**Taha Yasir YEŞİL / tahayesil4040@gmail.com**

**Veri Analizi Raporu**

**1. Veri Kümesi Özeti**

Bu veri kümesi, kullanıcılara ait ilaç kullanımı, yan etkiler ve kronik hastalıklar gibi tıbbi bilgileri içermektedir. Veri kümesinde bulunan sütunlar şunlardır:

* Kullanici\_id
* Cinsiyet
* Dogum\_Tarihi
* Uyruk
* Il
* Ilac\_Adi
* Ilac\_Baslangic\_Tarihi
* Ilac\_Bitis\_Tarihi
* Yan\_Etki
* Yan\_Etki\_Bildirim\_Tarihi
* Alerjilerim
* Kronik Hastaliklarim
* Baba Kronik Hastaliklari
* Anne Kronik Hastaliklari
* Kiz Kardes Kronik Hastaliklari
* Erkek Kardes Kronik Hastaliklari
* Kan Grubu
* Kilo
* Boy

Analizde kullanmak üzere de ekstra sütunlar eklenmiştir. Bu sütunlar şunlardır:

Günümüz tarihi - Dogum\_Tarihi = Yaş sütunu

Ilac\_Bitis\_Tarihi - Ilac\_Baslangic\_Tarihi = Ilac\_Kullanim\_Suresi (günlük)

Yan\_Etki\_Bildirim\_Tarihi - Ilac\_Baslangic\_Tarihi = Yan\_Etki\_Ilac\_kullanim\_tarihi (günlük)

**2. Veri Temizleme Adımları**

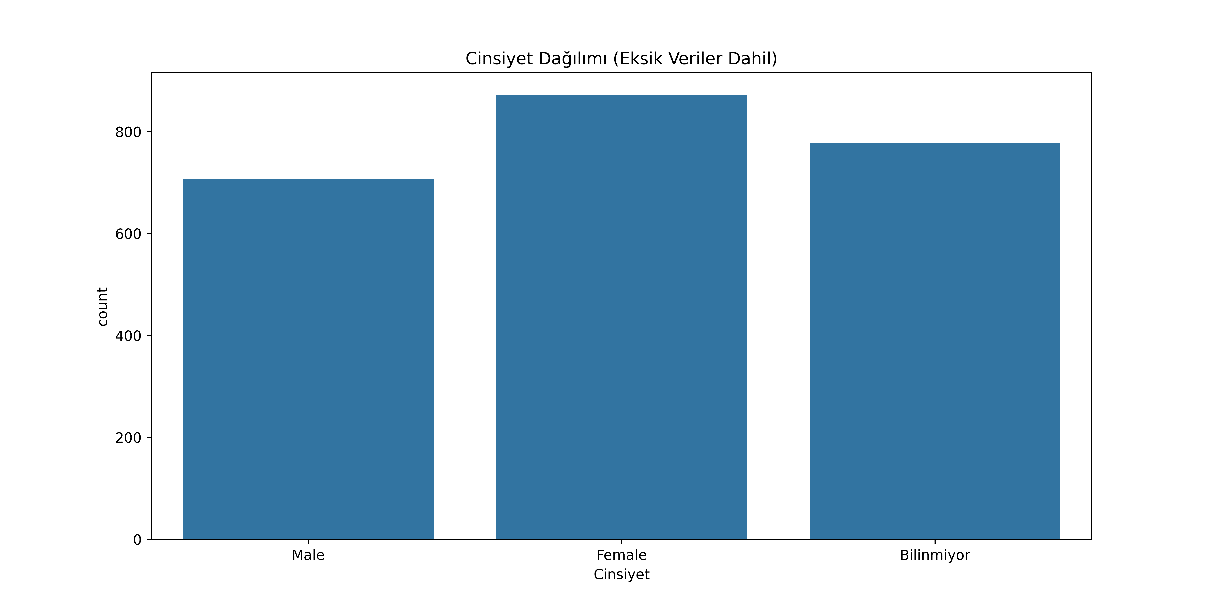
**Aynı Veriler**

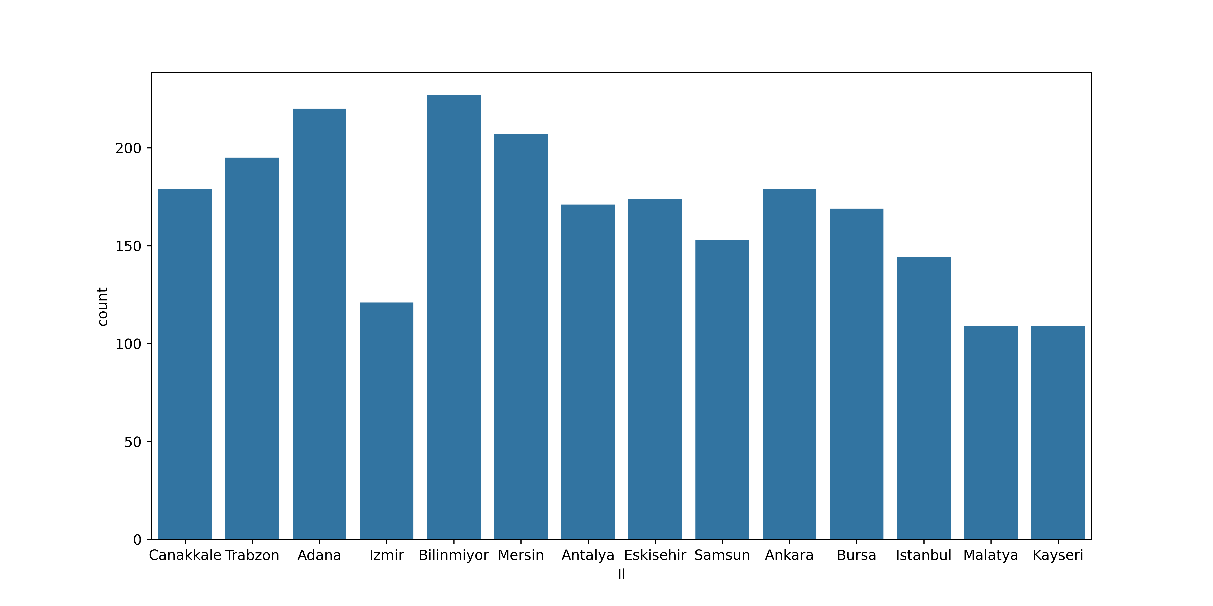
Veri setindeki Uyruk sütunundaki değerlerin hepsi aynı olduğu için bu sütun veri analizi için anlamlı bir bilgi sunmamaktadır. Bu durumdan dolayı Uyruk sütunu veri setinden kaldırıldı çünkü değişkenlik göstermeyen bir sütun, modelin öğrenme sürecine katkı sağlamaz ve gereksiz karmaşıklık yaratabilir.

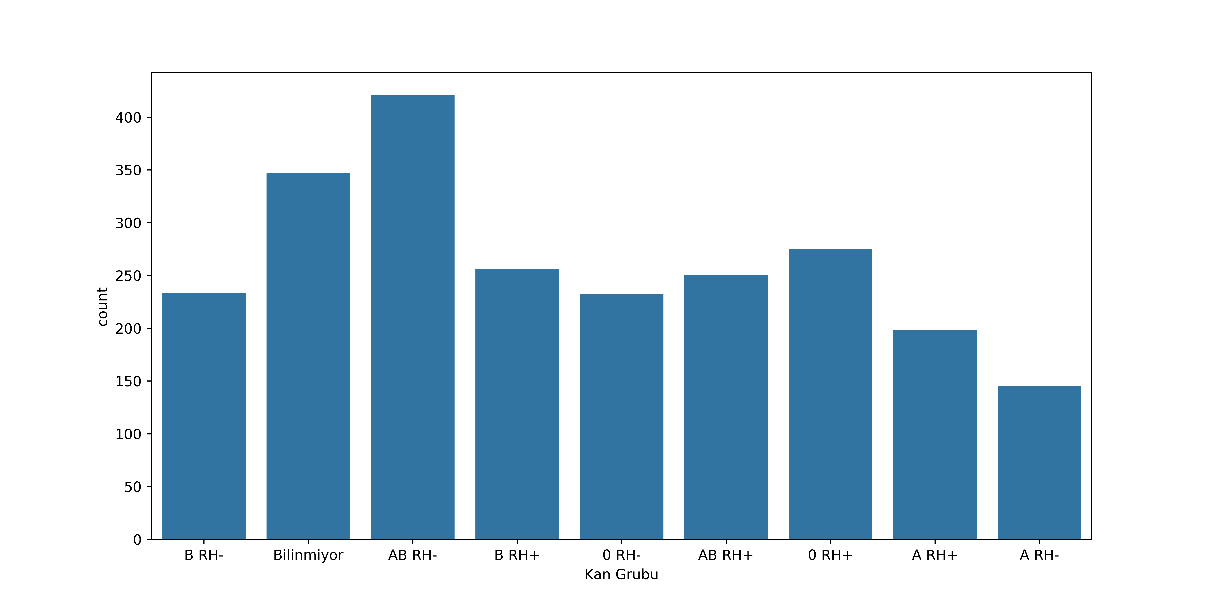
**Eksik Veriler**

Veri setinde bazı sütunlarda eksik değerler bulunmaktaydı. Özellikle aşağıdaki alanlarda eksik veriler mevcuttu:

* Cinsiyet: 1579 dolu gözlem (778 eksik)
* İl: 2130 dolu gözlem (227 eksik)
* Alerjilerim (484 eksik)
* Kronik Hastaliklarim (392 eksik)
* Baba Kronik Hastaliklari (156 eksik)
* Anne Kronik Hastaliklari (217 eksik)
* Kiz Kardes Kronik Hastaliklari (97 eksik)
* Erkek Kardes Kronik Hastaliklari (121 eksik)
* Kan Grubu: 2010 dolu gözlem (347 eksik)
* Kilo: 2064 dolu gözlem (293 eksik)
* Boy: 2243 dolu gözlem (114 eksik)







