



Kümeleme Analizi

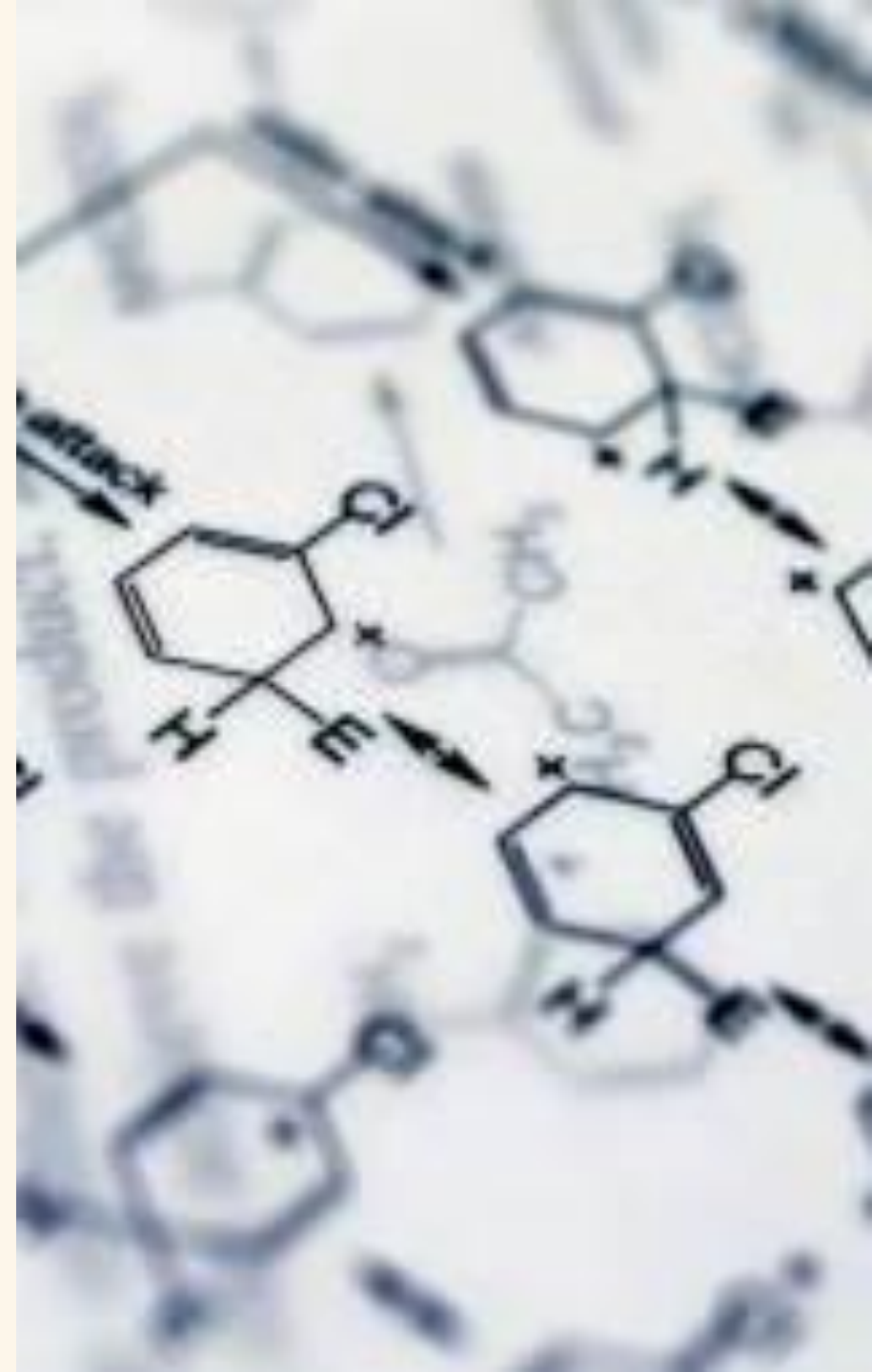
Nurcan Kurtoğlu

Kümeleme Analizi Nedir?

Kümeleme analizi gözlem birimlerini benzerliklerine göre homojen alt gruplar (kümeler) altında toplamayı amaçlayan çok değişkenli istatistiksel bir tekniktir. Bu yöntemle, veriler herhangi bir ön bilgi (etiket) olmadan gruplandırılır. Amaç, grup içi benzerliğin yüksek, gruplar arası benzerliğin ise düşük olduğu kümeler elde etmektir.

Bu analiz genellikle şu amaçlarla kullanılır:

- Pazar segmentasyonu
- Müşteri profilleme
- Sosyal bilimlerde örnek kümeleri belirleme
- Biyolojide tür sınıflandırması
- Makine öğrenmesinde örüntü tanıma





Kümeleme Analizinin Varsayımları

- Gözlemler Bağımsız Olmalıdır:**

Her gözlem birbiriyle ilişkili olmamalıdır.

- Uygun Değişken Seçimi Gerekir:**

Analizde kullanılan değişkenler anlamlı ve ölçülebilir olmalıdır.

- Ölçek Uyumu Önemlidir:**

Değişkenler farklı ölçeklerdeyse, veriler analizden önce standartlaştırılmalıdır.

- Küme Sayısı Belirlenmelidir:**

Hiyerarşik olmayan yöntemlerde kaç küme olacağı önceden belirlenmelidir.

- Aykırı Değerler Etkileyebilir:**

Verideki uç değerler küme yapısını bozabilir, bu yüzden ön analiz yapılmalıdır.

- Veri Türü Uygun Olmalıdır:**

Kümeleme genellikle sürekli (sayısal) verilerle daha iyi sonuç verir.

- Benzerlik/ Uzaklık Ölçütü Seçilmelidir:**

Veriler arasındaki ilişkiyi ölçmek için uygun bir benzerlik veya uzaklık ölçüsü kullanılmalıdır (örneğin Euclidean mesafesi).

Kümeleme Analizi Uygulama Aşamaları

1) Araştırma Probleminin Belirlenmesi

- Amaç Nedir?
- Kullanılacak değişkenler seçilir.

2)Araştırma Planı

- Araştırma tasarımı
- Uzaklık ve benzerlik ölçülerinin seçimi.

3)Varsayımlar

- Ölçme hatası olmamalıdır.
- Örneklem anakütleyi temsil etmelidir.

4)Kümeleme tekniğinin seçimi

- Hiyerarşik Kümeleme: Küme sayısı bilinmiyorsa kullanılır.
- Hiyerarşik Olmayan Kümeleme: Küme sayısı biliniyorsa kullanılır.
- İki Aşamalı Kümeleme: Birinci aşamada Hiyerarşik ikinci aşamada hiyerarşik olmayan kümeleme tekniğini kullanma esasına dayanır.

5)Küme Sayısı Belirlenmesi

6)Analiz Sonuçlarının Gözden Geçirilmesi

- Aykırı değer var mı?
- Çok küçük küme var mı?

7)Kümelerin Yorumlanması

8)Küme Profiline Belirlenmesi.



Uzaklık ve Benzerlik Ölçüleri

Uzaklık: İki gözlemin birbirinden ne kadar farklı olduğunun göstergesidir.

Benzerlik: İki gözlem incelenen özellik açısından birbirine ne kadar yakınsa benzerlik değeri de o kadar büyük çıkacaktır.

1. Öklid (Euclidean) Uzaklığı

İki nokta arasındaki düz (doğrusal) mesafeyi ölçer.

Formül:

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

Sayısal verilerde en yaygın kullanılan uzaklık ölçüsüdür.

2. Manhattan (City Block) Uzaklığı

İki nokta arasındaki mesafeyi yalnızca yatay ve dikey yollarla ölçer.

Formül:

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

Adını şehir bloklarından alır. En kısa yol değil, "köşeli" yol ölçülür.

3. Minkowski Uzaklığı

Öklid ve Manhattan uzaklıklarının genel bir biçimidir.

Formül:

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

$p = 1$ için Manhattan, $p = 2$ için Öklid olur.

4. Cosine Benzerliği

İki vektör arasındaki açıyı ölçerek benzerliği hesaplar.

Formül:

$$\text{Cosine Similarity} = \frac{x \cdot y}{\|x\| \cdot \|y\|}$$

Açı küçüldükçe benzerlik artar. Değer aralığı: -1 ile 1 .

5. Jaccard Benzerliği

İki kümenin ortak öğeleri ile birleşiminin oranı alınır. Genellikle kategorik/veri madenciliği uygulamalarında kullanılır.

Formül:

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

0 ile 1 arasında değer alır. 1 = tam benzerlik.

Kümeleme Analizinin Yöntemleri

1.Grafik Yöntemi

- İki ya da daha fazla 3 boyut olduğunda kullanılır.
- Değişken sayısı da gözlem sayısı da çok olmamalıdır. Gözlem sayısının az değişken sayısının çok olduğu durumlarda boyut indirgeme yapılarak grafik yöntem kullanılır.

2. Hiyerarjik Yöntem

- Gözlemler arası **uzaklık minimum** olmalıdır.
- Küme sayısının bilinmediği** durumlarda kullanılır.
- Hiyerarjik yöntemlerde ilk olarak **her gözlem bir küme olacak** şekilde analize başlanır. Yani n adet bireyden n adet küme olacaktır.
- Daha sonra **en yakın 2 küme** yani **dij değeri minimum** olan **2 küme birleştirilerek** analize devam edilir.
- Her aşamada küme sayısı **1 indirgenerek** yeniden uzaklık matrisi hesaplanır.
- Kullanılan hiyerarjik kümeleme tekniğine göre uzaklık matrisleri ve kümeler değişir.
- Bu adımlar (n-1) kez tekrarlanarak n gözlemi de kopyacak tek bir küme elde edilene kadar analize devam edilir. Elde edilen bu yapıya (ağaç diyagramı) ya da dendogram adı verilir.

