
Python İle Veri Bilimi Mart 3. Hafta Ödevi

1. Elde ettiğiniz çıktıların ekran görüntülerini ekleyin.

a) Veri Setinin ilk 5 satırı;

```
   age  sex  cp  trestbps  chol  fbs  ...  exang  oldpeak  slope  ca  thal  target
0   52   1   0     125    212   0  ...    0     1.0     2   2     3     0
1   53   1   0     140    203   1  ...    1     3.1     0   0     3     0
2   70   1   0     145    174   0  ...    1     2.6     0   0     3     0
3   61   1   0     148    203   0  ...    0     0.0     2   1     3     0
4   62   0   0     138    294   1  ...    0     1.9     1   3     2     0

[5 rows x 14 columns]
```

b) Eksik değerler sayısı;

```
age      0
sex      0
cp       0
trestbps 0
chol     0
fbs      0
restecg  0
thalach  0
exang    0
oldpeak  0
slope    0
ca       0
thal     0
target   0
dtype: int64
```

c) İstatiksel Özet;

	age	sex	...	thal	target
count	1025.000000	1025.000000	...	1025.000000	1025.000000
mean	54.434146	0.695610	...	2.323902	0.513171
std	9.072290	0.460373	...	0.620660	0.500070
min	29.000000	0.000000	...	0.000000	0.000000
25%	48.000000	0.000000	...	2.000000	0.000000
50%	56.000000	1.000000	...	2.000000	1.000000
75%	61.000000	1.000000	...	3.000000	1.000000
max	77.000000	1.000000	...	3.000000	1.000000

[8 rows x 14 columns]

2. Eksik verilerle ilgili aldığınız aksiyonları açıklayın.

Öncelikle **“df.isnull().sum()”** komutu kullanarak boş veri var mı yok mu onu bakıyoruz.

Kodun çıktısını inceleyelim;

```
age      0
sex      0
cp       0
trestbps 0
chol     0
fbs      0
restecg  0
thalach  0
exang    0
oldpeak  0
slope    0
ca       0
thal     0
target   0
dtype: int64
```

Çıktı sonucunda sütunlarda herhangi bir boş veri olmadığını görürüz. Eksik verilerle ilgili herhangi bir aksiyon almayız.

3. Temel istatistikler hakkında kısa bir yorum yapın.

Şimdi **“df.describe()”** komutunu kullanıyoruz. Ve çıktımızı getiriyoruz.

Çıktı;

	age	sex	...	thal	target
count	1025.000000	1025.000000	...	1025.000000	1025.000000
mean	54.434146	0.695610	...	2.323902	0.513171
std	9.072290	0.460373	...	0.620660	0.500070
min	29.000000	0.000000	...	0.000000	0.000000
25%	48.000000	0.000000	...	2.000000	0.000000
50%	56.000000	1.000000	...	2.000000	1.000000
75%	61.000000	1.000000	...	3.000000	1.000000
max	77.000000	1.000000	...	3.000000	1.000000

[8 rows x 14 columns]

Çıktıyı incelediğimizde toplam veri kümesinde count, mean, std, min, max bölümlerini inceleyelim.

Count:

"Count, bize veri kümesinde kaç adet veri sütünü olduğunu söyler."

→ "Count, her sütundaki toplam veri sayısını gösterir. Eksik veri olup olmadığını anlamak için kullanılabilir."

Mean (Ortalama):

"Mean, ise sütunların ortalamasını alır."

→ " Mean, her sütundaki değerlerin aritmetik ortalamasını gösterir. Örneğin, yaş sütununun ortalama değeri 54.4'tür, bu da veri kümesindeki bireylerin ortalama yaşını ifade eder."

Target Bölümü;

"Ve Target sütununu incelediğimizde buradaki hastaların yarısından fazlasının kalp hastalığı olduğunu görebiliriz."

→ "Target sütununun ortalama değeri 0.51 olduğuna göre, veri kümesindeki bireylerin yaklaşık %51'inin kalp hastası olduğunu söyleyebiliriz."

Standart Sapma (std):

"Std (Standart sapma), verilerin ortalamadan ne kadar saptığını gösterir. "

→ "Örneğin, age sütununun standart sapması 9.07 olduğuna göre, bireylerin yaşları ortalama yaş etrafında ± 9 yıl kadar değişmektedir. Daha yüksek bir std değeri, o sütundaki verilerin daha geniş bir aralıkta dağıldığını gösterir."

Minimum (min) ve Maksimum (max):

"Age sütununu incelediğimizde veri kümesinde en küçük kişinin 29 yaşında olduğunu anlayabiliriz."

→ "Min değeri, ilgili sütundaki en küçük değeri gösterir. Örneğin, age sütunundaki en küçük yaş 29'dur. Benzer şekilde, max değeri en büyük değeri gösterir ve age sütununda en büyük yaş 77'dir."
