019): Bu çalışmada havaalanlarının Dünya üzerindeki marka algıları konusunda bilgi edinilmesi amacıyla sosyal medya bağlantıları verileri kullanılarak 46 ülkeden 118 havaalanı seçilmiştir. 2016-2018 yılları arasında toplanan verilerle hiyerarşik olmayan kümelem yöntemlerinden birisi olan k-ortalamlar yöntemiyle kümeleme analizi yapılmıştır. Bu analiz için havaalanı endüstrisi için önemli kabul edilecek lüks imkânların mevcudiyeti, çevre faktörü, engelli kişiler için sağlanan kolaylıklar konusundaki göstergeler dikkate alınmıştır. Analiz sonucunda bu havaalanları, belirlenen 5 küme içersinde sınıflandırılmıştır. Bu 5 küme içersinde çevresel özellikler bakımından yüksek bir ortalama puanıyla öne çıkan 23 adet gözleme sahip kuzey Amerika havaalanları kümesi ve yüksek engelli dostu değeriyle öne çıkan 8 adet havaalanının barındığı Avustralya ve İngiltere kümesi benzer özellikleriyle ön plana çıkmışlardır. Coğrafya da kümelerde etkisi olan bir konumdadır. Çalışmadaki analiz işlemleri için R programından faydalanılmıştır.

Haustein, Nielsen (2016): Bu makale çalışmasında Avrupa'nın ulaşım karakteristiğini belirleme amacıyla 28 AB ülkesinde 2014 yılı 15 yaş üstündeki yaklaşık 28000 kişiyle görüşme gerçekleştirilmiş ve kullandıkları ulaşım araç ve sistemleriyle ilgili sorular sorulmuştur. Bu görüşmeler sonucu toplanan veriler hiyerarşik olmayan kümeleme yöntemlerinden birisi olan k-ortalamlar yöntemiyle analiz edilmiş ve 6 küme tespit edilmiştir. Bu kümelerde tanımlanan karakteristikler arasında konforu ön planda tutan ulaşım aracı kullanıcılarının yoğunlukta olduğu

İngiltere, Fransa, İtalya kümesi; Fiyat odaklı, ekonomik olarak çok zengin olmayan toplu taşımanın yoğun olduğu İber ve doğu Avrupa ülkeleri, Bisiklet kullanımının yoğun olduğu çevreci taşıtlar kullanan Batı Almanya, Hollanda, Belçika'nın bulunduğu ülkeler grubu; toplu ulaşımın zayıf olduğu ada ülkeleri olan Kıbrıs, İrlanda kümesi; tüm seyahat imkânlarında çevreci bilincin yüksek olduğu İsveç, Doğu Almanya, Avusturya ülkeleri kümesi belirlenmiştir. Çalışmadaki kümeleme analizi için SPSS 21.0 programı kullanılmıştır.

Karaoğlu (2018): Bu çalışmada sosyoekonomik yönden benzerlik gösteren illeri belirlemek amacıyla Türkiye İstatistik Kurumu internet sitesinden alınmış sosyo-ekonomik dinamikleri kuvvetli 55 değişken temel alınarak 2002, 2008 ve 2013 verileriyle Türkiye'deki 81 il için yedi farklı klasik yöntemle kümeleme analizi uygulanmış ve 5 küme belirlenmiştir. Ayrıca bu analizlerin istatistiksel olarak birbirleriyle ilişkileri de incelenmiştir. Uygulanan analiz sonuçlarında özellikle iktisadi gelişmişliği benzerlik gösteren illerin durumunda sosyoekonomik açıdan da aynı paralellikte sonuçlara ulaşıldığı gözlemlenmiş, örneğin İstanbul ili genellikle tüm analizlerde farklı şekilde tek bir küme olarak ortaya çıkmıştır. Analiz için gerekli hesaplamalarda SPSS 25 istatistik programından faydalanılmıştır.

Klinger, Kenworthy, Lanzerdorf (2013): Almanya'da bulunan 44 şehir için ulaşım karakteristiklerini belirlemek amacıyla en güncel olarak seçilen 23 gösterge esas alınıp (kentsel göstergeler, sosyoekonomik karakteristikler, ulaştırma altyapı ve tedariki, Ulaştırma talebi göstergeleri / seyahat davranışları, Mobilite ile ilgili algı ve değerlendirmeler) faktör analizi yapılmış, ward ve k-ortalamlar yöntemleri kullanılarak kümeleme analizleri uygulanmış ve 6 küme tespit edilmiştir. Bu kümeler bisiklet şehirleri kümesi, transit metropoller kümesi, otomobil odaklı şehirler kümesi, çok türlü ulaşım potansiyeli olan yürümeye uygun şehirler kümesi ve transit şehirler kümesi olarak adlandırılmışlardır. Çalışmada analizlerin yapılması için SPSS istatistiksel paket programından faydalanılmıştır.

Lachapelle, Burke, Brotherton, Leung (2018): Avustralya'daki 6 yönetim bölgesinden seçilen 45 adet kasa otomatının (koli göndermek için kullanılan) bölgesel konum ve özelliklerini değerlendirip raporlamak amacıyla kümeleme analizi yöntemi uygulanmıştır. Analizde hiyerarşik kümeleme analizi yöntemlerinden biri olan ward yöntemi kullanılmış olup otomatların bölge erişim ve karakteristik özellikleriyle ilgili değişkenler kullanılmıştır. Analiz sonucunda bu 45 adet otomat için 9 adet şehir konumundakiler kümesi, 17 adet banliyö kümesi, 9 adet

banliyölerdeki posta merkezleri kümesi ve 10 adet alışveriş merkezi kümesi olmak üzere dört adet sınıflamada değerlendirilmiştir. Yapılan analiz çalışmalarında STATA 14 programı ve ArcGIS 9.0 programlarından faydalanılmıştır.

Levent (2016): Bu çalışmada Türkiye'deki bir havayolu şirketine ait 2016 yılındaki yolcu sayısı, kalkış, varış bilgileri, uçak tipi, yolcu kapasitesi ve uçağa yüklenen ikram sayısı değişkenleri kullanılmıştır. Bu değişkenler üzerinden yapılan kümeleme analizi çalışmasında da sırasıyla üç ve dört adet küme belirlenerek k-ortalamlar yöntemi kullanılmış ve gözlemler bu kümelere aktarılmıştır. Bu analiz sonucunda fazla ve az ikramlı gözlemler birbirinden ayrıştırılarak gruplandırılmış ve böylece şirketin maliyetlerinin azaltılmasına olumlu katkı yapılması hedeflenmiştir. Bu çalışmadaki analizlerin uygulanmasında SPSS 20.0 programı kullanılmıştır.

Malkoç (2018): Bu çalışmada 2002 yılı Kayseri, Sivas ve Yozgat illerine bağlı 46 ilçeye ait bazı sosyal ve ekonomik göstergeleri ifade eden 15 değişkenlik bir veri seti oluşturulmuştur. Bu veri seti üstünde küme geçerlilik indeksleri yardımıyla 4 adet k küme değeri belirlenmiş; k-ortalamlar yöntemiyle kümeleme analizi uygulanmıştır. Bu çalışma sonucunda küme geçerlilik indekslerinin kümeleme analizi yöntemleri için gerekli olan uygun küme sayısını belirlemede başarılı olduğu ortaya çıkmıştır.

Marta (2017): Bu makalede 2000-2014 yılları arasındaki Dünya'nın çeşitli bölgelerindeki 838 havaalanı ortalama bağlantı yöntemi, 4 küme ve 8 küme ile k-ortalamlar yöntemi, iki aşamalı kümeleme yöntemi ile incelenmiş ve bu incelemede hâkim coğrafi konum, havaalanı sayıları ve taşınan yolcu sayısı esas alınmıştır. Bu analiz sonucunda hava ulaşımı için söz konusu gözlemlerin bölgesel konumlarından ziyade sınıflandırmaların hava trafiğini etkileyen olaylar ve mevsimsel etkilere göre gerçekleşeceği sonucuna ulaşılmıştır. Çalışmadaki analizlerde ise SPSS programından faydalanılmıştır.

Mayer (2016): Bu çalışmada farklı ülkelerden 114 adet havaalanı 2013 yılı için kargo işletme özelliklerine göre öklid uzaklık ve Ward yöntemiyle hiyerarşik kümeleme analizi uygulanmıştır. Bu havaalanlarındaki yıllık toplam kargo hacmi, iş yükü yüzdesi, ticari uçak hareket yüzdelik verileri ve uluslararası kargo yükünün toplam yükteki yüzdelikleri olmak üzere 4 adet değişken seçilmiştir. Bu analiz sonucunda coğrafi konum ve işlevlerine göre sekiz küme belirlenerek bu sekiz kümede havaalanları gruplandırılmıştır. Bu gruplar kargo yükü taşınmaya bağlı olarak işlev gören Avrupa havaalanları, Kargo yüklenen birincil Kuzey Amerika havaalanları, çift yönlü Avrupa Grubu, ikincil Kuzey Amerika kargo havaalanları,

uluslararası birincil aktarmalılar, uluslararası ikincil aktarmalılar, yolcu ağırlıklı uluslararası havaalanları ve yolcu ağırlıklı yerel havaalanlarıdır. İlk bahsedilen 4 grupta merkezi konum olarak daha avantajlı gruplara ait havaalanları toplanmıştır. Analiz ve hesaplamalar için SPSS programından faydalanılmıştır.

Nasri ve Zang (2018): Bu çalışmada metropolitan alanları kentsel yapılarına ve arazi kullanım düzenlerine göre gruplamak ve buna göre politika belirlemek için kümeleme analizi yapılmıştır. ABD’de 2010 yılı verilerine ait 50 yerleşim bölgesi üzerinde 4 farklı alanda (sosyoekonomi ve demografi, ağ ve hedef erişilebilirliği, çeşitlilik ve kentsel tasarım, yoğunluk ve merkezîyet) 50 adet değişken belirlenmiştir. Belirlenen bu veriler ışığında öklid mesafesi kullanılarak hiyerarşik yöntemlerden ortalama bağlantı, tam bağlantı ve ward yöntemiyle hiyerarşik olmayan yöntemlerden k-ortalamalar yöntemi kullanılmış ve kümeleri belirlemek için 3 seviye belirlenmiştir. Bunlar yoğunluğu ve kent erişilebilirliği yüksek yerleşim yerlerinin toplandığı A grubu, genel olarak iş erişilebilirliği yeterli ve orta yoğunluklu gözlemlerin bulunduğu B grubu, son olarak düşük yoğunluklu ve düşük iş ile yürüme erişilebilirliği yeterli olmayan C grubunda kümeler ayrıştırılıp yerleşimler gruplandırılmıştır. Çalışmada kullanılan hesaplamalarda ArcGIS 10.1 programından faydalanılmıştır.

Nezerenko, Koppel, Tuisk (2017): Ulaştırma alanında ortak projeler gibi çalışmalar için kullanılabilecek, ulaştırmada benzer karakteristiklere sahip ülkeleri tespit edip sınıflandırma amacıyla Baltık deniziyle yakından etkileşime sahip 9 ülkede 2004-2011 dönemi arasındaki kara, deniz, demiryolu olmak üzere ayrı veriler ışığında ward yöntemi kullanılarak hiyerarşik kümeleme analizi yapılmıştır. Analiz sonuçlarına göre ulaştırma türlerindeki karakteristiklerde özellikle Almanya ve Polonya’nın tüm ulaştırma türleri için diğer ülkelere göre ciddi farklılıkları bulunduğu tespit edilmiştir. Çalışmadaki analizlerde ise SPSS programından faydalanılmıştır.

Öcalır, Akünal, Erol (2016): Aralarında Türkiye’nin de bulunduğu Dünya Ticaret Örgütü üyesi 148 ülke arasında taşımacılık sektörüyle ilgili bir kümeleme analizi yapılmıştır. Çalışmada Dünya Bankası verilerinden 1998-2010 yılları aralığındaki verilerden yararlanılarak havayolu, denizyolu, demiryolu ve karayolu taşımacılığıyla sağlanan yük, yolcu, ithalat ve ihracat değerlerinin de bulunduğu ulaşım hizmetlerini kapsayan göstergeler kullanılmış, 8 farklı değişkenin kullanımıyla birlikte toplam sayısı 2-5 arasında belirlenen küme sınıflaması değerleriyle ward yöntemi

seçilmiştir. Bu 8 adet ölçüt ışığında uygulanan kümeleme analizleri sonucunda da Türkiye'nin bu sektörde gelişen bir pozisyonda olmasına rağmen güçlü ülkeler sınıfında bulunmadığı sonucuna ulaşılmıştır. Bu çalışmadaki analizler için SPSS programı kullanılmıştır.

Öztürk, Çiftçi, Hırlak (2017): Türkiye'deki illerde kadınların yaşam standartlarının incelenmesi amacıyla yapılan bu çalışmada regresyon analiziyle birlikte TÜİK kaynaklı 2014-2015 yıllarında yayınlanmış 10 adet göstergeden faydalanılarak yedi coğrafi bölge ve iller odaklı hiyerarşik (Ward Yöntemi) kümeleme analizi uygulanmıştır. Analiz sonucu genel olarak ele alındığında Güneydoğu Anadolu Bölgesi ve bu bölgeye bağlı illerin kadınların yaşam standartları yönü itibariyle diğer bölge ve illere göre farklı bir profil çizdiği belirlenmiştir. Çalışmada SPSS programından faydalanılmıştır.

Park, Lee, Sohn (2019): Kore'de başkent Seul şehrindeki sabah oluşan yoğun trafik problemine bir çözüm getirilmesi amacıyla taksilerin 2016 yılı mart ayındaki saat 07:00-09:00 arasındaki hareket vektörlerinden yararlanılarak kümeleme analizi ve tam sayılı programlama yaklaşımı uygulanmıştır. Kümeleme analizinde hiyerarşik olmayan yöntemlerden k-ortalamlar yöntemi uygulanmış ve bu yöntem uygulanırken 10 adet taksi yoğunluğu bulunan alan sınıflandırılmıştır. Analiz sonucu ulaşılan bulgularda ağırlıklı olarak aktarma istasyonları ve okullar gibi konumların bulunduğu, mevcut rotalardan farklı güzergâhlar önerilmiştir. Bu çalışmada gerekli olan algoritmaların kullanılması için c++ programlama dilinden faydalanılmıştır.

Ren, Li (2018): Bu çalışmada Çin ve ABD havayollarındaki ağları analiz etmek maksadıyla kümeleme analizi yapılmış ve bu çalışmada 45 adet ABD ve 40 adet Çin havaalanı çifti dâhil edilmiştir. Ağ kullanım modellerini belirlemek için hiyerarşik olmayan kümeleme yöntemlerinden birisi olan k-ortalamlar yöntemi uygulanmış ve Çin ile Amerikan havaalanları için sırasıyla 4 ve 7 adet küme sayısı belirlenmiştir. Bu kümeler günün belli vakitlerindeki hava trafiği yoğunluğuna göre belirlenmiştir. Analiz sonucunda Amerikan hava ağının Çin hava trafiği ağına göre yoğunluklara karşı daha dayanıklı, esnek bir trafiğe sahip olduğu belirlenmiştir. Çalışmadaki analizlerde ise R programından faydalanılmıştır.

Rocha ve diğerleri (2020): Dünya üzerindeki otobüslerle yapılan toplu ulaşımın planlanmasında kolaylaştırıcı yorumlar getirebilmek maksadıyla 2016 yılına ait 45 ülkeden 206 şehir ve 44 değişkenden oluşan 296 adet otobüs hattıyla ilgili bir veri

seti kullanılmıştır. Bu verilerden yola çıkılarak ward yöntemi ve k-ortalamlar yöntemiyle 4 adet küme sayısı belirlenerek kümeleme analizi uygulanmıştır. Yapılan incelemeler sonucunda otobüsle toplu ulaşım genel olarak temel hatlar (küme 1), geliştirilmiş hatlar (küme 2), hızlı otobüs Transit (BRT), yüksek hizmet seviyeli otobüs (BHLS) sistemleri (küme 3), ekspres ve sınırlı duraklı hizmetler (küme 4) olarak gruplandırılmıştır. Çalışmada matlab programı kullanılmıştır.

Sönmez, Er (2006): Bu çalışmada 2004 yılı için Türkiye'deki illerinde meydana gelen göç hareketleri bulanık kümeleme ve medoid yöntemine göre incelenmiştir. Bu inceleme sonucunda her ne kadar büyük şehirlerden köye göç etme teşvik ediliyor olsa da büyük şehirlerden göç hareketlerinin köylerden ziyade İstanbul, Ankara, İzmir harici diğer şehirlere doğru olduğu tespit edilmiştir.

Wang ve Pham (2020): Bu çalışmada Vietnam'daki 2015 ve 2016 yılları arasındaki uluslararası nitelikte olan ve sağlıklı veri alınabilen 28 havaalanı kullanılmıştır. Öklid uzaklığı ve ward yöntemi kullanılarak bazı performans göstergelerine göre (Check-in işlemi, biniş süreci, personel tutumu, bagajın hatalı kullanımı ve kişisel belgelerin hatalı kullanımı) kümeleme analizi uygulanmıştır. Analiz sonucunda her yıl için çeşitli değerlendirme ölçütlerinde performans analizi için hava alanları birbirine en yakın olacak şekilde 5 farklı gruplandırma ve toplamda 10 adet gruplama yapılmıştır. Bu her yıl için beş farklı grupta havaalanlarının performans değerlendirmelerine dönük hedefe ulaşım yüzdeleriyle artı ve eksi yönlerde çeşitli değerlendirmeler yapılmıştır. Bu değerlendirmelere göre eksiklerin tamamlanarak bu havaalanlarının daha iyi performans sergilenmesi amaçlanmıştır. Çalışmada matlab programının kullanılmış olduğu değerlendirilmektedir.

Wolf, Schröder (2019): Almanya'da İnsanların ulaştırma türleri üzerindeki seyahat davranışlarını incelemek amacıyla yaşları 16-69 arasında değişen yaklaşık 6000 kişiyle kümeleme analizi yapılmıştır. 15 ulaştırma biçimi üzerinden yapılan bu analizde hiyerarşik kümeleme yöntemlerinden ward yöntemi ve hiyerarşik olmayan kümeleme yöntemlerinden k-ortalamlar yöntemi kullanılmış ve 6 temel davranış sınıfı içerisinde gözlem verileri gruplanmıştır. Bu küme sınıfları kamuoyu oluşturan kesimler, inovasyon odaklı yenilikçiler, maliyetçi ve fayda odaklı kesimler, konfor odaklı bireyler, fayda odaklı şehir sakinleri ve son olarak riskten kaçınan gelenekçi kesim grupları olarak tanımlanmıştır. Maliyetçi ve fayda odaklı kesimler, araştırmadaki en kalabalık sınıftır. Çalışmadaki analizlerde ise SPSS programından faydalanılmıştır.

Yalçın, Ayyıldız (2018): Türkiye’de aktif kullanılan 55 adet havaalanında uçak, yolcu ve yük göstergelerinin dâhil olduğu aralık ayı 2017 verileri kullanılarak ward yöntemi ve k-ortalamalar yöntemiyle kümeleme analizi uygulanmış ve bu havaalanları k-ortalamalar yönteminde belirlenen kriterler doğrultusunda birbirlerine en benzer şekilde 6 kümede gruplandırılmıştır. Adana ve Trabzon havaalanları bu profildeki araştırmada diğerlerinden daha farklı bir karakteristik göstererek diğerlerinden ayrılmış ve ikisi tek bir grup oluşturmuşlardır. Çalışmada SPSS programı kullanılmıştır.

Yaşar (2009): Bu doktora tezi çalışmasında 1980-2005 yılları arasında 4 dönem periyodu halinde yabancı doğrudan yatırım yapılarına göre ülkelerin durumunu belirleyip belli gruplarda sınıflamak amacıyla 55 değişken ve 135 ülke belirlenmiştir. Standardize edilen bu verilere hiyerarşik kümeleme yöntemlerinden ward yöntemi uygulanmış ve 5 küme belirlenmiştir. Bu kümeler düşük, alt orta, orta düzey, üst orta düzey ile üst düzey doğrudan yabancı yatırım yapısındakiler olarak isimlendirilmiş ve ülkeler bu gruplamalara dağıtılmıştır. Türkiye bu analizde incelenen dönemlere göre üst orta ve üst düzey yabancı yatırım yapısı olan ülkeler grubunda yer almıştır. Kümeleme analizinin yapılması için SPSS 13 istatistik programından faydalanılmıştır.

Yılmaz, Uzgören (2014): Bu makale çalışmasında Türkiye’deki illerin bankacılık faaliyetleri bakımından performanslarını incelemek amacıyla hiyerarşik olmayan yöntemlerden birisi olan k-ortalamalar yöntemi kullanılmış ve analizde kullanılacak küme sayısı 6 olarak belirlenmiştir. Bu analizde kullanılmak üzere 9 adet değişken belirlenmiş ve bu değişkenler sırasıyla şube başına nüfus, kredi mevduat oranı, il başına kredi, il başına nüfus, şube başına ortalama mevduat, şube başına ortalama kredi, nüfus başına ortalama mevduat ve nüfus başına ortalama değerler olarak belirlenmiştir. Yapılan analiz sonuçlarına göre İstanbul ve Ankara şehirleri özel sektörün ve kamu sektörünün finans merkezleri olmaları nedeniyle özel olarak farklı birer grup oluşturdukları gözlemlenmiştir. Analizlerin uygulanmasında Minitab 14 ve SPSS 13 paket programları kullanılmıştır.

Zemp, Stauffacher, Jang, Scholz (2011): Bu makale çalışmasında İsviçre’de bulunan 1700 adet tren istasyonunun durumunun değerlendirilmesi ve stratejik olanlarının tespiti amacıyla faktör analizi ve kümeleme analizi yapılmıştır. İsviçre Federal Demiryolları kurumundan alınmış verilerle nüfus, tren ve mesafe odaklı 10

adet gösterge hiyerarşik kümeleme analizi yöntemlerinden biri olan ward yöntemi ve karesi alınmış öklid uzaklık kullanılarak incelenmiştir. Yapılan analiz sonucunda 7 küme belirlenmiş ve bu 7 kümeye 1406 istasyon tanımlanmıştır. Bu kümeler en geniş ve merkezi istasyonlar, geniş bağlantıya sahip olan noktalar, orta boyutlu banliyö beslemeler, küçük boyutlu banliyö beslemeler, ufak turist istasyonları, izole edilmiş turizm noktaları ve son olarak uzak yerlere ait istasyonlar olarak sınıflandırılmıştır. Yapılan analiz sırasında SPSS programından faydalanılmıştır.

3.3. KULLANILAN DEĞİŞKENLER VE YÖNTEMLER

Veri seti oluşturmak amacıyla Türkiye İstatistik Kurumu (TÜİK), Boru Hatları İle Petrol Taşıma Anonim Şirketi (BOTAŞ), Deniz Ticareti Genel Müdürlüğü, Devlet Hava Meydanları İşletmesi Genel Müdürlüğü (D.H.M.İ.) ve Karayolları Genel Müdürlüğü (K.G.M.) kurumlarından faydalanılmıştır. Karayolu, denizyolu, demiryolu, boru hattı ve havayolu ulaştırma türlerinden 2004-2018 yılları arasındaki 40 farklı değişkenden oluşturulmuş veri seti kullanılarak klasik kümeleme analizi yöntemlerinden 8 tanesi uygulanmıştır.

Kullanılan veri setinin içeriğindeki değişkenler ve bu veri setiyle ilgili bazı tanımlayıcı istatistikler aşağıda Tablo 6 ve Tablo 7’de belirtilmiştir. Tablo 6’da beş farklı ana başlıkta gruplanmış ulaşım faaliyetlerine ait 40 adet değişken bulunmaktadır. Tablo 6’daki her bir renk değişimi ayrı bir ulaşım faaliyeti ana başlığını temsil etmektedir.

Tablo 6: Kullanılan Değişkenler

V1	(Havayolu Ulaşımı) İniş-kalkış yapan uçak sayısı / Türk
V2	(Havayolu Ulaşımı) İniş-kalkış yapan uçak sayısı / Yabancı
V3	(Havayolu Ulaşımı) İniş-kalkış yapan uçak sayısı / Diğer
V4	(Havayolu Ulaşımı) Yolcu sayısı / İç hat gelen
V5	(Havayolu Ulaşımı) Yolcu sayısı / İç hat giden
V6	(Havayolu Ulaşımı) Yolcu sayısı / Dış hat gelen
V7	(Havayolu Ulaşımı) Yolcu sayısı / Dış hat giden
V8	(Havayolu Ulaşımı) Taşınan yük / İç hat gelen
V9	(Havayolu Ulaşımı) Taşınan yük / İç hat giden
V10	(Havayolu Ulaşımı) Taşınan yük / Dış hat gelen
V11	(Havayolu Ulaşımı) Taşınan yük / Dış hat giden
V12	Havalimanı Sayısı
V13	Sahillerde Mevcut Liman Başkanlığı Sayısı
V14	Limn Başkanlıkları Bazında Toplam İhracat Konteyner Miktarı (TEU)
V15	Limn Başkanlıkları Bazında Toplam İthalat Konteyner Miktarı (TEU)
V16	Limn Başkanlıkları Bazında Toplam Kabotaj Konteyner Miktarı (TEU)
V17	Limn Başkanlıkları Bazında Toplam Transit Konteyner Miktarı (TEU)
V18	Toplam İthalat Yık Miktarı
V19	Toplam Kabotaj Yık Miktarı
V20	Toplam Transit Yık Miktarı
V21	Toplam ihracat Yık Miktarı
V22	Otoyol-Taşıt-Km-Değeri
V23	DevletYolu-Taşıt-Km-Değeri
V24	Otoyol-Yolcu-Km-Değeri
V25	DevletYolu-Yolcu-Km-Değeri
V26	Otoyol-Yık-Km-Değeri
V27	DevletYolu-Ton-Km-Değeri
V28	Yol uzunlukları: Otoyol (km)
V29	Yol uzunlukları: İl ve devlet yolu (km)
V30	Yol uzunlukları: Köy yolu (km)
V31	Motorlu kara taşıtları sayısı: Otobüs
V32	Motorlu kara taşıtları sayısı: Minibüs
V33	Motorlu kara taşıtları sayısı: Otomobil
V34	Motorlu kara taşıtları sayısı: Kamyonet
V35	Motorlu kara taşıtları sayısı: Kamyon
V36	Motorlu kara taşıtları sayısı: Motosiklet
V37	Motorlu kara taşıtları sayısı: Özel amaçlı taşıtlar
V38	Motorlu kara taşıtları sayısı: Traktör
V39	Yol uzunlukları: Demir yolu (km)
V40	Mevcut Konvansiyonel Boru Hattı Sistem Türleri Sayısı (petrol, doğalgaz, maden)

Tablo 7: Veri Setindeki Bazı Tanımlayıcı İstatistikler

Değişkenler	N	Ortalama	Standart Sapma	En Küçük Değer	En Büyük Değer	Uzaklık
V1	81	8936,338	41902,38	0	362624,7	362624,7
V2	81	1897,033	9572,047	0	61724,33	61724,33
V3	81	1672,526	3971,549	0	25317,67	25317,67
V4	81	377746,9	1395694	0	11462625	11462625
V5	81	376822,8	1391248	0	11418987	11418987
V6	81	362507,1	2004288	0	15732527	15732527
V7	81	366499,4	2032342	0	15979628	15979628
V8	81	3722,668	14265,15	0	118408,2	118408,2
V9	81	3722,944	14303,54	0	118698,4	118698,4
V10	81	9542,78	62665,75	0	547864	547864
V11	81	10772,17	71671,38	0	628794,5	628794,5
V12	81	0,595885	0,560292	0	2,066667	2,066667
V13	81	0,618107	1,1572	0	4,333333	4,333333
V14	81	32779,03	137667,7	0	982687,9	982687,9
V15	81	33297,19	143629,8	0	1073160	1073160
V16	81	4658,421	18357,16	0	141052,2	141052,2
V17	81	7707,098	55139,14	0	484274,4	484274,4
V18	81	1018212	2919289	0	17142944	17142944
V19	81	2156514	6448652	0	38521913	38521913
V20	81	541919,5	1513222	0	8270949	8270949
V21	81	591843,5	4348708	0	38733467	38733467
V22	81	197221,9	633567,5	0	5001385	5001385
V23	81	778652,5	691379	57561,13	3037262	2979701
V24	81	629647,5	1968889	0	15428644	15428644
V25	81	2073452	1881367	130909,1	8549220	8418310
V26	81	553325,8	1568221	0	11763577	11763577
V27	81	1858332	1562660	138402,9	7908617	7770214
V28	81	104911,2	259291,2	2147,533	2064338	2062190
V29	81	779,3613	452,1067	129,6	3047,4	2917,8
V30	81	4984,484	8231,503	507,2667	69334	68826,73
V31	81	26,2642	53,99785	0	243,1333	243,1333
V32	81	2514,75	5983,494	43,46667	49007,4	48963,93
V33	81	3176,388	1681,749	164	9264,933	9100,933
V34	81	31677,55	65651,27	1117,133	527025,9	525908,7
V35	81	9287,723	16336,03	383,5333	127027,9	126644,3
V36	81	29919,86	46920,42	196,6667	218574,6	218377,9
V37	81	493,7778	927,0655	61,06667	7215,267	7154,2
V38	81	18573,68	16272,84	46,73333	72526,87	72480,13
V39	81	116,9004	127,8615	0	551,6	551,6
V40	81	1,08642	0,503371	0	2	2

Tablo 7’de görüldüğü gibi her bir değişken için gözlemler arasında bazı tanımlayıcı istatistiklere yer verilmiştir. Bu tanımlayıcı istatistikler il sayısı, ortalama, standart sapma, en büyük-küçük değerleriyle bunların arasındaki uzaklık değerleridir. Bu hesaplamalar R 3.5.1 programında eklenebilen “psych” paketiyle yapılmıştır (Revelle, 2015: 1).

