



Türkiye'ye Yönelik Dış Turizm Talebi Açısından Ülkelerin Kümeleme Analizi ile Sınıflandırılması

Ece Aydoğdu Ulukan^{1*}, Serhat Peker²

^{1*} İzmir Bakırçay Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, İzmir, Turkey, (ORCID: 0000-0003-2867-5221), ece.aydogdu1@gmail.com

² İzmir Bakırçay Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, İzmir, Turkey, (ORCID: 0000-0002-6876-3982), serhat.peker@bakircay.edu.tr

(İlk Geliş Tarihi 3 Ekim 2022 ve Kabul Tarihi 11 Kasım 2022)

(DOI: 10.31590/ejosat.1183700)

ATIF/REFERENCE: Aydoğdu Ulukan, E., & Peker, S. (2022). Türkiye'ye Yönelik Dış Turizm Talebi Açısından Ülkelerin Kümeleme Analizi ile Sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (41), 384-392.

Öz

Turizm ülkelerin ekonomik gelişimi için en önemli unsurlardan biridir. Ülkelere gelen yabancı turistlerin verilerinin analiz edilmesi bu gelişime katkı sağlaması açısından büyük önem taşısa da uluslararası turizme yönelik Türkiye'de yeteri kadar çalışma bulunmamaktadır. Bu çalışmanın amacı, farklı ülkelerden Türkiye'ye olan dış turizm talebini kümeleme analizi kullanarak incelemek ve Türkiye'ye turist gönderen bu ülkeleri sınıflandırmaktır. Bu bağlamda, ülkelerin gelir düzeyleri, ülkelerden çıkan turist sayıları, çıkan turist sayılarında Türkiye'nin payı ve turistlerin Türkiye'de konaklama süresi gibi faktörler dikkate alınmış ve iki aşamalı kümeleme yöntemi kullanılarak ülkeler gruplandırılmıştır. Elde edilen ülke grupları, kullanılan değişkenler ışığında karakterize edilmiştir. Bu çalışma sonucunda oluşturulan ülke profillerinin, politika yapıcılarının etkin stratejiler geliştirmesinde yardımcı olacağına inanılmaktadır.

Anahtar Kelimeler: Veri madenciliği, kümeleme analizi, veri yönetimi, iş zekâsı ve analitiği, uluslararası turizm talebi.

Classifying Countries in Terms of Inbound Tourism Demand Using Cluster Analysis: The Case of Turkey

Abstract

Tourism is one of the most important elements for the economic development of countries. Although the analysis of the data of foreign tourists coming to the countries is one of the great importance in terms of contributing to this development, there are not enough studies in Turkey for international tourism. The aim of this study is to examine the foreign tourism demand coming to Turkey from different countries by using cluster analysis and to classify these countries that send tourists to Turkey. In this context, factors such as the income levels of the countries, the number of tourists leaving the countries, Turkey's share in the number of tourists and the length of stay of the tourists in Turkey are considered, and the countries are clustered using the two-stage clustering method. The resulting country groups were characterized in the light of the features used. It is believed that the generated country profiles will help policy makers to develop effective strategies.

Keywords: Data mining, cluster analysis, data management, business intelligence and analytics, international tourism demand.

* Sorumlu Yazar: ece.aydogdu1@gmail.com

1. Giriş

Turizm, ülkelerin gelişimi için önemli bir etkiye sahip olmasıyla birlikte uluslararası ticareti de düzenleyen ve ülkelere ekonomik gelir artışı sağlayan sektörlerden biridir (Kandemir, 2018). Turizm sektörü, ülkelere olan gelir katkısı, sağladığı döviz girdisi ve iş imkanlarıyla ülke ekonomisini büyük ölçüde etkilemektedir (Özbek, 1991). Uluslararası turizmden, seyahat ve yolcu taşımacılığı alanında elde edilen ihracat kazançları, dünya çapında hem gelişmekte olan hem de gelişmiş birçok ekonomi için önemli bir dış gelir kaynağı oluşturmaktadır. Bu kazançlar ülke içerisinde iş imkânı yaratmaya, girişimciliği teşvik etmeye ve yerel ekonomileri de geliştirmeye yardımcı olur. Uluslararası turizm aynı zamanda, dış ticaret açığını azaltıp, diğer mal ve hizmetlerden elde edilen daha zayıf ihracat gelirlerini telafi ederek ülkelerin ödeme dengesini iyileştirmeye katkı sağlar. Bu sebeplerle turizm, toplumlar arası kültürel etkileşimin artması, ekonomik dengenin sağlanması ve iş olanakları yaratması sebebiyle ülkeler için oldukça önemli bir kavramdır (Garda & Temizel, 2016). Aynı zamanda bu katkıları nedeniyle bir hizmet sektörü olan turizmin, dünya çapında hızlı bir gelişim gösterdiği de fark edilmiştir (Bahar, 2006). Turizm sektörünün gelişimi, istihdam olanakları sağlaması açısından, yüksek işsizlik sorunları yaşayan ülkeler için önemlidir (Aydın et al., 2015). Türkiye’de dış turizm ülkeye gelir sağlayan, döviz girdisine en büyük katkıda bulunan alanlardandır (Karaoğlu, 2019). Buna paralel olarak, ulusal ve uluslararası yapılmış olan çalışmalardan turizmin ekonomik büyümeye etkisinin olumlu olduğu görülmektedir (Bahar, 2006).

Ulaşım imkanlarının gelişmesi ve teknolojik ilerlemelerle birlikte Türkiye için de turizm sektörü önemli bir noktaya ulaşmış, dış turist talebi ekonomiyi etkileyen faktörlerin başlarında yer almıştır (Zortuk & Bayrak, 2013). Gelişmekte olan ülkeler başta olmak üzere birçok ülke için dış turizm talebinin ekonomiyeye katkısının önemi giderek artmış ve bu ülkelerde turizm hızlı bir yükselişe geçmiştir (Özcan, 2021). UNWTO 2020 (“Int. Tour. Highlights, 2020 Ed.,” 2021) raporuna göre, yakıt ve kimyasaldan sonra turizm, dünyanın üçüncü en büyük ihracat kategorisi olmakla birlikte uluslararası turizmden yapılan toplam ihracat 2019’da 1,7 trilyon ABD dolarına ulaşmıştır. Türkiye’de ise 2019 yılında 26 milyar ABD doları ile dünyanın en büyük 5. seyahat fazlası (uluslararası turizm gelirleri (gelen) ile uluslararası turizm harcamaları (giden) arasındaki fark) kaydeden ülkesi olmuştur (“Int. Tour. Highlights, 2020 Ed.,” 2021). Yine 2019 yılında Türkiye, 50 milyondan fazla varış ile dünyanın ilk 10 destinasyonu arasında yer almakla birlikte, bu istatistiklerde dünyada en güçlü artışı (+%12) gösteren ülke olmuştur (“Int. Tour. Highlights, 2020 Ed.,” 2021). Bu gelişmeler sayesinde turizm, bilimsel çalışmalarında ilgisini üzerinde toplamış ve literatürde daha fazla yer almaya başlamıştır (Ersungur et al., 2017).

