PI DATA STRATEGY & CONSULTING

SERIES TEMPORALES: La clave para descifrar el futuro

Series Temporales:
La clave para descifrar el futuro

Johanna Frau
Data Scientist
jfrau@piconsulting.com.ar

Santiago Chiesa
Data Scientist
schiesa@piconsulting.com.ar

Agenda

- MOTIVACIÓN
- CONCEPTOS IMPORTANTES
- ARIMA
- SARIMA
- HOLT-WINTERS
- XGBOOST

MOTIVACIÓN

Finanzas

Detección de fraude

Planificación de la producción e inventario

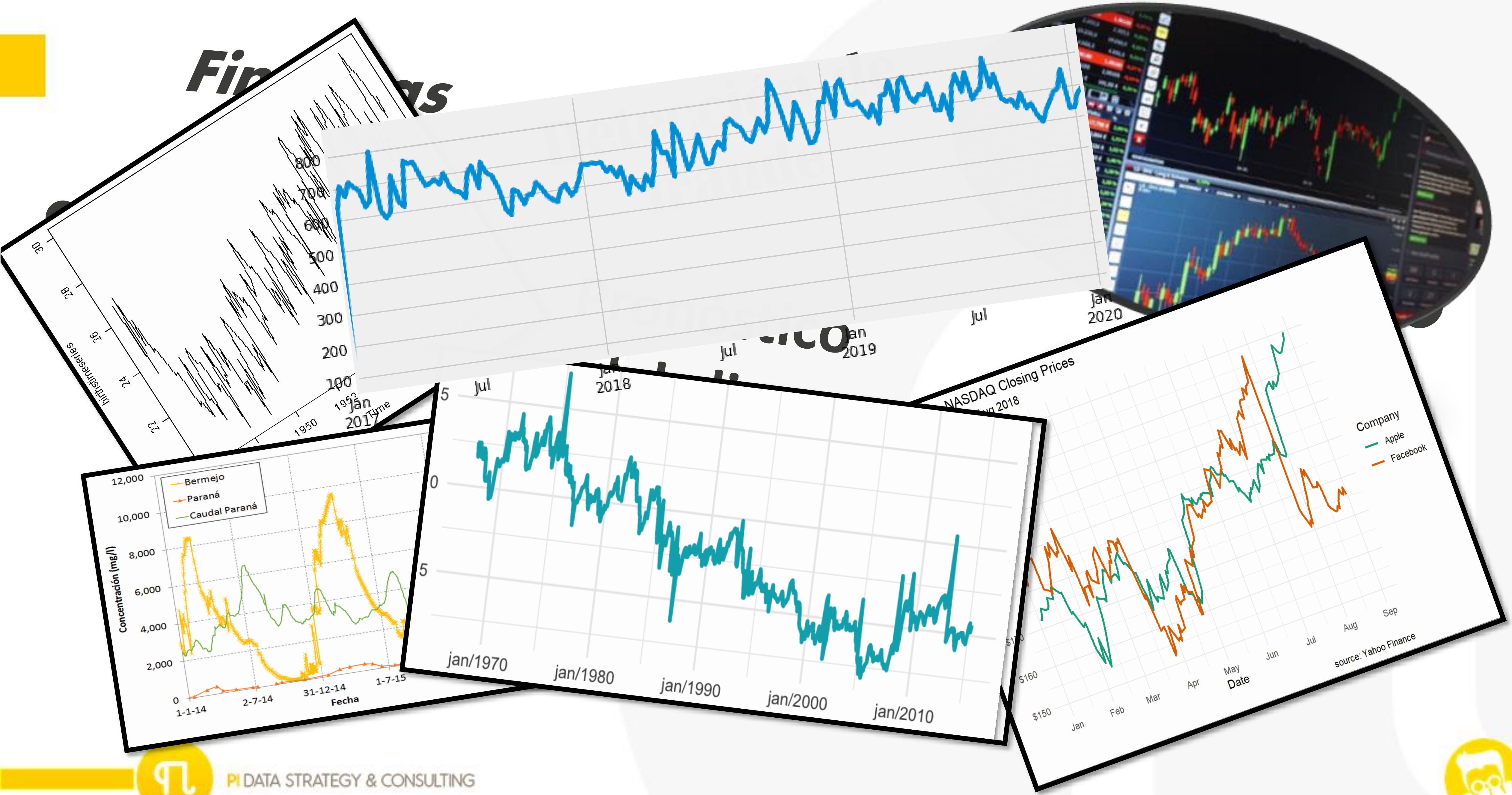
Gestión de cadena de suministro

Pronóstico del clima

Análisis sociales

Control de calidad

Monitoreo de la frecuencia cardíaca



NASDAQ Closing Prices
Company
Apple
Facebook
Date
source: Yahoo Finance
Bermejo
Paraná
Caudal Paraná
Concentración (mg/l)
Fecha
birthstimeseries
Time
jan/1970
jan/1980
jan/1990
jan/2000
jan/2010

***Series temporales** es una de las técnicas dentro de Data Science con mayor aplicación en problemas de la vida real y de gran sustento estadístico.*

CONCEPTOS IMPORTANTES

¿Qué es una serie temporal?

Es una colección de secuencias de ***valores igualmente espaciados*** *en el tiempo. Por ejemplo: días, semanas, quincenas, meses, años.*

¿Para qué estudiar series temporales?

Para explicar las variaciones observadas en
la serie del
pasado, identificando patrones detrás de
los datos que fuerzan a
la serie a tener cierta tendencia o comportamiento.

Predecir valores futuros de las variables utilizando el patrón detectado y tomar decisiones en base a los valores predichos.

Tipos de predicciones

PREDICCIONES DE SERIES TEMPORALES UNIVARIADAS

Solo se usan los valores previos de la serie para predecir valores futuros.

PREDICCIONES DE SERIES TEMPORALES MULTIVARIADAS

A los valores previos se le agregan otras variables para realizar las predicciones.

Descomposición de una serie temporal

Las series de tiempo pueden exhibir diferentes patrones.

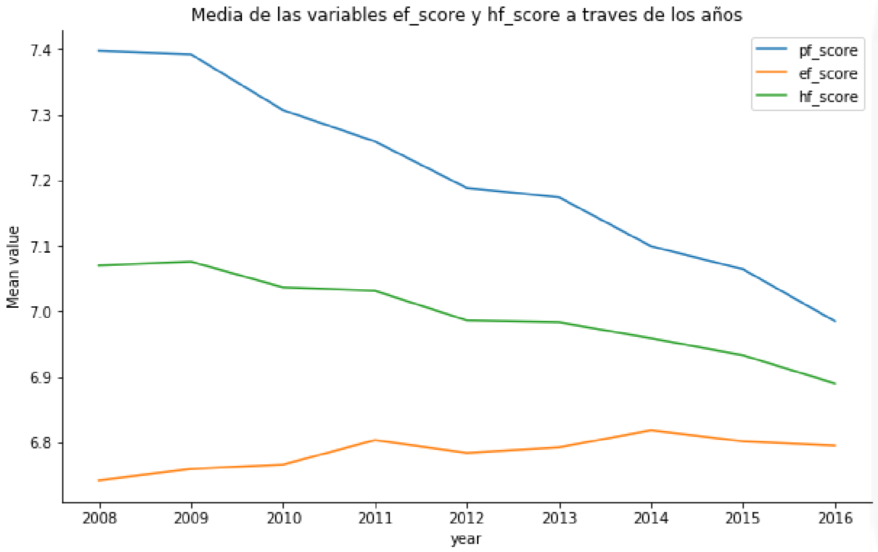
TENDENCIA

CICLOS

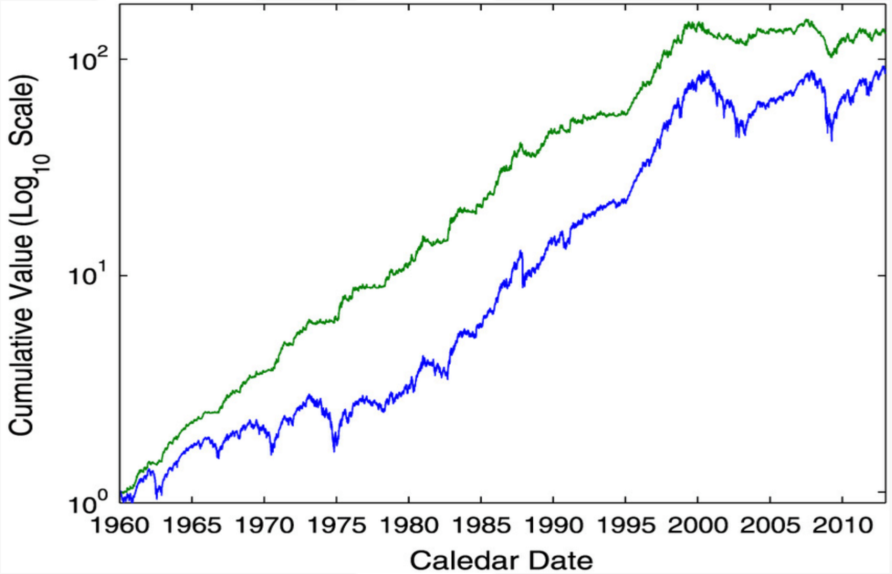
ESTACIONALIDAD

Descomposición de una serie temporal

- ***Tendencia (T):*** Refleja la proyección de la serie a largo plazo. Esta existe cuando hay un incremento o disminución que se mantiene en el tiempo. No es necesariamente lineal.



