



Ojalá que lluevan series temporales

I AM

Elena Abril Medina



Ing. Telecomunicación

5 años trabajando en Data Science & Machine Learning en varias empresas para clientes de distintos sectores y con varios lenguajes

<https://www.linkedin.com/in/elenaabrilmedina/>

@eabrilmedina

Los materiales están disponibles en:

<https://github.com/eabrilmedina/pycones2018>

Preguntas que vamos a resolver hoy



- ¿Qué es una serie temporal?
- ¿Qué componentes tiene una serie temporal?
- ¿Cómo preparamos nuestra serie para poder trabajar con ella?
- ¿Qué modelos de predicción podemos plantear?

Nuestros datos



AEMET publica de forma diaria los datos de 700 estaciones. Sólo son accesibles los últimos 7 días.

<https://datosclima.es/index.htm>

Recopilación de datos

- Desde Mayo de 2013
- Zip mensual con fichero Excel diario con filas en blanco al principio

datosclima.es -Base de datos Meteorológica-

Twitter G+ Me gusta 0

Inicio

- Inicio
- Contactar
- DATOS AEMET -hasta 2012-**
- Situación Estaciones
- Datos diarios
- Temperaturas
- Presión
- Viento
- Precipitación y Sol
- Descarga Datos
- DATOS AEMET Desde Mayo 2013**
- Situación Estaciones
- Temperaturas
- Precipitación
- Viento
- Descarga Datos diarios
- Datos Por Estación
- DATOS MUNDIALES NOAA-GHCN**



Datos Históricos Meteorológicos

¿Cuántas veces has sentido curiosidad por conocer el tiempo que hizo en el pasado en alguna zona de España o del Mundo? ¿Será cierto aquello del cambio climático? ¿Es cierto que en tiempos de nuestros abuelos hacía más frío? ¿Que cantidad de agua cayó en aquellas inundaciones de las que tanto hablan nuestros abuelos?

Estas y otras muchas curiosidades climáticas y meteorológicas las podrás encontrar en esta página.

Consulta on-line de datos de Temperaturas, precipitaciones, vientos, presiones, acaecidas en el pasado, gracias a la Base de datos que implementamos en esta web.

Datos AEMET (hasta 2012):

Hasta Noviembre de 2012, la web de AEMET, estuvo proporcionando datos históricos de más de 100 estaciones meteorológicas de España, repartidas por toda su geografía. A partir de entonces, modificaron su web y el formato de los datos ofrecidos. La época de la que datan los históricos, varía de unas estaciones a otras, algunas de ellas comienzan en los años 1920. Puedes ver la localización de las estaciones meteorológicas que componen esta parte de la web en el menú correspondiente de la izquierda. En total están introducidos más de 2.000.000 de líneas de datos históricos, correspondientes a registros diarios de variables como son:

Datos de Temperatura: Seleccionando una estación meteorológica y un período de tiempo, obtendrás los datos y la gráfica con información de Temperaturas máxima y mínima diarias. Además tendrás información sobre valores extremos y curiosidades acontecidas entorno a las Temperaturas, para esa estación meteorológica.

Datos de Presión: Seleccionando una estación meteorológica y un período de tiempo, obtendrás los datos y la gráfica con información de Presiones máximas y mínimas registradas por la estación meteorológica. Además tendrás información sobre valores extremos y curiosidades acontecidas entorno a las Presiones, para esa estación meteorológica.

Datos de Viento: Podrás consultar datos sobre velocidades medias del viento según los parámetros seleccionados de estación y período de tiempo, así como las

Nuestros datos



- Primeras filas
- Campos con valores y horas
- Campos vacíos

España											
Actualizado: miércoles, 08 mayo 2013 a las 09:00											
Fecha: martes, 07 mayo 2013											
Estación	Provincia	Temperatura máxima (°C)	Temperatura mínima (°C)	Temperatura media (°C)	Racha (km/h)	Velocidad máxima (km/h)	Precipitación 00-24h (mm)	Precipitación 00-06h (mm)	Precipitación 06-12h (mm)	Precipitación 12-18h (mm)	Precipitación 18-24h (mm)
Estaca de Bares	A Coruña	19.5 (14:40)	14.2 (03:40)	16.8	95 (15:40)	69 (15:40)	0.4	0	0.2	0.2	0
As Pontes	A Coruña	17.8 (15:10)	12.9 (01:10)	15.3			16.8	10.8	3	3	0
A Coruña	A Coruña	19.7 (13:30)	15.0 (00:00)	17.4	66 (19:10)	31 (14:50)	1.2	1	0	0.2	0
A Coruña Aeropuerto	A Coruña	19.2 (13:30)	14.8 (00:00)	17	68 (19:30)	39 (19:20)	0.1	0	0	0.1	0
Carballo, Depuradora	A Coruña	18.2 (14:40)	13.2 (00:00)	15.7			3.6	1.8	1.2	0.6	0
Cabo Vilán	A Coruña	15.3 (13:40)	13.2 (00:00)	14.3	81 (17:50)	56 (17:50)	22.1	9.9	8.1	2.5	1.6
Limia	A Coruña	15.7 (14:30)	12.1 (23:59)	13.9			39.8	17	14.4	3	5.4
Fisterra	A Coruña	13.3 (01:30)	12.2 (15:30)	12.8	97 (20:00)	74 (20:00)	18.4	6.8	6.4	3.2	2
Mazaricos	A Coruña	14.6 (14:40)	11.5 (23:30)	13.1			68	37.2	17.2	3.4	10.2
Sobrado	A Coruña	16.6 (15:50)	11.3 (22:20)	14			2.6	1.4	0.4	0.4	0.4
Santiago de Compostela Aeropuerto	A Coruña	17.8 (16:00)	11.7 (21:20)	14.7	58 (15:20)	35 (17:50)			2.5	0.7	2.2