Talep tahmini birçok sektörde olduğu gibi turizm sektöründe de önemlidir (Sevimli Deniz, 2019). Günümüzde turizmin hızlı gelişimiyle birlikte verinin depolanması ve ihtiyaç anında ulaşılması gibi durumlarda enformasyon sistemlerine olan ihtiyacı arttırmıştır (Akgün & Çizel, 2016). Bu sistemler sayesinde ekonomide stratejik planlar hazırlanması mümkün kılınabilirken aynı zamanda yatırımcılardan elde edilen kaynaklardaki israfları da en aza indirmek mümkün olabilmektedir (Çuhadar, 2015). Herhangi bir bölgedeki turist hareketleri incelenerek reklam kampanyaları hazırlanabilir,

sonraki sezonda turist sayısını yükseltmek için planlamalar yapılabilmektedir (Akgün & Çizel, 2016). Bununla birlikte turistlerin nereden geldikleri, geldikleri bölgelerdeki harcama bilgileri gibi verilerle pazarlama stratejileri de belirlenebilmektedir (Magnini et al., 2003).

Dijitalleşmeyle birlikte turizm endüstrisinde de çevrimiçi ve çevrimdışı kanallar aracılığıyla veri toplamak daha kolay ve mümkün hale gelmiştir. Bu da veriye dayalı iş analitiği modellerinin kullanımına olanak sağlamaktadır. Bu bağlamda, veri hacmi arttıkça ve veri özellikleri daha karmaşık hale geldikçe, veri madenciliği ve yapay zekâ uygulamalarının önemi de artmıştır. Veri madenciliği, büyük veri setlerinde örtük olan ilginç, değerli bilgi ve kalıplarından anlamlı sonuçlar çıkarabilmek için bir takım matematiksel ve istatistikî yöntemlerin kullanılmasını kapsamaktadır (Akgün & Çizel, 2016), (Bardak & Sözen, 2018). Veri bilimiyle elde edilen bu anlamlı sonuçlar ile firmalar farklı pazarlarda büyük başarılar elde edebilirler (Bardak & Sözen, 2018). Veri madenciliği ve yapay zeka uygulamaları, turizm alanında yeni olmalarına rağmen önemli bir etken haline geleceği tahmin edilmektedir (Doborjeh et al., 2022).

Bu çalışmanın amacı, Türkiye’ye yönelik dış turizm talebini, en yaygın veri madenciliği tekniklerinden biri olan kümeleme analizi kullanarak incelemek ve Türkiye’ye turist gönderen ülkeleri, bu açıdan karakterize edip sınıflandırmaktır. Bu anlamda, ülkelerin gelir düzeyleri, ülkelerden çıkan turist sayıları, çıkan turist sayılarında Türkiye’nin payı ve turistlerin Türkiye’de konaklama süresi gibi faktörler dikkate alınmıştır. Bu değişkenlere ilişkin, 2019 ve 2020 yıllarına ait veriler toplanmıştır. Toplanan veriler ışığında, ülkeler iki aşamalı kümeleme yöntemi ile her iki yıl için ayrı ayrı sınıflandırılmış ve oluşan kümeler, kullanılan değişkenler özelinde incelenmiştir. Çalışmamızın geri kalanı, dört bölüm halinde düzenlenmiştir. İkinci bölümde konu ile ilgili literatür taraması yapılmış, üçüncü bölümde ise kullanılan materyaller ve yöntem açıklanmıştır. Uygulama ve çalışmanın bulguları hakkında bilgilere dördüncü bölümde yer verilirken, çalışmanın sonuçlarına gelecekteki çalışmalara önerilerle birlikte beşinci bölümde yer verilmiştir.

2. Turizmde Veri Madenciliği ve Yapay Zeka Uygulamaları

Birçok sektör gibi turizm sektörü de talep tahmini çalışmalarını güvenilir veriler ışığında yapmaya ihtiyaç duymaktadır (Çuhadar, 2015). Turizm politikaları belirleme ve turizm planlarının yapılmasında etkisi büyük olan talep tahmini belirleme konusu çok sayıda araştırmanın ilgi odağı olmuştur (Kaya & Canlı, 2013). Huang ve Hao (2021), çalışmalarında Hong Kong turist gelişlerini tahmin etmede talep tahmini doğruluğunu geliştirmek için iki aşamalı yöntem önermişlerdir. Yöntemin ilk aşamasında arama sorgularındaki anahtar kelimelerle çift artırmalı bir algoritmadan faydalanılmış, ikinci aşamada ise olası doğrusal olmayan ilişkiyi tespit etmek ve derin öğrenme kombinasyonu yoluyla tahmin performansını iyileştirmek için Destek Vektör Regresyon (SVR) tabanlı Derin İnanç Ağı (DBN) yaklaşımı kullanılmıştır. Zhang ve diğerleri (2021), çalışmada modelleme doğruluğunu daha da geliştirmeyi ve yapay zeka tabanlı turizm talep tahmin yöntemlerini geliştirmeyi amaçlamıştır.

Akın (2015), çalışmada yapısal zaman serisi modellemesini kullanarak verilen zaman serisinin bileşenlerini

belirlemiş, belirlenen bu bileşenleri kullanarak bir dizi karar ağacı oluşturmuş ve model seçimi için bir kural seti elde etmiştir. Bir başka çalışmada ise Höpken ve diğerleri (2021), turist varış tahmini için gezginlerin web arama trafiğini harici girdi özelliği olarak dahil ederek otoregresif tahmin modellerini geliştirmek için bir yaklaşım sunmuştur. Ayrıca bu çalışmada, ilgili arama terimlerini belirlemek ve bunları bir otoregresif tahmin yaklaşımının ek girdisi olarak kullanılan bir bileşik web arama indeksinde toplamak için yeni bir yöntem önerilmiştir.

