

**IZV701**

**Veri Analitiğinin Temelleri**

**Vize Projesi**

**Sude Gül ÜZÜM**

**-6006165-**

**07.04.2028**

## İçindekiler

<b>Özet</b> .....	
<b>Giriş</b> .....	
<b>Yöntem</b> .....	
<b>Bulgular</b> .....	
<b>Sonuç</b> .....	

## 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 ( $r \approx 0.3$ )** 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 deęerler 0 ile dolduruldu.
- **Görselleřtirme:**
  - **Boxplot** (Daęılım 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. Arařtırmalarım 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 mevcut 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.

Güven_Derecesi	y2011	y2012	y2013	y2014	y2015	y2016	y2017	y2018	y2019	y2020
T_Cok_Guvenli	11.2	10.2	8.59	9.1	7.05	7.84	5.67	7.91	8.05	11.07
T_Guvenli	48.6	48.6	51.30	50.9	49.47	51.23	51.41	51.14	53.47	48.10
T_Notr	15.4	17.8	14.52	15.1	16.28	14.72	15.00	15.07	15.05	21.74
T_Guvensiz	19.2	18.3	18.46	17.8	20.62	20.42	21.84	19.99	18.67	13.97
T_Cok_Guvensiz	5.7	5.1	7.13	7.2	6.58	5.79	6.08	5.89	4.77	5.11
E_Cok_Guvenli	15.4	14.1	12.66	12.6	10.55	10.82	8.57	11.39	11.36	15.77
E_Guvenli	55.0	56.8	59.17	57.1	58.83	60.16	60.27	60.64	62.46	54.11
E_Notr	16.1	16.8	14.30	15.3	15.74	13.85	14.78	13.74	13.51	19.40
E_Guvensiz	11.7	10.1	10.85	11.5	11.94	12.97	13.87	12.17	10.68	8.07
E_Cok_Guvensiz	1.8	2.2	3.02	3.4	2.94	2.20	2.52	2.05	2.00	2.65
K_Cok_Guvenli	7.1	6.4	4.63	5.7	3.65	4.94	2.84	4.53	4.84	6.48
K_Guvenli	42.4	40.7	43.67	44.9	40.39	42.55	42.77	41.90	44.73	42.23
K_Notr	14.7	18.7	14.73	14.8	16.80	15.57	15.21	16.35	16.54	24.04
K_Guvensiz	26.4	26.2	25.85	23.8	29.05	27.67	29.62	27.60	26.44	19.74
K_Cok_Guvensiz	9.4	8.0	11.12	10.8	10.11	9.29	9.56	9.62	7.45	7.52

Tablo 1: Güven veri seti

Sucun_Islendigi_Il	Yil	Cinsiyet	Oldurme	Yaralama	Cinsel_Suclar	Kisiyi_Hurriyetinden_Yoksun_Kilma	Hakaret	Hirsizlik
İstanbul	2020	Erkek	1026	5055	820	811	520	6306
İstanbul	2020	Kadın	26	123	6	16	32	487
Tekirdag	2020	Erkek	71	453	76	60	46	509
Tekirdag	2020	Kadın	3	7	0	2	1	23
Edirne	2020	Erkek	39	260	39	28	28	226
Edirne	2020	Kadın	0	7	0	1	2	11
Kırklareli	2020	Erkek	22	221	36	21	16	193
Kırklareli	2020	Kadın	1	3	0	1	0	14
Balıkesir	2020	Erkek	101	939	110	95	85	649
Balıkesir	2020	Kadın	3	29	1	4	5	73
Canakkale	2020	Erkek	30	271	23	28	35	245
Canakkale	2020	Kadın	2	5	0	1	1	21

Uyusturucu_Madde_Imal_ve_Ticareti	Uyusturucu_Madde_Kullanma_Satin_Alma	Kotu_Muamele	Kacakcilik	Silah_ve_Bicakli_Suclar	Tehdit
1954	2621	14	310	1655	916
131	20	25	12	9	25
63	125	0	16	65	97
10	1	0	1	0	2
58	70	0	44	32	49
8	0	0	0	1	1
39	34	2	10	19	56
2	1	0	0	0	1
119	95	2	33	137	213
14	0	0	3	0	6
21	55	0	13	29	87

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?

Cinsiyet <chr>	Güven_Seviyesi <chr>	Çarpıklık <dbl>	Basıklık <dbl>	Shapiro_p <chr>	Normal_Dagilim <chr>	Çarpıklık_Yonu <chr>
Erkek	Cok_Guvenli	0.125	2.156	0.76383	Evet	Saga
Erkek	Guvenli	-0.278	2.035	0.78185	Evet	Sola
Erkek	Notr	1.114	3.573	0.1638	Evet	Saga
Erkek	Guvensiz	-0.541	3.149	0.86893	Evet	Sola
Erkek	Cok_Guvensiz	0.425	1.997	0.63611	Evet	Saga
Kadin	Cok_Guvenli	-0.120	2.137	0.87243	Evet	Sola
Kadin	Guvenli	0.124	2.131	0.61148	Evet	Saga
Kadin	Notr	1.858	5.449	0.0019018	Hayir	Saga
Kadin	Guvensiz	-1.159	3.933	0.14597	Evet	Sola
Kadin	Cok_Guvensiz	-0.203	1.898	0.39633	Evet	Sola

