

- Veriye genel bir bakış yapıldığında, bazı verilerin sayısal verilere dönüştürülebileceği fark edilmektedir; örneğin, 'Uygulama Süresi' ve 'Tedavi Süresi' gibi değişkenler sayısal verilere çevrilebilir.

```
===== VERİ SETİ GENEL BİLGİLER =====
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2235 entries, 0 to 2234
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   HastaNo                2235 non-null   int64  
1   Yas                    2235 non-null   int64  
2   Cinsiyet               2066 non-null   object  
3   KanGrubu               1560 non-null   object  
4   Uyrak                  2235 non-null   object  
5   KronikHastalik         1624 non-null   object  
6   Bolum                  2224 non-null   object  
7   Alerji                 1291 non-null   object  
8   Tanilar                2160 non-null   object  
9   TedaviAdi              2235 non-null   object  
10  TedaviSuresi           2235 non-null   object  
11  UygulamaYerleri        2014 non-null   object  
12  UygulamaSuresi         2235 non-null   object  
dtypes: int64(2), object(11)
memory usage: 227.1+ KB
None
```

- İlk iki satır incelendiğinde, 'KronikHastalık', 'Bolum', 'Alerji', 'Tanilar', 'TedaviAdi' ve 'UygulamaYerleri' sütunlarının metin verisi içerdiği ve bir hücrede birden fazla bilgi barındırabildiği görülmektedir. Bu nedenle, söz konusu verilerin listeleme ve normalizasyon işlemlerine tabi tutulması mantıklı olacaktır. Böylece veriler daha düzenli bir şekilde saklanabilir ve model eğitimi için daha uygun hale gelir. (Bkz. Görsel1: orijinal hali, Görsel2: listelenmiş hali)

	HastaNo	KronikHastalik	Bolum	Alerji	Tanilar	TedaviAdi	UygulamaYerleri
0	145134	Becker Musküler Distrofisi, Hipotiroidizm, Ka...	Fiziksel Tıp Ve Rehabilitasyon,Solunum Merkezi	TOZ	Ayak bileği ve ayağın yüzeysel yaralanması	Ayak Bileği	Ayak Bileği
1	145135	Duchenne Musküler Distrofisi, Myastenia gravis...	Fiziksel Tıp Ve Rehabilitasyon,Solunum Merkezi	NaN	Omuzun darbe sendromu,DORSALJİ, DİĞER, SERVİKO...	Dorsalji -Boyun+trapez+skapular	Boyun
2	145135	Duchenne Musküler Distrofisi, Myastenia gravis...	Fiziksel Tıp Ve Rehabilitasyon,Solunum Merkezi	NaN	Omuzun darbe sendromu,DORSALJİ, DİĞER, SERVİKO...	Dorsalji -Boyun+trapez+skapular	Boyun,Sırt

Görsel 1

	HastaNo	KronikHastalik	Bolum	Alerji	Tanilar	TedaviAdi	UygulamaYerleri
0	145134	['becker musküler distrofisi', 'hipotiroidizm...	['fiziksel tıp ve rehabilitasyon', 'solunum me...	['toz']	['ayak bileği ve ayağın yüzeysel yaralanması']	['ayak bileği']	['ayak bileği']
1	145135	['duchenne musküler distrofisi', 'myastenia gr...	['fiziksel tıp ve rehabilitasyon', 'solunum me...	NaN	['omuzun darbe sendromu', 'dorsalji', 'diğer', ...	['dorsalji boyun+trapez+skapular']	['boyun']
2	145135	['duchenne musküler distrofisi', 'myastenia gr...	['fiziksel tıp ve rehabilitasyon', 'solunum me...	NaN	['omuzun darbe sendromu', 'dorsalji', 'diğer', ...	['dorsalji boyun+trapez+skapular']	['boyun', 'sırt']