Dursun ve Caber (2016), kümeleme algoritmaları ve RFM modeli ile karlı otel müşterilerinin profilini çıkarmaya

odaklanmıştır. RFM modelinde, müşterilerin doğasını keşfetmek için yenilik, frekans ve parasal göstergeler kullanılmıştır. Turist profillerinin incelendiği farklı bir çalışmada Yıldırım ve İnce (2022), Türkiye'deki turist profilini veri madenciliği yöntemleriyle analiz etmeyi amaçlamıştır. Bu veriler üzerinde k-ortalama kümeleme algoritması kullanılarak profillerin anahtar parametreleri elde edilmiş ve kişiler özelliklerine göre gruplara ayrılmıştır. Yukarıda bahsi geçen be son yıllarda yapılan diğer benzer çalışmalar da Tablo 1'de özetlenmiştir.

Tablo 1. Literatür özeti
(Table 1. Literature review)

Makale	Yıl	Amaç	Kullanılan Yöntemler
(Akın, 2015)	2015	Farklı ülkelerden gelen turist verileri üzerindeki performansları karşılaştırarak model seçimi için bir kural oluşturmak	Yapısal zaman serisi modelleme, Ortalama Kare Hatası (MSE), Ortalama Mutlak Yüzde Hatası (MAPE), SARIMA, Sinir Ağı Modelleme, Destek Vektör Makineleri (SVM)
(Dursun & Caber, 2016)	2016	Müşteri davranışlarını inceleyerek pazarlama stratejilerini geliştirmek için müşteri segmentasyonu yapmak	K-ortalama algoritması ve kendini düzenleyen haritalar (SOM)
(Shapoval et al., 2018)	2018	Turistlerin davranışlarını analiz ederek pazarlama stratejilerine destek olmak	Karar Ağaçları
(Law et al., 2019)	2019	Aylık Makao turist varış hacimlerini tahmin etmek	Derin Öğrenme, Destek Vektör Regresyonu, Yapay Sinir Ağı Modelleri
(Kulshrestha et al., 2020)	2020	Turizm talep tahmini yapmak	Uzun-Kısa Süreli Bellek (BILSTM)
(Akgün et al., 2021)	2021	İlgili seyahat acentesi müşterilerinin profili hakkında bilgi edinmek	Karar ağaçları sınıflandırma algoritması
(Nguyen et al., 2021)	2021	Aylık uluslararası turist sayısını tahmin etmek	Yapay Sinir Ağı
(Huang & Hao, 2021)	2021	İlgili makine öğrenmesi tekniklerini kullanarak turizm talep tahmin başarısını arttırmak	Derin İnanç Ağı (DBN), Destek Vektör Regresyonu (SVR)
(Zhang et al., 2021)	2021	Yapay zekâ tabanlı turizm talep tahmin yöntemleri geliştirmek	Derin Öğrenme, STL, DADLM
(Höpken et al., 2021)	2021	Oto agresif tahmin modellerini geliştirmek	Oto agresif Hareketli Ortalamalar Modeli (ARIMA), Yapay Sinir Ağı
(Yıldırım, 2022)	2022	Türkiye'deki turist profilini veri madenciliği yöntemleriyle analiz etmek	K-ortalama algoritması

Son yıllarda veri madenciliği ve yapay zeka uygulamaları turizm sektöründe sık kullanılsa da, bu araçların ve teknolojilerin kullanıldığı farklı turizm dinamiklerine yönelik çalışmalara da ihtiyaç duyulmaktadır (Ünlüöner & Şen Küpeli, 2021). Ayrıca literatürdeki çalışmalarda Türkiye açısından dış turizm sektörü yeterince ilgi görememiş ve bu alan üzerine çok sayıda çalışma yapılmamıştır. Çalışmamız dış turizm alanında verimi arttırmak ve bu alandaki çalışmalara destek olmayı hedeflemektedir.

3. Materyal ve Yöntem

Bu bölümde, veri toplama ve veri analiz teknikleri gibi çeşitli konular tartışılmaktadır. İlk alt bölümde, araştırmanın verilerinin hangi kaynaklardan elde edildiği açıklanmış, ilgili internet siteleri paylaşılmıştır. Daha sonra değişken tanımları tablo halinde verilmiştir. İkinci alt bölümde ise veri setine uygulanan ön işlemler anlatılmış ve oluşturulan modellerin tanımlayıcı istatistik verilerini içeren bilgiler tablo halinde sunulmuştur. Son olarak üçüncü alt bölümde ise kullanılan analiz yöntemi açıklanmıştır.

3.1. Niteliklerin Tasviri ve Verinin Toplanması

Bu çalışmada dış turizm talebi açısından ülkeler sınıflandırılırken dört temel nitelik dikkate alınmıştır. Tablo 2’de çalışan bütün nitelikler ve tanımlar sunulmuştur.

Tablo 2. Değişken tanımları
(Table 2. Variable Definitions)

Değişken Adı	Tanım
Kişi başına düşen GSYİH	Yıl bazında ülkelerin ekonomik gelir düzeylerini temsil eder (dolar)
Çıkan Turist Sayısı	Yıl içinde toplam ülkeden çıkan kişi sayısı
Türkiye’nin Payı	Ülkeden çıkan turistlerin yüzde kaçının Türkiye’yi tercih ettiği bilgisi
Türkiye’de Konaklama süresi	Türkiye’de ortalama konaklama süresi (gün)

Bu özelliklerden ilki olan ülkelerin yıllık ortalama dolar bazında gelirleri, yabancı ülkeler hakkında güvenilir ekonomik verilere ihtiyaç duyan araştırmacılara, iş adamlarına, akademisyenlere ve yatırımcılara hizmet veren theglobeconomy.com (GDP per Capita, Current Dollars by Country, around the World | TheGlobalEconomy.Com, n.d.) internet sitesinden çekilmiştir. Ülkelerin ekonomik gelir düzeylerine ek olarak, turist çıkış sayılarına Dünya Bankası’ndan (International Tourism, Number of Departures | Data, n.d.) ve bu turistlerin Türkiye’de konaklama sürelerine T.C. Kültür ve Turizm Bakanlığı (İşletme (Bakanlık) Belgeli Tesis Konaklama İstatistikleri, n.d.) veri tabanlarından ulaşılmıştır. Ülkelerden çıkan ve Türkiye’ye gelen turist sayıları ile Türkiye’nin bu pastadaki payı elde edilmiştir. Son olarak ülkelerin kıta verilerine ise statisticstimes.com (List of Countries by Continent - StatisticsTimes.Com, n.d.) internet sitesinden ulaşılmıştır.

