

# Zaman Serileri Üzerine Bir Uygulama

**Matematik Mühendisliği Bitirme Çalışması**

**Zehra GÖL**

**18052042**

**Dr. Öğr. Üyesi Mert BAL**

# İçerik

- Giriş
- Çalışmanın Amacı
- Kullanılan Zaman Serisi Veri Seti
- Kullanılan Modeller
- Veri Ön İşleme

# Projenin Tanıtılması



# Giriş

- Günümüzde **makine öğrenmesi** birçok probleme yaklaşımında kullanılmaktadır
- Her geçen gün **farklı veri tipleri** için farklı çözümler geliştirilmektedir
- **Zaman serileri** gibi zamana bağlı değişen ve bir akış içinde olan veriler için de makine öğrenmesi farklı çözümler önermektedir
- Bu çalışmada da bir zaman serisi verisi ele alınıp birçok farklı açıdan analiz edilecektir



# Çalışmanın Amacı

- Amerika'da bir şehir olan **Chicago'da** işlenen suçların **sosyolojik olarak, zamana bağlı** bir şekilde incelenmesi
- Zamana bağlı değişkenleri belirleyip bir **zaman serisi modeli** ile oluşabilecek **suç yoğunluklarını** önceden tahmin etme
- Makine öğrenmesi modelleri ile oluşabilecek suçların **konumuna ve zamanına** göre türünü tahmin etme
- Veri üzerinden elde edilen analizlerle ve ortaya çıkan modellerin sonuçları ile suç yoğunluğunun arttığı bölgeleri ortaya çıkarma



# Kullanılan Zaman Serisi Veri Seti

- Chicago şehrine ait bir veri portalı web sitesi [1] üzerinden yayınlanan veri setleri incelenmiştir
- 2022 yılına ait **suç kayıtlarının** tutulduğu veri seti [2] çalışmada kullanılmıştır

- 238 bin adet suç kaydı
- 1 Ocak 2022 - 1 Ocak 2023 tarihleri arası
- Konum ve zaman bilgisi
- Suç türü ve tanımı bilgisi



Chicago Data Portal

Crimes - 2022  
Based on [Crimes - 2001 to Present](#)  
Records from the Crimes - 2001 to Present dataset for the indicated year.

ID	Case Number	Date	Block	IUCR	Primary Type	Description
13024128	JG202134	12/31/2022 11:55:00 ...	055XX W ADAMS ST	1477	WEAPONS VIOLATION	RECKLESS FIREA
12938426	JG100085	12/31/2022 11:50:00 ...	055XX W WASHINGTO...	0486	BATTERY	DOMESTIC BATTI
12938337	JG100016	12/31/2022 11:50:00 ...	014XX W LUNT AVE	0910	MOTOR VEHICLE THEFT	AUTOMOBILE
12938351	JG100087	12/31/2022 11:47:00 ...	091XX S LAFLIN ST	0560	ASSAULT	SIMPLE
12938340	JG100005	12/31/2022 11:47:00 ...	091XX S LAFLIN ST	143A	WEAPONS VIOLATION	UNLAWFUL POSS
12938578	JG100071	12/31/2022 11:45:00 ...	001XX E RANDOLPH ST	0460	BATTERY	SIMPLE

- Güncel ve polis departmanları tarafından onaylanmış veriler

# Veri Ön İşleme ve Analizi



# Veri Seti

ID	Case Number	Date	Block	IUCR	Primary Type					
13024128	JG202134	12/31/2022 11:55:00 ...	055XX W ADAMS ST	1477	WEAPONS VIOLATION					
12938426	JG100085	12/31/2022 11:50:00 ...	055XX W WASHINGTON...	0486	BATTERY					
12938337	JG100016	12/31/2022 11:50:00 ...	014XX W LUNT AVE	0910	MOTOR VEHICLE THEFT					
12938351	JG100087	12/31/2022 11:47:00 ...	091XX S LAFLIN ST	0560	ASSAULT					
Description	Location Description	Arrest	Domestic	Beat						
RECKLESS FIREARM DISCHARGE	RESIDENCE - YARD (FRONT / BACK)	False	False	1522						
DOMESTIC BATTERY SIMPLE	APARTMENT	False	True	1523						
AUTOMOBILE	ALLEY	False	False	2431						
District	Ward	Community	FBI Code	X Coordinate	Y Coordinate	Year	Updated	Latitude	Longitude	Location
001	4	32	03	0	0	2022	2022 Oct 09 ...	36.619446395	-91.686565...	(36.619446395°, -91.686565°)
001	4	32	11	0	0	2022	2022 Oct 14 ...	36.619446395	-91.686565...	(36.619446395°, -91.686565°)
031	41	76	08B	1091242	1941464	2022	2023 Jan 03 ...	41.996235316	-87.939732...	(41.996235316°, -87.939732°)

# Veri Seti

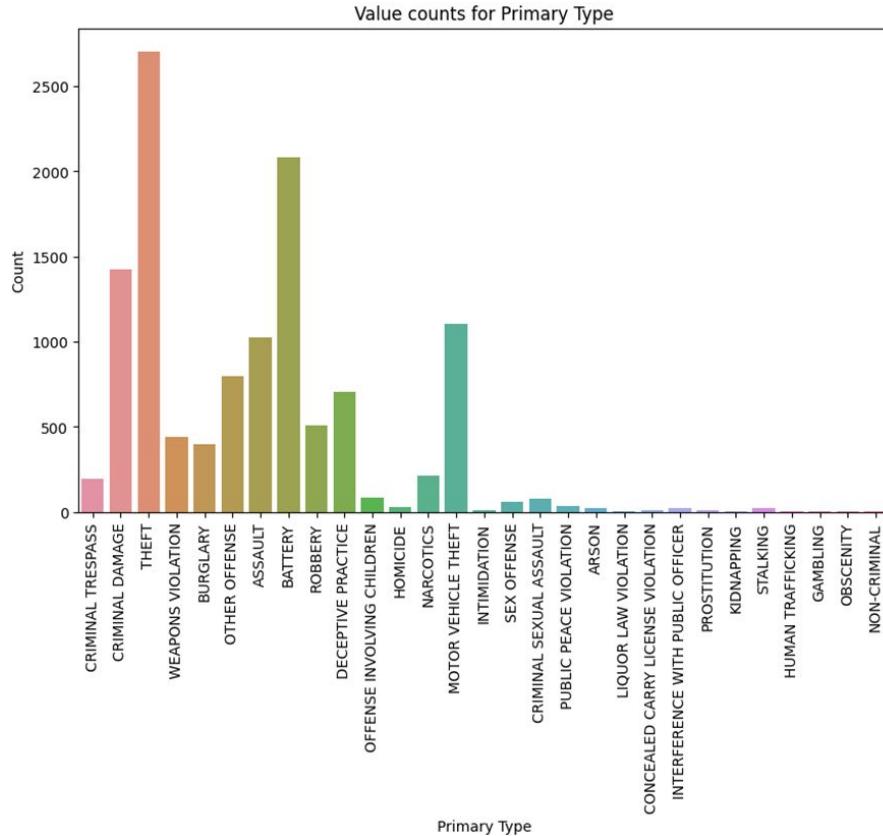
- Veri seti 12 adet **kategorik**, 2 adet **zaman**, 6 adet ise **sayısal** kolondan oluşmaktadır.
- Bu özelliklere sahip **232 bin** adet veri içermektedir.
- 
- Bu veri işlenen suçun **zamanı**, **konumu** ve **türü** gibi bilgilerin dışında **işlendiği polis bölgesi**, **tutuklanma**, **aile içi şiddet** veya **suçun tanımı** gibi bilgileri de içermektedir.
- 2001 yılından bu yana tutulan veri setinin 2022 yılına ait kısmını içermektedir.

# Ön İşleme Adımları

- Veri setinden eksik veri olan satırlar çıkartılmıştır
- Kullanılmayacak olan veya tekrarlanan verilerin bulunduğu kolonlar veriden çıkarılmıştır
  - *ID, Case Number, FBI Code, Block, IUCR, Updated On, Location*
- Tarihin tutulduğu *Date* kolonu pandas veri tiplerinden biri olan *datetime* veri tipine dönüştürülmüştür.
- Son olarak, aynı tarih ve saatte gerçekleşmiş olan suçlar tek bir satır olarak **gruplanmıştır** (*group by*)
- Bu birleştirme sırasında kolonlar farklı yöntemler ile (ortalama alma, modunu alma, sayma vs.) birleştirilmiştir.
- Hedef kolon olarak belirlenen **Suç Türü** ise iki farklı şekilde; modunu alarak ve adetini sayarak veriye kaydedilmiştir.
- Modu alınan suç türü kolonu **Random Forest** modeli ile adeti alınan kolon ise **LSTM** ve **ARIMA** ile analiz edilecek ve modeller geliştirilecektir.

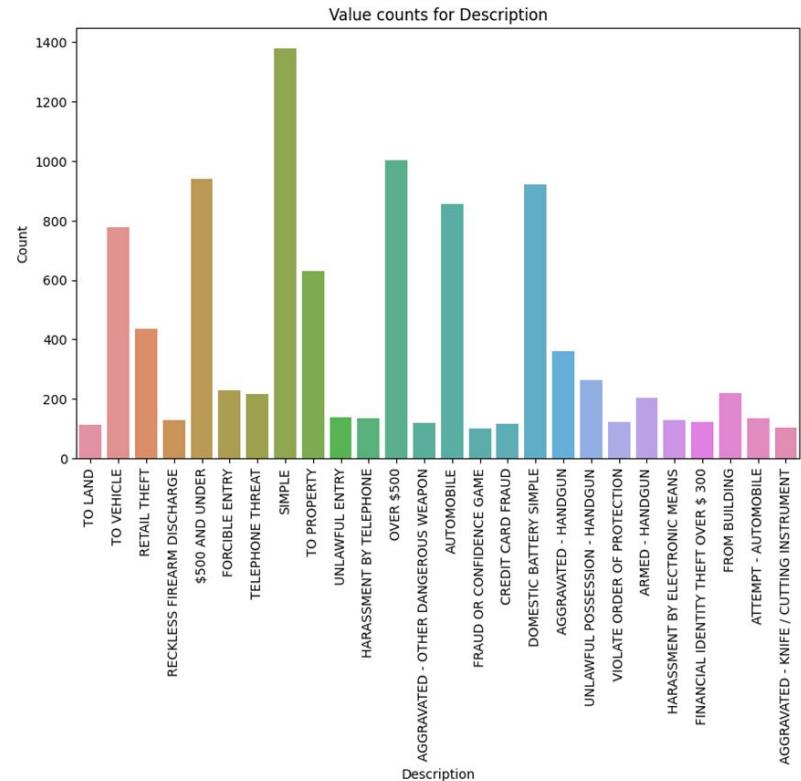
# Veri Analizi - Suç Tipleri

- Burada **suç türüne göre ilgili suçun işlenme sayısı** üzerine bir analiz gerçekleştirilmiştir.
- Görüldüğü gibi ilk üç en fazla suç işlenen suç tipi **hırsızlıktır**
- Hırsızlık dışında en çok işlenen suç tipleri ise **hasar verme** ve **saldırı** suçlarıdır



