

Mihriban Özdemir

mihriban.ozdemirr@outlook.com

Pusula Academy

Sağlık Veri Seti Üzerinde EDA ve Veri Ön İşleme

Bu doküman, side_effect_data veri seti üzerinde gerçekleştirilen **Exploratory Data Analysis (EDA)** ve veri ön işleme işlemlerini kapsamaktadır. Bu süreçte veri setindeki demografik bilgiler, ilaç kullanımı, kronik hastalıklar ve yan etki başlangıç tarihleri gibi bilgiler analiz edilmiştir. Ayrıca, hem kategorik hem de sayısal veriler üzerinde grafiksel analizler yapılmıştır.

Veri Yükleme ve İlk İnceleme

Veri seti yüklendikten sonra verinin genel yapısı incelenmiştir. Bu süreçte, veri türleri, eksik değerler ve veri dağılımı gibi temel bilgiler elde edilmiştir. Aşağıdaki çıktılar, veri setinin bu ilk aşamalarında alınan temel bulguların bir özetidir.

Veri Seti Genel Bilgileri:

- **Toplam Gözlem Sayısı:** 2357
- **Değişken Sayısı:** 19
- **Kategorik Değişkenler:** 6
- **Sayısal Değişkenler:** 5
- **Eksik Veriler:** Eksik veriler **missingno** kütüphanesi ile görselleştirilmiş ve eksik değerlerin hangi sütunlarda bulunduğu tespit edilmiştir

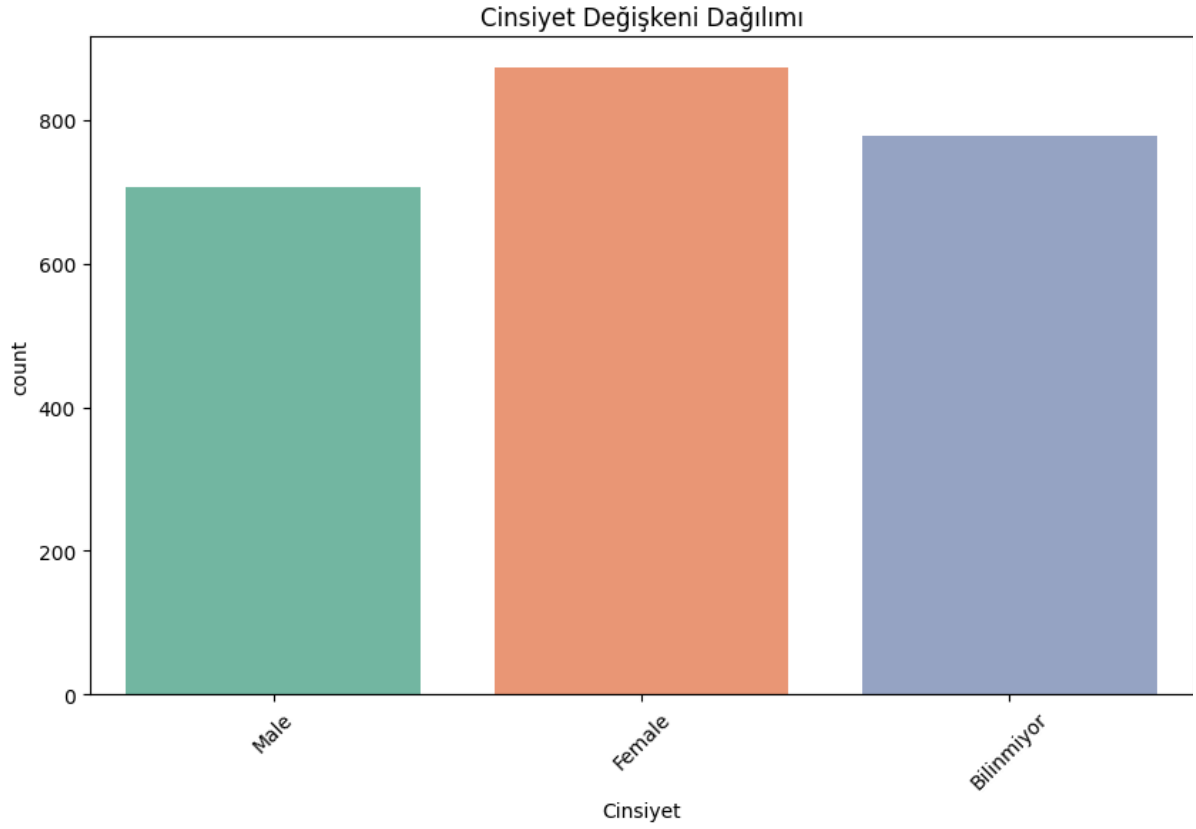
Kategorik Değişkenlerin Analizi

Kategorik değişkenlerin frekans dağılımı incelenmiş ve görselleştirilmiştir. Bu görseller üzerinden çıkarımlar yapılmıştır.

Cinsiyet Değişkeni Dağılımı

Cinsiyet değişkeninde kadın ve erkek oranları birbirine oldukça yakın olup, “Bilinmiyor” kategorisinde de belirli bir yoğunluk gözlemlenmiştir.

