**IZV701**

**Veri Analitiğinin Temelleri**

**Vize Projesi**

**Sude Gül ÜZÜM**

**-6006165-**

**07.04.2028**

İçindekiler

**Özet**

**Giriş**

**Yöntem**

**Bulgular**

**Sonuç**

**1. Özet**

Bu rapor, **TÜİK**'ten alınan iki veri seti kullanılarak "Bireylerin gece yalnız yürürken kendini güvende hissetme oranı ile suç oranları arasındaki ilişkinin incelenmesini amaçlamaktadır.

* **Veri Setleri:**
  + **Güven Verisi (2011-2020):** Cinsiyet ve yıllara göre "Çok Güvenli" ile "Çok Güvensiz" arasındaki dağılım.
  + **Suç Verisi (2011-2020):** İllere göre cinayet, yaralama, hırsızlık gibi suç türlerinin sayıları.
* **Ana Bulgular:**
  + **Kadınlar**, erkeklere kıyasla daha yüksek oranda **"Güvensiz"** hissetmektedir.
  + **İstanbul, Ankara ve Bursa**, en yüksek suç oranlarına sahip illerdir.
  + Güven ve suç oranları arasında **zayıf pozitif korelasyon (r ≈ 0.3)** bulunmuştur.
* **Sonuç:** Suç oranları tek başına güven algısını açıklamakta yetersizdir; başka faktörler de etkilidir.

**2. Giriş**

**Amaç**

* Güven hissi ile suç istatistikleri arasındaki ilişkiyi analiz etmek.

**Kapsam ve Sınırlılıklar**

* **Kapsam:** 2011-2020 yılları arası Türkiye verileri.

**3. Yöntem**

**Veri Kaynakları**

1. **Güven Verisi:**
   * **Değişkenler:** Cinsiyet, Yıl, Güven Seviyesi (Çok Güvenli → Çok Güvensiz)
   * **Boyut:** 15 satır × 11 sütun
2. **Suç Verisi:**
   * **Değişkenler:** İl, Yıl, Cinsiyet, Suç Türleri (Cinayet, Yaralama, Hırsızlık vb.)
   * **Boyut:** 1.617 satır × 15 sütun

**Analiz Teknikleri**

* **Veri Temizleme:** Eksik değerler 0 ile dolduruldu.
* **Görselleştirme:**
  + **Boxplot** (Dağılım analizi)
  + **Korelasyon Matrisi** (Pearson)
  + **Haritalandırma** (sf paketi ile Türkiye il haritası)
* **İstatistiksel Testler:**
  + **Çarpıklık-Basıklık** (Normallik testi)
  + **Zaman Serisi Analizi**

Bu projede iki veri setini kullanarak analizi yapmak istedim. Araştırmalarım sonucu TÜİK’den aldığım iki excel tablosunu kendi projemin amacına uygun olarak satırları ve sütunların üzerinde temizlemeler yaparak csv formatlarına çevirdim.

Ana Veri setim; Cinsiyete göre bireylerin yaşadığı çevrede gece yalnız yürürken kendini güvende hissetme durumu, 2011-2020

Kısaca güven veri setim 2011 ve 2020 yılları arasında Çok Güvenli, Güvenli, Nötr, Güvensiz, Çok Güvensiz şeklinde ayrımları mecvut ve yıl bazında oranları hakkında bir veri seti.

T=Toplam

E=Erkek

K=Kadın olarak veri setimi düzenledim. Yaşadığımız çevrede gece yürürken kendimizi güvende hissetmeme nedenlerini düşündüm ve buna uygun veri setleri arayışına girdim. Birçok farklı nedenden dolayı kendimi güvende hissetmeyiz bunlardan birisi de yaşadığımız çevredeki suç oranlarının çok olmasıyla ilgili bir ilişkisi olabilir mi düşüncesi ile 3. Düzeyde, Suç Türü ve Suçun İşlendiği ile Göre Ceza İnfaz Kurumuna Giren Hükümlüler, 2011-2020 veri setini buldum ve kısaca güven veri setini düzenledim.

metin, ekran görüntüsü, sayı, numara, paralel içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Tablo 1: Güven veri seti

metin, sayı, numara, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, sayı, numara, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Tablo 2: Suç veri setindeki sütunlar

Suçun işlendiği il, yıl ve cinsiyetlere göre suç türlerinden oluşmakta.