# Veri Analizi - Suç Tipleri

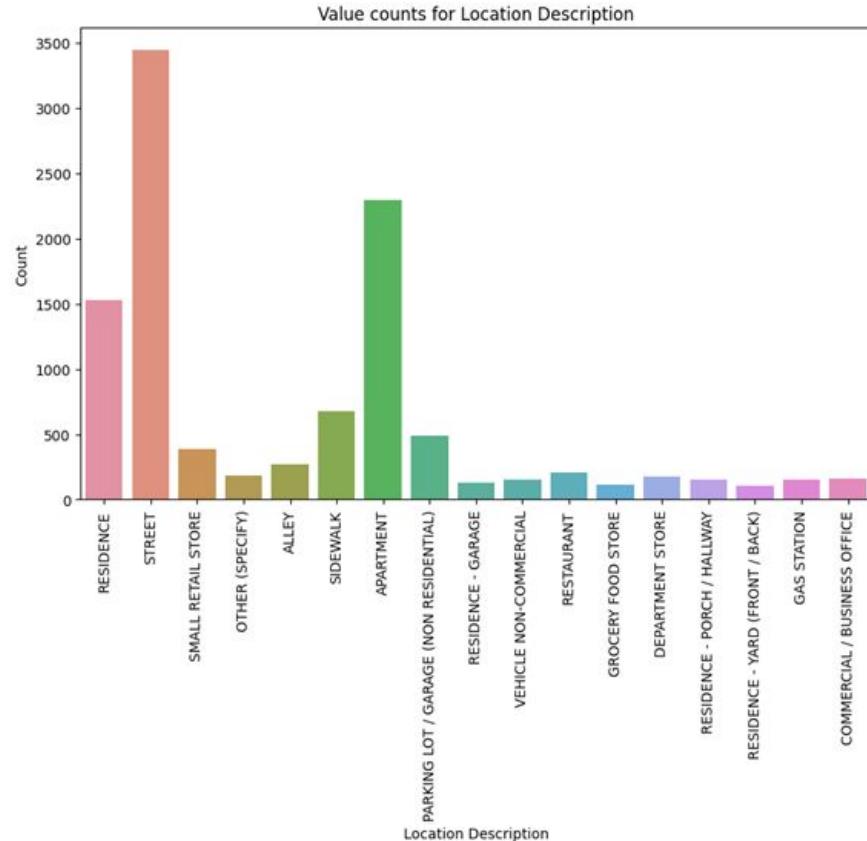
- Şehirde en çok işlenen suç türü olan hırsızlık özelinde, bu suç türünün **alt kategorileri** histogram olarak gösterilmiştir.
- Simple** sınıfına giren hırsızlık en fazla işlenmişken, **500\$ üstü para çalma**, **otomotiv hırsızlığı** ve **akü hırsızlığı** bunları takip etmektedir



# Veri Analizi - Suç Tipleri

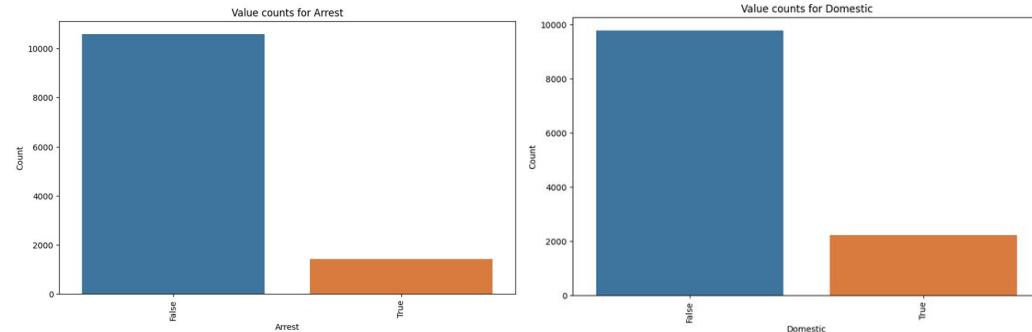
- Suçun işlendiği **yerler** üzerine bir analiz gerçekleştirilmiştir
- **Sokak, apartman ve rezidans** en çok suç işlenen yerler olarak dikkat çekmektedir.

Location Description	
STREET	3440
APARTMENT	2297
RESIDENCE	1525
SIDEWALK	675
PARKING LOT / GARAGE (NON RESIDENTIAL)	490
SMALL RETAIL STORE	390
ALLEY	271
RESTAURANT	205
OTHER (SPECIFY)	180
DEPARTMENT STORE	172

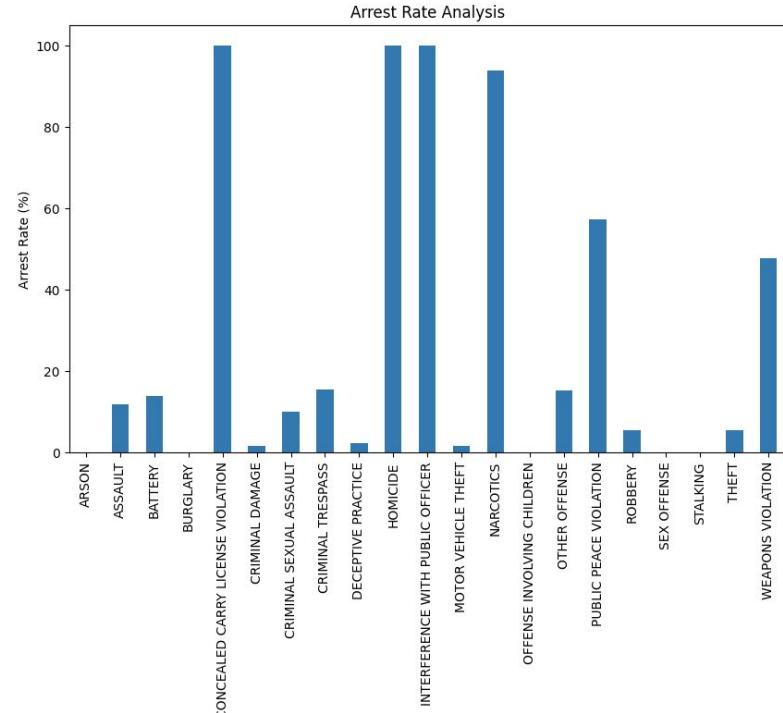


# Veri Analizi - Tutuklanma Oranları

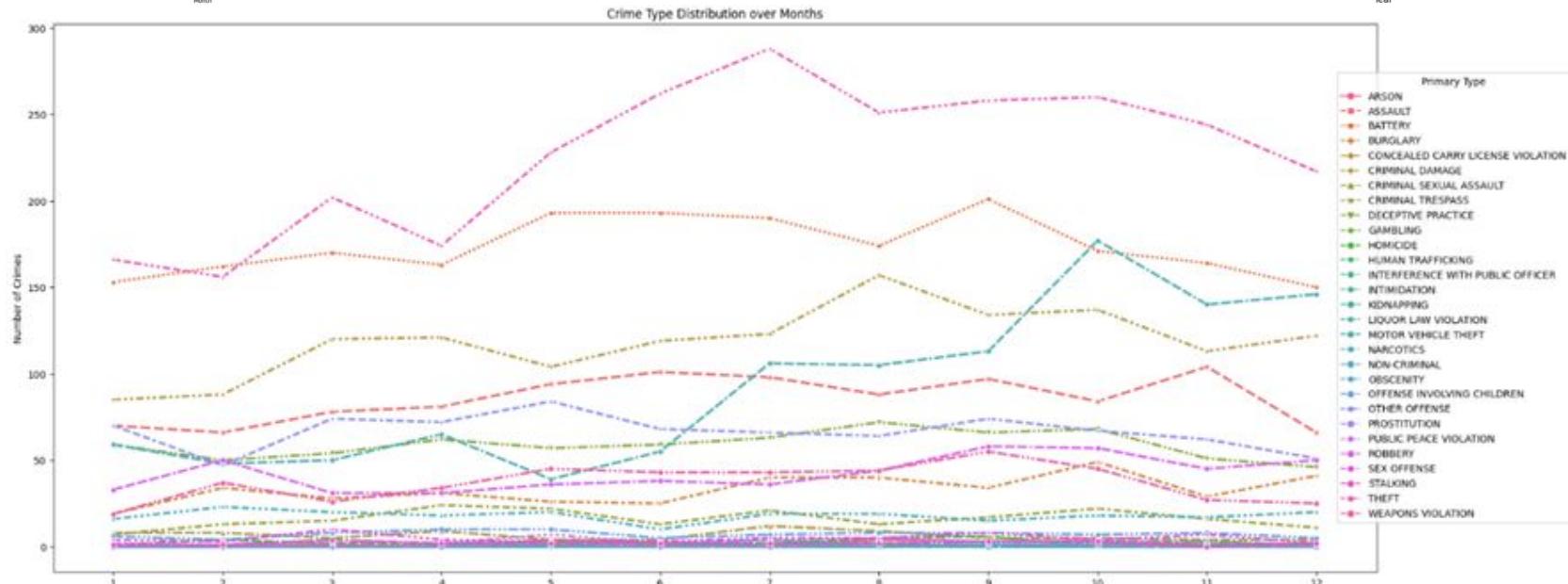
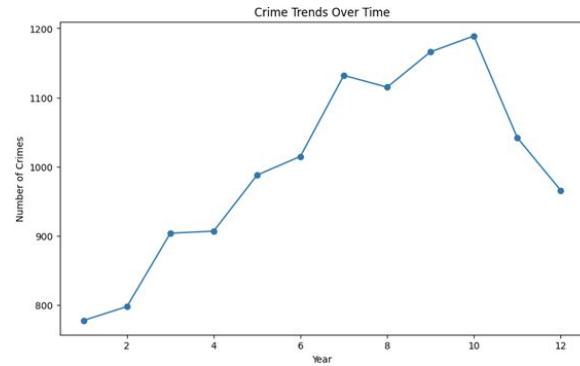
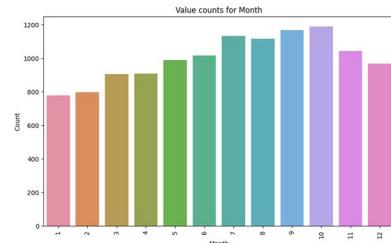
- Bu analizde işlenen suçların yaklaşık **%20'si** aile içi suç olduğu gösterilmiştir
- Aile içi suçlarda tutuklanma oranı **%13'dir**



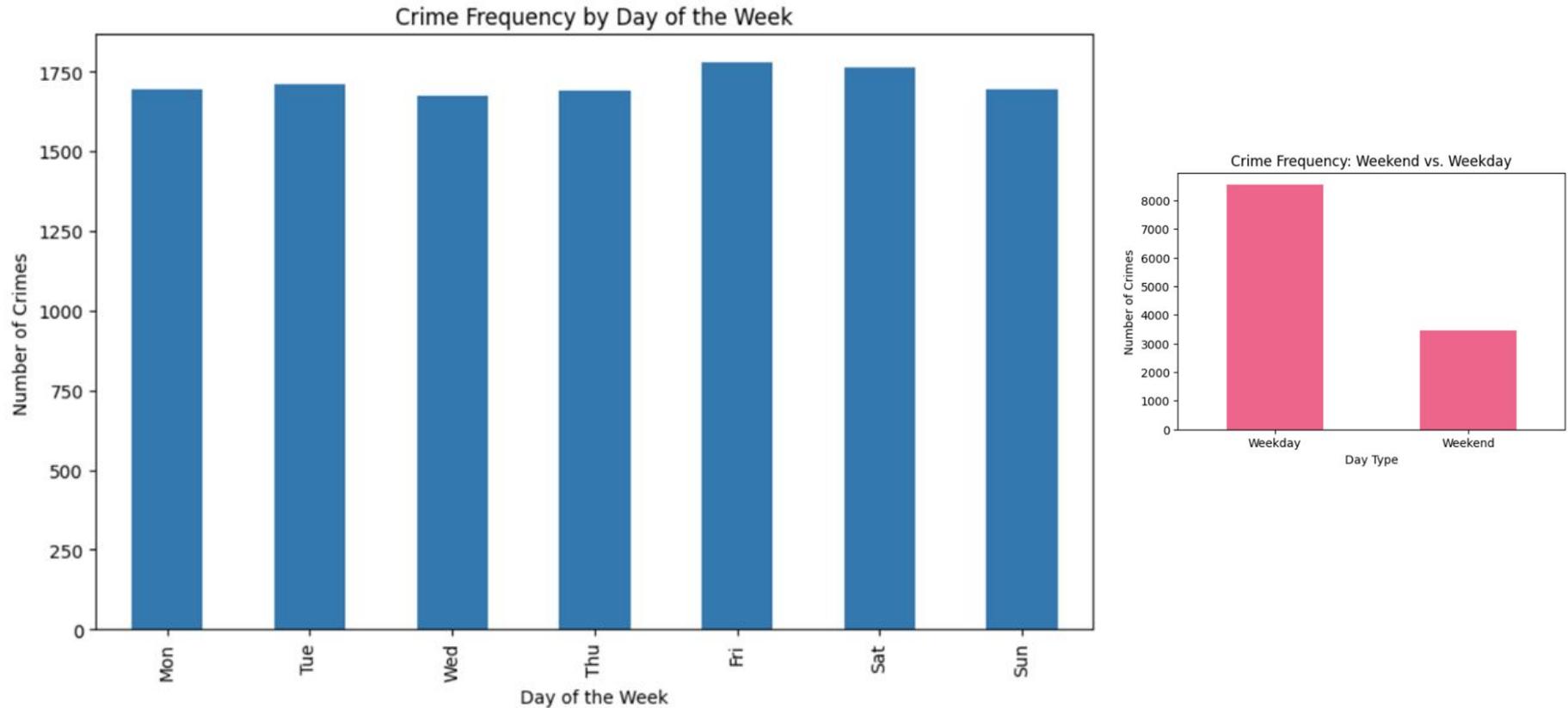
- Şehirdeki suçlara **cezai yaptırımlar** uygulanıp uygulanmadığını gösteren genel bir görselleştirme şeklindeki gibidir
- Hangi suçlara ne kadar **tutuklanma** yapıldığı ve bunların diğer suç türleriyle karşılaştırıldığı görsel ise şeklindeki gibidir