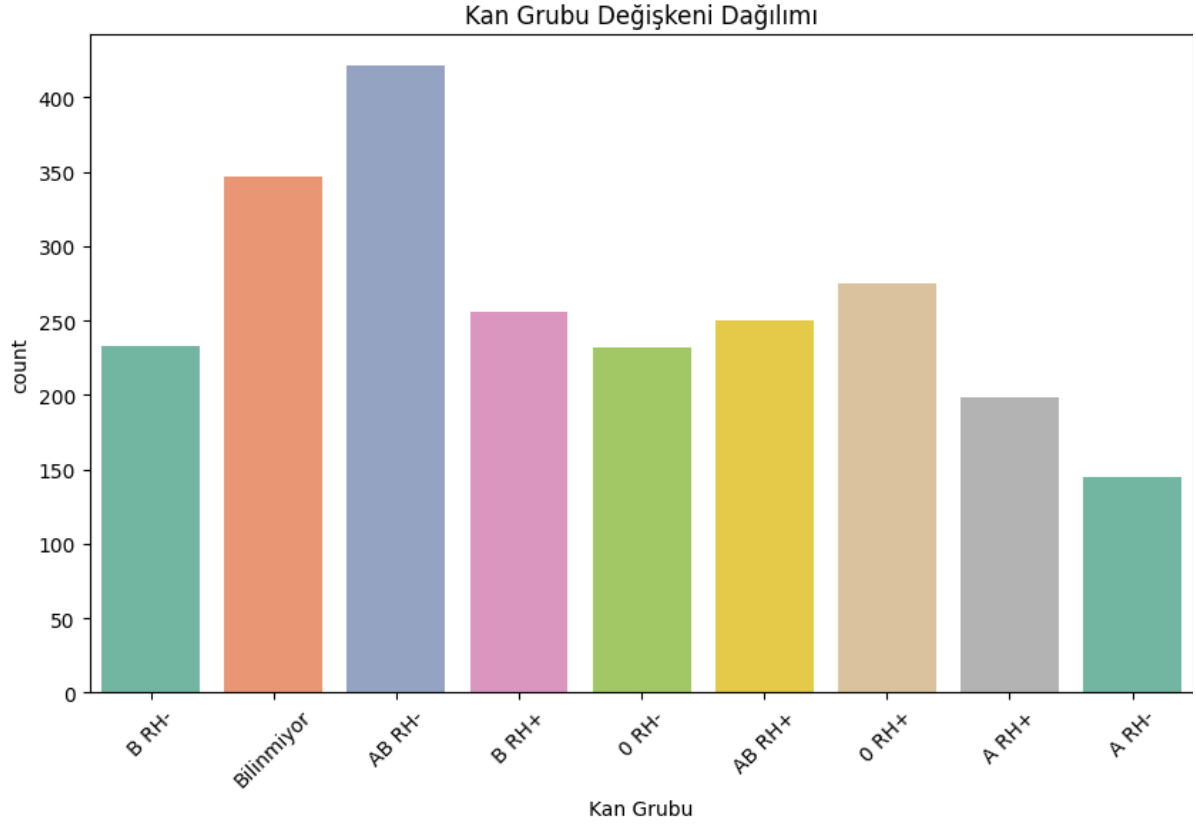
Cinsiyet dağılımı oldukça dengeli gözükmemektedir, bu yüzden cinsiyetin analizlerde dengeli bir değişken olduğu söylenebilir.



Kan Grubu Değişkeni Dağılımı

AB RH- grubu en az gözlemlenen kan grubu olurken, A RH+ ve 0 RH+ grupları en yaygın kan gruplarıdır.

Kan grubu dağılımları, genel toplumda gözlemlenen dağılımlarla paralellik göstermektedir.



3. Sayısal Değişkenlerin Analizi

Sayısal değişkenler üzerinde merkezi eğilim ölçütleri ve dağılımlar incelenmiştir.