2.1 Tek Bağlantı Tekniği(En yakın komşuculuk)

- Uzaklık matrisi incelenerek birbirine en benzer gözlemlerin aynı kümede toplanması prensibine dayanır.**İlk adımda en benzer gözlemler birleştirilerek bir küme oluşturulduktan sonra, sonraki uzaklık matrisi oluşturulurken bir gözlemin oluşturulan kümeye olan gözlemin değeri alınarak yapılır.

2.2 Tam Bağlantı Tekniği (En uzak komşuculuk)

- Tek bağlantı tekniğinden farklı olarak yeni uzaklık değerleri bulunurken bir gözlemin bir kümeye olan uzaklık değerinden maximum olanı alınır.** Yine uzaklık değeri en küçük olan değerler birleştirilir.Sadece yeni uzaklık matrisi oluşturulurken maksimum olan uzaklık seçilir.Bu iki teknik, serideki uç değerlerden çokça etkilenir.

2.3 Ortalama Bağlantı Tekniği

- Kümeyi oluşturan verilerin birbirleri arasındaki **uzaklıkların ortalama** değerini dikkate alarak gerçekleşen birleşme işlemidir.Çok fazla önerilmez. Çünkü ortalama aldıkça iki gözlem arasındaki farklılık merkezleştirilir ve gerçeklikten sapar.

2.4 Ward Tekniği

- En çok önerilen tekniktir ancak elle hesaplama yapılamaz.**Tüm bağlantıları ve olası kümeleri ele alıp sapma kareler toplamı en küçük olan gözlemleri kümeler.Böylelikle minimum varyans elde edilmiş olur.**En homojen kümeler Ward Tekniği ile elde edilir.

2.5 Merkezi Bağlantı

2.6 Medyan

Kümeleme Analizinin Yöntemleri

HİYERARŞİK OLMAYAN TEKNİKLER

Küme sayısının bilindiği durumlarda kullanılır.

Örneğin ülkelerin gelişmişlik düzeyine göre sınıflandırmak istediğimizde literatürde 4 küme ile işleme başlanabilir.

İki farklı yöntemle yapılır:

-K ortalamalar

-En çok olabilirlik yöntemi

K-Ortalamalar Yöntemi

Bu yöntemi kullanabilmek için verilerin mutlaka metrik olması gerekir.

UYGULAMA AŞAMALARI

- 1) n tane gözlem için k tanesi tesadüfi olarak seçilip her biri bir gözlemlili küme haline getirilir.
- 2) Kalan n-k gözlem içinden bu kümelere en yakın olanlar atanır ve küme ortalamaları (centroidleri) hesaplanır.
- 3) Her aşamada küme ortalamasına en yakın gözlemler ilgili kümeye atandıktan sonra atamalar yeniden hesaplanır.
- 4) Kümeler arası geçiş durana kadar işleme devam edilir. Varyans minimum oluncaya kadar geçiş devam eder.
- 5) En son elde edilen kümeler arası anlamlı fark olup olmadığını tespit etmek için ANOVA yapılır.
 - İyi bir kümeleme analizinde küme merkezlerinin birbirinde olabildiğince farklı olması beklenir.Bunu test etmek için de Anova kullanılır.
 - Ayrım gücü olmayan değişkenler analizden çıkarılarak analiz yeniden yapılır.

Kümeleme Analizinin Yöntemleri

En Çok Olabilirlik Yöntemi

Her bir gözlemin bağımsız ve aynı olasılık yoğunluk fonksiyonuna sahip olduğu varsayımdan hareketle gözlemleri birçok olabilirlik fonksiyonu boyunca maksimum değeri verecek şekilde kümelere ayırma amacı taşır.

Sık kullanılmaz.

UYGULAMA AŞAMALARI

1)Küme sayısının belirlenmesi

2)Mariott Yöntemi

3) Calinski-Harabsz Yöntemi

-Küme sayısını en doğru şekilde belirleme yöntemi literatür taramasıdır.

İKİ AŞAMALI KÜMELEME ANALİZİ (TWO STEP)

Birinci aşamada Hiyerarşik ikinci aşamada hiyerarşik olmayan kümeleme tekniğini kullanma esasına dayanır.

İlk adımda küme sayısı belirlenir.

İkinci adımda gözlemler kümelere atanır.

Yöntem Akaike ve Schwarz bilgi kriterlerine bağlı olarak kümelere ayrılır. Yöntem genellikle gözlemleri iki kümeye atama eğilimindedir. Küme sayısının belirleme işlemi uzaklık ölçüleri oranının maksimum olduğu noktada durur. Bu da genelde iki küme durumunda ortaya çıkar.

Önemli olabilecek bazı alt kümeleri vermez.

K-Ortalamalar Kümeleme Yöntemi ile G-7 Ülkelerinin Ekonomik Özgürlükler Açısından Karşılaştırılması

G-7 ülkeleri, ekonomik özgürlük endeksi içinde yer alan 12 faktör kullanılarak, Kümeleme analizi yöntemlerinden hiyerarşik olmayan **K-Ortalamalar tekniği** ile irdelenmiştir.

- Hiyerarşik olmayan Kümeleme yöntemlerinde ayrıştırılmak istenen küme sayısı, araştırmacı tarafından önceden belirlenir.
- Yöntemin **ilk adımında** da bu sayı kadar **rastlantısal olarak seçilen başlangıç merkez veriler yardımıyla** analiz uygulanır.
- Bu çalışmada, veri seti içinden olası **tüm başlangıç merkez verilerle** ve ayrıca bir yöntemle veri seti dışından elde edilen yeni verilerle K-Ortalamalar tekniği uygulanarak **G-7 ülkeleri iki kümeye** ayrıştırılmıştır.

Temel bulgular; 2018 yılında, Amerika ve Japonya'nın tüm küme gruplarında birlikte yer aldığıdır. Ayrıca 2019 yılı için elde edilen altı farklı küme grubunun beşinde aynı kümede yer alması oldukça dikkat çekicidir. Birçok farklı küme başlangıcı ile başlanan iterasyonlar ile çoğunlukla aynı durum ortaya çıkması, iki ülkenin kullanılan göstergelere göre benzerlikler gösterdiğini ortaya çıkarmaktadır.

KULLANILAN YÖNTEM

Bu çalışmada, G-7 ülkelerinin (Amerika, Kanada, Almanya, İtalya, Fransa, Japonya ve İngiltere) 2017-2019 dönem aralığı için ekonomik özgürlük endeksi açısından gruplandırılması (kümelere ayrıştırılması) amaçlanmıştır. Çalışmada, çok değişkenli istatistiksel analiz yöntemlerinden **Kümeleme analizi kullanılmıştır**. Kümeleme analizinde ise hiyerarşik olmayan Kümeleme yöntemlerinden **K-Ortalamlar** yöntemi tercih edilmiştir. Ülkeler ekonomik özgürlük endeksi içerisinde yer alan **12 faktör** dikkate alınarak iki kümeye ayrıştırılmıştır. Dönem aralığı ise 12 değerlendirme faktörünün değerlerine ulaşılabildiği 2017-2019 dönem aralığıdır.

Tablo 1: Kategorilere Göre Değerlendirme Faktörleri

	Notasyon	Tanım		Notasyon	Tanım
Hukukun Üstünlüğü	K ₁	Mülkiyet Hakları	Etkin Düzenleme	K ₇	İş Özgürlüğü
	K ₂	Yargının Etkinliği		K ₈	İşgücü Özgürlüğü
	K ₃	Devletin Bütünlüğü		K ₉	Parasal Özgürlük
Devletin Büyüklüğü	K ₄	Vergi Yüğü	Açık Piyasa Yapısı	K ₁₀	Ticaret Özgürlüğü
	K ₅	Kamu Harcaması		K ₁₁	Yatırım Özgürlüğü
	K ₆	Mali Özgürlük		K ₁₂	Finansal Özgürlük

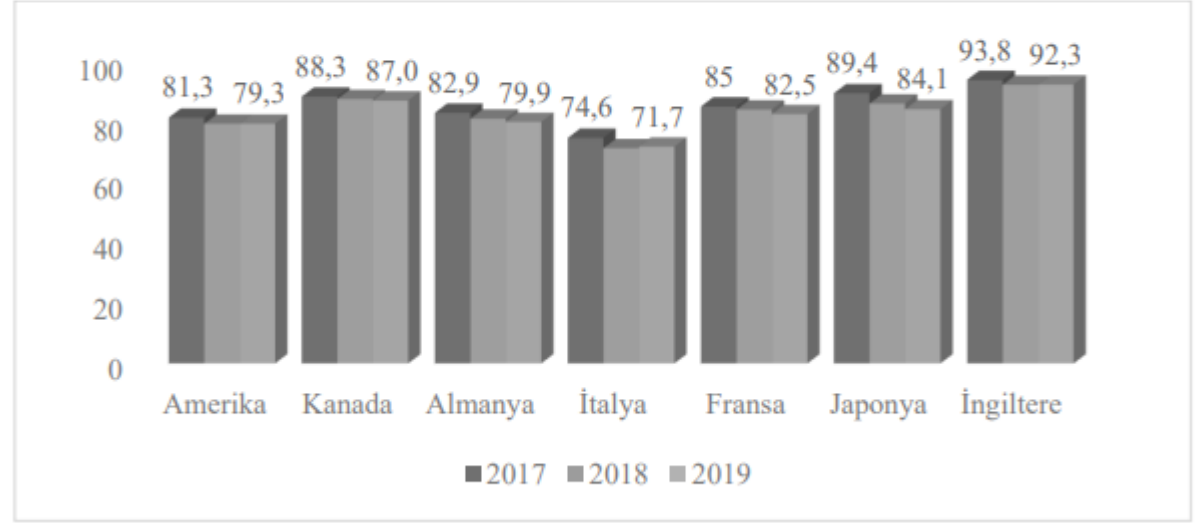
Kaynak: McMahon, 2018, <https://www.heritage.org/index/book/methodology>, (Erişim Tarihi: 21.07.2019).