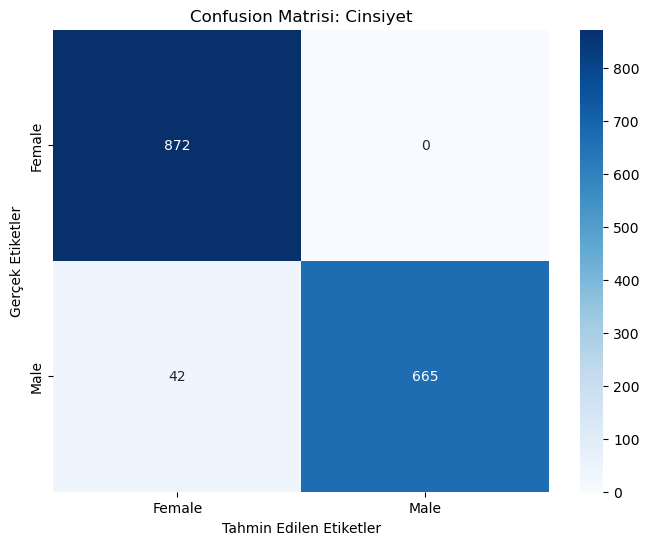
**Eksik veriler için uygulanan adımlar:**

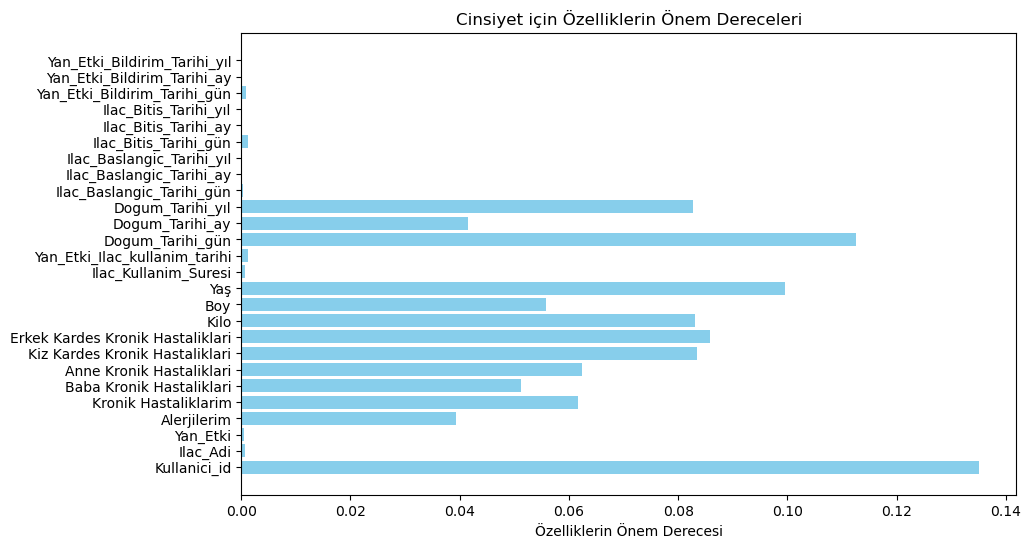
**Normalizasyon:** Eksik verilerin bulunması için sayısal veriler önden normalize edilmektedir. (Kilo, Boy, Yaş, Yan\_Etki\_Ilac\_kullanim\_tarihi Ilac\_Kullanim\_Suresi)

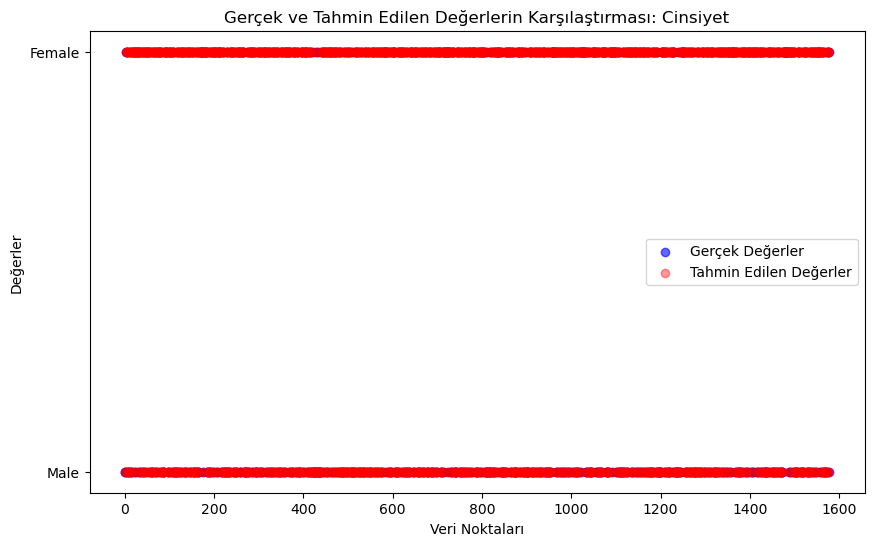
**Kategorik Sütunlar:** Alerjilerim, Kronik Hastaliklarim, Baba Kronik Hastaliklari, Anne Kronik Hastaliklari, Kiz Kardes Kronik Hastaliklari, Erkek Kardes Kronik Hastaliklari sütunlarındaki eksik verilere, hastaların bu özelliklere sahip olmadığı varsayılarak “yok” yazıldı.

Cinsiyet, Il ve Kan Grubu değerleri için ise Random Forest algoritması kullanıldı. Random Forest parametrelerine ise max\_depth=5, min\_samples\_split=30 değerleri verildi.

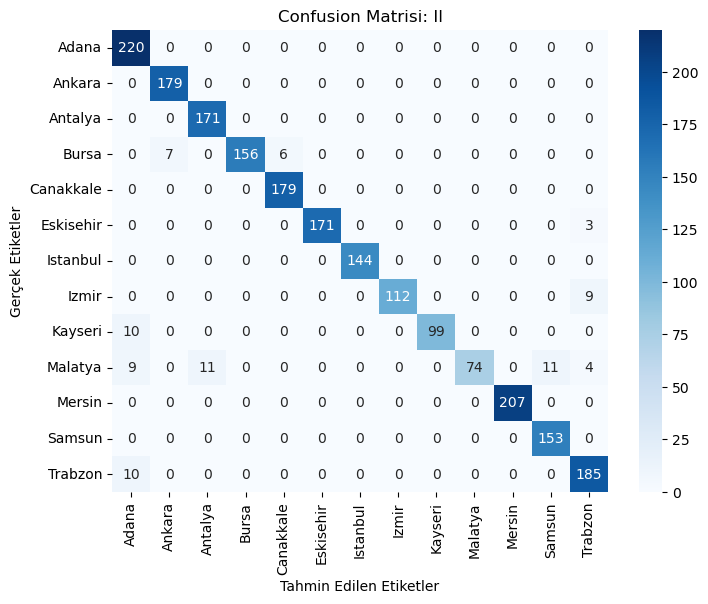
**Cinsiyet:** Random Forest algoritması ile cinsiyet sütunundaki eksik veriler tahmin edilerek dolduruldu. Ayrıca çapraz doğrulama (cross-validation) ile modelin doğruluğu ölçülerek 0.9690'lık bir başarı elde edildi.

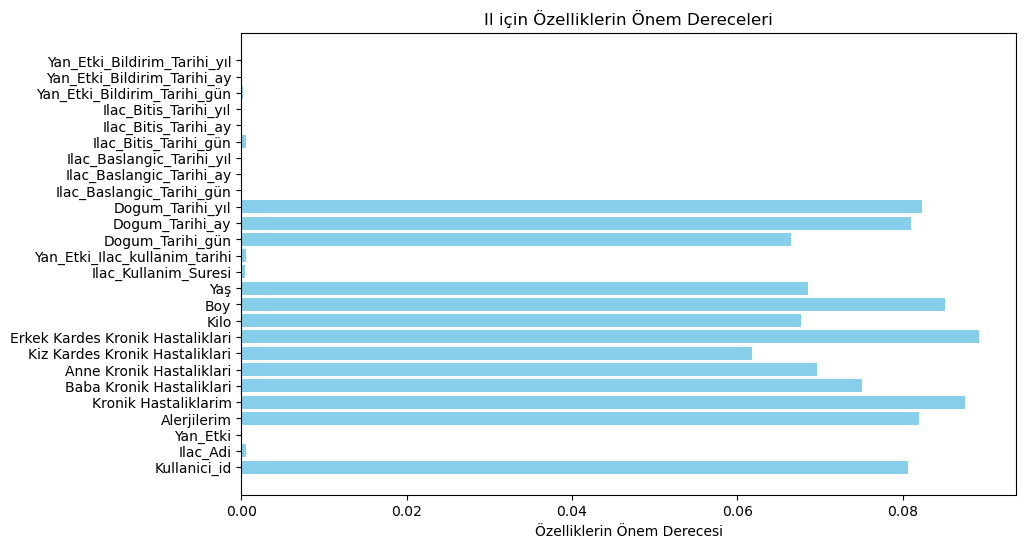






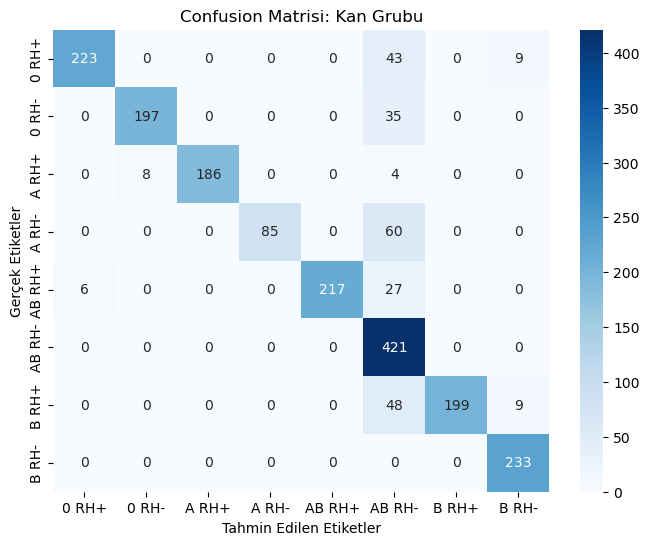
**İl:** Random Forest algoritması ile cinsiyet sütunundaki eksik veriler tahmin edilerek dolduruldu. Ayrıca çapraz doğrulama (cross-validation) ile modelin doğruluğu ölçülerek 0.8991'lik bir başarı elde edildi.