Kilo Dağılımı

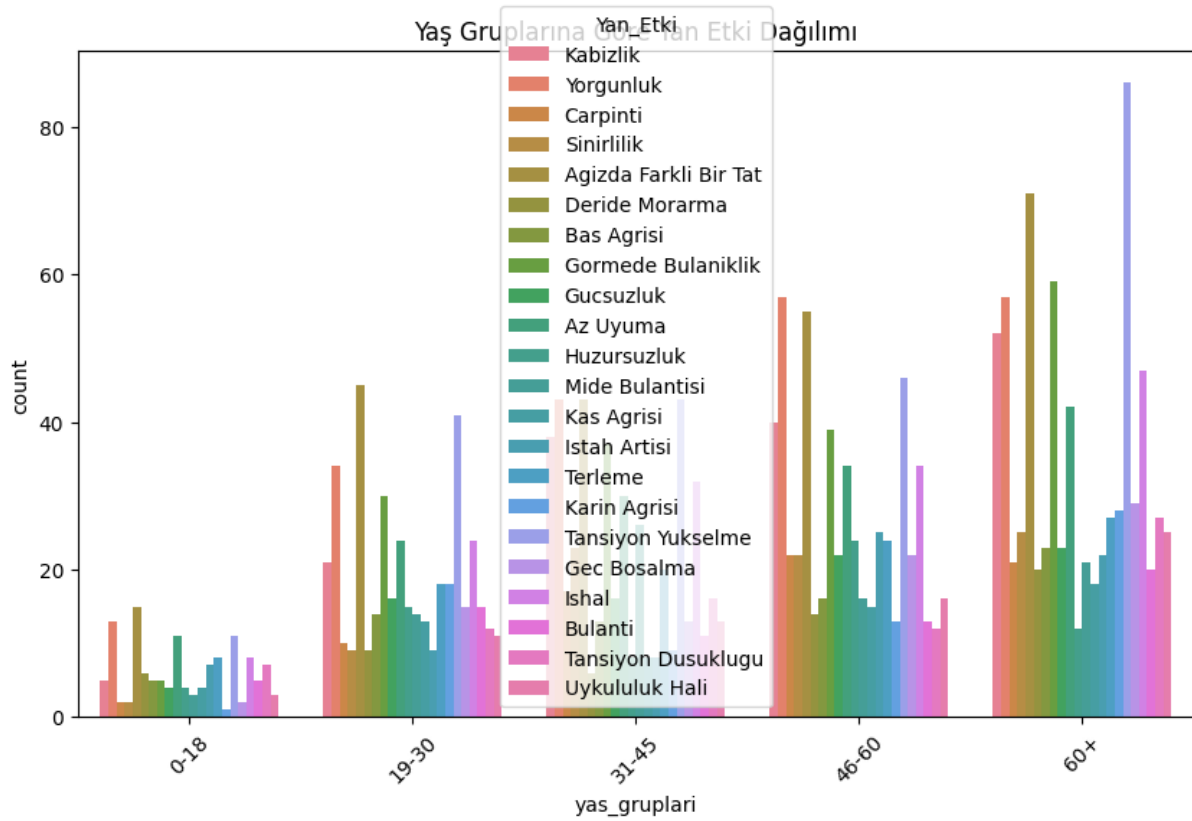
- **Dağılım:** Kilo değişkeni genel olarak 55-110 aralığında olup, 80 kg civarında bir tepe noktası bulunmaktadır.
- **Çıkarım:** Kilo değişkeni, büyük bir aralıkta dağılmıştır, ancak veri setinde aşırı uçlar bulunmaktadır.
- **Boy Dağılımı:** Boy değişkeni, 150-200 cm aralığında yoğunlaşmıştır. 170-180 cm aralığı, en fazla gözlemlenen gruba oluşturmaktadır.
- **Çıkarım:** Boy değişkeni, beklenen normal dağılıma sahip olup analizlerde güvenilir sonuçlar sunmaktadır.

Yaş Grupları ve Yan Etkilerin Gözlenmesi

- **0-18 yaş grubu:** Bu grupta yan etkilerin görülme sıklığı oldukça düşük. En sık karşılaşılan yan etkiler "Yorgunluk" ve "Sinirlilik" olarak öne çıkıyor.
- **19-30 yaş grubu:** Bu yaş grubunda özellikle "Sinirlilik", "Tansiyon Yükselmesi" ve "Yorgunluk" gibi yan etkiler sıkça karşılaşılan problemler arasında yer alıyor.

- **31-45 yaş grubu:** Bu grupta ise "Karin Ağrısı", "Huzursuzluk" ve "Yorgunluk" gibi yan etkiler daha fazla öne çıkıyor.
- **46-60 yaş grubu:** Bu grupta daha fazla yan etki gözlemleniyor, özellikle "Yorgunluk", "Tansiyon Yükselme" ve "Mide Bulantısı" gibi etkiler dikkat çekiyor.
- **60+ yaş grubu:** En fazla yan etkinin görüldüğü yaş grubudur. Özellikle "Ağızda Farklı Bir Tat", "Gücsüzlük", "Tansiyon Yükselmesi" ve "Yorgunluk" öne çıkan yan etkiler arasındadır.

Bu görsel, yaşın yan etkiler üzerinde önemli bir faktör olduğunu ve yaş grupları ilerledikçe bazı yan etkilerin daha sık görüldüğünü göstermektedir. Yan etkilerde “yorgunluk” her yaş grubunda yüksek bir değer içerdiğini söylemek mümkün.



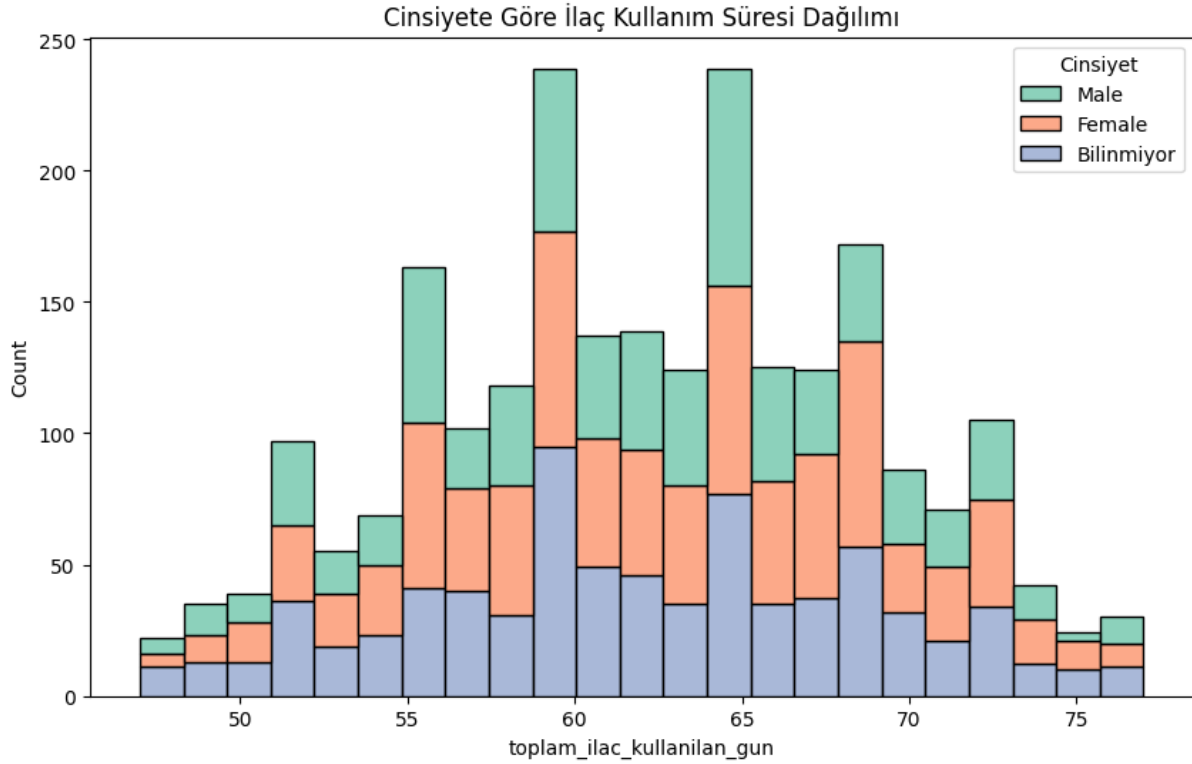
Cinsiyet ve İlaç Kullanım Süresi Dağılımı