MÜLKİYET HAKKI

Ekonomik özgürlük endeksinde yer alan 12 faktörden her biri 0-100 arasında puanlanır.

Bu puanlamanın görüldüğü ve mülkiyet hakkının G-7 ülkeleri için incelendiği Şekil 1’den mülkiyet hakları oranının **en yüksek olduğu ülkenin İngiltere** olduğu görülür. İngiltere’de özel mülkiyet haklarının ve sözleşmelerin oldukça güvenli olması, bu oranı yükselten unsurlardan biridir. Aynı **oran İtalya’da diğer ülkelere göre daha düşüktür.** Ülkede mülkiyet hakları ve sözleşmelerin güvenliğinde bir problem görülmezken, mahkeme prosedürlerinin yavaş olması bu oranı negatif yönde etkilemektedir

Şekil 1: Mülkiyet Hakkı: G-7 Ülkeleri



YARGININ ETKİNLİĞİ

Yargı etkinliği özellikle gelişmekte olan ülkeler için ekonomik büyümenin temellendirilmesinde oldukça önemlidir.

Ekonomik özgürlüğün ekonomik büyüme üzerindeki etkilerini inceledikleri çalışmada, bu teoriye uygun bir sonuca ulaşmış, ekonomik özgürlüklerin iktisadi büyüme sürecini hızlandığını belirtmişlerdir.

Yargının etkinliği bakımından **en gelişmiş ülke**

Şekil 2’de belirtildiği gibi **İngiltere’dir**.

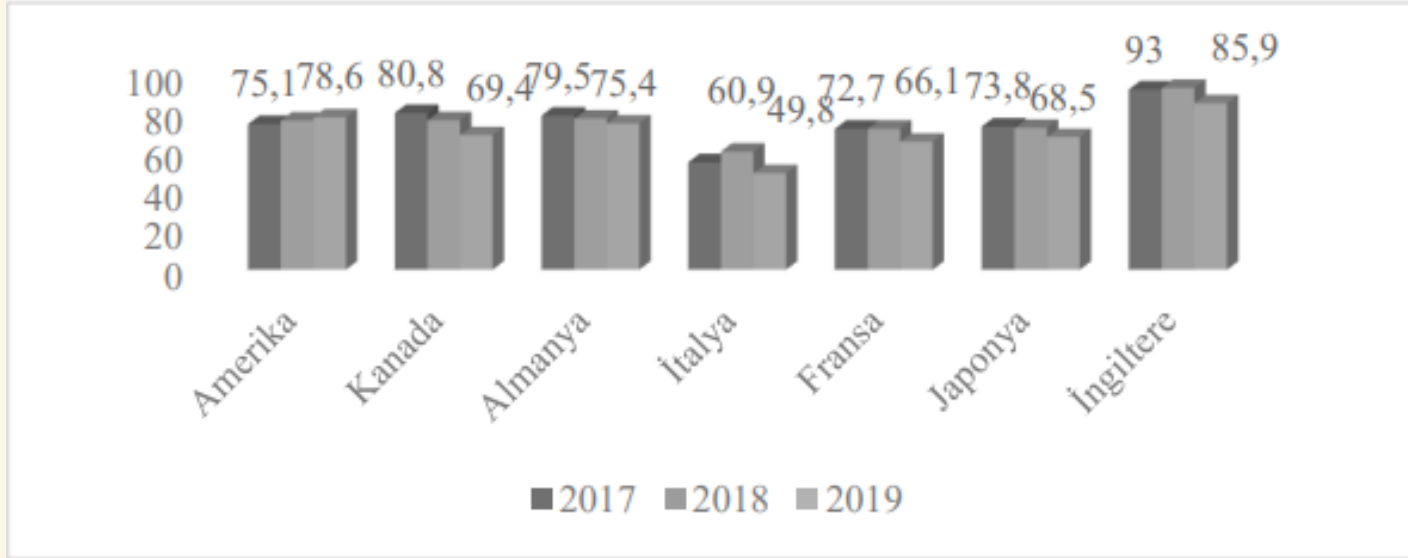
- **2017** yılında bu oran **93** iken,
- **2019** yılında ise **85,9** seviyesinde gerçekleşmiştir.

Bu oranın **en düşük olduğu ülke İtalya’dır**.

- **2018** yılında bu oran **60,9** iken,
- **2019** yılında **49,8** ile **önemli bir düşüş** göstermiştir.

İtalya’da yargı etkinliği endeksinde yaşanan bu düşüşün, **ülkede yargının ağır bir şekilde işlemesi olarak görülmektedir**.

Şekil 2: Yargının Etkinliği: G-7 Ülkeleri



DEVLETİN BÜTÜNLÜĞÜ

Bir diğer kavram olan devletin bütünlüğü, devlet kurumlarında meydana gelen kayırma, rüşvet gibi sistematik yolsuzluklar vb. alt göstergeleriyle ölçülür. Şekil 3'te G-7 ülkelerinin devlet bütünlüğü oranı sunulmuştur.

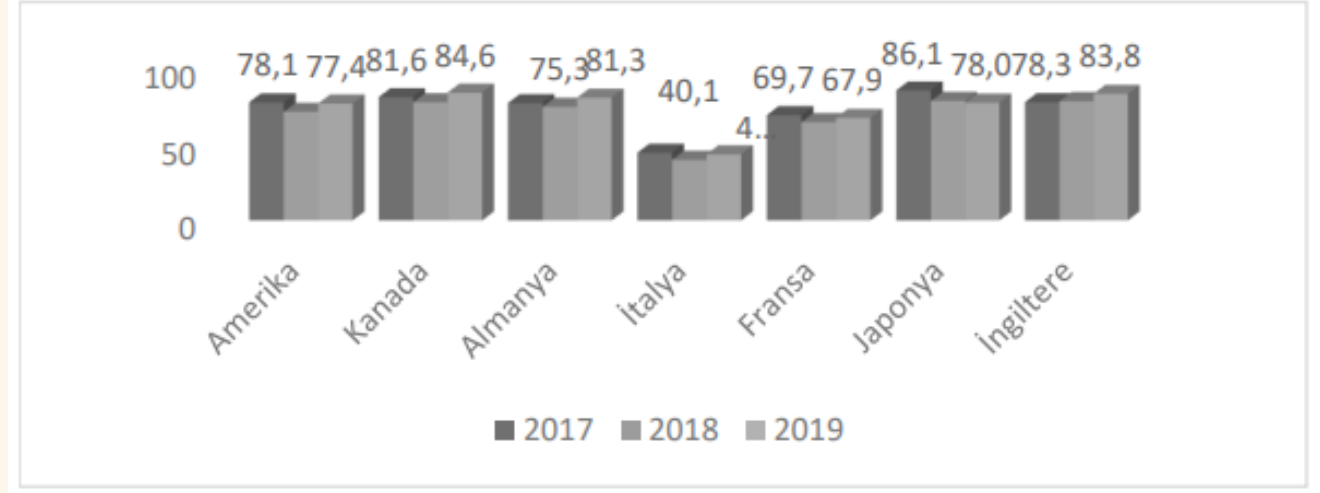
Şekilde diğer ülkeler ile karşılaştırıldığında **en düşük oranın İtalya'ya ait olduğu görülebilir.**

- 2018 yılında 40,1 olan bu oran
- 2019 yılında 43,7 seviyesinde gerçekleşmiştir.

Ekonomik Özgürlük 2019 raporuna göre elde edilen bu sonucun nedeni,

- Euro bölgesinin üçüncü en iyi ekonomisi olan İtalya'nın siyasi müdahaleler, yolsuzluk ve ülkenin kamu maliyesinin zayıf yönetimine maruz kalması olarak açıklanmaktadır.