https://github.com/jfrau/DiploDatos2019/blob/master/Practico_Analisis_y_Visualizacion.ipynb

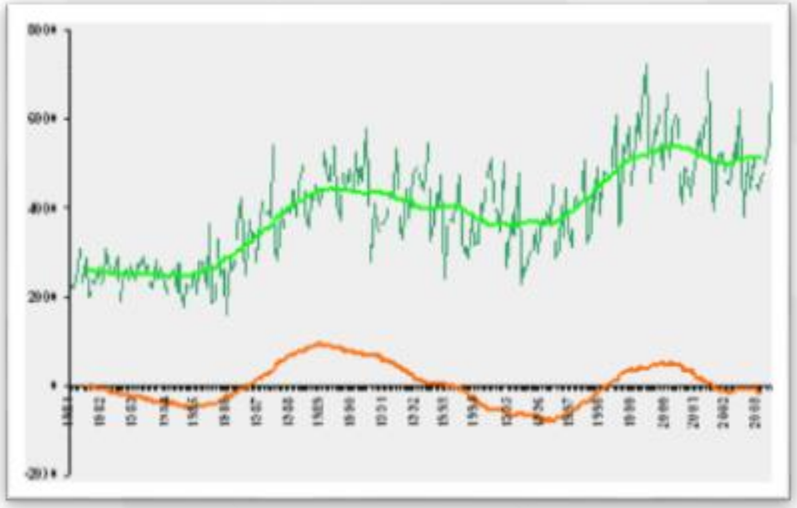


https://www.researchgate.net/figure/Time-Series-Plot-of-Cumulative-Values-of-the-Buy-and-Hold-and-the-Combination-Moving_fig4_309138994%E2%80%8B

Descomposición de una serie temporal

- ***Ciclos (C):*** Refleja aumentos y caídas en los valores de los datos que se repiten pero no en períodos de tiempo específicos. La duración de estos períodos es de al menos 1 año.

En el ámbito económico se observan muchos ejemplos relacionados con la inflación, el PBI, el riesgo país, etc.

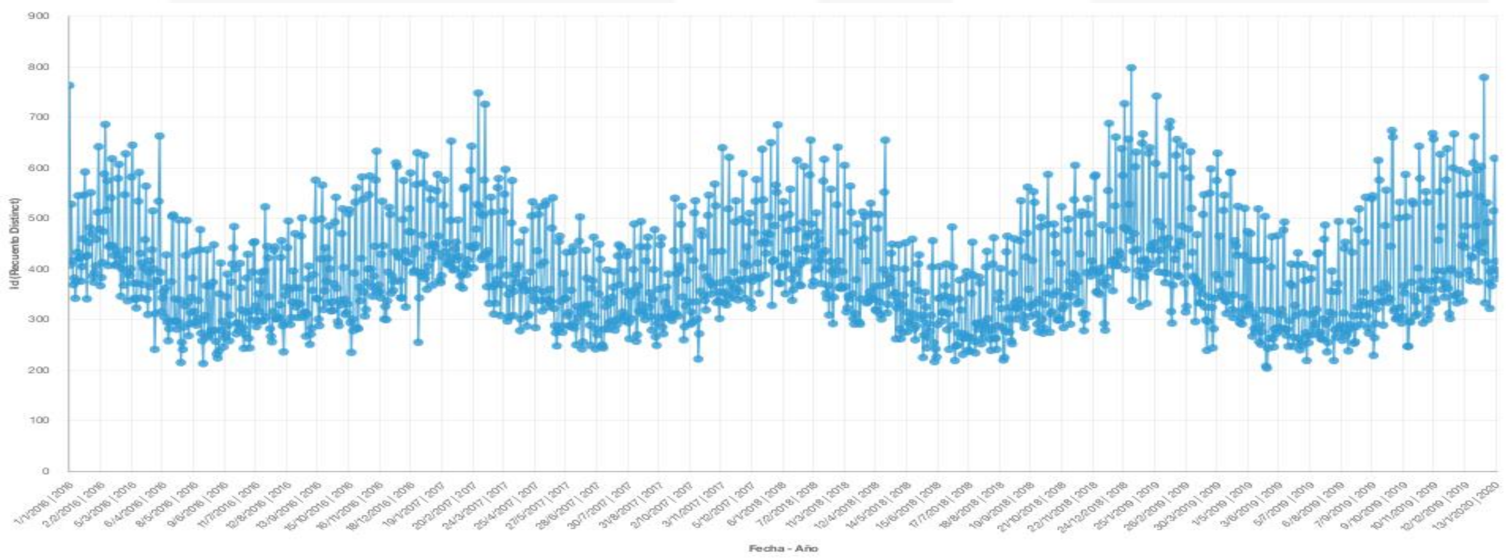


http://www5.uva.es/estadmed/datos/series/series1.htm

Descomposición de una serie temporal

- ***Estacionalidad (S):*** Refleja la periodicidad presente en la serie como variaciones en franjas de tiempo específicas. Los períodos son menores a un año y pueden ser diarios, semanales, quincenales, mensuales, etc.

Por ejemplo, la compra de cervezas (o helado) es mayor en verano.



Descomposición de una serie temporal

- ***Aleatoriedad (R):*** Refleja movimientos de la serie que no son explicados por las demás componentes, generalmente asociados a sucesos impredecibles y no períodicos (aleatorios).

Por ejemplo: clima inusual, huelgas, crisis, etc.

Descomposición de una serie temporal

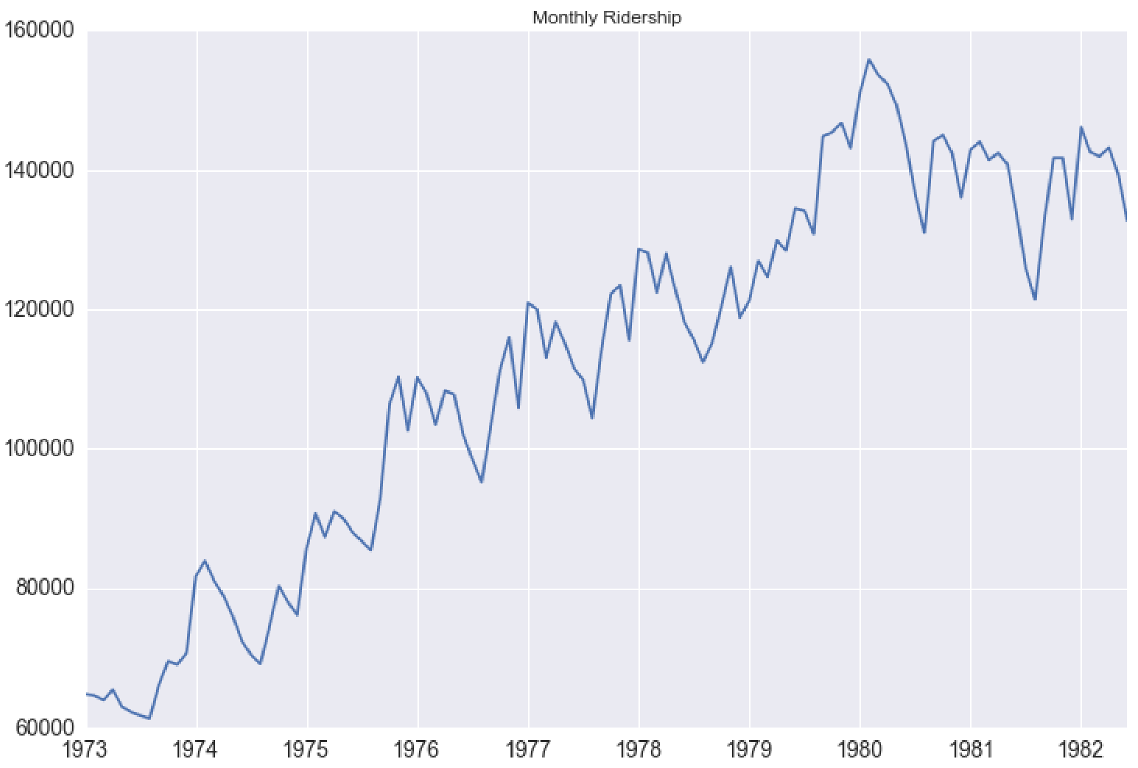
Las series de tiempo pueden exhibir diferentes patrones.