Bu grafikte cinsiyete göre ilaç kullanım süresi dağılımı gösterilmektedir. Grafikte ilaç kullanım sürelerinin cinsiyetler arasındaki dağılımını inceleyebiliriz.

- **Erkekler (Male):** Genel olarak erkekler, 55-65 gün arasında daha fazla ilaç kullanma eğilimindedir. Özellikle 60 ve 65 günlük ilaç kullanım sürelerinde erkeklerin sayısında belirgin bir artış gözlemlenmektedir.
- **Kadınlar (Female):** Kadınlar arasında da ilaç kullanım sürelerinde artış gözlemlenmektedir, ancak erkeklere göre dağılım daha dengeli görünmektedir. Kadınların ilaç kullanım süresi daha çok 55-70 gün aralığında yoğunlaşmaktadır.

- **Bilinmeyen Cinsiyet (Bilinmiyor):** Cinsiyeti bilinmeyen kişilerin ilaç kullanım süreleri daha düşük günlerde (50-60 gün) yoğunlaşmakta olup, bu grup genel dağılımda daha az yer kaplamaktadır.

Grafikte dikkat çeken bir diğer nokta, hem erkek hem de kadınların ilaç kullanım süresinde 60 ve 65 günlerde büyük bir zirve görülmektedir. Bu, belirli bir tedavi sürecinin bu sürelerde tamamlandığını veya bu sürenin ilaç kullanımında kritik bir nokta olduğunu gösterebilir.