Güven veri seti: 11 sütun ve 15 satırdan oluşmakta.

Suç veri seti: 15 sütun ve 1617 satırdan oluşmakta.

Güven veri setinde hiç boş değer yoktu. Fakat suç veri setinde boş değerler vardı bu boş satırları nasıl doldurmam gerektiğiyle ilgili düşünürken fark ettim ki o satırlar boş değer değil o suç türünden hükümlü birisinin olmadığı anlamına geliyor. İstatiksel hesaplama yaparken boş satırlar ve 0 değerinin yazması arasında farklar oluşacağına karar verdiğim için boş olan satırları 0 değerini girerek doldurdum. Bu şekilde verilerimi temizledim ve doldurdum.

Bu veri setlerine sormak istediğim soru: Bireyin yaşadığı çevrede gece yalnız yürürken kendini güvende hissetmeme oranı ile suç oranları arasında ilişki var mı, nasıl bir ilişki var?

metin, sayı, numara, yazı tipi, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 3: Güven veri seti için sürekli değişkenlerin dağılım analizi

**Çarpıklık**: Her bir grubun dağılımının simetriden ne kadar uzaklaştığını gösteren bir ölçüdür.

* Pozitif değerler dağılımın sağa çarpık olduğunu, yani kuyruğun sağ tarafa doğru uzadığını ve verilerin çoğunun daha düşük değerlerde toplandığını gösterir.

Bu veriler;

* Erkek - Notr: 1.114,
* Erkek - Cok\_Guvensiz: 0.425,
* Kadın - Guvenli: 0.124,
* Kadın –Notr: 1.858
* Negatif değerler dağılımın sola çarpık olduğunu, yani kuyruğun sol tarafa doğru uzadığını ve verilerin çoğunun daha yüksek değerlerde toplandığını gösterir.

Bu veriler;

* Erkek - Guvenli: -0.278,
* Erkek - Guvensiz: -0.541,
* Kadın - Cok\_Guvenli: -0.120,
* Kadın - Guvensiz: -1.159,
* Kadın - Cok\_Guvensiz: -0.203
* Sıfıra yakın değerler, dağılımın daha simetrik olduğunu düşündürür.

Sonuç: Bu tablo, farklı cinsiyet ve güven seviyelerindeki veri dağılımlarının temel şekil özelliklerini özetlemektedir. Çoğu grubun verisi normal dağılıma uygun görünse de, **Kadın - Notr** grubu normal dağılmamaktadır. Çarpıklık ve basıklık değerleri, dağılımların simetri ve "sivrilik" açısından farklılıklar gösterdiğini ortaya koymaktadır.

metin, diyagram, yazı tipi, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 4: Güven veri setinin zaman serisi

* **Çok Güvenli:**
* Erkeklerin "Çok Güvenli" diyen yüzdesi genel olarak düşüş göstermiş son yıllarda yükselmiş.
* Kadınların "Çok Güvenli" diyen yüzdesi de dalgalı bir seyir izlese de son yıllarda artmış gibi.
* **Çok Güvensiz:**
* Erkeklerin "Çok Güvensiz" diyen yüzdesi dalgalı bir şekilde seyretmiş, belirgin bir artış ya da azalış yok gibi.
* Kadınların "Çok Güvensiz" diyen yüzdesi de benzer şekilde dalgalı bir seyir izlemiş.
* **Güvenli:**
* Erkeklerin "Güvenli" diyen yüzdesi genel olarak artış göstermiş son yılda biraz düşmüş.
* Kadınların "Güvenli" diyen yüzdesi de dalgalı bir seyir izlese de son yıllarda artmış gibi.
* **Güvensiz:**
* Erkeklerin "Güvensiz" diyen yüzdesi genel olarak düşüş göstermiş.
* Kadınların "Güvensiz" diyen yüzdesi de dalgalı bir seyir izlese de son yıllarda azalmış gibi.
* **Nötr:**
* Erkeklerin "Nötr" diyen yüzdesi dalgalı bir seyir izlemiş, son yıllarda hafif bir düşüş olabilir.
* Kadınların "Nötr" diyen yüzdesi de dalgalı bir seyir izlemiş, özellikle son yılda belirgin bir artış olmuş.