Veri seti için kullanılacak değişkenler belirlendikten sonra bu değişkenlerin içeriklerinin normal dağılımlı olup olmadığını belirlememiz gerekmektedir. Bunun sebebi de ileride değişkenlerin içeriğinin kümelemedeki etkisini araştırarak yöntemle karar vermek ve ağırlık, uzunluk, hacim ölçüleri gibi farklı ölçü birimlerinin bir araya getirildiği bir veri seti oluşturulmuş olması dolayısıyla yapmamız gereken standartlaştırma şeklini buna göre belirlememiz gerekmektedir. Veri setindeki değişkenlerin normal dağılımlı olup olmadığına karar vermek için çeşitli yöntemler bulunmaktadır. Bu konuda uygulanabilecek ideal yöntemlerden birisi de değişkenlerdeki içeriklerin çarpıklık ve basıklık değerlerini araştırmaktır. Bu amaçla R 3.5.1 programında eklenebilen “moments” paketinin içeriğinde bulunan “skewness” ve “kurtosis” fonksiyonları kullanılarak aşağıdaki değerler elde edilmiştir (Komsta ve Novomestky, 2015: 1).

Tablo 8: Değişkenlerin Çarpıklık ve Basıklık Değerleri

Değişkenler	Çarpıklık	Basıklık	Değişkenler	Çarpıklık	Basıklık
V1	7,6629062	64,401099	V21	8,5104906	74,930492
V2	5,7162599	34,764049	V22	5,8339191	42,577878
V3	3,7273617	19,149967	V23	1,6764035	5,335097
V4	6,6017602	51,026009	V24	5,6981285	41,142236
V5	6,5907846	50,892547	V25	1,7299356	5,650127
V6	6,6230102	47,794034	V26	5,0571369	34,114629
V7	6,6393904	48,030367	V27	1,8276765	6,93315
V8	6,7785421	53,350508	V28	6,0447756	43,376707
V9	6,7772335	53,316978	V29	2,0912302	10,138578
V10	8,0858109	69,16371	V30	6,174887	47,688426
V11	8,1584967	70,16661	V31	2,286244	7,761013
V12	0,4401903	2,433192	V32	6,1974769	46,810115
V13	1,9809046	6,018073	V33	1,048528	4,55636
V14	5,1951067	32,168737	V34	5,7727275	41,905137
V15	5,599325	37,317589	V35	5,3029591	36,089358
V16	5,7805873	40,404157	V36	2,2533824	7,624555
V17	8,2114075	70,808655	V37	5,7070644	38,643404
V18	3,6353714	16,784577	V38	1,2186428	4,265728
V19	3,8388104	18,613318	V39	1,3923548	4,906353
V20	3,8978628	18,778491	V40	0,5215559	2,465522

Tablo 8'deki değerlere bakıldığında, değişkenlerin kümeleri ayırmadaki başarısını ölçmede parametrik yöntem olan ANOVA testi yerine alternatif olan Kruskal-Wallis testi kullanılmına yönelmiştir. Küme sayısını her bir değişken için k-ortalamalar yönteminde k adet grup olarak belirlediğimizde bu hiyerarşik olmayan kümeleme yöntemi için kurulacak hipotezler şu şekilde olacaktır:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

H_1 : En az bir tanesi farklıdır.

H_0 hipotezinde beş küme arasında anlamlı bir fark olmadığı, alternatif hipotezde ise en az iki kümenin birbirinden anlamlı derecede farklı olduğu ifade edilmiştir. Veri seti $k=5$ iken $\alpha=0.05$ güven aralığında incelendiğinde sonuçlar Tablo 33'de gösterildiği gibi 40 değişkende sadece V40 değişkeni olan boru hattı sistem sayısı değişkeni sonucu diğerlerinden farklıdır. Onun dışındaki tüm değişkenlerin en az iki kümeyi birbirinden başarıyla ayırabildiği ve H_0 hipotezinin reddedildiği sonucuna ulaşılmıştır.

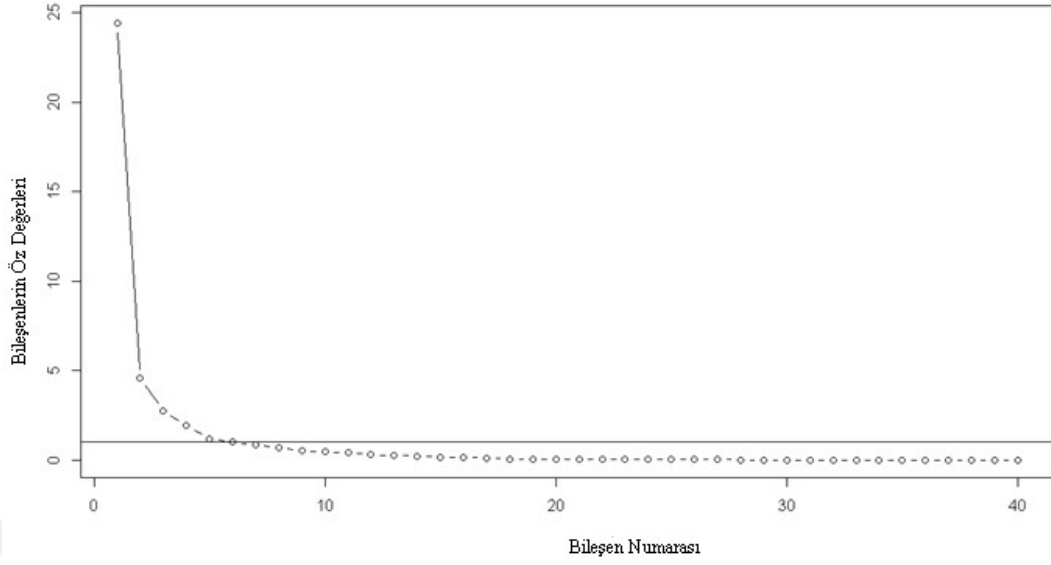
3.4. KÜMELENME YAPISININ GÖRÜLMESİNDE ANA BİLEŞEN SKOR GRAFİKLERİ

Gözlemlerdeki olası küme yapılarını görüntülemek amacıyla 2 boyutlu ve 3 boyutlu grafikler kullanmak araştırmacının karar vermesinde son derece faydalı bir yaklaşımdır. Günümüz programlarında iki ve üç boyutlu grafiklerle veri incelemesi kolaylıkla yapılabilmektedir. Ana bileşen skorları yardımıyla iller verisi incelendiğinde Tablo 9'da ilk iki bileşenin toplam varyansın %72.191'i, ilk 3 bileşenin toplam varyansın da %79,12'sini temsil ettiği görülmektedir.

Tablo 9: Veri Setine Ait Bileşenler için Özdeğer ve Önem Değerleri

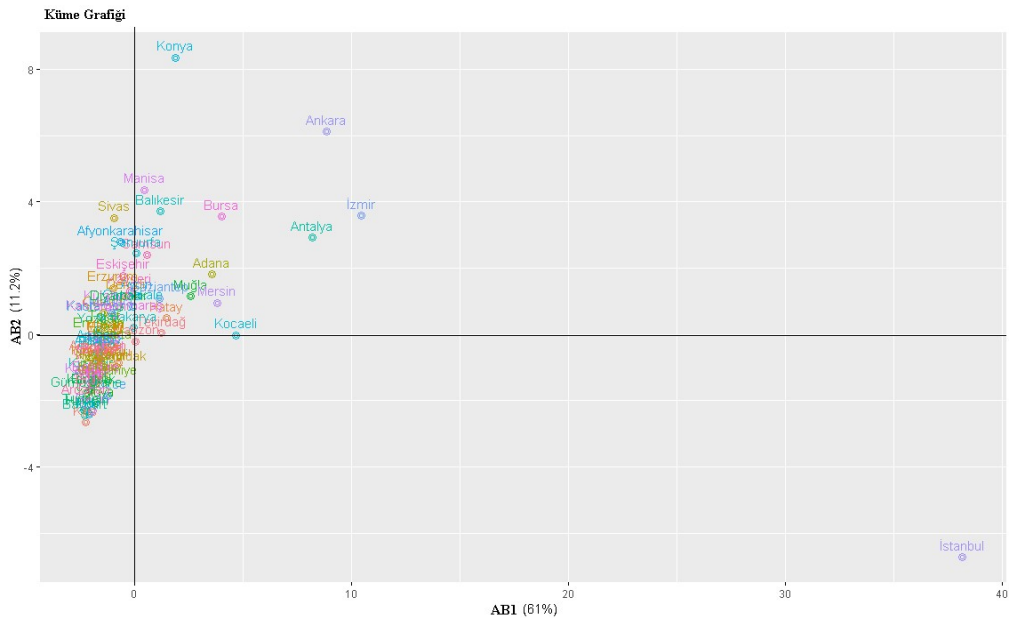
Ana Bileşen	Özdeğer	Açıklanan Varyans Yüzdesi	Kümülatif Toplam
1	24,384	60,959	60,959
2	4,493	11,232	72,191
3	2,771	6,929	79,12
4	1,927	4,818	83,938
5	1,188	2,969	86,907
6	0,991	2,479	89,386
7	0,83	2,075	91,461
8	0,687	1,717	93,178
9	0,517	1,291	94,469
10	0,44	1,101	95,57
11	0,403	1,008	96,578
12	0,278	0,695	97,273
13	0,25	0,624	97,897
14	0,224	0,56	98,457
15	0,161	0,403	98,86
16	0,149	0,374	99,233
17	0,098	0,245	99,478
18	0,053	0,133	99,611
19	0,047	0,117	99,728
20	0,038	0,094	99,822
21	0,024	0,061	99,883
22	0,011	0,028	99,911
23	0,009	0,023	99,933
24	0,009	0,022	99,955
25	0,005	0,013	99,968
26	0,005	0,013	99,981
27	0,003	0,008	99,989
28	0,001	0,004	99,993
29	0,001	0,004	99,997
30	0,001	0,002	99,998
31	0	0,001	99,999
32	0	0	100
33	6,22E-05	0	100
34	1,50E-05	3,75E-05	100
35	1,01E-05	2,53E-05	100
36	3,04E-06	7,60E-06	100
37	1,29E-06	3,22E-06	100
38	4,53E-07	1,13E-06	100
39	3,49E-07	8,72E-07	100
40	1,92E-08	4,80E-08	100

Şekil 17: Veri Seti Karakteristiklerine Ait Yığın Grafiği



Şekil 17'deki yığın grefiği incelendiğinde, sistemin toplam varyansını açıklamak için ilk üç anabileşenin yeterli olduğu görülmektedir. Yığın grafiğinde, ilk üç özdeğerden sonra grafiğin yatay eğilimli düz bir çizgi şeklini aldığı dikkati çekmektedir. Gözlemlerin Şekil 18, Şekil EK: 2 ve Şekil EK: 7'daki iki boyutlu ve üç boyutlu ana bileşen uzayındaki saçılımı incelendiğinde İstanbul, Ankara, İzmir, Antalya gibi illerin genel topluluktan uzak ve düzensiz bir konumda aykırı gözlem gibi davrandıkları ve gözle görülebilecek kadar keskin şekilde ayrılan büyük gruplar olmadığı gözlemlenmektedir.

Şekil 18: Veri Setinde İllerin İki Boyutlu Ana Bileşen Saçılım Grafiği



3.5. KÜMELEME ANALİZİ UYGULAMALARI

Hiyerarşik Kümeleme Yöntemlerinden tek bağlantı, tam bağlantı, ortalama bağlantı, ward yöntemi, ağırlık merkezi yöntemi, medyan yöntemi uygulanan veri setinde bu yöntemler için gerekli uzaklık ölçülerinden en yaygın yöntemlerden birisi olan öklid uzaklık yöntemi kullanılmıştır. İllerin birbiriyle uzaklık ilişkisini gösteren öklid uzaklık matrisi ısı grafiği EK: 3'te gösterilmiştir. Bu grafikte diagonal bölgelerdeki koyu mavi renk minimum uzaklık ilişkisi olan 0 değerine (kendisi) eşittir. Koyu kırmızı da maksimum uzaklığı temsil etmektedir. Şekil dikkatli şekilde incelendiğinde İstanbul başta olmak üzere birçok büyükşehir gözlemlerinin diğer illerle ilişkisinin kırmızı ve kırmızıya yakın renklerle ifade edildiği, diğer illerin ise maviye yakın renklerle ifade edildiği gözlemlenmektedir. Bu da İstanbul başta olmak üzere büyükşehirlerle diğer illerin aralarında daha az bir benzerliğe sahip olduğunu göstermektedir. Veri seti farklı ölçme düzeyinde değişkenler barındırdığından ve normal dağılım göstermediği de göz önünde bulundurularak değişkenlerdeki gözlemler için 0-1 arasında min-maks maksimizasyon yöntemiyle bir standartlaştırma işlemi yapılmıştır. Hiyerarşik kümeleme analizlerinin kullanılmasında R programında “cluster” paketi kullanılmıştır (Maechler ve diğerleri, 2019: 1).

Hiyerarşik olmayan kümeleme analizi yöntemlerinde de k-ortalamlar yöntemi ve k-medoid yöntemiyle kümeleme analizi uygulanmıştır. Yöntemlerin kullanılması için R programında “stats” paketi indirilip kullanılmıştır (Team, 2013: 193).

k-ortalamlar ve k-medoid yöntemleri için küme sayılarına karar vermek amacıyla $(n/2)^{1/2}$ formülünden, ana bileşen skor grafikleri ve küme geçerlilik indekslerinden faydalanılmıştır. $(n/2)^{1/2}$ formülü $(81/2)^{1/2}=41^{1/2}$, den 6,4 sonucu vermiş ve 6 adet küme sayısı olması gerektiğini ifade etmektedir. Buna ek olarak çalışmada 26 adet küme geçerlilik indeksi kullanılmış olmasına rağmen karar vermede 14 adet küme geçerlilik indeksinin vermiş olduğu sonuçlar dikkate alınmıştır. Bu indeksler için İstanbul, İzmir, Ankara, Antalya gibi büyükşehirlerin farklı konumlardaki sapan gözlemler olarak ayrı küme oluşturacağı $(n/2)^{1/2}$ formülü de dikkate alınmıştır. Bu indekslerde küme sayısı olarak alt limit için $(n/2)^{1/2}-1$ 'den 5 küme ve üst limit uygun küme sayısı da 12 olarak belirlenmiştir. Yapılan hesaplamalarda tek bağlantı, tam bağlantı, ortalama bağlantı, ward yöntemi, ağırlık

merkezi yöntemi, medyan yöntemi ve k-ortalamlar yöntemlerinin tamamında önceden karar verilen 14 adet geçerlilik indekslerin her biri uygulanarak tüm yöntemler incelendiğinde ağırlıklı olarak ideal küme sayısının 5 olduğu görülmüştür ("Calinski Harbarz-CH", "Küçük Kümeleme Kriteri-CCC", "C", "DB", "Silhouette", "Duda", "Pseudo t^2 ", "Beale", "Ratkowsky", "Ptbiserial", "Frey", "Mcclain", "Dunn", "SDindex"). Bu geçerlilik indekslerin sonuçları da Tablo 11, 12, 13, 14, 15, 16, 17' de gösterilmiştir. Geçerlilik indekslerinin hesaplanmasında R programının "NbClust" paketi kullanılmıştır (Charrad ve diğerleri, 2013:4-5).

En iyi kümeleme algoritmasının seçiminde ise literatürde genel bir ortak yöntem olmasa da araştırmacı inisiyatifinde, çeşitli yöntemler kullanılarak hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleriyle ayrıntılı inceleme yapılabilir. Örneğin hiyerarşik kümeleme yöntemleri arasında kofenetik korelasyon katsayısı kullanılarak uygun yönteme karar vermek mümkündür. Hiyerarşik yöntemlerde kofenetik korelasyon katsayıları bu çalışmada karşılaştırıldığında en yüksek korelasyon katsayısı değeri hiyerarşik yöntemlerde tek bağlantı yöntemine ait çıkmıştır (Tablo 10).

Hiyerarşik olmayan yöntemler hiyerarşik yöntemlere göre özellikle yüksek sayıda gözleme sahip verilerde daha az işlem yükü gerektirmesi ve aykırı gözlemlere karşı daha az duyarlı olması sebebiyle daha sık tercih edilmektedir.

Tablo 10: Uygulamada Kullanılan Hiyerarşik Kümeleme Yöntemlerindeki Kofenetik Korelasyon Katsayıları

Yöntem	Kofenetik Korelasyon Katsayısı
Tek Bağlantı Yöntemi	<u>0.968526</u>
Tam Bağlantı Yöntemi	0.917049
Ortalama Bağlantı Yöntemi	0.965494
Ağırlık Merkezi Yöntemi	0.964410
Ward Bağlantı Yöntemi	0.525156
Medyan Yöntemi	0.955254

Tablo 10'da gösterildiği gibi hiyerarşik yöntemler arasında en yüksek kofenetik korelasyon katsayısı değeri 0.968526 değeri ile tek bağlantı yöntemine ait çıkmıştır. Bu katsayı sonucuna göre çalışmaya ait veri seti için en başarılı hiyerarşik kümeleme analizi yöntemidir.

Tablo 11: Tek Bağlantı Yöntemi İçin Küme Geçerlilik İndeksleri Sonuçları

	Uygun Küme Sayısı	Kritik Değer
CH	5	18,3491
CCC	5	2,3175
Cindex	7	0,3458
DB	5	0,2945
Silhouette	5	0,5652
Duda	5	1,016
PseudoT2	5	-0,1204
Beale	5	-0,0449
Ratkowsky	5	0,341
PtBiserial	5	0,7411
Frey	-	-
McClain	5	0,0335
Dunn	8	0,5984
SDindex	5	1,3961

Tablo 11'e göre tek bağlantı yönteminde mevcut çoğunluğa göre küme sayısının 5 olması gerektiğine karar verilmiştir. Sadece C-indeksi, Frey indeksi, Dunn indeksi uygun küme sayısı için 5'ten farklı sonuçlar vermiştir.

Tablo 12: Tam Bağlantı Yöntemi için Küme Geçerlilik İndeksleri Sonuçları

	Uygun Küme Sayısı	Kritik Değer
CH	7	22,35
CCC	12	8,2781
Cindex	7	0,3844
DB	6	0,6268
Silhouette	5	0,4922
Duda	5	1,7453
PseudoT2	5	-0,427
Beale	5	-6,0548
Ratkowsky	5	0,3544
PtBiserial	6	0,7148
Frey	6	4,9379
McClain	5	0,0703
Dunn	6	0,3985
SDindex	6	4,1119

Tablo 13: Ortalama Bağlantı Yöntemi Küme Geçerlilik İndeksleri Sonuçları

	Uygun Küme Sayısı	Kritik Değer
CH	5	20,8911
CCC	9	5,4078
Cindex	7	0.3751
DB	7	0,464
Silhouette	5	0,5065
Duda	5	1.9253
PseudoT2	5	-0,4806
Beale	5	-6.8141
Ratkowsky	5	0.3482
PtBiserial	5	0,735
Frey	-	-
McClain	5	0.0579
Dunn	11	0,5093
SDindex	6	2,0029

Tablo 12 ve 13’de tam bağlantı ve ortalama bağlantı yöntemlerindeki küme geçerlilik indekslerindeki hesaplama sonuçları gösterilmiştir. Bu sonuçların gösterdiği ortak çoğunluğa göre küme sayısının 5 olması gerektiğine işaret edilmektedir.

Tablo 14: Ward Yöntemi için Küme Geçerlilik İndeksleri Sonuçları

	Uygun Küme Sayısı	Kritik Değer
CH	5	29,3614
CCC	12	10,7636
Cindex	11	0,2172
DB	12	1,1183
Silhouette	12	0,2283
Duda	10	1,4066
PseudoT2	10	-0,5781
Beale	10	-5,4644
Ratkowsky	5	0,3622
PtBiserial	5	4221
Frey	5	15,3538
McClain	5	0,8438
Dunn	12	0,1754
SDindex	12	8,5689

Tablo 15: Ağırlık Merkezi Yöntemi Küme Geçerlilik İndeksleri Sonuçları

	Uygun Küme Sayısı	Kritik Değer
CH	5	20,5058
CCC	5	3,4334
Cindex	6	0,3907
DB	6	0,3081
Silhouette	5	0,5475
Duda	5	7,5596
PseudoT2	5	0
Beale	-	-
Ratkowsky	5	0,3486
PtBiserial	7	0,7453
Frey	-	-
McClain	5	0,0445
Dunn	7	0,5984
SDindex	7	1,6499

Yukarıda Tablo 14'e göre Ward yönteminde uygun küme sayılarını gösteren indeks sayısında 5 ve 12 sayısında bir eşitlik söz konusudur (5'er adet). Tablo 15'e göre ise uygun küme sayısının 5 olması gerektiğine işaret edilmektedir.

Tablo 16: Medyan Yöntemi Küme Geçerlilik İndeksleri Sonuçlar Tablosu

	Uygun Küme Sayısı	Kritik Değer
CH	5	18,3491
CCC	5	2,3175
Cindex	9	0.333
DB	5	0.2945
Silhouette	5	0.5652
Duda	5	1,0016
PseudoT2	5	-0,1204
Beale	5	-0,0449
Ratkowsky	5	0,341
PtBiserial	5	0,7411
Frey	-	-
McClain	5	0,0335
Dunn	5	0,597
SDindex	5	1,2964

Tablo 17: k-Ortalamalar Yöntemi için Küme Geçerlilik İndeksleri Sonuçları

	Uygun Küme Sayısı	Kritik Değer
CH	5	28,0122
CCC	11	10,5674
Cindex	8	0,2064
DB	11	1,1178
Silhouette	11	0,2175
Duda	5	0,9796
PseudoT2	5	0,6234
Beale	5	0,5625
Ratkowsky	5	0,3595
PtBiserial	6	0,3636
Frey	-	-
McClain	6	1,2673
Dunn	10	0,1791
SDindex	11	9,3149

Tablo 16 ve 17’de medyan ve k-ortalamalar yöntemlerinde uygulanan küme geçerlilik indekslerindeki sonuçları gösterilmiştir. Bu sonuçların gösterdiği ortak çoğunluğa göre küme sayısının 5 olması gerektiğine işaret edilmektedir.

Bütün geçerlilik indeksi tablolarına bakıldığında genel bir yorum yapmak gerekirse belirgin bir çoğunlukla olması gereken küme sayısının 5 olması gerektiği anlaşılmaktadır.

3.5.1. Hiyerarşik Kümeleme Analizi Yöntemleri Uygulamaları

Tek Bağlantı Yöntemi: Tablo 18’de görüldüğü gibi veri setine tek bağlantı yöntemi kullanılarak kümeleme analizi uygulandığında belirlenen 5 adet kümeye büyükşehir unvanına sahip 4 ilin birer kümeyi sahiplendiği gözlemlenmekte ve diğer illerin birinci kümede toplandığı dendrogram sonucu görülmektedir. Aykırı gözlemler olan İstanbul, Ankara, İzmir ve Antalya diğer dört kümeye tek olacak şekilde yerleşmiştir. Kofenetik korelasyon katsayısındaki değerinin (0.9685) diğer hiyerarşik kümeleme algoritmalarına göre en yüksek seviyede olması sebebiyle veri setini en başarılı kümeleyen hiyerarşik kümeleme yöntemi olduğu söylenebilir. 1. Küme’deki iller dışındaki illerin bu şekilde kümelenmesi Türkiye’nin en büyük beş ilinden 4’ünün yine söz konusu iller olması ve dolayısıyla ulaşım faaliyetlerinin bu duruma yansımalarıyla açıklanabilir. Ayrıca tek bağlantı yönteminin yapısının da bu duruma yol açtığı değerlendirilmektedir.