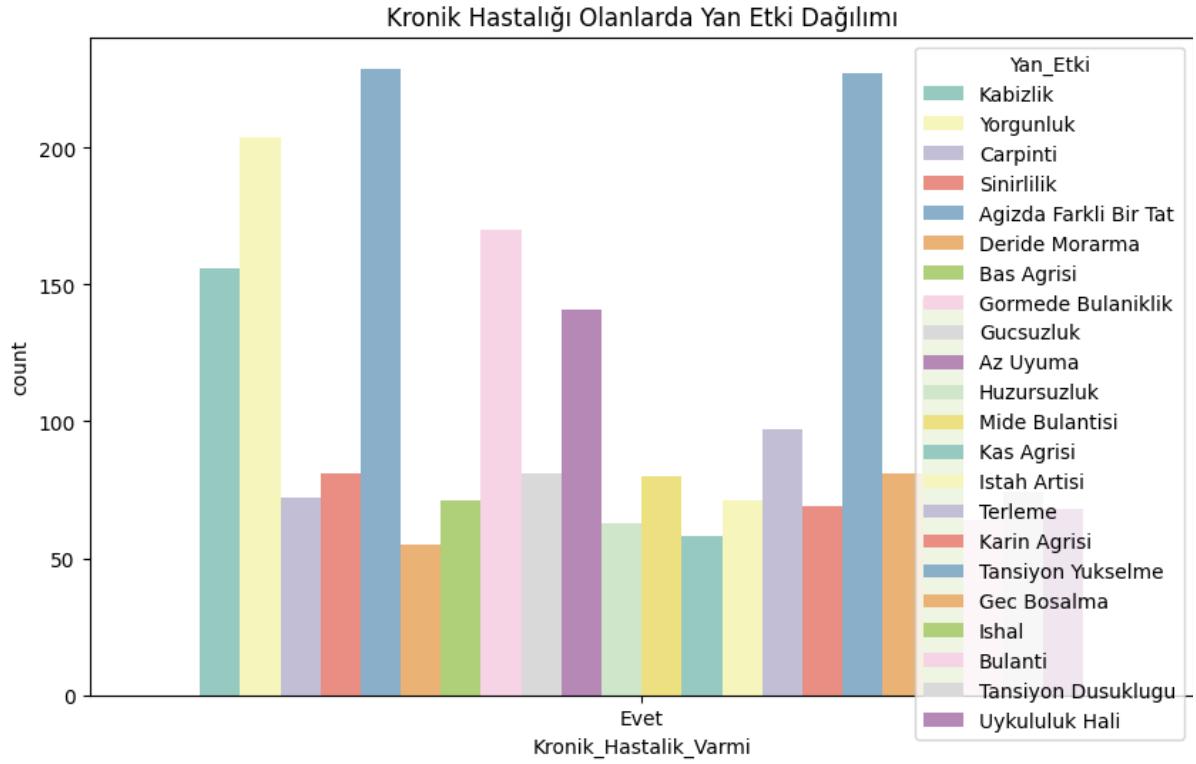


Kronik Hastalığı Olanlarda Yan Etki Dağılımı

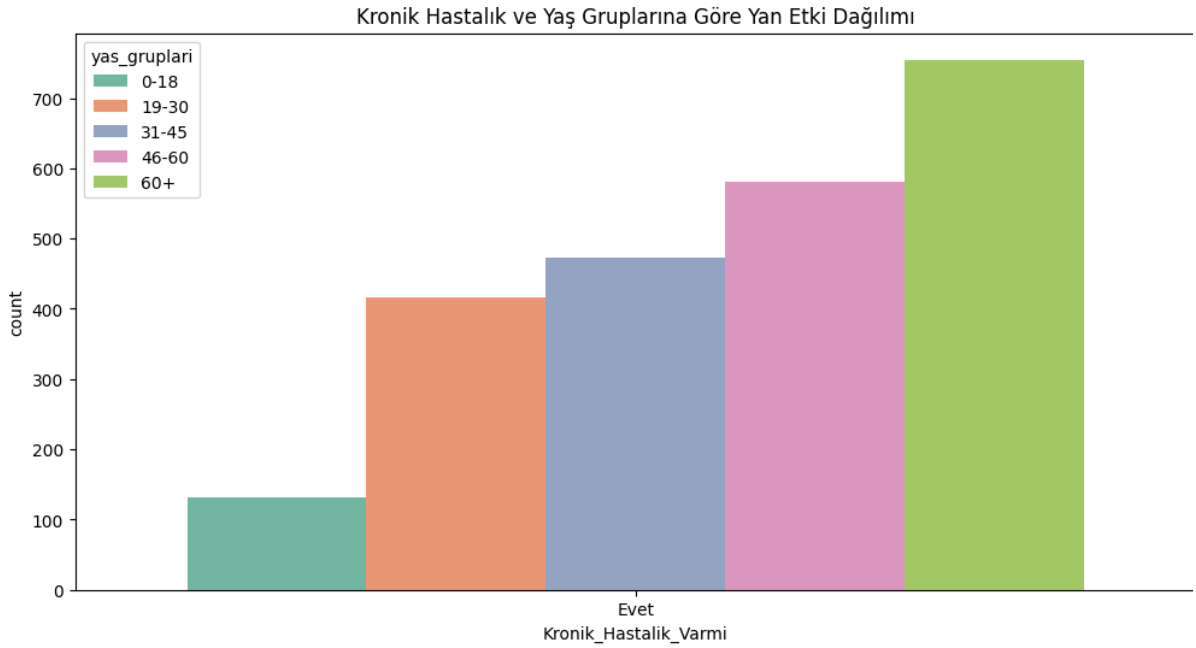
Bu grafikte kronik hastalığı olan bireylerde görülen yan etkilerin dağılımı gösterilmektedir.

- **Kabızlık, Yorgunluk ve Çarpıntı, Ağızda Farklı Bir Tat ve Tansiyon Yükselmesi:** Kronik hastalığı olan kişilerde en yaygın görülen yan etkiler kabızlık, yorgunluk ve çarpıntıdır. Kabızlık ve yorgunluk yan etkileri, yaklaşık 200'den fazla bireyde görülmektedir, çarpıntı ise en yaygın yan etki olarak öne çıkmaktadır.
- **Diğer Yaygın Yan Etkiler:** Sinirlilik, baş ağrısı ve mide bulantısı gibi yan etkiler de kronik hastalığı olan kişilerde sıkça karşılaşılan şikayetler arasındadır. Özellikle sinirlilik ve baş ağrısı belirli bir grupta oldukça yaygındır.
- **Daha Az Görülen Yan Etkiler:** Gecikmiş boşalma, ishal ve kas ağrısı gibi yan etkiler, daha az sayıda bireyde rapor edilmiştir.

Grafik, kronik hastalığa sahip bireylerde, özellikle çarpıntı gibi ciddi yan etkilerin daha yüksek oranda görüldüğünü ve bu bireylerin geniş bir yan etki yelpazesıyla karşılaştığını göstermektedir.