3.2. Veri Setinin Hazırlanması ve Tanımlayıcı İstatistikler

Veri setinin hazırlanması, veri toplama işlemlerinin tamamlanması sonrasında yapılan ilk aşamadır. Bu çalışmada veri hazırlama aşamasında veri setinde boş alanlar temizlenmiş, tutarsız veriler veri setinden çıkarılmıştır. T.C. Kültür ve Turizm Bakanlığı (İşletme (Bakanlık) Belgeli Tesis Konaklama İstatistikleri, n.d.) veri tabanından ulaşılan konaklama bilgileri veri setinde toplam 90 adet ülkeye ait veri bulunmaktadır. Bu 90 ülke baz alındığında theglobeconomy.com (GDP per Capita, Current Dollars by Country, around the World | TheGlobalEconomy.Com, n.d.) internet sitesinden alınan kişi başına düşen GSYİH verilerinde ise 6 ülkenin veri seti bulunmamaktadır. Bu ülkeler; Venezuela, Kuzey Kıbrıs Türk Cumhuriyeti, Yemen, Güney Kıbrıs Rum Kesimi, Kosova ve Türkmenistan olup çalışma veri setine dahil edilmemiştir. Ek olarak, Moldova ülkesinin verilerinde de tutarsızlık olduğundan dolayı çalışma veri setinden çıkartılmıştır. Türkiye’ye gelen turist sayıları, bu çalışma kapsamında analiz edildiğinden, Türkiye ülke olarak veri setine dahil edilmemiş ve veri setinde toplamda 82 ülke kalmıştır. Bu 82 ülke içerisinde, Dünya Bankası (International Tourism, Number of Departures | Data, n.d.) veri tabanından elde edilen, ülkelerden çıkan turist sayılarını içeren verilerde 2019 yılında 17 adet ve 2020 yılında 32 adet ülkenin verisi eksik olduğundan, bu ülkeler de veri setinden çıkarılmıştır. Sonuç olarak 2019 ve 2020 yılları için sırasıyla 65 ve 50 adet ülke analizlere dahil edilmiş, elde edilen tüm bu veriler bir Excel dosyasına alınmıştır. Ülkelerden çıkan ve Türkiye’ye gelen turist sayılarından, Türkiye’nin sahip olduğu pay elde edilmiştir. Tablo 3 veri setinde çalışmaya dahil edilen, 2019 yılı için niteliklere ilişkin tanımlayıcı istatistikleri, Tablo 4 ise veri setinde çalışmaya dahil edilen, 2020 yılı için niteliklere ilişkin tanımlayıcı istatistikleri sunmaktadır.

Tablo 3. 2019 yılı verilerine ilişkin tanımlayıcı istatistik değerleri
(Table 3. The descriptive statistics for 2019 dataset)

	Maks.	Min.	Ort.	Std. Sap.
Kişi başına düşen GSYİH	117.197,48	1.308,14	28.716,45	25.482,34
Çıkan Turist Sayısı	170.930.000,00	311.000,00	21.529.015,38	33.054.100,90
Türkiye’nin Payı	30,88	0,08	2,76	4,72
Türkiye’de Konaklama süresi	4,86	1,47	3,12	0,94

Tablo 4. 2020 yılı verilerine ilişkin tanımlayıcı istatistik değerleri
(Table 4. The descriptive statistics for 2020 dataset)

	Maks	Min	Ort	Std. Sap.
Kişi başına düşen GSYİH	113.218,71	3.114,62	28.783,58	25.236,56
Çıkan Turist Sayısı	60.550.000,00	130.000,00	6.400.200,00	10.755.135,98
Türkiye’nin Payı	24,59	0,03	2,56	4,46
Türkiye’de Konaklama süresi	3,65	1,60	2,70	0,51

3.3. Kümeleme Analizi

Bu çalışmada ülkeleri gruplara ayırmak için iki aşamalı bir küme analizi yöntemi uygulanmıştır. Bu yöntemde ilk aşamada, uygun k küme sayısını saptamak için Ward yöntemi, ikinci aşamada verileri belirlenen sayıda (k) gruba ayırmak için K-ortalama kümeleme algoritması kullanılmaktadır. Ward yönetiminde veri matrisinin satırları arasındaki mesafeler öklid yöntemi ile hesaplanmıştır. Ward yönteminden dendrogram elde edilmiş ve bunun üzerinden küme sayısı belirlenmiştir. Uygun küme sayısı K-ortalama yönteminde k değeri olarak belirlenmiştir.

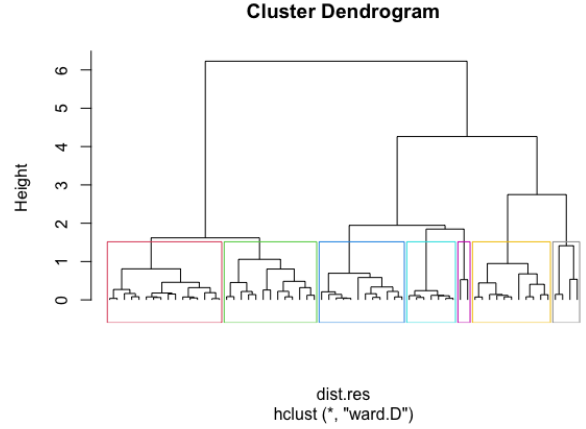
Kümeleme analizinden önce, standardizasyon işlemi yapılmıştır. Standardizasyon, her özelliğin ortalama değerini kaldırarak verileri merkezileştirmek ve ardından (sabit olmayan) özellikleri standart sapmalarına bölerek ölçeklendirmek için kullanılır. Verileri standartlaştırdıktan sonra ortalama sıfır olacak ve standart sapma bir e eşit olacaktır.

4. Uygulama ve Bulgular

Bu bölümde, kullanılan yöntemlerin uygulaması ve analizlerden elde edilen bulgular 2019 ve 2020 yılları için ayrı ayrı verilmiştir.

4.1. 2019 Yılı Analizi

2019 yılı verileri üzerinde Ward hiyerarşik kümeleme yönteminin çalıştırılmasıyla elde edilen dendrogram Şekil 1'de verilmiştir. Dendrogramda görüldüğü gibi optimum küme sayısı yedi olarak belirlenmiştir.