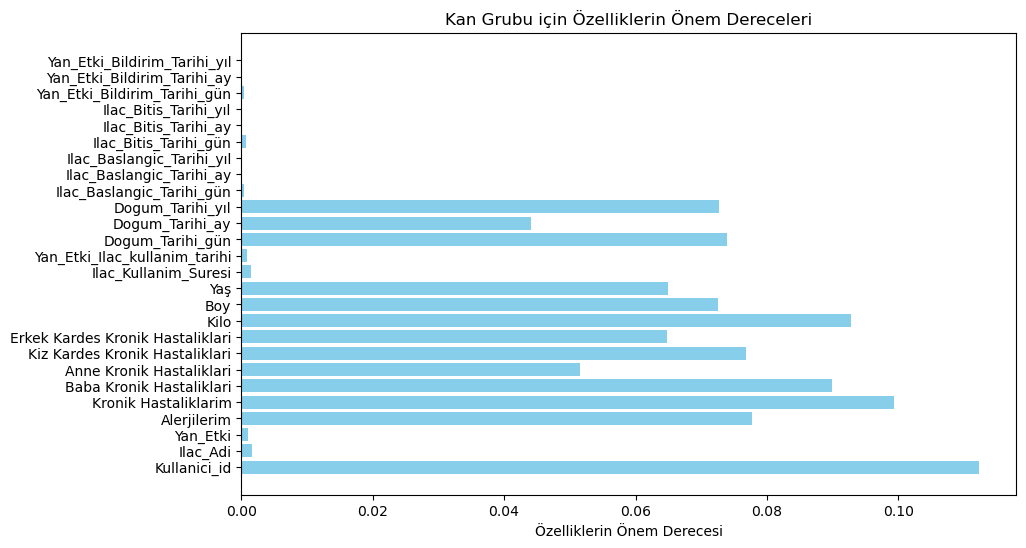


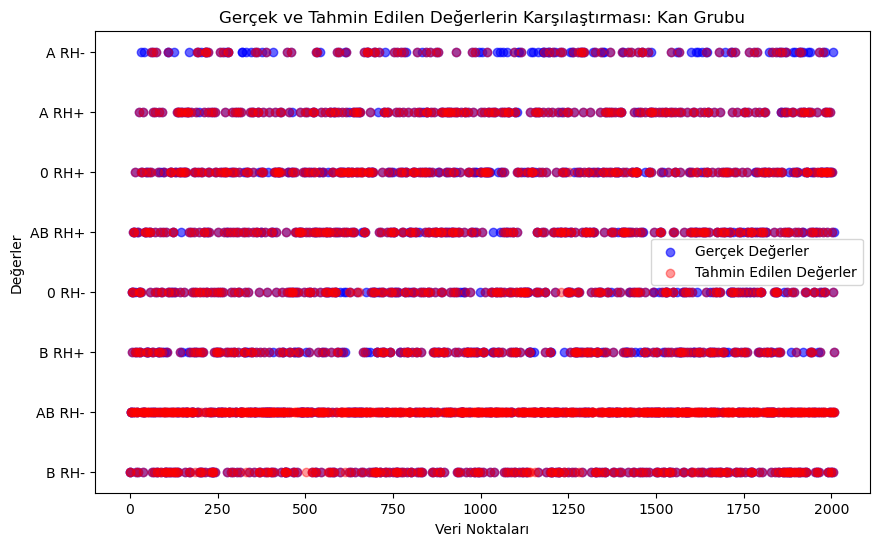




**Kan Grubu:** Random Forest algoritması ile cinsiyet sütunundaki eksik veriler tahmin edilerek dolduruldu. Ayrıca çapraz doğrulama (cross-validation) ile modelin doğruluğu ölçülerek 0.8542'lik bir başarı elde edildi.

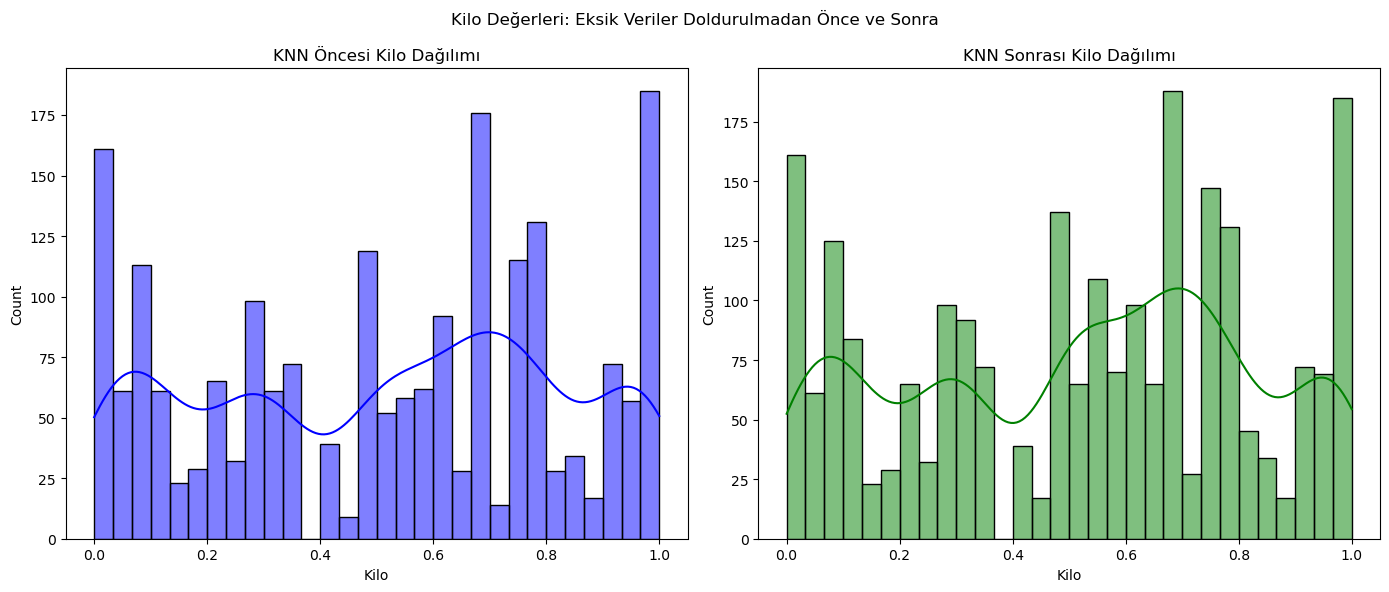


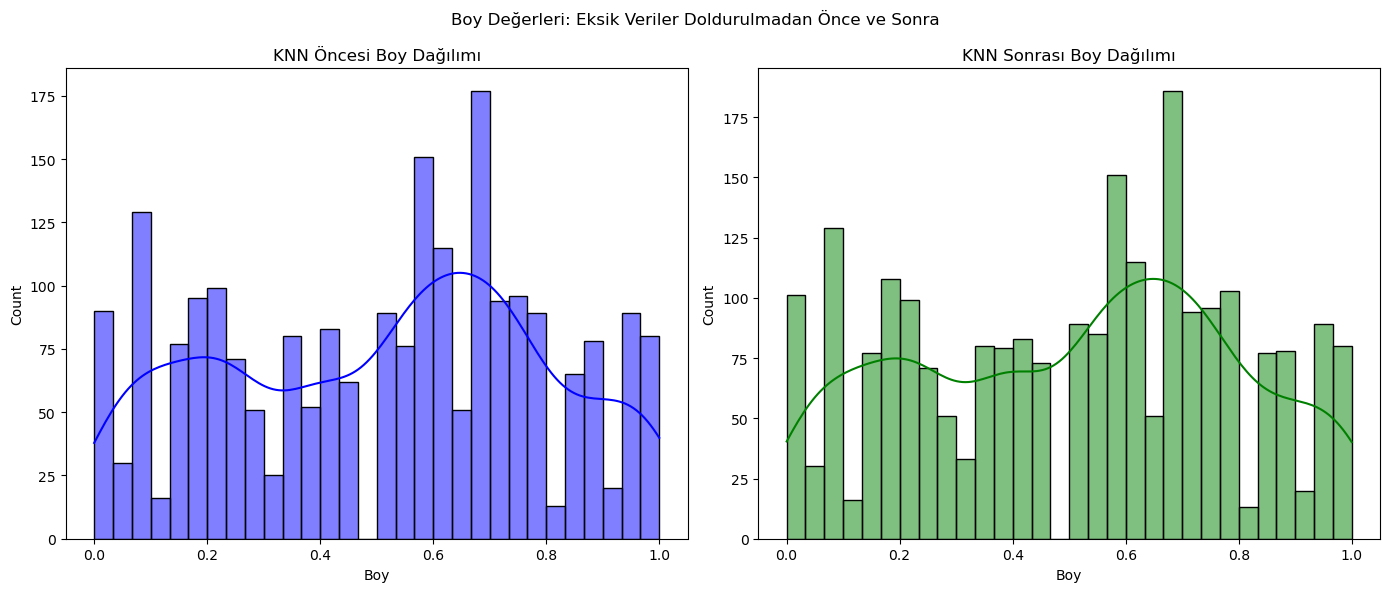




Random Forest parametrelerini değiştirerek daha uygun veriler elde etmek mümkündür.

**Numerik Sütunlar:** Kilo ve boy sütunlarındaki eksik değerler KNNImputer ile dolduruldu.





Knn öncesi ve Knn sonrası dağılım grafikleri gösterilmektedir. Grafikler incelendiğinde Knn sonrası dağılımların minimal olarak değiştiği görülmektedir.