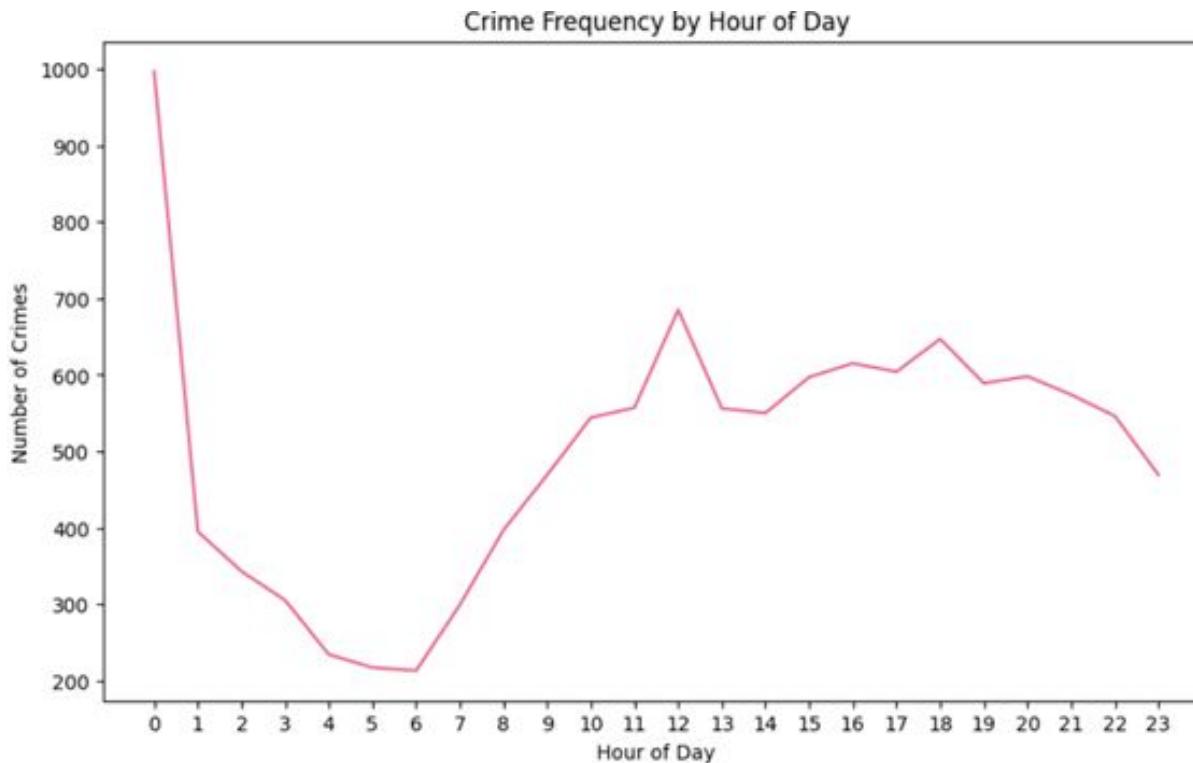
# Veri Analizi - Aylara Göre Değişim



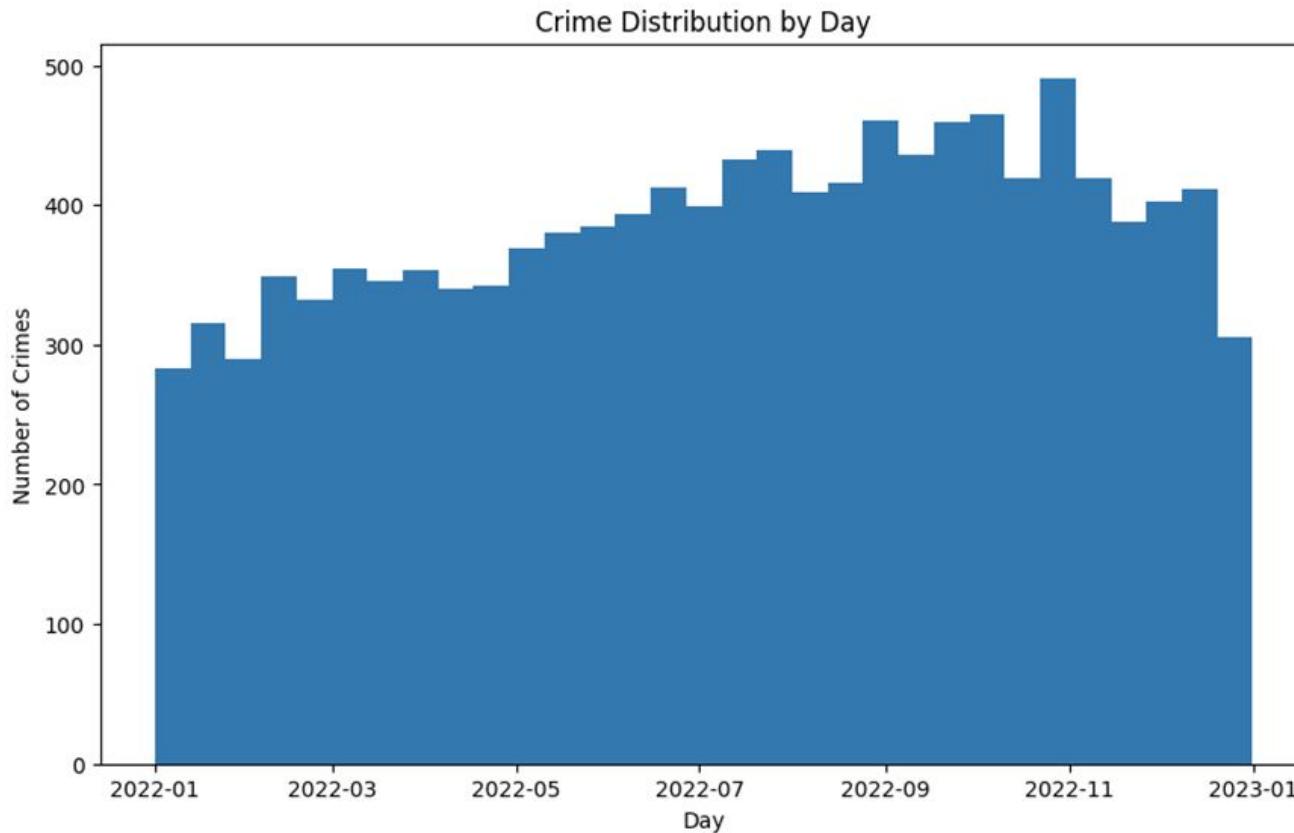
# Veri Analizi - Günlere Göre Değişim



# Veri Analizi - Saatlere Göre Değişim



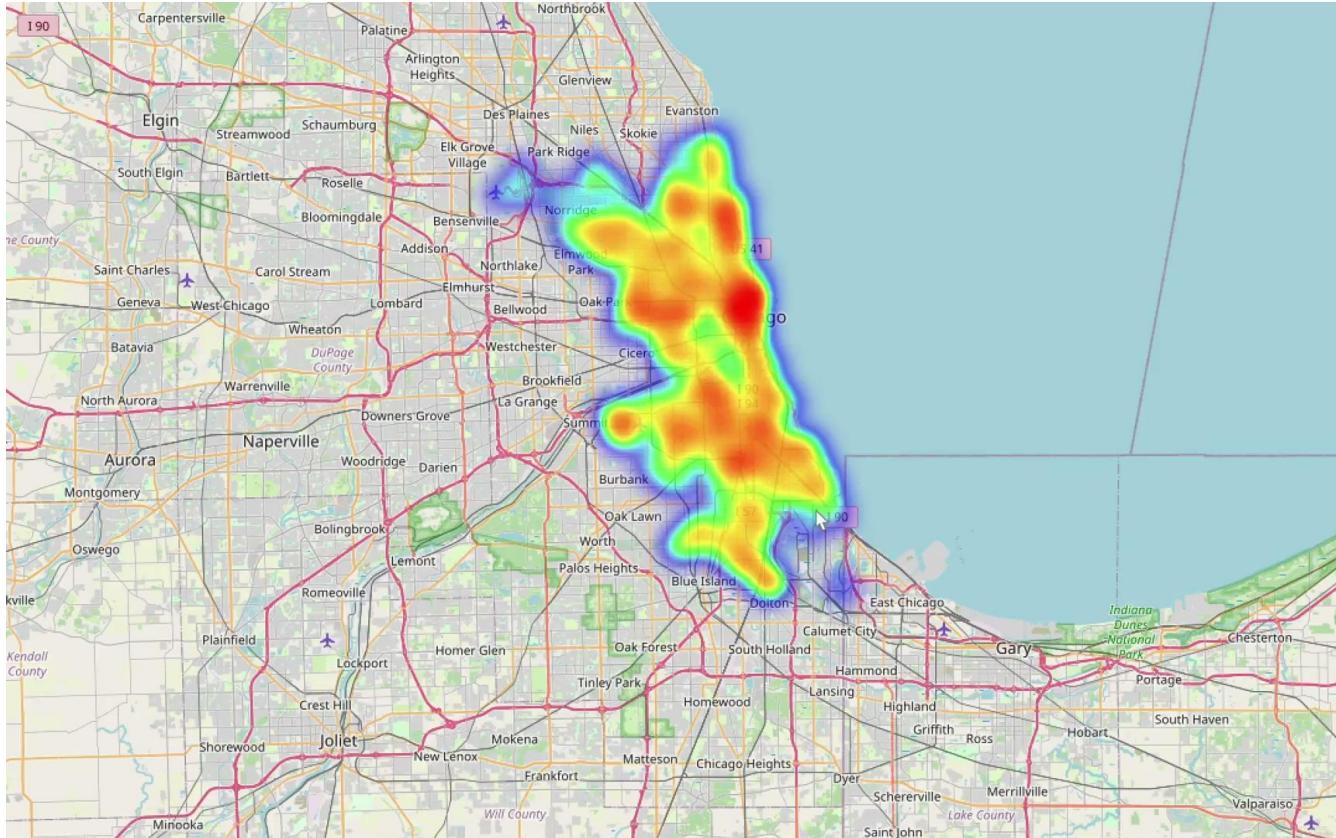
# Veri Analizi - Zamana Göre Genel Değişim