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;

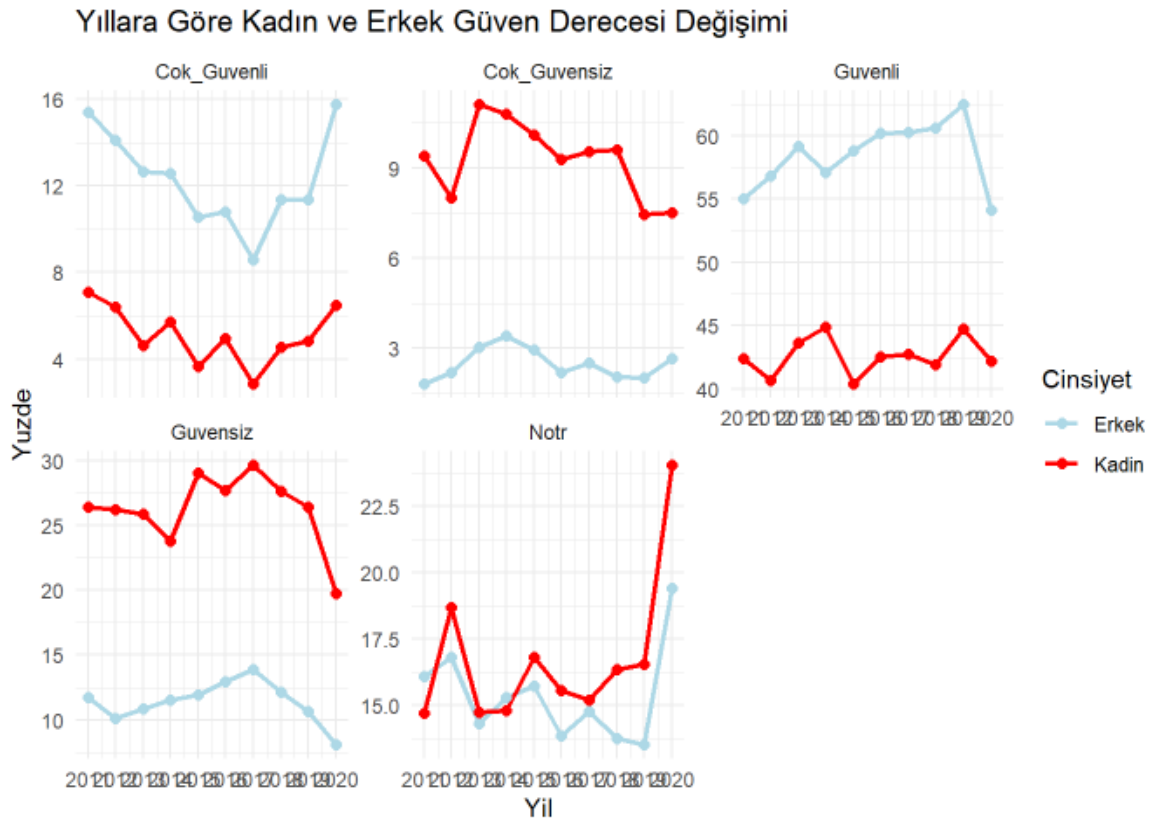
- Erkek - Guvenli: -0.278,
- Erkek - Guvensiz: -0.541,
- Kadın - Cok\_Guvenli: -0.120,
- Kadın - Guvensiz: -1.159,

➤ Kadın - Çok\_Guvenli: -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.

## Zaman Serisi Grafikleri



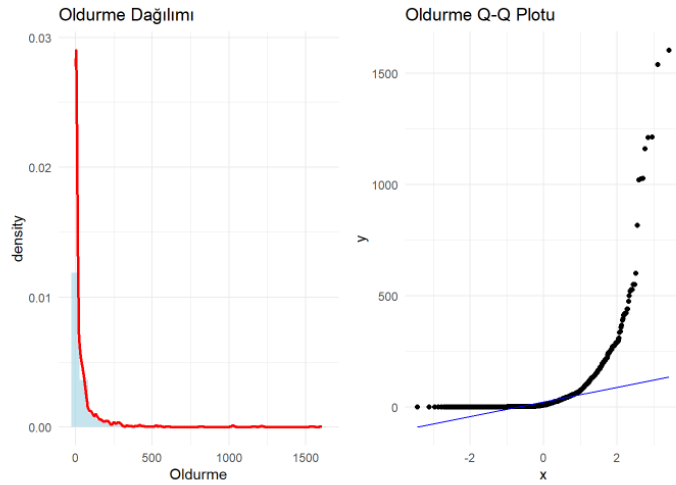
Görsel 4: Güven veri setinin zaman serisi

- **Çok Güvenli:**
  - Erkeklerin "Çok Güvenli" diyen yüzdesi genel olarak düşüş göstermiş son yıllarda yükselmiştir.
  - 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.