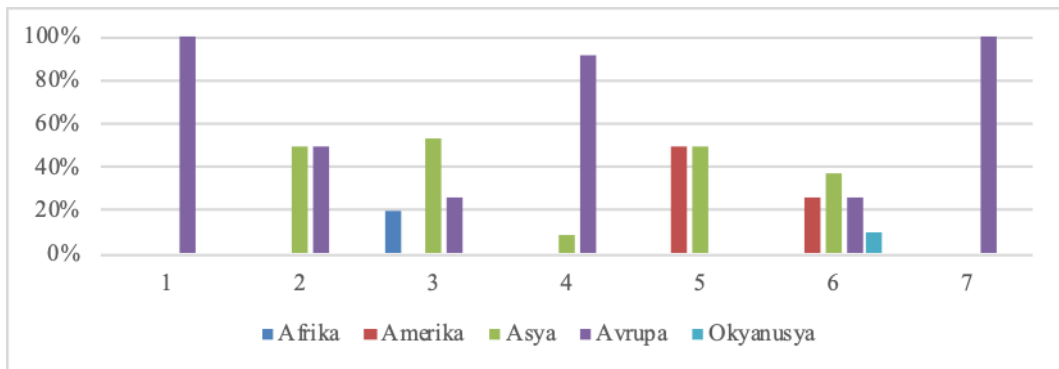


Şekil 1. 2019 yılı verileri için üretilen dendrogram (Figure 1. Dendrogram produced based on 2019 dataset)

Ward yöntemi ile tespit edilen optimal $k=7$ değeriyle, K-ortalama yöntemi çalıştırılmış ve elde edilen analiz sonuçları Tablo 5'te gösterilmiştir. Şekil 2'de ise elde edilen bu kümelerdeki ülkelerin kıtalara göre dağılımı gösterilmiştir.

Tablo 5. 2019 yılı verileri için K-means kümeleme analizi sonuçları (Table 5. Clustering results by K-means method for 2019 dataset)

Küme	Adet	Kişi Başına Düşen GSYİH (\$)	Çıkan Turist Sayısı	Türkiye'nin Payı (%)	Türkiye'de Konaklama süresi (gün)
1	10	49.337,82	33.780.100,00	2,63	3,86
2	2	4.269,26	924.000,00	26,14	3,60
3	15	9.843,23	6.316.400,00	2,32	2,94
4	12	14.811,45	17.888.083,33	3,70	4,21
5	2	36.484,88	162.781.000,00	0,50	1,92
6	19	27.731,81	20.188.210,53	0,80	2,09
7	5	87.878,50	8.239.200,00	1,10	3,81
Avg		28.716,45	21.529.015,00	2,76	3,12



Şekil 2. 2019 Yılı analizi kümelerinin kıtalara göre dağılımı (Figure 2. Clusters of analysis based on 2019 dataset by continents)

Tablo 5'te verilen 2019 yılı genel durumuna göre gelir düzeyinin en yüksek olduğu iki küme sırasıyla, Küme 7 ve Küme 1'dir. Her iki küme de de gelir düzeyleri gibi Türkiye'de konaklama süreleri yüksektir. Küme 7'nin turist potansiyeli ve pazar payı düşükken, Küme 1'de ise sadece pazar payı düşüktür. Küme 1'de Danimarka, İsveç, Almanya gibi gelir düzeyi yüksek Avrupa ülkeleri yer alırken, Küme 7'de İrlanda, İsviçre, İzlanda, Lüksemburg ve Norveç ülkeleri bulunmaktadır. Şekil 2'de en yüksek gelir düzeyine sahip bu kümelerin tamamının Avrupa kıtasında olduğu görülmektedir.

Tablo 5'te yer alan analizde turist potansiyeli olarak en yüksek iki küme ise Küme 5 ve Küme 1'dir. Küme 5'in gelir düzeyi yüksek olmasına rağmen pazar payı ve konaklama süresi düşüktür. Küme 5'te Amerika Birleşik Devletleri ve Çin Halk Cumhuriyeti ülkeleri yer alırken, Şekil 2'de görüleceği üzere %50 Asya ve %50 Amerika kıtaları ülkelerinden oluştuğu görülmektedir. Bu analizlerden yola çıkarak gelir düzeyi en yüksek olan Küme 7 olmasına rağmen turist potansiyeli açısından düşük kalmıştır. Buna rağmen Küme 5 ve Küme 1 hem gelir düzeyi olarak hem de turist potansiyeli yüksek olan ülkelerden oluşmaktadır.

Tablo 5'te verilen bilgilere göre en yüksek pazar payına sahip kümeler sırasıyla Küme 2 ve Küme 4 olmasına rağmen Küme 2'nin 26,14 pazar payı değeri ile ortalamadan çok yüksek bir değere sahip olduğu görülür. Küme 4 ve Küme 2 diğer özellikler yönünden de benzerlik gösterir. Her iki kümenin de gelir düzeyi ve turist potansiyeli düşük olmasına rağmen, pazar payı ve konaklama süresi yüksektir. Küme 4'te Çekya, Litvanya, Slovakya, Belarus gibi ülkeler yer alırken Küme 2'de Moldova ve Ürdün ülkeleri yer almaktadır. Şekil 2'den görüleceği üzere Küme 4'te yer alan ülkelerin %8,33'ü Asya, %91,67'si ise Avrupa kıtasındadır. Bu oran Küme 2'de ise %50 Asya ve %50 Avrupa kıtalarından oluşmaktadır.

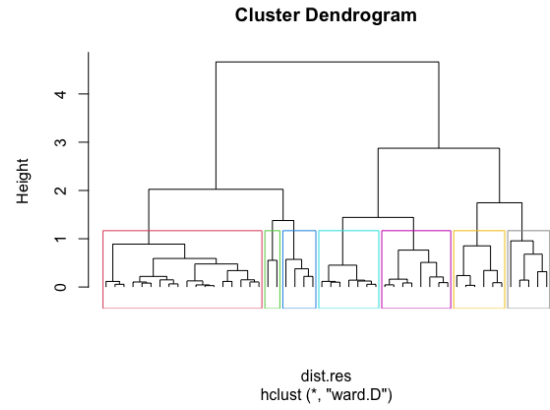
Turistlerin Türkiye'deki konaklama süreleri incelendiğinde Tablo 5'ten de görüleceği üzere en yüksek konaklama süresine sırasıyla Küme 4 ve Küme 1 sahiptir. Küme 1'de diğer özelliklere oranla sadece pazar payı düşüktür. Küme 4'te ise gelir düzeyi ve turist potansiyeli düşük, pazar payı ve konaklama süreleri ise yüksektir. Tablo 5'teki 2019 Yılı analiz sonuçlarına göre oluşturulan kümelerin profilleri özet olarak Tablo 6'da sunulmuştur.