Nuestros datos

Preparamos un csv con los datos agregados para cada día:

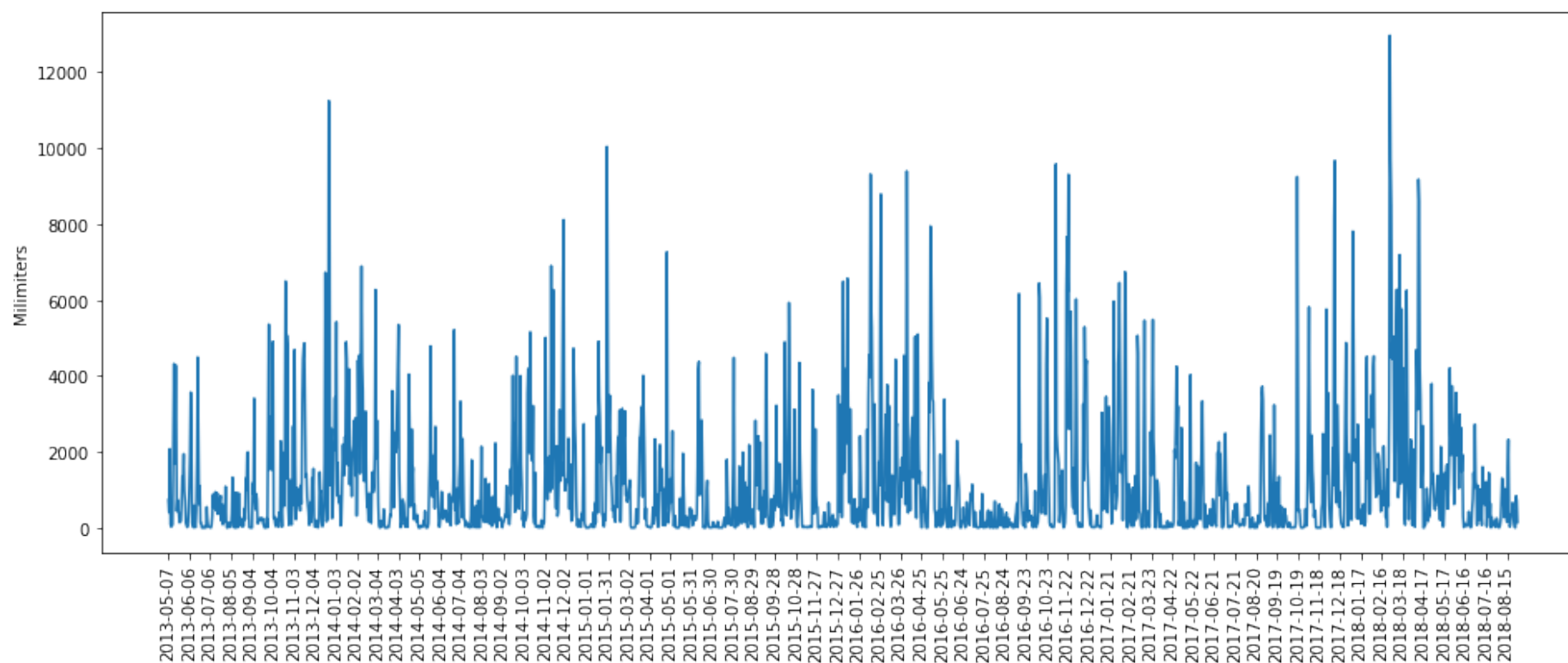
	rain_all	rain_00_06	rain_06_12	rain_12_18	rain_18_24	rn	date	year	month
0	733.7	322.1	190.7	63.7	172.6	0	2013-05-07	2013	5
1	402.3	73.1	155.7	123.0	68.5	1	2013-05-08	2013	5
2	2075.8	89.9	740.3	813.2	469.1	2	2013-05-09	2013	5
3	88.3	46.3	19.8	9.6	13.6	3	2013-05-10	2013	5
4	23.8	6.2	3.6	9.8	4.4	4	2013-05-11	2013	5

- Del 07-05-2013 al 30-08-2017 hay 1.942 días
- Solo 1.936 registros

Nuestros datos

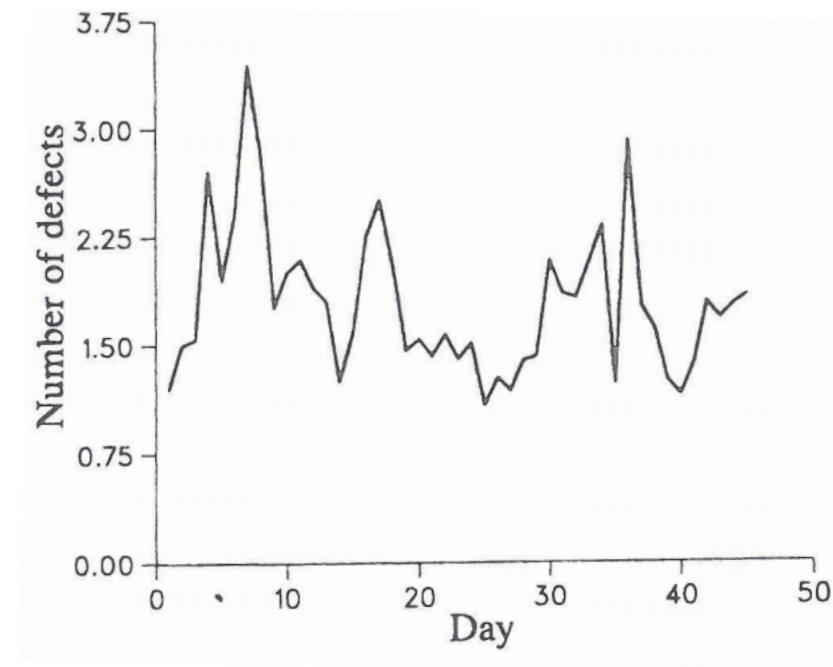


Amount of rain per day



¿Qué es una serie temporal?