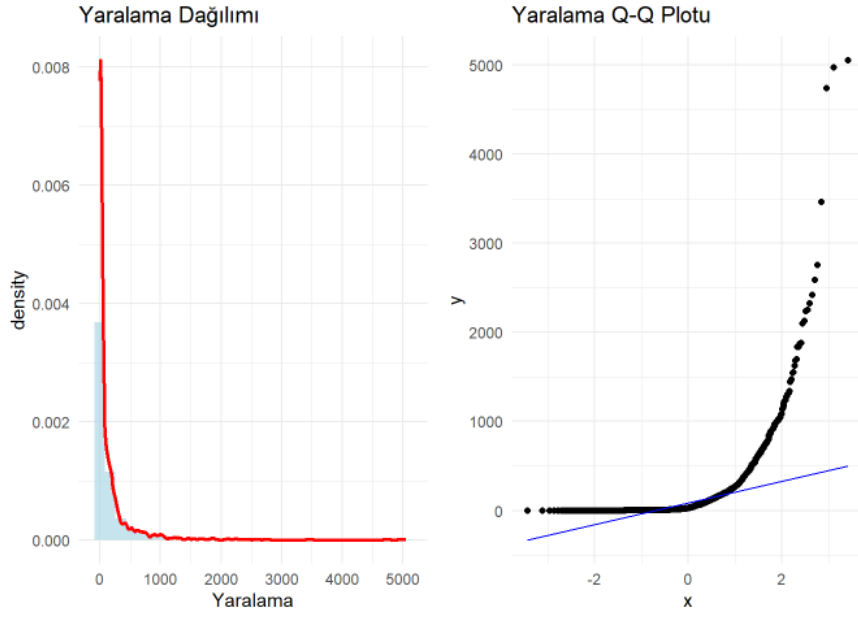
## Suç Veri Seti Üzerine Analizler

Her Bir Suç Türü için Sürekli Değişkenlerin Dağılım ve Q-Q Plot Analizi

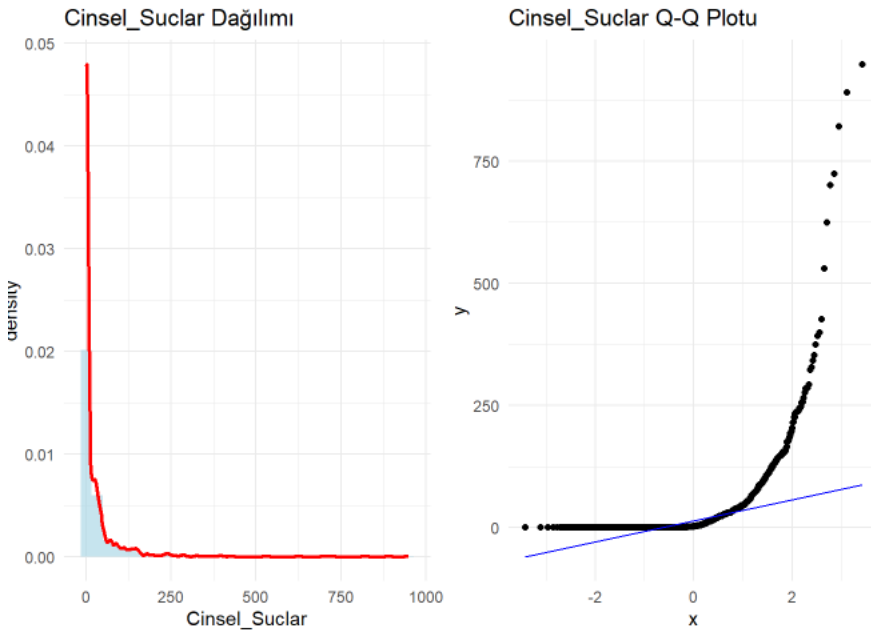


Görsel 5: Öldürme suçu dağılımı

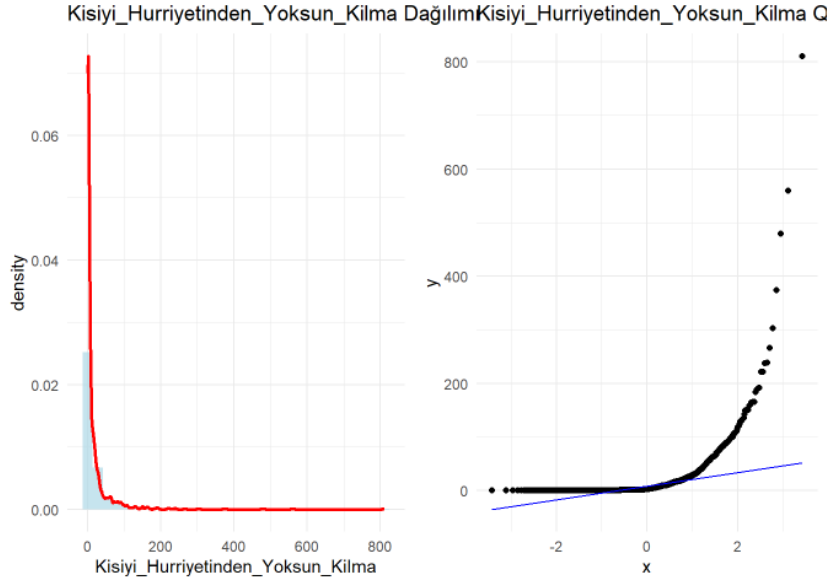




Görsel 6: Yaralama suçu dağılımı

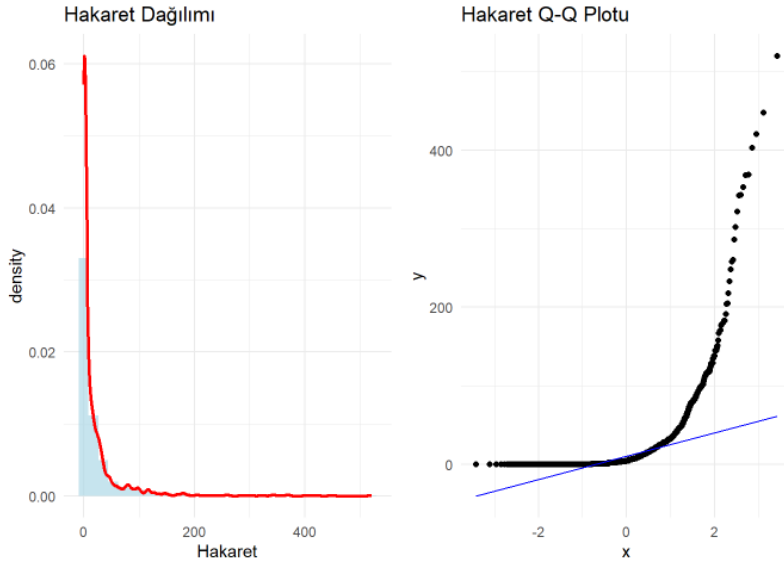