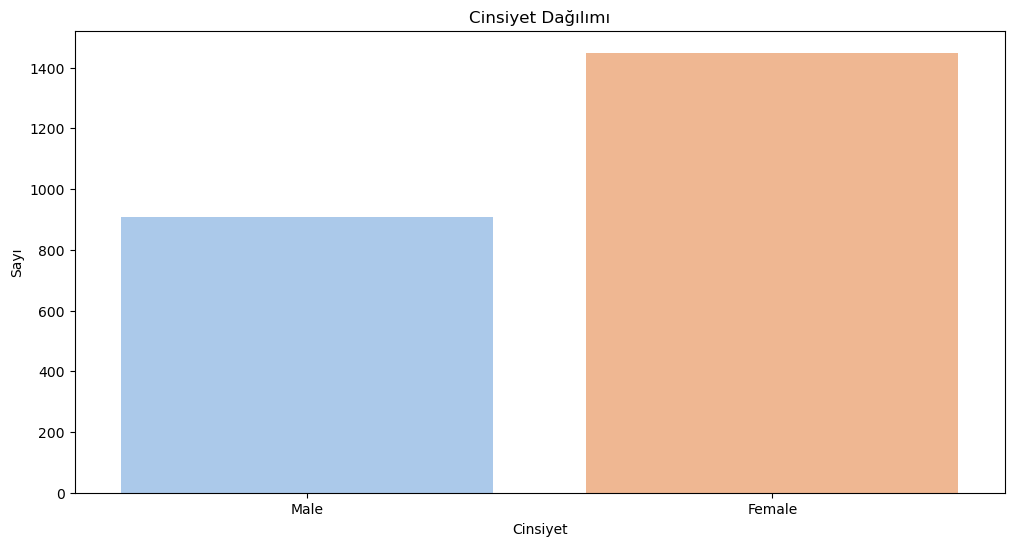
**Aykırı Değerler:** Özellikle Kilo ve Boy sütunlarındaki aykırı değerler incelendi ama düzeltme yapılmasına gerek olmadığı sonucuna varıldı.

**3. Keşifsel Veri Analizi (EDA)**

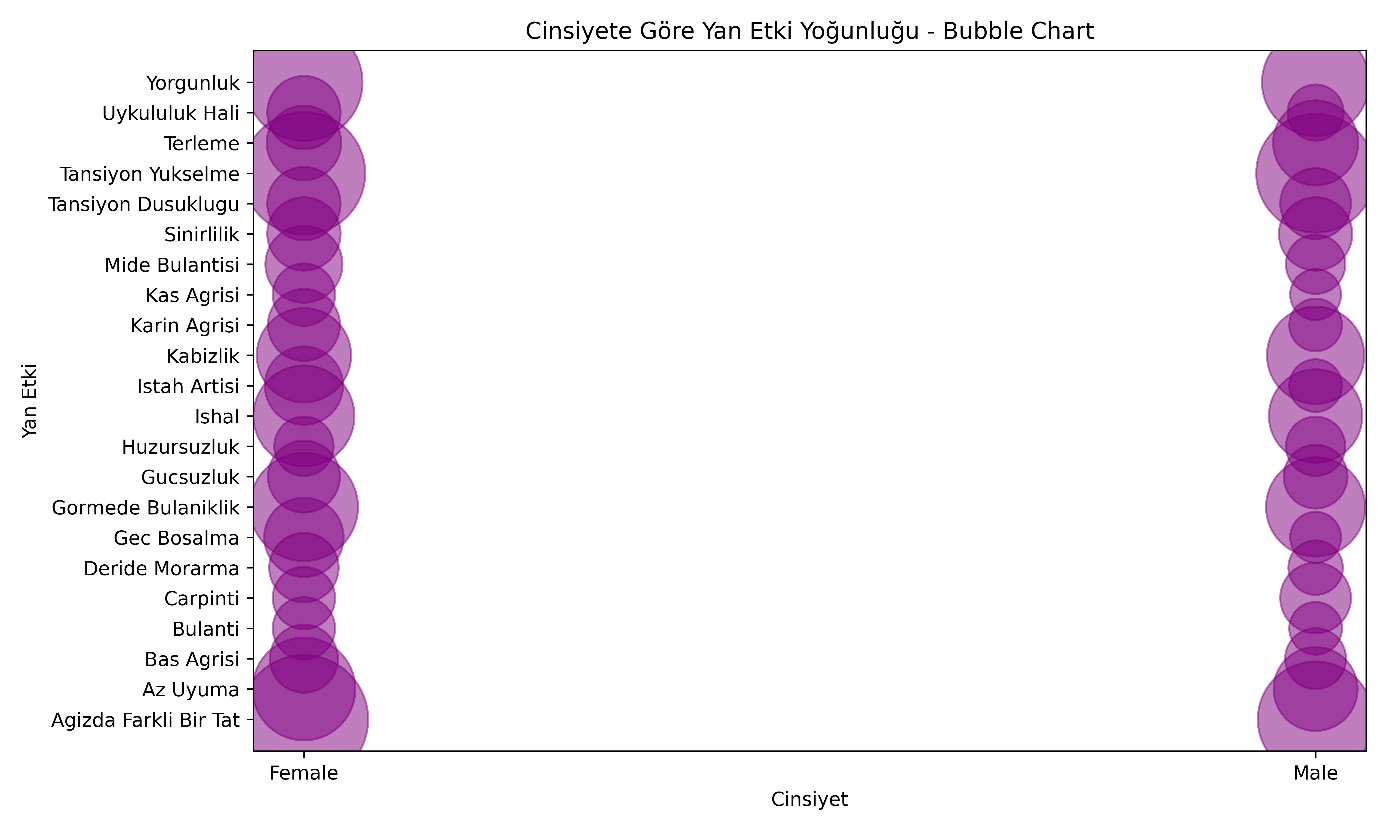
Verilerin keşifsel analizi sırasında aşağıdaki adımlar izlendi:

**3.1 Kategorik Değişkenlerin İncelenmesi**

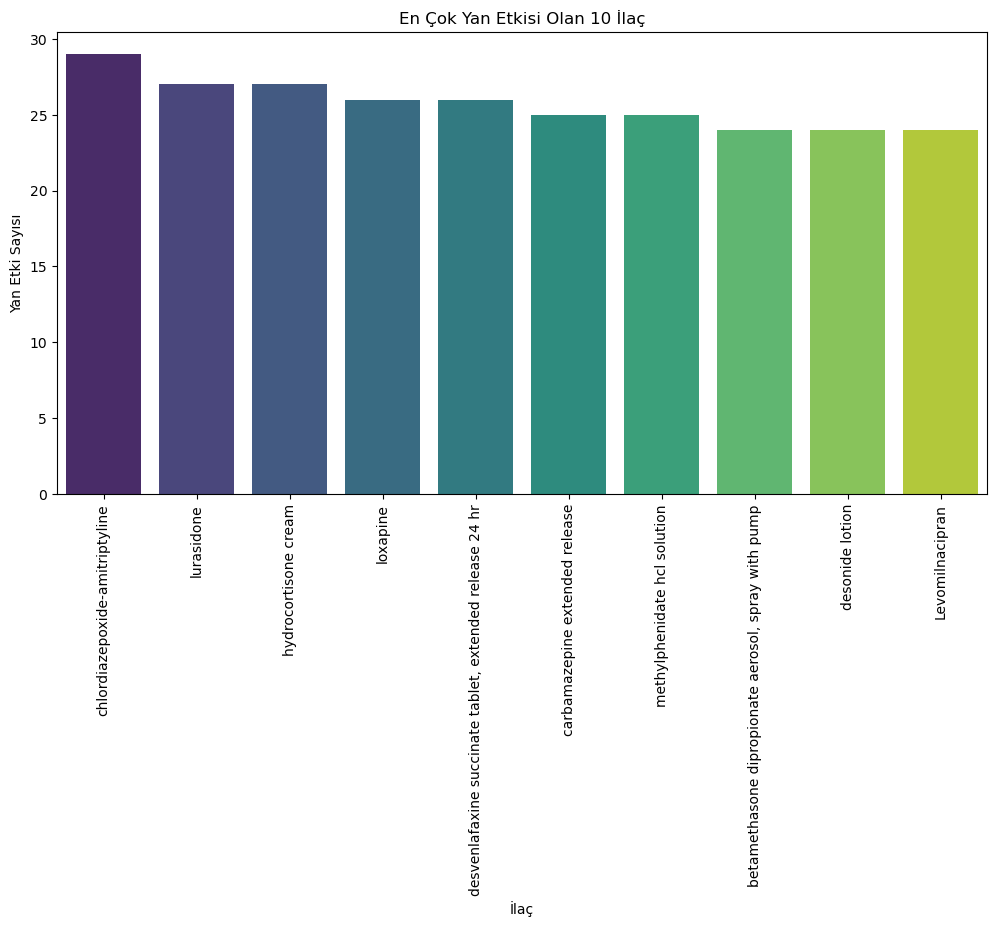
**Cinsiyet:** Verilerde erkek ve kadın cinsiyeti arasındaki dağılım incelendi. Yaklaşık olarak yüzde 66'sı dolu olan bu sütunda eksik veriler Random Forest ile dolduruldu.