TENDENCIA – CICLO (T)

ESTACIONALIDAD (S)

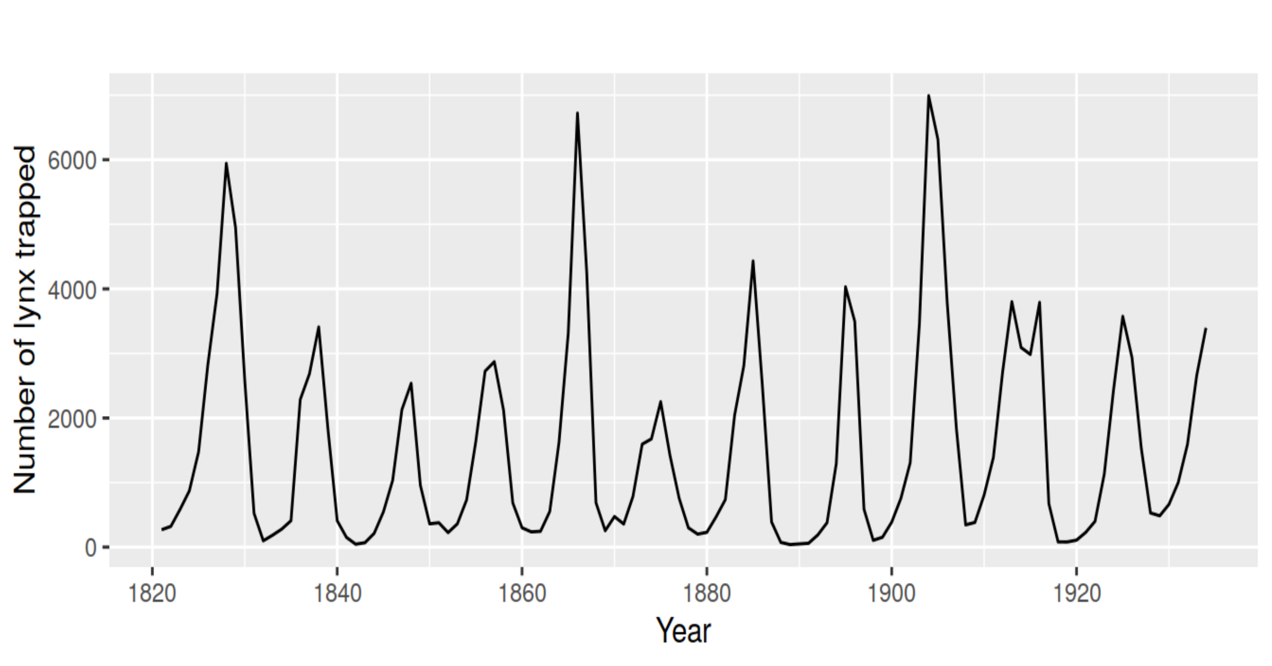
ALEATORIEDAD (R)

Más ejemplos



Tendencia creciente con períodos estacionales

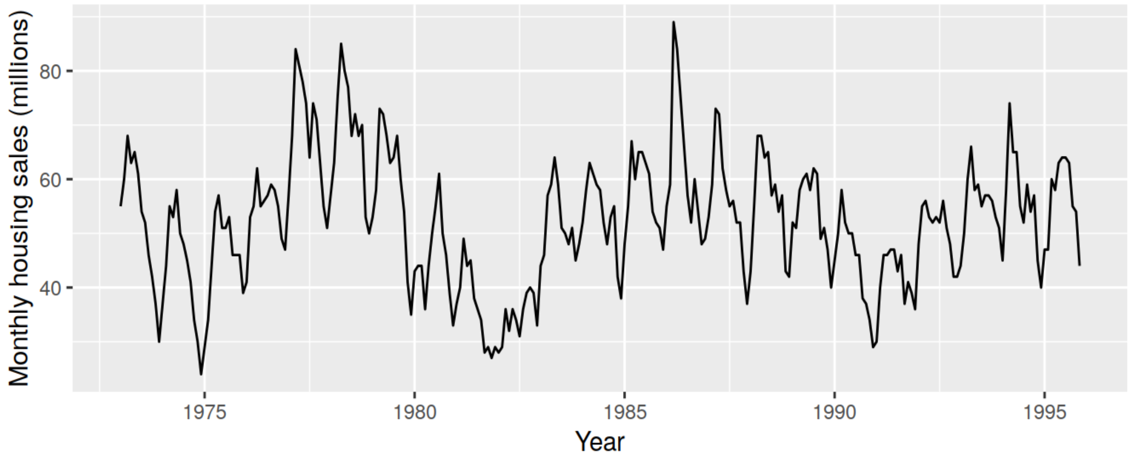
http://www.seanabu.com/2016/03/22/time-series-seasonal-ARIMA-model-in-python/



Presencia de ciclos de aproximadamente 10 años

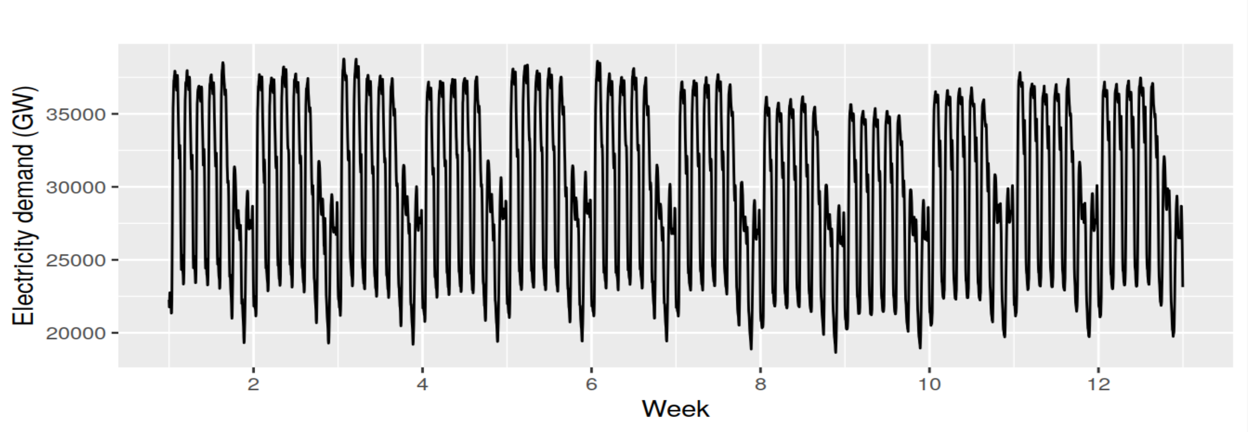
ttps://robjhyndman.com/hyndsight/cyclicts/

Más ejemplos



Estacionalidad anual y comportamiento cíclico en períodos de 6-10 años

https://robjhyndman.com/hyndsight/cyclicts/



Estacionalidad diaria y semanal

https://robjhyndman.com/hyndsight/cyclicts/

Descomposición de una serie temporal

ADITIVA

$ST = T + S + R$

La serie tiene una tendencia más lineal y la variación de la componente estacional es constante a través del tiempo.

Ejemplo: todos los meses el consumo de energía aumenta 1000 unidades.

