

15 DECEMBER 2019

# **ANALYSIS AND FORECASTING CRIME IN BOSTON WITH ARIMA SARIMA**

Indira Insiyah Nastiti - 1706986006  
Probabilitas-Stokastik 03

# 911

## Why We Need?



## WORLD OF CRIMES

## Background

Kriminalitas merupakan suatu perbuatan atau tindakan yang mengacu pada perlakuan melawan hukum dan merugikan pihak lain. Kriminalitas menjadi suatu masalah sosial yang tidak dapat dihindari dalam kehidupan bermasyarakat. Tindak kejahatan ini pun menjadi ancaman bagi masyarakat luas dan cukup meresahkan. Berdasarkan rekam jejak kejadian kriminalitas, dapat diketahui bahwa tingkat kemunculannya terkadang dapat mengalami penurunan atau peningkatan. Oleh karena itu diperlukannya suatu informasi untuk dapat mengetahui kemungkinan-kemungkinan kejahatan yang akan terjadi selanjutnya agar dapat dilakukannya tindakan pencegahan baik masyarakat maupun aparat hukum.

Penelitian ini mengembangkan model prediksi terhadap kemungkinan tindakan kriminal di Boston dengan mengkombinasikan metode ARIMA dan SARIMA. metode pramalan yang digunakan ialah berdasarkan data timeseries untuk menghitung probabilitas kemungkinan dimasa mendatang baik dalam jangka pendek (ARIMA) dan jangka panjang (SARIMA).

# CASE AND SOURCE

Database CRIME IN BOSTON merupakan kumpulan data yang disediakan oleh Departemen Kepolisian Boston (BPD), Analyze Boston. Berisikan rincian insiden yang ditangani oleh BPD. Data ini merupakan catatan dari sistem laporan insiden kejahatan yang kemudian diperkecil fokusannya dalam jenis insiden, waktu serta lokasi terjadinya insiden.

Kumpulan data didapat dari hasil rekap jejak yang dimulai pada 1 Januari 2018 hingga 3 September 2018.

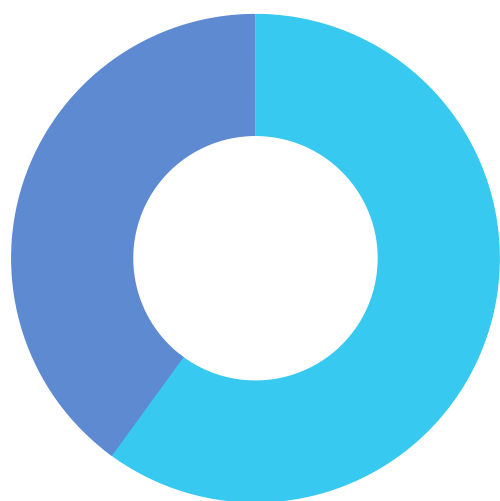
Dalam pemaparan ini, akan dilakukannya proses stokastik pada Database CRIME IN BOSTON dengan melakukan pendekatan pada metode ARIMA (Autoregressive Integrated Moving Average) dan SARIMA (Seasonal ARIMA) dalam membuat prediksi. Berdasarkan konsepnya ARIMA serta SARIMA menggunakan nilai masa lalu dan saat ini dari deret waktu secara statistik yang berhubungan satu sama lain untuk memprediksi jangka pendek (ARIMA) dan jangka panjang (SARIMA).



# 911

## Data Crime in BOSTON

Offense\_Codes  
40%



Crimeinboston  
60%

## DATA WE USED

```
#Import file offense_codes
oc = pd.read_csv (r'https://raw.githubusercontent.com/IndiraInsiyah/Crime-Boston1/master/offense\_codes.csv', encoding = "ISO-8859-1")
print (oc)
```

	CODE	NAME
0	612	LARCENY PURSE SNATCH - NO FORCE
1	613	LARCENY SHOPLIFTING
2	615	LARCENY THEFT OF MV PARTS & ACCESSORIES
3	1731	INCEST
4	3111	LICENSE PREMISE VIOLATION
..	...	...
571	1806	DRUGS - CLASS B TRAFFICKING OVER 18 GRAMS
572	1807	DRUGS - CLASS D TRAFFICKING OVER 50 GRAMS
573	1610	HUMAN TRAFFICKING - COMMERCIAL SEX ACTS
574	2010	HOME INVASION
575	1620	HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE

[576 rows x 2 columns]

## OFFENSE\_CODES

```
#Import file crimesinboston
cib = pd.read_csv (r'https://raw.githubusercontent.com/IndiraInsiyah/Crime-Boston1/master/crimeinboston.csv', encoding = "ISO-8859-1")
cib = cib.fillna(0)
print (cib)
```