Tablo 18: Tek Bağlantı Yöntemine Göre Kümeler

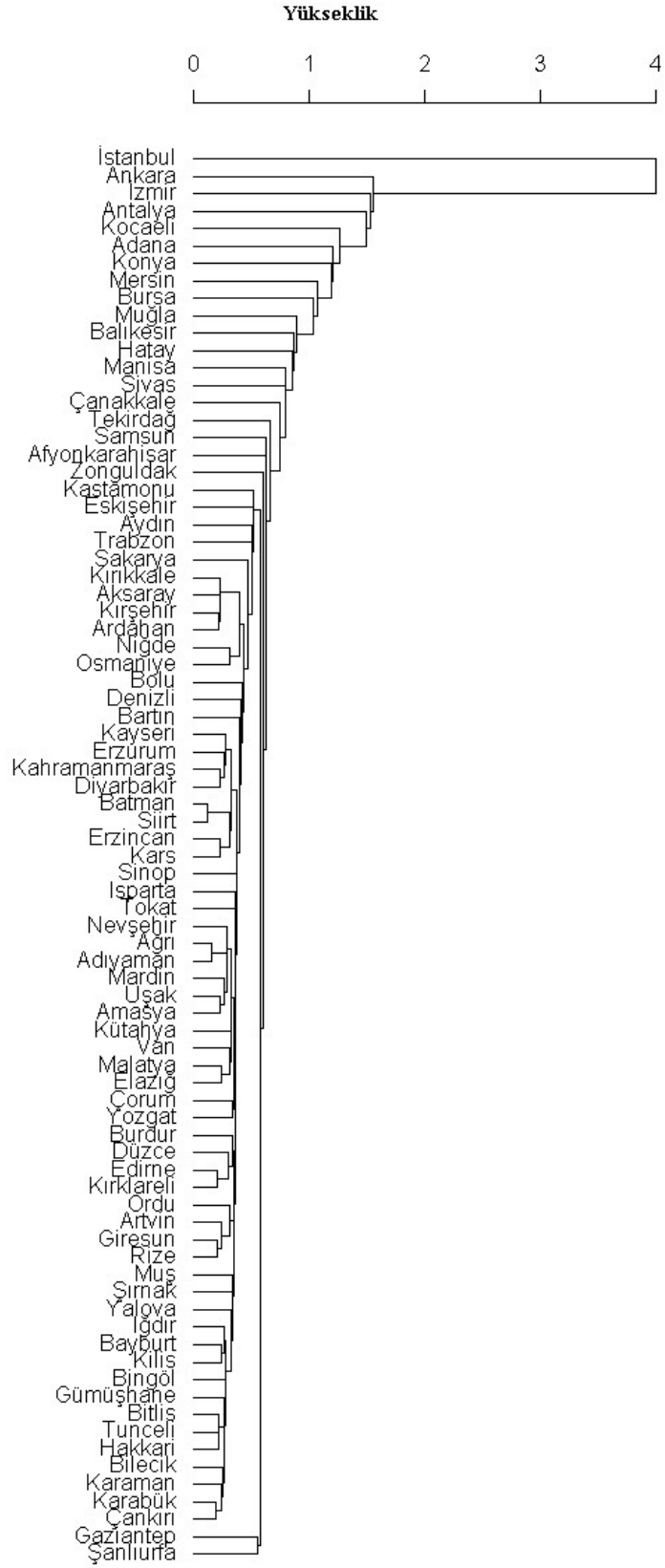
Kümeler	İller
1.Küme	Adana, Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Balıkesir, Bartın, Bayburt, Bilecik, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, Kahramanmaraş, Karabük, Karaman, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mersin, Muğla, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Ardahan, Batman, Bingöl, Bitlis, Diyarbakır, Hakkari, Iğdır, Kars, Kilis, Mardin, Muş, Osmaniye, Siirt, Şanlıurfa, Şırnak, Van
2.Küme	İzmir
3.Küme	Antalya
4.Küme	Ankara
5.Küme	İstanbul

Tablo 19: Tek Bağlantı Yöntemi için Uygulanan Aglomerasyon Tablosu

Aşamalar	Birleştirilen Kümeler		Katsayılar	Aşamalar	Birleştirilen Kümeler		Katsayılar
1	79	81	0,1219689	41	26	45	0,3403353
2	68	75	0,1600712	42	35	41	0,3489765
3	23	41	0,1931543	43	30	40	0,3510251
4	3	5	0,2002505	44	38	43	0,3525250
5	32	36	0,2099609	45	42	44	0,3639398
6	62	72	0,2159063	46	30	45	0,3639414
7	44	69	0,2193186	47	53	46	0,3677522
8	66	6	0,2227195	48	27	47	0,3743192
9	58	77	0,2257004	49	36	48	0,3790264
10	42	7	0,2269183	50	13	33	0,3954091
11	17	29	0,2308597	51	24	49	0,3965198
12	60	67	0,2312925	52	16	51	0,4162770
13	43	10	0,2317250	53	20	52	0,4282076
14	37	5	0,2385833	54	50	53	0,4381794
15	35	74	0,2395485	55	11	54	0,4751966
16	49	3	0,2399406	56	34	55	0,5024148
17	59	61	0,2461962	57	13	56	0,5121446
18	10	16	0,2497269	58	38	57	0,5205363
19	71	15	0,2619360	59	25	58	0,5235254
20	8	18	0,2665167	60	73	76	0,5527108
21	64	9	0,2694691	61	59	60	0,5844115
22	78	11	0,2698212	62	22	61	0,6029023
23	33	20	0,2700523	63	19	62	0,6221940
24	48	21	0,2751844	64	28	63	0,6260271
25	63	23	0,2769503	65	4	64	0,6629572
26	19	25	0,2821106	66	1	65	0,7483639
27	2	22	0,2912245	67	50	66	0,7971375
28	47	27	0,2947727	68	15	67	0,7995215
29	21	4	0,3081032	69	56	68	0,8601528
30	31	14	0,3118597	70	2	69	0,8642530
31	1	12	0,3149758	71	14	70	0,8898019
32	70	17	0,3153152	72	7	71	1,0356333
33	46	57	0,3166369	73	54	72	1,0763584
34	18	32	0,3209127	74	39	73	1,1968712
35	28	34	0,3273065	75	55	74	1,2109937
36	24	31	0,3279636	76	9	75	1,2696853
37	8	26	0,3317883	77	51	76	1,4978283
38	52	29	0,3346062	78	12	77	1,5357910
39	80	37	0,3366741	79	40	78	1,5529414
40	65	39	0,3367027	80	6	79	3,9979058

Tablo 18'i oluştururken kullanılan yukarıda Tablo 19'daki aglomerasyon tablosu ve Şekil 19'daki dendrogram göz önüne alındığında sapan gözlemlerin oluşturduğu büyükşehir illerinin neden olduğu görüntü net bir biçimde anlaşılabilir. Özellikle İstanbul ilinin bulunduğu birleşim noktası Tablo 19'daki 80. Aşamada görülebilir.

Şekil 19: Tek Bağlantı Yöntemi için Kümeleme Analizi Dendrogramı

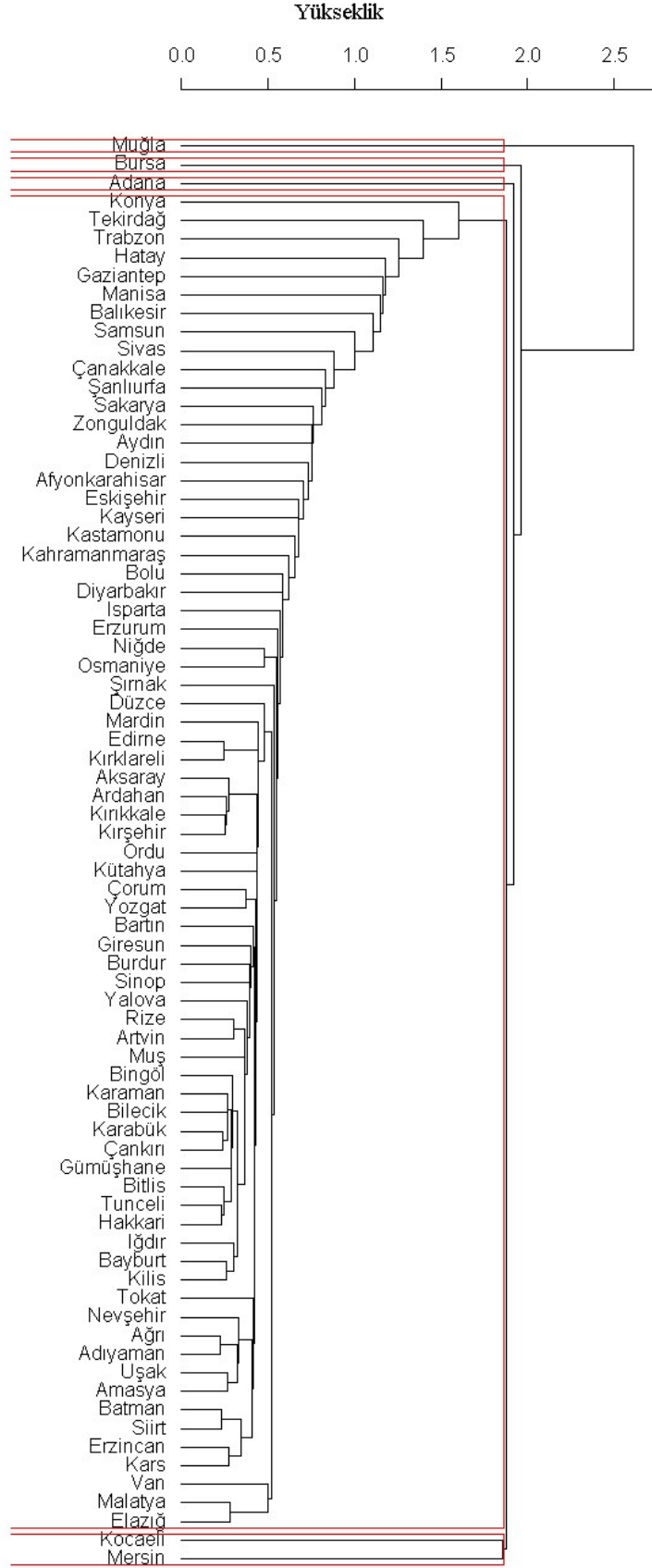


Tablo 18 ve Şekil 19'daki mevcut sonuçlar tek bağlantı yönteminin yapısını yansıtmaktadır. Tekil kümeleri oluşturan İstanbul İzmir, Ankara ve Antalya sapan gözlemlerini oluşturan iller veri setinden çıkarılıp sonuçlar tekrar incelendiğinde kalan 77 ilin oluşturduğu veri seti için Tablo 19.1 ve Şekil 19.1'deki sonuçlarla karşılaşılmıştır. Bu sonuçlarda Adana, Bursa, Muğla, Konya illeri tekil kümeler oluşturmuştur. Bu durum mevcut illerin diğer illere göre ulaşım faaliyetleri bakımından belirgin farkları bulunmasından ve kümeleme analizi yönteminin sapan gözlemlere karşı duyarlılığından kaynaklanmaktadır. Konya ilinin coğrafi olarak diğer illere göre düz olan yapısının ulaşımaya yansıması, Muğla ilinin turizme dönük ulaşım faaliyetleri, Adana ilinin deniz ulaşımı gibi ulaşım faaliyetinin farklı durumu ve Bursa ilinin sanayideki durumu ve Türkiye'nin 4. büyük ili olarak öneminin bu sonuçlarda payı bulunduğu değerlendirilmektedir. Sonuçlardaki asıl ilginç durum Kocaeli ve Mersin illerinin aynı kümede yer almasıdır. Bu illerin ulaşım faaliyetleri incelendiğinde özellikle ihracat ve ithalat geçişlerinin yoğunluğu ve buna yönelik tesislerin varlığının bu benzerliğin oluşmasında mutlak katkı sağladığı değerlendirilmektedir. Kalan iller ise anlamlı olamayacak şekilde tek bir kümede yığılmıştır.

Tablo 19.1: 77 Adet İlden Oluşan Veri Setiyle Tek Bağlantı Yöntemine Göre Kümeler

Kümeler	İller
1.Küme	Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Balıkesir, Bartın, Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, Kahramanmaraş, Karabük, Karaman, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Konya, Kütahya, Malatya, Manisa, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Ardahan, Batman, Bingöl, Bitlis, Diyarbakır, Hakkari, Iğdır, Kars, Kilis, Mardin, Muş, Osmaniye, Siirt, Şanlıurfa, Şırnak, Van
2.Küme	Kocaeli, Mersin
3.Küme	Bursa
4.Küme	Muğla
5.Küme	Adana

Şekil 19.1: 77 İlden Oluşan Veri Setiyle Tek Bağlantı Yöntemi Dendrogramı



Tam Bağlantı Yöntemi: Tablo 20’de görüldüğü gibi bu yöntem veri seti üzerinde uygulandığında 5 küme için İstanbul ve Antalya illeri yine 3. Küme ve 5. Küme olarak aykırı gözlemler olarak ayrı kümelerde yer almıştır. Fakat bununla beraber Türkiye’nin ithalat ve ihracat kapıları niteliğinde yoğun ve gelişmiş ulaşım sistemlerine sahip İzmir, Kocaeli ayrı bir küme olarak birlikte yer almıştır. Aynı şekilde Türkiye’nin en gelişmiş ve yoğun ulaşım hatlarına sahip kara illeri konumlarında bulunan Ankara, Konya ve Sivas ayrı bir kümede yer almış, kalan iller de birinci kümede toplanmıştır. Bu yöntemde başarılı gruplanmalar meydana gelmiş olsa da kofenetik korelasyon katsayı değeri yüksek görünmesine rağmen (0.917) ward yönteminden sonra en düşük seviyeye sahip olması nedeniyle bu çalışma için kılavuz alınabilecek bir yöntem olmaktan uzak bir profil çizmektedir.

Tablo 20: Tam Bağlantı Yöntemine Göre Kümeler

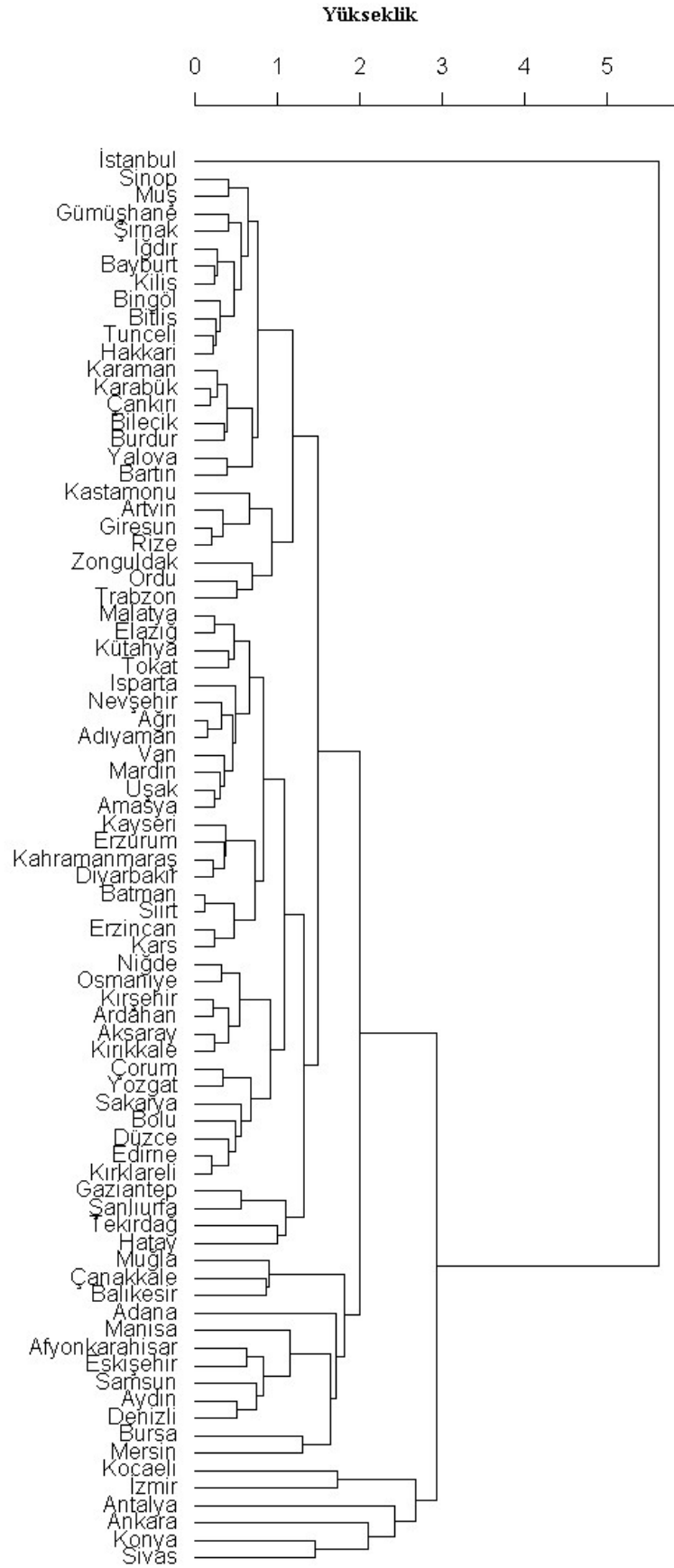
Kümeler	İller
1.Küme	Adana, Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Balıkesir, Bartın, Bayburt, Bilecik, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, Kahramanmaraş, Karabük, Karaman, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Kütahya, Malatya, Manisa, Mersin, Muğla, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Sinop, Tekirdağ, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Ardahan, Batman, Bingöl, Bitlis, Diyarbakır, Hakkari, Iğdır, Kars, Kilis, Mardin, Muş, Osmaniye, Siirt, Şanlıurfa, Şırnak, Van
2.Küme	İzmir, Kocaeli
3.Küme	Antalya
4.Küme	Ankara, Konya, Sivas
5.Küme	İstanbul

Tablo 21: Tam Bağlantı Yöntemi için Uygulanan Aglomerasyon Tablosu

Aşamalar	Birleştirilen Kümeler		Katsayılar	Aşamalar	Birleştirilen Kümeler		Katsayılar
1	79	81	0,1219689	41	13	16	0,5121446
2	68	75	0,1600712	42	19	30	0,5432677
3	23	41	0,1931543	43	73	76	0,5527108
4	3	5	0,2002505	44	32	36	0,5565929
5	32	36	0,2099609	45	11	39	0,5624285
6	62	72	0,2159063	46	19	38	0,6221940
7	44	69	0,2193186	47	31	44	0,6457813
8	58	77	0,2257004	48	25	21	0,6545659
9	17	29	0,2308597	49	37	38	0,6655753
10	60	67	0,2312925	50	22	45	0,6862389
11	42	43	0,2364977	51	22	40	0,6996855
12	35	74	0,2395485	52	27	28	0,7025247
13	59	61	0,2461962	53	26	35	0,7307260
14	66	6	0,2576786	54	28	41	0,7466768
15	71	12	0,2664338	55	47	52	0,7659524
16	49	3	0,2784805	56	46	54	0,8311184
17	78	9	0,3038776	57	49	53	0,8349251
18	63	14	0,3112963	58	1	2	0,8642530
19	46	57	0,3166369	59	14	58	0,9052216
20	47	2	0,3280356	60	42	50	0,9123952
21	37	5	0,3388611	61	48	51	0,9341987
22	26	45	0,3403353	62	4	56	0,9947644
23	64	8	0,3544286	63	57	60	1,0876826
24	10	52	0,3553196	64	43	62	1,0946932
25	70	17	0,3584048	65	15	56	1,1518499
26	48	23	0,3729540	66	55	61	1,1930593
27	16	24	0,3907045	67	7	54	1,3040660
28	8	24	0,3965198	68	63	64	1,3164579
29	18	30	0,4015217	69	39	50	1,4558651
30	7	11	0,4024276	70	66	68	1,4839243
31	27	65	0,4046914	71	65	67	1,6406944
32	33	80	0,4063306	72	55	71	1,7127443
33	21	4	0,4075269	73	9	12	1,7224787
34	20	25	0,4585215	74	59	72	1,8207098
35	1	10	0,4699145	75	70	74	2,0001702
36	15	18	0,4771133	76	40	69	2,1048830
37	13	29	0,4777415	77	51	76	2,4190405
38	53	34	0,4974246	78	73	77	2,6677213
39	20	33	0,4983228	79	75	78	2,9227877
40	31	34	0,5024148	80	6	79	5,6168792

Tablo 21 ve Şekil 20 incelendiğinde özellikle tek bağlantı yöntemine göre sapan gözlemler biraz daha geri planda kalmıştır. En yüksek seviyede birleşim noktası yine 5.62 seviyesinde İstanbul ili olmuştur.

Şekil 20: Tam Bağlantı Yöntemi için Kümeleme Analizi Dendrogramı



Ortalama Bağlantı Yöntemi: Bu yöntemde Tablo 22’de görüldüğü gibi coğrafi olarak birbirine benzer özelliklere sahip Ankara, Konya ve Manisa illeri bir kümededir. Antalya, İstanbul ve İzmir illeri diğer illere göre aykırı değer özellikleri gösterdiği için yine ayrı kümelerde yer almışlardır. Kalan iller de bir kümede toplanmış durumdadır. Kofenetik korelasyon katsayı değeri tek bağlantı yönteminden sonra diğer hiyerarşik yöntemlere göre en yüksek değere (0.9655) sahiptir. Bu da verileri gerçekçi olarak gruplamada hiyerarşik kümeleme yöntemlerinde ikinci tercih edilebilir bir yöntem yapmaktadır.

Tablo 22: Ortalama Bağlantı Yöntemine Göre Kümeler

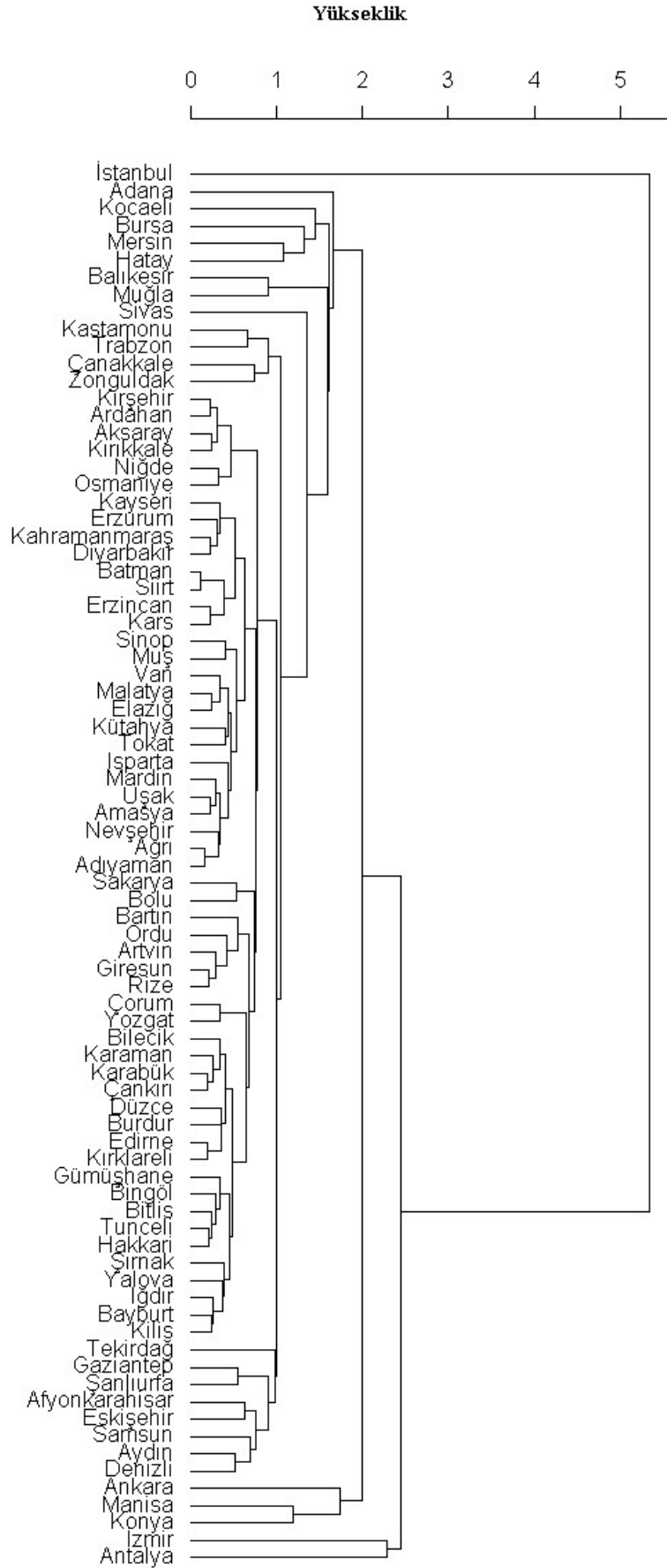
Kümeler	İller
1.Küme	Adana, Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Balıkesir, Bartın, Bayburt, Bilecik, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, Kahramanmaraş, Karabük, Karaman, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Kütahya, Malatya, Mersin, Muğla, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Ardahan, Batman, Bingöl, Bitlis, Diyarbakır, Hakkari, Iğdır, Kars, Kilis, Mardin, Muş, Osmaniye, Siirt, Şanlıurfa, Şırnak, Van
2.Küme	İzmir
3.Küme	Antalya
4.Küme	Ankara, Konya, Manisa
5.Küme	İstanbul

Tablo 23: Ortalama Bağlantı Yöntemi için Uygulanan Aglomerasyon Tablosu

Aşamalar	Birleştirilen Kümeler		Katsayılar	Aşamalar	Birleştirilen Kümeler		Katsayılar
1	79	81	0.1219689	41	13	16	0,5121446
2	68	75	0.1600712	42	19	30	0,5432677
3	23	41	0.1931543	43	73	76	0,5527108
4	3	5	0.2002505	44	32	36	0,5565929
5	32	36	0,2099609	45	11	39	0,5624285
6	62	72	0,2159063	46	19	38	0,6221940
7	44	69	0,2193186	47	31	44	0,6457813
8	58	77	0,2257004	48	25	21	0,6545659
9	17	29	0,2308597	49	37	38	0,6655753
10	60	67	0,2312925	50	22	45	0,6862389
11	42	43	0,2364977	51	22	40	0,6996855
12	35	74	0,2395485	52	27	28	0,7025247
13	59	61	0,2461962	53	26	35	0,7307260
14	66	6	0,2576786	54	28	41	0,7466768
15	71	12	0,2664338	55	47	52	0,7659524
16	49	3	0,2784805	56	46	54	0,8311184
17	78	9	0,3038776	57	49	53	0,8349251
18	63	14	0,3112963	58	1	2	0,8642530
19	46	57	0,3166369	59	14	58	0,9052216
20	47	2	0,3280356	60	42	50	0,9123952
21	37	5	0,3388611	61	48	51	0,9341987
22	26	45	0,3403353	62	4	56	0,9947644
23	64	8	0,3544286	63	57	60	1,0876826
24	10	52	0,3553196	64	43	62	1,0946932
25	70	17	0,3584048	65	15	56	1,1518499
26	48	23	0,3729540	66	55	61	1,1930593
27	16	24	0,3907045	67	7	54	1,3040660
28	8	24	0,3965198	68	63	64	1,3164579
29	18	30	0,4015217	69	39	50	1,4558651
30	7	11	0,4024276	70	66	68	1,4839243
31	27	65	0,4046914	71	65	67	1,6406944
32	33	80	0,4063306	72	55	71	1,7127443
33	21	4	0,4075269	73	9	12	1,7224787
34	20	25	0,4585215	74	59	72	1,8207098
35	1	10	0,4699145	75	70	74	2,0001702
36	15	18	0,4771133	76	40	69	2,1048830
37	13	29	0,4777415	77	51	76	2,4190405
38	53	34	0,4974246	78	73	77	2,6677213
39	20	33	0,4983228	79	75	78	2,9227877
40	31	34	0,5024148	80	6	79	5,6168792

Tablo 23'teki aglomerasyon tablosu ve Şekil 21'deki dendrogram incelendiğinde İstanbul ilinin sapan gözlem olarak ayrılma durumu ve Ankara, İzmir, Antalya illeri gibi büyükşehir illerinin sapan gözlem olarak ayrılma çabası net bir şekilde anlaşılabilmektedir.