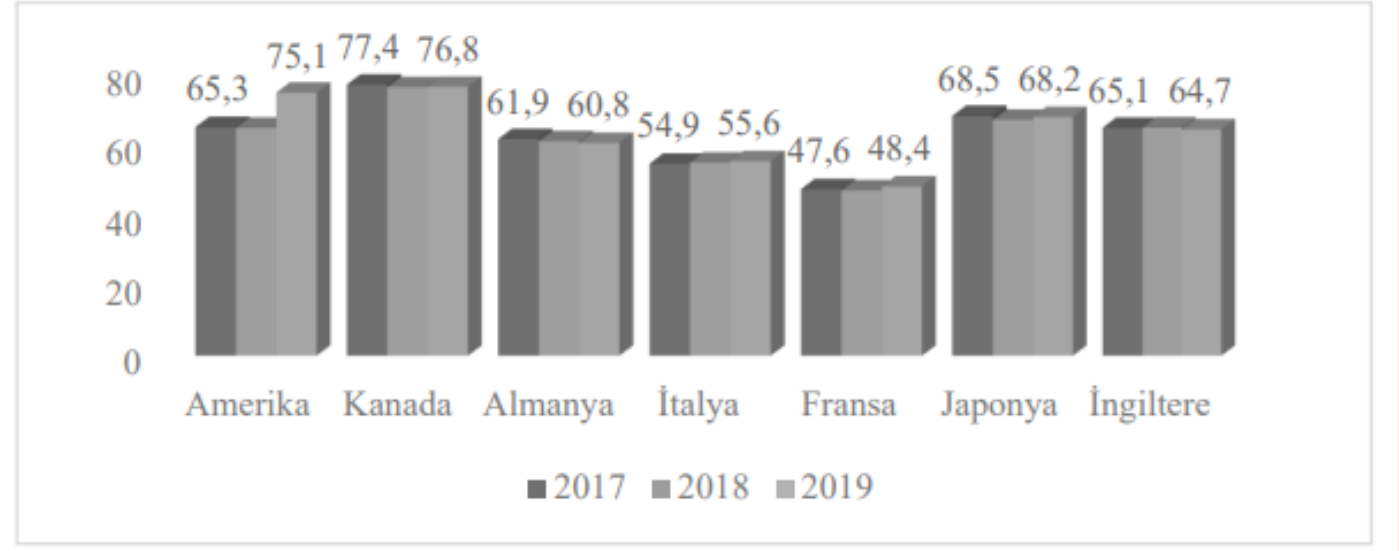
Şekil 3: Devletin Bütünlüğü: G-7 Ülkeleri



VERGİ YÜKÜ

- Bütün hükümetler vergi ve borçlanma yoluyla ekonomik faaliyetlere bir mali yük yüklemektedir. Yüksek vergi oranları, bireylerin ve firmaların piyasada hedeflerine ulaşma yeteneklerini azalttığı gibi özel sektörün faaliyetlerini de kısımtadır. Bu da vergi yükü kavramı ile açıklanır.
- Şekil 4’de ülkelerin vergi yükü incelendiğinde, özellikle Amerika’da 2017’de gerçekleştirilen vergi reformu ile birlikte Amerikan şirketlerine yönelik vergi oranı indiriminin şirketlere kolaylık sağlaması bakımından ekonomik özgürlükler endeksinde **Amerika’nın sıralamasının iyileşmesinde oldukça önemli bir rol oynamıştır.**

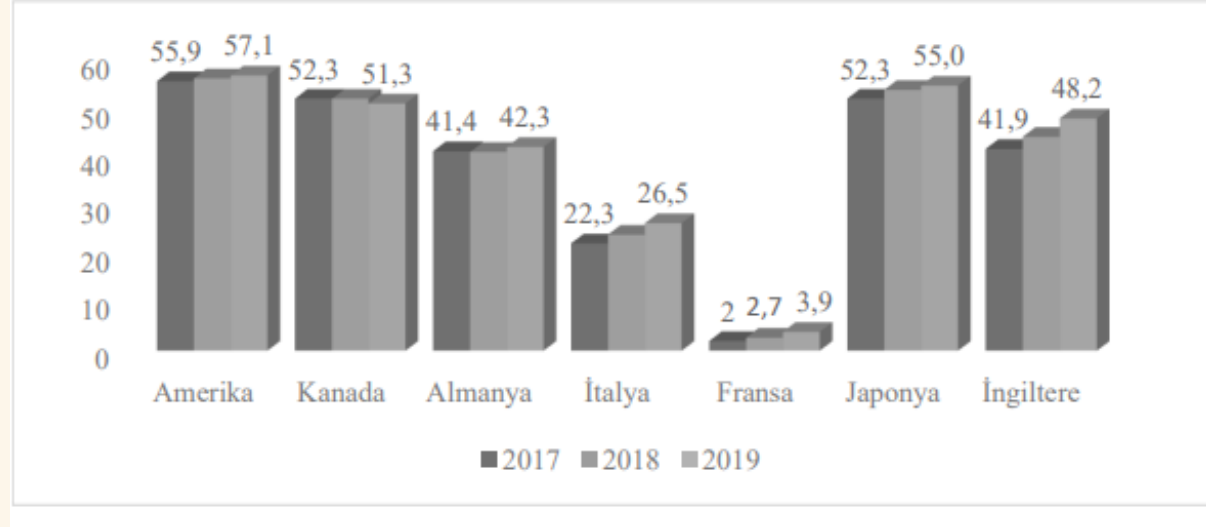
Şekil 4: Vergi Yükü: G-7 Ülkeleri



KAMU HARCAMASI

- Devletin büyüklüğünün açıklandığı bir diğer gösterge kamu harcamalarıdır. **Kamu harcamalarının tümü ekonomik özgürlüğe zarar vermeyebilir.** Örneğin alt yapı çalışmaları için yapılan harcamalar bir yatırım olarak kabul edilebilir. Fakat daha sonra tüm kamu harcamalarının vergilendirme ile finanse edilmesi durumunda fırsat maliyeti ortaya çıkar.
- Bu maliyet, kaynakların özel sektörde kalması durumunda gerçekleşecek olan tüketim ve yatırımın değerine eşittir.
- Ülkelerin kamu harcaması bakımından değerlendirildiği şekil 5'te **dikkat çeken ülke Fransa'dır.** 2017 yılında 2 olan kamu harcaması oranı, 2019'da 3,9 oranında gerçekleşerek diğer ülkeler ile kıyaslandığında **Fransa'nın kamu harcamalarının oldukça fazla olduğu anlamına gelmektedir.** OECD verilerine göre Fransa'nın 2017 yılında kamu harcamasının gayri safi yurtiçi hasıladaki oranı %56,4'tür.

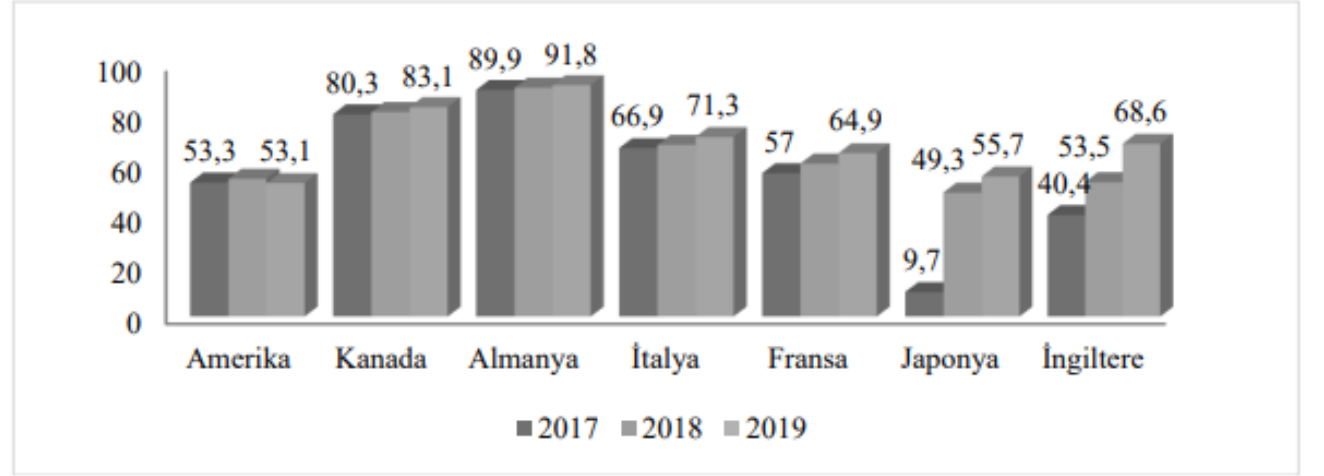
Şekil 5: Kamu Harcaması: G-7 Ülkeleri



MALİ ÖZGÜRLÜK

Devletin mali özgürlüğü, ekonomik özgürlük kavramları içinde ayrıca önem taşır. Çünkü genişleyen bütçe açığı ve artan borç yükü bir ülkenin genel mali özgürlüğünün bozulmasına neden olur. Güçlü mali yapıdan sapmalar meydana geldiğinde makroekonomik istikrar bozulur, ekonomik belirsizlik tetiklenir ve böylece ekonomik özgürlükler sınırlanır.