Tablo 6. 2019 Yılı veri setine dayalı küme karakteristikleri
(Table 6. Cluster characteristics based on 2019 dataset)

Küme	Ekonomik Gelir Düzeyi	Turist Potansiyeli	Pazar Payı	Konaklama süresi
1	Yüksek	Yüksek	Düşük	Yüksek
2	Düşük	Düşük	Yüksek	Yüksek
3	Düşük	Düşük	Düşük	Düşük
4	Düşük	Düşük	Yüksek	Yüksek
5	Yüksek	Yüksek	Düşük	Düşük
6	Düşük	Düşük	Düşük	Düşük
7	Yüksek	Düşük	Düşük	Yüksek

4.2. 2020 Yılı Analizi

2020 yılı verileri üzerinde Ward hiyerarşik kümeleme yönteminin çalıştırılmasıyla elde edilen dendrogram Şekil 3'te verilmiştir. Dendrogram incelendiğinde 2019 yılı verileri gibi 2020 yılı verileri içinde optimum küme sayısı yedi olarak belirlenmiştir.

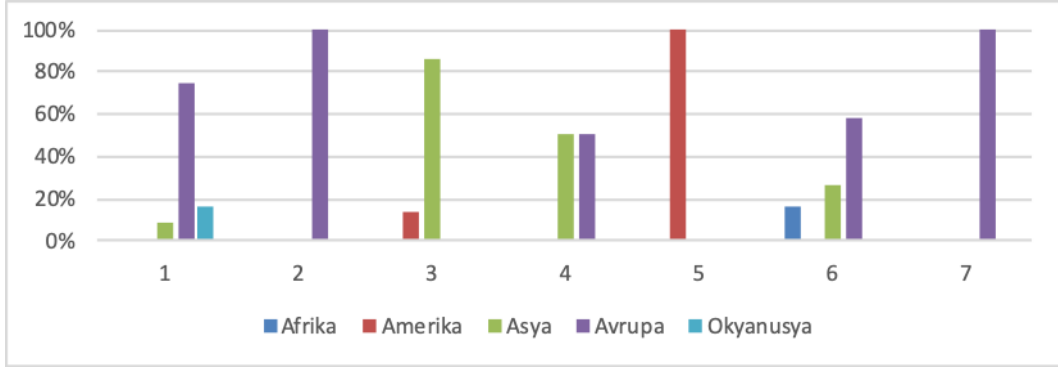


Şekil 3. 2020 yılı verileri için üretilen dendrogram
(Figure 3. Dendrogram produced based on 2020 dataset)

Ward yöntemi ile tespit edilen k=7 değeriyle, K-ortalamlar yöntemi analiz sonuçları Tablo 7'de gösterildiği gibidir. Şekil 4'te ise kümelerde yer alan ülkelerin kıta dağılımlarına yer verilmiştir.

Tablo 7. 2020 yılı verileri için K-means kümeleme analizi sonuçları
(Table 7. Clustering results by K-means method for 2020 dataset)

Küme	Adet	Kişi Başına Düşen GSYİH (\$)	Çıkan Turist Sayısı	Türkiye'nin Payı (%)	Türkiye'de Konaklama süresi (gün)
1	12	44.282,76	6.817.083,33	0,63	2,64
2	4	17.874,41	9.644.500,00	4,73	3,60
3	7	23.220,64	5.000.428,57	0,98	1,77
4	2	7.951,57	6.383.500,00	21,18	3,07
5	2	37.614,99	48.303.000,00	0,15	2,26
6	19	12.828,30	2.612.894,74	2,46	2,85
7	4	84.718,22	1.401.500,00	1,37	2,89
Avg		28.783,58	6.400.200,00	2,56	2,70



Şekil 4. 2020 Yılı analizi kümelerinin kıtalara göre dağılımı
(Figure 4. Clusters of analysis based on 2020 dataset by continents)

Tablo 7’de verilen 2020 yılı genel durumu analizine göre gelir düzeyinin en yüksek olduğu iki küme sırasıyla, Küme 7 ve Küme 1’dir. Küme 7’nin turist potansiyeli ve pazar payı düşük olmasına rağmen konaklama süresi yüksektir. Küme 1’de ise turist potansiyeli yüksek, pazar payı ve konaklama süresi düşüktür. Küme 1’de Avustralya, Avusturya, Yeni Zelanda gibi ülkeler yer alırken, Küme 7’de İrlanda, İzlanda, Lüksemburg ve Norveç ülkeleri bulunmaktadır. Şekil 4’te Küme 1’in %8,33’ünü Asya, %75’ini Avrupa, %16,67’sini Okyanusya ülkelerinden, Küme 7’nin ise sadece Avrupa ülkelerinden oluştuğu görülmektedir.

Tablo 7’de yer alan analizde turist potansiyeli olarak en yüksek iki küme ise Küme 5 ve Küme 2’dir. Küme 2’nin gelir düzeyi düşük iken turist potansiyeli, pazar payı ve konaklama süresinin yüksek olduğu görülür. Küme 5’in ise gelir düzeyi ve turist potansiyeli yüksek olmasına rağmen pazar payı ve konaklama süresi düşüktür. Küme 5’te Amerika Birleşik Devletleri ve Meksika ülkeleri yer alırken, Küme 2’de Belarus, İngiltere, Letonya ve Ukrayna ülkeleri olduğu görülmektedir. Şekil 4’ten görüleceği üzere Küme 5’te yer alan tüm ülkelerin Amerika, Küme 2’nin sadece Avrupa ülkelerinden oluştuğu görülmektedir.

Tablo 7 pazar payı yönünden incelendiğinde en yüksek iki kümenin Küme 4 ve Küme 2 olduğu görülmektedir. Küme 4’ün pazar payı değeri 21,18 ile ortalamadan çok üzerindedir. Küme 4’ün gelir düzeyinin ve turist potansiyelinin düşük, konaklama süresinin yüksek olduğu görülür. Küme 2’de ise gelir düzeyi düşük, turist potansiyeli ve konaklama süresi yüksek olan Avrupa ülkeleri yer almaktadır. Küme 4’te Rusya ve Ürdün ülkeleri bulunurken, Küme 2’nin aksine %50’sinin Avrupa, %50’sinin Asya ülkelerinden oluşmaktadır.

Turistlerin Türkiye’deki konaklama süreleri incelendiğinde Tablo 7’den de görüleceği üzere en yüksek konaklama süresine sırasıyla Küme 2 ve Küme 4 sahiptir. Bu iki kümenin pazar payı ve konaklama süreleri en yüksek değere sahipken, ekonomik gelir düzeylerinin de ortalamanın altında olması dikkat çekicidir. Tablo 7’deki 2020 yılı analiz sonuçlarına göre oluşturulan kümelerin profilleri özet olarak Tablo 8’de sunulmuştur.