Conjunto de valores que representan una misma **variable ordenados** en el **tiempo** en **intervalos regulares**





Creando nuestra serie temporal

```
dateparse = lambda dates: pd.datetime.strptime(dates, '%Y-%m-%d')
data = pd.read_csv('./PyCon/aemet_all.csv', sep=';', parse_dates=['date'], index_col='date', date_parser=dateparse)
data.index
```

```
DatetimeIndex(['2013-05-07', '2013-05-08', '2013-05-09', '2013-05-10',
               '2013-05-11', '2013-05-12', '2013-05-13', '2013-05-14',
               '2013-05-15', '2013-05-16',
               ...,
               '2018-08-21', '2018-08-22', '2018-08-23', '2018-08-24',
               '2018-08-25', '2018-08-26', '2018-08-27', '2018-08-28',
               '2018-08-29', '2018-08-30'],
              dtype='datetime64[ns]', name='date', length=1936, freq=None)
```

```
data = data.asfreq('d')
data.index
```

```
DatetimeIndex(['2013-05-07', '2013-05-08', '2013-05-09', '2013-05-10',
               '2013-05-11', '2013-05-12', '2013-05-13', '2013-05-14',
               '2013-05-15', '2013-05-16',
               ...,
               '2018-08-21', '2018-08-22', '2018-08-23', '2018-08-24',
               '2018-08-25', '2018-08-26', '2018-08-27', '2018-08-28',
               '2018-08-29', '2018-08-30'],
              dtype='datetime64[ns]', name='date', length=1942, freq='D')
```