Descomposición de una serie temporal

MULTIPLICATIVA

$ST = T * S * R$

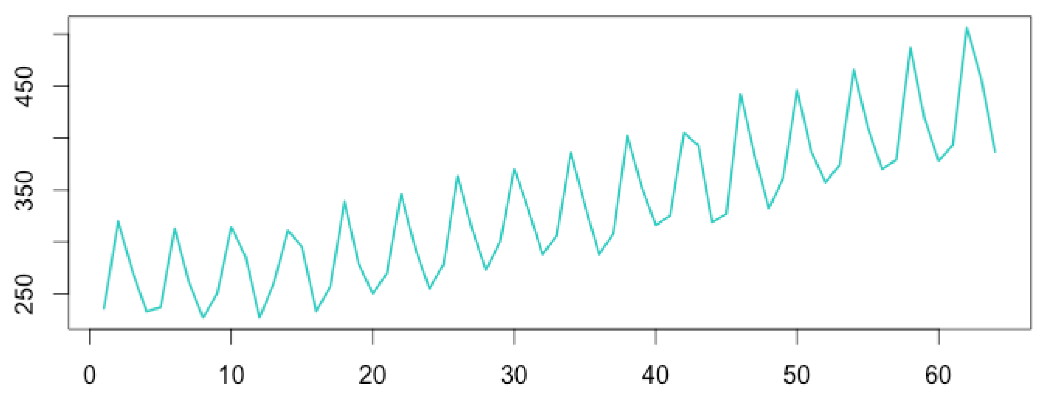
La serie incrementa o disminuye a una tasa no lineal.
La variación del patrón estacional, o la variación de la tendencia-ciclo, es proporcional al nivel de la serie de tiempo.

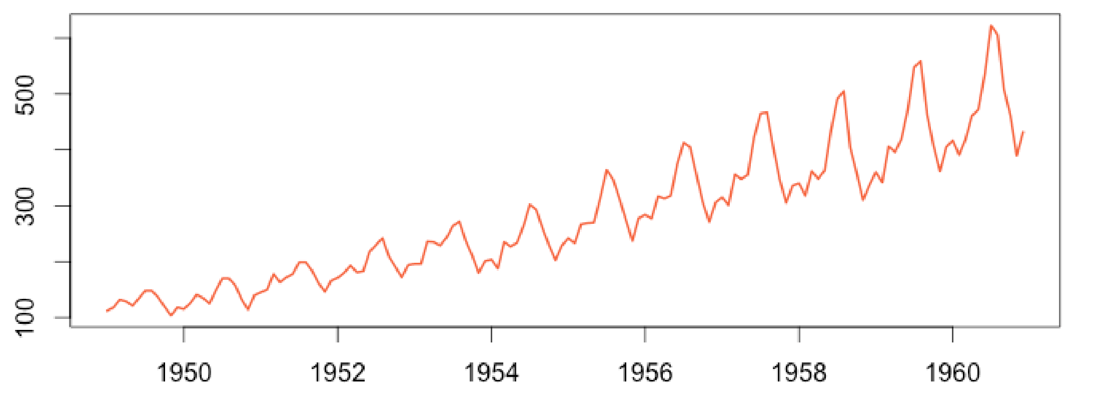
Ejemplo: cada año se duplica la cantidad de producción de energía.

