**1. Proje Amacı**

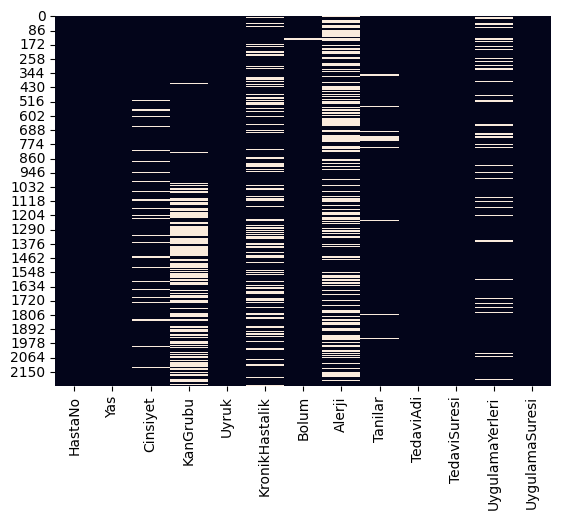
Bu çalışma, Talent Academy tarafından sağlanan hasta verilerini analiz ederek eksik değerlerin giderilmesi, kategorik verilerin sayısallaştırılması ve sayısal değişkenlerin ölçeklenmesini kapsamaktadır. Hedef değişken: TedaviSuresi.

**2. Veri Seti ve Genel Bilgiler**

* Veri seti Excel formatında sağlanmıştır: Talent\_Academy\_Case\_DT\_2025.xlsx.
* Veri seti **satır ve sütun sayısı**: df.shape
* Veri tipleri karışıktır; sayısal (int64, float64) ve kategorik (object) değişkenler mevcuttur.
* Bazı sütunlarda eksik değerler bulunmaktadır.

**3. Eksik Değer Analizi**

* Eksik değerler sütun bazında tespit edildi: df.isnull().sum()
* TedaviSuresi, UygulamaSuresi ve bazı kategorik sütunlar eksikti.
* Eksik değerlerin görselleştirilmesi için ısı haritası kullanıldı (sns.heatmap(df.isnull(), cbar=False))(Görsel 1).



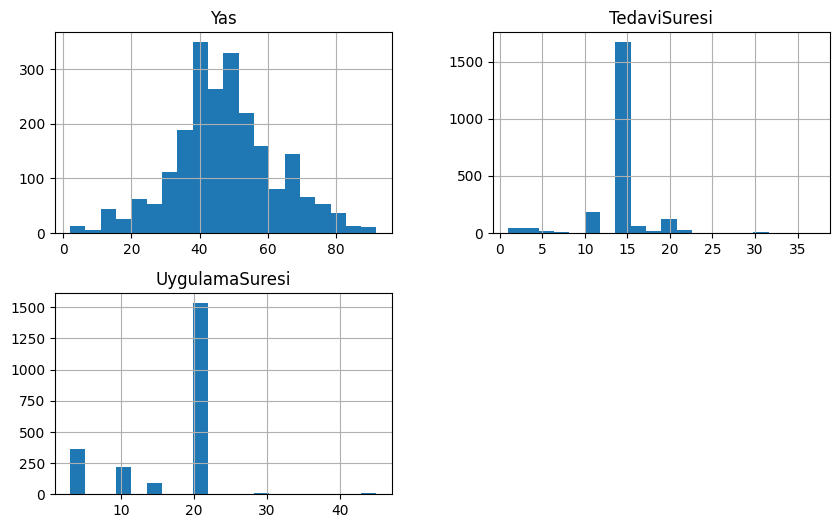
Görsel 1

**4. Veri Temizleme ve Dönüşümler**

* **Sayısal dönüşümler:**
  + TedaviSuresi ve UygulamaSuresi sütunları sayısal değere çevrildi.
  + String içerisindeki sayılar çıkarıldı, dönüştürülemeyenler NaN olarak bırakıldı.
* **Eksik değerlerin doldurulması:**
  + Sayısal sütunlar KNN imputer ile dolduruldu (n\_neighbors=5).

**5. Sayısal Değişkenlerin Görselleştirilmesi**

* Yas, TedaviSuresi ve UygulamaSuresi histogramları ile dağılımları incelendi(Görsel 2).
* Dağılımların normal veya çarpık olup olmadığı gözlemlendi.



Görsel 2

**6. Kategorik Değişkenlerin İşlenmesi**

* **One-Hot Encoding:** Cinsiyet, KanGrubu, Uyruk, Bolum
* **Label Encoding:** KronikHastalik, Alerji, Tanilar, UygulamaYerleri
* Böylece kategorik değişkenler makine öğrenmesi algoritmalarına uygun hâle getirildi.

**7. Sayısal Değişkenlerin Ölçeklenmesi**

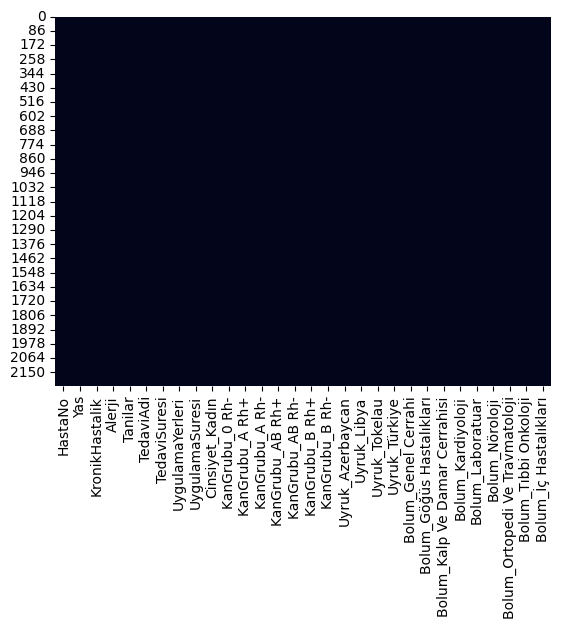
* Yas, UygulamaSuresi ve TedaviSuresi standartlaştırıldı (StandardScaler).
* Ölçekleme, özellikle regresyon ve mesafe tabanlı algoritmalar için önemlidir.

**8. Özellik ve Hedef Değişkenler**

* **Özellikler (X):** HastaNo ve TedaviSuresi hariç tüm sütunlar.
* **Hedef değişken (y):** TedaviSuresi
* Veri seti model eğitimine hazır hâle getirildi.

**9. Son Kontroller**

* İşlem sonrası veri setinde eksik değer kalmadığı doğrulandı (sns.heatmap(df.isnull(), cbar=False))(Görsel 3).



Görsel 3

**10. Özet ve Sonuçlar**

* Veri seti başarıyla incelendi ve eksik değerler giderildi.
* Sayısal ve kategorik sütunlar uygun şekilde dönüştürüldü ve ölçeklendirildi.
* Kategorik değişkenler encoding ile sayısallaştırıldı.
* Sayısal değişkenler normalize edildi.
* Model eğitimi için veri seti (X ve y) hazırlandı.
* EDA ve veri ön işleme süreci sonucunda veri temiz, analiz ve modellemeye uygun hâle getirildi.