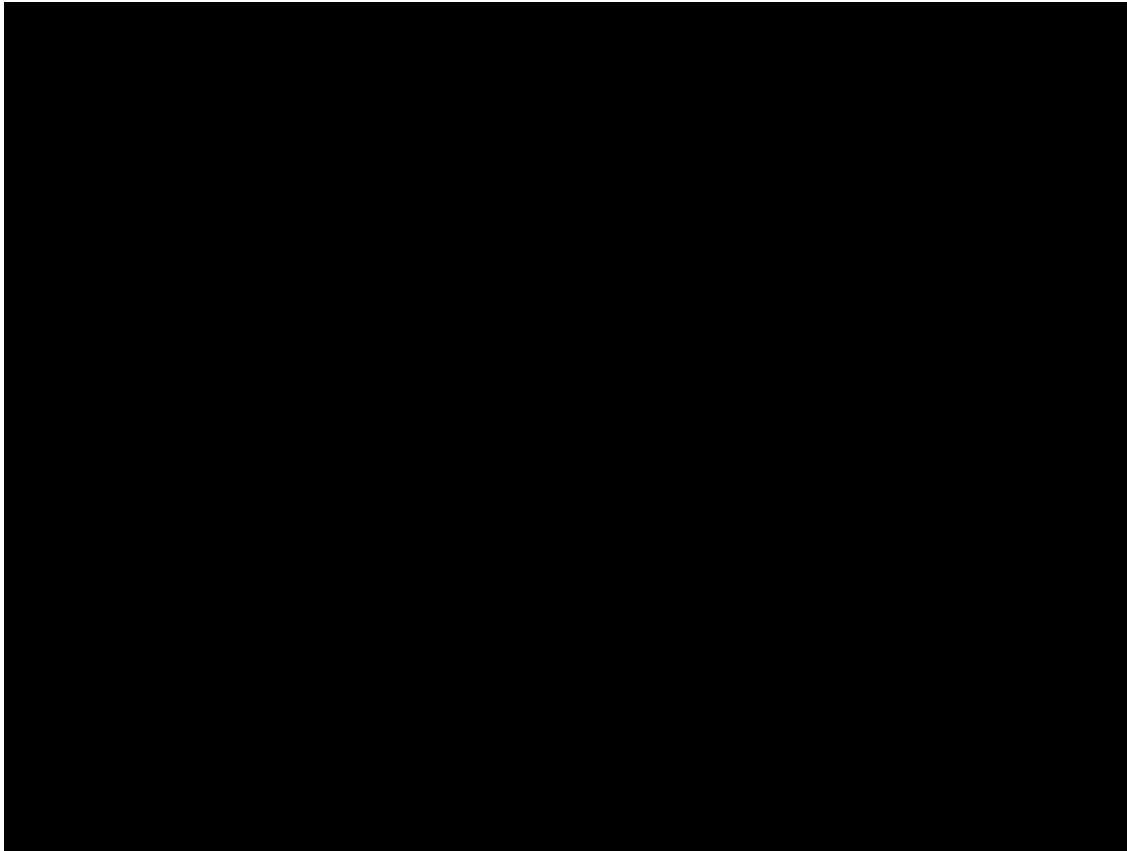
# Harita Üzerinde Görselleştirmeler



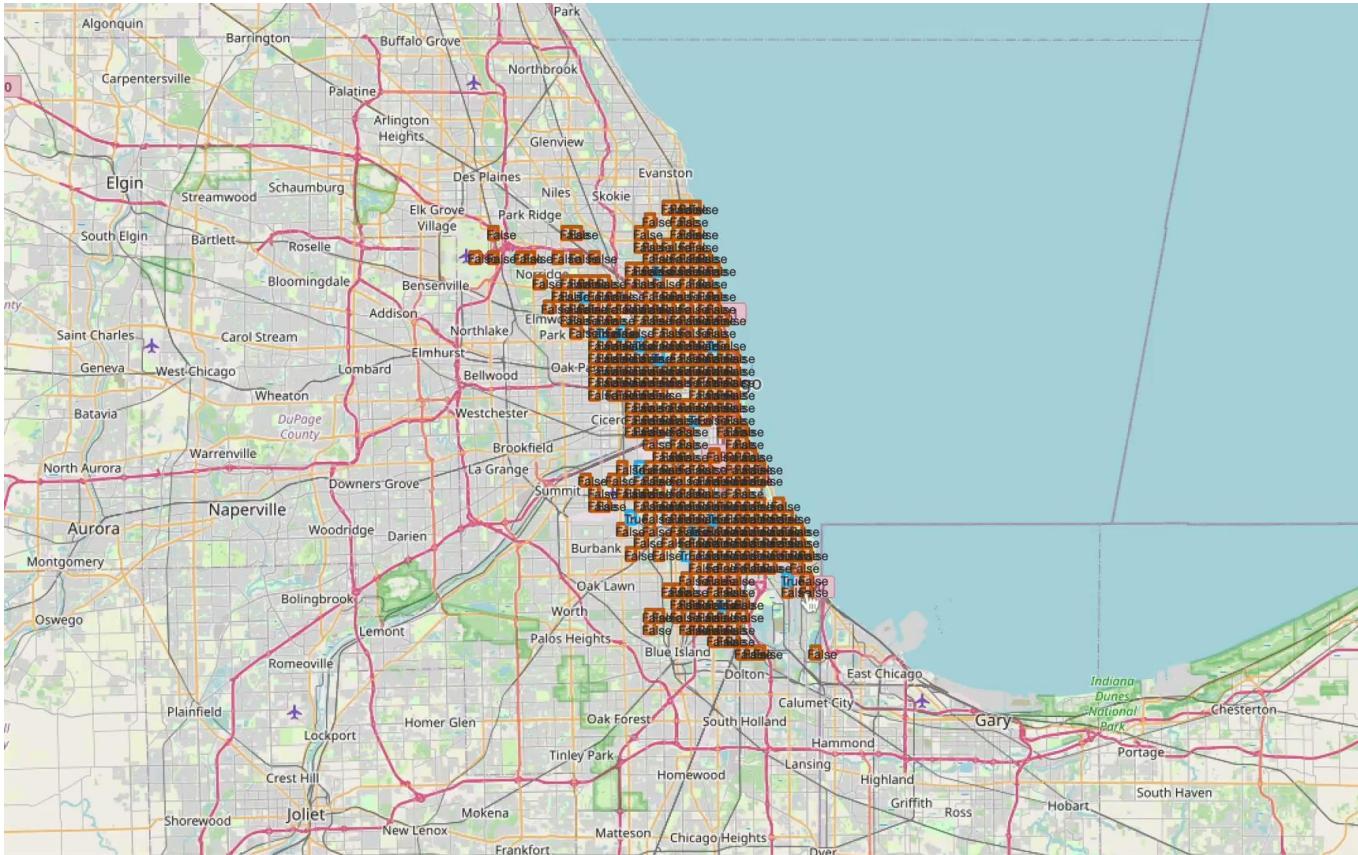
# Harita Görselleştirmesi - Suç Yoğunlukları



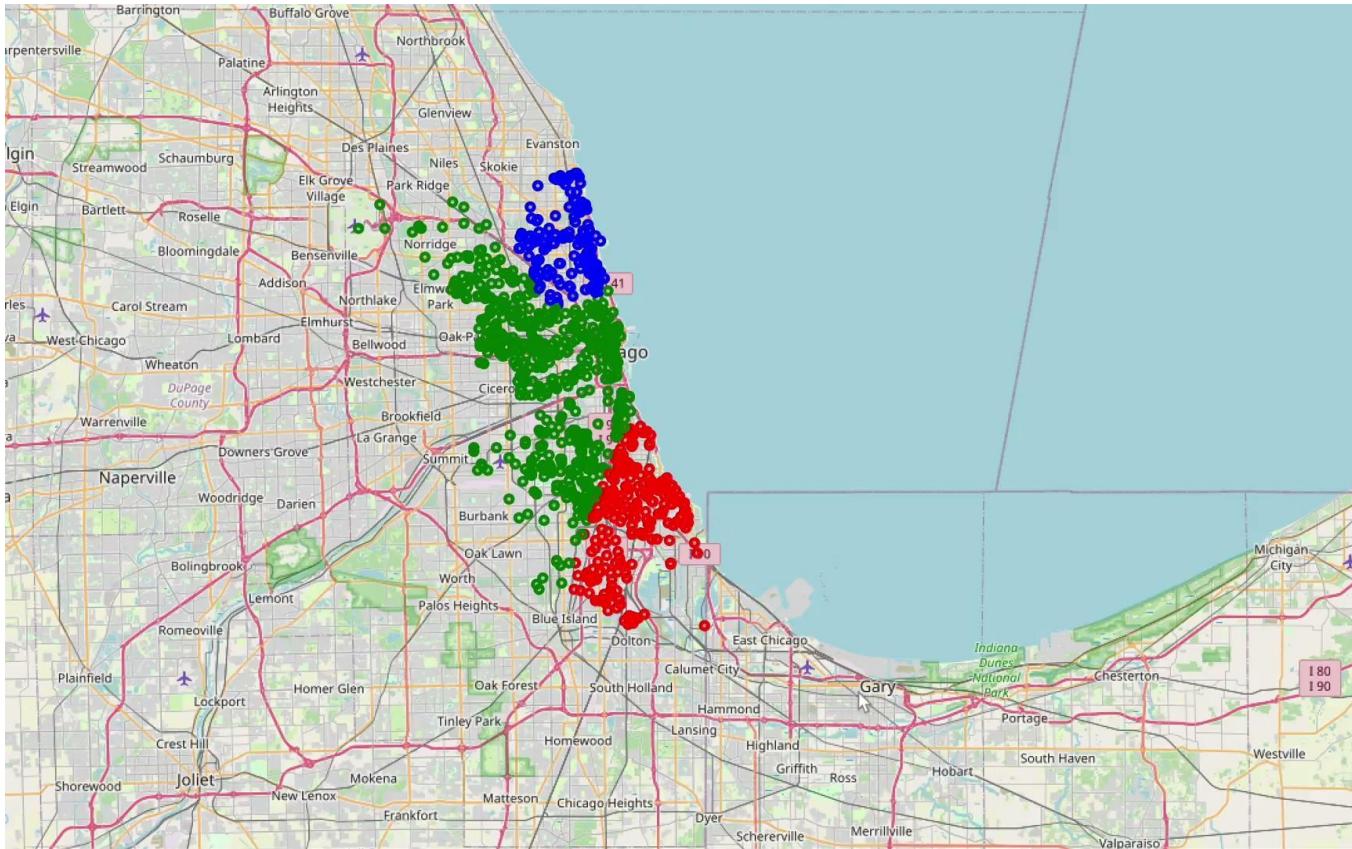
# Harita Görselleştirmesi - Bölgelere Göre En Çok İşlenen Suçlar



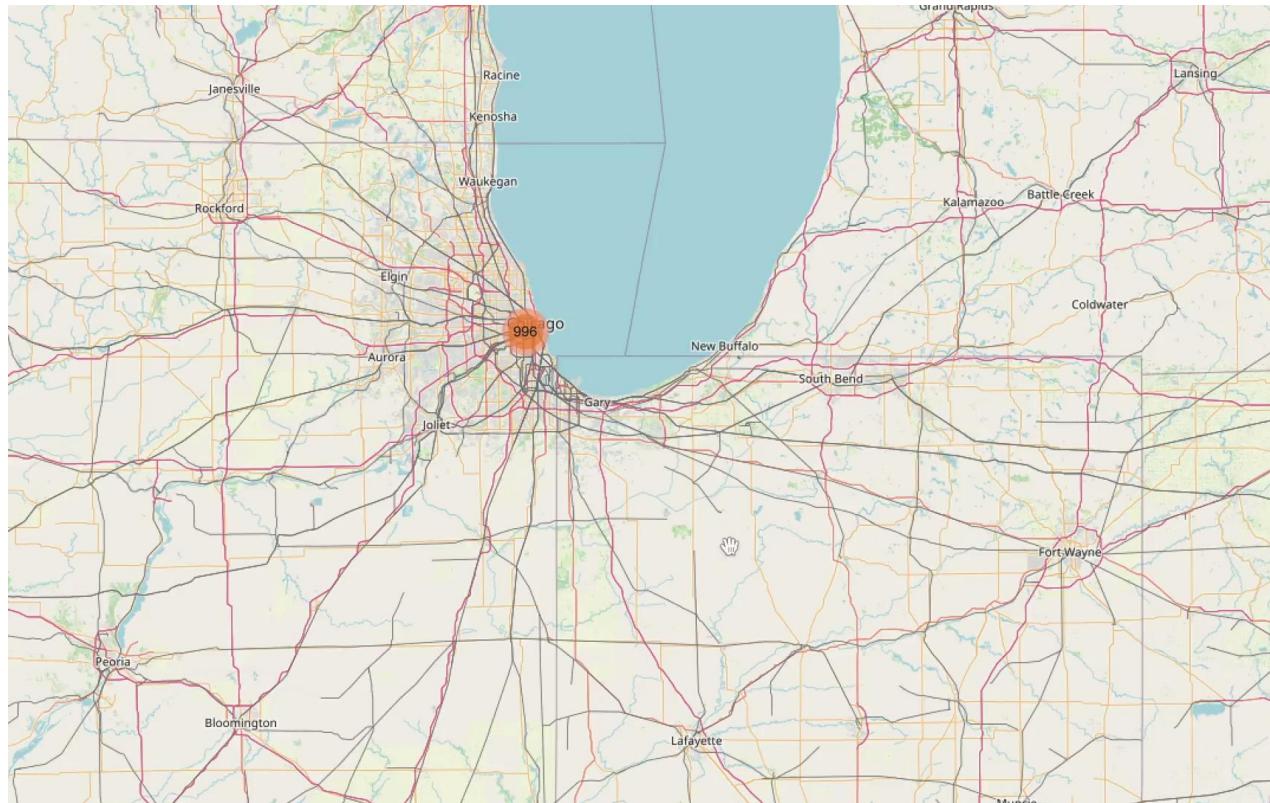
# Harita Görselleştirmesi - Tutuklanma Oranları