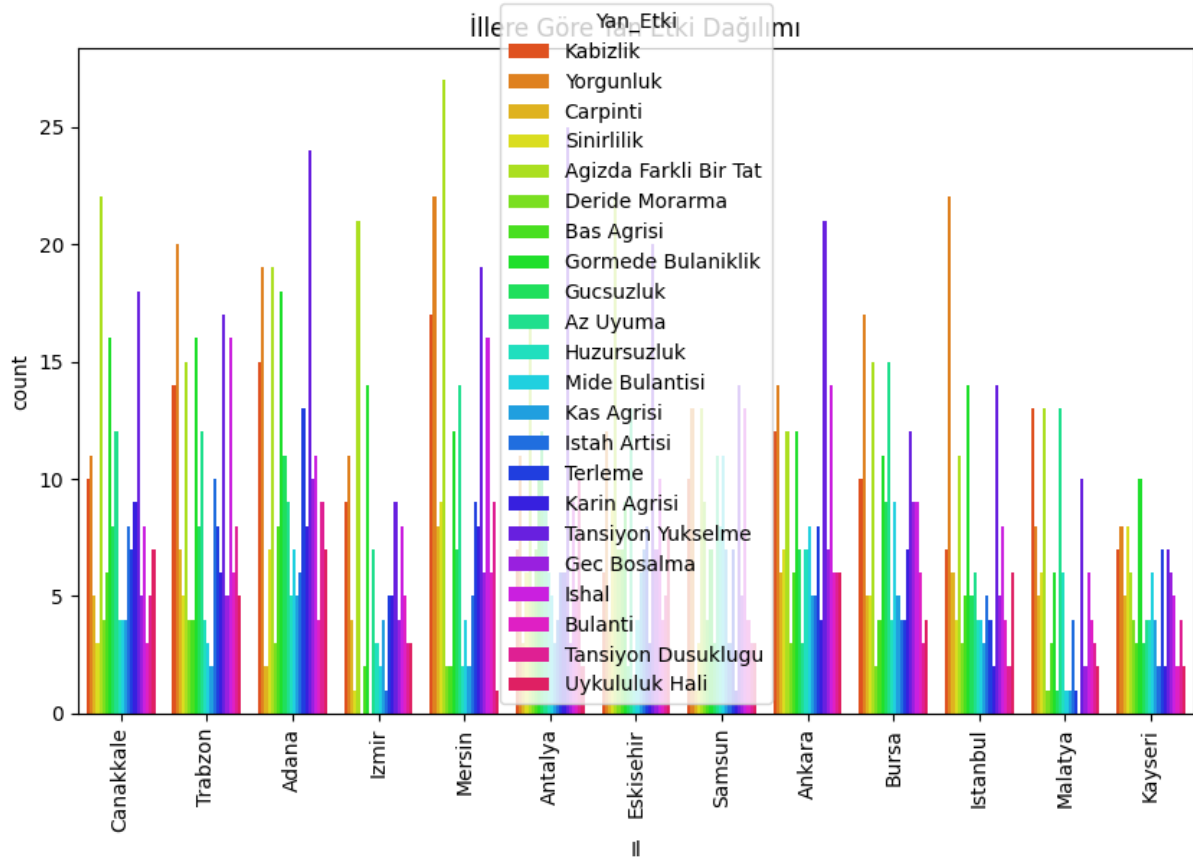
Aynı zamanda bu grafiği yaş gruplarına göre görselleştirdiğimizde kronik hastalıkların yaş ile doğru orantıda arttığını gözlemlemiş olduk.



İllere Göre Yan Etki Dağılımı

Farklı illerde, çeşitli yan etkilerin görülme sıklığı analiz edilmiştir.

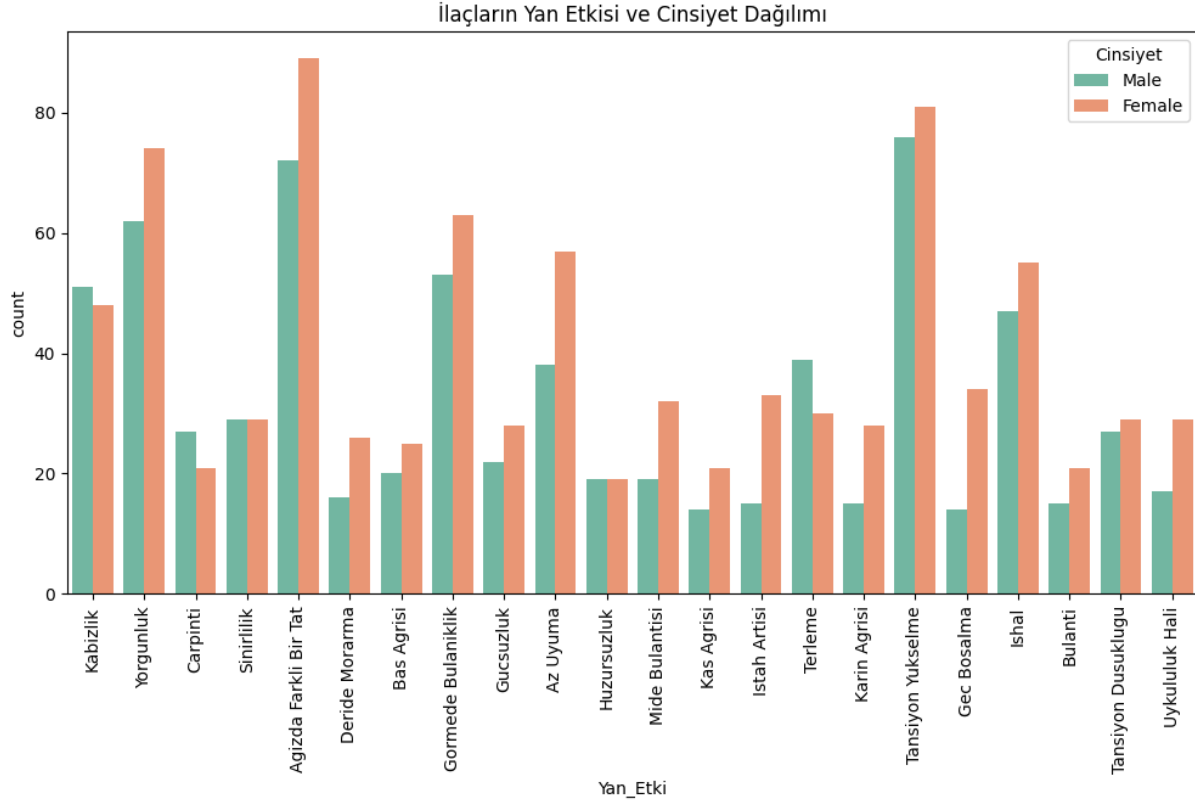
- Ağızda farklı bir tat, yorgunluk Özellikle Mersin’de en yüksek sıklıkta görüldüğü gözlemlenmiştir.
- Tansiyon yükselmesi en çok Adana, Ankara ve Çanakkale’de öne çıkmaktadır.:
- Daha Az Görülen Yan Etkiler: İshal, tansiyon düşüklüğü ve uykululuk hali gibi yan etkiler genellikle daha az sayıda rapor edilmiştir.
- Genel olarak grafiğe baktığımızda ise sinirlilik, baş ağrısı, tansiyon yükselmesi, kabızlık en çok görülen yan etkilerdendir yorumu yapılabilir.