Şekil 21: Ortalama Bağlantı Yöntemi için Kümeleme Analizi Dendrogramı

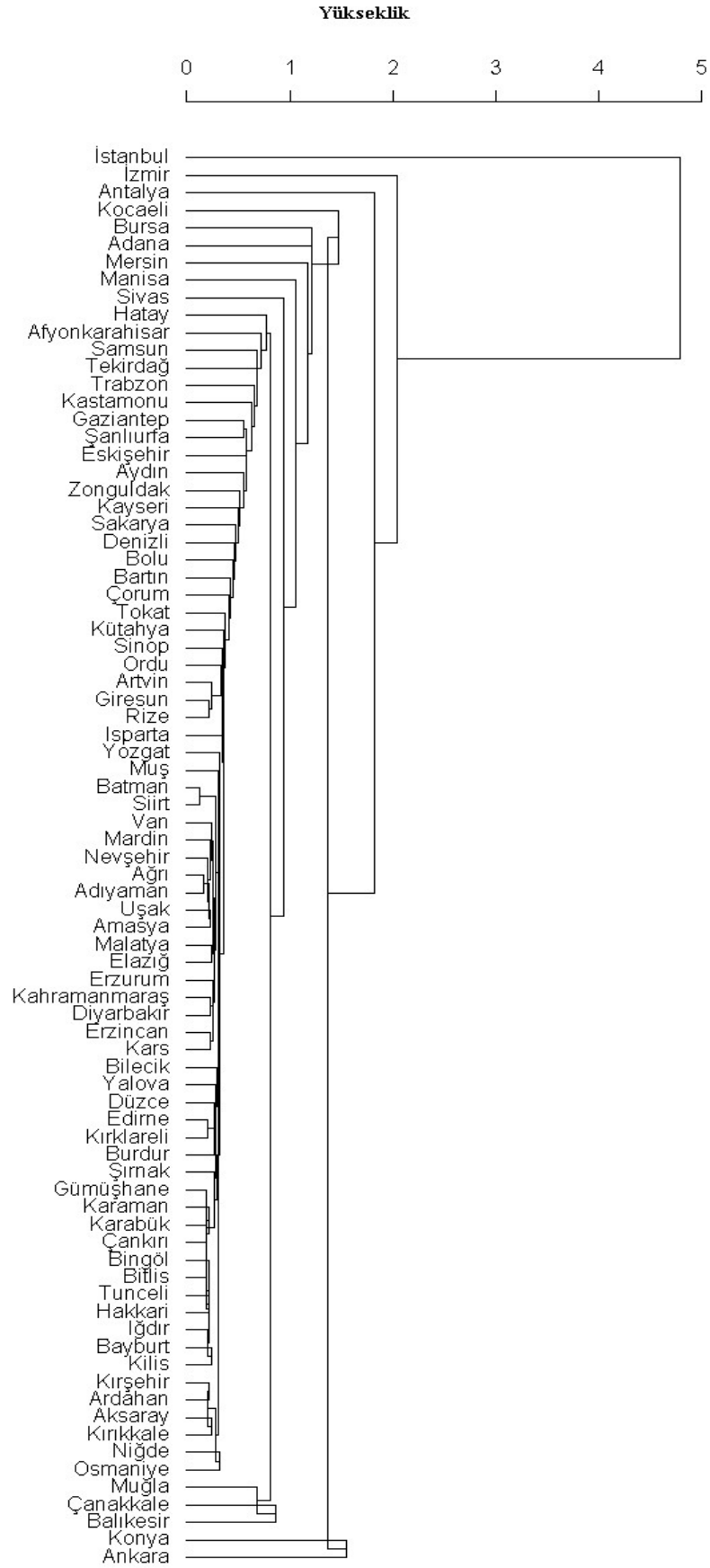


Ağırlık Merkezi Yöntemi: Dendrogram sonucu 5 küme için incelendiğinde tersine dönme durumu dikkate alınmazsa Ankara, Kocaeli ve Konya illeri bir kümede, İzmir, Antalya, İstanbul illeri de ayrı kümelerde olmak üzere 3 adet kümede tek gözlemlilik gruplar oluşturmuştur. Yine de tersine dönme sebebiyle oluşan dendrogramdaki bazı bozulmalar göz önüne alındığında çalışmada tercih edilebilir yöntem değildir. Bu durum Tablo 24'teki aglomerasyon tablosunda Şekil 22'deki dendrogramda da anlaşılabilmektedir.

Tablo 24: Ağırlık Merkezi Yöntemi için Uygulanan Aglomerasyon Tablosu

Aşamalar	Birleştirilen Kümeler		Katsayılar	Aşamalar	Birleştirilen Kümeler		Katsayılar
1	79	81	0,1219689	41	19	40	0,2797360
2	68	75	0,1600712	42	39	41	0,2995409
3	23	41	0,1931543	43	45	42	0,3140898
4	3	5	0,2002505	44	31	17	0,3265165
5	32	36	0,2099609	45	53	43	0,3532601
6	49	3	0,2109220	46	44	45	0,3584888
7	62	72	0,2159063	47	27	46	0,3449306
8	66	7	0,1862225	48	18	47	0,3632186
9	63	8	0,2154359	49	30	48	0,3710382
10	44	69	0,2193186	50	26	49	0,4118950
11	58	77	0,2257004	51	24	50	0,4171622
12	17	29	0,2308597	52	20	51	0,4511410
13	2	12	0,2164137	53	16	52	0,4675179
14	47	13	0,2082973	54	11	53	0,4741077
15	78	14	0,2289209	55	48	54	0,5063451
16	60	67	0,2312925	56	22	55	0,5184464
17	37	5	0,2362320	57	13	56	0,5455500
18	42	43	0,2364977	58	73	76	0,5527108
19	10	18	0,1954859	59	38	57	0,5840323
20	35	74	0,2395485	60	58	59	0,5783470
21	71	20	0,2042978	61	25	60	0,6258721
22	9	21	0,2211375	62	34	61	0,6544251
23	6	22	0,1886998	63	4	62	0,6851186
24	33	23	0,1938194	64	28	63	0,6829362
25	70	15	0,2422290	65	19	64	0,7183997
26	59	61	0,2461962	66	56	65	0,7758567
27	25	26	0,2474595	67	1	2	0,8642530
28	11	16	0,2481636	68	14	67	0,6814485
29	64	28	0,2548799	69	66	68	0,8119271
30	27	29	0,2617563	70	50	69	0,9452748
31	80	24	0,2710832	71	15	70	1,0601839
32	1	30	0,2771466	72	54	71	1,1711744
33	52	31	0,2914749	73	55	72	1,2146471
34	4	33	0,2753076	74	7	73	1,2113897
35	21	34	0,2727319	75	9	74	1,4773841
36	8	35	0,2827156	76	39	40	1,5529414
37	10	36	0,2936841	77	75	76	1,3659778
38	32	37	0,3031319	78	51	77	1,8282971
39	65	38	0,3029364	79	12	78	2,0419401
40	46	57	0,3166369	80	6	79	4,7951610

Şekil 22: Ağırlık Merkezi Yöntemi için Kümeleme Analizi Dendrogramı

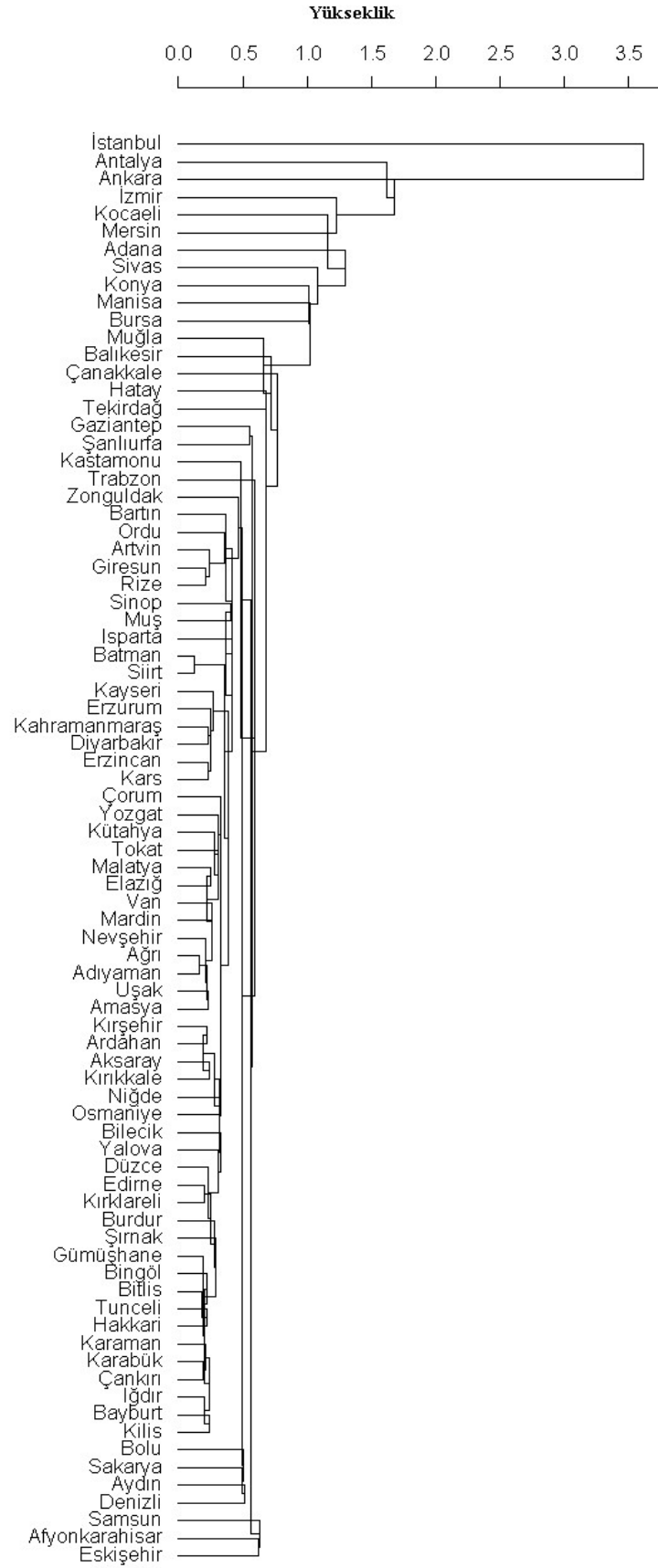


Medyan Yöntemi: Dendrogram sonucu 5 küme için incelendiğinde tersine dönme göz önüne alınırsa Antalya, Ankara, İzmir, İstanbul illeri 4 adet kümede tek gözlemlilik gruplar oluşturmuş, kalan 77 il de ayrı bir küme grubu olarak birinci kümede toplanmıştır. Ayrıca dendrogramdaki tersine dönmelerin sebep olduğu çok ciddi bozulmalar sebebiyle çalışmada tercih edilebilir bir yöntem değildir. Bu bozulmalar Tablo 25'teki Aglomerasyon tablosu ve Şekil 23'teki dendrogramdan da rahatlıkla anlaşılabilir.

Tablo 25: Medyan Yöntemi için Uygulanan Aglomerasyon Tablosu

Aşamalar	Birleştirilen Kümeler		Katsayılar	Aşamalar	Birleştirilen Kümeler		Katsayılar
1	79	81	0,1219689	41	26	38	0,3260330
2	68	75	0,1600712	42	10	37	0,3298739
3	23	41	0,1931543	43	40	42	0,3215905
4	3	5	0,2002505	44	41	43	0,3305377
5	32	36	0,2099609	45	31	16	0,3566461
6	49	3	0,2109220	46	30	44	0,3886459
7	62	72	0,2159063	47	1	46	0,3607899
8	66	7	0,1862225	48	27	65	0,4046914
9	63	8	0,2185990	49	53	47	0,4136778
10	44	69	0,2193186	50	48	49	0,3671381
11	58	77	0,2257004	51	45	50	0,4163601
12	17	29	0,2308597	52	24	51	0,3689360
13	2	12	0,2164137	53	22	52	0,4613803
14	47	13	0,2082973	54	13	16	0,5121446
15	60	67	0,2312925	55	11	54	0,4949963
16	37	5	0,2362320	56	20	55	0,5057115
17	42	43	0,2364977	57	53	56	0,4905195
18	10	17	0,1954859	58	73	76	0,5527108
19	35	74	0,2395485	59	34	57	0,5927033
20	71	19	0,2042978	60	25	59	0,4864902
21	6	20	0,2362896	61	19	38	0,6221940
22	9	21	0,1966724	62	28	61	0,6324998
23	33	22	0,1934420	63	60	62	0,5673437
24	59	61	0,2461962	64	58	63	0,5716555
25	11	15	0,2481636	65	4	64	0,6759021
26	64	25	0,2548799	66	56	65	0,6824971
27	78	14	0,2569190	67	1	66	0,7668434
28	70	27	0,2586415	68	2	67	0,7148260
29	24	28	0,2255145	69	14	68	0,6562910
30	48	26	0,2702800	70	7	69	1,0245327
31	80	23	0,2935308	71	15	70	1,0197063
32	52	31	0,2808204	72	39	71	1,0135951
33	4	32	0,2481271	73	50	72	1,0808478
34	21	33	0,2303396	74	55	73	1,3016996
35	30	29	0,3041408	75	54	74	1,1594071
36	18	35	0,2828038	76	9	75	1,1644434
37	8	34	0,3055765	77	12	76	1,2322792
38	45	36	0,3126389	78	40	77	1,6762528
39	46	57	0,3166369	79	51	78	1,6157828
40	18	39	0,2797360	80	6	79	3,6175761

Şekil 23: Medyan Yöntemi için Kümeleme Analizi Dendrogramı



Ward Yöntemi: Tablo 26’da kümeleme sonuçları görüldüğü gibi 5 küme olarak objelerin dağılımı kontrol edildiğinde hiyerarşik yöntemlerde aykırı gözlemlerin en az hissedildiği yöntem bu olmaktadır. Fakat kofenetik korelasyon katsayısı değeri 0,52 ile diğer yöntemlere göre en düşük kümeleme algoritmasının bu olması çalışmadaki güvenilirliğini olumsuz yönde etkilemektedir. Tablo 26’da 1. Küme için ulaşımın güç ve hafif yoğunlukta olduğu Karadeniz Bölgesi ve Doğu Anadolu ağırlıklı illeriyle iç bölgelerdeki bazı küçük iller yer almıştır. Bu iller Amasya, Artvin, Bartın, Bayburt, Bilecik, Bolu, Burdur, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Giresun, Gümüşhane, Isparta, Karabük, Karaman, Kastamonu, Kırklareli, Kütahya, Malatya, Nevşehir, Ordu, Rize, Sakarya, Sinop, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Bingöl, Bitlis, Hakkari, Iğdır, Kilis, Mardin, Muş, Şırnak, Van’dır. 2. Küme ise Türkiye’nin yoğun ve gelişmiş ulaşım sistemlerini barındıran Marmara, Ege ve Akdeniz kıyı illeri çoğunluktadır. Bu iller Adana, Antalya, Balıkesir, Bursa, Çanakkale, Hatay, İzmir, Kocaeli, Mersin, Muğla, Tekirdağ’dır. 3. Küme ulaşım göstergeleri bakımından genellikle orta büyüklükte ama belli başlı öneme sahip ulaşım tesisleri barındıran (doğu ekspresi gibi bazı demiryolu hatları, petrol ve doğalgaz boru hatları) iller yer almıştır. Bu iller Aksaray, Ardahan, Batman, Diyarbakır, Erzincan, Erzurum, Kahramanmaraş, Kars, Kayseri, Kırıkkale, Kırşehir, Niğde, Osmaniye, Siirt’tir.

Tablo 26: Ward Yöntemine Göre Kümeler

Kümeler	İller
1.Küme	Amasya, Artvin, Bartın, Bayburt, Bilecik, Bolu, Burdur, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Giresun, Gümüşhane, Isparta, Karabük, Karaman, Kastamonu, Kırklareli, Kütahya, Malatya, Nevşehir, Ordu, Rize, Sakarya, Sinop, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Bingöl, Bitlis, Hakkari, Iğdır, Kilis, Mardin, Muş, Şırnak, Van
2.Küme	Adana, Antalya, Balıkesir, Bursa, Çanakkale, Hatay, İzmir, Kocaeli, Mersin, Muğla, Tekirdağ
3.Küme	Aksaray, Ardahan, Batman, Diyarbakır, Erzincan, Erzurum, Kahramanmaraş, Kars, Kayseri, Kırıkkale, Kırşehir, Niğde, Osmaniye, Siirt
4.Küme	Afyonkarahisar, Ankara, Aydın, Eskişehir, Gaziantep, Konya, Manisa, Samsun, Sivas, Şanlıurfa
5.Küme	İstanbul

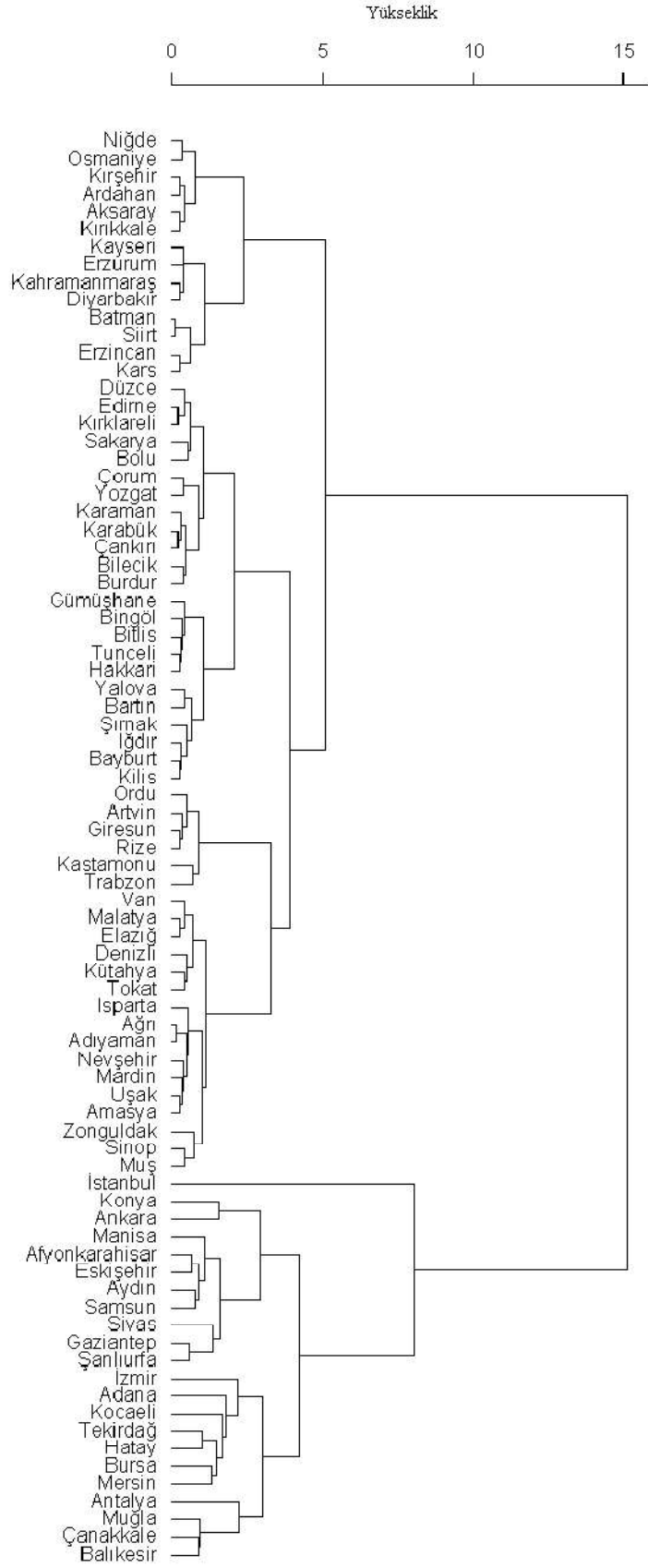
4. Küme ise Aydın, Samsun dışında tamamen kara illeri ağırlıklı fakat yine de belli seviyede gelişmiş ve işlek ulaşım göstergelerine sahip illerin toplanmış olduğu görülmektedir. Bu iller Afyonkarahisar, Ankara, Aydın, Eskişehir, Gaziantep, Konya, Manisa, Samsun, Sivas, Şanlıurfa'dır. İstanbul ise tek başına küme oluşturan bir sapan gözlemdir.

Tablo 27: Ward Yöntemi için Uygulanan Aglomerasyon Tablosu

Aşamalar	Birleştirilen Kümeler		Katsayılar	Aşamalar	Birleştirilen Kümeler		Katsayılar
1	79	81	0,121969	41	32	38	0,598989
2	68	75	0,160071	42	1	10	0,601170
3	23	41	0,193154	43	19	38	0,622194
4	3	5	0,200251	44	29	34	0,625380
5	32	36	0,209961	45	25	34	0,665064
6	62	72	0,215906	46	26	36	0,695018
7	44	69	0,219319	47	22	31	0,725381
8	58	77	0,225700	48	19	28	0,745963
9	17	29	0,230860	49	13	28	0,746677
10	60	67	0,231293	50	1	2	0,864253
11	42	43	0,236498	51	43	49	0,882229
12	35	74	0,239549	52	21	33	0,887799
13	59	61	0,246196	53	37	45	0,897697
14	66	6	0,248297	54	14	50	0,908598
15	71	12	0,272397	55	4	56	0,994764
16	49	3	0,281229	56	39	47	1,008600
17	78	9	0,305513	57	41	52	1,038493
18	37	5	0,314976	58	27	44	1,051707
19	46	57	0,316637	59	15	51	1,089350
20	63	14	0,323154	60	25	42	1,101863
21	26	45	0,340335	61	46	56	1,144916
22	64	8	0,340698	62	7	54	1,304066
23	10	52	0,355320	63	50	40	1,313767
24	47	17	0,360141	64	55	62	1,442528
25	48	22	0,36438	65	39	40	1,552941
26	70	13	0,380584	66	59	63	1,587629
27	33	20	0,384899	67	9	64	1,652850
28	7	11	0,390972	68	55	67	1,766141
29	8	24	0,396520	69	57	58	2,059964
30	18	30	0,401522	70	12	68	2,178374
31	27	65	0,404691	71	51	54	2,245433
32	21	4	0,410337	72	48	60	2,356368
33	16	23	0,431469	73	65	66	2,923435
34	80	15	0,455078	74	70	71	3,008979
35	2	24	0,481985	75	53	61	3,299344
36	16	30	0,483115	76	69	75	3,911483
37	31	18	0,489775	77	73	74	4,236456
38	11	20	0,527311	78	72	76	5,088287
39	53	35	0,533242	79	6	77	8,050810
40	73	76	0,552711	80	78	79	15,11187

Tablo 27'deki aglomerasyon tablosu incelendiğinde 79 ve 80. aşamalara kadar istikrarlı bir artış söz konusudur.

Şekil 24: Ward Yöntemine için Kümeleme Analizi Dendrogramı



Şekil 24'teki dendrogram incelendiğinde yorumlanması kolay ve belli bir küme sayısı belirlenmesine (örneğin 12) uygun bir dendrogram ortaya çıkmıştır. Bunun yorumlanması araştırmacılar tarafından farklı şekillerde ele alınabilir.

3.5.2. Hiyerarşik Olmayan Kümeleme Analizi Yöntemleri Uygulamaları

k-Ortalamalar Yöntemi: İncelenen verilerin ana bileşen skorlarının görsel olarak incelenmesi ve geçerlilik indekslerinin göz önünde bulundurulması sonucu k değeri 5 olarak uygulanan bu yöntem sonucunda Tablo 28'de görüldüğü gibi 1. Küme için Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa, Muğla illeri olmak üzere 8 il yer almıştır. Bu iller ulaşım faaliyet ve sistemleri bakımından oldukça işlek ve gelişmiş ulaşımına sahiptir. Bununla birlikte bu kümedeki bazı illerin önemli ulaşım göstergeleri çok zayıf veya olmamasına rağmen yine de sahip oldukları diğer ulaşım göstergelerinin etkisiyle kendilerine özel farklı ve gelişkin konumda bu gruba dâhil oldukları ifade edilebilir. 2. Küme için ise sanayisi ve limanları gelişmiş olan ve ithalat ve ihracat faaliyetlerinin odak noktalarına sahip kıyı illeri bu grupta yer bulmuştur. Bu iller ise Adana, Hatay, İzmir, Kocaeli, Mersin olmuştur. 3. Küme'de ise birinci, ikinci ve beşinci kümeler kadar yoğun ve gelişmiş bir ulaşım faaliyeti görülüyor olsa da bu gruptaki diğer illerin de zayıf olmayan ve gelişmeye müsait iller olduğu sonucu çıkartılabilir. Bu kümede yer alan illerin genelinde önemli bir demiryolu hattı üzerinde veya önemli bir liman kenti olma özelliği göstermesi dikkat çekicidir. Örneğin doğu ekspresi ve bu ekspresin olduğu bazı önemli boru hatlarının geçtiği iller bu küme içindedir. Önemli otoyol ve demiryolu ağına sahip İç Ege ve Aydın illeri de bu küme içinde yer almıştır. Ayrıca önemli liman bölgeleri olan Samsun, Trabzon, Çanakkale, Tekirdağ bu kümededir. 3. Küme'deki gözlemler Afyonkarahisar, Aydın, Çanakkale, Denizli, Diyarbakır, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Isparta, Kahramanmaraş, Kars, Kayseri, Kütahya, Malatya, Mardin, Samsun, Sivas, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Uşak, Van illeridir. 4. Küme'deki iller ise genellikle içinde bulunduğu coğrafi sebepler ve dezavantajlı konumları nedeniyle yeterli ulaşım faaliyet ve sistemlerine sahip olamamış gözlemlerden oluştuğu söylenebilir. Bu grubun dahil olduğu iller genellikle coğrafi şartların zorladığı Karadeniz Bölgesi, Doğu Anadolu, Kuzey Marmara ve Toros Dağlarının yüksek olduğu noktalarda toplanmışlardır. Batı Karadeniz ve Ankara'nın

doğusundaki bazı illerin bu kümeye katılması genel görüntüyü bozsa da bu gruptaki illerin ulaşım faaliyetlerinin zayıflığına işaret ettiğini göstermektedir.

Tablo 28’de 4. Küme’nin gözlemlerini oluşturan iller ise Adıyaman, Ağrı, Aksaray, Amasya, Ardahan, Artvin, Bartın, Batman, Bayburt, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çankırı, Çorum, Düzce, Edirne, Giresun, Gümüşhane, Hakkari, Iğdır, Karabük, Karaman, Kastamonu, Kırıkkale, Kırklareli, Kırşehir, Kilis, Muş, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Siirt, Sinop, Şırnak, Tunceli, Yalova, Yozgat, Zonguldak’tır.

Tablo 28: k-Ortalamalar Yöntemine Göre Kümeler

Kümeler	İller
1.Küme	Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa, Muğla
2.Küme	Adana, İzmir, Kocaeli, Hatay, Mersin
3.Küme	Afyonkarahisar, Aydın, Çanakkale, Denizli, Diyarbakır, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Isparta, Kahramanmaraş, Kars, Kayseri, Kütahya, Malatya, Mardin, Samsun, Sivas, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Uşak, Van
4.Küme	Adıyaman, Ağrı, Aksaray, Amasya, Ardahan, Artvin, Bartın, Batman, Bayburt, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çankırı, Çorum, Düzce, Edirne, Giresun, Gümüşhane, Hakkari, Iğdır, Karabük, Karaman, Kastamonu, Kırıkkale, Kırklareli, Kırşehir, Kilis, Muş, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Siirt, Sinop, Şırnak, Tunceli, Yalova, Yozgat, Zonguldak
5.Küme	İstanbul

k-ortalamar yönteminde uygulanan iterasyonlar sonucu oluşan küme merkezleri her değişken için Tablo 29’da aşağıdaki gibi oluşmuştur. Küme merkezleri, algoritma işleyişi sırasında gözlemlerin atanacağı kümelerin belirlenmesinde doğrudan bir etkiye sahiptir.