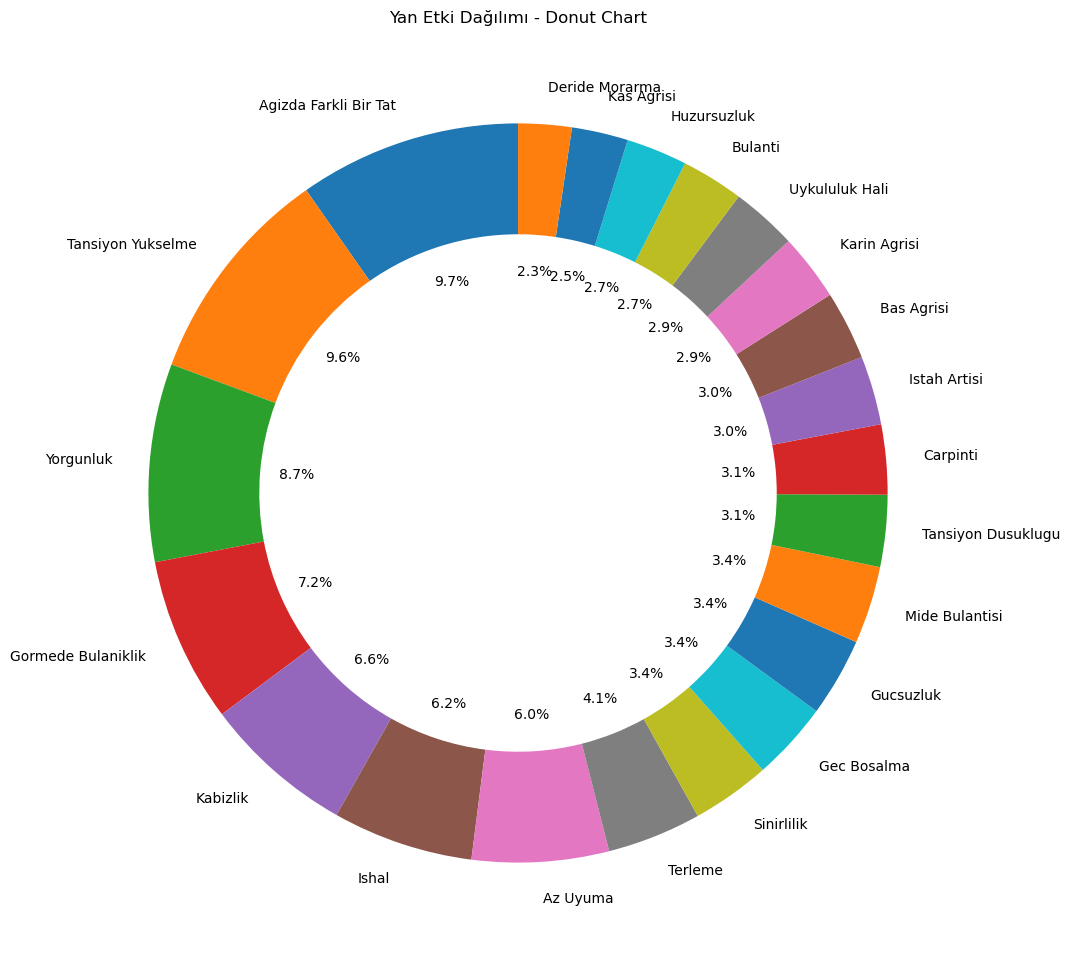


**Cinsiyete Göre Yan Etki Yoğunluğu:** Cinsiyete göre yan etkilerin yoğunlukları incelendi.

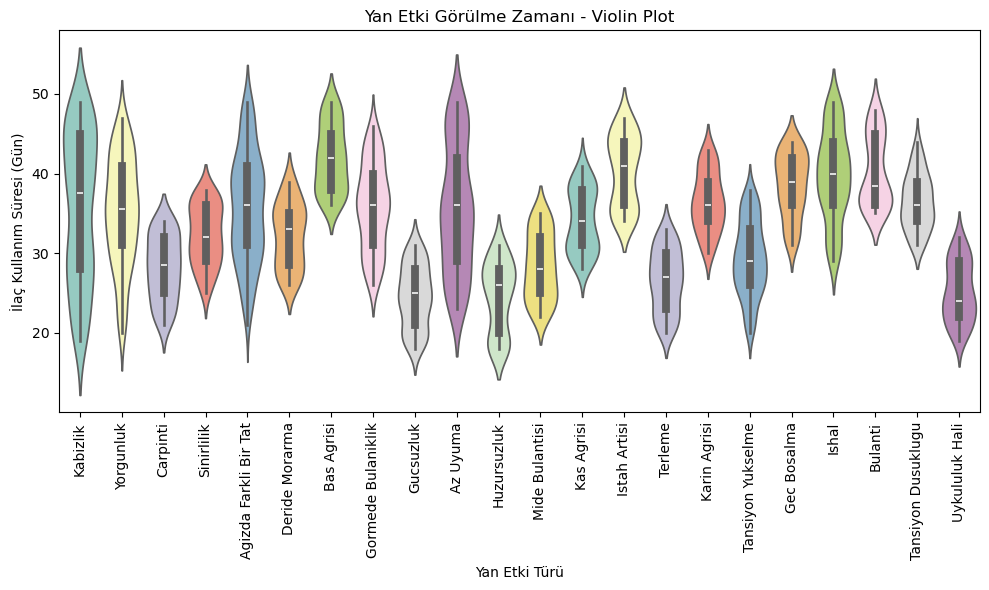


**İlaçlar ve Yan Etkiler:** Farklı ilaçların yan etkileri analiz edildi.

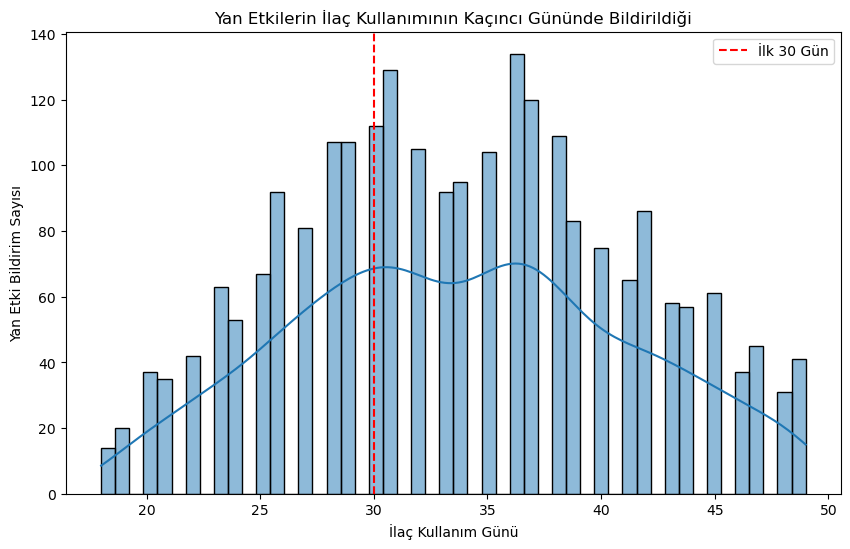




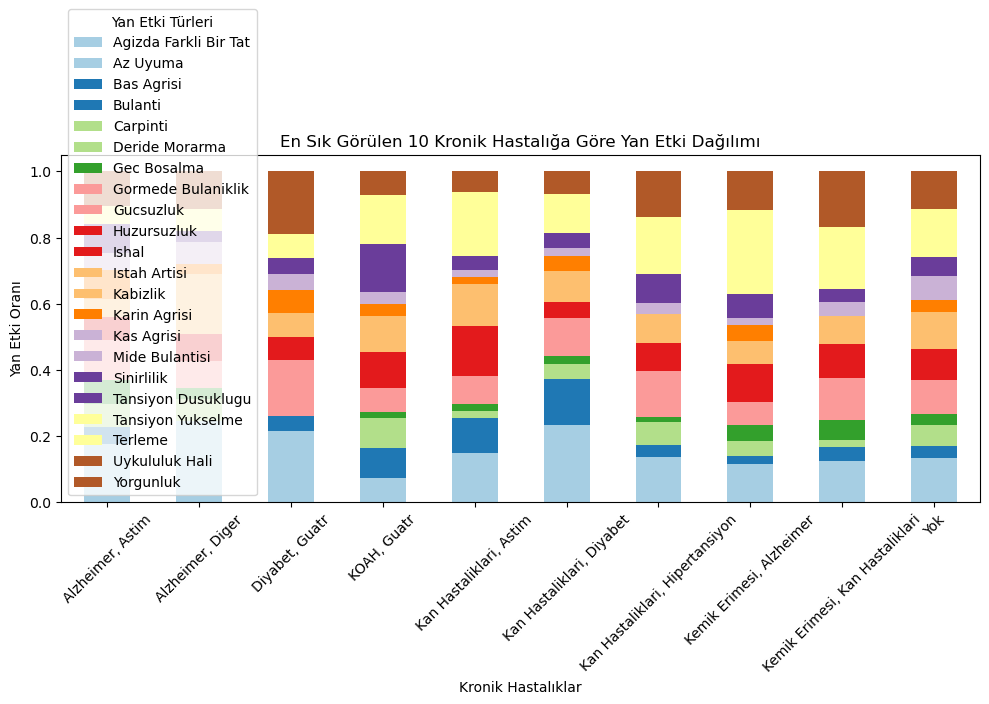
**Yan Etki Görülme Zamanı:** Yan etkilerin ilaç kullanım süresine göre zamanı analiz edildi.