Componentes de una serie temporal



Serie Temporal

=

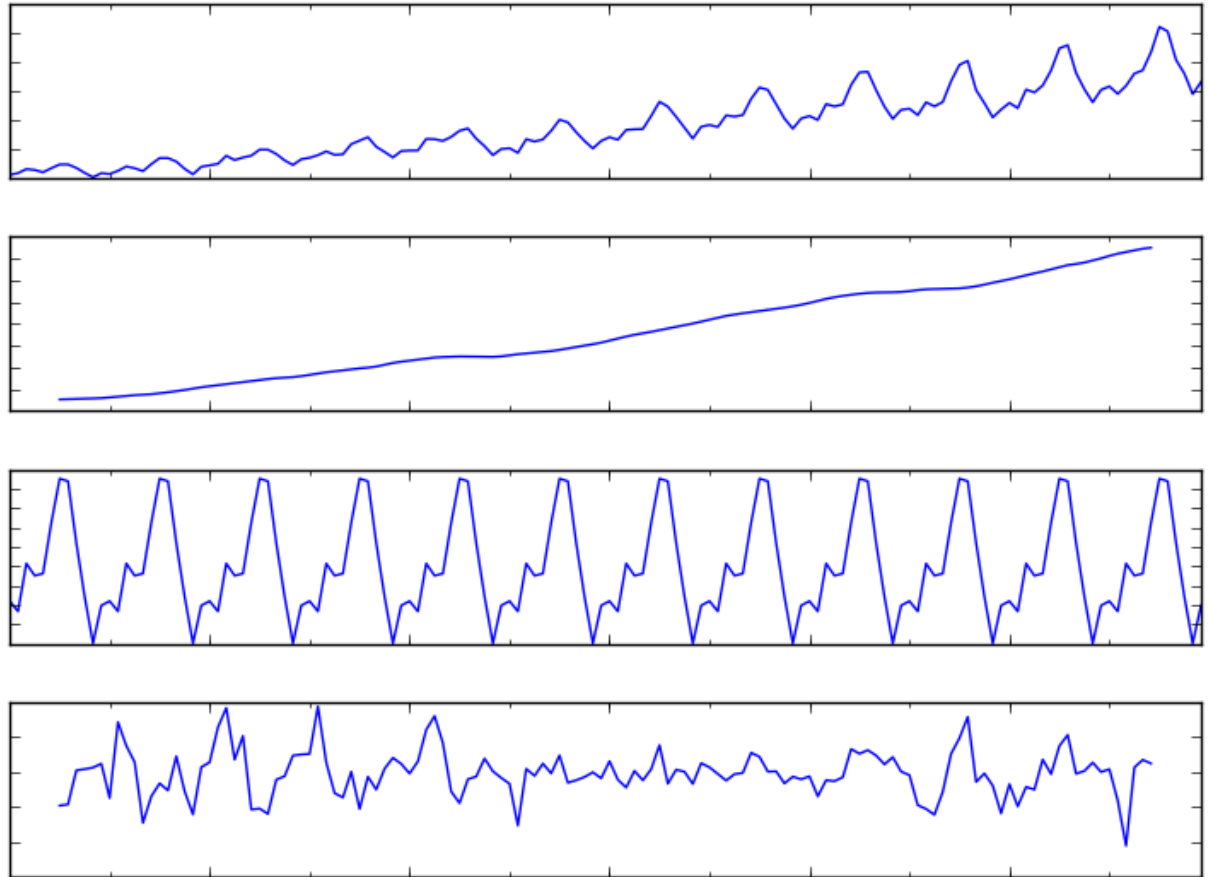
Tendencia

+

Estacionalidad

+

Resto



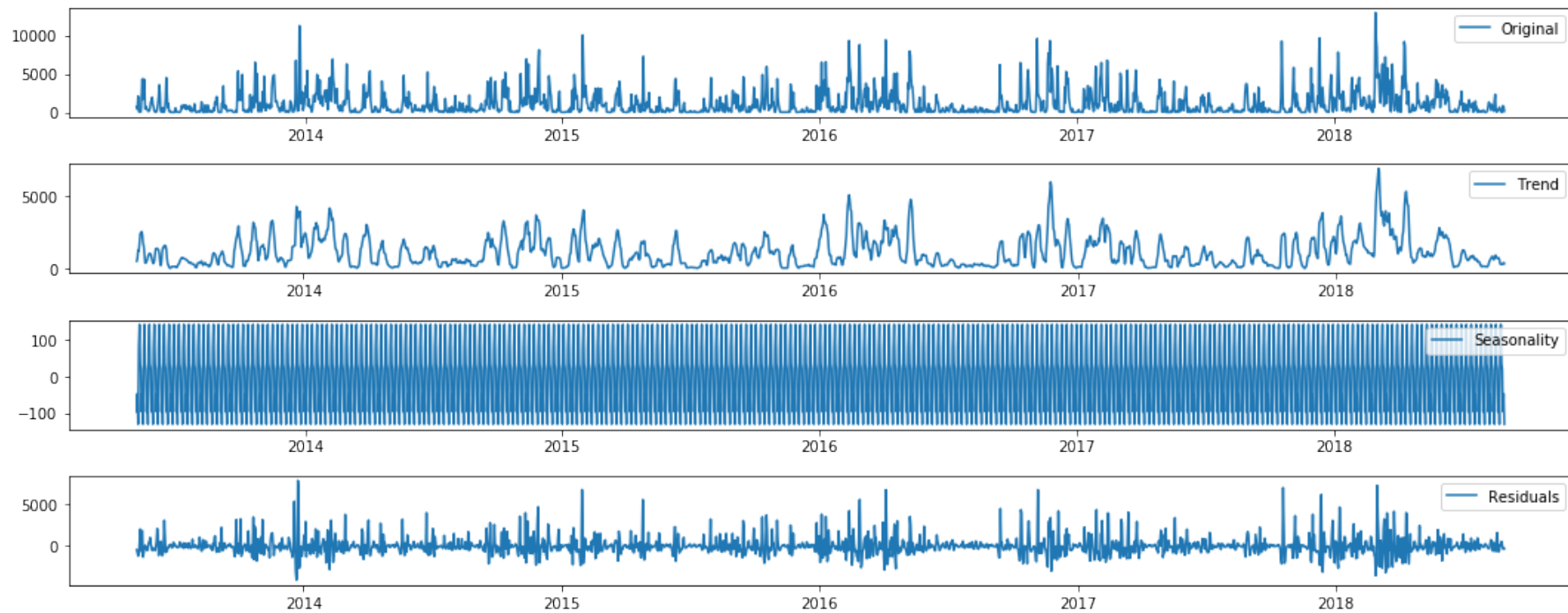
Componentes de la serie temporal



```
decomposition = seasonal_decompose(data['rain_all'])
trend = decomposition.trend
seasonal = decomposition.seasonal
residual = decomposition.resid
```

```
mpl.rc("figure", figsize=(15,6))
plt.subplot(411)
plt.plot(data['rain_all'], label='Original')
plt.legend(loc='best')
plt.subplot(412)
plt.plot(trend, label='Trend')
plt.legend(loc='best')
plt.subplot(413)
plt.plot(seasonal, label='Seasonality')
plt.legend(loc='best')
plt.subplot(414)
plt.plot(residual, label='Residuals')
plt.legend(loc='best')
plt.tight_layout()
```

Componentes de la serie temporal



Vamos a agregar a nivel año-mes

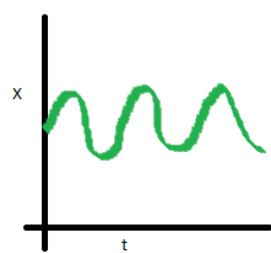
Serie agregada a nivel año-mes



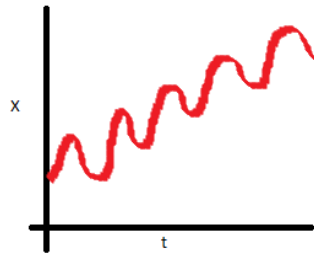
Time decomposition of time series



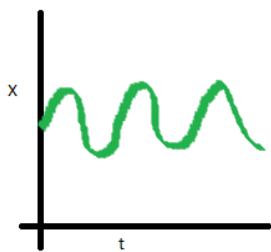
¿Cómo necesitamos que sea la serie?



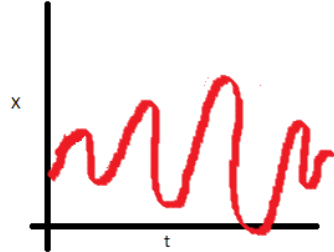
Stationary series



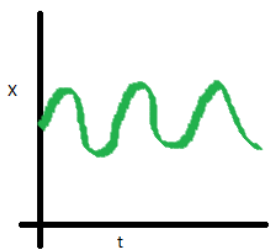
Non-Stationary series



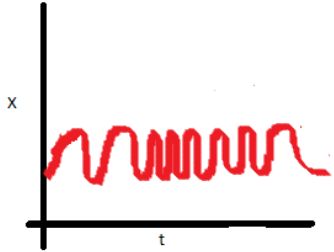
Stationary series



Non-Stationary series



Stationary series



Non-Stationary series

Serie estacionaria

1. Media constante en el tiempo
2. Varianza constante en el tiempo
3. Covarianza constante en el tiempo

En una serie estacionaria sus propiedades no dependen del instante de observación.