Görsel 7: Cinsel suç dağılımı



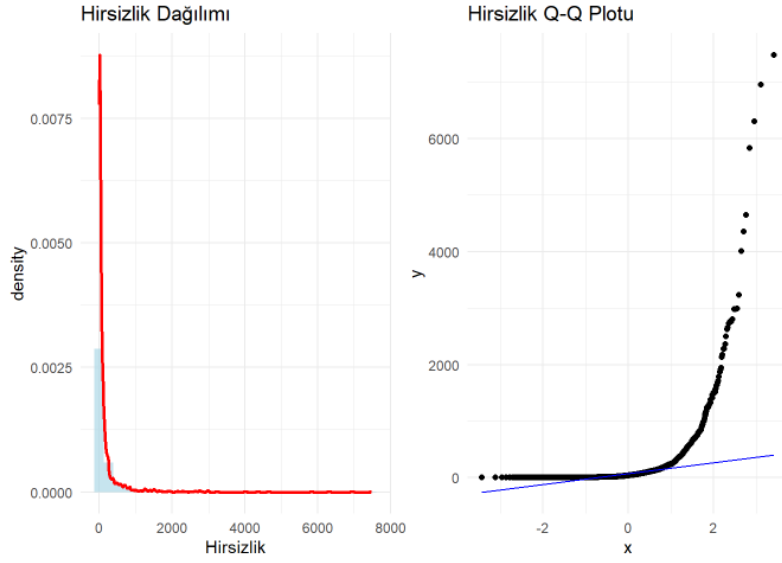
Görsel 8: Kişiyi hürriyetinden yoksun kılma suçu dağılımı

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**.



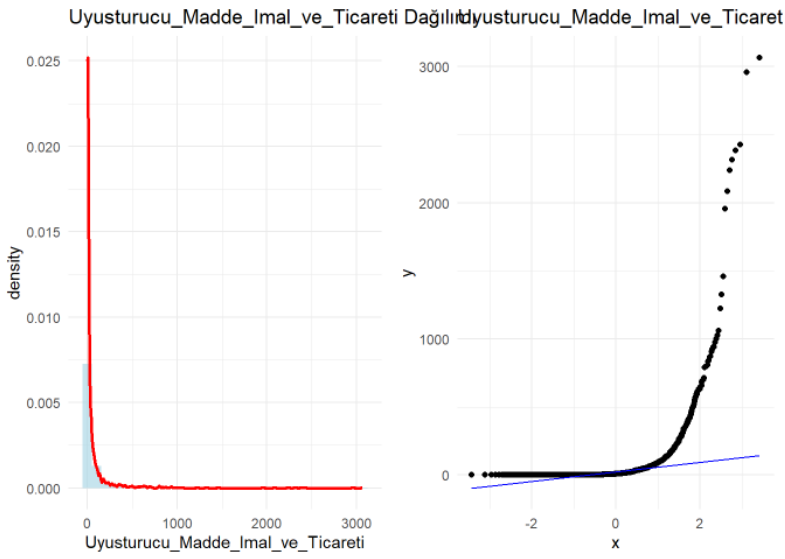
Görsel 9: Hakaret suçu dağılımı

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**.