**Yan Etki ve İlaç Kullanımı:** Yan Etkilerin ilaç kullanımının en çok kaçıncı gününde görüldüğü incelendi.

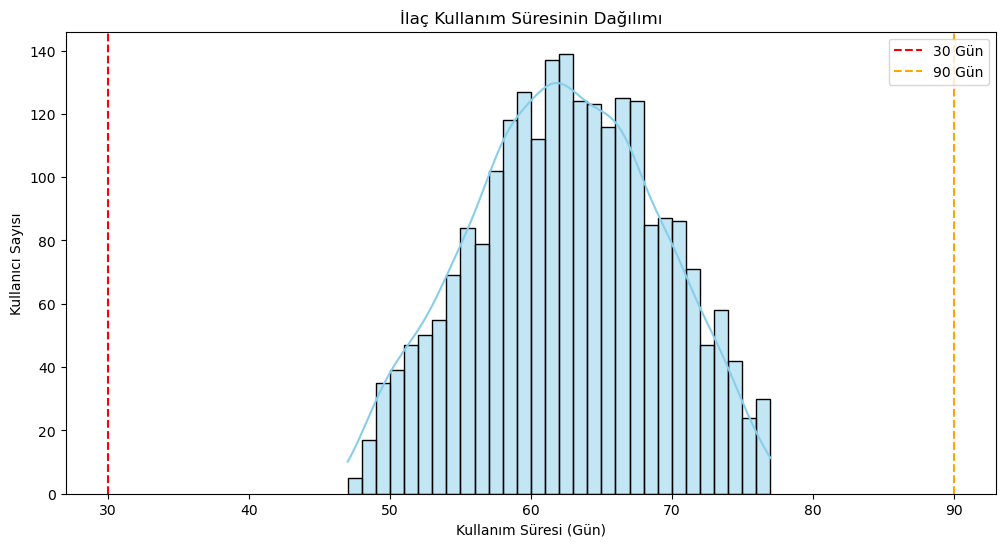


**Kronik Hastalık ve Yan Etki Dağılımı:** En sık görülen 10 kronik hastalığa göre yan etki dağılımı incelendi.



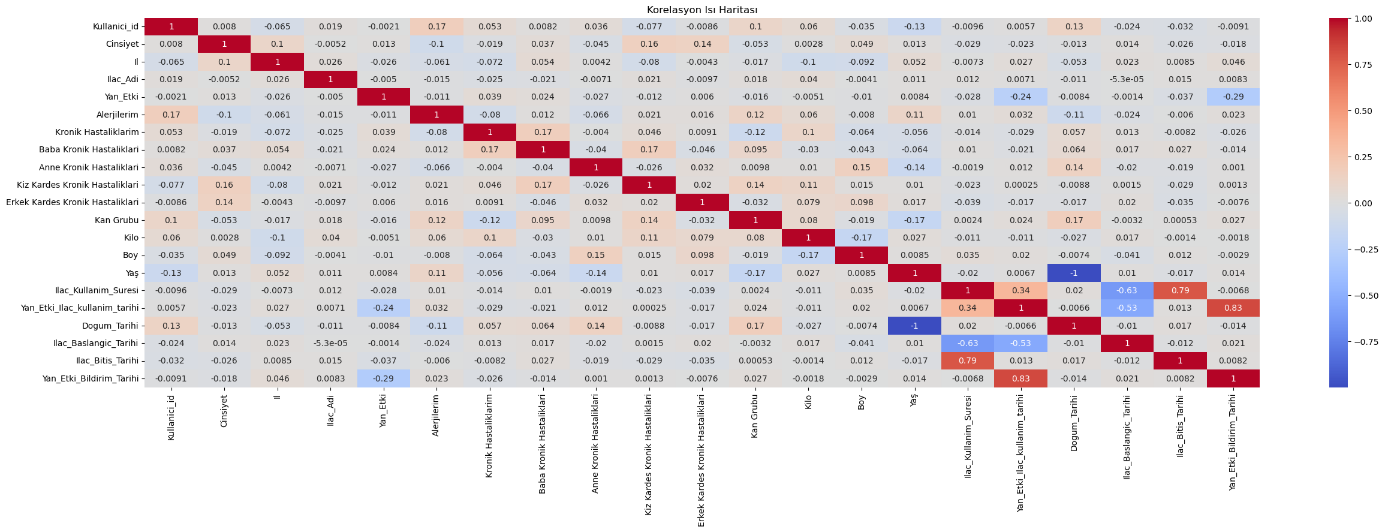
**3.2 Sayısal Değişkenlerin İncelenmesi**

**İlaç Kullanım Süresi:** İlaç kullanım sürelerinin dağılımı incelendi.



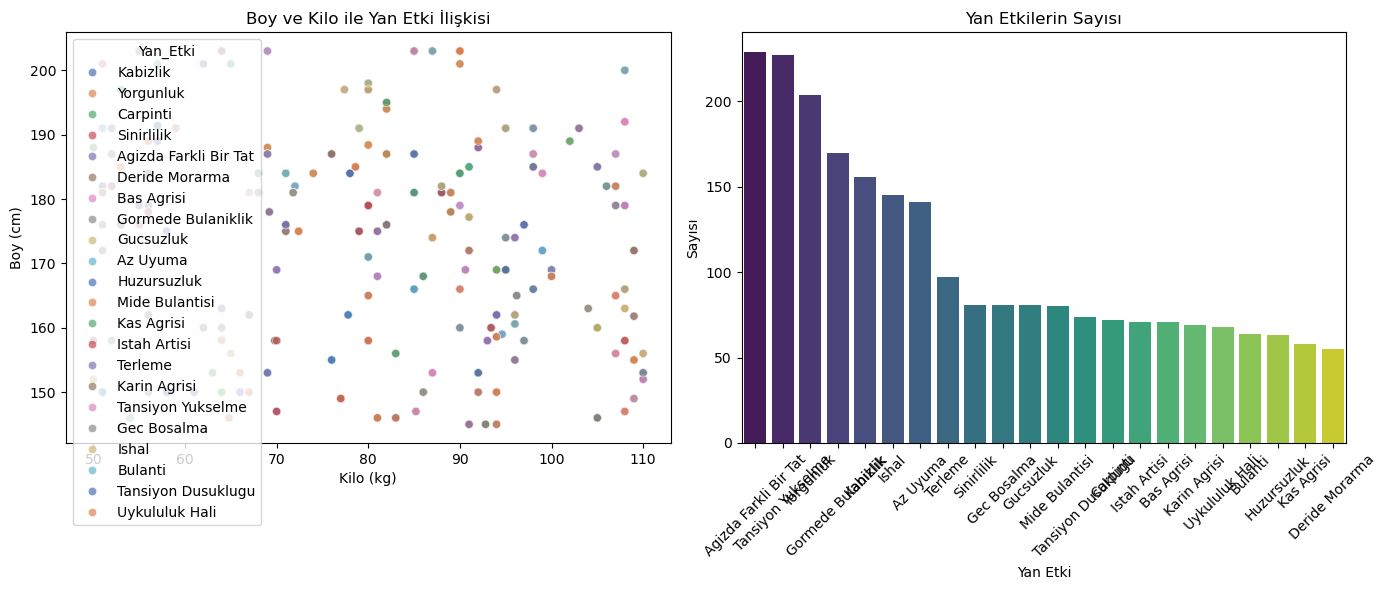
**3.3 Korelasyon Analizleri**

**Korelasyon Isı Haritası:** Tüm değişkenler arasındaki ilişkiler incelendi.

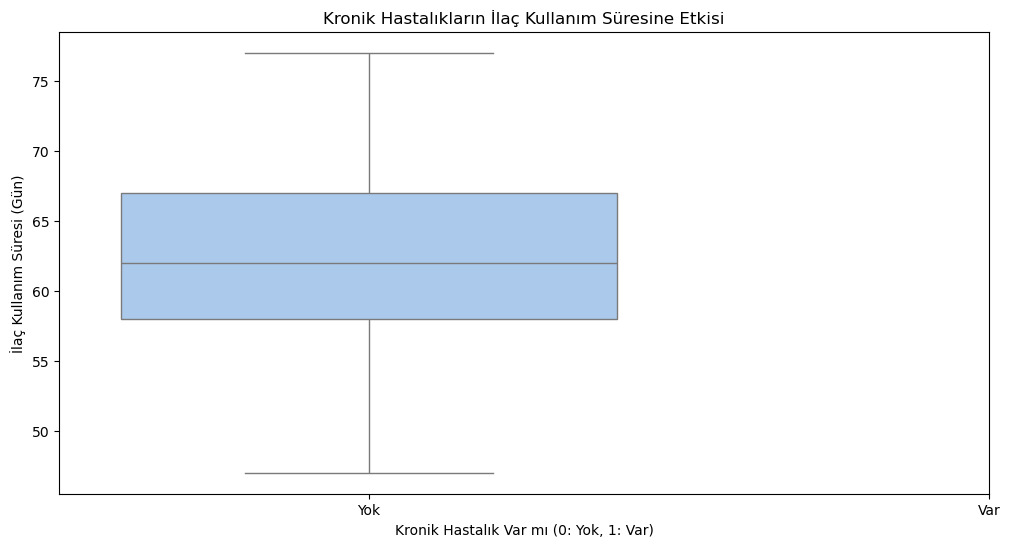


**3.4 Yan Etkiler ve İlişkiler**

Boy ve Kilo ile Yan Etki İlişkisi: Bu iki değişkenin yan etkilerle olan ilişkisi analiz edildi.