Şekil 6: Mali Özgürlük: G-7 Ülkeleri

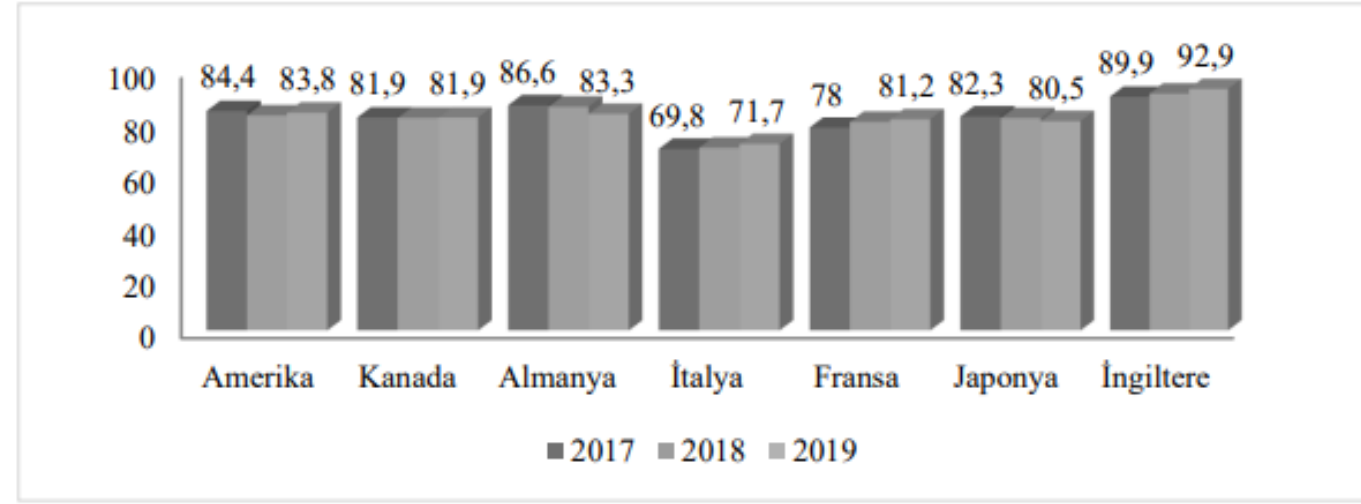


- Şekil 6’da ülkelerin mali özgürlük oranları sunulmuştur. Burada **dikkat çeken ülke Japonya’dır**; çünkü **2017** yılında mali özgürlük oranı 9,7 seviyesinde gerçekleştirmiş olup bu oranın diğer ülkelere kıyasla oldukça düşük olduğu görülür.
- **2018** yılında bu oran önemli bir yükseliş göstererek 49,3 ve 2019 yılında ise 55,7 oranında seviyesinde gerçekleşmiştir.
- **2017** yılında OECD verilerine göre Japonya’nın kamu borcunun gayri safi yurtiçi hasıladaki oranı %234,3 seviyesinde gerçekleşmiştir.
- Aynı oran Almanya için incelendiğinde 2017 yılında mali özgürlük oranı 89,9 seviyesinde gerçekleşmiştir.
- Aynı yıl için Almanya’da kamu borcunun gayri safi yurtiçi hasıladaki oranı 71,5 olup, **Japonya ile kıyaslandığında düşük seviyede olduğu görülebilir.**

İŞ ÖZGÜRLÜĞÜ

- Bir ekonomide bireylerin işletme kurma ve işletme özgürlüklerini oluşturan iş özgürlüğü kavramı ekonomik özgürlük endeksi içinde yer alır.
- Şekil 7’de ülkelerin iş özgürlüğü verileri incelenmiştir. **İş özgürlüğü** oranının **en yüksek olduğu ülke İngiltere** olmakla birlikte, bu oran 2019 yılında 92,9 seviyesindedir.
- **2019 yılı ekonomik özgürlükler raporuna göre İngiltere**, dünyanın en verimli iş ve yatırım ortamlarına sahiptir.
- Bu sebeple, özgür iş ve yatırım ortamında bireylerin girişimcilik faaliyetlerinin oldukça fazla olmasına bağlı olarak iş özgürlüğü de oldukça yüksektir.

Şekil 7: İş Özgürlüğü: G-7 Ülkeleri



İŞGÜCÜ ÖZGÜRLÜĞÜ

Bir ekonomide işgücü özgürlüğünün olması, bireylerin iş bulma fırsatlarına sahip olması anlamına gelir. Aynı şekilde işletmelerin emek istihdam ederken serbestçe sözleşme yapabilmeleri, verimliliği arttırmada ve ekonomik büyümeye katkı sağlamada önem taşır.

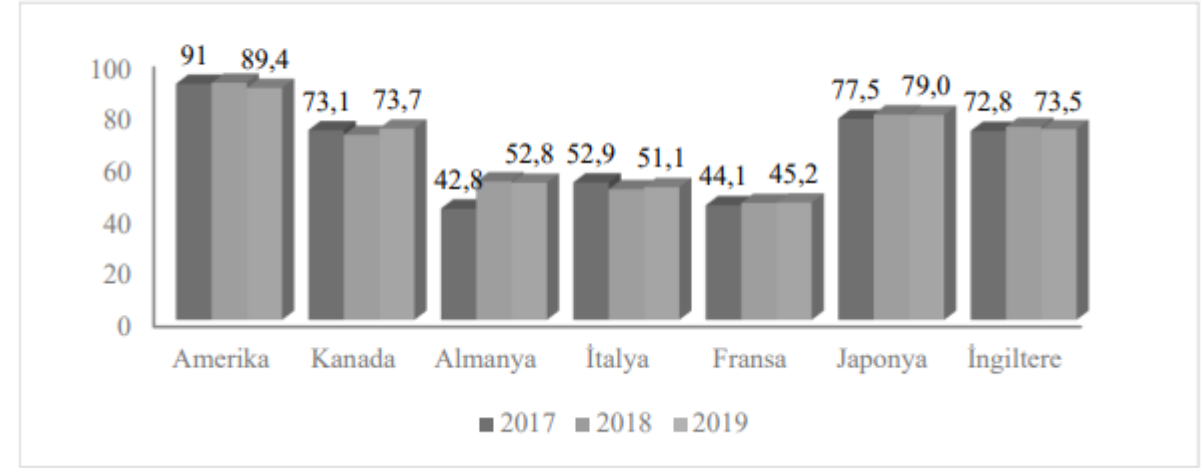
İşgücü özgürlüğünün sunulduğu şekil 8’de;

- İşgücü özgürlüğünün **en yüksek olduğu ülke** 2019 yılında 89,4 ile **Amerika** olduğu görülürken,
- 2019 yılında **en düşük oran** ise işgücü piyasasında katı düzenlemelere maruz kalan **Fransa**’ya aittir.

Amerika’da işgücü özgürlüğünün yüksek olmasının nedeni,

- Zorunlu sendikalaşma konusundaki değişimler olması ve bu değişimin çalışma haklarını genişleterek, işgücü özgürlüğüne pozitif katkı sağlamasıdır.

Şekil 8: İşgücü Özgürlüğü: G-7 Ülkeleri

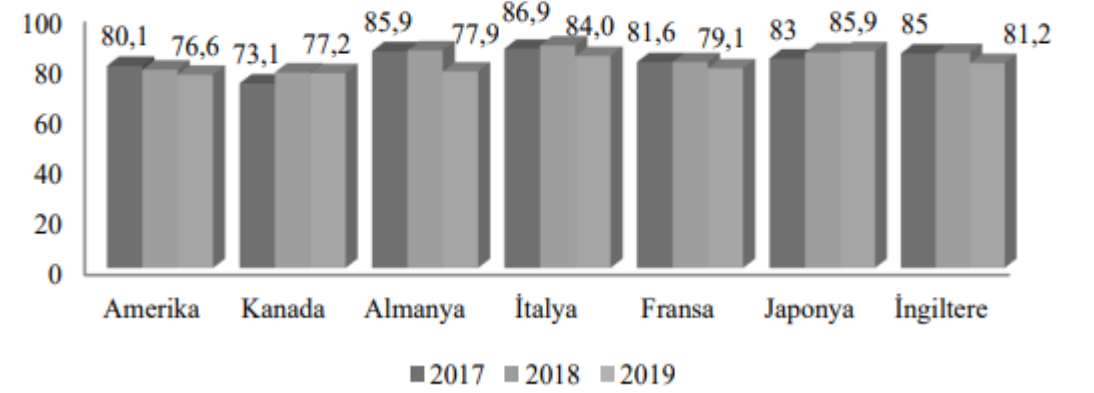


PARASAL ÖZGÜRLÜK

İşgücü özgürlüğünün yanında parasal özgürlük de ekonomik özgürlük kavramında önemli bir faktördür. Parasal özgürlük, istikrarlı bir para birimini ve istikrarlı fiyatı gerektirir. İster tüketici ister girişimci olsun, ekonomik olarak özgür olan bireyler güvenilir para birimine ihtiyaç duyar. Parasal özgürlük olmadan uzun vadeli bir değer yaratmak veya sermaye biriktirmek zordur.

- Şekil 9’da ülkelerin parasal özgürlük oranları verilmiştir. **2019 yılında 85,9 ile Japonya birinci** sırada yer alırken, onu **84 ile İtalya** ve **81,2 ile İngiltere** takip etmektedir.