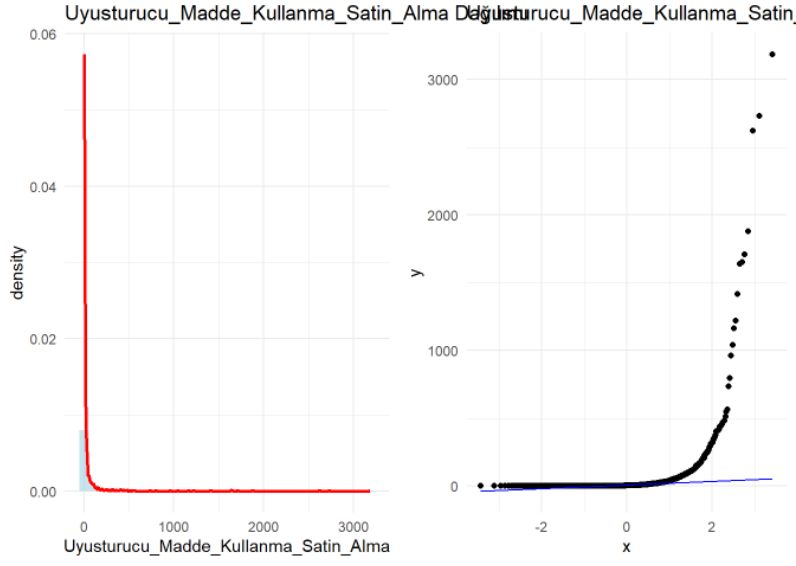


Görsel 10: Hırsızlık suçu dağılımı

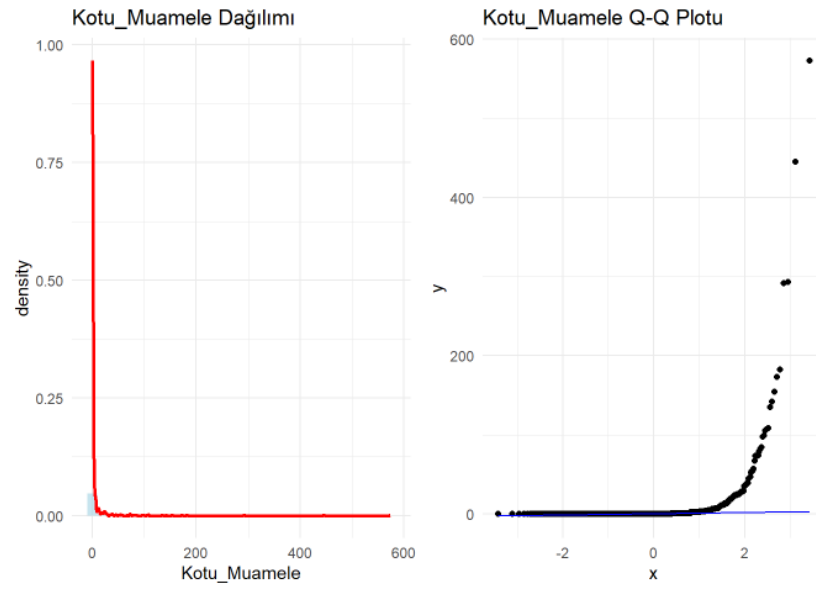
Bu grafiklere göre, "hırsızlık " suçuyla ilgili bu sayısal veri, **çok sayıda düşük değere ve az sayıda yüksek değere sahip** ve bu veri **normal dağılım göstermiyor**.



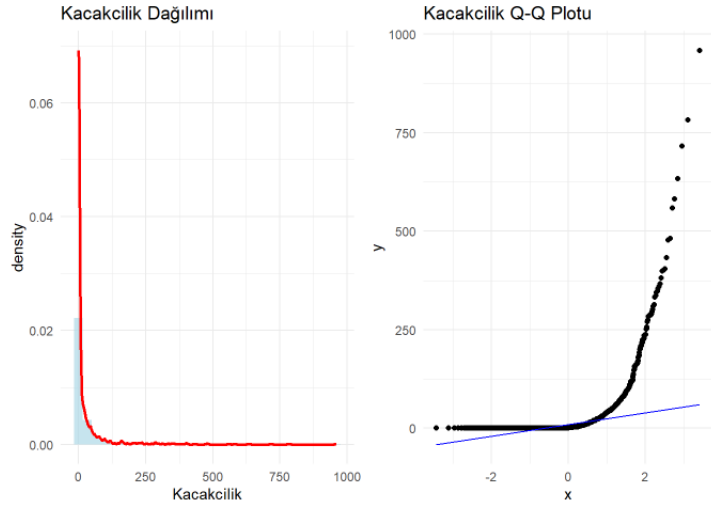
Görsel 11: Uyuşturucu madde imal ve ticareti suçu dağılımı



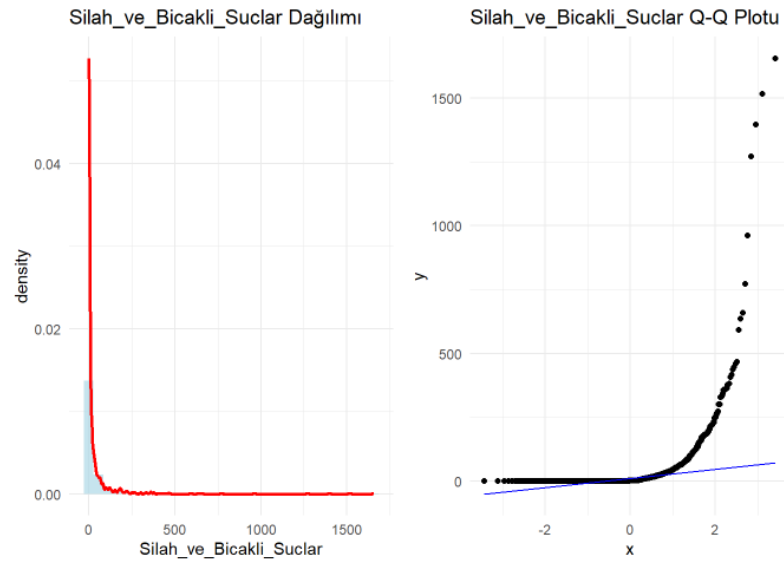
Görsel 12: Uyuşturucu madde kullanma ve satın alma suçu dağılımı