	INCIDENT_NUMBER	OFFENSE_CODE	...	Long	Location
0	I182070945	619	...	-71.139371	(42.35779134, -71.13937053)
1	I182070943	1402	...	-71.060300	(42.30682138, -71.06030035)
2	I182070941	3410	...	-71.072429	(42.34658879, -71.07242943)
3	I182070940	3114	...	-71.078664	(42.33418175, -71.07866441)
4	I182070938	3114	...	-71.090361	(42.27536542, -71.09036101)
...	...	...	...	...	...
65502	I182000021	724	...	-71.090898	(42.28976255, -71.09089843)
65503	I182000020	2405	...	-71.048724	(42.35325379, -71.04872410)
65504	I182000019	3301	...	-71.112333	(42.26133580, -71.11233325)
65505	I182000017	3304	...	-71.111603	(42.33225457, -71.11160342)
65506	I182000016	3831	...	0.000000	(0.00000000, 0.00000000)

[65507 rows x 17 columns]



## METHOD AND RESEARCH

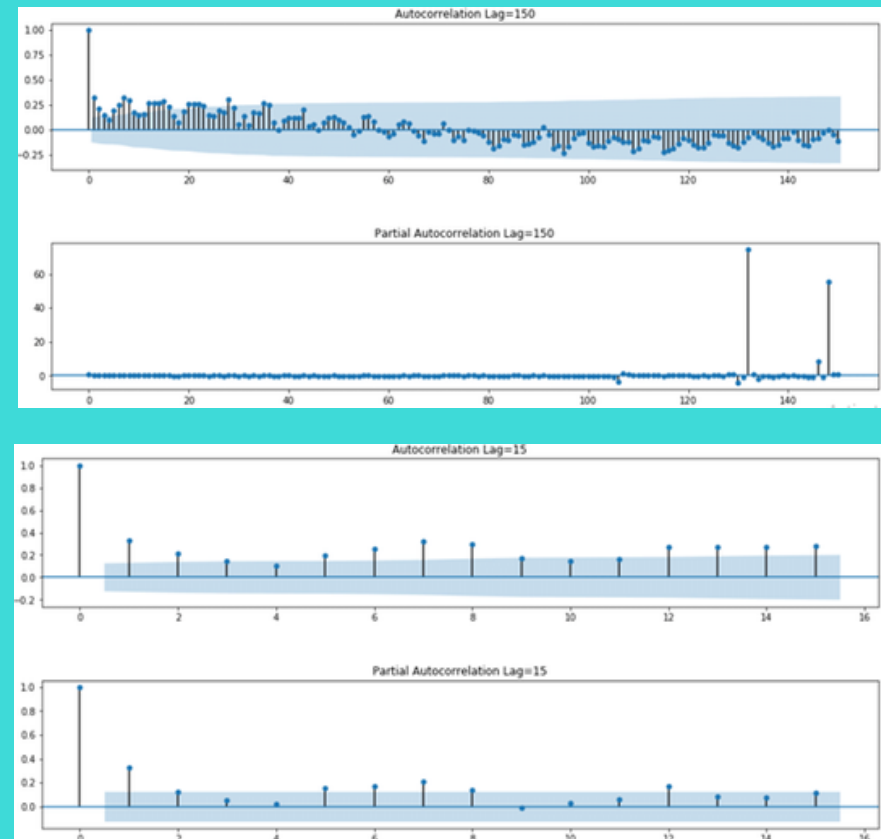
---

ARIMA (Autoregressive Integrated Moving Average) dan SARIMA (Seasonal Autoregressive Integrated Moving Average) merupakan salah satu permodelan stokastik dalam menentukan peramalan masa mendatang menggunakan data timeseries sebagai variabel yang diacu dalam perhitungan, yaitu banyak kasus, lokasi, tanggal dan waktu kejadian.