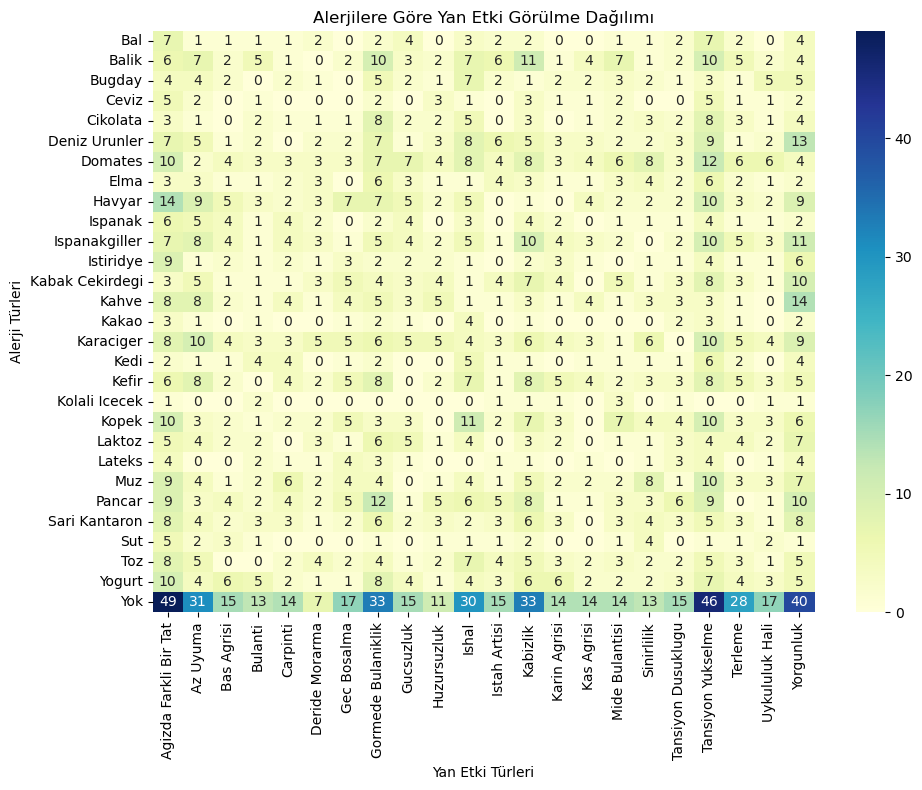


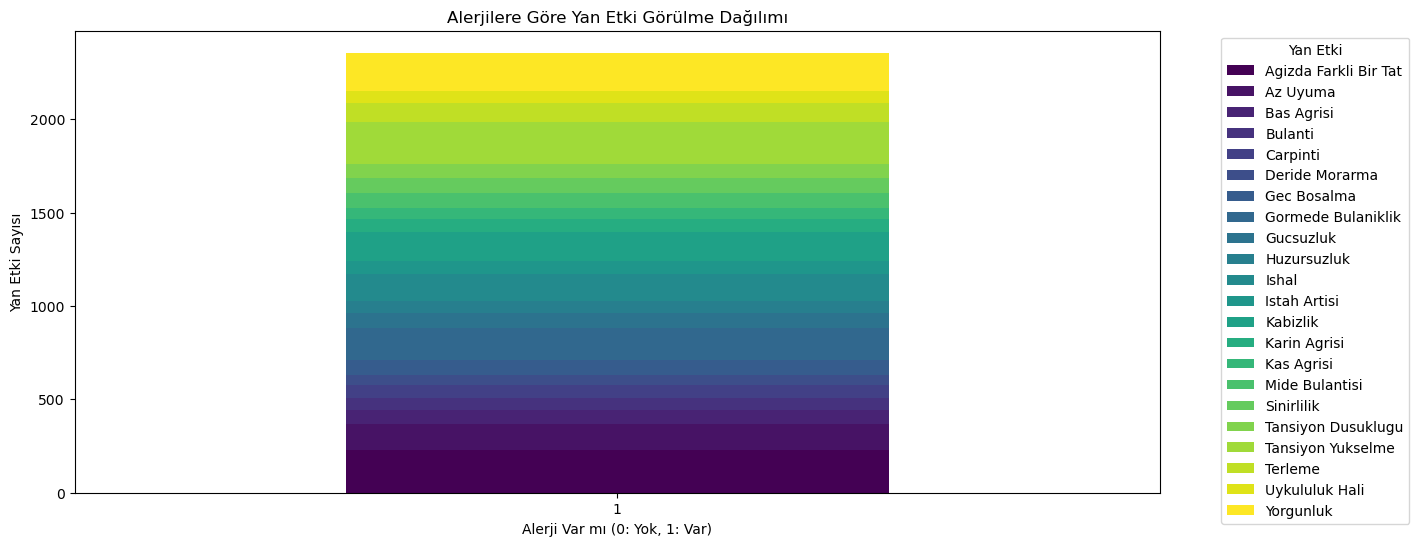
**Kronik Hastalıklar ve İlaç Kullanım Süresi ile Yan Etki:** Kronik hastalıkların ilaç kullanım sürelerine etkisi incelendi.



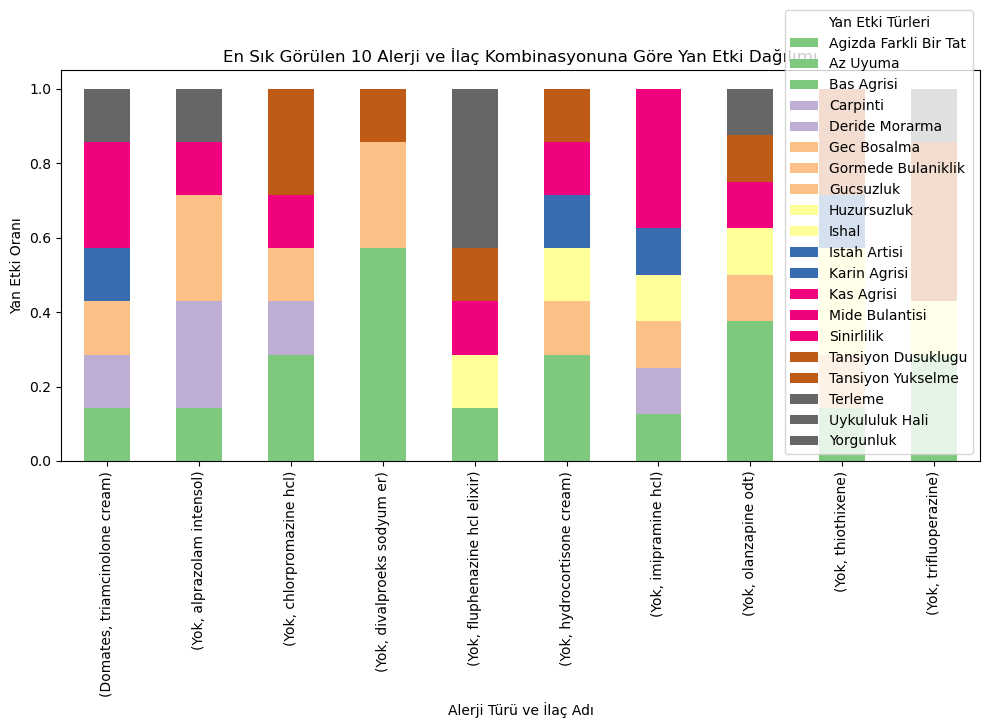
**3.5 Alerji ve Yan Etkiler**

**Alerjilere Göre Yan Etki Görülme Dağılımı:** Alerjilerin yan etkiler üzerindeki etkileri analiz edildi.



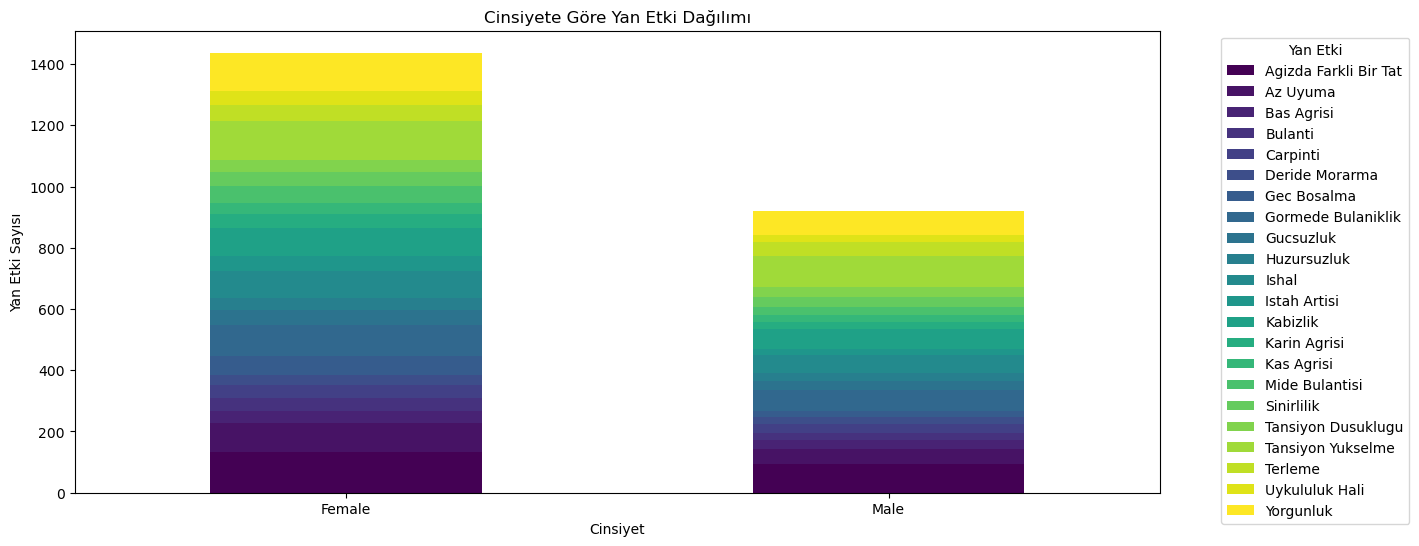


**En Sık 10 Alerji ve İlaç Yan Etkileri:** Belirli alerjilere sahip kullanıcıların yaşadığı yan etkilerin sıklığı belirlendi.