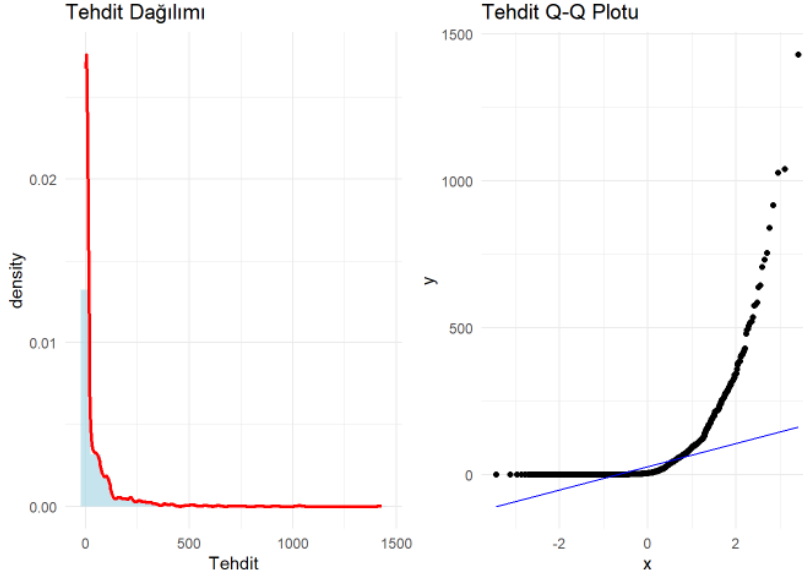
Görsel 13: kötü muamele suçu dağılımı



Görsel 14: Kaçakçılık suçu dağılımı



Görsel 15: Silah ve bıçaklı suçlar dağılımı



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.

## Kategorik Değişkenlerin Frekanslarını Hesaplama

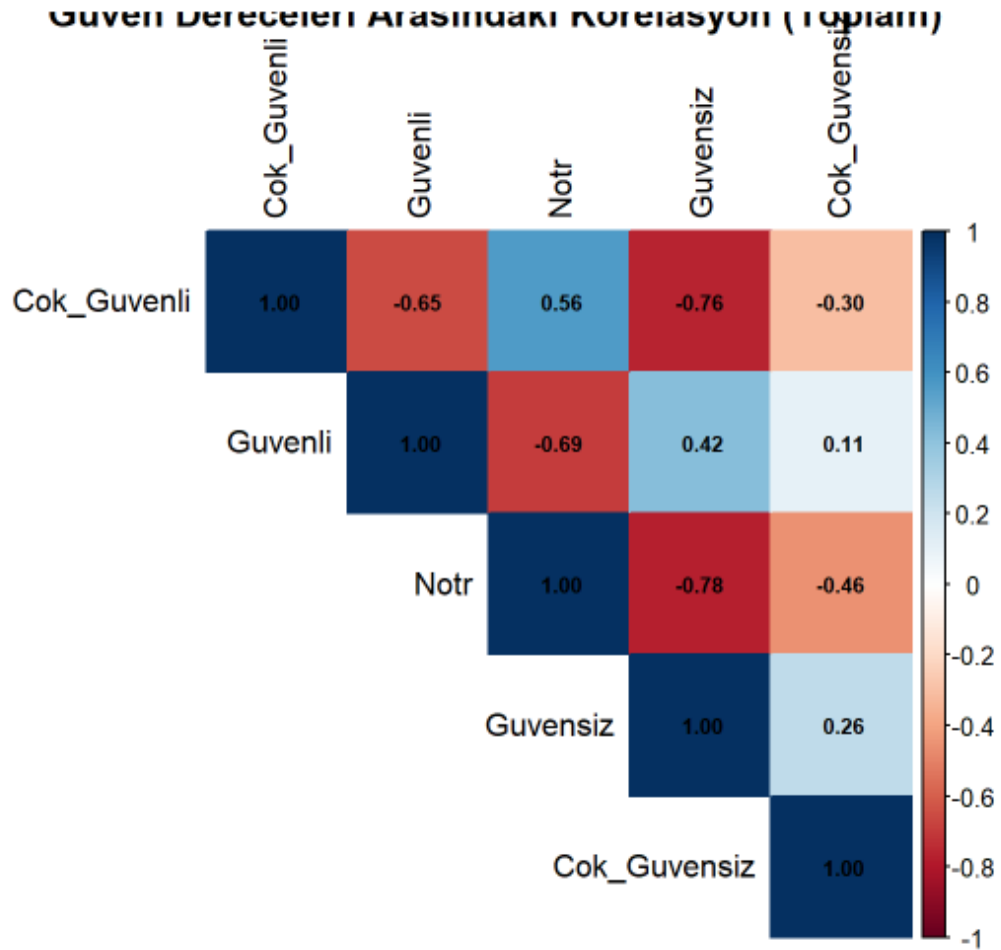
```
## # A tibble: 15 x 4
##   Cinsiyet   Guven_Derecesi Ortalama_Yuzde Std_Dev
##   <chr>      <chr>             <dbl>   <dbl>
## 1 "Erkek"    Cok_Guvenli        12.3    2.26
## 2 "Erkek"    Cok_Guvensiz        2.48    0.518
## 3 "Erkek"    Guvenli             58.5    2.65
## 4 "Erkek"    Guvensiz            11.4    1.60
## 5 "Erkek"    Notr                15.4    1.79
## 6 "Kad\u0131n" Cok_Guvenli         5.11    1.33
## 7 "Kad\u0131n" Cok_Guvensiz        9.29    1.28
## 8 "Kad\u0131n" Guvenli         42.6    1.50
## 9 "Kad\u0131n" Guvensiz        26.2    2.82
## 10 "Kad\u0131n" Notr         16.7    2.85
## 11 "Toplam"  Cok_Guvenli         8.67    1.76
## 12 "Toplam" Cok_Guvensiz        5.94    0.835
## 13 "Toplam" Guvenli            50.4    1.68
## 14 "Toplam" Guvensiz         18.9    2.14
## 15 "Toplam" Notr            16.1    2.21
```