# IDENTIFICATION TIME SERIES

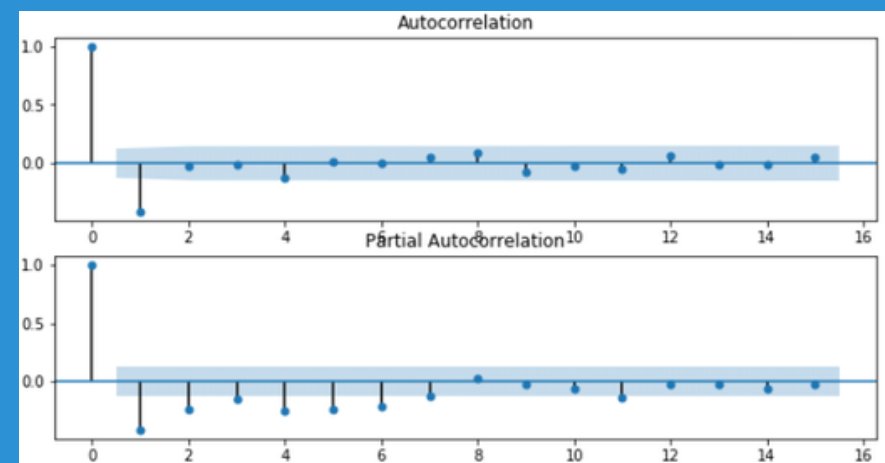
## CRIME BY DATE

ACF AND PACF TEST



## CRIME BY TIME

ACF AND PACF TEST



## TIME SERIES

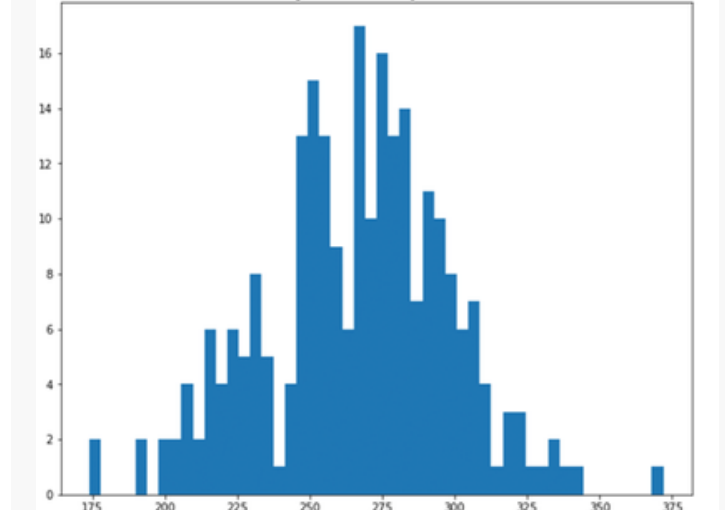
STATIONARY TEST

Average= -0.30612244897959184  
Std= 37.021374022220485  
SE= 2.365208826689423

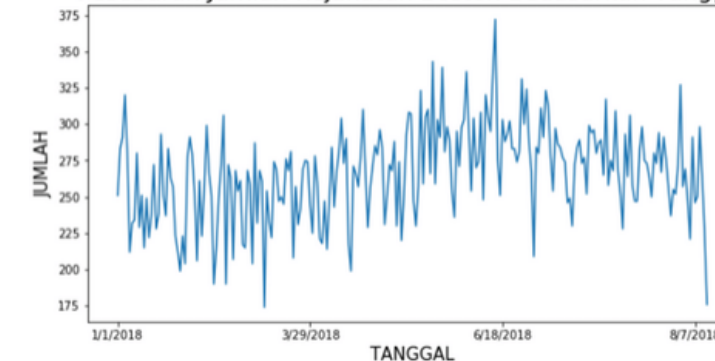
## Number of Crimes per Day

	TANGGAL	JUMLAH
2/25/2018	2/25/2018	207
2/11/2018	2/11/2018	206
2/7/2018	2/7/2018	204
1/7/2018	1/7/2018	204
1/5/2018	1/5/2018	199
4/16/2018	4/16/2018	199
2/18/2018	2/18/2018	190
2/22/2018	2/22/2018	190
9/3/2018	9/3/2018	176
3/11/2018	3/11/2018	174