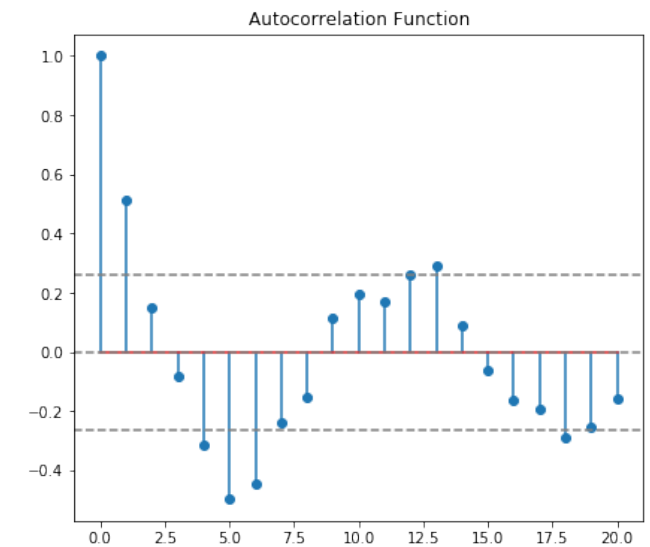
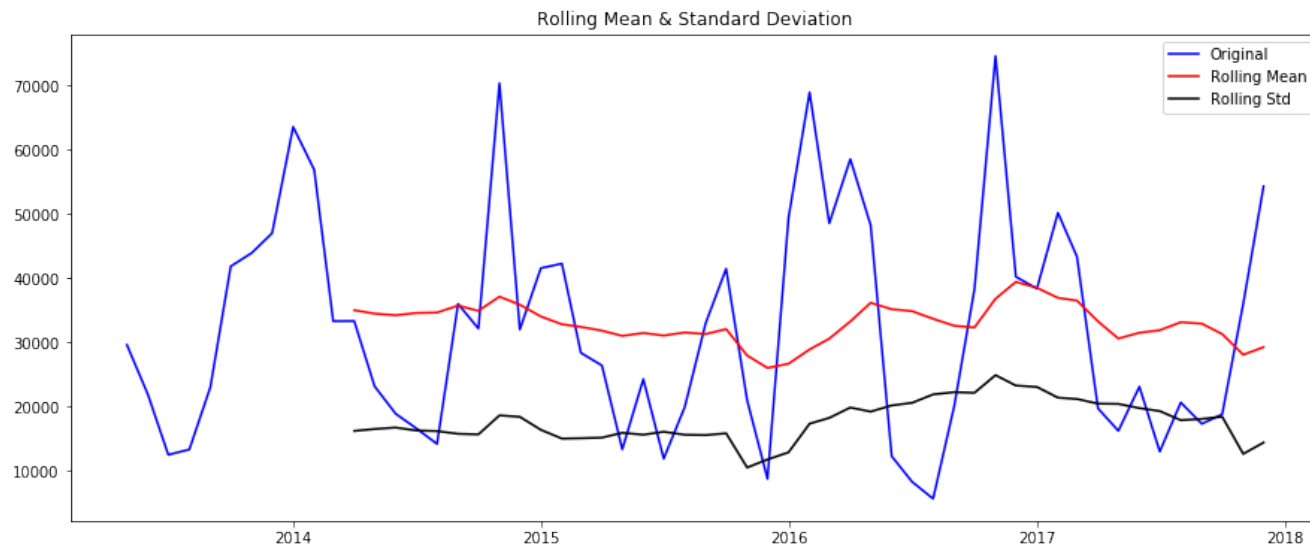
Una serie estacionaria tiene una autocorrelación que decae exponencialmente



¿Nuestra serie es estacionaria?

```
# code found in: https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/def
test_stationarity(tseries):
    #Determining rolling statistics
    rolmean = pd.rolling_mean(tseries, window=12)
    rolstd = pd.rolling_std(tseries, window=12)
    #Plot rolling statistics:
    orig = plt.plot(tseries, color='blue',label='Original')
    mean = plt.plot(rolmean, color='red', label='Rolling Mean')
    std = plt.plot(rolstd, color='black', label = 'Rolling Std')
    plt.legend(loc='best')
    plt.title('Rolling Mean & Standard Deviation')
    plt.show(block=False)
```

¿Nuestra serie es estacionaria?