Şekil 9: Parasal Özgürlük: G-7 Ülkeleri

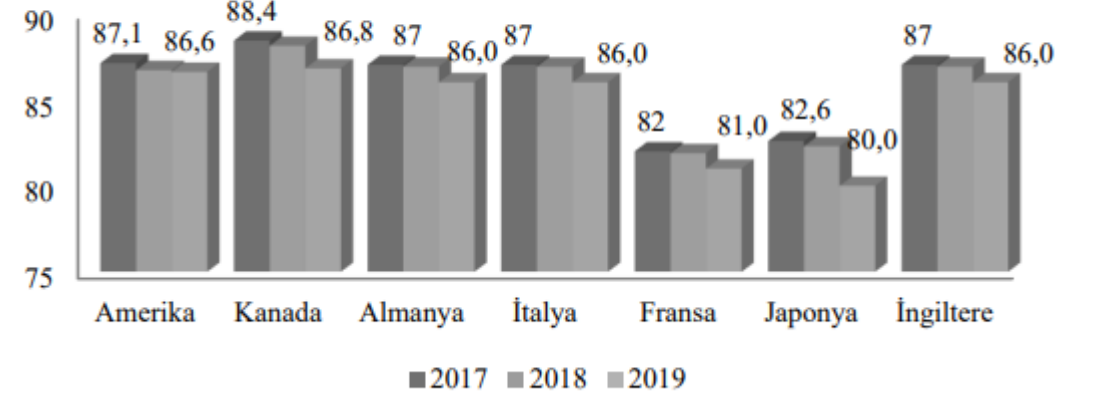


TİCARET ÖZGÜRLÜĞÜ

Bireyler ticaret yapabilme özgürlüğüne sahip olmadıkları takdirde bireylerin ekonomiye katkıları sınırlanır. Birçok hükümet, vatandaşlarının uluslararası pazarda alıcı veya satıcı olarak serbestçe etkileşim kurma hakkına kısıtlamalar getirir. Bunlar tarifeler, ihracat vergileri, ticari kotalar ve doğrudan ticari yasaklar biçiminde olabilmekle birlikte, uygulanan bu kısıtlamalar ekonomik etkinliğe ve ekonomik büyümeye engel olmaktadır.

- Şekil 10'da 2019 yılına bakıldığında **en yüksek oranın 86,8 oranında Kanada'da** gerçekleştiği görülmektedir. Kanada'nın ticaret yapısı incelendiğinde, ihracat ve ithalatın toplam değeri gayri safi yurtiçi hasılanın %64,1'ine eşittir ve ortalama uygulanan tarife oranı %1,6'dır.
- **Ticaret özgürlüğünün yüksek olduğu ülkelerden biri de Almanya'dır.** Almanya'nın ihracat ve ithalatın toplam değeri gayri safi yurtiçi hasılanın %86,9'una eşittir. Ortalama uygulanan tarife oranı ise %2'dir.

Şekil 10: Ticaret Özgürlüğü: G-7 Ülkeleri

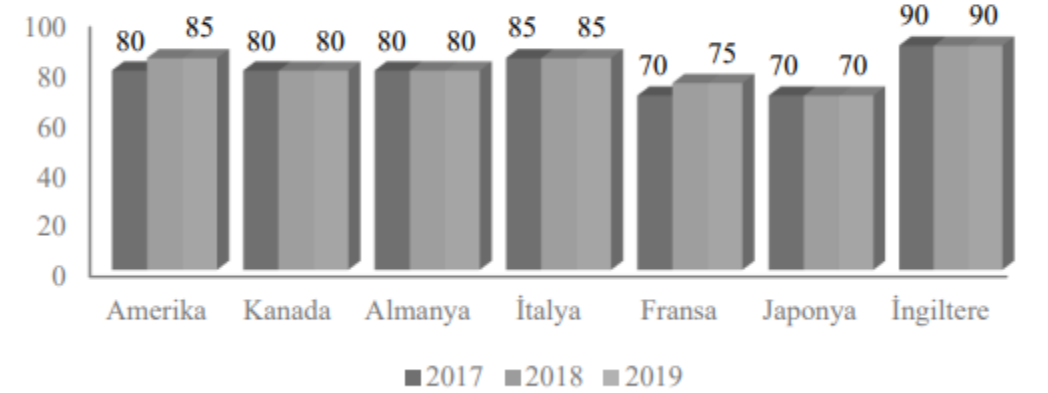


YATIRIM ÖZGÜRLÜĞÜ

Özgür bir yatırım ortamı ise ekonomik faaliyetlerin genişletilmesinde, verimliliğin artırılmasında ve girişimcilik faaliyetlerinin yaygınlaşmasında oldukça önemlidir. Hem ulusal hem de uluslararası sermaye hareketliliğinin kısıtlanması, kaynakların etkin kullanımına engel olurken, ayrıca verimliliği de düşürür.

- Ülkelerin yatırım özgürlüklerinin değerlendirildiği şekil 11’de, **en yüksek oran 90 ile İngiltere’ye aittir.**
- İş özgürlüğünde de bahsedildiği gibi **İngiltere**, dünyanın en verimli iş ve yatırım ortamına sahip olması nedeniyle **ülkede yatırım özgürlüğü oranı yüksektir.**
- Yatırım özgürlüğünün **yüksek olduğu bir diğer ülke İtalya’da** da genel olarak yabancı yatırımlara kısıt konmaması ve ekonominin çoğu sektörünün yatırıma açık olması sebebiyle ülkede özgür bir yatırım ortamı mevcuttur. **Aksine Japonya’da** devlet bazı sektörlerle yatırımları kısıtlamakta, bazı yatırımlar devletin kontrolü altındadır.
- Bu sebeple yatırım özgürlüğünün diğer ülkelerle kıyaslandığında **Japonya’da daha düşük** olduğu görülmektedir.

Şekil 11: Yatırım Özgürlüğü: G-7 Ülkeleri

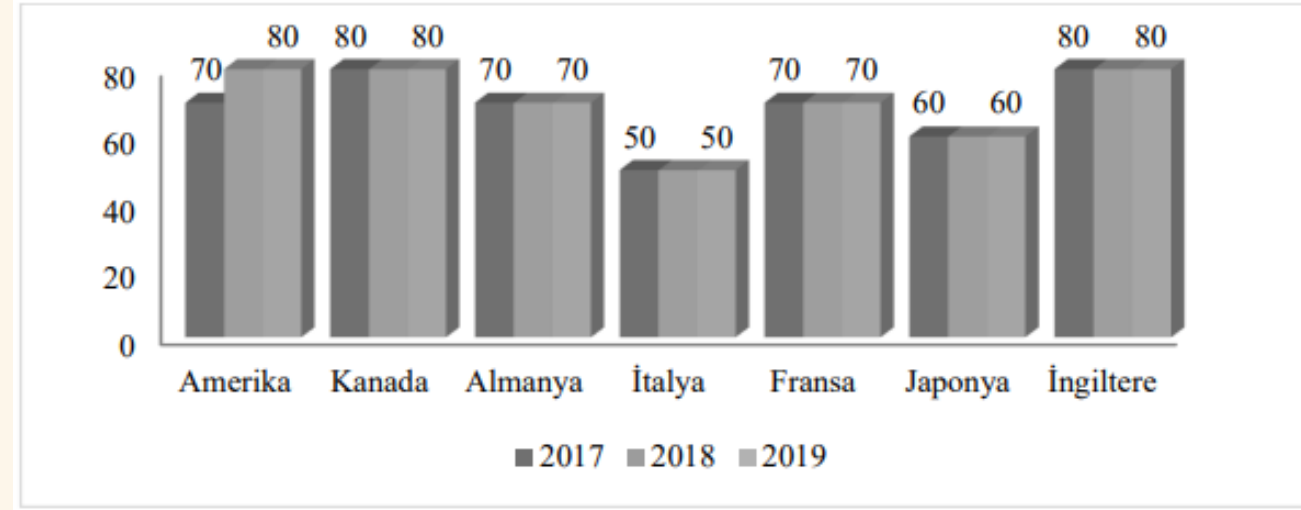


FİNANSAL ÖZGÜRLÜK

Tüm bunların yanında finansal özgürlüğün olmadığı bir ortamda devletin finansal piyasalarda şeffaflık ve dürüstlük güvencesinin önüne geçen bankacılık ve finansal düzenlemeleri verimliliği engelleyebilmekte, girişimcilik faaliyetlerini finanse etme maliyetlerini arttırabilmekte ve rekabeti sınırlayabilmektedir.

- Ülkelerin Finansal özgürlüklerinin verildiği şekil 12’de **İtalya’da finansal özgürlüğün oranının 50 olduğu** görülür.
- Ülkede finansal özgürlüğü etkileyen bankacılık faaliyetlerinden kaynaklı sorunlar yaşanmakta ve geri ödenmeyen kredilerin fazla olması, bu sorunları derinleştiren faktörler olarak görülmektedir.
- **Finansal özgürlüğün yüksek olduğu ülkelerden biri olan İngiltere**, dünyada en gelişmiş finansal sektöre sahiptir. Gayri safi yurtiçi hasıla içinde özellikle bankacılık ve sigorta hizmetleri önemli kalemleri oluşturur.

Şekil 12: Finansal Özgürlük: G-7 Ülkeleri



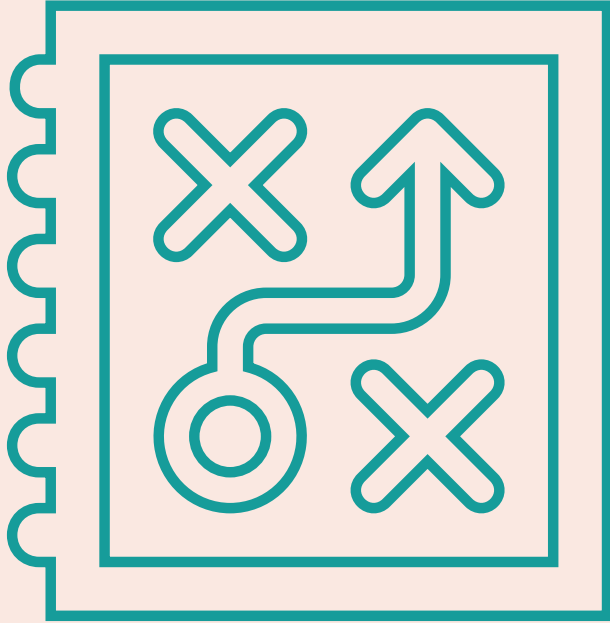
KULLANILAN ANALİZ YÖNTEMİ

- Hiyerarşik Kümeleme yönteminde kullanılan teknikler, tek bağlantı tekniği, ortalama bağlantı tekniği ve ward's tekniğidir.
- Hiyerarşik olmayan yöntemler ise bu çalışmada kullanılan K-Ortalamalar yöntemi ve en çok olabilirlik yöntemidir.