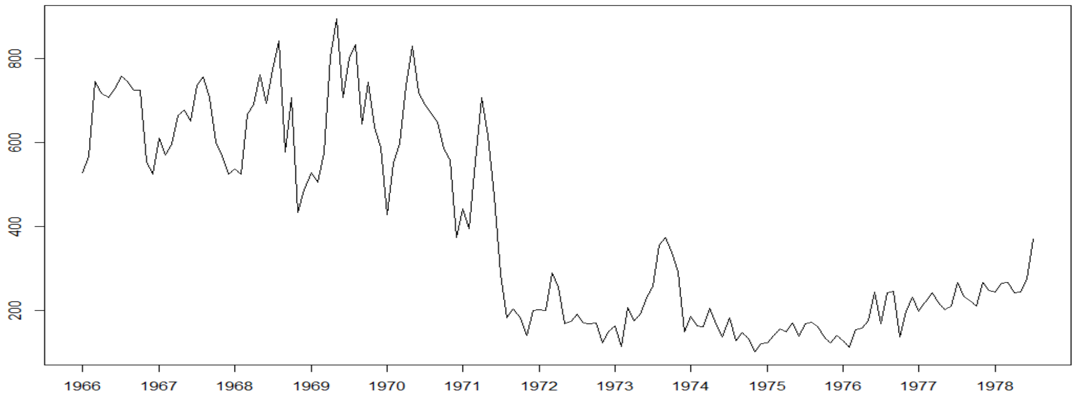
Descomposición de una serie temporal

ADITIVA

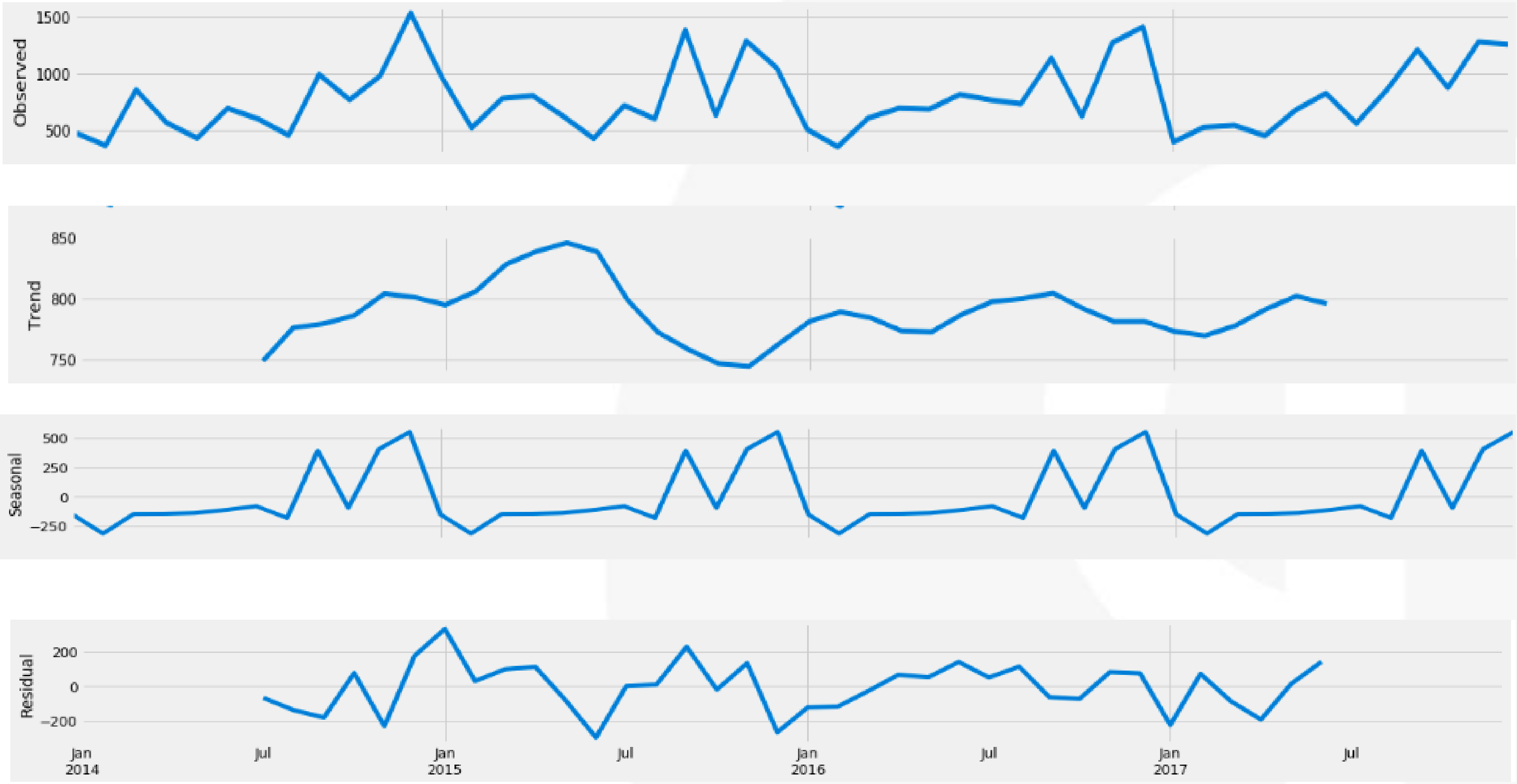
MULTIPLICATIVA





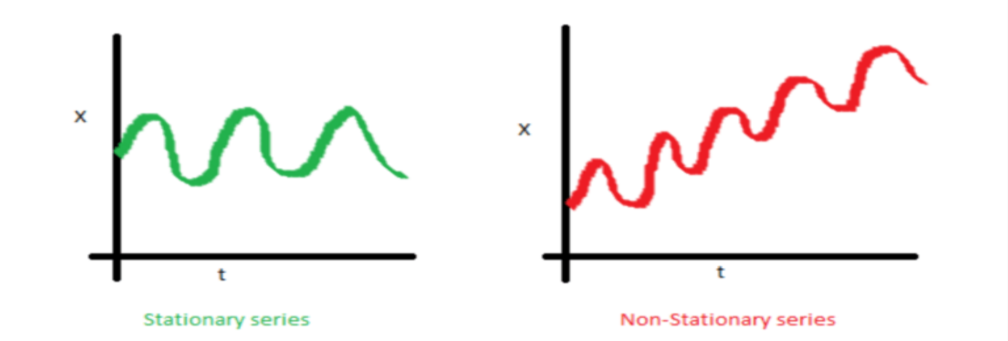


Descomposición de una serie temporal



Estacionariedad

Las series estacionarias tienen las siguientes propiedades que se mantienen a lo largo del tiempo:



- Media constante
- Varianza constante.
- No suelen tener una tendencia marcada.

Las propiedades de la serie no dependen del momento en que se capturan.

La importancia de la estacionariedad

- Si es posible entender el patrón presente detrás de los datos, entonces podemos decir con una alta probabilidad que la serie de tiempo seguirá ese mismo patrón en el futuro.

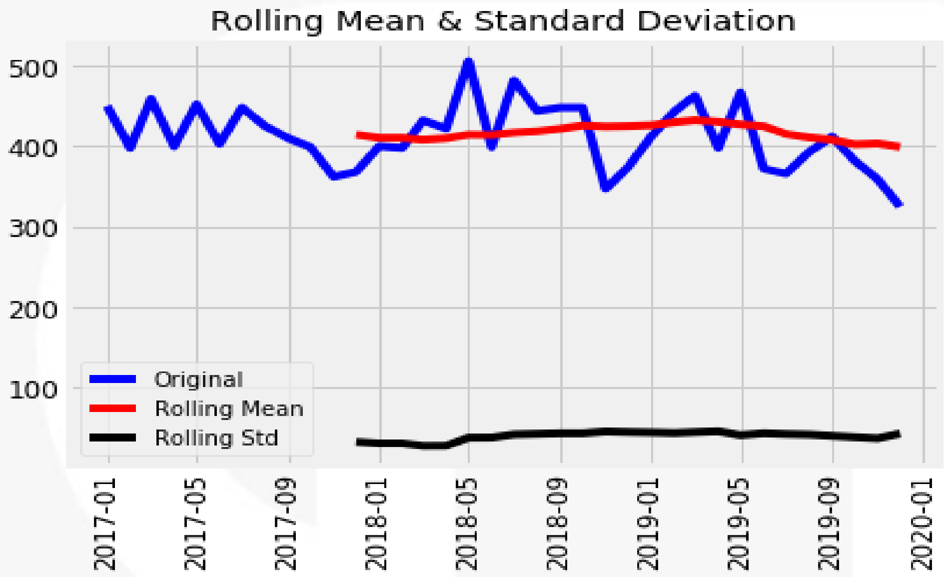
- Usando datos estacionarios los modelos pueden realizar predicciones basandose en el hecho de que la media y la varianza será la misma en los períodos futuros.

Estacionariedad: Método Visual

Promedio y desviación estándar móvil.

SERIE NO ESTACIONARIA

http://www.seanabu.com/2016/03/22/time-series-seasonal-ARIMA-model-in-python/



SERIE ESTACIONARIA

Estacionariedad: Test estadístico

Augmented Dickey–Fuller (ADF) test

- Es uno de los métodos estadísticos más populares para chequear si la serie es estacionaria o no. Se conoce también como test de la raíz unitaria (Unit Root Test).

- El test determina la intensidad con la cual una serie temporal se define por una tendencia.

Estacionariedad: Test estadístico

Augmented Dickey–Fuller (ADF) test

- **Hipótesis nula (H0):**

La serie de tiempo NO ES ESTACIONARIA (esto significa que tiene cierta estructura dependiente del tiempo, o en otras palabras, tiene una raíz unitaria).

- **Hipótesis alternativa (HA):**

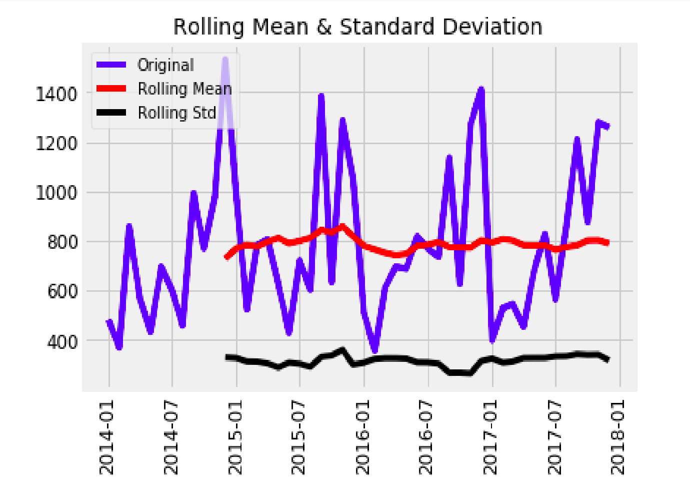
La serie de tiempo ES ESTACIONARIA.

Estacionariedad: Test estadístico

Augmented Dickey–Fuller (ADF) test

El ADF test devuelve como resultado:

- Estadístico de prueba.
- p-value.
- Valores críticos para diferentes niveles de significancia (1%, 5% y el 10%).




```
Results of Dickey-Fuller Test:
Test Statistic                  5.191070
p-value                         0.000009
#Lags Used                     10.000000
Number of Observations Used    37.000000
Critical Value (1%)             3.620918
Critical Value (5%)             2.943539
Critical Value (10%)            2.610400
dtype: float64
```

Estacionariedad: Test estadístico

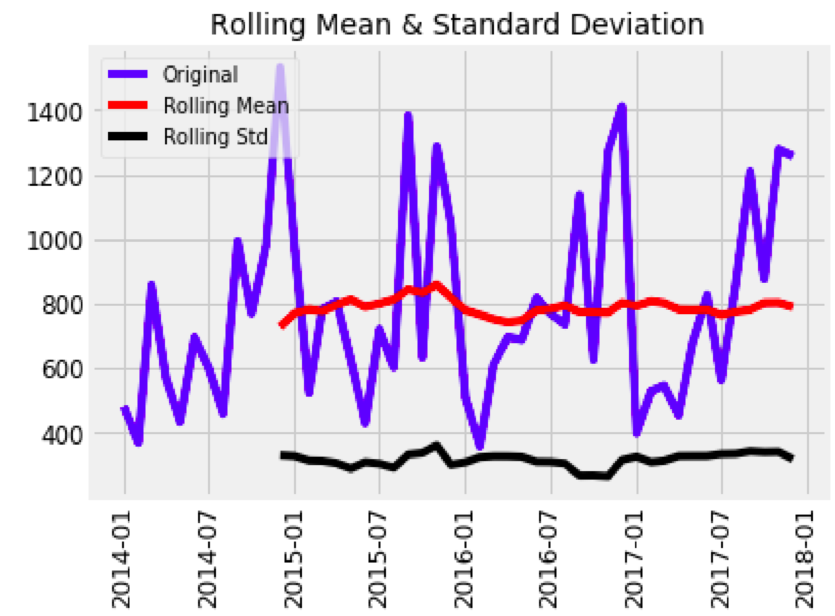
Augmented Dickey–Fuller (ADF) test

Criterios

1. Si p<=0.05, se rechaza la hipótesis nula y la serie es estacionaria con un 5% de significancia.

2. Si el valor del estadístico de prueba es más chico que los valores críticos entonces se rechaza la hipótesis nula y la serie es estacionaria con algún nivel de significancia.