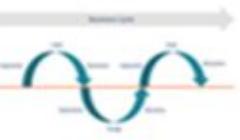
# Harita Görselleştirmesi - KMeans ile Suç Bölgeleri



# Harita Görselleştirmesi - Suç Sayıları ve Kapladığı Alanlar



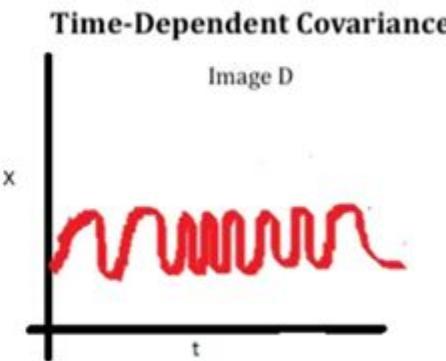
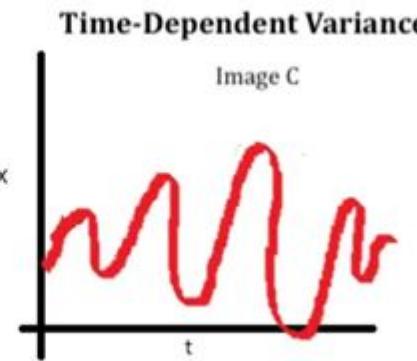
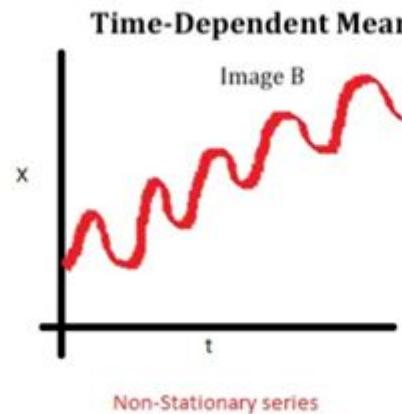
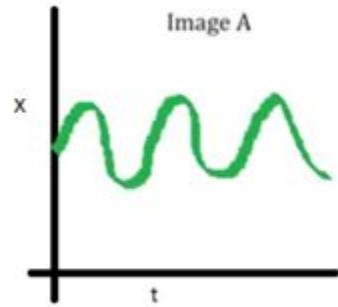
# Zaman Serisi Kavamları

	Trend	Seasonality	Cyclical	Irregularity
Time	Fixed Time Interval	Fixed Time Interval	Not Fixed Time Interval	Not Fixed Time Interval
Duration	Long and Short Term	Short Term	Long and Short Term	Regular/Irregular
Visualization				
Nature - I	Gradual	Swings between Up or Down	Repeating Up and Down	Errored or High Fluctuation
Nature – II	Upward/Down Trend	Pattern repeatable	No fixed period	Short and Not repeatable
Prediction Capability	Predictable	Predictable	Challenging	Challenging
Market Model				Highly random/Unforeseen Events – along with white noise.

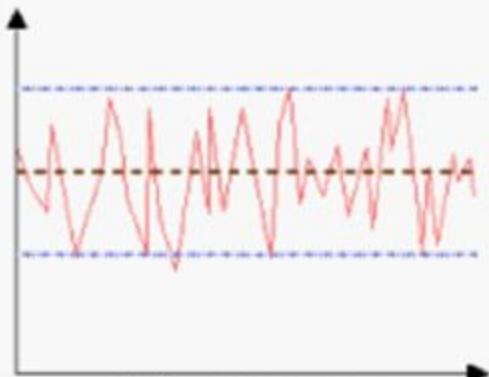
Designed by Author (Shanthababu)

# Durağanlık

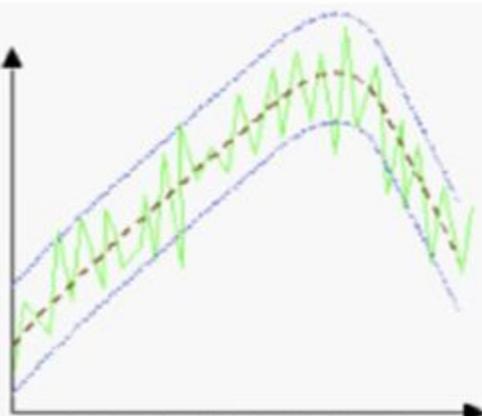
## The Principles of Stationarity



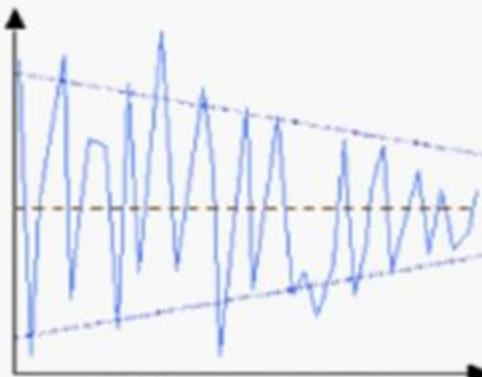
Random Walk



**stationary mean**  
**stationary variance**



**non-stationary mean**  
**stationary variance**



**stationary mean**  
**non-stationary variance**

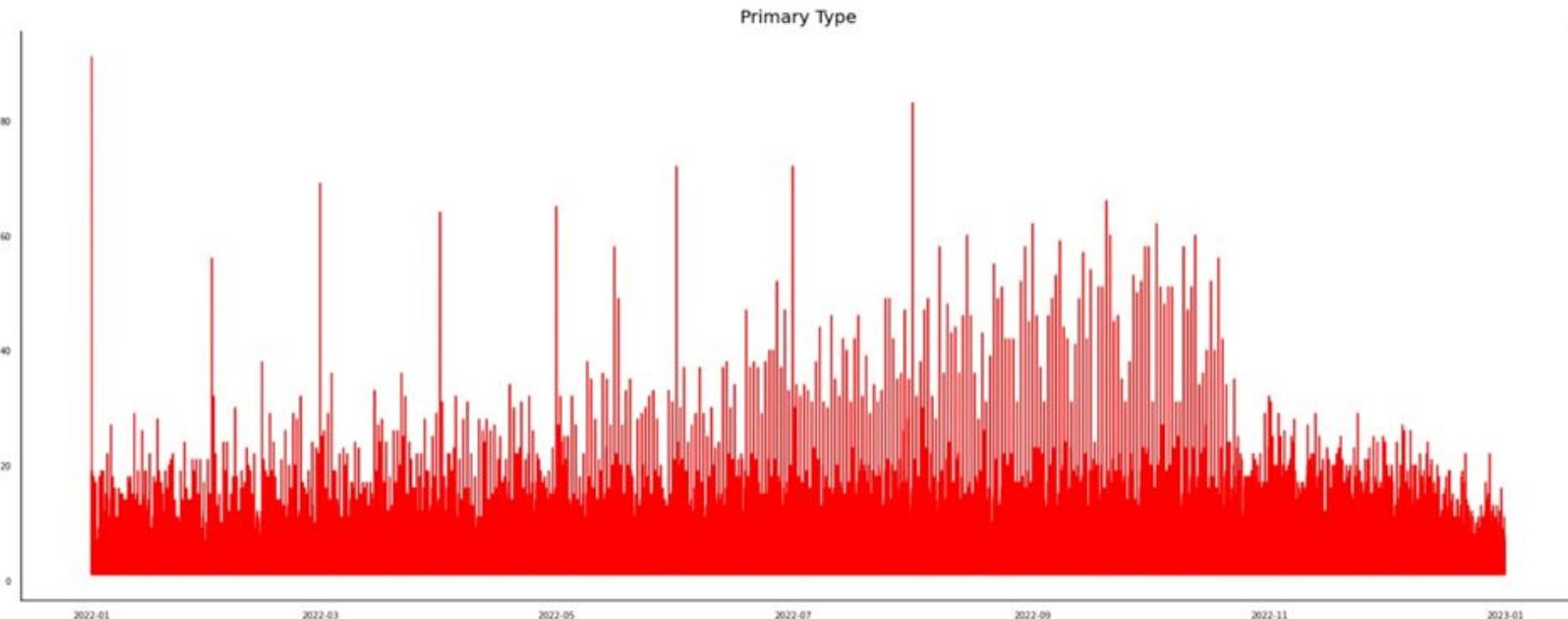
## Veriyi Durağanlıktan Arındırmalıyız, Neden?

### Durağanlık Testleri

-> Veri Görselleştirme

- ADF Test
- KPSS Test

# Veri Görselleştirme

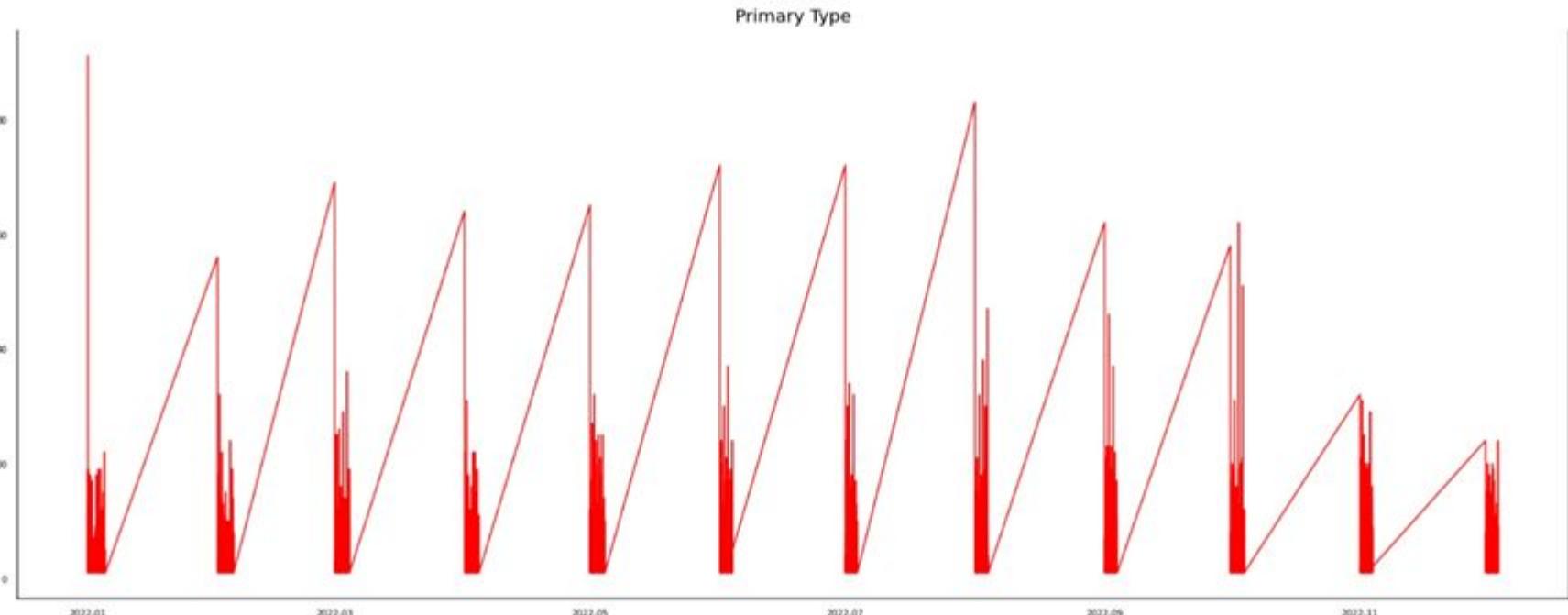


### **6.1.2.2 Tüm Veri için Durağanlık Testi**

	<b>ADF Testi</b>	<b>KPSS Testi</b>
<b>Test İstatistiği</b>	<b>-75.564171</b>	<b>19.878922</b>
<b>p değeri</b>	<b>0.000000</b>	<b>0.010000</b>
<b>Kritik Değer (%1)</b>	<b>-3.430409</b>	<b>0.739000</b>
<b>Kritik Değer (%5)</b>	<b>-2.861566</b>	<b>0.463000</b>
<b>Kritik Değer(%10)</b>	<b>-2.566784</b>	<b>0.347000</b>
<b>Sonuçlar</b>	<b>DURAĞAN</b>	<b>DURAĞAN DEĞİL</b>

**Tablo 6.1.2.2.1** Tüm Verinin Kullanıldığı Durağanlık Testi Sonuçları

```
data_50 = grouped['Primary Type'].groupby(pd.Grouper(freq='M')).head(50)
data_1000 = grouped['Primary
Type'].groupby(pd.Grouper(freq='M')).head(1000)
data_3000 = grouped['Primary
Type'].groupby(pd.Grouper(freq='M')).head(3000)
```



# ADF Testi Yorumlanması

Sıfır Hipotezi (HO): Seri durağan değildir veya seri birim köke sahiptir.

