

Veri Duzenleme Odev

Kubra Atalay Kabasakal

04 12 2021

Değerli Öğrenciler, aşağıda yer alan soruları istediğiniz R paketlerini kullanarak cevaplandırabilirsiniz. Alternatif çözümleri bulmak adına tüm çözüm yollarını da ele alabilirsiniz. Odevde kullandığımız R paketlerini lütfen komut satırlarının başına ekleyiniz.

```
library(dplyr)
library(data.table)
```

- İlk 9 madde “DATA2.csv” veri seti,
 - 10.-15. maddeler arasında ggplot2 paketinde yer alan diamonds veri seti ile
 - 16.-21. maddeler arasında ise DT veri setini üzerinden cevaplayınız.
1. “DATA2.csv” dosyasını fread() fonksiyonu ile R ortamına df adı ile aktarınız. Oluşan nesnenin türünü sorgulayınız.
 2. Bu uygulamaya farklı şehirlerde yer alan öğrenciler katılmıştır. Veri setinde ISTANBUL, KONYA ve VAN’dan katılan kaç öğrenci olduğunu farklı kod satırları ile bulunuz.
 3. Uygulamaya Konya’dan katılan öğrencilerin Okuma tutum ölçeğine ait cevaplarından bir veri seti oluşturunuz. Okuma tutum maddeleri 11 tane olup OKT olarak isimlendirilmiştir.
 4. Uygulamaya katılan öğrencilerin cinsiyete göre dağılımı nasıldır?
 5. Uygulamaya katılan öğrencilerin bölgelere göre dağılımı nasıldır?
 6. Her bir bölgede uygulamaya katılan öğrencilerin TurkcePuan değişkeninin ortalamasını hesaplayıp, ortalamaları büyükten küçüğe sıralayınız.
 7. df veri Setini atama operatörü kullanmadan TurkcePuan değişkenine göre büyükten küçüğe sıralayınız. Veri setinin ilk üç satırını yazdırınız.
 8. df veri Setini atama operatörü kullanmadan Bolge değişkenine göre alfabetik olarak sıralayınız. Oluşan veri setinden sadece MARMARA bölgesinin olduğu satırlarını Bolge değişkeni/sütun adını kullanmadan filtreleyiniz.
 9. Veri setinden Ege bölgesinde bulunan Şehirleri (toplam 60) veri seti ve vektör olarak ayrı ayrı nesneler olacak şekilde kaydediniz.
 10. diamonds veri setini ?diamonds komutu ile inceleyiniz. 10.000 doların üzerinde fiyata sahip olan pırlanta sayısını bulunuz.
 11. diamonds veri setini 10.000 doların üzerinde olan kaç farklı renk pırlanta vardır ve bu pırlantaların renklere göre dağılımı nasıldır?
 12. diamonds veri setini kesim türü “Ideal” olan E ve J renklerindeki pırlantalardan oluşan veri setini fiyat değişkenine göre büyükten küçüğe sıralayınız.
 13. diamonds veri setinden price ve cut olmak üzere iki sütunundan oluşan yeni bir veri seti oluşturunuz.

14. diamonds veri setinden x, y ve z değişkenleri sırası ile uzunluk genişlik ve derinlik ölçülerini vermektedir. Uzunluk, genişlik ve derinlik toplamı 18mm'den büyük olan pirlantalardan yeni bir veri setini oluşturunuz. Bu koşula uyan kaç pirlanta olduğunu belirtiniz.

15. “Ideal” ve “Premium” kesimdeki pirlantaların ortalama fiyatını bulunuz.

16. Aşağıdaki kod ile oluşturulan DT veri setinde, satır bazında seçim işlemlerini bolge değişkenine göre kolay yapabilecek ve bolge değişkenine göre alfabetik sıralama yapacak kodu yazınız.

```
DT <- data.table( no = 1:500,  
                  cinsiyet = sample(1:2, size = 500, replace = T),  
                  kitapcik = sample(LETTERS[1:4], 500, replace = T),  
                  bolge = rep(c("Ege", "Akdeniz", "Karadeniz", "Marmara"), 125),  
                  puan = rnorm(500, 60, 15))
```

17. Bolge değişken adını kullanmadan Akdeniz ve Ege Bölgesine ait verileri filtreleyiniz.

17. Satır bazında sıralama işlemi bolge ve kitapcik değişkenlerine göre yaptıktan sonra data.table yapısı içinde kullanılan mult argümanı ile Akdeniz ve Ege bölgesinin ilk sırada yer alan verilerini yazdırınız. Aynı işlemi her iki bölgenin son satırı içinde yapınız.

18. Satır bazında sıralama işlemi bolge ve kitapcik değişkenlerine göre yapıldıktan sonra, “A” kitapçık türüne göre ortalama puanı hesaplayınız.

19. Satır bazında sıralama işlemi bolge ve kitapcik değişkenlerine göre yapıldıktan sonra, her bir bölge için “A” kitapçık türünü alan bireylerin ortalama puanı hesaplayınız.

20. Satır bazında sıralama işlemi bolge ve kitapcik değişkenlerine göre yapıldıktan sonra, “C” kitapçığını alan öğrencilerin cinsiyetlerine göre ortalamalarını hesaplayınız.

21. Cinsiyet ve bölgenin her bir düzeyi için ortalama puan alıp bir sütun olarak DT veri setine ekleyiniz.

Geçmiş olsun:)