Tablo 29: k-Ortalamalar Yöntemi için Küme Merkezleri

Değişkenler	1.Küme	2.Küme	3.Küme	4.Küme	5.Küme
V1	0,069237	0,009859	7,40E-04	0,046633	1
V2	0,191258	0,00191	5,42E-05	0,020116	1
V3	0,219894	0,084159	2,68E-03	0,118526	1
V4	0,094322	0,020753	1,54E-03	0,084811	1
V5	0,094798	0,020686	1,54E-03	0,085173	1
V6	0,10738	0,000975	3,67E-05	0,017755	1
V7	0,106439	0,000945	3,04E-05	0,017555	1
V8	0,093146	0,018543	1,37E-03	0,074405	1
V9	0,088575	0,018126	1,36E-03	0,081796	1
V10	0,049415	0,000732	2,12E-05	0,009153	1
V11	0,045619	0,000974	1,35E-05	0,008678	1
V12	0,59447	0,452903	1,20E-01	0,341935	1
V13	0,413187	0,085538	7,87E-02	0,433846	0,969231
V14	0,046348	0,001161	4,63E-05	0,269284	1
V15	0,04088	0,001524	3,32E-05	0,237504	1
V16	0,085775	0,014405	4,97E-05	0,142488	1
V17	0,001021	0,009519	7,81E-07	0,008787	1
V18	0,112701	0,015015	3,71E-03	0,566903	0,652884
V19	0,037995	0,020742	8,03E-03	0,586261	0,473303
V20	0,076125	0,030704	1,46E-02	0,47547	1
V21	0,00071	0,001705	3,39E-05	0,205969	0,158766
V22	0,054213	0,009359	1,35E-02	0,200372	1
V23	0,765897	0,258858	1,07E-01	0,450862	0,900044
V24	0,054921	0,010455	1,45E-02	0,207008	1
V25	0,742451	0,248123	1,03E-01	0,407351	0,816179
V26	0,059798	0,015642	1,93E-02	0,234373	1
V27	0,669591	0,251566	1,12E-01	0,331815	0,480658
V28	0,160828	0,03306	1,14E-02	0,118947	1
V29	0,482989	0,280489	1,44E-01	0,26515	0,120502
V30	0,15142	0,063125	2,30E-02	0,128652	1
V31	0,172823	0,06187	4,95E-02	0,572745	1
V32	0,149918	0,03249	1,20E-02	0,141714	1
V33	0,398175	0,384596	3,08E-01	0,236279	0
V34	0,176387	0,043337	1,45E-02	0,152982	1
V35	0,194522	0,05583	2,43E-02	0,178652	1
V36	0,489877	0,111264	3,29E-02	0,505348	0,871267
V37	0,191168	0,04536	1,82E-02	0,128577	1
V38	0,710105	0,297106	1,35E-01	0,433663	0,314363
V39	0,40844	0,354054	8,20E-02	0,309089	0,386391
V40	0,604762	0,625333	4,82E-01	0,58	0,5

Tablo 30: k-Ortalamalar Yönteminde Veri Setindeki Sonuç Küme Merkezleri Arasındaki Öklid Uzaklık Değerleri

Kümeler	1	2	3	4	5
1	-				
2	1.2854258	-			
3	1.1845439	1.3678609	-		
4	1.6167570	1.5530568	0.5829115	-	
5	4.7407495	4.5666325	5.2997600	5.4756576	-

Tablo 30’da k-ortalamalar yönteminde elde edilen küme merkezleri arasındaki öklid uzaklık matrisi incelendiğinde en yakın kümeler 0.58 değeriyle 3. Küme ve 4. Küme; en uzak kümeler ise 5.48 değeriyle 4. Küme ve 5. Küme olmuştur. 5. Küme’nin İstanbul ili olduğu 4. Küme’nin ise Türkiye’nin ulaştırma göstergeleri bakımından en düşük durumda olan illerden olduğu göz önüne alındığında bu son derece beklenen bir durumdur. Aynı şekilde geri kalmış iller ile orta büyüklükteki illerin birbirine çok yakın durumda olduğu da bu matriste belli olmaktadır. Tablo 31’de kümelerin sahip olduğu gözlem sayıları görülmektedir.

Tablo 31: k-Ortalamalar Yönteminde Kümelerde Bulunan İllerin Sayısı

Kümeler	Üye Sayısı
1	5
2	7
3	25
4	43
5	1
Toplam	81

k-Medoid Yöntemi: k-ortalamalar yönteminde olduğu gibi küme sayısı olan $k=5$ olarak belirlendikten sonra elde edilen gruplar, Tablo 32’de görüldüğü gibi 1. Küme için Çanakkale, Uşak, Zonguldak, Kastamonu, Çorum, Sinop, Amasya, Tokat, Ordu, Giresun, Trabzon, Artvin, Yozgat, Nevşehir, Isparta, Kahramanmaraş, Malatya, Erzincan, Elazığ, Erzurum, Muş, Kars, Ağrı, Van, Adıyaman, Diyarbakır, Mardin, Batman, Siirt illeridir. 2. Küme için Balıkesir, Tekirdağ, Sakarya, Aydın, Muğla, Manisa, Denizli, Kütahya, Afyonkarahisar, Samsun, Eskişehir, Konya, Ankara, Kayseri, Sivas, Antalya, Adana, Hatay, Gaziantep, Şanlıurfa’dır. 3. Küme için ise Edirne, Kırklareli, Yalova, Bilecik, Bolu, Düzce, Karabük, Bartın, Gümüşhane,

Bayburt, Rize, Çankırı, Aksaray, Kırıkkale, Kırşehir, Niğde, Karaman, Burdur, Osmaniye, Tunceli, Bingöl, Bitlis, Ardahan, Iğdır, Hakkari, Kilis, Şırnak ve 4. Küme’de İstanbul yine tek başına yerini almıştır. 5. Küme’de Bursa, İzmir, Kocaeli, Mersin illeri yer almaktadır.

Tablo 32: k-Medoid Yöntemine Göre Kümeler

Kümeler	İller
1. Küme	Çanakkale, Uşak, Zonguldak, Kastamonu, Çorum, Sinop, Amasya, Tokat, Ordu, Giresun, Trabzon, Artvin, Yozgat, Nevşehir, Isparta, Kahramanmaraş, Malatya, Erzincan, Elazığ, Erzurum, Muş, Kars, Ağrı, Van, Adıyaman, Diyarbakır, Mardin, Batman, Siirt
2. Küme	Balıkesir, Tekirdağ, Sakarya, Aydın, Muğla, Manisa, Denizli, Kütahya, Afyonkarahisar, Samsun, Eskişehir, Konya, Ankara, Kayseri, Sivas, Antalya, Adana, Hatay, Gaziantep, Şanlıurfa
3. Küme	Edirne, Kırklareli, Yalova, Bilecik, Bolu, Düzce, Karabük, Bartın, Gümüşhane, Bayburt, Rize, Çankırı, Aksaray, Kırıkkale, Kırşehir, Niğde, Karaman, Burdur, Osmaniye, Tunceli, Bingöl, Bitlis, Ardahan, Iğdır, Hakkari, Kilis, Şırnak
4. Küme	İstanbul
5. Küme	Bursa, İzmir, Kocaeli, Mersin

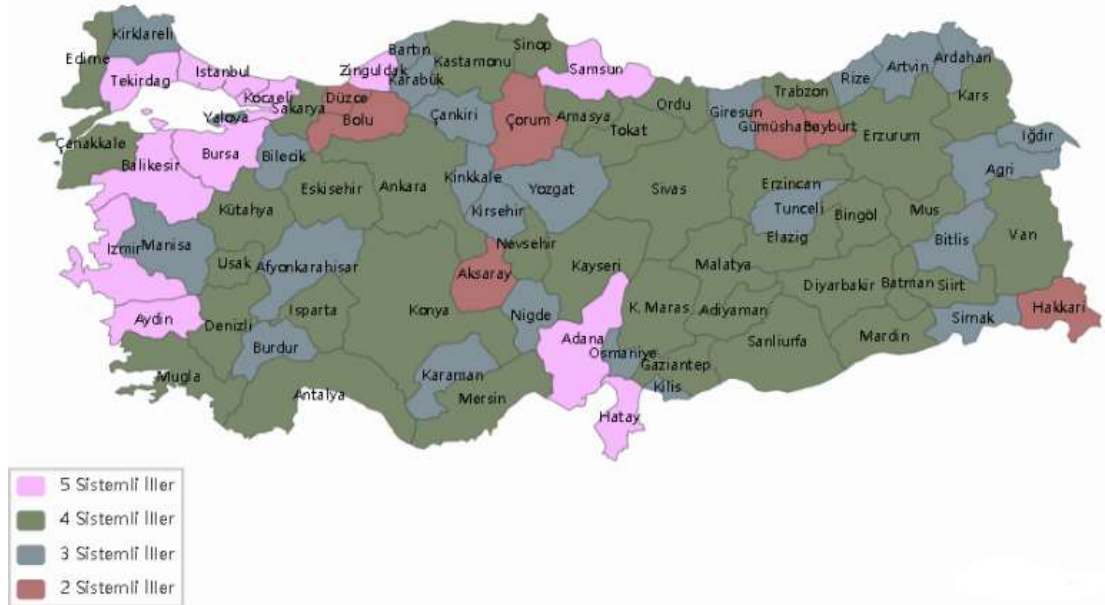
Yine Tablo 32’deki 2. Küme genellikle Türkiye ortalamasının üstünde olan ve çoğunluğu büyükşehir ünvanına sahip belli seviyde ulaşım sistemi ve yoğunluktaki illerden oluşmaktadır. 5. Küme içinde yine Türkiye’nin en gelişmiş durumdaki büyükşehir ünvanına sahip illerin yer aldığı, genellikle en işlek ticaret ve turizm potansiyelleri olan illerin toplandığı görülmektedir. 5. Küme’de bulunan İzmir, Bursa, Kocaeli ve Mersin illeri Türkiye’nin ithalat ve ihracat faaliyetlerine yönelik güçlü ulaşım tesislerine sahiptir. 1. Küme ise uluslararası boru hattı sistemlerine sahip ve işlek biçimde kullanılan kara ve demiryolu hatlarının üzerinden geçtiği iller yer almaktadır. 3. Küme ise dezavantajlı coğrafi konumlara sahip yeterince ulaşım altyapı ve yoğunluğa sahip olamayan iller toplanmıştır. Bu kümede genellikle yüzölçümü küçük illerin yer almış olması da mevcut kümede dikkat çeken bir diğer faktördür. İstanbul ili diğer tüm kümeleme algoritmalarında olduğu gibi

Türkiye'nin en yoğun ve gelişmiş ulaşım sistemine sahip ili olarak 4. Kümede yer alması oldukça beklenebilir bir durumdur.

Tablo EK: 9 ve Şekil 25'e göre karayolu, denizyolu, demiryolu, havayolu ve boru hattı ulaştırma türlerinin illerde aktif olarak kullanılıp kullanılmadığı belirtilmiştir. Tablo EK: 9'deki 1 değerleri o ulaşım türünün ilde aktif olarak kullanıldığı göstermektedir. Bu Tablo ve haritaya göre toplam 11 ilde ulaştırma türlerinin tamamı aktif olarak faaliyettedir. Bu iller Adana, Balıkesir, Aydın, Bursa, Hatay, İstanbul, İzmir, Kocaeli, Samsun, Tekirdağ, Zonguldak illeri olup kümeleme analizlerinde ulaşım faaliyetleri bakımından görece zayıf illerin bulunduğu kümelerde yer almamaktadırlar.

Şekil 25: İl Ulaştırması Faaliyet Durumu Türkiye İl Haritası

İl Bazılı Aktif Sistem Sayısı



Tablo 33: Kruskal Wallis Testi Sonuç Tablosu -2 (k-Ortalamalar Yöntemi)

DEĞİŞKENLER	Ki-Kare Değeri	Serbestlik Derecesi	P-Değeri
V1	35,620	4	,000
V2	37,794	4	,000
V3	47,954	4	,000
V4	33,942	4	,000
V5	34,067	4	,000
V6	40,405	4	,000
V7	38,941	4	,000
V8	32,721	4	,000
V9	33,407	4	,000
V10	40,761	4	,000
V11	40,041	4	,000
V12	40,131	4	,000
V13	20,955	4	,000
V14	36,195	4	,000
V15	24,805	4	,000
V16	28,162	4	,000
V17	33,761	4	,000
V18	28,207	4	,000
V19	27,105	4	,000
V20	25,673	4	,000
V21	27,865	4	,000
V22	26,740	4	,000
V23	49,404	4	,000
V24	26,740	4	,000
V25	47,680	4	,000
V26	26,095	4	,000
V27	42,219	4	,000
V28	48,532	4	,000
V29	38,371	4	,000
V30	45,149	4	,000
V31	27,605	4	,000
V32	45,943	4	,000
V33	10,576	4	,032
V34	51,317	4	,000
V35	45,065	4	,000
V36	38,623	4	,000
V37	53,741	4	,000
V38	35,526	4	,000
V39	31,911	4	,000
V40	5,472	4	,242

R 3.5.1 programında bulunan moments paketindeki çarpıklık ve basıklık değerlerini ölçmemizi sağlayan skewness ve kurtosis fonksiyonları yardımıyla Tablo 8'deki değerler elde edilmişti. Bu değerler ışığında veri setindeki değişkenlerin tamamı daha önce incelenen çarpıklık ve basıklık değerlerine göre normal dağılım göstermediği için parametrik olmayan istatistiksel yöntemler için kullanılan Kruskal-

Wallis testi kullanılmıştır. Bu test sonucuna göre $\alpha=0.05$ değerine göre sadece boru hattı sistem sayısı (V40) değişkeninin en az iki kümeyi ayıracak biçimde anlamlı olmadığı, diğer değişkenlerin en az iki kümeyi birbirinden ayırabilecek biçimde anlamlı olduğu belirlenmiştir. k-ortalamlar yöntemiyle ayrılmış bulunan bu grupların Kruskal-Wallis testi sonuçları Tablo 33 ve Tablo Ek-8’da gösterilmiştir.

SONUÇ VE DEĞERLENDİRME

Kümeleme analizi istatistik biliminde çok değişkenli istatistik yöntemlerinden birisidir. Kümeleme analiziyle illerin ulaştırma faaliyetlerini oluşturan göstergeler kullanılarak kümeleme yapısını belirlemenin ne derece olabileceği bu çalışmayla ortaya konulmaya çalışılmıştır. Bir bölgedeki ulaştırma faaliyetlerine bağlı göstergelerin sosyal, kültürel, ekonomik, askeri alanlarla doğrudan ilişkisi bulunmaktadır. Modern ve yaygın ulaşım sistemine sahip bir bölgede, söz konusu bu alanlarda olumlu bir tabloyu beraberinde getirmesi kaçınılmazdır. Bu çalışmada, Türkiye’deki illerin durumunun ulaşım faaliyetlerine bağlı göstergelere göre yorumlanarak doğru şekilde değerlendirilmesi amaçlanmıştır.

Bu amacı gerçekleştirmek için 2004-2018 yılları arasındaki çeşitli kamu kurumlarında bulunan karayolu, denizyolu, demiryolu, havayolu ve boru hattı verilerinden faydalanılmıştır. Ulaştırma faaliyetlerine bağlı göstergelerin içerdiği bu veriler Türkiye İstatistik Kurumu (T.Ü.İ.K.), Türkiye Cumhuriyeti Devlet Demiryolları (T.C.D.D.), Boru Hatları ile Petrol Taşıma Anonim Şirketi (BOTAŞ), Deniz Ticareti Genel Müdürlüğü, Devlet Hava Meydanları İşletmesi (D.H.M.İ.), Karayolları Genel Müdürlüğü (K.G.M.) kurumlarının internet ortamlarından alınmıştır. Bu veriler kullanılarak yapılan klasik kümeleme analizi yöntemleri için beş adet küme sayısı belirlenerek 81 ilin sınıflandırılması yapılmıştır. Küme sayısının belirlenmesinde küme geçerlilik indeksleri ve oluşturulan grafiklerden yararlanılmıştır. Veri setinde, uzaklık ölçüsü olarak öklid uzaklık ölçüsü kullanılmış ve 0-1 arası min-maks yöntemi kullanılarak standartlaştırma yapılmıştır. Verilerin çarpıklık ve basıklık değerleri incelenerek normal dağılıma uyup uymadığı incelenmiş ve kullanılan değişkenlerin tamamının normal dağılıma uygun olmadığına karar verilmiştir. Bu durum, değişkenlerin kümeleri ayırmadaki başarısını ölçmek için kullanılacak analizin Kruskal-Wallis testi olmasına neden olmuştur. Ayrıca değişkenlerin ölçü birimlerindeki farklılıklardan ötürü kullanılacak

standartlaştırma yönteminin 0-1 aralığında min-maks yöntemi olmasında da verilerin normal dağılmamış olmasının payı bulunmaktadır. Kruskal - Wallis testi sonucunda veri setindeki boru hattı sistemleri sayısı (V40) göstergesinin gruplara ayrılmasında bir önemi olmadığı ortaya çıkmış, diğer göstergelerin en az iki grubu başarıyla ayırdığı görülmüştür.

Çalışmanın yüksek doğrulukla tamamlanabilmesi için veri yapısının görüntülenmesi oldukça önemlidir. Bu sebeple kümeleme yapısının görüntülenebilmesi amacıyla çalışmada grafiklerden yoğun şekilde faydalanılmıştır. Bu konuda ilk başvuru olan yöntem, gözlemlerin konumlarının görüntülenebilmesi amacıyla iki boyutlu ve üç boyutlu saçılım grafiklerinin oluşturulmasıdır. İki veya üç değişkenli verilerin bu yöntemle veri yapılarının tam doğrulukta görüntülenebilmesi mümkün olsa da daha fazla değişkenin söz konusu olduğu veri setlerinde ana bileşenleri kullanarak belli doğrulukta bu grafikleri oluşturabilmek mümkün olmaktadır. Nitekim bu çalışmadaki 40 adet değişkene sahip veri setinde ana bileşenler analizi uygulanmıştır. Sonuç olarak, en büyük iki bileşenden veri setinin toplam varyansının %72'sini, üçüncü bileşenle birlikte toplam varyansının %79'unu açıklayabilecek şekilde iki ve üç boyutlu grafikler elde edilmiştir.

Sonraki aşamada, kümeleme yapısı grafiklerle ortaya konulduktan sonra hiyerarşik kümeleme analizi yöntemlerinden altısı ve hiyerarşik olmayan kümeleme yöntemlerinden ikisi kullanılmıştır. Hiyerarşik yöntemlerde, kümeleme analizi yöntemlerindeki başarı düzeyini ölçmek amacıyla kofenetik korelasyon katsayıları incelenmiştir. Bu yöntemler arasında en yüksek kofenetik korelasyon katsayısına sahip kümeleme yöntemi 0.96'dan fazla bir değer ile tek bağlantı yöntemi olmuştur. Bu da elde edilen kümeleme sonuçlarının hiyerarşik yöntemler arasında en başarılı yöntemin tek bağlantı yöntemi olduğunu göstermektedir. Tek bağlantı yöntemi sonuçlarına göre, İstanbul, Ankara, İzmir ve Antalya illeri sapan gözlem şeklinde ayrı kümeler olarak öne çıkmıştır. Diğer iller de başka bir kümede toplanmıştır. Bu sonuçların oluşması, bahsi geçen illerin diğer illere göre ulaştırma göstergeleri bakımından bariz farklılıklara işaret etmektedir. Bununla birlikte, hiyerarşik yöntemler arasındaki tek bağlantı yönteminin sapan gözlemlere karşı yapısı gereği hassas olmasının da bu sonucun ortaya çıkmasındaki payı büyüktür.

Hiyerarşik olmayan yöntemler, çok sayıda gözlem ve değişken içerikli büyük verilerde rahatlıkla ve hızlı biçimde analiz edebilmesi ve aykırı değerlere karşı daha az hassas olması nedeniyle araştırmacılar için genelde daha uygulanabilir

yöntemlerdir; Ancak, küme sayısının ayrıca belirlenmesinin gerekliliği ve ön değerlendirme amaçlı bir analiz olarak kullanılamaması gibi dezavantajları bulunmaktadır. Hiyerarşik olmayan kümeleme analizinde literatürde en sık ve yaygın şekilde uygulanan yöntemlerden birisi k-ortalamlar yöntemidir. Bu yöntemin çalışmadaki sonuçlarına göre İstanbul diğer yöntemlerde olduğu gibi tek kümede yer almıştır. İhracat-ithalat faaliyetlerinin yoğun olarak yapıldığı, gelişmiş liman ve tesislere sahip sanayileşmiş durumdaki Adana, Hatay, İzmir, Kocaeli ve Mersin illeri bir kümede, yine belli bir gelişmişlikte sanayi kollarına sahip Ankara, Antalya, Balıkesir, Bursa, Konya, Manisa ve Muğla illerinin ayrı bir kümede benzer gözlemler olarak diğer illere göre farklı bir konumda olduğu görülmektedir. Bu da illerdeki sanayileşme ile ulaşım faaliyetleri arasındaki güçlü ilişkinin bir işareti olarak yorumlanabilir.

Kullanılan veri seti tüm kümeleme analizi yöntemleriyle birlikte değerlendirildiğinde mutlak bir şekilde İstanbul ili, sahip olduğu devasa ulaştırma göstergeleriyle tek başına özel bir konumda ayrı bir küme olarak öne çıkmaktadır. Bu da diğer illere göre olağanüstü büyüklükteki ulaşım ağı ve alt yapı yoğunluğundan kaynaklanmakta, diğer illerle arasında büyük bir uçurum olduğu söylenebilir. Bunun yanında uygulanan analizler sonucunda sıkça öne çıkan iller Marmara Bölgesi'nde Balıkesir, Bursa, Kocaeli illeri, Ege Bölgesi'nde İzmir, Manisa, Muğla illeri, Akdeniz Bölgesi'nde Antalya ili, İç Anadolu Bölgesi'nde Ankara ve Konya illeri, Karadeniz Bölgesi'nde ise Samsun ilidir. Doğu ve Güneydoğu Anadolu bölgelerinde sapan gözlem özelliği gösteren il çıkmasa da özellikle coğrafi özelliklerin zorlu seviyelerde olduğu Hakkâri, Şırnak, Bitlis, Iğdır gibi iller ulaşım göstergeleri açısından oldukça benzer bir görüntü çizmektedir. Doğu Karadeniz bölgesi için de aynı durum söz konusudur. Genellikle bu bölgelerdeki benzer coğrafi şartlar ve yer şekillerinin bu sonuca etki ettiği değerlendirilebilir. Bir diğer önemli husus ise Ege, Akdeniz ve Marmara'ya kıyısı olan illerin uygulanan analiz çıktılarında diğer illere göre avantajlı bir konuma erişebildiği, ancak Karadeniz'e kıyısı olan illerde denizin sağladığı avantajların bu illere tam olarak yansımadağı gözlemlenmektedir. bu bölgede yalnızca Samsun ilinin uygun ulaştırma sistemlerine sahip olmasının da etkisiyle bölgede bir adım öne çıkmış görünmektedir. Bu durum, medoid yönteminde net bir biçimde gözlemlenebilmektedir.

literatürdeki iller bazında yapılmış birçok çalışma bulguları karşılaştırıldığında, mevcut çalışmada elde edilenlerden çok farklı olmadığı

görülmüştür. Birçok Doğu ve Güneydoğu Anadolu ili ile Doğu Karadeniz ili aynı ortak kümelerde toplanmıştır. Bu iller Türkiye'nin genel anlamda biraz daha geri kalmış illeri olarak bilindiği için ulaşım konusunda gelişime ihtiyaç duyduğu fikri ağır basmaktadır. Başta İstanbul ili başta olmak üzere özellikle Türkiye'nin en büyük kent merkezlerini barındıran Ankara, İzmir, Bursa, Antalya illeri ulaştırma göstergeleri baz alındığında sapan gözlem özelliklerine sahip olduğu ve diğerlerinden farklı olacak şekilde ayrı kümeler oluşturma eğiliminde oldukları gözlemlenmiştir. Sanayi, ticaret, finans merkezi durumunu almış illerin genelde aynı kümelerde yer almış olmaları, ulaştırma göstergelerinin bu durumla bir bağlantısı olduğu izlenimini sağlamıştır. Elde edilen sonuçlarda, mevcut illere ait özel konum ve yer şekillerinin de yakından ilişkili olduğu değerlendirilmektedir. Örneğin Konya ilinin coğrafi yapısının düz olması ve büyük nüfuslu bölgelere yakınlığı bu sonuçlara yansıdığı söylenebilir. Karadeniz Bölgesi'nde de özellikle hiyerarşik olmayan kümeleme analizi yöntemlerinden birisi olan medoid yönteminde Samsun ilinin bölgedeki diğer illere göre farklı belirgin şekilde gösterilmiştir. Samsun, Karadeniz Bölgesi'ndeki mevcut konumuyla en büyük kent durumundadır.

Bütün bu veriler ışığında çalışmada ön planda yer alan bazı sonuçlar değerlendirildiğinde hiyerarşik kümeleme analizi yöntemleri arasında en yüksek kofenetik korelasyon katsayısına sahip tek bağlantı yöntemi sonucunda çok istenen bir durum olmasa da İstanbul, Ankara, İzmir ve Antalya illeri bir objelik kümeler oluşturmuştur. Mevcut illerin Türkiye'de beş büyük il arasında olması ve tek bağlantı yönteminin sapan gözlemlere karşı olan duyarlılığının da bunda etkili olduğu değerlendirilmiştir. Fakat, Türkiye'nin en büyük beş kent merkezi arasında bulunan Bursa'nın diğer 76 ile göre bir farklılık oluşturmaması oldukça ilginçtir. Çünkü, Bursa ilinin ulaşım verileri göz önüne alındığında tıpkı Antalya, İzmir, Ankara illeri gibi büyükşehir ili olarak bir farklılık oluşturmaması beklenmektedir. Bursa ili gerek nüfusu, gerek sanayisi, gerek ulaşım verileri ve altyapısıyla bu duruma uygun görünmektedir.