Alternatif Hipotez (HA): Seri durağandır veya birim kökü yoktur.

Sıfır hipotezi reddedilemezse, bu test serinin durağan olmadığını dair kanıt sağlayabilir.

Sıfır Hipotezini Reddetme Koşulları (H0):

Test istatistiği  $<$ Kritik Değer ve p değeri  $<0,05$  ise – Sıfır Hipotezi (HO) reddedilir, yani zaman serisinin birim kökü yoktur bu da zaman serisinin durağan olduğu anlamına gelir. Bir diğer tabirle zamana bağlı bir yapısı yoktur.

ADF testine göre p değeri  $<0.5$  ve Test İstatistiği tüm güven aralıklarında kritik değerlerden daha düşük olduğu için H (0) hipotezi reddedilir. H (1) kabul edilir. Bu teste göre seri durağandır.

# KPSS Testi

Sıfır Hipotezi (HO): Seri trend durağandır veya serinin birim kökü yoktur.

Alternatif Hipotez (HA): Seri durağan değildir veya seri birim köke sahiptir.

Sıfır hipotezi reddedilemezse, bu test serinin durağan olduğuna dair kanıt sağlayabilir.

Sıfır Hipotezini Reddetme Koşulları (HO):

Test İstatistiği  $<$ Kritik Değer ve p değeri  $<0,05$  ise Sıfır Hipotezi (HO) Reddedilemez, yani zaman serisinin birim kökü yoktur, yani trend durağandır anlamına gelmektedir.

KPSS Testine göre  $p<0,05$  ve Test İstatistiği tüm güven aralığında kritik değerlerden daha büyük olduğundan sıfır hipotezi reddedilir. Dolayısıyla serinin KPSS testine göre durağan olmadığı sonucuna varılır.

# Veriyi Durağanlaştırma

- Detrending
- Fark Alma
  - serinin birim kökünü ortadan kaldırma
- Dönüşüm



```
grouped_copy= grouped.copy()

grouped_copy["Primary_Type_diff"] = grouped_copy["Primary Type"] - grouped_copy["Primary Type"].shift()
grouped_copy["Primary_Type_diff_30"] = grouped_copy["Primary Type"] - grouped_copy["Primary Type"].shift(30)
grouped_copy["Primary_Type_diff_100"] = grouped_copy["Primary Type"] - grouped_copy["Primary Type"].shift(100)
grouped_copy["Primary_Type_diff_630"] = grouped_copy["Primary Type"] - grouped_copy["Primary Type"].shift(630)
```

Primary Type shift shift\_diff

Date

2022-01-01 00:00:00	91	NaN	NaN
2022-01-01 00:01:00	35	91.0	-56.0
2022-01-01 00:03:00	1	35.0	-34.0
2022-01-01 00:04:00	1	1.0	0.0
2022-01-01 00:05:00	3	1.0	2.0

<u>Primary_type_diff</u>	ADF Testi	KPSS Testi
Test İstatistiği	<b>-68.624302</b>	<b>0.312981</b>
p değeri	<b>0.000000</b>	<b>0.10000</b>
Kritik Değer (%1)	<b>-3.430409</b>	<b>0.739000</b>
Kritik Değer (%5)	<b>-2.861566</b>	<b>0.463000</b>
Kritik Değer(%10)	<b>-2.566784</b>	<b>0.347000</b>
Sonuçlar	<b>DURAĞAN</b>	<b>DURAĞAN DEĞİL</b>

Tablo 6.1.2.3.1 1 Kaydırma ile Fark İşlemi Alınan Senaryo ve Durağanlık Testi  
Sonuçları

<u>Primary_type_diff_30</u>	ADF Testi	KPSS Testi
Test İstatistiği	<b>-66.639192</b>	<b>0.015214</b>
p değeri	<b>0.000000</b>	<b>0.10000</b>
Kritik Değer (%1)	<b>-3.430409</b>	<b>0.739000</b>
Kritik Değer (%5)	<b>-2.861566</b>	<b>0.463000</b>
Kritik Değer(%10)	<b>-2.566784</b>	<b>0.347000</b>
Sonuçlar	<b>DURAĞAN</b>	<b>DURAĞAN DEĞİL</b>

Tablo 6.1.2.3.2 30 Kaydırma ile Fark İşlemi Alınan Senaryo ve Durağanlık Testi  
Sonuçları

<u>Primary_type_diff_100</u>	ADF Testi	KPSS Testi
Test İstatistiği	<b>-56.214668</b>	<b>0.008152</b>
p değeri	<b>0.000000</b>	<b>0.010000</b>
Kritik Değer (%1)	<b>-3.430409</b>	<b>0.739000</b>
Kritik Değer (%5)	<b>-2.861566</b>	<b>0.463000</b>
Kritik Değer(%10)	<b>-2.566784</b>	<b>0.347000</b>
Sonuçlar	<b>DURAĞAN</b>	<b>DURAĞAN</b>

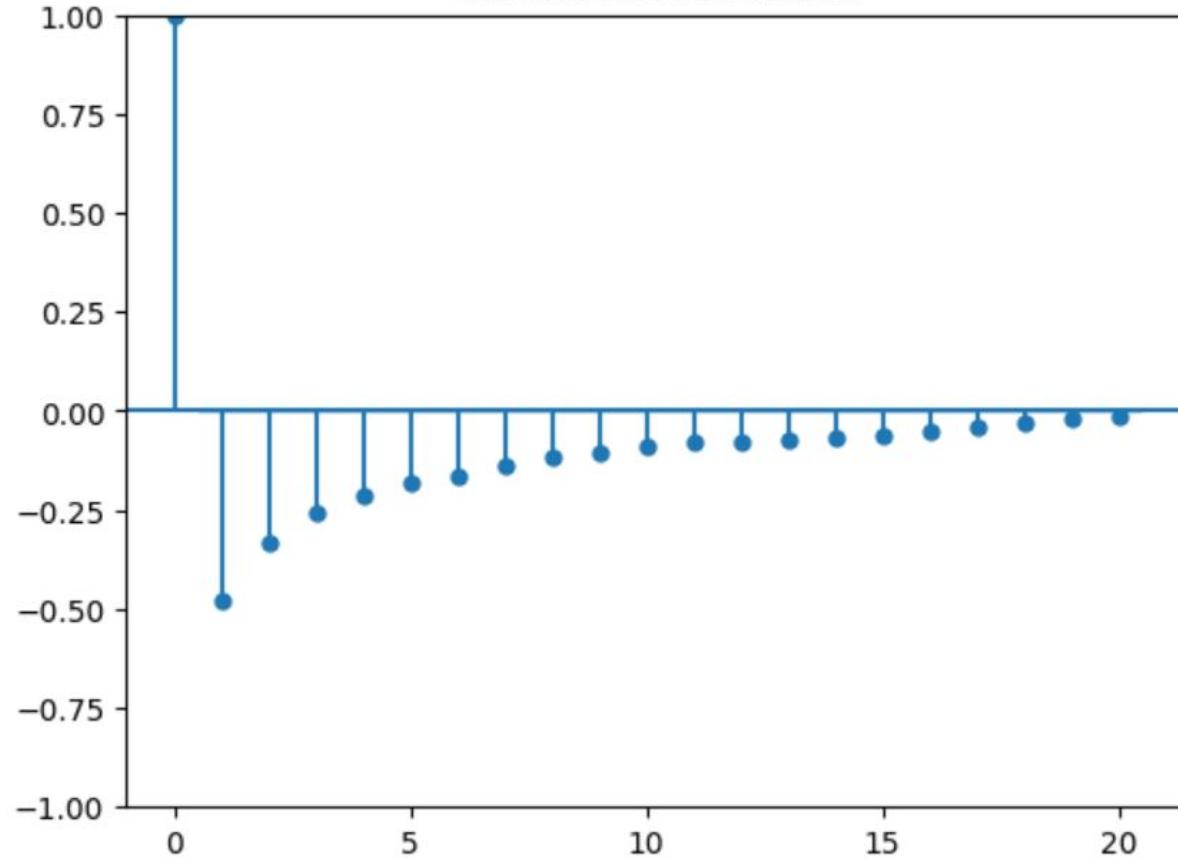
Tablo 6.1.2.3.3 100 Kaydırma ile Fark İşlemi Alınan Senaryo ve Durağanlık Testi  
Sonuçları

<u>Primary_type_diff_630</u>	ADF Testi	KPSS Testi
Test İstatistiği	<b>-45.927002</b>	<b>0.032496</b>
p değeri	<b>0.000000</b>	<b>0.010000</b>
Kritik Değer (%1)	<b>-3.430409</b>	<b>0.739000</b>
Kritik Değer (%5)	<b>-2.861566</b>	<b>0.463000</b>
Kritik Değer(%10)	<b>-2.566784</b>	<b>0.347000</b>
Sonuçlar	<b>DURAĞAN</b>	<b>DURAĞAN</b>

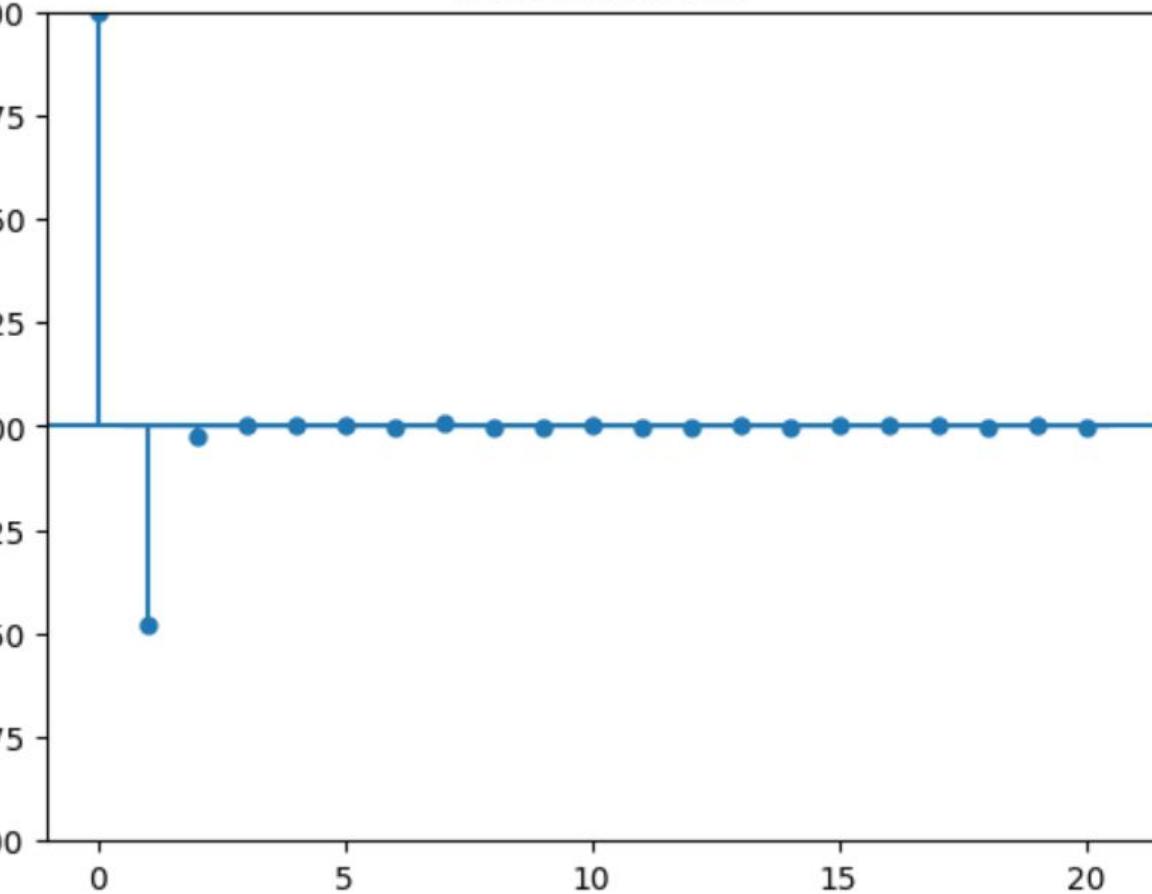
Tablo 6.1.2.3.4 630 Kaydırma ile Fark İşlemi Alınan Senaryo ve Durağanlık Testi  
Sonuçları