K-ORTALAMA YÖNTEMİ NEYDİ?

- K-Ortalamalar yöntemi, n tane veriyi k adet kümeye ayıştırmak için, ilk başta n adet veriden rastlantısal şekilde k adet veri seçer ($k < n$).
- Bu verilerden her biri, bir kümenin merkezini ya da orta noktasını temsil eder.
- Tüm veriler, seçilen k adet merkeze olan uzaklıkları baz alınarak en yakın olan kümeye dağılır.
- Diğer adımda her küme için ortalama hesaplanır ve hesaplanan bu değer o kümenin yeni merkezi olur.
- Bu yöntem merkez noktanın kümeyi temsil etmesi ana fikrine dayanır.

K- ORTALAMALAR YÖNTEMİ



- **K-Ortalamlar Kümeleme yönteminin adımları aşağıdaki gibidir.**
- **Adım 1:** Ayırıştırılmak istenen küme sayısını (k) belirle,
- **Adım 2:** Rastlantısal olarak k adet başlangıç küme merkezi belirle,
- **Adım 3:** Her veri ile Adım 2’de belirlenen başlangıç küme merkezlerinin uzaklıklarını hesapla ve her veriyi uzaklığı en yakın olan kümeye ata,
- **Adım 4:** Tüm veriler atandığında, oluşan kümelerde yer alan veriler yardımıyla k adet merkezi yeniden hesapla,
- **Adım 5:** Adım 3 ve Adım 4’ü kümede yer alan veriler sabitleninceye kadar tekrarla.

VERİ SETİ

Çalışmada yer alan G-7 ülkelerine ait 12 değerlendirme faktörünün 2019 yılı için değerleri aşağıdaki gibidir.

Tablo 3: Veri Seti: 2019

	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀	K ₁₁	K ₁₂
Amerika	79,3	78,6	77,4	75,1	57,1	53,1	83,8	89,4	76,6	86,6	85	80
Kanada	87,0	69,4	84,6	76,8	51,3	83,1	81,9	73,7	77,2	86,8	80	80
Almanya	79,9	75,4	81,3	60,8	42,3	91,8	83,3	52,8	77,9	86,0	80	70
İtalya	71,7	49,8	43,7	55,6	26,5	71,3	71,7	51,1	84,0	86,0	85	50
Fransa	82,5	66,1	67,9	48,4	3,9	64,9	81,2	45,2	79,1	81,0	75	70
Japonya	84,1	68,5	78,0	68,2	55,0	55,7	80,5	79,0	85,9	80,0	70	60
İngiltere	92,3	85,9	83,8	64,7	48,2	68,6	92,9	73,5	81,2	86,0	90	80

Yöntemin ikinci adımda belirlenmesi gereken **başlangıç küme merkezleri**, rastlantısal olarak veri seti içinden ve/veya veri seti dışından seçilebilir.

Bu çalışmada ise yöntem, veri seti içinden seçilebilecek olası tüm başlangıç merkez verilerle uygulanacak ve oluşan küme grupları incelenecektir.

Buna ek olarak veri seti dışından başlangıç küme merkezleri ile farklı ve/veya yeni küme gruplarının oluşup oluşmadığını incelemek adına, diğer sayfadaki **algoritma yardımıyla veri seti dışından yeni veriler oluşturulacaktır.**

VERİ SETİ DIŞINDAN ELDE EDİLEN YENİ VERİLER

Adım 1: Tüm verilerin birbirine olan uzaklıkları hesaplanır,

Adım 2: Veri sayısı (n) dörde bölünerek elde edilen değer yukarı yuvarlanır,

Adım 3: Adım 2'den elde edilen sayının m olduğu durum için:

- Birbirine en uzak nesneler belirlenir. Bu iki nesneye uzaklık değerinin n'de biri eklenerek veri setinde yer almayan ve yoğunluğun olmadığı düzlemde yeni nesneler bulunur. Aynı işlem ikinci, üçüncü, ..., ve m. en uzak olan nesnelere de uygulanır. Böylelikle 2m tane veri seti dışından yeni nesne bulunur,

Adım 4: Adım 3'ün aynısı birbirine en yakın olan nesneler için yapılır, fakat burada yeni nesnelere uzaklık değerinin n'de 1'i çıkartılarak bulunur.

Bu yöntemle elde edilen aynı zamanda veri seti içinde yer almayan ve başlangıç merkez veri olarak kullanılacak yeni veriler yan taraftadır.

Tablo 4: Veri Seti Dışından Elde Edilen Yeni Veriler: 2019

K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀	K ₁₁	K ₁₂
90,58	89,88	88,68	86,38	68,38	64,38	95,08	100,68	87,88	97,88	96,28	91,28
82,98	61,08	54,98	66,88	37,78	82,58	82,98	62,38	95,28	97,28	96,28	61,28
90,49	89,79	88,59	86,29	68,29	64,29	94,99	100,59	87,79	97,79	96,19	91,19
93,69	77,29	79,09	59,59	15,09	76,09	92,39	56,39	90,29	92,19	86,19	81,19
82,70	65,10	80,30	72,50	47,00	78,80	77,60	69,40	72,90	82,50	75,70	75,70
88,00	81,60	79,50	60,40	43,90	64,30	88,60	69,20	76,90	81,70	85,70	75,70
82,37	64,77	79,97	72,17	46,67	78,47	77,27	69,07	72,57	82,17	75,37	75,37
75,27	70,77	76,67	56,17	37,67	87,17	78,67	48,17	73,27	81,37	75,37	65,37

Sadece mevcut ülke verilerini değil, aynı zamanda uzaklık analizlerine göre oluşturulan 8 yeni sentetik veriyle de K-Ortalamalar analizini gerçekleştirildi. Bu veriler:
En uzak ülke çiftlerinden: + uzaklık eklenerek
En yakın ülke çiftlerinden: – uzaklık çıkarılarak hesaplandı.
Böylece başlangıç merkezlerinin konumuna duyarlılığı test ettik.

BAŞLANGIÇ MERKEZLERİNİ RASTGELE SEÇELİM

(Veri seti içinden)

Örnek olarak:

Merkez 1 = Amerika (ABD)

Merkez 2 = İtalya

HER ÜLKE İÇİN UZAKLIKLARI HESAPLAYALIM

Kanada'nın K1–K12 değerleri:

87.0 69.4 84.6 76.8 51.3 83.1 81.9 73.7 77.2 86.8 80 80

a) Kanada – ABD uzaklığı:

$$D = \sqrt{(87.0 - 79.3)^2 + (69.4 - 78.6)^2 + \dots + (80 - 80)^2}$$

Kısa yol: her farkın karesini alıp toplayalım:

$$(7.7)^2 + (-9.2)^2 + (7.2)^2 + (1.7)^2 + (-5.8)^2 + (30)^2 + (-1.9)^2 + (-15.7)^2 + (0.6)^2 + (0.2)^2 + (-5)^2 + (0)^2$$

$$= 59.29 + 84.64 + 51.84 + 2.89 + 33.64 + 900 + 3.61 + 246.49 + 0.36 + 0.04 + 25 + 0 = 1407.8$$

$$D_{Kanada-ABD} \approx \sqrt{1407.8} \approx 37.52$$

b) Kanada – İtalya uzaklığı:

Aynı işlemleri İtalya ile de yaparız. Sonuç:

$$D_{Kanada-İtalya} \approx 49.89$$

Sonuç: Kanada, ABD'ye daha yakın → Küme 1'e atanır.

YENİ KÜME MERKEZLERİ HESAPLANIR

Her kümedeki ülkelerin K1–K12 değerlerinin ortalaması alınır:

Örnek: Küme 1 (ABD, Kanada, Japonya, İngiltere)

Yeni merkez:

$$K1 = (79.3 + 87.0 + 84.1 + 92.3)/4 = 85.675$$

$$K2 = (78.6 + 69.4 + 68.5 + 85.9)/4 = 75.6$$

... ve böyle devam eder.

Bu ortalama 12 faktör için yapılır.

UZAKLIKLAR TEKRAR HESAPLANIR → Yeni kümeler belirlenir

Yeni merkezlere göre tüm ülkelerin uzaklıkları tekrar hesaplanır, atamalar yeniden yapılır. Bu işlem merkezler sabitlenene kadar sürer (genelde 2–4 iterasyonda durur).

Bu süreci makale **49 kez** (21 içeriden, 28 dışarıdan) tekrarlamıştır.

En sık çıkan küme yapısı Tablo 5’ te gösterilmiştir.

KÜME SAYISININ BELİRLENMESİ

- Kümeleme analizinde, kullanılacak Kümeleme yöntemi her ne olursa olsun, küme sayısının belirlenmesi önemli bir sorunsaldır, ancak veri setinin az olduğu durumlarda veri setinin yarısının karekök değeri kadar kümeye ayırtırmak mümkündür.
- Olası tüm **veri seti içinden** başlangıç merkezlerle **21 kere** K-Ortalamlar tekniği uygulandığında, **4 farklı küme** grubunun elde edildiği gözlemlenmiştir.
- Diğer teknikle **veri seti dışından** başlangıç merkez verisi olarak kullanılmak üzere elde edilen yeni 8 veri ile analiz **28 kere** uygulandığında ise **6 farklı küme** grubunun elde edildiği gözlemlenmiştir.
- Tablo 5’de ise oluşan tüm küme grupları ile birlikte görülme sıklıkları yer almaktadır.