Hiyerarşik olmayan kümeleme yöntemleri yapıları gereği özellikle sapan değerlerin çok fazla olduğu verilerde daha avantajlı olduğu söylenebilir. Hiyerarşik olmayan yöntemler konusunda ilk ön plana çıkan yöntem k-ortalamlar yöntemidir. Bu yöntemin çalışmadaki sonuçları değerlendirildiğinde belirlenen beş küme içinde Türkiye'nin en büyük nüfusa sahip ili olan, ekonomik, tarihi ve sosyokültürel açıdan farklı konumdaki İstanbul ili tek başına küme olarak kalmıştır. Antalya, Ankara,

Bursa, Balıkesir, Konya, Manisa, Muğla illeri de ayrı bir küme oluşturmuştur. Bu kümede gelişmiş ulaşım tesislerinin bulunması ve yoğun ulaşım faaliyetleri icra edilmesinin yanında bu faaliyetlerin varlığının bir işareti olarak gelişmiş sanayi ve turizme dönük tesisler bulunmaktadır. Adana, İzmir, Kocaeli, Hatay, Mersin illerinin oluşturduğu küme yoğun ithalat ve ihracat faaliyetlerinin gerçekleştirildiği bir kümedir. Bir diğer küme ise Afyonkarahisar, Aydın, Çanakkale, Denizli, Diyarbakır, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Isparta, Kahramanmaraş, Kars, Kayseri, Kütahya, Malatya, Mardin, Samsun, Sivas, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Uşak, Van illerinin oluşturduğu kümedir. Sözü edilen küme incelendiğinde birçok önemli ulaştırma unsurları göze çarpmaktadır. Doğu Karadeniz Bölgesi altındaki illerde, Doğu Ekspresi'nin yer aldığı demiryolu hattı, ayrıca yurt dışına çıkan önemli boru hatları ve demiryolu hatları geçmektedir. Karadeniz Bölgesi'ndeki Samsun ve Trabzon illeri, Marmara Bölgesi'ndeki Edirne ve Çanakkale illeri önemli liman geçişlerinin icra edildiği illerdir. Bu kümedeki İç Ege illeri de yoğunlaşmış demiryolu ve karayolu ağına ev sahipliği yapmaktadır. Bunların dışında kalan kümede ise ön planda bir ulaşım unsurunun göze çarpmadığı ve zayıf profilde illerin ağırlıkta olduğu (Hakkari, Ardahan, Ağrı vs.) büyük yüz ölçümüne sahip olmayan iller görülmektedir.

Analiz sonuçları göz önüne alınarak uygun veri setinde söz konusu yöntemler uygulandığında bilhassa k-ortalamlar yönteminin 2020 yılının başında ortaya çıkan Covid-19 pandemisinin Türkiye'deki başlangıç yayılımına yakın benzerlikte gruplama sonuçları verdiği görülmektedir. Dolayısıyla bu yöntem kullanılarak bir virüs yayılımının benzetiminin ulaştırma sistemlerinin analiz edilmesiyle yapılabileceği fikri ortaya çıkabilir. Bu benzetimin Covid-19 veya başka pandemilerde özellikle ilk yayılış safhalarında başarılı sonuçlar gösterebileceği düşünülmektedir. Kullanılan veri setinde yayılıma sebep olan ulaştırma göstergelerinin tamamının kullanılması halinde ulaşılan sonuçların benzer yayılım gösterecek illerin belirlenmesinde kullanılabilir. Bu tespitlere dayanarak her ilde daha uygun önlem ve politikalar geliştirilebilir.

İstanbul'dan farklı bir noktaya yatırım yapmak isteyen bir işletmeci mantığıyla çalışma sonuçları değerlendirildiğinde, örneğin yurt dışına kolayca bağlantı kurulabilecek konumda bir tesis yatırımı için k-ortalamlar yöntemindeki sonuçlara göre yurt dışına yoğun ulaşım faaliyetlerinin göze çarptığı Adana, İzmir, Kocaeli, Hatay, Mersin illeri kümesi ya da sanayi ve turizm tesislerinin yoğun

şekilde toplandıđı, bunun sebebinin de ulaşına bađlı olduđu deđerlendirilen Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa, Muđla illerinin toplandıđı küme titizlikle deđerlendirilebilir. Ayrıca yine önemli ulaşım faaliyetleri ve tesislerinin bulunduđu k-ortalamar yönteminde ortaya çıkan 3. Küme illeri de ayrıca bu sonuçlara göre ele alınabilir.

Araştırmacıların, bu çalışmanın sonuçlarını dikkate alarak, hiyerarşik yöntemlerde tekil küme oluşturan illeri, ya da diđer illerden farklılaşması beklendiđi halde bu beklentiyi karşılamayan illerde bunun altında yatan olası farklı sebeplerini araştırabilecekleri söylenebilir. Hiyerarşik yöntemlerde dendrogram görüntülerinin incelenmesiyle, çalışmayla ilgili farklı araştırmalar yapmak ve sonuçlara ulaşmak mümkün olabilir. Farklı küme sayılarıyla yapılacak araştırmalar da bu seçeneklerinden biridir. Örneđin Ward yöntemi ile olası küme sayısını tek başına 12 olarak belirleyerek bu konuda farklı bir çalışma yapılabilir. Aynı şekilde çalışmada ağırlık merkezi ve medyan yöntemlerinde meydana gelen tersine dönme durumları da ayrı bir araştırma konusu olabilir.

KAYNAKÇA

Aggarwal, C. C. ve Reddy, C. K. (2014). *Data Clustering: Algorithms and Applications*. Londra: Chapman ve Hall/CRC Data Mining and Knowledge Discovery Series.

Akgüngör, A. P. ve Demirel, A. (2011). Türkiye'deki Ulaştırma Sistemlerinin Analizi ve Ulaştırma Politikaları. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 10(3), 423-430.

Alonso, A., Monzón, A. ve Cascajo, R. (2015). Comparative Analysis of Passenger Transport Sustainability in European Cities. *Ecological Indicators*, 48, 578-592.

Alpar, R. (2017). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*. Ankara: Detay Yayıncılık.

Anderberg, M. R. (1973). The Broad View Of Cluster Analysis. *Cluster Analysis For Applications*, 1-9.

Aras, H. (2012). Deniz ve Suyolu Ulaşımı. *Ulaştırma Sistemleri*, 106-129. Eskişehir: Anadolu Üniversitesi

Atalay, A. ve Tortum, A. (2013). Trafik Hareketliliğine Göre Türkiye'de İllerin Kümelenmesi. *5. Karayolu Trafik Güvenliği Sempozyumu ve Sergisi*, 204.

Aydın, N. ve Seven, A.N. (2015). İl Nüfus ve Vatandaşlık Müdürlüklerinin İş Yoğunluğuna Göre Hibrid Kümeleme İle Sınıflandırılması. *Yönetim ve Ekonomi Araştırmaları Dergisi*, 13 (2), 181-201.

Aytaç, B. P., Çelik, F. ve Türe, F. (2007). Ülkemiz Ulaştırma Politikalarının Doğu Karadeniz Bölgesi'nin Kalkınması Üzerindeki Etkileri. *İnşaat Mühendisleri Odası*, 7, 137-143.

Baculakova, K. (2018). Cluster Analysis of Creative Industries in the Regions and Districts of Slovakia. *Theoretical and Empirical Researches in Urban Management*, 13(3), 74-89.

Bakanlığı, U. ve Üniversitesi, İ. T. (2005). Ulaştırma Ana Planı Stratejisi–Sonuç Raporu. *Ankara, Ulaştırma ve Ulaşım Araçları Uyg-Ar Merkezi, Şubat.*

Bakırcı, M. (2012) Ulaşım Coğrafyası Açısından Türkiye’de Havayolu Ulaşımının Tarihsel Gelişimi ve Mevcut Yapısı. *Marmara Coğrafya Dergisi*, (25), 340-377.

Beale, E.M.L. (1969). ‘*Euclidean Cluster Analysis*’, Bulletin of the International Statistical Institute 43, Book 2.

Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques. In *Grouping Multidimensional Data* (Pp. 25-71). Springer, Berlin, Heidelberg.

Bezdek, J. C. (1973). Cluster Validity with Fuzzy Sets.

Billard, L. ve Diday, E. (2019). *Clustering Methodology for Symbolic Data*. ABD: John Wiley ve Sons.

Brody, H., Rip, M. R., Vinten-Johansen, P., Paneth, N. ve Rachman, S. (2000). Map-making and Myth-Making in Broad Street: The London Cholera Epidemic, 1854. *The Lancet*, 356(9223), 64-68.

Bulut, H., (2018). *R Uygulamaları ile Çok Değişkenli İstatistiksel Yöntemler*. Nobel Akademik Yayıncılık, Ankara, Türkiye.

Bulut, H. (2019). Türkiye'deki İllerin Yaşam Endekslerine Göre Kümelenmesi. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23 (1) , 74-82.

Cabral, A. ve De Sousa Ramos, F. (2014). Cluster Analysis of the Competitiveness of Container Ports in Brazil. *Transportation Research Part A: Policy and Practice*, 69, 423-431.

Caliński, T. ve Harabasz, J. (1974). A Dendrite Method for Cluster Analysis. *Communications in Statistics-Theory and Methods*, 3(1), 1-27.

Carvalho, P. R., Munita, C. S. ve Lapolli, A. L. (2019). Validity Studies Among Hierarchical Methods of Cluster Analysis Using Cophenetic Correlation Coefficient. *Brazilian Journal of Radiation Sciences*, 7(2A).

- Cengiz, D. ve Öztürk, F. (2012). Türkiye’de İllerin Eğitim Düzeylerine Göre Kümeleme Analizi ile İncelenmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 14(1), 69-84.
- Charrad, M., Ghazzali, N., Boiteau, V. ve Niknafs, A. (2013). Nbclust R Package: Finding the Relevant Number of Clusters in a Dataset. *J. Stat. Softw.*
- Cheetham, A. H. ve Hazel, J. E. (1969). Binary (presence-absence) similarity coefficients. *Journal of Paleontology*, 1130-1136.
- Cormack, R. M. (1971). A Review of Classification. *Journal of the Royal Statistical Society: Series A (General)*, 134(3), 321-353.
- Çakmak, Z., Uzgören, N. ve Keçek, G. (2005). Kümeleme Analizi Teknikleri ile İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerin İncelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (12).
- Çelik, Ş., Şengül, T., Şengül, Ö. ve İnci, H. (2019). Türkiye’de İllere Göre Hayvansal ve Bitkisel Ürünlerin Kümeleme Analizi İle İncelenmesi. *Journal of Awareness*, 3 (5) , 385-398.
- Çilingirtürk, A. M. (2011). *İstatistiksel Karar Almada Veri Analizi*. Ankara: Seçkin Yayıncılık.
- Davies, D. L. ve Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2), 224-227.
- Duda, R. O., Hart, P. E. ve Stork, D. G. (1973). *Pattern Classification and Scene Analysis* (Vol. 3, Pp. 731-739). New York: Wiley.
- Dunn, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, 4(1), 95-104.
- Emran, S. M. ve Ye, N. (2001, June). Robustness Of Canberra Metric in Computer Intrusion Detection. In *Proc. IEEE Workshop on Information Assurance and Security, West Point, NY, USA* (Pp. 80-84).

Er, F. (2006). 2004 Yılında Türkiye'ye Gelen Yabancıların Taşıt Araçlarına Göre Dağılımının İstatistiksel Analizi. *Anatolia: Turizm Araştırmaları Dergisi*, 17(2), Güz: 184-188.

Erilli, N. A. (2009). Kümeleme Analizine Bulanık Yaklaşım Algoritmaları ve Uygulamaları, 19 Mayıs Üniv. *Fen Bilimleri Enst., Yüksek Lisans Tezi, Samsun. Bilimleri Enst., Samsun.*

Erilli, N. A. (2019). Comparison of Fuzzy Clustering Methods in Economic Freedom Ranking in Asia-Pacific. *Jurnal Perspektif Pembiayaan Dan Pembangunan Daerah*, 7(2), 157-168.

Erilli, N. A. ve Gündoğdu, Ö., (2019). Classification of Sub-Levels in Turkey with Gustafson-Kessel Method. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 20(2), 248-260.

Everitt, B. S. Landau, S., Leese, M. ve Stahl, D. (2011). *Cluster Analysis. 5th Edition*. United Kingdom: John Wiley and Sons. Ltd. 352

Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H. ve Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. In *Colloquium mathematicum* (Vol. 2, No. 3-4, pp. 282-285).

Frey, T. ve Van Groenewoud, H. (1972). A Cluster Analysis of the D2 Matrix of White Spruce Stands in Saskatchewan Based on the Maximum-Minimum Principle. *The Journal Of Ecology*, 873-886.

George, D. ve Mallery, P. (2019). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*. Routledge.

Gitto, S. ve Mancuso, P. (2019). Brand Perceptions of Airports Using Social Networks. *Journal of Air Transport Management*, 75, 153-163.

Gower, J. C. ve Ross, G. J. (1969). Minimum Spanning Trees and Single Linkage Cluster Analysis. *Journal Of The Royal Statistical Society: Series C (Applied Statistics)*, 18(1), 54-64.

Grabmeier, J. ve Rudolph, A. (2002). Techniques of Cluster Algorithms in Data Mining. *Data Mining and Knowledge Discovery*, 6(4), 303-360.

Grimm, L. G. ve Yarnold, P. R. (1995). *Reading and Understanding Multivariate Statistics*. USA: American Psychological Association.

Hair, J. F. Black, W. C., Babin, B. J., Anderson, R. E. ve Tatham, R. L. (2014). *Multivariate Data Analysis*. USA: Pearson.

Halkidi, M., Vazirgiannis, M. ve Batistakis, Y. (2000, September). Quality Scheme Assessment İn The Clustering Process. In *European Conference On Principles Of Data Mining And Knowledge Discovery* (Sayfa 265-276). Springer, Berlin: Heidelberg.

Hartigan, J. A. (1981). Consistency of Single Linkage for High-Density Clusters. *Journal of the American Statistical Association*, 76(374), 388-394.

Haustein, S. ve Nielsen, T. A. S. (2016). European Mobility Cultures: A Survey-Based Cluster Analysis Across 28 European Countries. *Journal of Transport Geography*, 54, 173-180.

Höft, M., Barik, S. K. ve Lykke, A. M. (1999). Quantitative Ethnobotany. *Applications of Multivariate and Statistical Analyses in Ethnobotany. People and Plants Working Paper*, 6, 1-49.

Hubert, L. J. ve Levin, J. R. (1976). A General Statistical Framework for Assessing Categorical Clustering in Free Recall. *Psychological Bulletin*, 83(6), 1072.

Irani, J., Pise, N. ve Phatak, M. (2016). Clustering Techniques and the Similarity Measures Used in Clustering: A Survey. *International Journal of Computer Applications*, 134(7), 9-14.

İncekırık, A. (2005). *Çok Değişkenli İstatistiksel Bir Boyut İndirgeme Yöntemi Olarak Kümeleme Analizi ve Bir Uygulama*. (Yüksek Lisans Tezi). İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

İncekırık, A. (2014). *Türkiye’de Faaliyet Gösteren Farklı Ölçeklerdeki Şirketlerin Fuzzy Kümeleme Analizi ile Gruplandırılması*. (Doktora Tezi). İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Johnson, D. E. (1998). *Applied Multivariate Methods for Data Analysts*. Duxbury Resource Center.

Johnson, R. A. ve Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis 6th Edition*. Upper Saddle River, NJ: Prentice Hall.

Kalaycı, Ş. (2014). *SPSS Uygulamaları Çok Değişkenli İstatistik Teknikleri*, Ankara: Asil Yayınevi, 6. Baskı.

Karaoğlu, Y. (2018). Farklı Kümeleme Analizi Yöntemleri İle Türkiye’deki İllerin Sosyoekonomik Özelliklerine Göre Sınıflandırılması ve Karşılaştırılması. (Yüksek Lisans Tezi). Sivas: Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü.

Kassambara, A. ve Mundt, F. (2017). R Package ‘Factoextra’. *Extract and Visualize the Results of Multivariate Data Analyses*, 76.

Kaya, E. (2012). Ulaştırma Kavramı ve Önemi. *Ulaştırma Sistemleri*, 2-23. Eskişehir: Anadolu Üniversitesi

Klinger, T., Kenworthy, J. R. ve Lanzendorf, M. (2013). Dimensions of Urban Mobility Cultures—A Comparison of German Cities. *Journal of Transport Geography*, 31, 18-29.

Korkmaz, M. ve Alacahan, N. D. (2013). Karayolu Yolcu Taşımacılığı Katma Değer Tahmininin Detaylı Analizi. *Karadeniz Uluslararası Bilimsel Dergi*, 1 (20) , 85-101.

Komsta, L. ve Novomestky, F. (2015). R Package Version, 14. *Moments, Cumulants, Skewness, Kurtosis And Related Tests*.

Lachapelle, U., Burke, M., Brotherton, A. ve Leung, A. (2018). Parcel Locker Systems In A Car Dominant City: Location, Characterisation and Potential Impacts On City Planning and Consumer Travel Access. *Journal of Transport Geography*, 71, 1-14.

Lessig, V. P. (1972). Comparing Cluster Analyses with Cophenetic Correlation. *Journal Of Marketing Research*, 9(1), 82-84.

Levent, E. B. (2016). *Veri Madenciliği ve Havacılık Sektöründe Bir Uygulama*. (Yüksek Lisans Tezi). İstanbul: Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü.

Ligges, U., Maechler, M., Schnackenberg, S. ve Ligges, M. U. (2018). R Package 'Scatterplot3d'. *Recuperado De [https://Cran.Rproject.Org/Web/Packages/Scatterplot3d/Scatterplot3d. Pdf](https://cran.r-project.org/web/packages/Scatterplot3d/Scatterplot3d.pdf)*, (23.01.2021)

Mcclain, J. O. ve Rao, V. R. (1975). Clustisz: A Program to Test for the Quality of Clustering of a Set of Objects. *Journal of Marketing Research*, 456-460.

Macqueen, J. (1967, June). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, No. 14, Pp. 281-297).

Malkoç, H. (2018). Kümeleme Analizinde Yapay Sinir Ağlarının Küme Geçerlilik İndeksi Olarak Kullanılması. (Yüksek Lisans Tezi). Sivas: Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü.

Mardia, K. V., Kent, J. T. ve Bibby, J. M. (1979). *Multivariate Analysis*, New York: Academic Press, London.

Marriott, F. (1971). Practical Problems in a Method of Cluster Analysis. *Biometrics*, 501-514.

Markov, I., Arampatzis, A. ve Crestani, F. (2012, August). Unsupervised Linear Score Normalization Revisited. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Pp. 1161-1162).

Marta, Z. (2017). Cluster Analysis of World's Airports on the Basis of Number of Passengers Handled (Case Study Examining the Impact of Significant Events). *Statistika-Statistics And Economy Journal*, 97(1), 74-88.

- Mayer, R. (2016). Airport Classification Based on Cargo Characteristics. *Journal of Transport Geography*, 54, 53-65.
- Mcclain, J. O. ve Rao, V. R. (1975). Clustisz: A Program To Test For The Quality Of Clustering Of A Set Of Objects. *Journal Of Marketing Research*, 456-460.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. ve Studer, M. (2013). R Package ‘cluster’. *Dosegljivo na*
- Milligan, G. W. (1981). A Monte Carlo Study of Thirty Internal Criterion Measures for Cluster Analysis. *Psychometrika*, 46(2), 187-199.
- Milligan, G. W. ve Cooper, M. C. (1987). Methodology review: Clustering Methods. *Applied Psychological Measurement*, 11(4), 329-354.
- Morgan, B. J. ve Ray, A. P. (1995). Non-uniqueness and Inversions in Cluster Analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(1), 117-134.
- Müdürlüğü, T. İ. G. (2016). Demiryolu Sektör Raporu 2016. *Ankara: Türkiye Cumhuriyeti Devlet Demiryolları*. <http://www.tcdd.gov.tr/files/istatistik/stratejik.pdf> (Erişim tarihi: 26.08.2020)
- Nalçakan, M., Tutar, F. ve Tutar, E. (2012). Ulaştırma Sektörünün Yarattığı Dışsal Maliyetlerin Ekonomik Etkileri. *Journal Of Social Sciences/Sosyal Bilimler Dergisi (2146-4561)*, 2(4).
- Nasri, A. ve Zhang, L. (2018). A Multi-Dimensional Multi-Level Approach to Measuring the Spatial Structure of U.S. Metropolitan Areas. *Journal of Transport and Land Use*, 11(1), 49-65.
- Nezerenko, O., Koppel, O. ve Tuisk, T. (2017). Cluster Approach in Organization of Transportation in the Baltic Sea Region. *Transport*, 32(2), 167-179.
- Öcalır, Akünal, E. ve Erol, S. (2016). Using Cluster Analysis to Define the Position of a Developing Country in Global Transportation Services Trade Environment. *Gazi University Journal of Science*, 29 (4) , 751-767.

Özeri, G. (2019). *İktisadi İşbirliği ve Gelişme Teşkilatına Üye Ülkeler Arasında Eğitim Bakımından Türkiye'nin Konumu*. (Yüksek Lisans Tezi). İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Özdamar, K. (2004). *Paket Programlar ile İstatistiksel Veri Analizi 5. Baskı*, Eskişehir: Kaan Kitabevi.

Öztürk, L., Çiftçi G. E. ve Hırlak B. (2017) Kümeleme Analizi Kullanarak Kadının Yaşam Kalitesi ve Profili Açısından İllerin Homojen Kümelere Ayrılması, *Gazi Üniversitesi Sosyal Bilimler Dergisi*, Cilt 4, Sayı 10, Sayfalar 301 – 318.

Park, C., Lee, J. ve Sohn, S. Y. (2019). Recommendation of Feeder Bus Routes Using Neural Network Embedding-Based Optimization. *Transportation Research Part A: Policy and Practice*, 126, 329-341.

Ratkowsky, D.A. ve Lance, G.N. (1978). Criterion for Determining the Number of Groups in a Classification. *Australian Computer Journal*. 1978; 10(3): 115-117 11 refs: <http://hdl.handle.net/102.100.100/300266?index=1>

Ren, P. ve Li, L. (2018). Characterizing Air Traffic Networks Via Large-Scale Aircraft Tracking Data: A Comparison Between China and the US Networks. *Journal of Air Transport Management*, 67, 181-196.

Rendon, E., Abundez, I., Arizmendi, A. ve Quiroz, E. M. (2011). Internal Versus External Cluster Validation Indexes. *International Journal of Computers and Communications*, 5(1), 27-34.

Revelle, W. ve Revelle, M. W. (2015). R Package 'Psych'. *The Comprehensive R Archive Network*, 337, 338.

Rocha, M., Silva, C. A. M., Junior, R. G. S., Anzanello, M., Yamashita, G. H. ve Lindau, L. A. (2020). Selecting the Most Relevant Variables Towards Clustering Bus Priority Corridors. *Public Transport*, 12(3), 587-609.

Rodrigue, J. P., Comtois, C. ve Slack, B. (2013). Transportation, Economy and Society. *The Geography of Transport Systems, 3rd Edn*. New York: Routledge.

Rousseeuw, P. J. (1987a). Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Rousseeuw, L. ve Kaufman, P. (1987b). Clustering by Means of Medoids.

Sapp, M., Obiakor, F. E., Gregas, A. J. ve Scholze, S. (2007). Mahalanobis Distance: A Multivariate Measure Of Effect in Hypnosis Research. *Sleep and Hypnosis*, 9(2), 67.

Saraçlı, S., Doğan, N. ve Doğan, İ. (2013). Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation. *Journal of Inequalities and Applications*, 2013(1), 1-8.

Sarle, W. S. (1983). *Cubic Clustering Criterion*. SAS Institute.

Seber, G. A. (2009). *Multivariate Observations*. USA: John Wiley ve Sons, Inc.

Sharma, S. (1996). *Applied Multivariate Techniques*. USA: John Wiley ve Sons, Inc.

Small, K. A., Verhoef, E. T. ve Lindsey, R. (2007). *The Economics of Urban Transportation*. Routledge.

Sokal, R. R. ve Michener, C.D. (1958). A Statistical Method for Evaluating Systematic Relationships. *Univ. Kansas, Sci. Bull.*, 38, 1409-1438.

Sneath, P. H. (1957). The Application of Computers to Taxonomy. *Microbiology*, 17(1), 201-226.

Sokal, R. R. ve Rohlf, F. J. (1962). The Comparison of Dendrograms by Objective Methods. *Taxon*, 11(2), 33-40.

Sorensen, T. A. (1948). A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. *Biol. Skar.*, 5, 1-34.

Sönmez, H. ve Er, F. (2007). Türkiye’de İllere Göre İç Göç Hareketlerinin Modern Kümeleme Teknikleri ile İncelenmesi. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi* , 20 (1) , 17-32.

Şahin, G., Sipahioğlu, A., Kamışlı Öztürk, Z., Ertek, G., Aba, B. ve Büyüközkan, G. (2012). *Uluslararası Lojistik*. Eskişehir: Anadolu Üniversitesi.

Team, R. C. (2013). R Package “Stats.”. *RA Lang. Environment Stat. Comput. Vienna, Austria: R Foundation for Statistical Computing*.

Tinsley, Howard E. A. ve Brown, S. D. (2000). *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, Academic Pr Inc.

Tryon, R. C. (1940). Genetic Differences in Mazelarning Ability in Rats. *Yearbook of the National Society for the Study of Education*.

UDHB (Ulaştırma Denizcilik ve Haberleşme Bakanlığı). (2017). İstatistiklerle Ulaştırma, Denizcilik ve Haberleşme (2003-2017). Ankara: UDHB.

Ulaştırma, D. ve Bakanlığı, H. (2013). 11. Ulaştırma, Denizcilik ve Haberleşme Şurası: Havacılık ve Uzay Teknolojileri Çalışma Grubu Raporu. *Ankara, Ulaştırma, Denizcilik ve Haberleşme Bakanlığı*.

Wang, T. C. ve Pham, Y. T. H. (2020). An Application of Cluster Analysis Method to Determine Vietnam Airlines’ Ground Handling Service Quality Benchmarks. *Journal of Advanced Transportation*, 2020.

Waters, D. (2003). *Logistics an Introduction to Supply Chain Management*. London: Palgrave Macmillan.

Wickham, H., Bryan, J., Kalicinski, M., Valery, K., Leittenne, C., Colbert, B., Hoerl, D., Miller, E. ve Bryan, M. J. (2019). R Package ‘readxl’.

Wickham, H. ve Seidel, D. (2016). Scales: Scale Functions for Visualization. *R Package Version 0.4.0*, <https://cran.r-project.org/web/packages/scales/index.html>, (23.01.2021)

Wierzchoń, S. T. ve Kłopotek, M. (2018). *Modern Algorithms of Cluster Analysis* (P. 421). Berlin, Germany: Springer.

Williams, W. T., Lambert, J. M. ve Lance, G. N. (1966). Multivariate Methods in Plant Ecology. *J. Ecol*, 54, 427-5.

Wolf, I. ve Schröder, T. (2019). Connotative Meanings of Sustainable Mobility: A Segmentation Approach Using Cultural Sentiments. *Transportation Research Part A: Policy and Practice*, 126, 259-280.

Xukuo, G. ve Qiong, W. (2013). Research on the Mode of Present Transportation in China and the Analysis of Railway Transportation. *6th International Conference on Information Management, Innovation Management and Industrial Engineering* (Vol. 3, Pp. 414-416). IEEE.

Yalçın, S. ve Ayyıldız, E. (2018). Analysis of Airports Using Clustering Methods: Case Study In Turkey. *Journal of Management Marketing and Logistics*, 5(3), 194-205.

Yaşar E. (2009). *Veri Doğrudan Yabancı Yatırımların Belirlenmesine Yönelik Kümeleme Analizi ve 1980–2005 Dönemi Homojen Ülke Gruplarının Belirlenmesi* (Doktora Tezi). Kütahya: Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü.

Yılmaz, Z. ve Uzgören, E. (2014). Türkiye’de İllerin Temel Bankacılık Faaliyetleri Yönünden Kümeleme Analizi Yöntemiyle Sınıflandırılması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 535-554.