ACF, PACF Fonksiyonu ile Model Parametrelerini belirlemek  
ARIMA Modeline Bunu Vermek

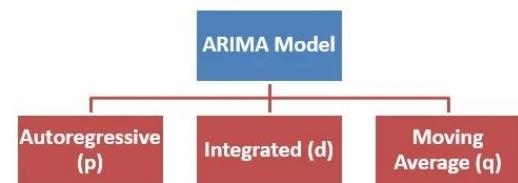
### Partial Autocorrelation



## Autocorrelation



```
model = ARIMA(train, order=(1, 1, 1))
```

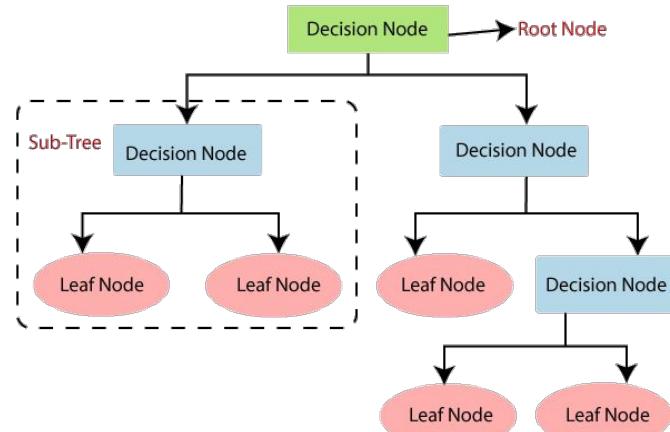
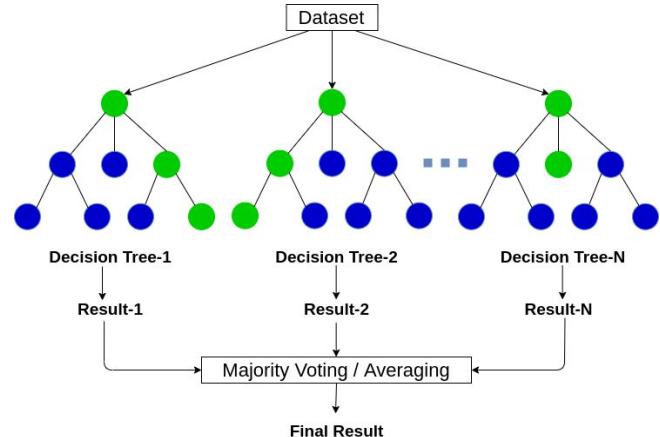


# Modeller ve Sonuçlar



# Kullanılan Modeller - Random Forest

- Random Forest (Rastgele Ormanlar) hem **sınıflandırma** hem de **regresyon** problemlerini çözmek için kullanılan bir makine öğrenmesi algoritmasıdır.
- Bu algoritma **karar ağacı (decision tree)** tabanlı çalışan bir algoritmadır.
- Veri setindeki özelliklerin rastgele seçilmesi ve örneklerin de rastgele seçilmesiyle birçok karar ağacı oluşturulur.
- Oluşturulan karar ağaçları **ensemble (toplu)** bir şekilde test örneklerinin sınıflarına karar verir.
- Random Forest, ensemble yapısından dolayı çok tercih edilen başarılı bir modeldir.



# Random Forest

- Veri Rastgele Orman algoritmasına uygulanmak üzere günlere göre gruplanmıştır.
- Bu gruplama ile sayısal kolonların **ortalaması**, kategorik kolonların **modu** alınarak gerçekleştirilmiştir.
- Hedef kolon olarak belirlenen **Primary Type** kolonu ise modu alınarak birleştirilmiştir.
- Ardından veri eğitim ve test verisi olarak ikiye ayrılmıştır.

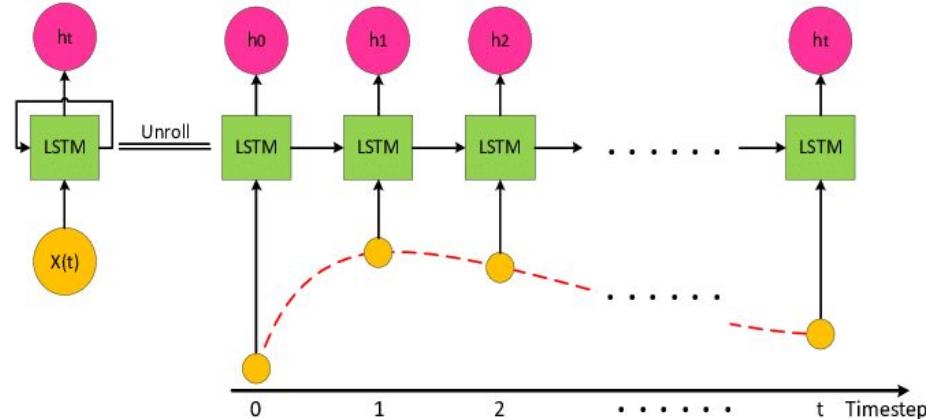
Date	Primary Type	Arrest	Domestic	Beat	District	Ward	Community Area	X Coordinate	Y Coordinate	Latitude	Longitude
2022-01-01 00:00:00	THEFT	0.032967	0.208791	0.949670	0.200000	0.102041	0.092105	0.968396	0.965280	0.965571	0.966822
2022-01-01 00:01:00	CRIMINAL DAMAGE	0.114286	0.200000	0.041667	0.700000	0.367347	0.921053	0.967559	0.961498	0.961827	0.965882
2022-01-01 00:03:00	CRIMINAL DAMAGE	0.000000	0.000000	0.371287	0.300000	0.469388	0.368421	0.953376	0.970254	0.970571	0.950930
2022-01-01 00:04:00	CRIMINAL DAMAGE	0.000000	0.000000	0.459158	0.366667	0.530612	0.355263	0.966468	0.974252	0.974474	0.964899
2022-01-01 00:05:00	CRIMINAL DAMAGE	0.000000	0.000000	0.124175	0.100000	0.142857	0.328947	0.969207	0.960383	0.960713	0.967611

# Random Forest

- Gruplamanın ardından veri sayısı 110 bine düşmüştür.
- Algoritma sayısal kolonlardaki verilerin dengesizliğinden etkilenmemesi için **normalize** edilmiştir.
- %1'lik bir kısım test verisi olarak ayrıldıktan sonra Random Forest modeli eğitilmiştir.
- Eğitilen model test veri seti üzerinde **30** adet sınıf üzerinde (**suç tipi sayısı**) **%30'luk** bir başarı sağlamıştır.
- Bu sayede veri setindeki zaman hariç kolonlar ile hedef olarak belirlenen kolon arasında bir **korelasyon** olduğu kesinleşmiştir.
- 30 sınıfta %30 başarı düşük gözükse de modelin veri setini **öğrendiği** anlamına gelmektedir.

# Kullanılan Modeller - LSTM

- LSTM (Long Short Term Memory) modelleri **akış (sequence)** verilerini analiz etmede kullanılmakta ve bu verilerin çözdüğü problemlere araç olmaktadır.
- Yazı bir akış verisi olduğundan genellikle **Doğal Dil İşleme** alanında kullanılan LSTM modelleri, benzer yapıya sahip Zaman Serisi Verilerinde de kullanılabilir.
- Her bir örnek bir **LSTM hücresına** verildiğinde model, geçmiş verideki bilgileri sonraki hücrelere aktarmakta ve bu sayede bir sonraki veriyi (test verisi) tahmin edebilmektedir.



# LSTM Modeli

- LSTM Model Yapısı:

```
model_lstm=tf.keras.Sequential()
model_lstm.add(tf.keras.layers.LSTM(128, input_shape= (win_length,num_features),return_sequences=True))
model_lstm.add(tf.keras.layers.LeakyReLU(alpha=0.5))

model_lstm.add(tf.keras.layers.LSTM(128,return_sequences=True))
model_lstm.add(tf.keras.layers.LeakyReLU(alpha=0.5))
model_lstm.add(tf.keras.layers.Dropout(0.3))

model_lstm.add(tf.keras.layers.LSTM(64,return_sequences=False))
model_lstm.add(tf.keras.layers.Dropout(0.3))

model_lstm.add(tf.keras.layers.Dense(1))
```

- **64 batch** boyutu ve pencere boyutu olarak 5 kullanılarak görülen LSTM modeli eğitilmiştir.

```
batch_size=64
win_length=5

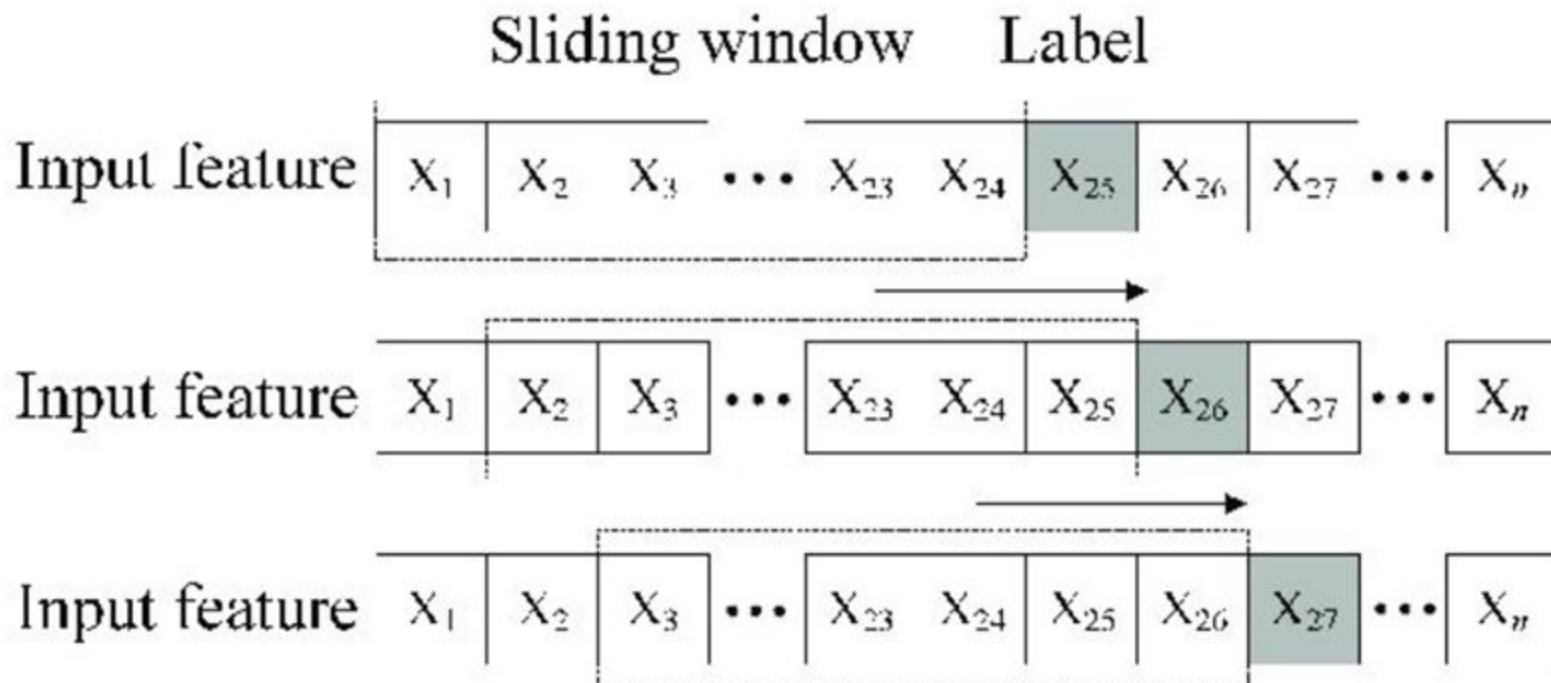
train_generator=TimeseriesGenerator(data=x_train,targets=y_train,length=win_length,batch_size=batch_size,sampling_rate=1)
test_generator=TimeseriesGenerator(data=x_test,targets=y_test,length=win_length,batch_size=batch_size,sampling_rate=1)
```

