

Veri Analitiği – Kümeleme Proje Ödevi

UNICEF 167 ülke için elde ettiği bazı veriler sayesinde ülkeleri az gelişmiş, gelişmekte ve gelişmiş ülkeler olarak kümeleyerek az gelişmiş ülkelerdeki çocuklara yardım etmek istemektedir. Bu anlamda eldeki veride 10 farklı bilgiye sahiptir. Bunlar;

- **Ulke:** Ülkenin adı.
- **Cocuk_olumu:** 1000 canlı doğumda 5 yaş altı çocuk ölümü sayısı.
- **Ihracat:** Kişi başına mal ve hizmet ihracatı.
- **Saglik:** Kişi başına toplam sağlık harcaması.
- **lthalat:** Kişi başına mal ve hizmet ithalatı.
- **Gelir:** Kişi başı net gelir.
- **Enflasyon:** Toplam GSYİH'nın yıllık büyüme hızının ölçümü.
- **Beklenen_yasam:** Mevcut ölüm oranları aynı kalacaksa, yeni doğan bir çocuğun yaşayacağı ortalama yıl sayısı.
- **Dogum_sayisi:** Mevcut yaş-doğurganlık oranları aynı kalırsa her kadının doğuracağı çocuk sayısı.
- **GSYİH:** Kişi başına düşen GSYİH.



Şekil 1. Ülkeler

Verilen bilgiler ışığında aşağıdaki 8 aşamayı gerçekleştiriniz.

- 1) Kullanacağınız modülleri ve veri setini yükleyiniz. Veri setinde bulunan sütun isimlerini yazdırınız. (5 Puan)
- 2) Veri setindeki sütunlarda kayıp/eksik veri var mı kontrol ediniz ve her sütun için kayıp eksik veri sayısını yazdırınız. (10 Puan)
- 3) Kayıp/eksik veri varsa bunları yazdırınız ve veri setinden siliniz. (10 Puan)

- 4) Veri setinde tekrar eden ülke ismi var mı kontrol ediniz ve tekrar eden ülke sayısını yazdırınız. (10 Puan)
- 5) Tekrar eden ülke varsa bu ülkeler ile ilgili satırları yazdırınız. Daha sonra ilk ülke satırı hariç geri kalan tekrar eden ülke satırlarını veri setinden siliniz. (10 Puan)
- 6) Z-skor normalizasyonu ile ülke verilerini standartlaştırınız. (10 Puan)
- 7) Normalizasyon yapılan veri ile ülkeleri 3 kategoride K-means algoritması ile kümeleyiniz ve küme etiketlerini orjinal veri setine "Kumesi" başlıklı sütun olarak ekleyiniz. (20 Puan)
- 8) Kümeleme sonucuna göre UNICEF tarafından desteklenmesine uygun gördüğünüz az gelişmiş ülkeleri alfabetik sırada yazdırınız ve toplam az gelişmiş ülke sayısını yazdırınız. (25 Puan)

Ülkeler ile ilgili veri seti dosyası:

<https://drive.google.com/file/d/1aloHYUKHcHDcrAK4FtyL3FrE8Q0f9WjX/view?usp=sharing>

Raporda İstenenler ve Rapor Formatı

1. Projeler Jupyter platformu üzerinden hazırlanmalıdır ve gruptaki bir öğrencinin numarası ile öğrenci_numaranız.jpynb şeklinde kaydedilerek <https://forms.gle/Qe9UpZstYvGYV8h88> linkindeki form üzerinden teslim edilmelidir. Başka yollarla üretilip gönderilen projeler geçersiz sayılacaktır.
2. Her aşamayı ayrı bir kod bloğu içerisinde kodlayınız ve yazdırınız. Toplamda 8 kod bloğu olmalıdır. Yazdığınız her satırdaki kodun yanına kodun işlevini (açıklamasını) # yorum işareti koyarak yazınız. Açıklaması bulunmayan kod satırı geçersiz sayılacaktır. Aşağıda örnek Jupyter not defteri şekli ve kodu gösterilmiştir.
3. Daha önce belirttiğiniz proje tipine göre ilgili projeyi çözünüz. Farklı proje tipini çözen öğrencilerin projeleri dikkate alınmayacaktır.
4. Oluşturduğunuz proje grubu için tek bir e-posta adresi belirleyerek sorularınızı bu e-posta adresi ile merhaday@erciyes.edu.tr adresine gönderebilir veya bu ders için belirlenmiş ofis saatlerinde Arş. Gör. Merhad Ay'a ofisinden ulaşabilirsiniz.
5. Proje son teslim tarihi **6 Haziran 2022 Pazartesi günü saat 12:00'**dir. Sonraki saat ve tarihlerde gönderilen proje raporları kabul edilmeyecektir. 6 Haziran 2022 Pazartesi günü saat 14:00'da bölüm web sayfamızda projeleri teslim alınan öğrencilerin listesi ilan edilecektir.

İPUCU: Kullanabileceğiniz modüller ve fonksiyonlardan bazıları aşağıdaki gibidir. Bunların dışındaki modülleri ve fonksiyonları da kullanabilirsiniz.

Modüller: pandas, sklearn.cluster-Kmeans, sklearn.preprocessing-StandardScaler

Fonksiyonlar: sort_values(), sum(), isnull(), any(axis), duplicated(), drop_duplicates(keep), dropna(), fit_transform()

Önemli Not: Proje ödevinin farklı mecralarda tespit edilmesi veya benzerlik durumlarında projeniz geçersiz sayılacaktır.

ÖRNEK JUPYTER DOSYASI

In [1]:

```
1 import pandas as pd # pandas modülü pd kısaltması ile yüklenmiştir
2
3
4
5
```

Yorum işareti ile her
kod satırının açıklaması

Yazdırılan çıktı

Index(['Ulke', 'Cocuk_olumu', 'Ihracat', 'Saglik', 'ithalat', 'Gelir',

In [2]:

1

Kod bloğu

In [3]:

1

2