Zemp, S., Stauffacher, M., Lang, D. J. ve Scholz, R. W. (2011). Classifying Railway Stations for Strategic Transport and Land Use Planning: Context Matters!. *Journal of Transport Geography*, 19(4), 670-679.

EKLER

EK: 1 (R Kodları)

ulastirma <- read_excel("C:/analiz2018.xlsx") # Excel dosyası okumak için kullanılır. Okuması için analiz2018.xlsx isimli dosyanın tam olarak bilgisayarın c hard diskinin içine atılması gerekmektedir. Bu dosyanın içeriği de “ulastirma” ifadesine aktarılır.

ulastirma_sayisal_veri<-ulastirma[,4:43] # kullanılacak sayısal verileri 4.sütundan 43. Sütuna kadar seçerek “ulastirma_sayisal_veri” ifadesine aktarır (40 değişken).

iller<-ulastirma[,3] # illerin analizlerde yazıya dökülmesi için 3. Sütun verileri seçilir

plot_icin_veri <- ulastirma_sayisal_veri # ”plot_icin_veri” ifadesine ulastirma_sayisal_veri ifadesi dosyasındaki veriler aktarılır. Bu ifade iki boyutlu ana bileşen skor grafiğine aktarım için kullanılacaktır.

plot_icin_veri <- scale(plot_icin_veri) # grafiksel okuma için plot_icin_veri ifadesinde hazırlama işlemi yapılır. Bu komut olmadan grafikler düzgün görüntülenememektedir.

labeladi<-iller\$`BÖLGE ADI` # iller\$`BÖLGE ADI` ifadesinden il isimlerini aktarır.

rownames(plot_icin_veri)<-labeladi # il isimlerinin bir üstteki ifadede atanmasıdır.

rownames(ulastirma_sayisal_veri)<-labeladi #ulastirma_sayisal_veri veri setinin gözlem isimlerini atar.

mattris<-data.matrix(ulastirma_sayisal_veri) #frameden matris formuna çevirir.

for (i in 1:length(ulastirma_sayisal_veri[,1])) {ulastirma_sayisal_veri[,i]<- scales::rescale(mattris[,i], to=c(0,1))}# ulastirma_sayisal_veri veri setinin sütun uzunluğu kadar for döngüsü yardımıyla min-maks yöntemi uygulanır.

mesafe<-dist(ulastirma_sayisal_veri) # Öklid uzaklık ölçüsü uygulanır ve bu değerleri “mesafe” ifadesine aktarır.

cl<-hclust(mesafe,"complete") #kümeleme analizi tam bağlantı yöntemi uygulanır.

plot(cl, hang = -1) #dendrogram grafiği oluşturur. “hang” ifadesi dendrogramdaki taban çizgilerinde görsel olarak değişiklikler yapar.

`aglomerasyoncomplete<-data.frame(cbind(abs(cl[["merge"]]),cl[["height"]]))` # tam bağlantı yöntemi aglomerasyon Tablosu oluşturur. “merge” gözlemleri, “height” ise birleştikleri noktanın yüksekliğini verir.

`cl<-hclust(mesafe,"average" )` #kümeleme analizi ortalama bağlantı yöntemi uygular.

`plot(cl, hang = -1)` #dendrogram oluşturur. “hang” ifadesi dendrogramdaki taban çizgilerinde görsel olarak değişiklikler yapar.

`aglomerasyonaaverage<-data.frame(cbind(abs(cl[["merge"]]),cl[["height"]]))` # Aglomerasyon Tablosu oluşturur. “merge” gözlemleri, height ise birleştikleri noktanın yüksekliğini verir.

`cl<-hclust(mesafe,"single" )` #kümeleme analizi tek bağlantı yöntemi uygular.

`plot(cl, hang = -1)` # dendrogram oluşturur. “hang” ifadesi dendrogramdaki taban çizgilerinde görsel olarak değişiklikler yapar.

`aglomerasyonsingle<-data.frame(cbind(abs(cl[["merge"]]),cl[["height"]]))` # Aglomerasyon Tablosu oluşturur. Merge gözlemleri, height ise birleştikleri noktanın yüksekliğini verir.

`cl<-hclust(mesafe,"ward.D" )` #kümeleme analizi tek bağlantı yöntemi uygular.

`plot(cl, hang = -1)` # dendrogram oluşturur.

`aglomerasyonward<-data.frame(cbind(abs(cl[["merge"]]),cl[["height"]]))` # Aglomerasyon Tablosu oluşturur. Merge gözlemleri, height ise birleştikleri noktanın yüksekliğini verir.

`cl<-hclust(mesafe,"centroid" )` # kümeleme analizi ağırlık merkezi yöntemi uygular.

`plot(cl, hang = -1)` #dendrogram

```
aglomerasyoncentroid<-data.frame(cbind(abs(cl[["merge"]]),cl[["height"]])) #  
Aglomerasyon Tablosu oluşturur. Merge gözlemleri, height ise birleştikleri noktanın  
yüksekliğini verir.
```

```
cl<-hclust(mesafe,"median" ) # kümeleme analizi medyan yöntemi uygular.
```

```
plot(cl, hang = -1) #dendrogram
```

```
aglomerasyonmedian<-data.frame(cbind(abs(cl[["merge"]]),cl[["height"]])) # üstteki  
yöntemin aglomerasyon Tablosunu oluşturur.
```

```
averege<-hclust(mesafe,"average" ) #kümeleme analizi ortalama bağlantı yöntemi  
uygular ve average ifadesine aktarır.
```

```
single<-hclust(mesafe,"single" ) #kümeleme analizi tek bağlantı yöntemi uygular ve  
single ifadesine aktarır.
```

```
complete<-hclust(mesafe,"complete" ) #kümeleme analizi tam bağlantı yöntemi  
uygular ve complete ifadesine aktarır.
```

```
ward<-hclust(mesafe,"ward.D" ) #kümeleme analizi ward yöntemi uygular ve  
“ward” ifadesine aktarır.
```

```
centroid<-hclust(mesafe,"centroid" ) #kümeleme analizi ağırlık merkezi yöntemi  
uygular ve “centroid” ifadesine aktarır.
```

```
median<-hclust(mesafe,"median" ) # kümeleme analizi medyan yöntemi uygular ve  
“median” ifadesine aktarır.
```

```
kofenetik<-cophenetic(single) #tek bağlantı yönteminde kofenetik uzaklıkları  
hesaplar ve “kofenetik” ifadesine aktarır.
```

```
cor(mesafe,kofenetik) #kofenetik uzaklık ile Öklid uzaklık verileri arasındaki  
korelasyonu hesaplar.
```

```
kofenetik<-cophenetic(complete) #tam bağlantı yönteminde kofenetik uzaklıkları  
hesaplar ve “kofenetik” ifadesine aktarır.
```

`cor(mesafe,kofenetik) #kofenetik uzaklık ile Öklid uzaklık verileri arasındaki korelasyonu hesaplar.`

`kofenetik<-cophenetic(ward) #ward bağlantı yönteminde kofenetik uzaklıkları hesaplar ve “kofenetik” ifadesine aktarır.`

`cor(mesafe,kofenetik) #kofenetik uzaklık ile Öklid uzaklık verileri arasındaki korelasyonu hesaplar.`

`kofenetik<-cophenetic(average) #ortalama bağlantı yönteminde kofenetik uzaklıkları hesaplar ve “kofenetik” ifadesine aktarır.`

`cor(mesafe,kofenetik) #kofenetik uzaklık ile Öklid uzaklık verileri arasındaki korelasyonu hesaplar.`

`kofenetik<-cophenetic(centroid) #ağırlık merkez bağlantı yönteminde kofenetik uzaklıkları hesaplar ve “kofenetik” ifadesine aktarır.`

`cor(mesafe,kofenetik) #kofenetik uzaklık ile Öklid uzaklık verileri arasındaki korelasyonu hesaplar.`

`kofenetik<-cophenetic(median) #medyan bağlantı yönteminde kofenetik uzaklık değerlerini hesaplar ve “kofenetik” ifadesine aktarır.`

`cor(mesafe,kofenetik) #kofenetik uzaklık ile Öklid uzaklık verileri arasındaki korelasyonu hesaplar.`

`labeladi<-iller$`BÖLGE ADI` # veri setinde etiket ismi aktarma`

`distance <- get_dist(ulastirma_sayisal_veri) # Öklid uzaklık matrisi oluşturulması ve distance ifadesine atanması`

`fviz_dist(distance, gradient = list(low = "#00AFBB", mid = "white", high = "#FC4E07")) #distance matrisi için ısı matris grafiği oluşturur.`

`rownames(ulastirma_sayisal_veri)<-labeladi # ulastirma_sayisal_veri ifadesinin satır isimlerini “labeladi” ifadesindeki karakter içeriklerinden oluşturur.`

`mattris<-data.matrix(ulastirma_sayisal_veri)` #frameden matrise çevirme işlemi yapar ve matris ifadesine aktarır.

`pc <- princomp(ulastirma_sayisal_veri[,1:4], cor=TRUE, scores=TRUE)` #
“ulastirma_sayisal_veri” ifadesinin ilk 4 sütunun alarak temel bileşenleri, korelasyon matrisi seçip skor gösterimini yaparak pc ifadesine aktarır.(pc=principal compenent)

`veriler<-data.frame(pc$scores[,1:3])` #ana bileşen skorlarını veri çerçevesine alarak veriler ifadesine atar.

`rownames(veriler)<-labeladi` # veriler ifadesinde bulunan veri çerçevesine gözlemlerin isimlerini “labeladi” adlı ifadeden alarak atar.

`with(data.frame(veriler), {`

`s3d <- scatterplot3d(Comp.1, Comp.2, Comp.3,` # x y ve z vektörleri için ilk 3
bileşeni atar.

`color="blue", pch=19,` # mavi renkli nokta şeklinde gözlemler

`type="h",` # x-y düzlemine dikey hatlar çizer.

`main="Üç Ana Bileşen Skor Grafiği- %79.12",`
`xlab="PCA1-%60,959",` `ylab="PCA2-%11,232",`
`zlab="PCA3-%6,929")` #main grafiğin başlığını x,y ve zlab ifadeleri de şeklin kenar taraflarına yazılacak yazıları belirtir.

`s3d.coords <- s3d$xyz.convert(Comp.1, Comp.2, Comp.3)` # 2 boyutlu
izdüşümleri üçüncü boyuta genişletir.

`text(s3d.coords$x, s3d.coords$y,` # x ve y koordinatları

`labels=row.names(veriler),` # etiketleri grafiğe aktarır.

`cex=.6, pos=4)` # metni %50 küçülterek gözlem noktalarının sağ tarafına koyar.)

`})`

`indekslersingle<-NbClust(data=ulastirma_sayisal_veri, distance="euclidean", min.nc=5, max.nc=12, method="single", index="all")` # “ulastirma_sayisal_veri” veri matrisini Öklid uzaklıkla tek bağlantı yöntemi için alt limit 5, üst limit 12 küme sayısı olmak üzere programdaki tüm geçerlilik indekslerini kullanarak kaç adet küme sayısı olması gerektiğini yazar. Ulaşılan sayısal sonuçlar da “indekslersingle” ifadesine aktarılır.

`eniyilersingle<-indekslersingle[["Best.nc"]]` # indeks sonuç ve değerleri Tablosu oluşturup “eniyilersingle” ifadesine aktarır.

`indekslerkmeans<-NbClust(data=ulastirma_sayisal_veri, distance="euclidean", min.nc=5, max.nc=12, method="kmeans", index="all")` # “ulastirma_sayisal_veri” veri matrisini Öklid uzaklıkla k-ortalamlar yöntemi için alt limit 5, üst limit 12 küme sayısı olmak üzere programdaki tüm geçerlilik indekslerini kullanarak kaç adet küme sayısı olması gerektiğini yazar. Ulaşılan sayısal sonuçlar da “indekslerkmeans” ifadesine aktarılır.

`eniyilerkmeans<-indekslerkmeans[["Best.nc"]]` # indeks sonuç ve değerleri Tablosu oluşturup “eniyilerkmeans” ifadesine aktarır.

`indeksleraverage<-NbClust(data=ulastirma_sayisal_veri, distance="euclidean", min.nc=5, max.nc=12, method="average", index="all")` # “ulastirma_sayisal_veri” veri matrisini Öklid uzaklıkla ortalama bağlantı yöntemi için alt limit 5, üst limit 12 küme sayısı olmak üzere programdaki tüm geçerlilik indekslerini kullanarak kaç adet küme sayısı olması gerektiğini yazar. Ulaşılan sayısal sonuçlar da “indeksleraverage” ifadesine aktarılır.

`eniyileraverage<-indeksleraverage[["Best.nc"]]` # indeks sonuç ve değerleri Tablosu oluşturup “eniyileraverage” ifadesine aktarır.

`indekslerward<-NbClust(data=ulastirma_sayisal_veri, distance="euclidean", min.nc=5, max.nc=12, method="ward.D", index="all")` # “ulastirma_sayisal_veri” veri matrisini Öklid uzaklıkla ortalama bağlantı yöntemi için alt limit 5, üst limit 12 küme sayısı olmak üzere programdaki tüm geçerlilik indekslerini kullanarak kaç adet

küme sayısı olması gerektiğini yazar. Ulaşılan sayısal sonuçlar da “indekslerward” ifadesine aktarılır.

```
eniyilerward<- indekslerward[["Best.nc"]] # indeks sonuç ve değerleri Tablosu  
oluşturup “eniyilerward” ifadesine aktarır.
```

```
indekslercomplete<-NbClust(data=ulastirma_sayisal_veri, distance="euclidean",  
min.nc=5, max.nc=12, method="complete", index="all") # “ulastirma_sayisal_veri”  
veri matrisini Öklid uzaklıkla ortalama bağlantı yöntemi için alt limit 5, üst limit 12  
küme sayısı olmak üzere programdaki tüm geçerlilik indekslerini kullanarak kaç adet  
küme sayısı olması gerektiğini yazar. Ulaşılan sayısal sonuçlar da “indekslerward”  
ifadesine aktarılır.
```

```
eniyilercomplete<- indekslercomplete[["Best.nc"]] # indeks sonuç ve değerleri  
Tablosu oluşturup “eniyilercomplete” ifadesine aktarır.
```

```
indekslermedian<-NbClust(data=ulastirma_sayisal_veri, distance="euclidean",  
min.nc=5, max.nc=12, method="ward.D", index="all") # “ulastirma_sayisal_veri”  
veri matrisini Öklid uzaklıkla ortalama bağlantı yöntemi için alt limit 5, üst limit 12  
küme sayısı olmak üzere programdaki tüm geçerlilik indekslerini kullanarak kaç adet  
küme sayısı olması gerektiğini yazar. Ulaşılan sayısal sonuçlar da “indekslerward”  
ifadesine aktarılır.
```

```
eniyilermedian<- indekslermedian[["Best.nc"]] # indeks sonuç ve değerleri Tablosu  
oluşturup “eniyilermedian” ifadesine aktarır.
```

```
labeladi<-iller$`BÖLGE ADI`
```

```
rownames(plot_icin_veri)<-labeladi # “labeladi” ifadesinde bulunan isimleri  
“plot_icin_veri” data matrisine atama işlemini yapar.
```

```
ulastirma_kmeanskort <- kmeans(ulastirma_sayisal_veri, 5, nstart = 25) # 5 adet  
küme sayısı belirlenerek k-ortalamalar yöntemi ve 25 farklı başlangıç kaynak noktası  
başlangıcı seçer ve uygular. Bu başlangıç noktalarının çok olması doğru sonuca  
ulaşılması olasılığını artırır.
```

`merkezler<-data.frame(ulastirma_kmeanskort$centers)` #k ortalamalar yönteminde ulaşılan kümelerin merkezlerini “merkezler” ifadesine aktarır.

`medoidyontemi<-pam(plot_icin_veri, k=5)` # küme sayısı 5 olacak şekilde medoid yöntemi uygular ve uygulama sonuçlarını “medoidyontemi” ifadesine aktarılır.

`kortkumemesafeleri<-dist(merkezler)` #k-ortalamalar yönteminde küme merkezleri arası mesafeyi öklid uzaklıkla hesaplayıp “kortkumemesafeleri” ifadesine aktarır.

`skewandkurt<-`

`data.frame(cbind(skewness(ulastirma[,4:43]),kurtosis(ulastirma[,4:43])))` # çarpıklık ve basıklık değerleri Tablosu oluşturarak “skewandkurt” ifadesine aktarır.

`Temelistatistikler<-describe(ulastirma[,4:43], skew = FALSE)` #iller veri seti için bazı temel istatistik verilerini Tablolaştırarak “Temelistatistikler” ifadesine aktarır.

`fit <- princomp(ulastirma_sayisal_veri, cor=TRUE)` #ulastirma_sayisal_veri içindekileri korelasyon matrisi üzerinden temel bileşen skorlarına dönüştürerek sonuçları fit ifadesine aktarır.

`screplot(fit, npcs = length(ulastirma_sayisal_veri[,1]), type = "lines")` # ulastirma_sayisal_veri ifadesindeki sütun sayısı kadar oluşacak fit ifadesindeki bileşen skorlarını grafik üzerinde oluşturur.

`fviz_pca_ind(fit)` # fit içeriğinde bulunan ilk iki temel bileşen skorunu kullanarak iki boyutl

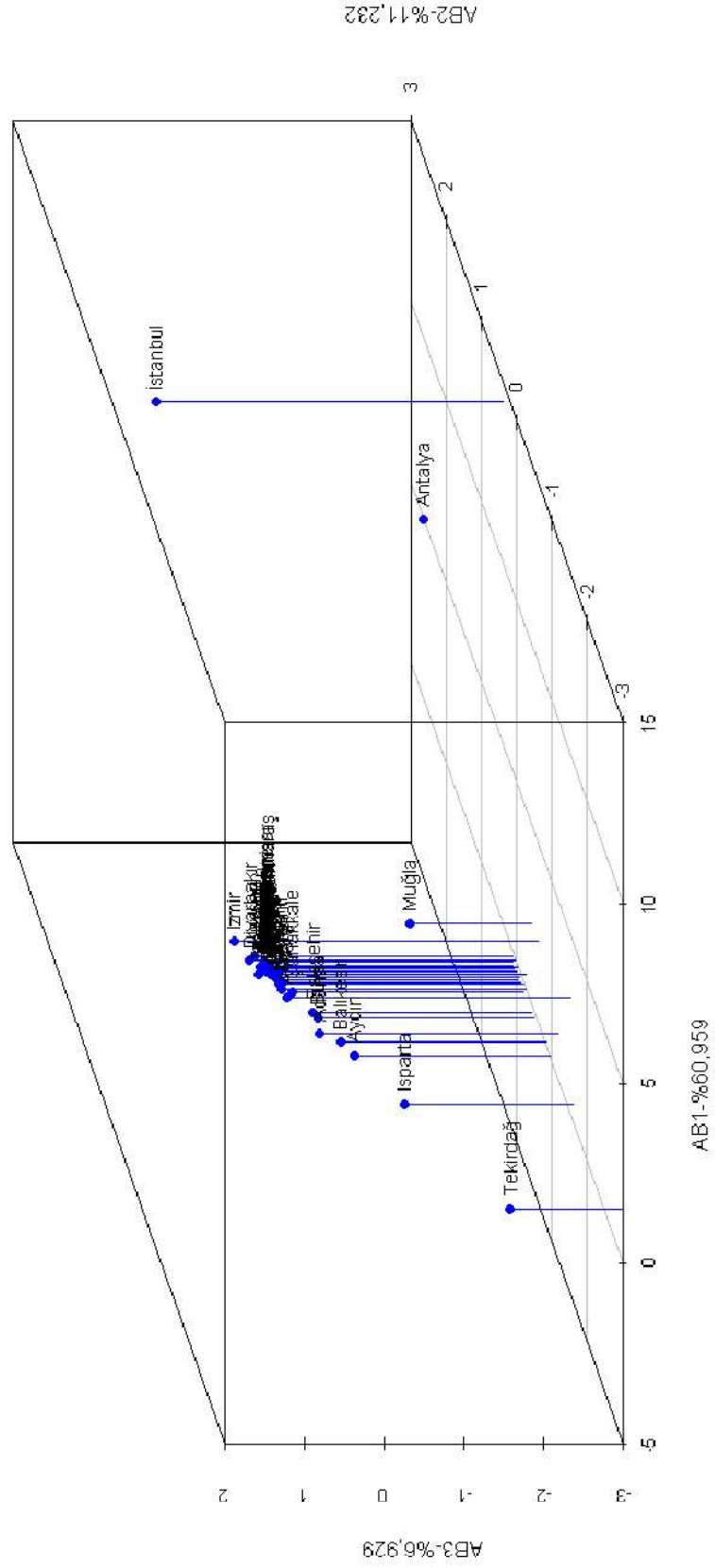
`plot_icin_kmeans <- kmeans(ulastirma_sayisal_veri, 5, nstart=25)` # k ortalamalar yönteminde 81 il gösterilir ve 25 farklı kaynak noktası başlangıcı seçer ve k-ortalamalar yöntemi uygular.

`fviz_cluster(plot_icin_kmeans, data = plot_icin_veri, repel = FALSE, shape = TRUE)`

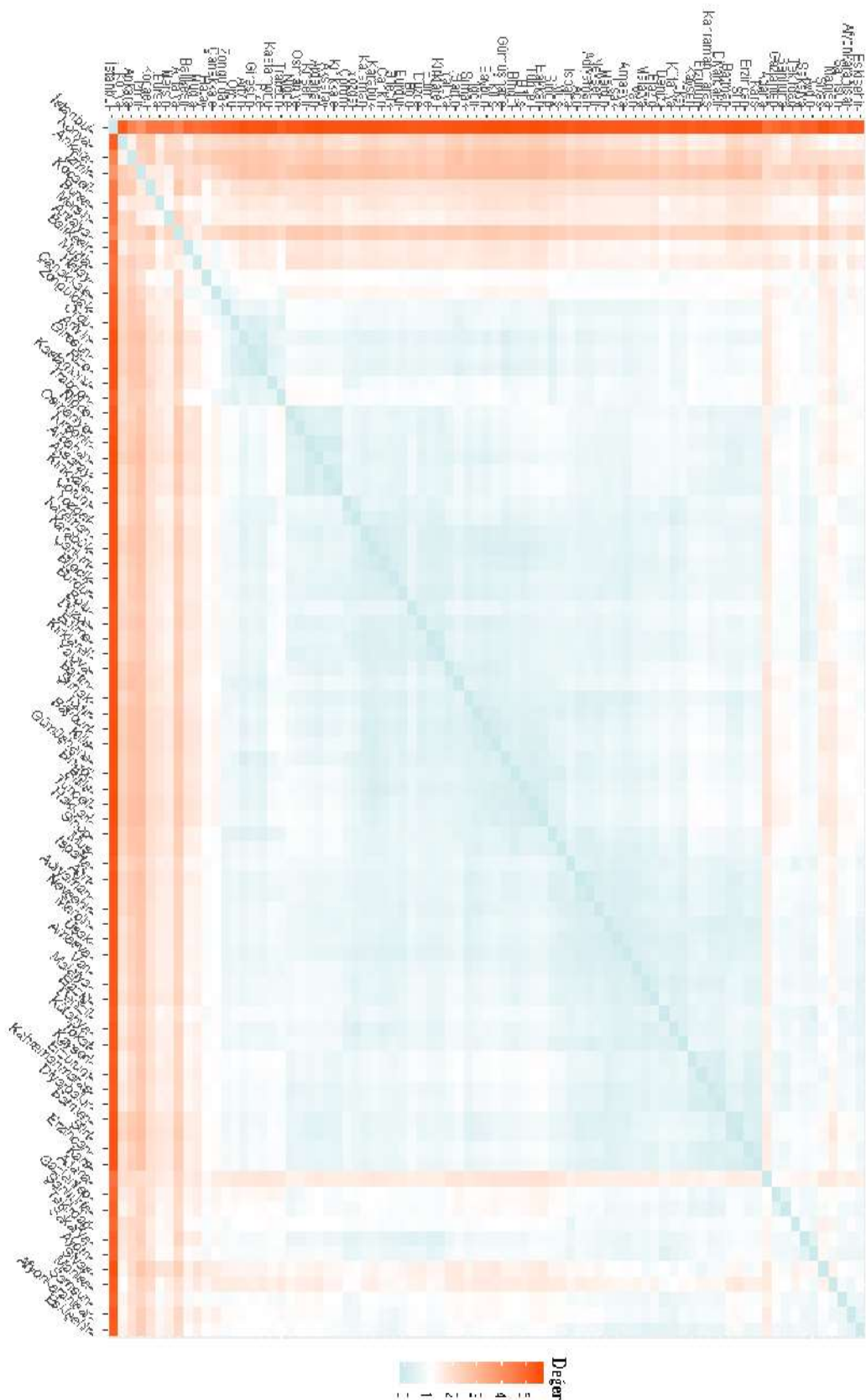
#plot_icin_kmeans grafik gösteriminde k-ortalamalar yöntemi algoritmasından faydalanacağını belirtir(fakat amaç sadece fviz_cluster fonksiyonunun çalışmasını sağlamak için yöntem seçmektir.) plot_icin_veri, veri girişidir, repel, metin etiketlerinin birbirine geçmesini engellemede yardımcı olur. “shape” ifadesi de gözlemlerin standart bir nokta şeklinde gösterimini sağlar ve dört bölgeli koordinat grafiği oluşturur.