Tablo 8. 2020 Yılı veri setine dayalı küme karakteristikleri
(Table 8. Cluster characteristics based on 2020 dataset)

Küme	Ekonomik Gelir Düzeyi	Turist Potansiyeli	Pazar Payı	Konaklama süresi
1	Yüksek	Yüksek	Düşük	Düşük
2	Düşük	Yüksek	Yüksek	Yüksek
3	Düşük	Düşük	Düşük	Düşük
4	Düşük	Düşük	Yüksek	Yüksek
5	Yüksek	Yüksek	Düşük	Düşük
6	Düşük	Düşük	Düşük	Yüksek
7	Yüksek	Düşük	Düşük	Yüksek

5. Sonuç

Bu çalışmada, ülkelerin Türkiye’ye olan dış turizm talebi kümeleme analizi kullanılarak incelenmiştir. Kümeleme analizinde, ülkelerin kişi başına düşen gayrisafi yurtiçi hasıla, ülkelere çıkan turist sayısı, bu sayılarda Türkiye’nin payı ve turistlerin Türkiye’de konaklama süresi özellikleri dikkate alınmıştır. Veriler, 2019 ve 2020 yılları için ayrı ayrı temin edilmiş ve iki ayrı kümeleme gerçekleştirilmiştir. İki aşamalı kümeleme yöntemi kullanılmış, bu yöntemde önce Ward algoritması ile optimum küme sayısı belirlenmiş, sonrasında ise bu değer girdi olarak kullanılarak K-ortalamlar tekniğiyle ülkeler gruplara ayrılmıştır. Elde edilen kümelerin profili çıkarılmış ve kümelerdeki ülkeler kıtalar açısından da analiz edilmiştir.

Değerlendirmeler sonucunda, her iki yıl analizlerine göre ülkeler 7 gruba ayrılmıştır. Sonuçlar detaylı incelendiğinde, Amerika Birleşik Devletleri ve Çin Halk Cumhuriyeti 2019 yılında Küme 5’i oluştururken, 2020 yılında Çin Halk Cumhuriyeti’nin kümedeki yerini Meksika alarak Küme 2’yi oluşturmuştur. 2019 yılında Küme 7’de İrlanda, İsviçre, İzlanda, Lüksemburg ve Norveç ülkeleri yer alırken benzer özelliklere sahip 2020 yılında Küme 7’yi oluşturan ülkelerin arasında 2020 yılı verilerinde yer almamasından dolayı İsviçre’nin olmadığı gözlemlenmiştir.

Bu çalışmanın sonucunda elde edilen bulgular ve küme profilleri ile belirli hedef ülkelere gelen turist sayısını arttırmak ve turistlerin konaklama sürelerini uzatabilmek için çeşitli pazarlama çalışmaları yapılabilir. Turizm tesisleri reklam çalışmalarını uygulamadan önce bu çalışma içerisinden elde ettiği bilgiler doğrultusunda doğru stratejiler belirleyebilir. Hali hazırda veri madenciliği ve yapay zekâ uygulamalarının turizm

sektöründe aktif olarak kullanılmasına karşın uluslararası turizmde bu uygulamalar literatürde yeterince yer bulamamıştır. Literatürde yeteri kadar çalışma bulunmadığından çalışmamız uluslararası turizm sektöründe planlama ve pazarlama konularında fayda sağlayacaktır.

Geçmişte Giray (2013), 2011 yılına ilişkin uluslararası turizm verilerini baz alıp benzer bir şekilde bulanık kümeleme ve klasik kümeleme (k ortalamalar) analizleri kullanarak ülkeleri sınıflandırmıştır. Bu çalışmada ülkelerin turizm harcamaları, turizm gelirleri ve geliş sayıları dikkate alınmış olup bu niteliklere ve kullanılan verilere göre Türkiye'nin hangi segmentte hangi ülkelerle birlikte yer aldığı saptanmıştır. Bu çalışmadan farklı olarak bizim çalışmamız daha güncel veriler ışığında Türkiye'ye olan dış turizm talebini incelemiştir. Ülkelerin kişi başına düşen gayrisafi yurtiçi hasıla, ülkelerden çıkan turist sayısı, bu sayılarda Türkiye'nin payı ve turistlerin Türkiye'de konaklama süresi gibi daha kapsamlı özellikler dikkate alınarak diğer ülkeler Türkiye talebi açısından gruplara ayrılmıştır. Bu bakımdan elde edilen bulgular ışığında, ülke gruplarına yönelik turizm sektöründe strateji ve hedefler daha net ve ayrıntılı bir şekilde oluşturulabilecektir.

Bu çalışmanın mevcut literatüre katkıları ve faydaları olmasına karşın bir dizi kısıtları da mevcuttur. Çalışmada kullanılan veri, beş özellik içermekte olup sadece iki yıla (2019-2020) ilişkindir. Gelecekteki çalışmalarda bu çalışmada kullanılan özelliklerin yanı sıra turistlerin ırk, cinsiyet, yaş gibi kişisel özelliklerini dikkate alabilir ve farklı ve daha geniş bir zaman dilimi için ilgili analizleri genişleterek bu çalışmanın bulgularını geçerleyebilirler. Bir diğer kısıt bilgi kaybından dolayı çalışmaya dahil edilen ülke sayısıdır. Bu açıdan gelecekte daha fazla ülkeyi içeren veri setleri üzerine benzer çalışmalar yapılabilir. Ayrıca farklı kümeleme yöntemleri ve algoritmaları kullanılarak gelecekte yine benzer çalışmalar yapılabilir, elde edilen sonuçlar ve bulgular bu çalışmanınki ile kıyaslanabilir.

Kaynakça

- Akgün, A., & Çizel, B. (2016). Günlük Tur Programları Oluşturmada Veri Madenciliği: A Grubu Seyahat Acentası Örneği. *Turizm Turizm ve Araştırma Dergisi*, 6(1), 73–87. <https://doi.org/10.1007/s11135-021-01234-3>
- Akgün, A., Çizel, B., & Ajanovic, E. (2021). Mining excursion tourist profile through classification algorithms. *Quality and Quantity*, 1–22. <https://doi.org/10.1007/s11135-021-01234-3>
- Akın, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management*, 48, 64–72. <https://doi.org/10.1016/j.tourman.2014.11.004>
- Aydın, A., Darıcı, B., & Taşçı, H. M. (2015). Uluslararası Turizm Talebini Etkileyen Ekonomik Faktörler: Türkiye Üzerine Bir Uygulama. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 45, 143–177. <https://doi.org/10.18070/euiibfd.85938>
- Bahar, O. (2006). Turizm Sektörünün Türkiye'nin Ekonomik Büyümesi Üzerindeki Etkisi: VAR Analizi Yaklaşımı, *Yönetim ve Ekonomi Dergisi* 13(2), 137–150.
- Bardak, T., & Sözen, E. (2018). Veri madenciliği ve önemi. 6th ASM International Congress of Agriculture and Environment, 2018 Proceeding Book.
- Çuhadar, M. (2015). Turizm Sektöründe Talep Tahmini İçin Yapay Sinir Ağları Kullanımı ve Diğer Yöntemlerle Karşılaştırmalı Analizi (Antalya İlinin Dış Turizm Talebinde