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).

## Güven Derecesi Korelasyon Matrisi

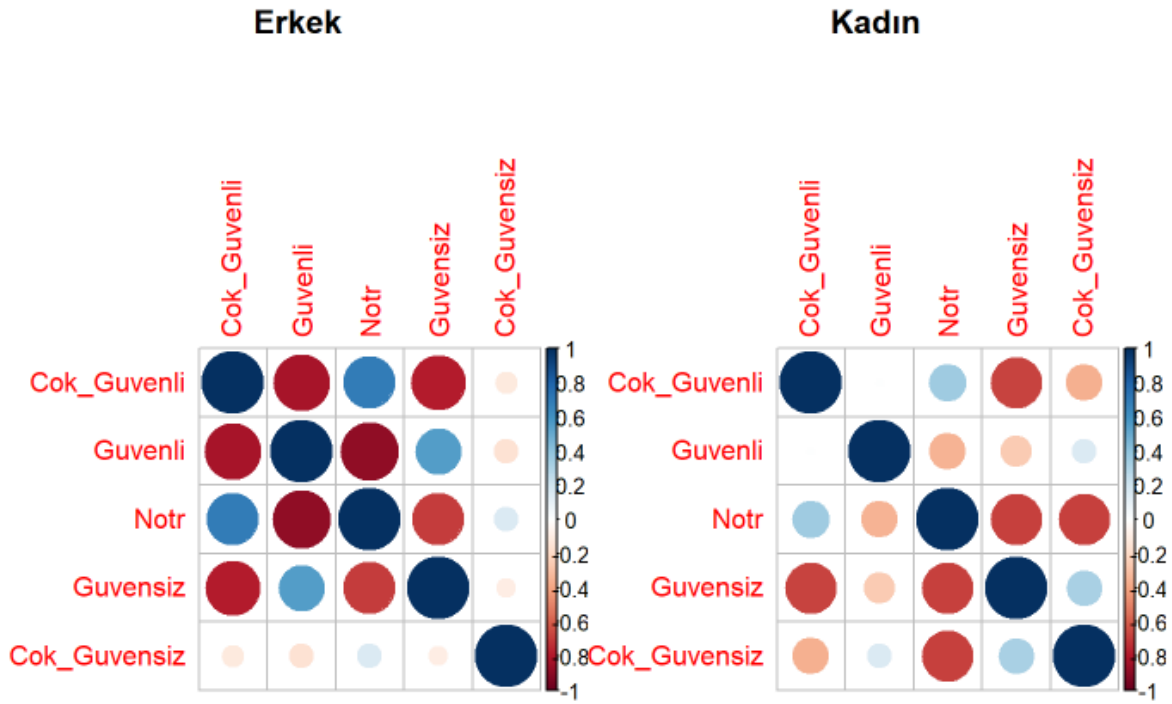


Görsel 18: Korelasyon matrisi

- **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.