EK: 2 (Veri Seti İçin Üç Boyutlu Ana Bileşen Skor Grafiği)

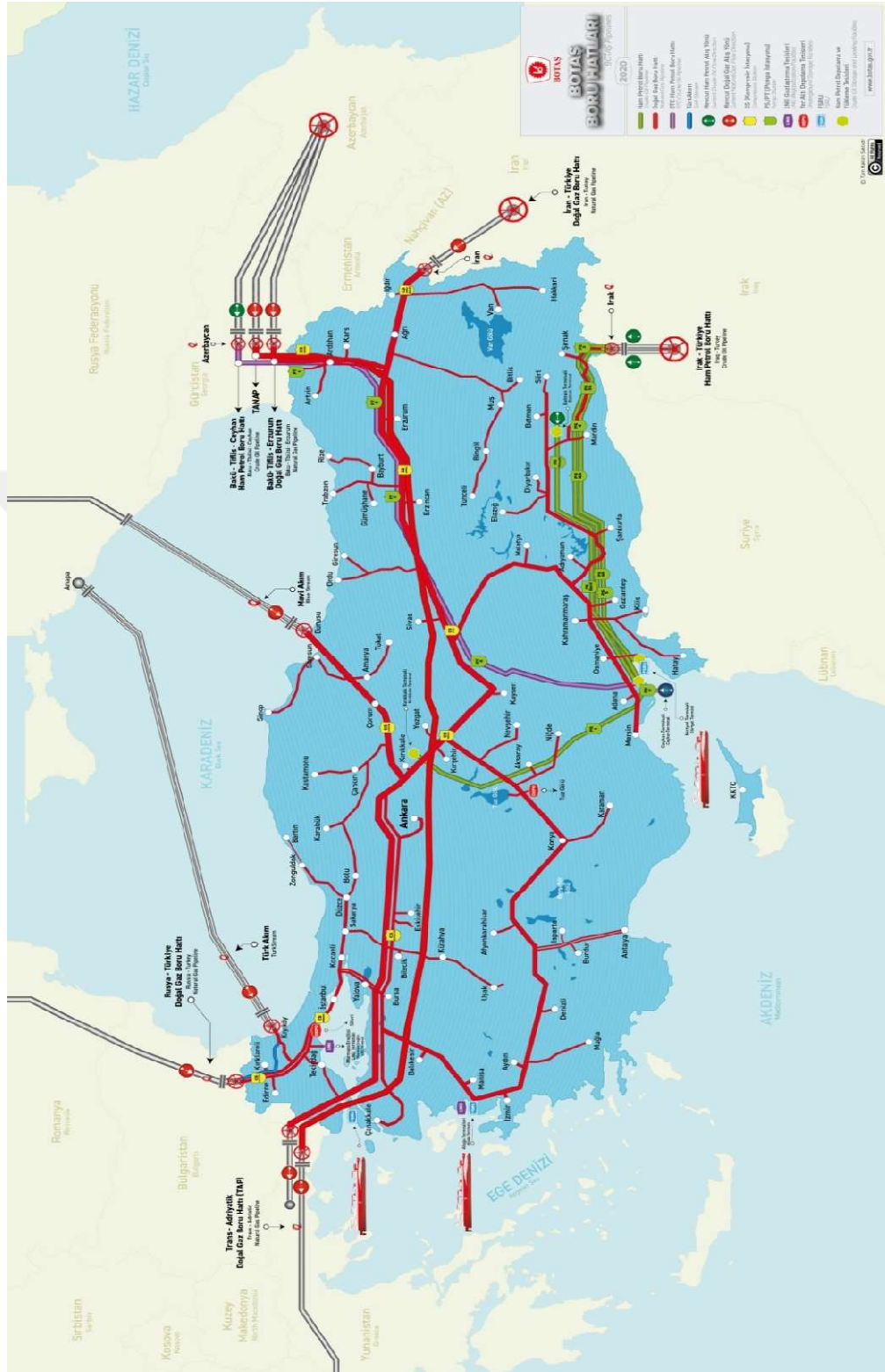
Üç Ana Bileşen Skor Grafiği- %79.12



EK: 3 (İller Arası sıcaklık Benzerlik Matrisi (Öklid Uzaklık))

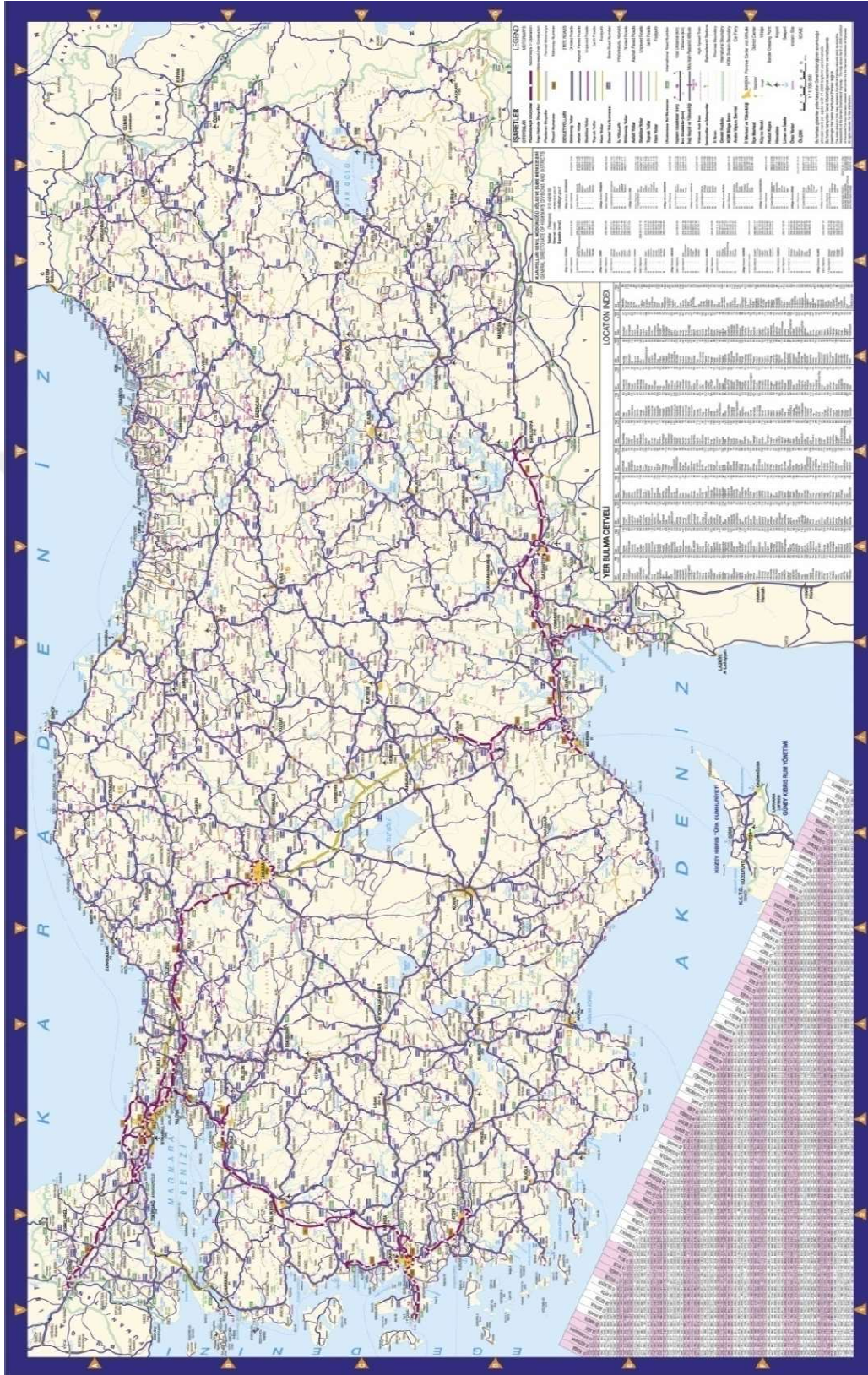


EK: 4 (Türkiye Boru Hatları Haritası)



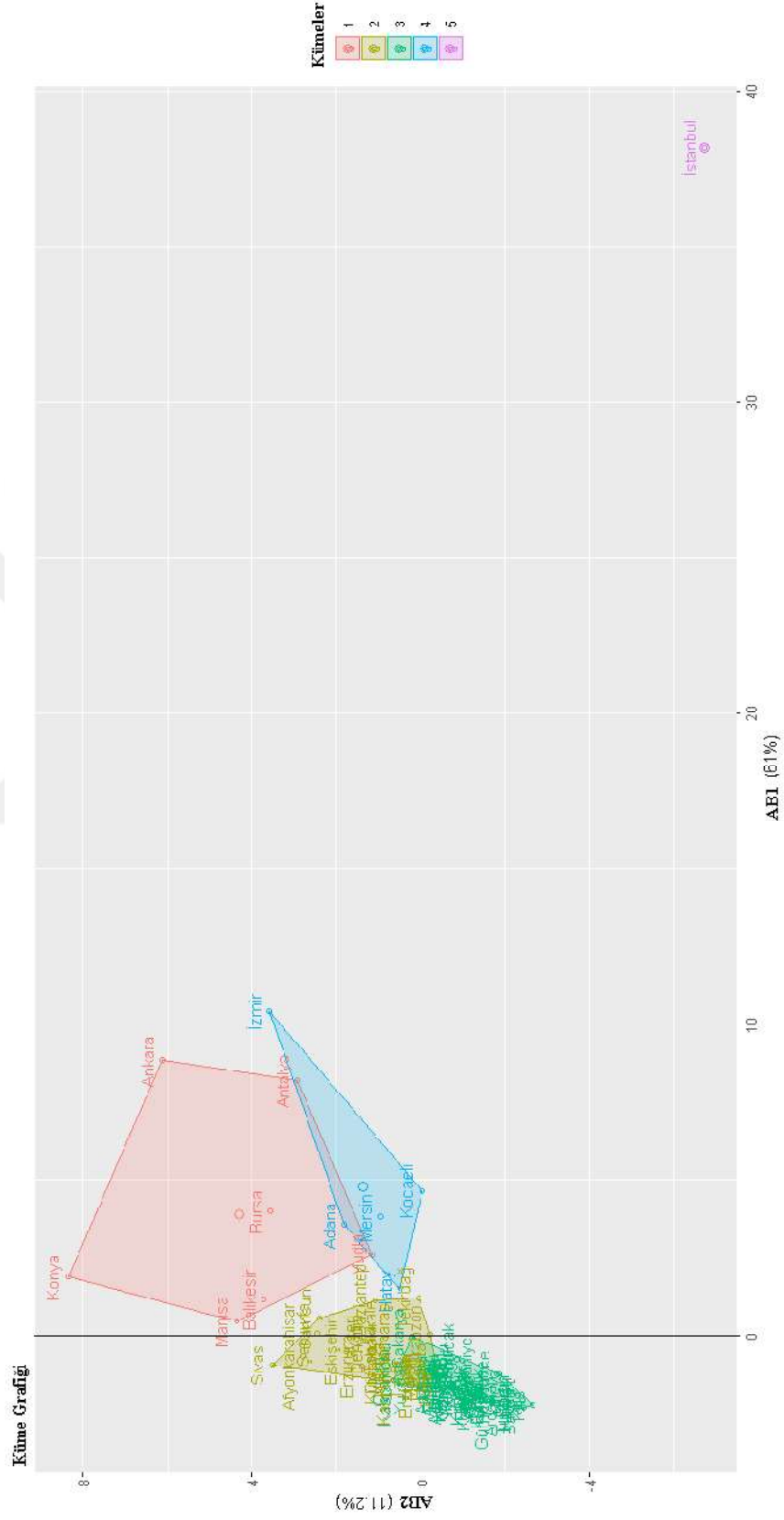
Kaynak: <https://www.botas.gov.tr/Sayfa/dogal-gaz-ve-petrol-boru-hatlari-haritasi/168> (20.11.2020)

EK: 5 (Türkiye Karayolları Haritası)



Kaynak: http://turkiyeharitasi.com/karayollari_haritasi (20.11.2020)

EK: 6 (k-Ortalamalar Yöntemi Küme Grafiği (k=5))



EK: 8 (Kruskal Wallis Test Sonuç Tablosu – 1)

	Gruplar	N	Ortalama Rank		Gruplar	N	Ortalama Rank
V1	1	7	60,86	V21	1	7	52,5
	2	5	54,9		2	5	77,2
	3	25	54,96		3	25	36,48
	4	43	27,1		4	43	36,64
	5	1	81		5	1	80
	Toplam	81			Toplam	81	
V2	1	7	63,64	V22	1	7	50,71
	2	5	55		2	5	76,2
	3	25	52,42		3	25	36,6
	4	43	28,12		4	43	36,95
	5	1	81		5	1	81
	Toplam	81			Toplam	81	
V3	1	7	64,71	V23	1	7	76,29
	2	5	54,4		2	5	67,2
	3	25	58,04		3	25	51,72
	4	43	24,74		4	43	25,09
	5	1	81		5	1	79
	Toplam	81			Toplam	81	
V4	1	7	59,93	V24	1	7	50,71
	2	5	54,7		2	5	76,2
	3	25	54,68		3	25	36,6
	4	43	27,44		4	43	36,95
	5	1	81		5	1	81
	Toplam	81			Toplam	81	
V5	1	7	60,07	V25	1	7	76,43
	2	5	54,7		2	5	65,4
	3	25	54,68		3	25	51,56
	4	43	27,42		4	43	25,4
	5	1	81		5	1	78
	Toplam	81			Toplam	81	
V6	1	7	63,64	V26	1	7	50,71
	2	5	58,9		2	5	75,6
	3	25	54,22		3	25	36,56
	4	43	26,62		4	43	37,05
	5	1	81		5	1	81
	Toplam	81			Toplam	81	
V7	1	7	63,5	V27	1	7	75
	2	5	58,5		2	5	62,2
	3	25	54,08		3	25	51,24
	4	43	26,77		4	43	26,28
	5	1	81		5	1	74
	Toplam	81			Toplam	81	

V8	1	7	59,36		V28	1	7	74
	2	5	54,7			2	5	73,3
	3	25	54,4			3	25	50,28
	4	43	27,7			4	43	25,55
	5	1	81			5	1	81
	Toplam	81				Toplam	81	
V9	1	7	59		V29	1	7	71,71
	2	5	55,3			2	5	51,9
	3	25	54,66			3	25	55,24
	4	43	27,53			4	43	26,99
	5	1	81			5	1	18
	Toplam	81				Toplam	81	
V10	1	7	63,79		V30	1	7	70,71
	2	5	58,7			2	5	70,4
	3	25	54,34			3	25	51,56
	4	43	26,55			4	43	25,67
	5	1	81			5	1	81
	Toplam	81				Toplam	81	
V11	1	7	63,29		V31	1	7	51,21
	2	5	57,8			2	5	76
	3	25	54,6			3	25	37,24
	4	43	26,58			4	43	36,52
	5	1	81			5	1	81
	Toplam	81				Toplam	81	
V12	1	7	61,64		V32	1	7	73,86
	2	5	46,9			2	5	73,2
	3	25	56,6			3	25	49,2
	4	43	26,95			4	43	26,21
	5	1	81			5	1	81
	Toplam	81				Toplam	81	
V13	1	7	55,21		V33	1	7	54,14
	2	5	68,9			2	5	32,2
	3	25	35,86			3	25	49,08
	4	43	37,53			4	43	36,12
	5	1	79,5			5	1	1
	Toplam	81				Toplam	81	
V14	1	7	55		V34	1	7	74,43
	2	5	75,8			2	5	72,8
	3	25	38,36			3	25	51,44
	4	43	35,28			4	43	24,86
	5	1	81			5	1	81
	Toplam	81				Toplam	81	
V15	1	7	54,57		V35	1	7	71,43
	2	5	68,8			2	5	75
	3	25	37,44			3	25	49,4
	4	43	36,7			4	43	26,28
	5	1	81			5	1	81

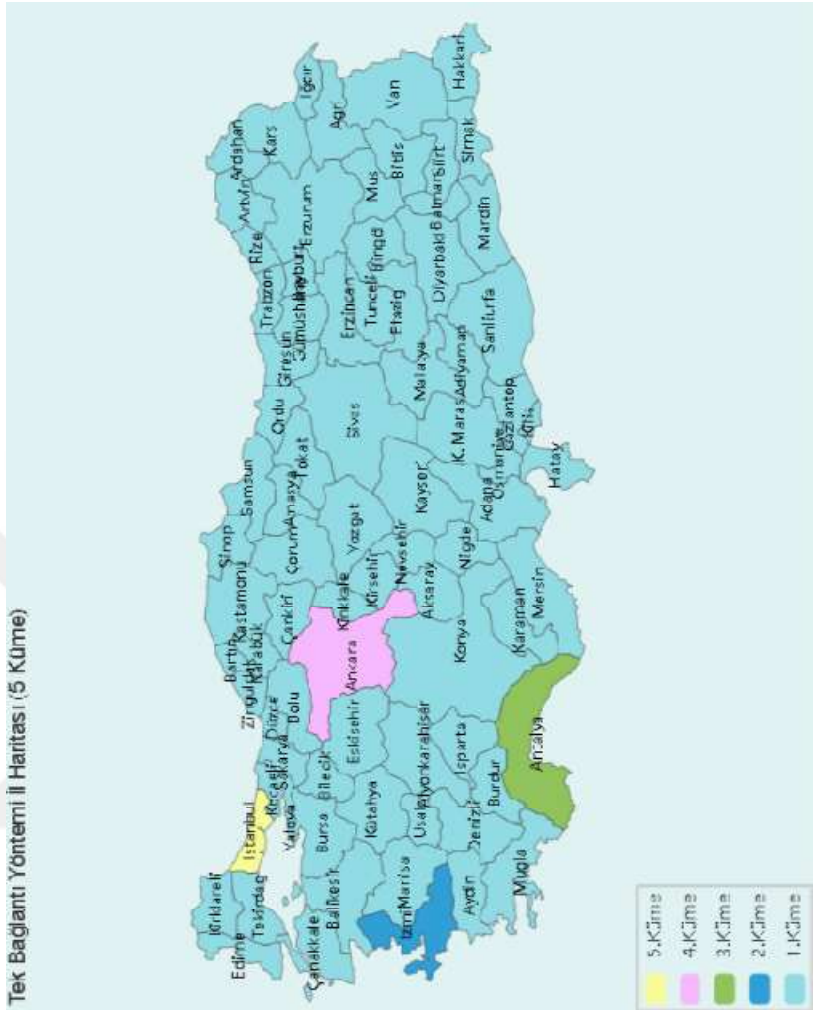
	Toplam	81			Toplam	81	
V16	1	7	55,64	V36	1	7	72,71
	2	5	67,5		2	5	71,4
	3	25	39,06		3	25	46,92
	4	43	35,73		4	43	27,95
	5	1	81		5	1	80
	Toplam	81			Toplam	81	
V17	1	7	56	V37	1	7	74,57
	2	5	68,4		2	5	73
	3	25	37,44		3	25	52,28
	4	43	36,51		4	43	24,33
	5	1	81		5	1	81
	Toplam	81			Toplam	81	
V18	1	7	53,36	V38	1	7	74,86
	2	5	78,2		2	5	61,2
	3	25	35,28		3	25	49,72
	4	43	37,08		4	43	27,77
	5	1	80		5	1	54
	Toplam	81			Toplam	81	
V19	1	7	50,79	V39	1	7	51
	2	5	78,4		2	5	55,4
	3	25	36,08		3	25	57,72
	4	43	37,06		4	43	27,4
	5	1	78		5	1	66
	Toplam	81			Toplam	81	
V20	1	7	51,5	V40	1	7	48,36
	2	5	76,6		2	5	46,2
	3	25	35,68		3	25	47,36
	4	43	37,31		4	43	35,35
	5	1	81		5	1	47,5
	Toplam	81			Toplam	81	

EK: 9 (İl Ulaştırması Faaliyet Durum Tablosu)

BÖLGE ADI	Hava	Kara	Demir	Deniz	Boru Hattı Sistemi (Petrol, Doğalgaz, Maden)	Toplam	Ulaştırma	İl Sayısı
Adana	1	1	1	1	1	5	5	11
Adıyaman	1	1	1	0	1	4	4	38
Afyonkarahisar	0	1	1	0	1	3	3	25
Ağrı	1	1	0	0	1	3	2	7
Aksaray	0	1	0	0	1	2	1	0
Amasya	1	1	1	0	1	4	0	0
Ankara	1	1	1	0	1	4	Toplam: 81	
Antalya	1	1	0	1	1	4		
Ardahan	0	1	1	0	1	3		
Artvin	0	1	0	1	1	3		
Aydın	1	1	1	1	1	5		
Balıkesir	1	1	1	1	1	5		
Bartın	0	1	0	1	1	3		
Batman	1	1	1	0	1	4		
Bayburt	0	1	0	0	1	2		
Bilecik	0	1	1	0	1	3		
Bingöl	1	1	1	0	1	4		
Bitlis	0	1	1	0	1	3		
Bolu	0	1	0	0	1	2		
Burdur	0	1	1	0	1	3		
Bursa	1	1	1	1	1	5		
Çanakkale	1	1	0	1	1	4		
Çankırı	0	1	1	0	1	3		
Çorum	0	1	0	0	1	2		
Denizli	1	1	1	0	1	4		
Diyarbakır	1	1	1	0	1	4		
Düzce	0	1	0	0	1	2		
Edirne	0	1	1	1	1	4		
Elazığ	1	1	1	0	1	4		
Erzincan	1	1	1	0	1	4		
Erzurum	1	1	1	0	1	4		
Eskişehir	1	1	1	0	1	4		
Gaziantep	1	1	1	0	1	4		
Giresun	0	1	0	1	1	3		
Gümüşhane	0	1	0	0	1	2		
Hakkari	1	1	0	0	0	2		
Hatay	1	1	1	1	1	5		
Iğdır	1	1	0	0	1	3		

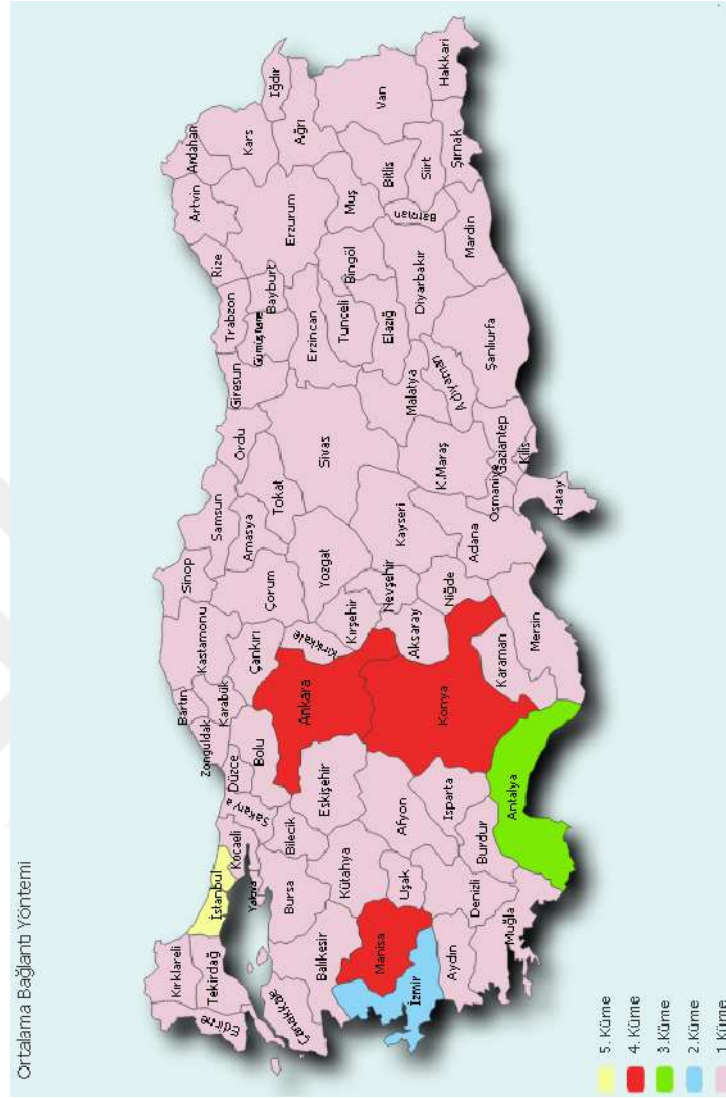
Isparta	1	1	1	0	1	4
İstanbul	1	1	1	1	1	5
İzmir	1	1	1	1	1	5
Kahramanmaraş	1	1	1	0	1	4
Karabük	0	1	1	0	1	3
Karaman	0	1	1	0	1	3
Kars	1	1	1	0	1	4
Kastamonu	1	1	0	1	1	4
Kayseri	1	1	1	0	1	4
Kırıkkale	0	1	1	0	1	3
Kırklareli	0	1	1	0	1	3
Kırşehir	0	1	1	0	1	3
Kilis	0	1	1	0	1	3
Kocaeli	1	1	1	1	1	5
Konya	1	1	1	0	1	4
Kütahya	1	1	1	0	1	4
Malatya	1	1	1	0	1	4
Manisa	0	1	1	0	1	3
Mardin	1	1	1	0	1	4
Mersin	0	1	1	1	1	4
Muğla	1	1	0	1	1	4
Muş	1	1	1	0	1	4
Nevşehir	1	1	1	0	1	4
Niğde	0	1	1	0	1	3
Ordu	1	1	0	1	1	4
Osmaniye	0	1	1	0	1	3
Rize	0	1	0	1	1	3
Sakarya	0	1	1	1	1	4
Samsun	1	1	1	1	1	5
Siirt	1	1	1	0	1	4
Sinop	1	1	0	1	1	4
Sivas	1	1	1	0	1	4
Şanlıurfa	1	1	1	0	1	4
Şırnak	1	1	0	0	1	3
Tekirdağ	1	1	1	1	1	5
Tokat	1	1	1	0	1	4
Trabzon	1	1	0	1	1	4
Tunceli	0	1	1	0	1	3
Uşak	1	1	1	0	1	4
Van	1	1	1	0	1	4
Yalova	0	1	0	1	1	3
Yozgat	0	1	1	0	1	3
Zonguldak	1	1	1	1	1	5

EK: 10 (Tek Bağlantı Yöntemine Göre Türkiye İl Haritası)





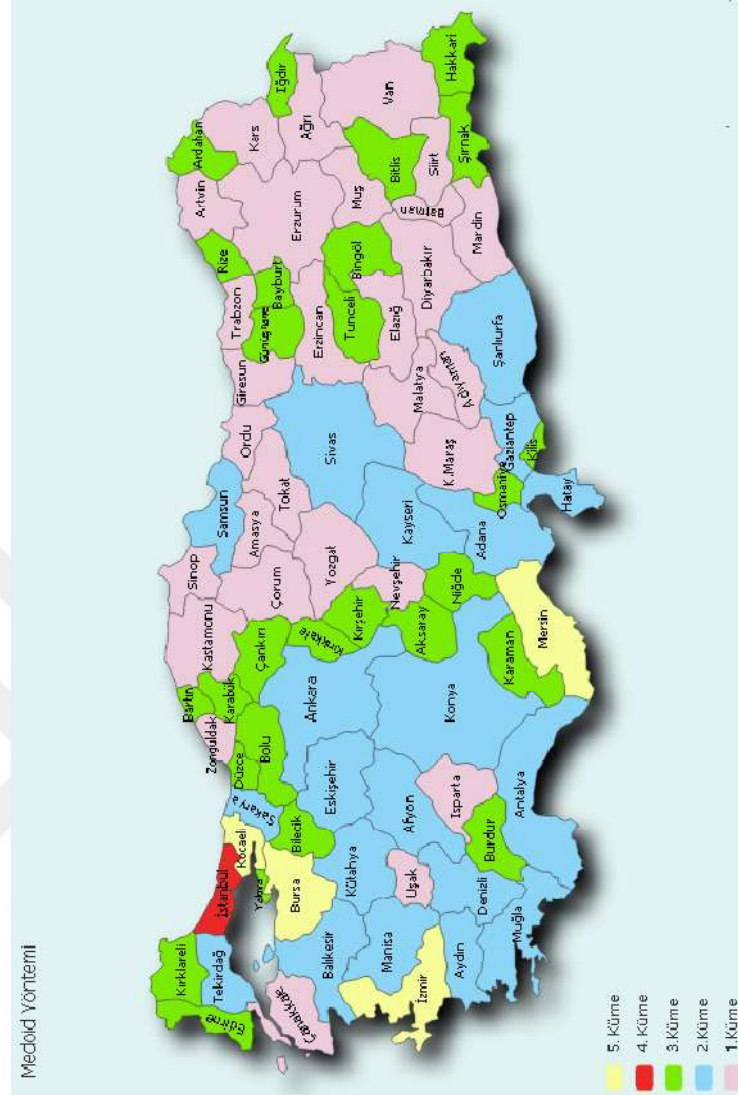
EK: 12 (Ortalama Bağlantı Yöntemine Göre 5 Küme İçin Türkiye İl Haritası)





[illegible]

EK: 15 (Medoid Yöntemine göre İl Haritası)



EK: 16 (Türkiye Limanları Haritası - 2021)



Kaynak: cografiyaharita.com

Kaynak: cografyaharita.com