- Uygulama), 38–49. <https://doi.org/10.13140/RG.2.1.4503.7288>
- Doborjeh, Z., Hemmington, N., Doborjeh, M., & Kasabov, N. (2022). Artificial intelligence: a systematic review of methods and applications in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 34(3), 1154–1176. <https://doi.org/10.1108/IJCHM-06-2021-0767>
- Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 18, 153–160. <https://doi.org/10.1016/j.tmp.2016.03.001>
- Ersungur, Ş. M., Doru, Ö., & Aslan, M. B. (2017). Türkiye'ye Yönelik Uluslararası Turizm Talebinin Sosyo-Ekonomik Belirleyicileri: Panel Veri Analiz. *International Conference on Eurasian Economies*, 433–441. <https://doi.org/10.16951/atauniiibd.769566>
- Garda, B., & Temizel, M. (2016). Sürdürülebilir Turizm Çeşitleri. *Selcuk University Journal of Social and Technical*, 12, 83–103.
- GDP per capita, current dollars by country, around the world | TheGlobalEconomy.com. (n.d.). Retrieved June 19, 2022, from https://www.theglobaleconomy.com/rankings/GDP_per_capita_current_dollars/
- Giray, S. (2013). Ülkelerin Turizm İstatistikleri Bakımından Farklı Kümeleme Analizi Metotları ile Sınıflandırılması ve Türkiye'nin Bu Oluşumdaki Yeri. 695–704.
- Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2021). Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach. *Journal of Travel Research*, 60(5), 998–1017. <https://doi.org/10.1177/0047287520921244>
- Huang, B., & Hao, H. (2021). A novel two-step procedure for tourism demand forecasting. *Current Issues in Tourism*, 24(9), 1199–1210. <https://doi.org/10.1080/13683500.2020.1770705>
- International tourism, number of departures | Data. (n.d.). Retrieved June 20, 2022, from https://data.worldbank.org/indicator/ST.INT.DPRT?end=2020&most_recent_value_desc=true&start=2019
- International Tourism Highlights, 2020 Edition. (2021). In *International Tourism Highlights, 2020 Edition*. <https://doi.org/10.18111/9789284422456>
- İşletme (Bakanlık) Belgeli Tesis Konaklama İstatistikleri. (n.d.). Retrieved June 19, 2022, from <https://yigm.ktb.gov.tr/TR-201121/isletme-bakanlik-belgeli-tesis-konaklama-istatistikleri.html>
- Kandemir, A. Ş. (2018). Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Konaklama İstatistiklerine Göre Sınıflandırılması. *Journal of Travel and Hospitality Management*, 15(3), 657–668. <http://dergipark.gov.tr/download/article-file/585914>
- Karaoğlu, N. (2019). Türkiye'de Uluslararası Turizm Talebinin Dinamik Panel Veri Analizi. *Bulletin of Economic Theory and Analysis*, 4(1), 85–101. <https://doi.org/10.25229/beta.561802>
- Kaya, A. A., & Canlı, B. (2013). Türkiye'ye Yönelik Uluslararası Turizm Talebinin Belirleyicileri: Panel Veri Yaklaşımı.
- Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand

- forecasting. *Annals of Tourism Research*, 83, 102925.
<https://doi.org/10.1016/j.annals.2020.102925>
- Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, 75(January), 410–423.
<https://doi.org/10.1016/j.annals.2019.01.014>
- List of Countries by Continent - *StatisticsTimes.com*. (n.d.). Retrieved June 19, 2022, from <https://statisticstimes.com/geography/countries-by-continent.php>
- Magnini, V. P., Honeycutt Jr, E. D., & Hodge, S. K. (2003). Data Mining for Hotel Firms. *Cornell Hotel and Restaurant Administration Quarterly*, 44(2), 94–105.
Doi:10.1177/0010880403442009
- Nguyen, L. Q., Fernandes, P. O., & Teixeira, J. P. (2021). Analyzing and Forecasting Tourism Demand in Vietnam with Artificial Neural Networks. *Forecasting*, 4(1), 36–50.
<https://doi.org/10.3390/forecast4010003>
- Özbek, T. (1991). Dünyada ve Türkiye’de termal turizmin önemi. *Anatolia*, 2, 17–18.
- Özcan, Ş. (2021). Turizm ve Ekonomik Büyüme İlişkisi: Üst-Orta Gelirli Ülkeler Üzerine Panel Veri Analizi.
- Sevimli Deniz, S. (2019). Veri Madenciliği Araçları Kullanılarak Türkiye’nin Turizm Gelirlerinin Aylara Göre Yapay Sinir Ağları İle Tahminlenmesi. *Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 241-255.
<https://dergipark.org.tr/pub/yyusbed/issue/51801/673555>
- Shapoval, V., Wang, M. C., Hara, T., & Shioya, H. (2018). Data Mining in Tourism Data Analysis: Inbound Visitors to Japan. *Journal of Travel Research*, 57(3), 310–323.
<https://doi.org/10.1177/0047287517696960>
- Ünlüönen, K., & Şen Küpeli, T. (2021). Veri Madenciliği ve Turizmde Veri Madenciliği Çalışmaları. *Journal of Tourism and Gastronomy Studies*, 9(1), 275–298.
<https://doi.org/10.21325/jotags.2021.7>
- Yıldırım, M. E. (2022). A Case Study : Unsupervised Approach for Tourist Profile Analysis by K-means Clustering in Turkey. *Journal of Internet Computing and Services*, 23(1), 11-17. <https://doi.org/10.7472/jksii.2022.23.1.11>
- Zhang, Y., Li, G., Muskat, B., & Law, R. (2021). Tourism Demand Forecasting: A Decomposed Deep Learning Approach. *Journal of Travel Research*, 60(5), 981–997.
<https://doi.org/10.1177/0047287520919522>
- Zortuk, M., & Bayrak, S. (2013). Seçilmiş Ülkelere Göre Türkiye’nin Turizm Talebi. *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, 19, 38–58.
<http://dergipark.gov.tr/iuekois/issue/8997/112139>