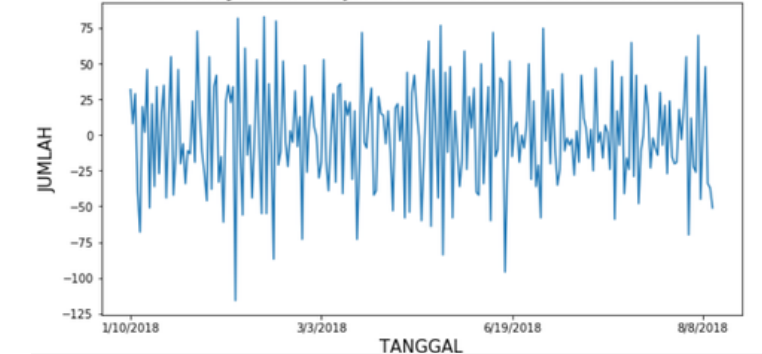
Pesebaran Jumlah Kejahatan Boston



Pesebaran Jumlah Kejahatan Boston Berdasarkan Tanggal



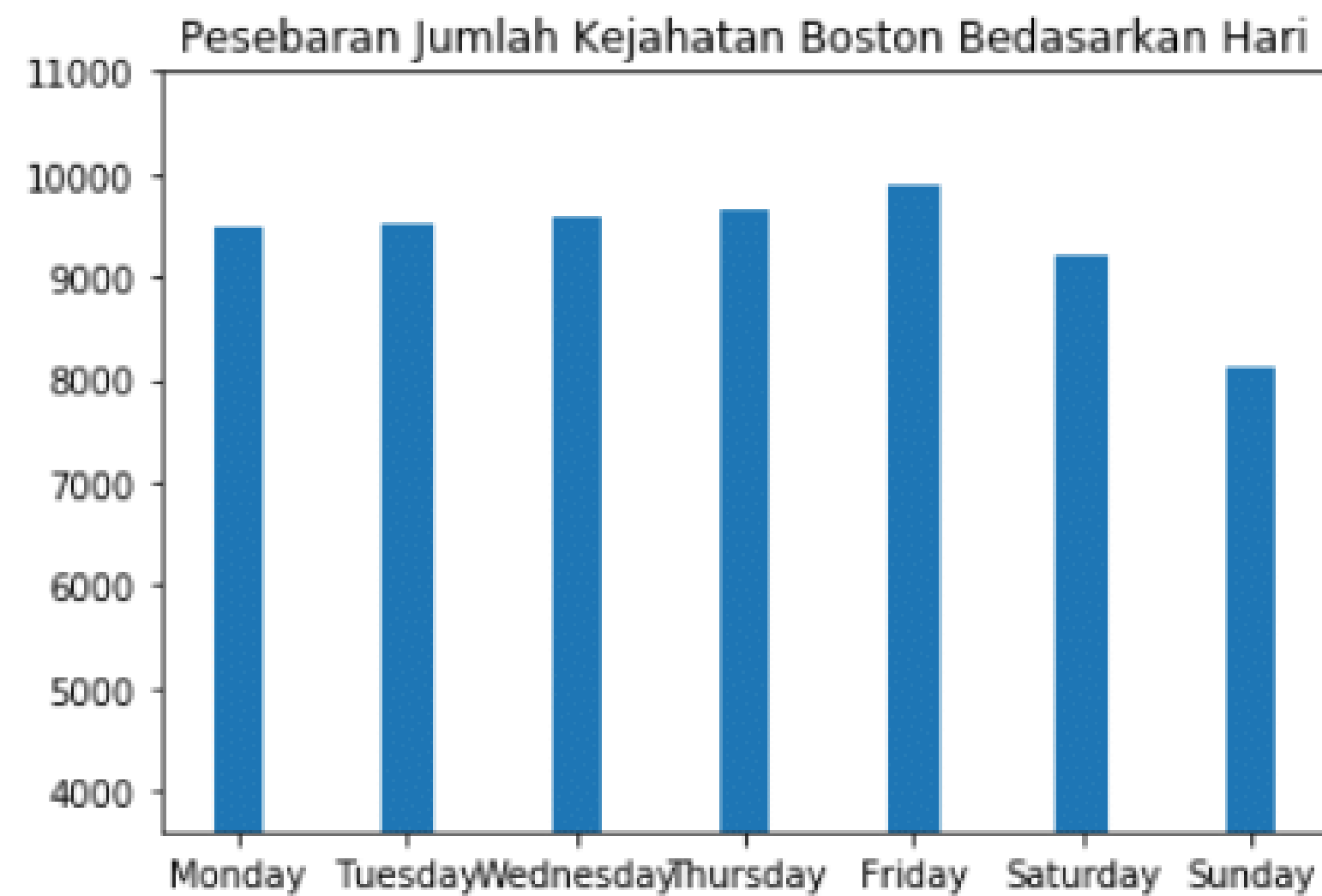
Pesebaran Jumlah Kejahatan Boston Berdasarkan Waktu



Bedasarkan pola grafik, data tidak menunjukkan nilai statis sehingga perlu dilakukan pengolahan ARIMA dan SARIMA. ACF menunjukkan pola dying down, dan PACF menunjukkan cut off, maka dapat dikatakan model ARIMA berupa AR murni.