- **Loss** olarak MSE
- **Optimize** edici olarak Adam
- **Metrik** olarak MSE ve RMSE
- **50 epoch**

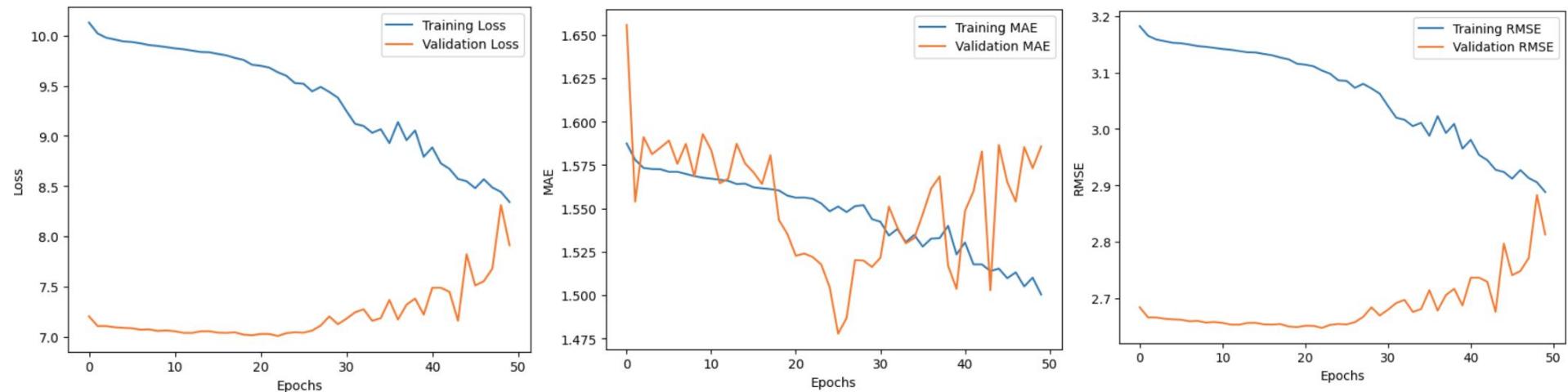
```
model_lstm.compile(loss= tf.losses.MeanSquaredError(),
                    optimizer= tf.optimizers.Adam(),
                    metrics=[tf.metrics.MeanAbsoluteError(), tf.metrics.RootMeanSquaredError()])

history= model_lstm.fit_generator(train_generator, epochs=50,
                                    validation_data= test_generator,
                                    shuffle=False)
```

# windowing techniques



# LSTM Modeli - Sonuçlar



- 50 döngü boyunca eğitilen model yaklaşık **20. döngüye** kadar **normal** seyredip sonrasında **aşırı öğrenme** eğiliminde davranışlar gerçekleştirmiştir.
- Eğitimlerin tamamlanmasının ardından **test verisi üzerinde en başarılı** olan model kaydedilmiştir.

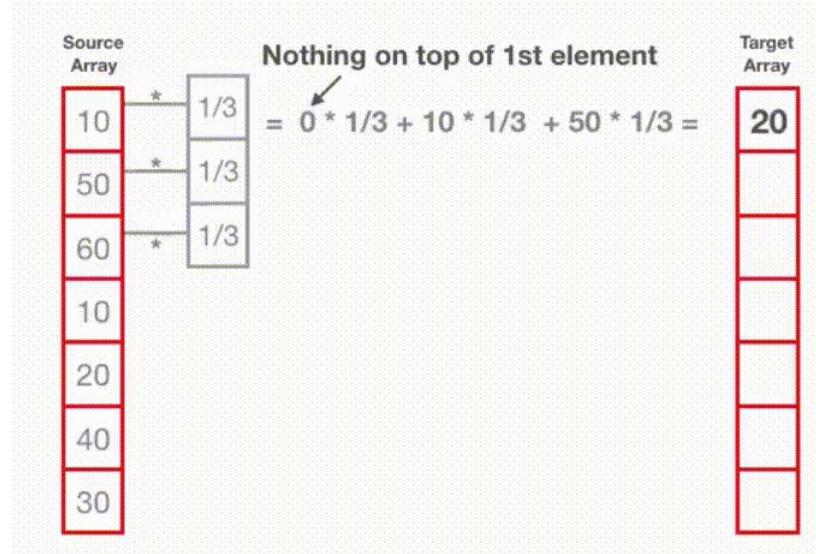
# LSTM Modeli - Çıktılar

- Test verisi olarak kullanılan kısımdan alınan örnekler, modelin tahminlemesini gözlemlemek için modele verilmiştir.
- Modelin tahminleri diğer örnekler'e göre daha çok suç işlenen günlerde **stabil kalmaya** devam etmiştir.
- Bu çıktılar modelin geçmişteki verilere bağlı olduğunu ve kendinden hemen önceki veriden dolayı **stabil durumunu bozmadığını** gösterir.

	Date	Actual	Predictions
2022-11-23 22:30:00	3	2.622087	
2022-11-23 22:38:00	1	2.538553	
2022-11-23 22:40:00	2	1.468421	
2022-11-23 22:42:00	1	1.571744	
2022-11-23 22:43:00	1	1.643446	
...	...	...	...
2022-11-24 02:00:00	9	2.693014	
2022-11-24 02:03:00	1	2.178812	
2022-11-24 02:08:00	2	2.809929	
2022-11-24 02:12:00	1	2.299984	
2022-11-24 02:15:00	2	2.006717	

# Kullanılan Modeller - 1D Convolution

- 1D konvolüsyonlar, zaman serilerindeki **örüntüleri** yakalamak için etkili bir yöntemdir.
- 1D konvolüsyonları yerel örüntüleri algılamakta **zaman ölçüğünü** dikkate aldığı için, parametre sayısını azaltır.
- Belirlenen **pencere boyutu** ile veri üzerinde gezinmekte ve örüntüleri **geçmişe bağlı** verilerin **geleceği nasıl etkilediğini** analiz ederek tespit etmektedir.



# 1D Konvolüsyon Modeli

- Modelin Yapısı:
- Girdilerimiz LSTM Modeli ile aynı olup katmanları farklılaştırılmıştır.
- Veri, 64 adet filtresi bulunan 3 katmandan, **kayan pencereler** yöntemi ile geçmiştir.
- Veri yükleme değişkenleri, epoch sayısı, optimize edici gibi hiperparametreler LSTM ile aynıdır.

```
input_layer = tf.keras.layers.Input(input_shape)

conv1 = tf.keras.layers.Conv1D(filters=64, kernel_size=3, padding="same")(input_layer)
conv1 = tf.keras.layers.BatchNormalization()(conv1)
conv1 = tf.keras.layers.ReLU()(conv1)

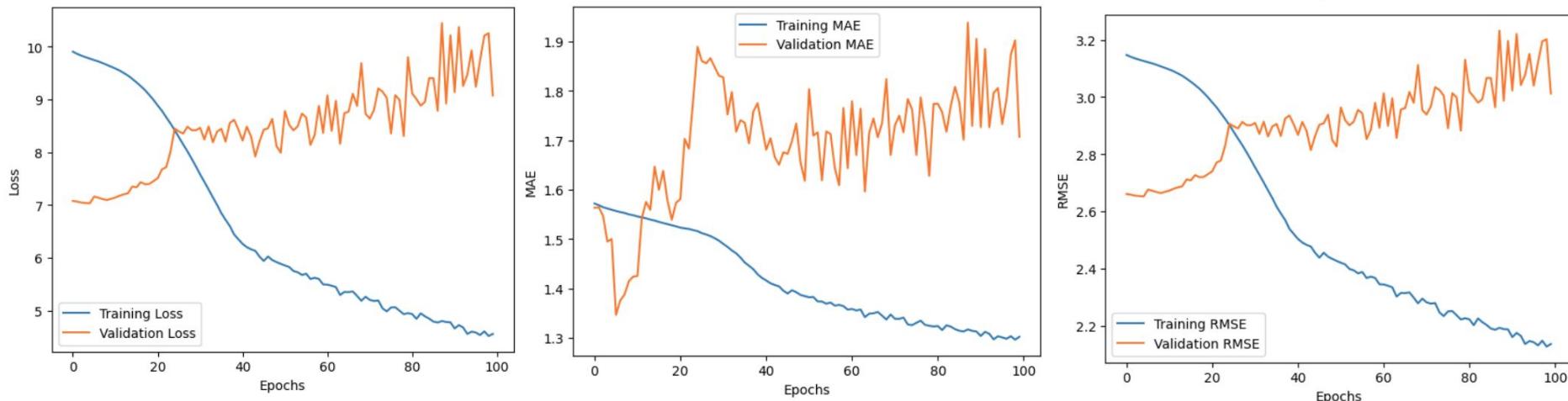
conv2 = tf.keras.layers.Conv1D(filters=64, kernel_size=3, padding="same")(conv1)
conv2 = tf.keras.layers.BatchNormalization()(conv2)
conv2 = tf.keras.layers.ReLU()(conv2)

conv3 = tf.keras.layers.Conv1D(filters=64, kernel_size=3, padding="same")(conv2)
conv3 = tf.keras.layers.BatchNormalization()(conv3)
conv3 = tf.keras.layers.ReLU()(conv3)

gap = tf.keras.layers.GlobalAveragePooling1D()(conv3)

output_layer = tf.keras.layers.Dense(1)(gap)
```

# 1D Konvolüsyon Modeli - Sonuçlar



- Loss değerlerine bakıldığında modelin **30. döngüye kadar** düzgün bir şekilde öğrendiği, sonrasında ise **aşırı öğrenme (overfit)** gerçekleştiği görülmüştür.
- MAE** ve **RMSE** metriklerinde de bu sürecin aynı şekilde tamamlandığı görülmektedir.
- Eğitim verisinde stabil, test verisinde ise dalgalı bir grafik izlemesi beklenen bir çıktı olup, tahminlemede kullanılan model test verisinde **en başarılı** olan döngüden alınarak kullanılmıştır.

# 1D Konvolüsyon Modeli - Çıktılar

- Test verisi olarak kullanılan kısımdan alınan örnekler, modelin tahminlemesini gözlemlemek için modele verilmiştir.
- Görüldüğü üzere model çoğu örnekte asıl çıktıya yaklaşmıştır.
- Asıl değerin **9** olduğu örnekten sonraki tahminini **8.67** olarak yapması ise **geçmiş verilere olan bağlılığını** kanıtlamaktadır.

		Actual	Predictions
	Date		
2022-11-23 22:30:00	3	3.426759	
2022-11-23 22:38:00	1	0.906691	
2022-11-23 22:40:00	2	1.826573	
2022-11-23 22:42:00	1	2.507490	
2022-11-23 22:43:00	1	2.353150	
...	...	...	...
2022-11-24 02:00:00	9	2.070238	
2022-11-24 02:03:00	1	8.671225	
2022-11-24 02:08:00	2	1.716118	
2022-11-24 02:12:00	1	0.621413	
2022-11-24 02:15:00	2	0.669833	

---

---

# Dinlediğiniz için teşekkürler!

— Soru/Cevap ve Tartışma Bölümü —

---

---