Augmented Dickey–Fuller (ADF) test




```
Results of Dickey-Fuller Test:
Test Statistic                 -5.191070
p-value                         0.000009
#Lags Used                     10.000000
Number of Observations Used    37.000000
Critical Value (1%)            -3.620918
Critical Value (5%)            -2.943539
Critical Value (10%)           -2.610400
dtype: float64
```

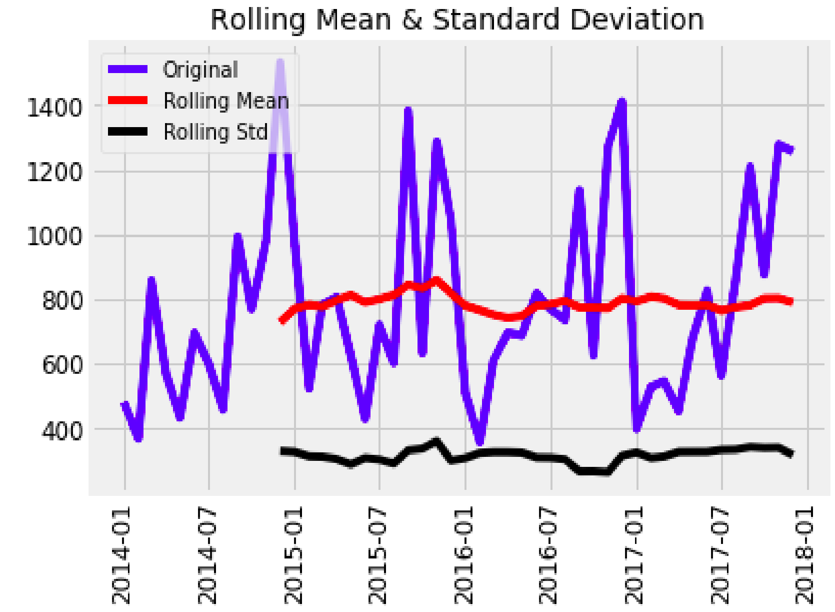
<u>Ejemplo 1</u>

Como

p-value=0.000009<0.05

la serie es **ESTACIONARIA** con 5% de significancia.

Augmented Dickey–Fuller (ADF) test




```
Results of Dickey-Fuller Test:
Test Statistic                 -5.191070
p-value                         0.000009
#Lags Used                     10.000000
Number of Observations Used    37.000000
Critical Value (1%)            -3.620918
Critical Value (5%)            -2.943539
Critical Value (10%)           -2.610400
dtype: float64
```

Ejemplo 1

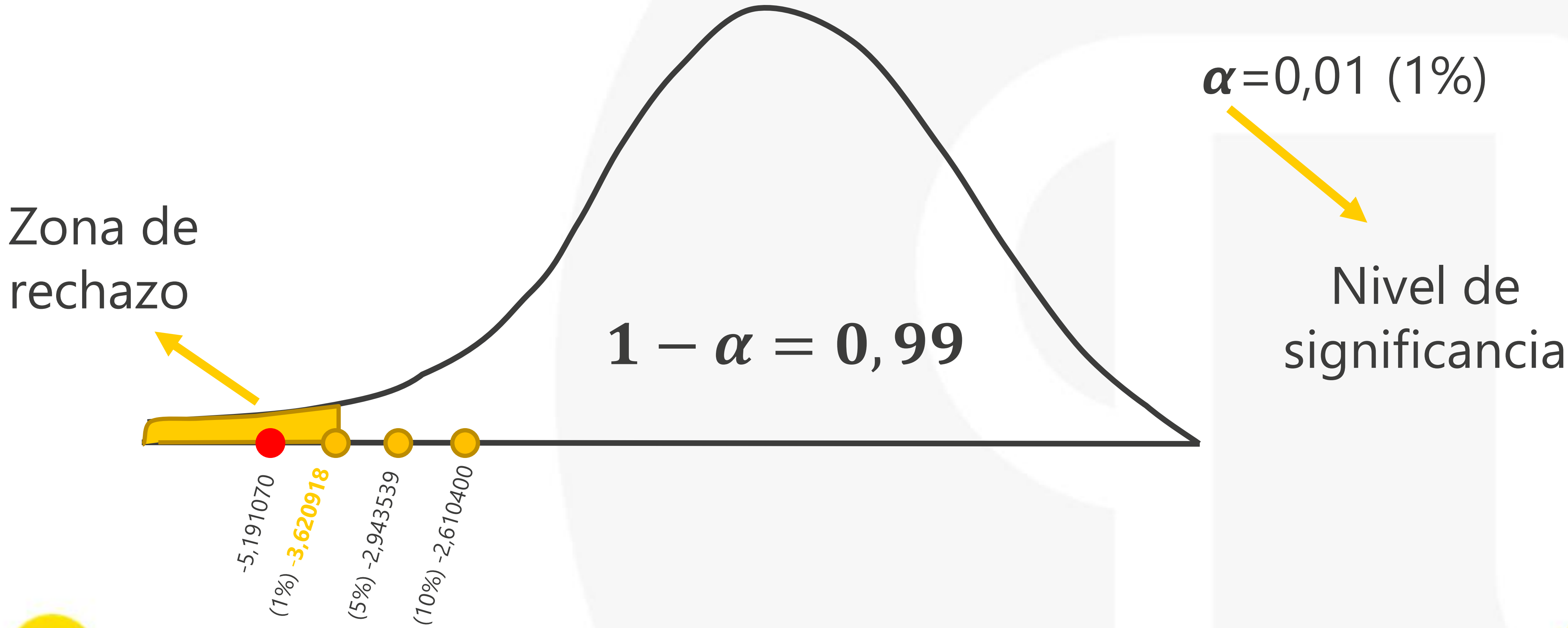
Por otro lado,

Estadístico=-5.191070<-3.620918 (1%)
Estadístico=-5.191070<-2.943539 (5%)
Estadístico=-3.366749<-2.610400 (10%)

la serie es **ESTACIONARIA** con 1% de significancia.

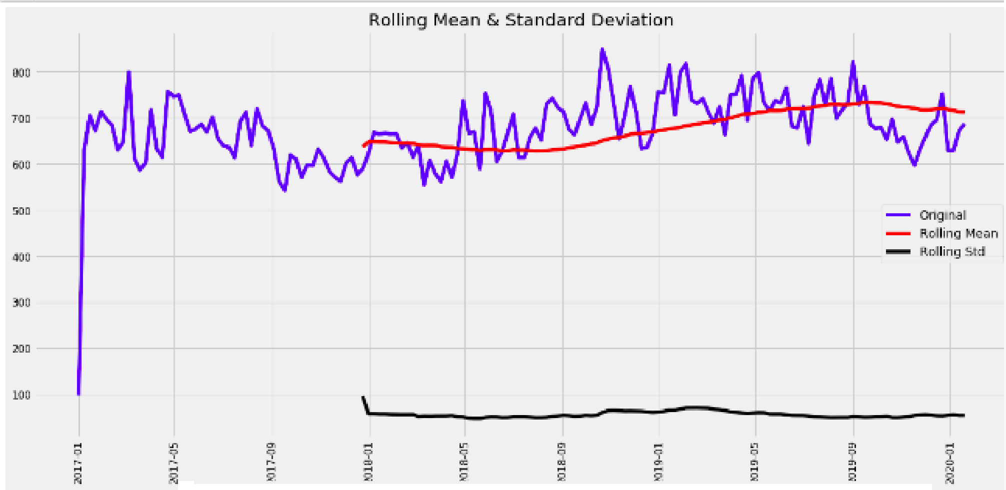
Augmented Dickey–Fuller (ADF) test

Ejemplo 1



Augmented Dickey–Fuller (ADF) test

Ejemplo 2




```
Results of Dickey-Fuller Test:
Test Statistic                  -3.366749
p-value                          0.012146
#Lags Used                       3.000000
Number of Observations Used    156.000000
Critical Value (1%)             -3.472979
Critical Value (5%)             -2.880252
Critical Value (10%)            -2.576747
dtype: float64
```

