

İKT442 Veri Madenciliği Vize Sınavı Soruları 06 Nisan 2022

Aşağıdaki soruların çözümlerini (kullanılan kodlar ve işlem sonucunun ekran kaydı) en geç 17 Nisan 2022 tarihine kadar Teams üzerinden sınav sonuçlarının yüklenmesi için özel olarak hazırlanmış forma yükleyiniz... (Olası sorun durumunda benimle iletişime geçerek mail adresime de gönderebilirsiniz).

1. Dünya kalkınma göstergelerinden (World Development Indicators- WDI) veya Penn World Table (PWT 10.0) veri setinden ülkelerin dahil oldukları gelir grubuyla ilgili olduğunu düşündüğünüz 2 değişken belirleyiniz... Kısa bir yorum yazınız... **(10 puan)**
2. Öncelikle bu değişkenlerin ayrı ayrı ortalamasına, medyanına, varyansına ve aralığına (range) bakınız... **(10 puan)**
3. Bu değişkenlerin değerlerine sadece sizin seçeceğiniz bir ülke için de bakınız (ortalamasına, medyanına, varyansına ve aralığına (range))... **(10 puan)**
4. Tek bir ülke için bu iki değişkeni x ekseninde yıllar ve y ekseninde bu değişkenler ayrı ayrı olacak şekilde sütun grafiği (bar chart) ile gösteriniz...**(10 puan)**
5. Bu iki değişken arasındaki kovaryans ve korelasyon nedir? **(10 puan)**
6. Bu iki değişkeni yıllara göre derleyerek, tüm dünyada bu yıllar için bu değişkenlerin ortalamasını ve toplamını sütun grafiği (bar chart) ile gösteriniz. **(15 puan)**
7. Bu iki değişkenin olduğu veri seti ile ülkelerin hangi gelir grubuna dahil olduğunu gösteren veri setini eşleştirerek (tek bir yıl, tüm ülkeler ve iki değişken); **(30 puan)**
 - a. Bu iki değişkeni bir ele aldığınızda kaç adet küme (cluster) mantıklı olabilir ve her küme merkezleri ile her kümedeki eleman sayısı nedir?
 - b. Gelir gruplarını kategorik değişken olarak ele aldığınızda, gelir grubu kadar kümeye verinizi böldürseniz, gelir gruplarına göre verinin kategorizasyonu ile kıyasladığınızda ne kadar uymaktadır. Yorumlayınız...
 - c. Bu iki değişkenin kümelere ayrıldığında elemanlarının dağılımı ve küme merkezlerinin gösterimini grafik üzerinde yapınız...
8. Bu derste öğrendiklerinizi dikkate alırsak, şunu da yapabilirsem dediğiniz bir şey yazınız... **(5 puan)**

Not: Eğer WDI veri setini kullanırsanız, veri setinin işlenmesiyle ilgili ekstra yapılacak işlemler de dikkate alarak **+10 puan** daha değerlendirmeye katılacaktır.