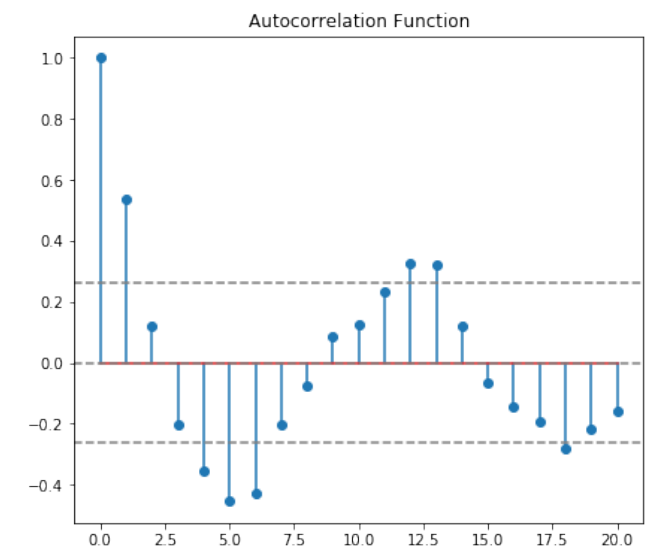
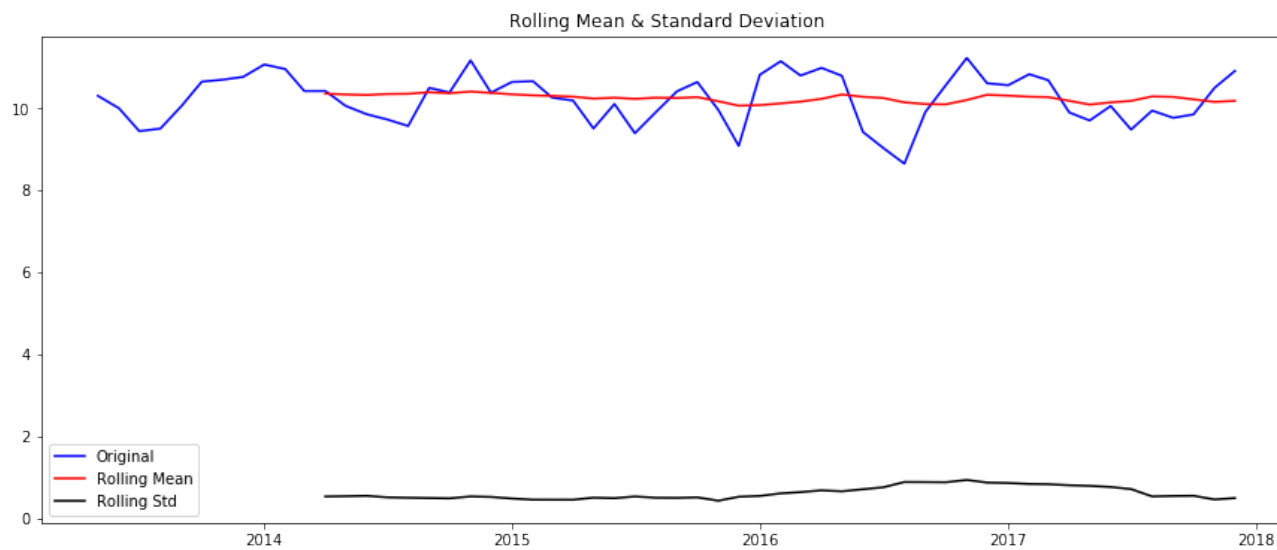


¿Cómo preparamos nuestra serie?

1. Transformación logarítmica ➤ `ts_log = np.log(ts)`
2. Diferenciación ➤ `ts_log = ts_log.shift()`
3. Ajuste de estacionalidad ➤ `ts_log = pd.rolling_mean(ts_log, 12)`
4. Suavizado EDA ➤ `ts_log = pd.ewma(ts_log, halflife=12)`

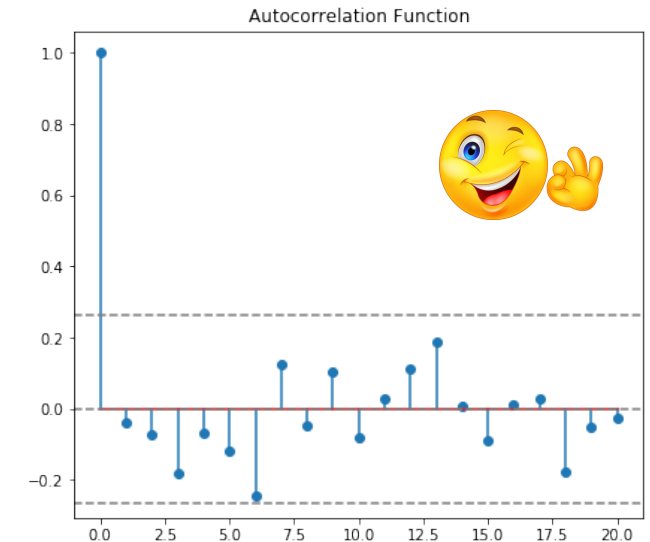
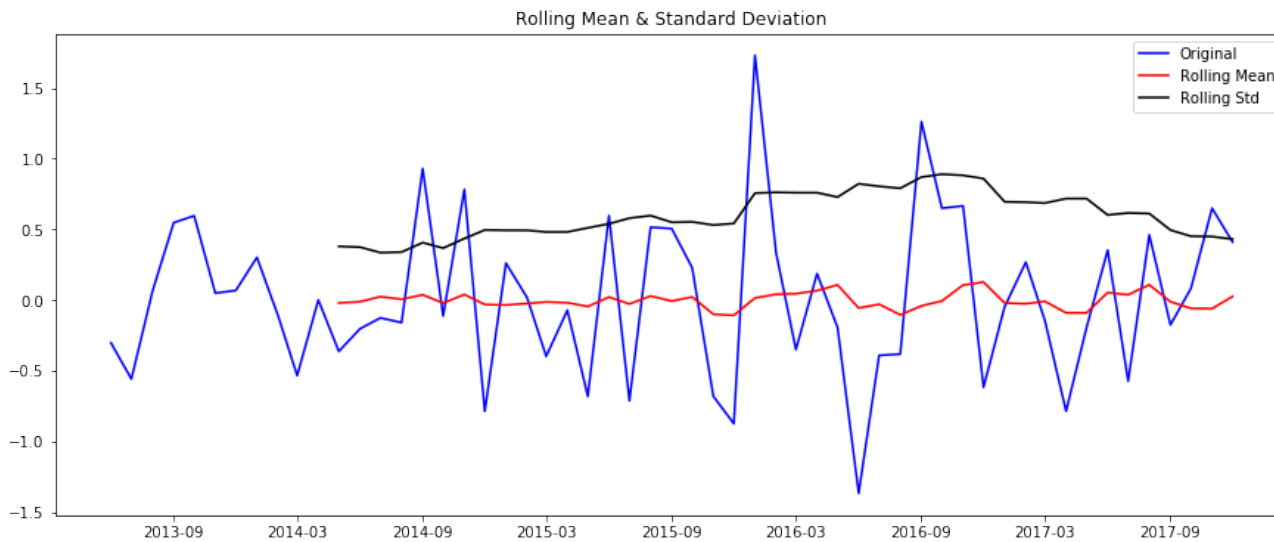
¿Cómo preparamos nuestra serie?