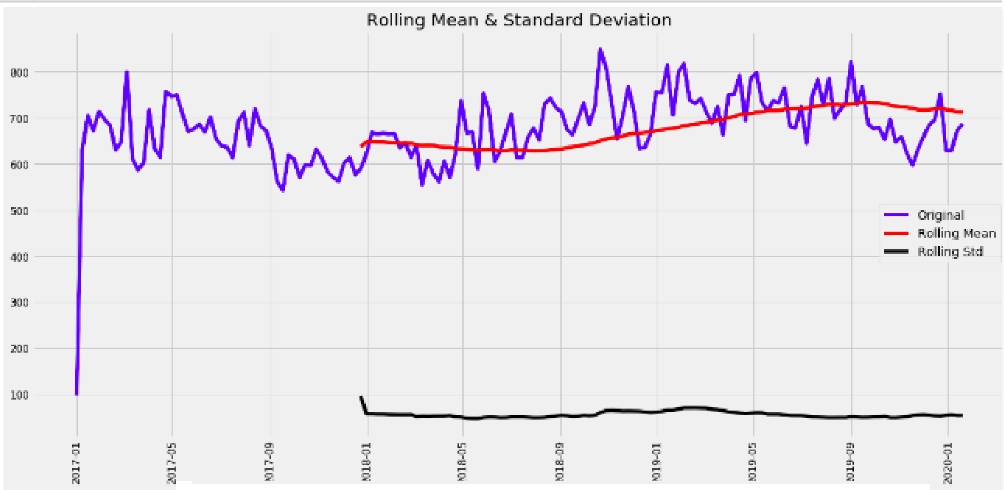
Como

p-value=0.012146<0.05

la serie es **ESTACIONARIA** con 5% de significancia.

Augmented Dickey–Fuller (ADF) test

Ejemplo 2




```
Results of Dickey-Fuller Test:
Test Statistic                  -3.366749
p-value                          0.012146
#Lags Used                       3.000000
Number of Observations Used    156.000000
Critical Value (1%)             -3.472979
Critical Value (5%)             -2.880252
Critical Value (10%)            -2.576747
dtype: float64
```

Por otro lado,

Estadístico=-3.366749<-2.880252 (5%)
Estadístico=-3.366749<-2.576747 (10%)

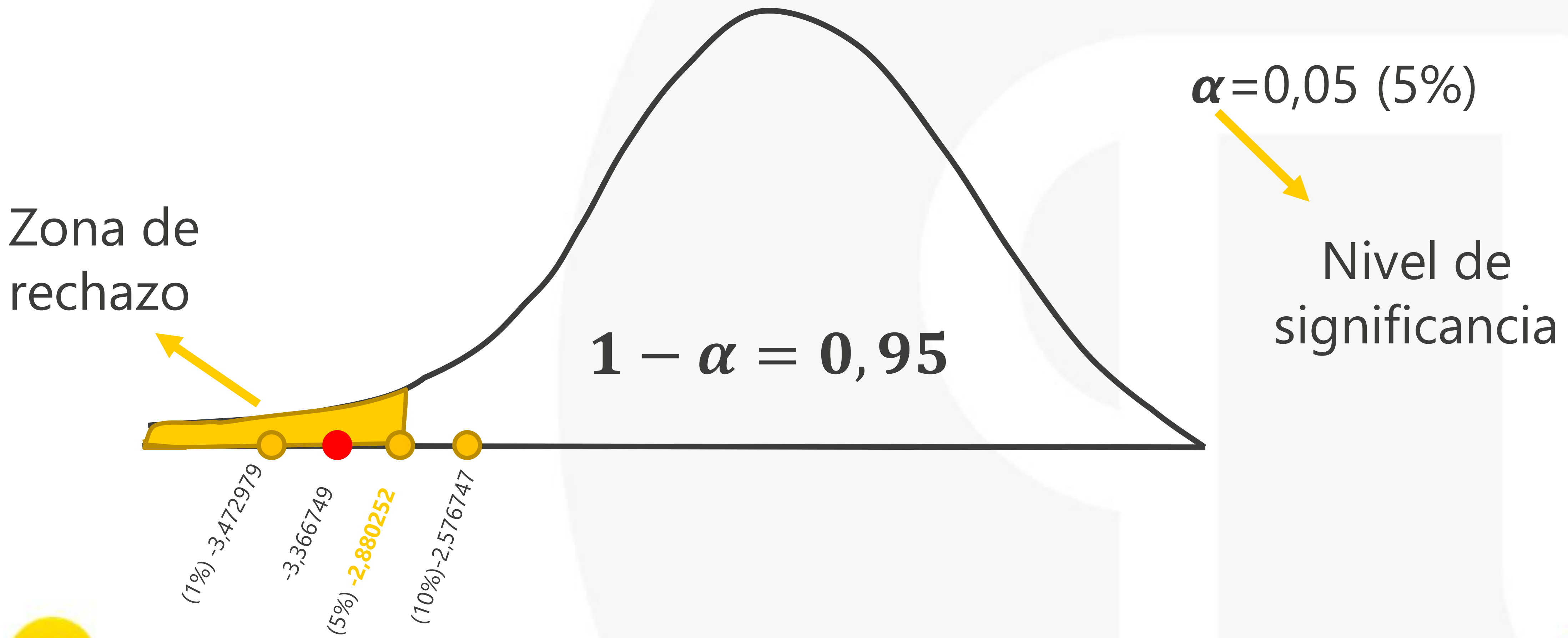
la serie es **ESTACIONARIA** con 5% de significancia.

Notar que

Estadístico=-3.366749>-3.472979 (1%)

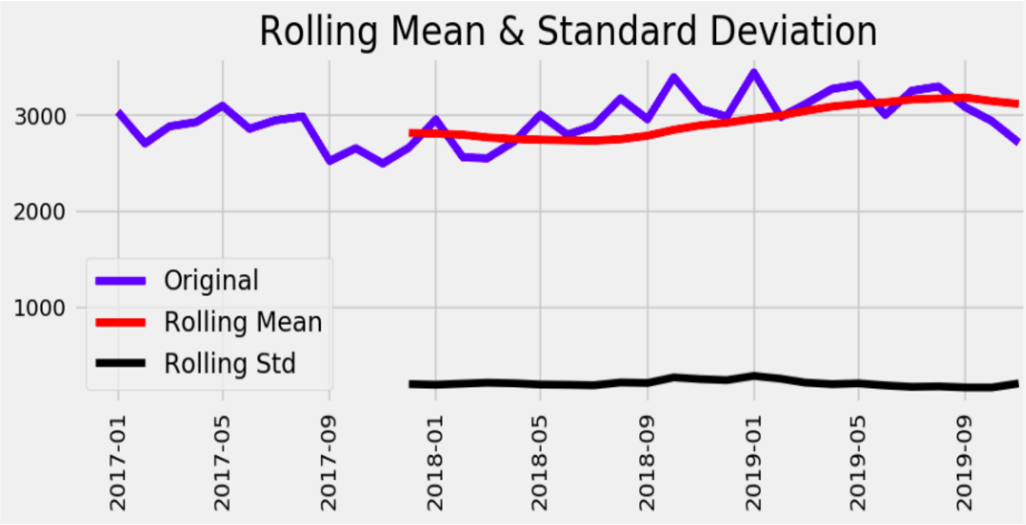
Augmented Dickey–Fuller (ADF) test

<u>*Ejemplo 2*</u>



Augmented Dickey–Fuller (ADF) test

Ejemplo 3




```
Results of Dickey-Fuller Test:
Test Statistic                 -2.269828
p-value                         0.181884
#Lags Used                      9.000000
Number of Observations Used    25.000000
Critical Value (1%)            -3.723863
Critical Value (5%)            -2.986489
Critical Value (10%)           -2.632800
dtype: float64
```

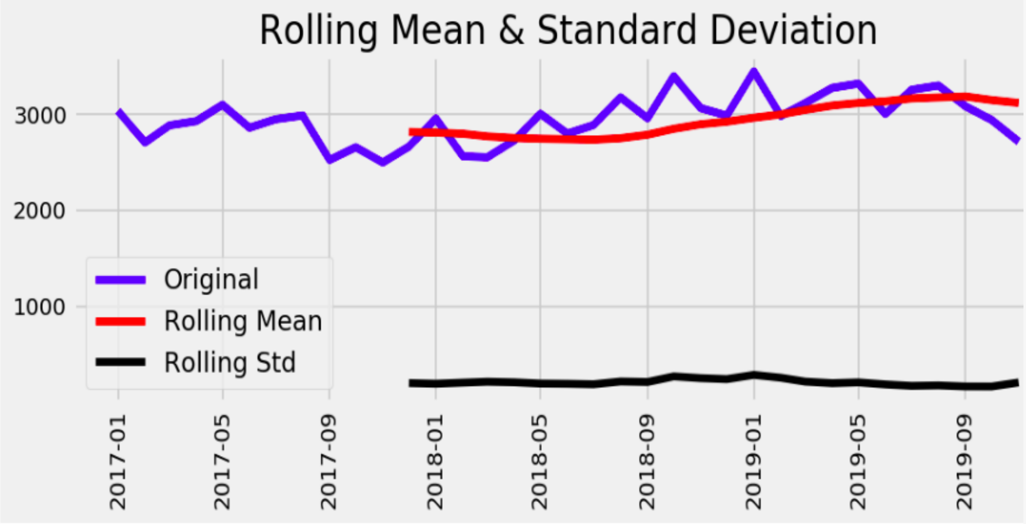
Como

p-value=0.181884>0.05

no se rechaza la hipótesis nula y la serie es **NO ESTACIONARIA**.