İlaçların Yan Etkisi ve Cinsiyet Dağılımı

- Yorgunluk, ağızda farklı bir tat, görmede bulanıklık, az uyuma gibi yan etkilerde kadınlarda daha fazla vaka görülmekte.
- Karın Ağrısı, Tansiyon Yükselme ve Baş Ağrısı gibi yan etkiler de yine kadınlarda daha yaygın.
- Kabızlık, terleme gibi bazı yan etkilerde ise erkeklerde daha fazla vaka tespit edilmiş.
- Huzursuzluk ve sinirlilik gibi yan etkilerde erkekler ve kadınlar arasında neredeyse eşit bir dağılım var.

Bu grafik, cinsiyete bağılı olarak bazı yan etkilerin daha fazla görüldüğünü ortaya koyuyor. Özellikle belirli ilaçların yan etkilerinin kadınlarda veya erkeklerde daha sık görülmesi, cinsiyet farklılıklarının ilaç yan etkileri üzerindeki etkilerini anlamak için önemli olabilir.

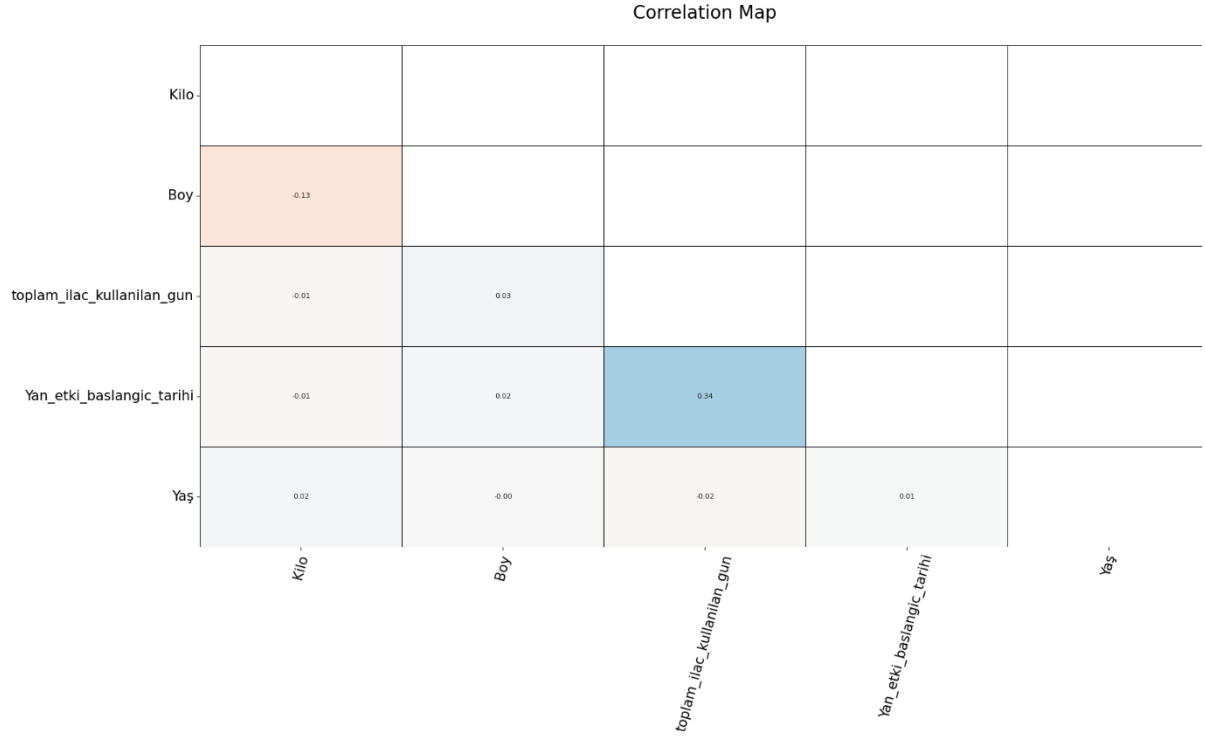


Korelasyon Haritası

Boy ve Kilo arasında pozitif bir korelasyon var. Korelasyon katsayısı 0.43, yani boy ve kilo birbirine bağılı değişkenlerdir. Uzun boylu bireyler genellikle daha yüksek kiloya sahip olma eğilimindedir.

Toplam İlaç Kullanım Günü ve Yan Etki Başlangıç Tarihi arasında 0.34'lük pozitif bir korelasyon var. Bu, ilacın uzun süre kullanılmasıyla birlikte yan etkilerin daha erken ortaya çıkabileceğini gösteriyor.

Diğer değişkenler arasında gözle görülür belirgin bir korelasyon bulunmamaktadır. Özellikle Yaş ve Yan Etki Başlangıç Tarihi, Toplam İlaç Kullanım Günü gibi değişkenler arasında anlamlı bir korelasyon gözlemlenmemiştir.