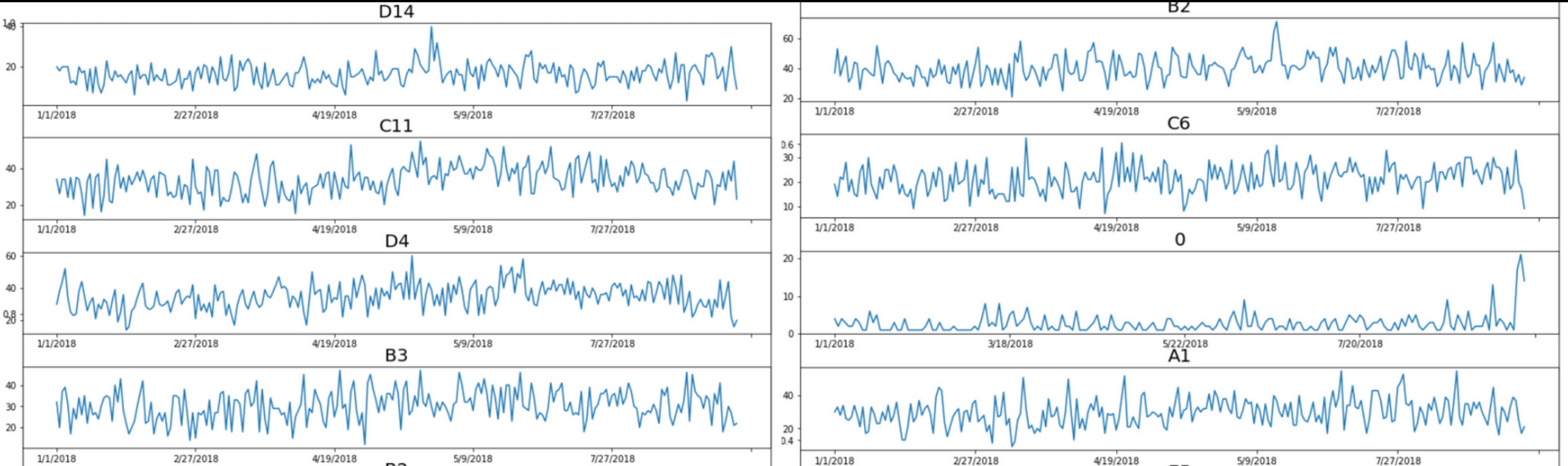
Tingkat kejahatan di Boston cenderung tinggi pada hari Jumat dan cenderung rendah pada hari Minggu.