Transformación logarítmica



¿Cómo preparamos nuestra serie?

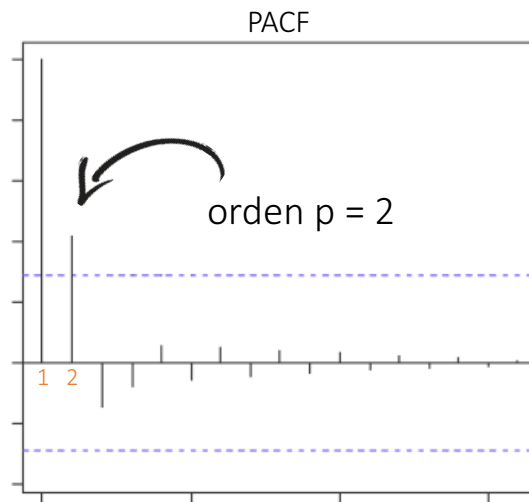
Transformación logarítmica + diferenciación



¿Qué modelos podemos plantear?

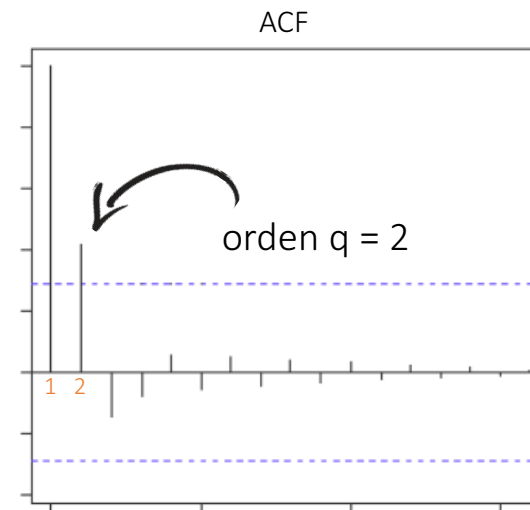
Modelo AR

Auto + Regresión
=
Auto + serie “retrasada” p veces

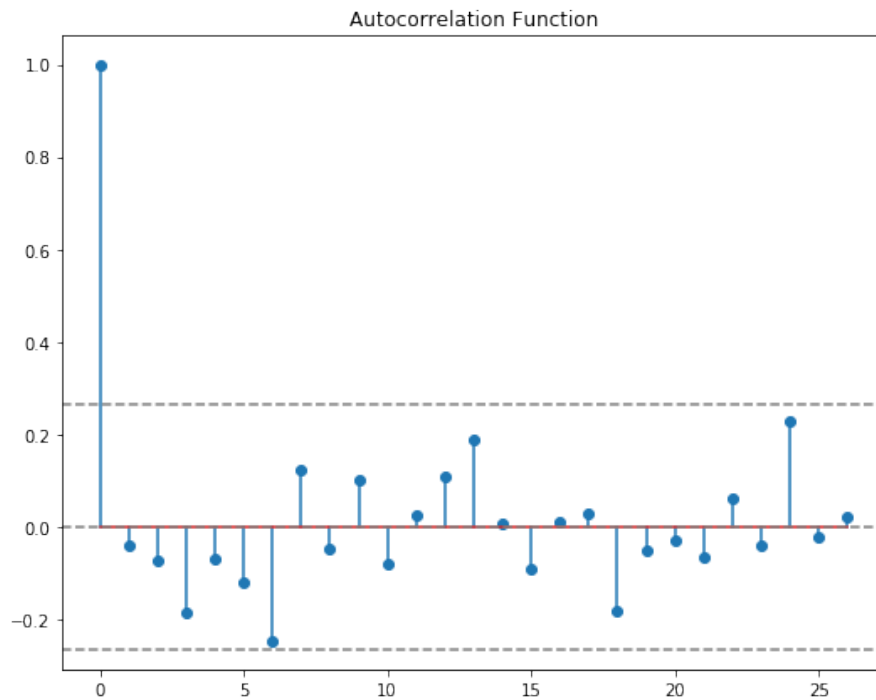


Modelo MA

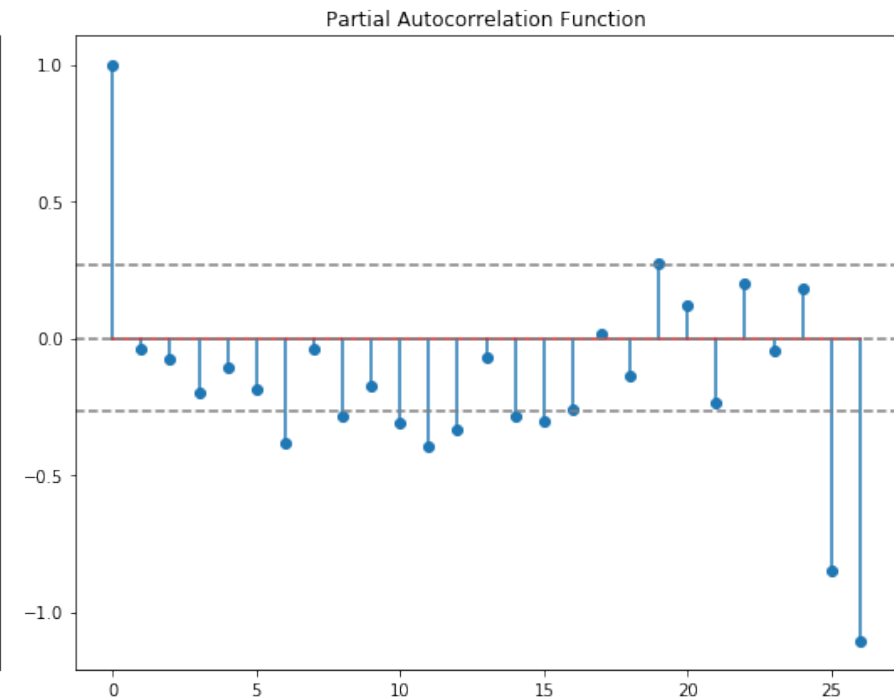
Un modelo de media móvil usa los errores de predicciones pasadas para ajustar un “modelo de tipo regresión”



¿Qué modelos podemos plantear?



$q = 6, 3 (\pm 1)$



$p = 6 (\pm 1)$

$P = 2 (\pm 1)$

¿Qué modelos podemos plantear?

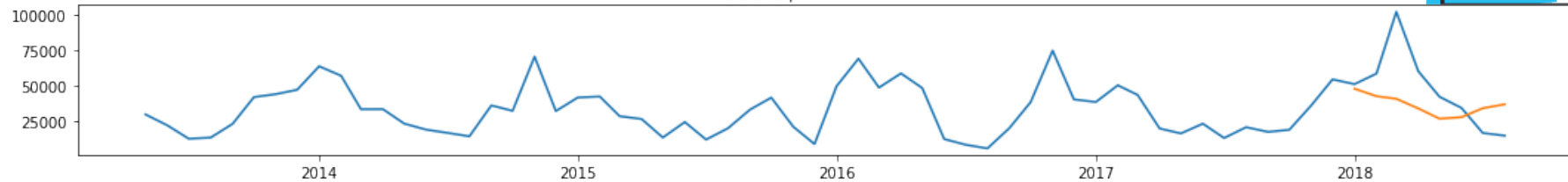


```
model = ARIMA(order=(p, d, q)).fit(y=ts)
model
forecast = model.predict(n_periods=k)