Tablo 5: K-Ortalamlar Yöntem Bulguları: 2019

No	Küme X	Küme Y	Sıklık-İçerden	Sıklık-Dışardan	Sıklık-İçerden ve Dışardan
1	Amerika, Kanada, Japonya, İngiltere	Almanya, İtalya, Fransa	8	14	27
2	Amerika, Kanada, Almanya, Japonya, İngiltere	İtalya, Fransa	11	8	20
3	Amerika, Fransa, İngiltere	Kanada, Almanya, İtalya, Japonya	1	1	0
4	Amerika, İtalya, Japonya	Kanada, Almanya, Fransa, İngiltere	1	1	1
5	Amerika, Kanada, İtalya, Japonya	Almanya, Fransa, İngiltere	0	1	0
6	Amerika, Kanada, Almanya, İtalya, Fransa, Japonya, İngiltere		0	3	8

5 ve 6 numaralı küme grupları sadece veri seti dışından başlangıç merkez verilerle elde edilmiştir.

6 numaralı küme grubu ise tüm ülkelerin benzer olduğu tek bir kümenin elde edildiği gruptur.

Oluşan 6 farklı küme grubunun 5’inde Amerika ve Japonya’nın aynı kümede yer aldığı gözlemlenirken, sadece 3 numaralı küme grubunda Amerika ve Japonya ayrı kümelerde yer almıştır.

Ülkelerin aynı kümede yer alması, ülkelerin 12 faktör açısından birbirine benzediği, aynı zamanda diğer kümede yer alan ülkelerle bu açıdan farklılaştığını gösterir.

- En sık gözlemlenen küme grupları ise 1 ve 2 numaralı küme gruplarıdır.

Bu iki küme grubunun farklılaştığı tek ülke ise Almanya’dır. Veri seti dışından başlangıç merkez verilerle en sık elde edilen küme grubunda Almanya, İtalya ve Fransa aynı kümede yer alırken,

Veri seti içinden başlangıç merkez verilerle en sık elde edilen küme grubunda ise İtalya ve Fransa aynı küme grubunda yer almaktadır.

Tablo 5: K-Ortalamalar Yöntem Bulguları: 2019

No	Küme X	Küme Y	Sıklık-İçerden	Sıklık-Dışardan	Sıklık-İçerden ve Dışardan
1	Amerika, Kanada, Japonya, İngiltere	Almanya, İtalya, Fransa	8	14	27
2	Amerika, Kanada, Almanya, Japonya, İngiltere	İtalya, Fransa	11	8	20
3	Amerika, Fransa, İngiltere	Kanada, Almanya, İtalya, Japonya	1	1	0
4	Amerika, İtalya, Japonya	Kanada, Almanya, Fransa, İngiltere	1	1	1
5	Amerika, Kanada, İtalya, Japonya	Almanya, Fransa, İngiltere	0	1	0
6	Amerika, Kanada, Almanya, İtalya, Fransa, Japonya, İngiltere		0	3	8

K-ORTALAMALAR YÖNTEM BULGULARI: 2018

- Aynı analiz **2018 yılı için** gerçekleştirildiğinde oluşan **7 farklı küme** grubu ile birlikte bu küme gruplarının görülme sıklıkları Tablo 6’da sunulmuştur.
- 2018 yılı için veri seti içinden ve veri seti dışından başlangıç merkez verilerle **en sık elde edilen küme grubu 1 numaralı küme grubu** olmakla birlikte, bu küme grubunda **Almanya, İtalya ve Fransa** aynı kümede yer almıştır.
- 2018 yılında oluşan tüm küme gruplarında **Amerika ve Japonya’nın** aynı kümede yer aldığı gözlemlenmiştir.

Tablo 6: K-Ortalamalar Yöntem Bulguları: 2018

No	Küme X	Küme Y	Sıklık-İçerden	Sıklık-Dışardan	Sıklık-İçerden ve Dışardan
1	Amerika, Kanada, Japonya, İngiltere	Almanya, İtalya, Fransa	9	20	29
2	Amerika, Kanada, İtalya, Japonya	Almanya, Fransa, İngiltere	1	2	0
3	Amerika, Kanada, Almanya, Japonya, İngiltere	İtalya, Fransa	8	3	19
4	Amerika, İtalya, Fransa, Japonya, İngiltere	Kanada, Almanya	1	0	0
5	Amerika, İtalya, Japonya	Kanada, Almanya, Fransa, İngiltere	2	0	1
6	Amerika, Kanada, Almanya, İtalya, Japonya	Fransa, İngiltere	0	2	1
7	Amerika, Kanada, Almanya, İtalya, Fransa, Japonya, İngiltere		0	1	6

K-ORTALAMALAR YÖNTEM BULGULARI: 2017

- Aynı analiz **2017 yılı** için gerçekleştirildiğinde oluşan **12 farklı küme grubu** ile birlikte bu küme gruplarının görülme sıklıkları Tablo 7’de sunulmuştur.
- Veri seti içinden ve veri seti dışından** başlangıç merkez verilerle **en sık gözlemlenen küme grubu** ise **1 numaralı** küme grubudur.
- Veri seti dışından başlangıç merkez verilerle en sık elde edilen** küme grubunda Kanada, Almanya, İtalya, Fransa aynı kümede yer alır.
- 2017 yılında oluşan 12 farklı küme grubundan 6 tanesinde Amerika ve Japonya’nın** aynı kümede yer aldığı gözlemlenmiştir.

Tablo 7: K-Ortalamalar Yöntem Bulguları: 2017

No	Küme X	Küme Y	Sıklık-İçerden	Sıklık-Dışardan	Sıklık-İçerden ve Dışardan
1	Amerika, Japonya, İngiltere	Kanada, Almanya, İtalya, Fransa	6	9	12
2	Amerika, Kanada, Japonya, İngiltere	Almanya, İtalya, Fransa	6	2	18
3	Amerika, Kanada, Almanya, İtalya, Fransa, İngiltere	Japonya	3	4	7
4	Amerika, Kanada, Almanya, İtalya	Fransa, Japonya, İngiltere	1	3	0
5	Amerika, Kanada, Almanya, Japonya, İngiltere	İtalya, Fransa	3	1	4
6	Amerika, Kanada, Almanya, İngiltere	İtalya, Fransa, Japonya	2	1	4
7	Amerika, İtalya, Fransa, Japonya, İngiltere	Kanada, Almanya	0	2	3
8	Amerika, Kanada, İtalya, Fransa, Japonya, İngiltere	Almanya	0	2	2
9	Amerika, Kanada	Almanya, İtalya, Fransa, Japonya, İngiltere	0	0	1
10	Amerika, Kanada, Almanya	İtalya, Fransa, Japonya, İngiltere	0	0	2
11	Amerika, Kanada, Almanya, İtalya, İngiltere	Fransa, Japonya	0	0	1
12	Amerika, Kanada, Almanya, İtalya, Fransa, Japonya, İngiltere		0	4	2

SONUÇ

- 2019 yılında oluşan bulgular değerlendirildiğinde, oluşan **6 farklı küme grubunun 5'inde Amerika ve Japonya'nın aynı kümede** yer aldığı görülmüştür.
- Amerika ve Japonya'nın aynı kümelerde yer alması ekonomik özgürlükler endeksinde yer alan 12 faktör bakımından **bu iki ülkenin birbirine benzerlikler gösterdiği anlamına gelmektedir.**
- **2018** yılında **Amerika ve Japonya'nın tüm küme gruplarında birlikte** yer aldığı görülürken, **2017** yılında **içerden ve dışardan sıklıkların** en yüksek olduğu kümeler olan 1 ve 2 numaralı kümelerde aynı küme grubunda yer almışlardır.
- Aynı şekilde 2019 yılında **Amerika ve Kanada'nın** içerden ve dışardan sıklıkların en yüksek olduğu 1 ve 2 numaralı küme grubu ile birlikte toplam 4 kümede birlikte yer alırken, **2018 yılında 5 kümede** ve **2017 yılında ise toplam 10 kümede** yer alarak ekonomik özgürlükler bakımından birbirine benzedikleri görülmüştür. Bu iki ülkenin ekonomik anlamda dünyadaki yerleri incelendiğinde, dünyadaki en büyük iki ülke ekonomisi olduğu belirlenmiştir.
- **İçerden ve dışardan sıklık derecelerinin en yüksek** olduğu kümelerde yer alan **Amerika ve Kanada** ile birlikte yer alan **bir diğer** ülkenin **İngiltere** olduğu görülür.
- **2019** ve **2017** yıllarında küme 1 ve 2, **2018** yılında ise küme 1'de yer almıştır.



KAYNAKÇA

- Coşar & Özarı (2019) – G7 Ülkeleri Kümeleme Makalesi
- Özarı vd. (2000) – Başlangıç Merkezi Oluşturma Yöntemi
- Jain (2010) – K-Means Kümeleme Tarihçesi
- Tatlıdil (1992) – Çok Değişkenli İstatistik
- Han, Kamber & Pei (2011) – Veri Madenciliği Kitabı



Teşekkürler...