# DISTRIBUTION CRIME IN DISTRICT

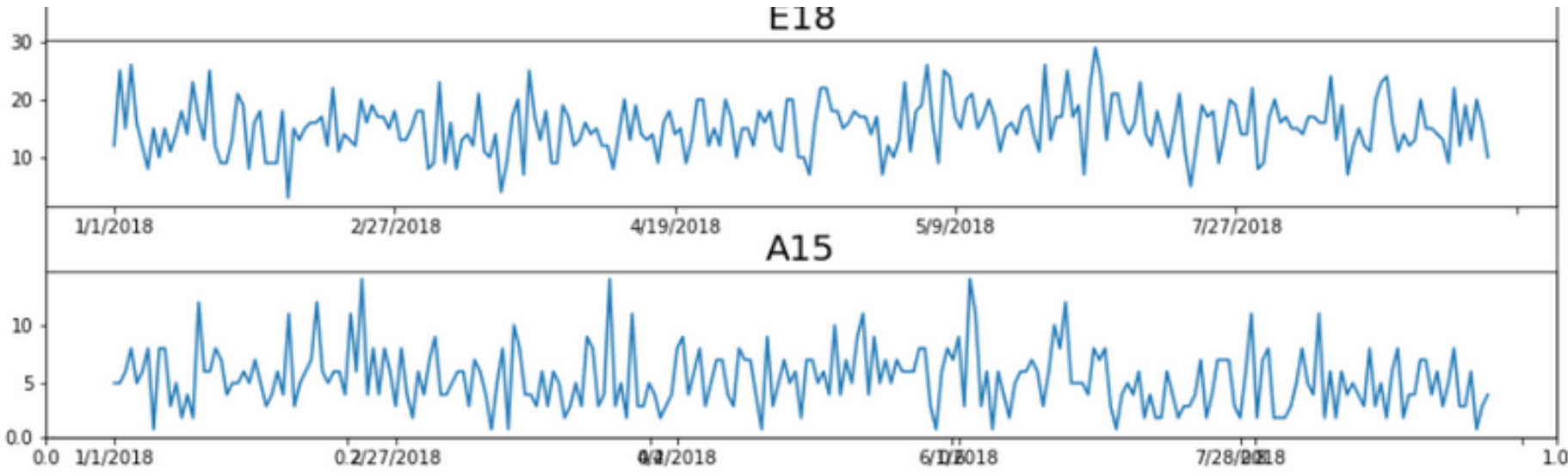
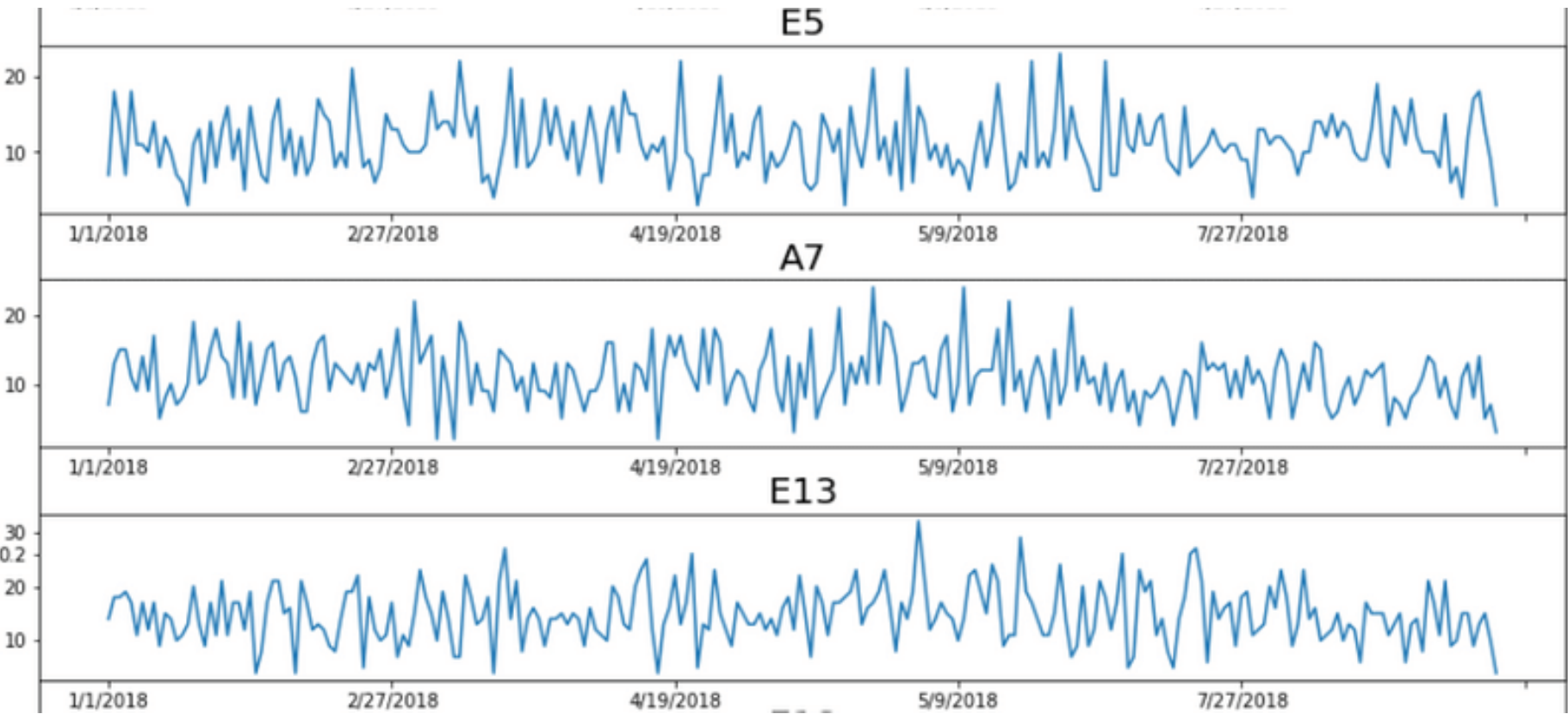
13 District