**Suç Veri Seti Üzerine Analizler**

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 5: Öldürme suçu dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 6: Yaralama suçu dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 7: Cinsel suç dağılımı

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 8: Kişiyi hürriyetinden yoksun kılma suçu dağılımı

Bu grafiklere göre, " kişiyi hürriyetinden yoksun kılma" suçuyla ilgili bu sayısal veri, **çok sayıda düşük değere ve az sayıda yüksek değere sahip** ve bu veri **normal dağılım göstermiyor**.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 9: Hakaret suçu dağılımı

Bu grafiklere göre, "hakaret " suçuyla ilgili bu sayısal veri, **çok sayıda düşük değere ve az sayıda yüksek değere sahip** ve bu veri **normal dağılım göstermiyor**.

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 10: Hırsızlık suçu dağılımı

Bu grafiklere göre, "hırsızlık " suçuyla ilgili bu sayısal veri, **çok sayıda düşük değere ve az sayıda yüksek değere sahip** ve bu veri **normal dağılım göstermiyor**.

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 11: Uyuşturucu madde imal ve ticareti suçu dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, sayı, numara içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 12: Uyuşturucu madde kullanma ve satın alma suçu dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 13: kötü muamele suçu dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 14: Kaçakçılık suçu dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 15: Silah ve bıçaklı suçlar dağılımı

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 16: Tehdit suçları dağılımı

Bu grafiklerin tümüne baktığımızda sayısal veri, **çok sayıda düşük değere ve az sayıda yüksek değere sahip** ve bu veri **normal dağılım göstermiyor**.

Bu suç türleriyle ilgili sayısal veriler tipik bir "çan eğrisi" şeklinde dağılmıyor. Çoğu olayda değerler düşük çıkarken, çok daha yüksek değerlere sahip seyrek olaylar da bulunuyor.

metin, ekran görüntüsü, sayı, numara, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 17: Kategorik değişkenlerin frekansları

**Güven Derecesi Dağılımı (Cinsiyete Göre):** Erkekler ve kadınlar arasında güven derecelerinin nasıl farklılaştığını gösteriyor. Örneğin, erkeklerde "Güvenli" diyenlerin ortalama yüzdesi kadınlara göre daha yüksekken, "Çok Güvenli" diyenlerin ortalama yüzdesi de erkeklerde daha yüksek. Ancak "Güvensiz" diyenlerin ortalama yüzdesi kadınlarda daha yüksek gibi görünüyor (26.2'ye karşı 11.4).

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 18: Korelasyon matrisi

* **Mavi Tonları (Pozitif Sayılar):** İki güven derecesi arasında pozitif bir ilişki olduğunu gösterir. Yani, birinde artış olduğunda diğerinde de artış eğilimi vardır. Sayı 1'e yaklaştıkça bu ilişki güçlenir.
* **Beyaz/Açık Tonlar (Sıfıra Yakın Sayılar):** İki güven derecesi arasında zayıf veya neredeyse hiç ilişki olmadığını gösterir.
* **Kırmızı Tonları (Negatif Sayılar):** İki güven derecesi arasında negatif bir ilişki olduğunu gösterir. Yani, birinde artış olduğunda diğerinde azalma eğilimi vardır. Sayı -1'e yaklaştıkça bu ilişki güçlenir.