Görsel 2

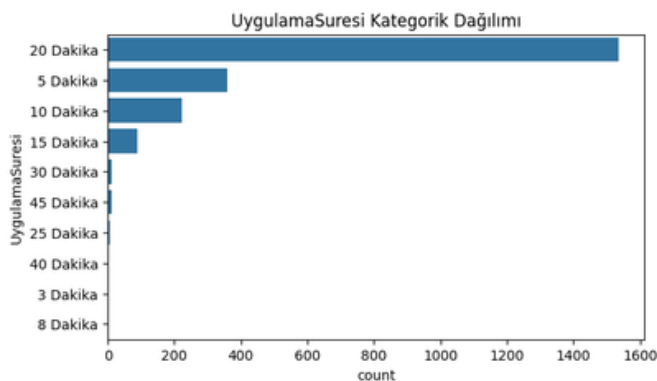
- Cinsiyet sütunu binary (ikili) şekilde encode edildi.

	HastaNo	Cinsiyet
0	145134	0.0
1	145135	1.0
2	145135	1.0
3	145135	1.0
4	145135	1.0

- Kan grubu sütunundan, "KanGrubuTürü" ve "RhFaktörü" adında iki yeni sütun oluşturuldu. Örneğin, bir satırda KanGrubu "0" olduğunda, KanGrubuTürü "Rh+" ve RhFaktörü de "Rh+" olacak şekilde veriler ayrıştırıldı. Bu işlem, özellikle bazı hastalıkların tanı ve tedavisinde Rh faktörünün önemli rolü olduğu için modelin bu özellikleri daha doğru öğrenmesini sağlar.

	HastaNo	KanGrubu	KanGrubuTürü	RhFaktörü
0	145134	0 Rh+	0	Rh+
1	145135	0 Rh+	0	Rh+
2	145135	0 Rh+	0	Rh+
3	145135	0 Rh+	0	Rh+
4	145135	0 Rh+	0	Rh+

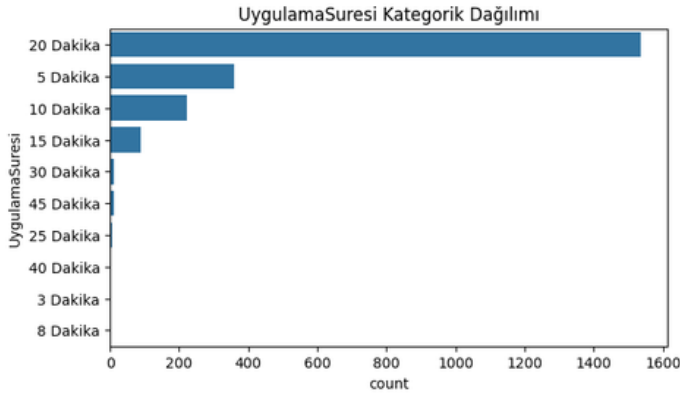
- Uyruk değişkeninde yalnızca 5 farklı kategori bulunduğu için, bu değişkeni temsil etmede One-Hot Encoding uygun ve mantıklı bir seçim olacaktır.



- Yaş değişkeninden, bireyleri yaş aralıklarına göre gruplandırarak yeni bir YaşGrubu değişkeni türetebiliriz. Bu gruplandırma şu şekilde yapılabilir: 0–14 yaş (Çocuklar ve Ergenler: 0), 15–24 yaş (Gençler ve Genç Yetişkinler: 1), 25–64 yaş (Yetişkinler: 2), 65–79 yaş (Yaşlılar: 3) ve 80 yaş ve üzeri (Çok Yaşlılar / Yaşlılık Sonrası Dönem: 4).

	HastaNo	Yas	Yas Grubu
0	145134	60	2
1	145135	28	2
2	145135	28	2
3	145135	28	2
4	145135	28	2

- UygulamaSuresi değişkeninde 10 farklı değer bulunduğunu ve herhangi bir tutarsızlık olmadığını gözlemledik. Ancak, modelin daha verimli çalışmasını sağlamak amacıyla bu değişkene normalizasyon işlemi uyguladık.



	HastaNo	UygulamaSuresi
0	145134	0.546788
1	145135	0.546788
2	145135	0.546788
3	145135	-1.846613
4	145135	0.546788