# DISTRIBUTION CRIME IN DISTRICT

13 District



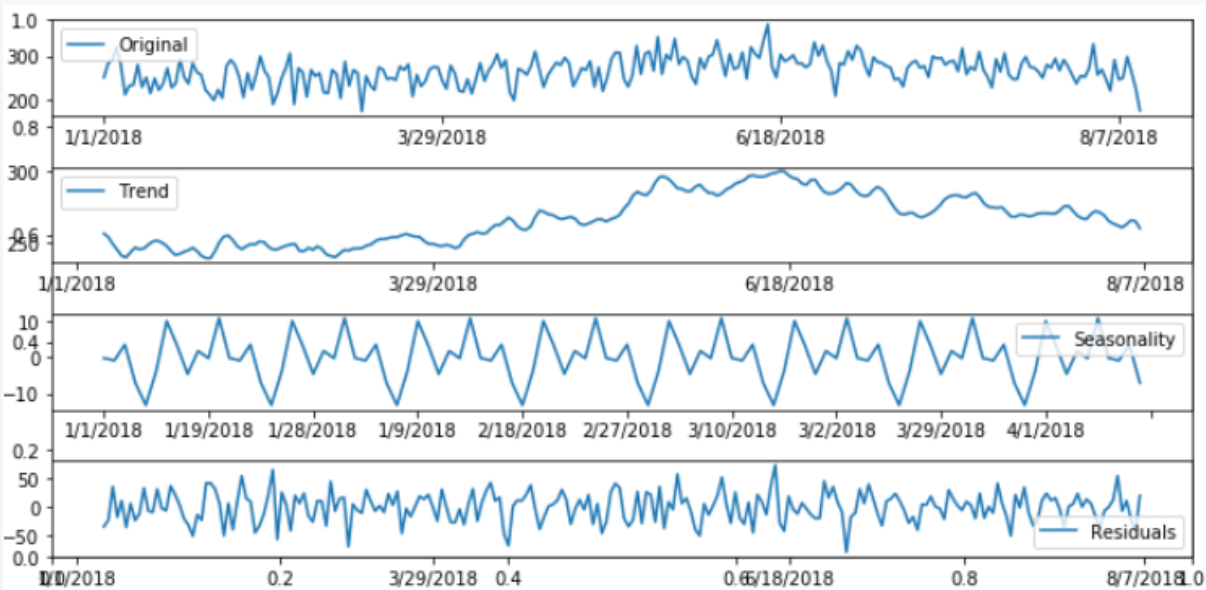
# ARIMA AND SARIMA

## Results: ARIMA

```
=====
del:          ARIMA          BIC:          2378.910
pendent Variable: D.JUMLAH    Log-Likelihood: -1167.5
te:           2019-12-15 14:42 Scale:        1.0000
. Observations: 245          Method:        css-mle
  Model:       7             Sample:        1
  Residuals:   238           S.D. of innovations: 28.268
nverged:      1.0000         HQIC:          2362.180
. Iterations: 20.0000
C:            2350.9002
=====
```

	Coef.	Std.Err.	t	P> t	[0.025	0.975
nst	-0.0064	0.1651	-0.0387	0.9692	-0.3299	0.31
.L1.D.JUMLAH	0.3304	0.2648	1.2477	0.2134	-0.1886	0.84
.L2.D.JUMLAH	-0.0844	0.0753	-1.1216	0.2632	-0.2320	0.06
.L3.D.JUMLAH	-0.1092	0.0683	-1.5986	0.1112	-0.2431	0.02
.L4.D.JUMLAH	-0.1681	0.0800	-2.1007	0.0367	-0.3249	-0.01
.L1.D.JUMLAH	-1.1526	0.2662	-4.3296	0.0000	-1.6744	-0.63
.L2.D.JUMLAH	0.2425	0.2428	0.9984	0.3191	-0.2335	0.71

	Real	Imaginary	Modulus	Frequency
AR.1	0.9283	-0.9390	1.3204	-0.1259
AR.2	0.9283	0.9390	1.3204	0.1259
AR.3	-1.2532	-1.3571	1.8472	-0.3687
AR.4	-1.2532	1.3571	1.8472	0.3687
MA.1	1.1419	0.0000	1.1419	0.0000
MA.2	3.6121	0.0000	3.6121	0.0000



## SARIMAX

	order	seasonal_order	AIC
27	(0, 1, 1)	(0, 1, 1, 7)	2284.353577
59	(1, 1, 1)	(0, 1, 1, 7)	2286.202742
31	(0, 1, 1)	(1, 1, 1, 7)	2286.304914
43	(1, 0, 1)	(0, 1, 1, 7)	2288.072394
63	(1, 1, 1)	(1, 1, 1, 7)	2288.132790

## Statespace Model Results

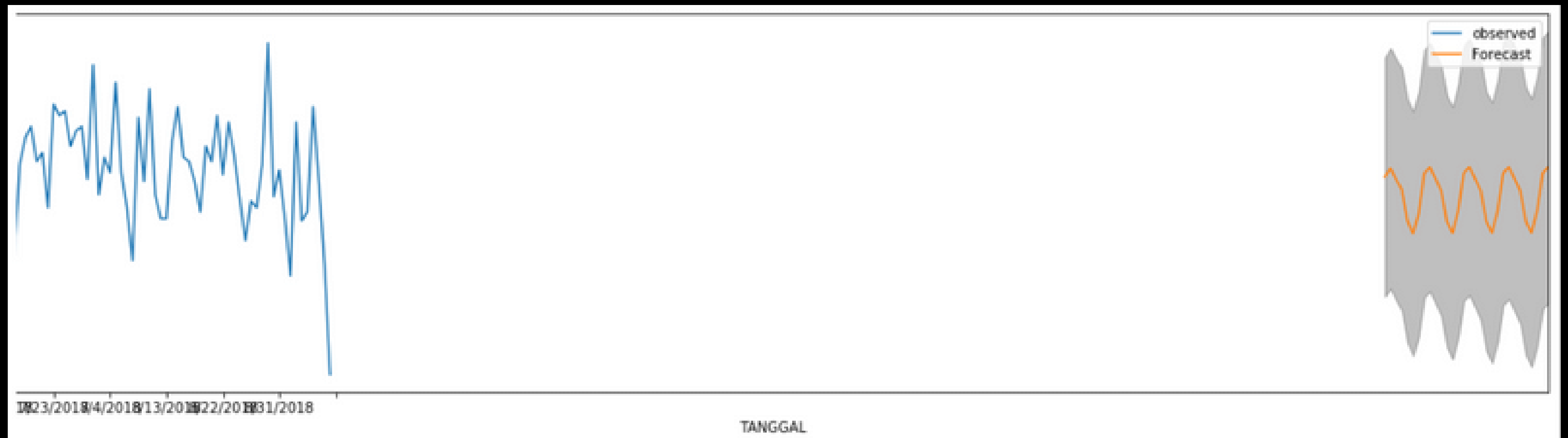
```
=====
Dep. Variable:          JUMLAH    No. Observations:          246
Model:                  SARIMAX(1, 1, 1)x(1, 1, 1, 7)    Log Likelihood          -1139.066
Date:                   Sun, 15 Dec 2019    AIC                    2288.133
Time:                   06:05:41    BIC                    2305.494
Sample:                 0    HQIC                    2295.130
                        - 246
Covariance Type:       opg
=====
```