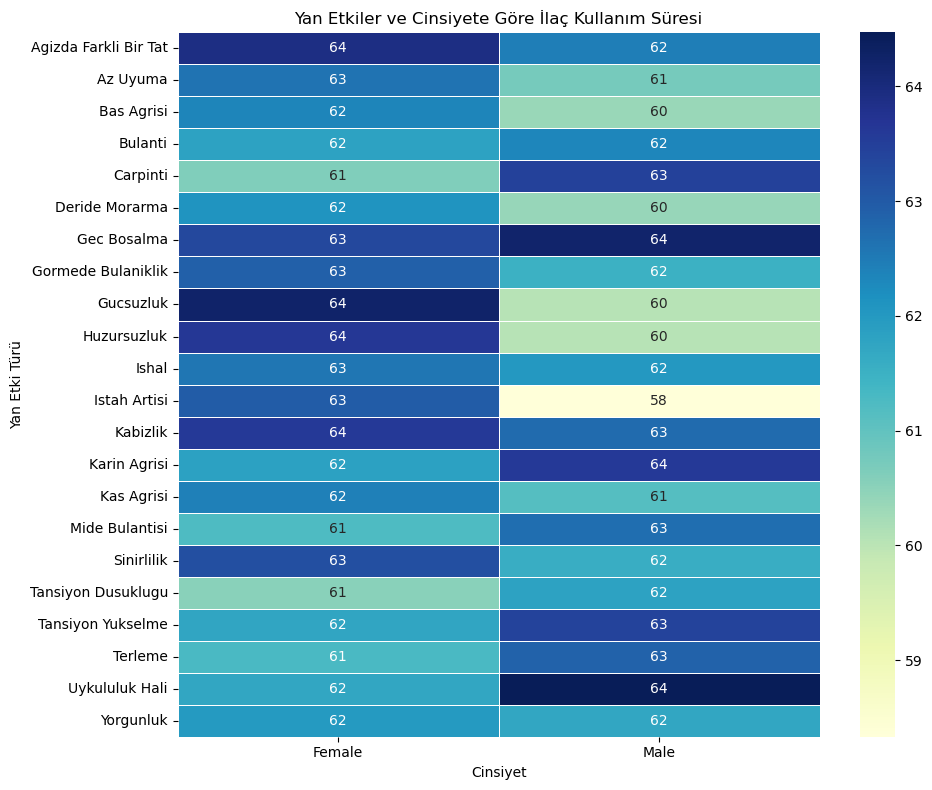


**3.6 Cinsiyet ve Yan Etkiler**

**Cinsiyete Göre Yan Etki Dağılımı:** Cinsiyetin yan etkilere etkisi analiz edildi.

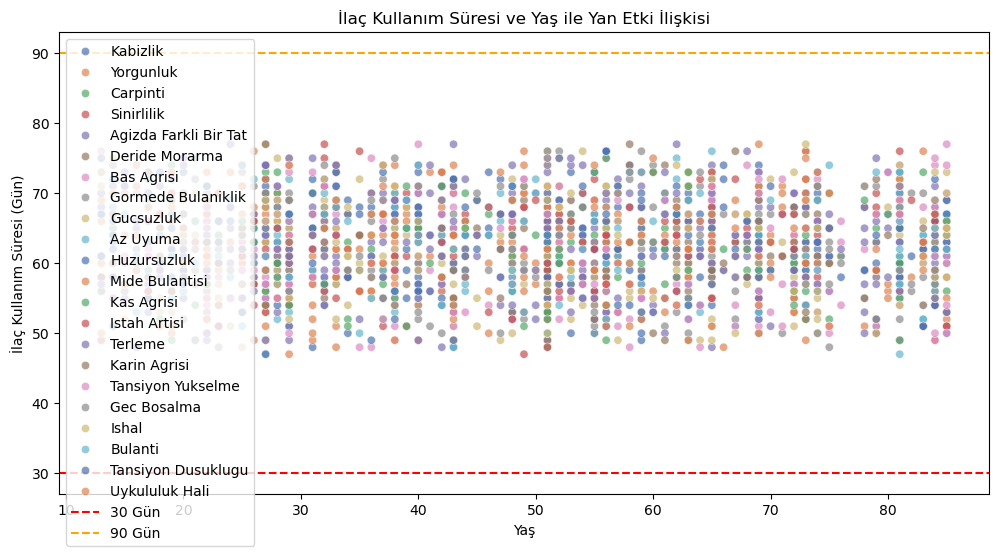


**Cinsiyete Göre Yan Etki ve İlaç Kullanım Süresi Dağılımı:** Cinsiyetin yan etkilere etkisi analiz edildi.

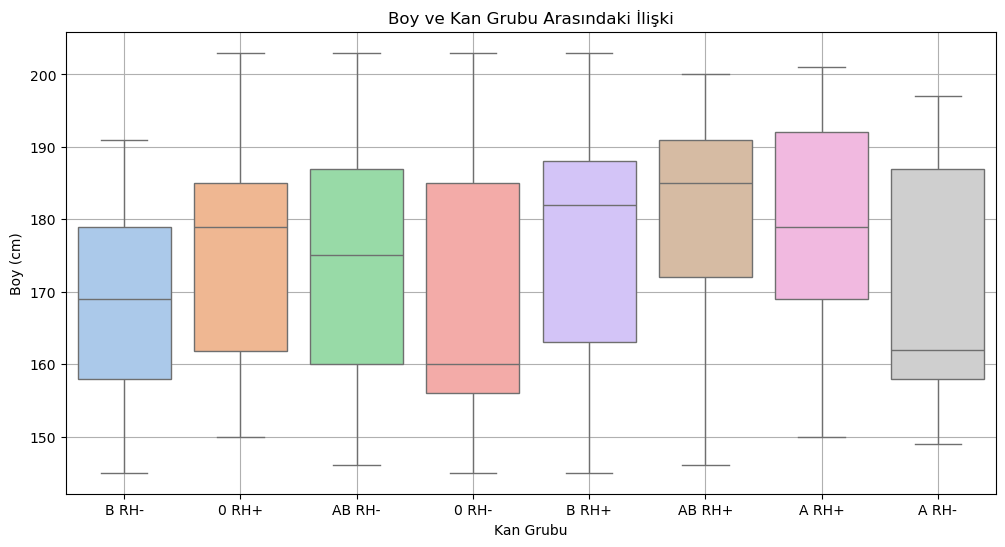


**3.7 Diğer Analizler**

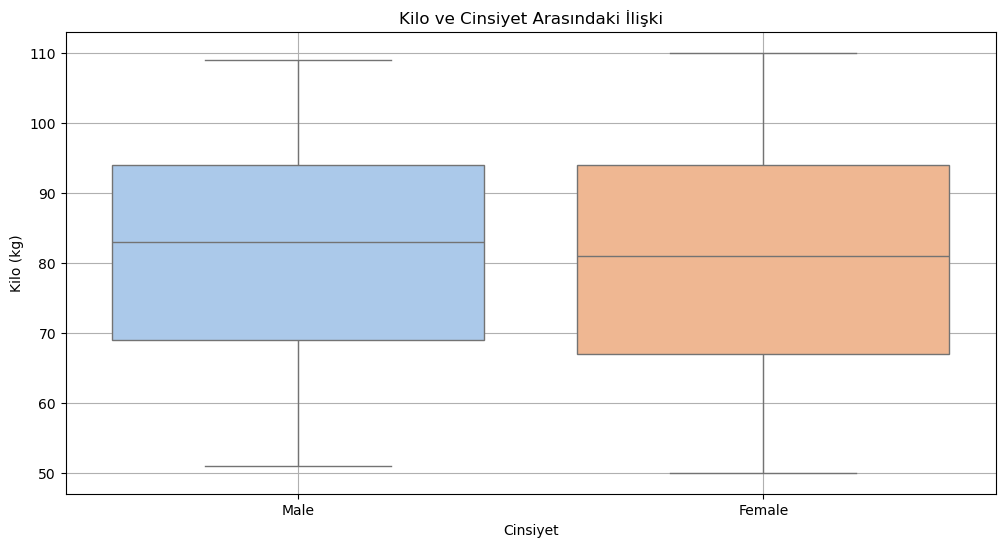
**İlaç Kullanım Süresi ve Yaş ile Yan Etki İlişkisi:** Yaşın ilaç kullanım süresi üzerindeki etkisi incelendi.



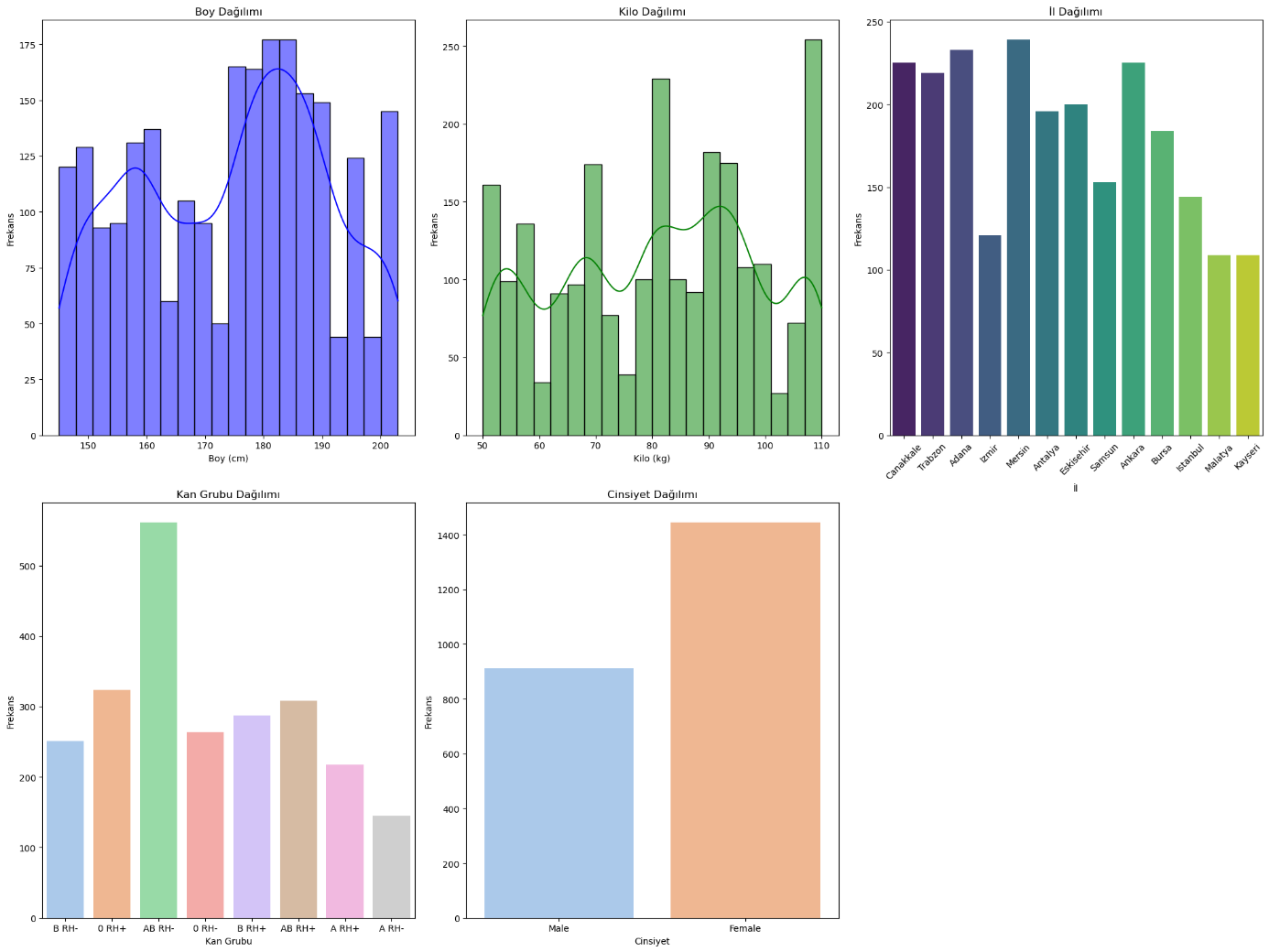
**Boy ve Kan Grubu Arasındaki İlişki:** Fiziksel özelliklerin ilişkisi analiz edildi.



**Kilo ve Cinsiyet Arasındaki İlişki:** Cinsiyetin kilo üzerindeki etkisi değerlendirildi.



**Dağılımlar:** Boy, kilo, il, kan grubu ve cinsiyet dağılımlarının görselleştirilmesi.



**4. Özet ve Bulgular**

EDA sürecinde elde edilen önemli bulgular:

İlaç ve Yan Etkileri: Chlordiazepoxide-amitriptyline, Lurasidone ve Hydrocortisone Cream en çok yan etkiye sahip 3 ilaç olarak belirlendi.

Yan Etki: Ağızda farklı bir tat, tansiyon yükselmesi ve yorgunluk en çok görülen yan etkiler olarak gözlemlendi.

Yan Etki ve İlaç Kullanım Süresi: İlaç kullanım süresine göre kabızlık, az uyuma ve ağızda farklı bir tat yan etkilerinin yüksek olduğu tespit edildi.

İlaç Kullanım Süresi: Kullanıcıların çoğunluğunun ilacı 30-90 gün arasında kullandığı görüldü.

Alerjilere Göre Yan Etki Görülme Dağılımı: Örneğin, havyar alerjisinin en çok ağızda farklı bir tat yan etkisi yaptığı görüldü.

**5. Veri İşleme Sonrası**

Veriler, eksik değerler giderildikten ve aykırı veriler düzeltildikten sonra analizler için hazır hale getirildi. Verilerin işlenmesi sonucu modelleme ve analiz için uygun bir formata dönüştürüldü.