İlaç - Cinsiyet Kırılımında Ortalama Yaş, İlaç Kullanım Süresi ve Yan Etki Süresi

- **Levominlacipran** için cinsiyetler arasında ilaç kullanım süresi ve yan etki başlangıç sürelerinde dikkate değer farklılıklar yok gibi görünüyor, ancak kadınlarda yan etki daha geç başlıyor ve kullanım süresi erkeklere kıyasla hafifçe daha kısa.
- **Mometasone furoate** için yan etki başlangıç tarihleri açısından belirgin bir fark gözlenmemektedir, ancak kadınlarda bu ilacın kullanım süresi erkeklere kıyasla daha uzun.

Her iki ilaç için de cinsiyetler arasında büyük farklar olmamakla birlikte, ilaç kullanım süresi açısından kadınlar genellikle daha uzun süre ilaç kullanmaktadır. Yan etki başlangıç tarihleri açısından ise kadınlar için bir miktar daha geç yan etkiler ortaya çıkabilmektedir, ancak bu farklar genellikle küçük düzeyde.

		Yan_etki_baslangic_tarihi	Yaş	toplam_ilac_kullanilan_gun
Ilac_Adi	Cinsiyet			
Levomilnacipran	Bilinmiyor	33.62500	52.62500	61.87500
	Female	38.28571	52.71429	65.28571
	Male	33.22222	53.77778	65.88889
Mometasone furoate	Bilinmiyor	32.75000	44.50000	62.50000
	Female	32.50000	41.50000	59.50000
...
ziprasidone vial	Female	27.00000	48.40000	61.40000
	Male	27.33333	37.00000	63.00000
zolpidem tablet, sublingual	Bilinmiyor	34.87500	53.25000	65.75000
	Female	35.16667	57.66667	64.66667
	Male	33.66667	63.66667	56.00000

Cinsiyet- Yaş Kırılımında Ortalama İlaç Kullanım Süresi ve Yan Etki Süresi

- **Bilinmeyen cinsiyet grubundaki bireylerde**, özellikle genç yaş grubunda (13-22), yan etki başlangıç süreleri yaş ilerledikçe genellikle azalırken, ilaç kullanım süresi de benzer şekilde değişkenlik göstermektedir.
- **Erkek cinsiyet grubunda ise** daha yaşlı bireylerde (80 ve üzeri), yan etki başlangıç süreleri genellikle daha kısa (30-36 gün civarında) ve ilaç kullanım süreleri 60-65 gün arasında gözlemlenmiştir.

Bu veriler, cinsiyet ve yaş gruplarına göre yan etki ve ilaç kullanım süresindeki farklılıkları analiz etmek için bir başlangıç noktası sağlayabilir.

		Yan_etki_baslangic_tarihi	toplam_ilac_kullanilan_gun
Cinsiyet	Yaş		
Bilinmiyor	13	30.00000	64.00000
	19	38.00000	60.14286
	20	38.90909	65.09091
	21	34.33333	62.50000
	22	32.80952	59.52381
...
Male	80	34.41667	61.16667
	81	36.15385	64.46154
	83	35.14286	60.57143
	84	32.50000	59.83333
	85	34.38889	61.11111

Outlier Analizi

Veri setindeki sayısal değişkenlerde aşırı uç (outlier) değerlerin olup olmadığı incelenmiştir. outlier değer bulunmamıştır

Genel Sonuç

Bu çalışma kapsamında sağlık veri seti üzerinde gerçekleştirilen Exploratory Data Analysis (EDA) ve veri ön işleme işlemleri sonucunda, veri setindeki demografik bilgiler, ilaç kullanımı, kronik hastalıklar ve yan etki bilgileri detaylı bir şekilde incelenmiştir.

Veri analizinde elde edilen önemli bulgular:

- **Yaş gruplarına göre yan etkiler:** Yaş ilerledikçe bazı yan etkilerin daha sık görüldüğü, özellikle ileri yaş gruplarında tansiyon yükselmesi ve mide bulantısı gibi etkilerin arttığı gözlemlenmiştir.
- **Cinsiyet ve ilaç kullanım süresi:** Erkeklerde ilaç kullanım süresi 60 ve 65 günlerde zirve yaparken, kadınlarda ilaç kullanım süresi daha dengeli bir dağılım göstermektedir.
- **Kronik hastalıklar ve yan etkiler:** Kronik hastalığı olan bireylerde çarpıntı, kabızlık ve yorgunluk gibi ciddi yan etkilerin daha yaygın olduğu tespit edilmiştir.

- **İllere göre yan etkiler:** Özellikle büyük şehirlerde bazı yan etkilerin daha sık görüldüğü belirlenmiştir. Mersin'de ağızda farklı bir tat hissi ve yorgunluk yaygınken, Ankara ve Adana'da tansiyon yükselmesi öne çıkmaktadır.

Bu bulgular, sağlık sektöründe ilaç kullanımı ve yan etkilerin anlaşılması açısından önemli veriler sunmaktadır. İlaç kullanımı ve yan etkiler üzerine yapılan bu analiz, medikal karar alma süreçlerine ve kişiye özel tedavi yaklaşımlarına ışık tutabilecek potansiyele sahiptir. Bu doğrultuda, daha ileri çalışmalarla elde edilen verilerin detaylandırılması ve klinik açıdan yorumlanması, sağlık hizmetlerinin iyileştirilmesine katkı sağlayacaktır.