**Yorum**: Genel olarak, birbirine zıt güven seviyeleri arasında negatif ilişkiler, benzer güven seviyeleri arasında ise (daha zayıf olsa da) pozitif ilişkiler görülmektedir.

metin, ekran görüntüsü, diyagram, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

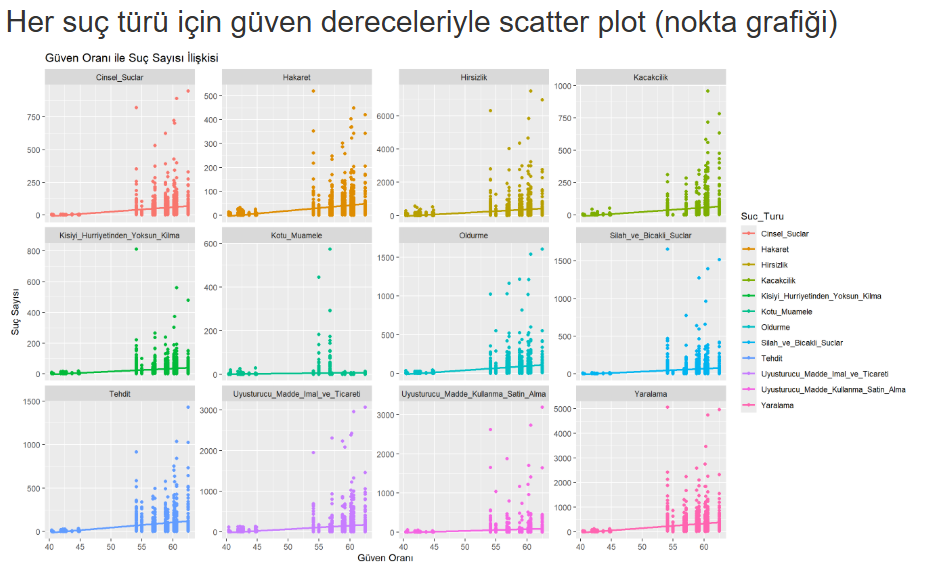
Görsel 19: Cinsiyetleri göre korelasyon matrisi

metin, ekran görüntüsü, yazı tipi, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 20: İllere göre toplam suç dağılımı

İl bazında toplam suç dağılımları koyu mavi yüksek suç sayısı açık renklere göre suç sayısı azlığını belirtir. En çok suçun İstanbul, Ankara, Antalya, Bursa, Adana diyebiliriz. Bu illerdeki gençler gece yalnız yürümekte güvende hissetmeyebilir.



Görsel 21: Suç türlerine göre güven dereceleri

**Yorum**: Bu grafiklerdeki trend çizgileri genellikle **yatay veya hafif yukarı/aşağı eğimli** görünüyor. Ayrıca, noktalar çizgi etrafında oldukça **dağınık** bir şekilde yayılmış durumda.

Bu suç türleri için "güven oranı" ile "suç sayısı" arasında **güçlü veya belirgin bir ilişki olmadığını** gözlemleniyor.



Görsel 22: Güven ve suç arasındaki korelasyon

Korelasyon değeri:

0.7 – 1.0 → güçlü pozitif

0.4 – 0.7 → orta pozitif

0.1 – 0.4 → zayıf pozitif

-0.1 – 0.1 → ilişki yok

-1.0 – -0.7 → güçlü negatif

Bizim korelasyon sonucumuza göre güven oranı ile suç sayısı arasında zayıf pozitiflik bulunmakta.

metin, ekran görüntüsü, sayı, numara, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Görsel 23: Yıllara göre ortalama güven oranı ile toplam suç sayısı arasındaki ilişki

**Ana Sonuç:**

Bu nokta grafiği bize şunu gösteriyor: **Ortalama güven skoru ile toplam suç sayısı arasında güçlü ve net bir ilişki görünmüyor.** Yıllar içinde hem güven skorunda hem de suç sayısında dalgalanmalar olmuş ve bu dalgalanmalar birbirini çok yakından takip etmemiş gibi duruyor.

Kırmızı çizgi, yüksek güven skorunun genel olarak biraz daha düşük suç sayısıyla ilişkili olabileceğine dair hafif bir işaret verse de, bu ilişki çok kuvvetli değil ve başka birçok faktörün de suç sayısını etkilediği anlaşılıyor.