print("Test RMSE: %.3f" % np.sqrt(mean_squared_error(data.loc['2018-01-01':'2018-08-01', 'rain_all'], forecast)))
pd.concat([pd.Series(data['rain_all']), pd.Series(forecast_ar, index=[pd.Timestamp('2018-01-01'),
    pd.Timestamp('2018-02-01'), pd.Timestamp('2018-03-01'),
    pd.Timestamp('2018-04-01'), pd.Timestamp('2018-05-01'),
    pd.Timestamp('2018-06-01'), pd.Timestamp('2018-07-01'), pd.Timestamp('2018-08-01')])], axis=1).plot()
```

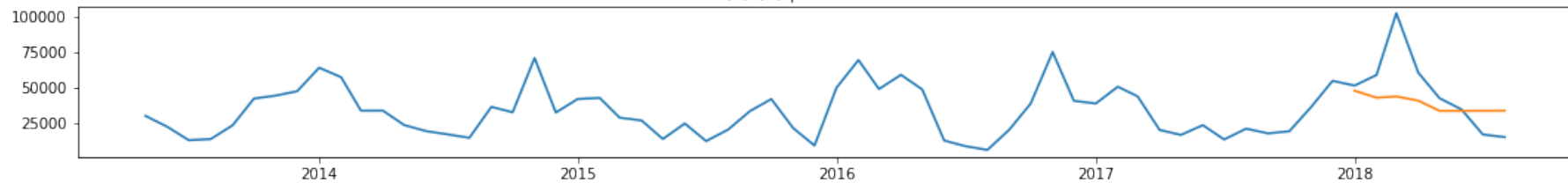
¿Qué modelos podemos plantear?

PYCONES
2018
MÁLAGA

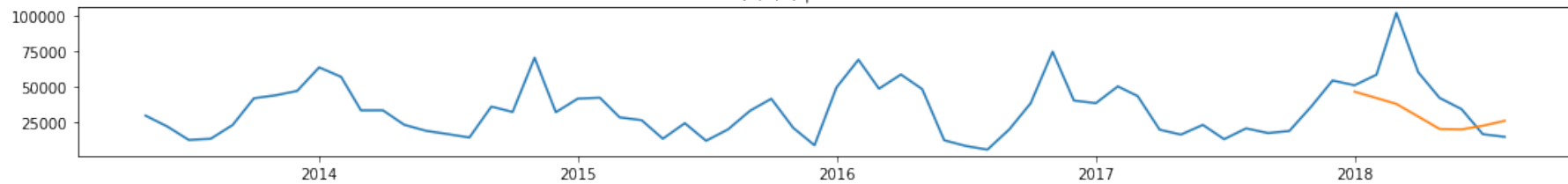
AR (6,1,0) | RMSE: 26911.237



MA (0,1,5) | RMSE: 24509.626



ARIMA (5,1,1) | RMSE: 27878.986



SARIMAX (5,1,1) (2,1,0,12) | RMSE: 21892.210







Thank You