## Cinsiyete Göre Ayır Analiz



Görsel 19: Cinsiyetleri göre korelasyon matrisi

## Toplam Suç Haritası (Her İl İçin Tüm Suçların Toplamı)

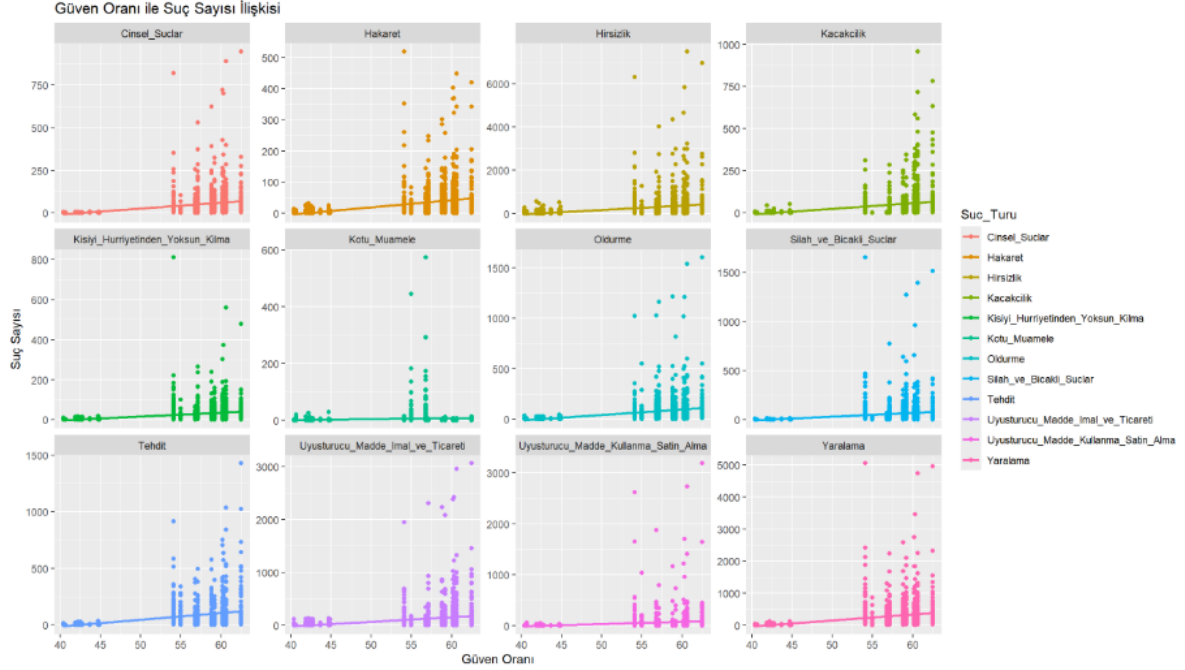


Görsel 20: İllere göre toplam suç dağılımı



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.

## Her suç türü için güven dereceleriyle scatter plot (nokta grafiği)



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** gözlemliyoruz.

```
## [1] "Guven ile Suc arasındaki korelasyon: -0.044"
```

Görsel 22: Güven ve suç arasındaki korelasyon

Korelasyon değeri:

0.7 – 1.0 → güçlü pozitif

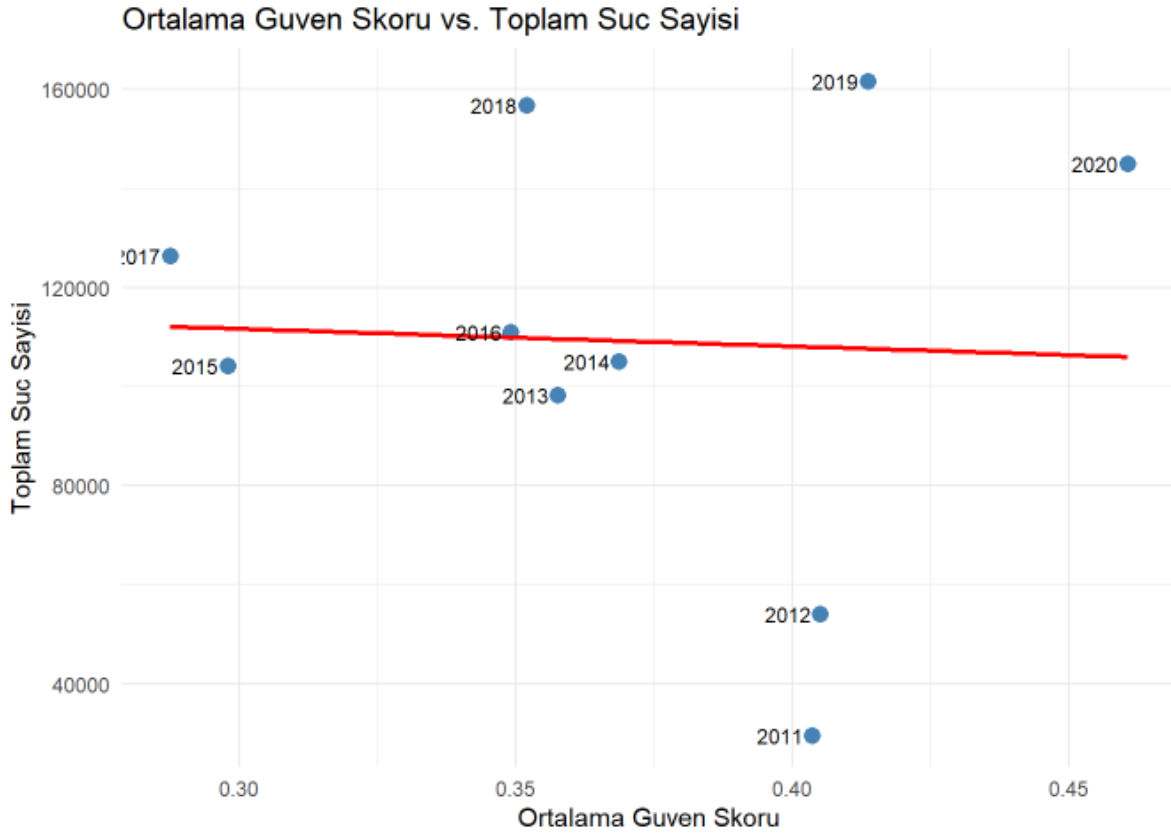
0.4 – 0.7 → orta pozitif

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 bulunmaktadır.



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 sayısı ile ilişkili olabileceğine dair hafif bir işaret verse de, bu ilişki çok kuvvetli değil ve başka birçok faktörün de suç sayısını etkilediği anlaşıyor.