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.0292	0.081	0.362	0.717	-0.129	0.187
ma.L1	-0.9128	0.042	-21.768	0.000	-0.995	-0.831
ar.S.L7	0.0185	0.070	0.264	0.792	-0.119	0.156
ma.S.L7	-0.9992	3.396	-0.294	0.769	-7.655	5.657
sigma2	748.6101	2506.848	0.299	0.765	-4164.722	5661.942

```
=====
Ljung-Box (Q):          44.23    Jarque-Bera (JB):          0.25
Prob(Q):                0.30    Prob(JB):              0.88
Heteroskedasticity (H): 0.72    Skew:                  0.08
Prob(H) (two-sided):    0.15    Kurtosis:              2.97
=====
```

# FORECAST RESULT

For the next 30 days

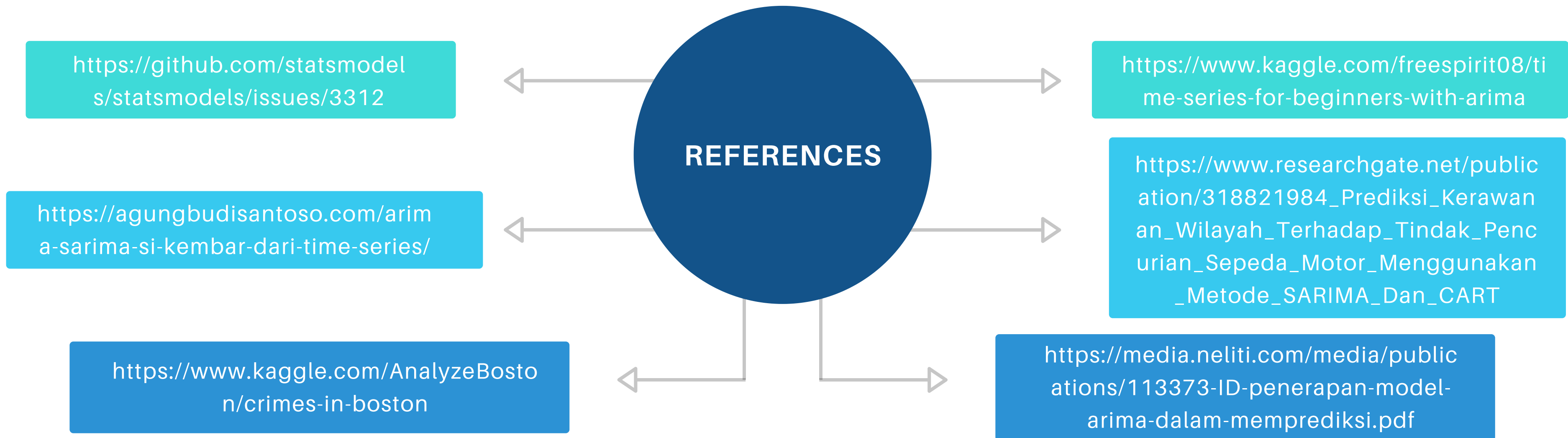


Data's  
Original

Prediksi data mengalami  
siklus yang cukup statis

Data's  
Forecast

# Web Most Used



**PROJECT'S LINK**  
**THANK YOU**

**[HTTPS://GITHUB.COM/INDIRAINSIYAH  
/CRIME-BOSTON1](https://github.com/INDIRAINSIYAH/CRIME-BOSTON1)**