Augmented Dickey–Fuller (ADF) test

Ejemplo 3




```
Results of Dickey-Fuller Test:
Test Statistic                 -2.269828
p-value                         0.181884
#Lags Used                      9.000000
Number of Observations Used    25.000000
Critical Value (1%)            -3.723863
Critical Value (5%)            -2.986489
Critical Value (10%)           -2.632800
dtype: float64
```

Observemos además que

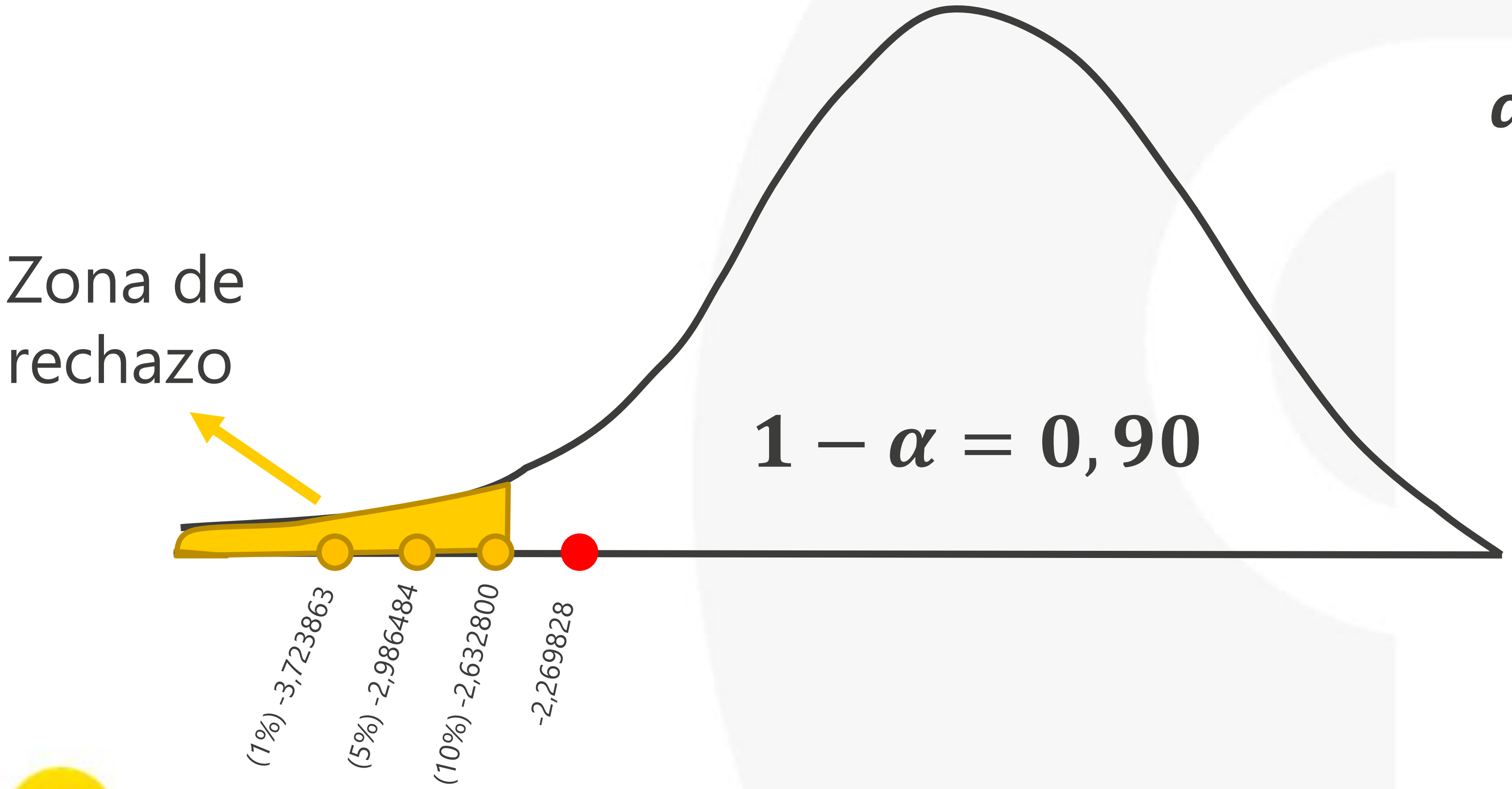
Estadístico=-2.269828>-3.723863 (1%)

Estadístico=-2.269828>-2.986489 (5%)

Estadístico=-2.269828>-2.632800 (10%)

Augmented Dickey–Fuller (ADF) test

Ejemplo 3



α=0,10 (10%)

Nivel de significancia

ARIMA

Algoritmo ARIMA

ARIMA (AutoRegressive Integrated Moving Average) es una clase de modelos que explican una serie de tiempo teniendo en cuenta sus propios valores pasados, es decir, sus propios retrasos (lags) y errores de pronóstico rezagados (lagged).

Algoritmo ARIMA

Un modelo ARIMA es caracterizado por tres términos:

- p: orden del término AR.
- d: orden del término I.
- q: orden del término MA.

ARIMA(p,d,q)

ARIMA(p,d,q)

p : Es la parte autoregresiva del modelo (AR) y son los valores pasados usados en la ecuación de regresión para la serie. Ayuda a ajustar la línea que está intentando pronosticar la serie.

"Es probable que haga calor mañana dado que ha hecho calor los últimos 3 días."

En este caso, p=3 y por lo tanto se usarán tres períodos previos de la serie en la porción autoregresiva.

ARIMA(p,d,q)

d: Es la parte integrada del modelo (I) y representa el mínimo número de veces que los datos tienen que ser diferenciados para producir datos estaciona rios.
Si la serie ya es estacionaria d=0.

"Es muy probable que haga la misma temperatura mañana si la diferencia de temperatura en los últimos tres días ha sido muy pequeña."

En este caso, la tercera diferencia es la diferencia entre el período de tiempo actual y los últimos tres días (d=3).

ARIMA(p,d,q)

q: Es la parte de media móvil (MA) del modelo.

Representa el error del modelo como una combinación lineal de términos de errores anteriores y es el número de valores anteriores o retrasados para el término de error que se agregan o restan a la serie En este caso, el error es una parte de la serie temporal no explicada por la tendencia o la estacionalidad.

ACF y PACF plots

Determinar parámetros **NUNCA** es una tarea fácil.

Autocorrelation function (ACF)
Parcial Autocorrelation function (PACF)

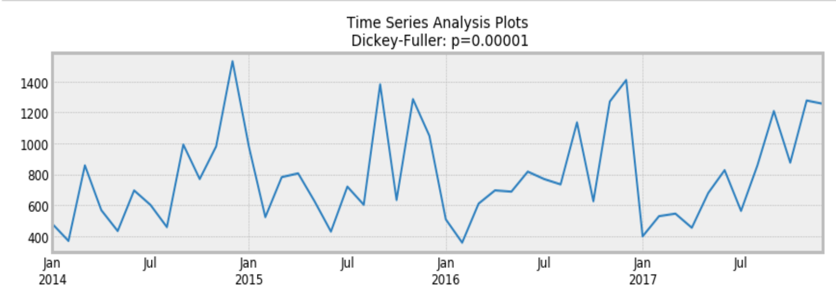
ACF y PACF plots

Autocorrelation function (ACF)

Mide la correlación entre la serie y sus valores lagged (retrasados).
Es útil para determinar el valor de q.

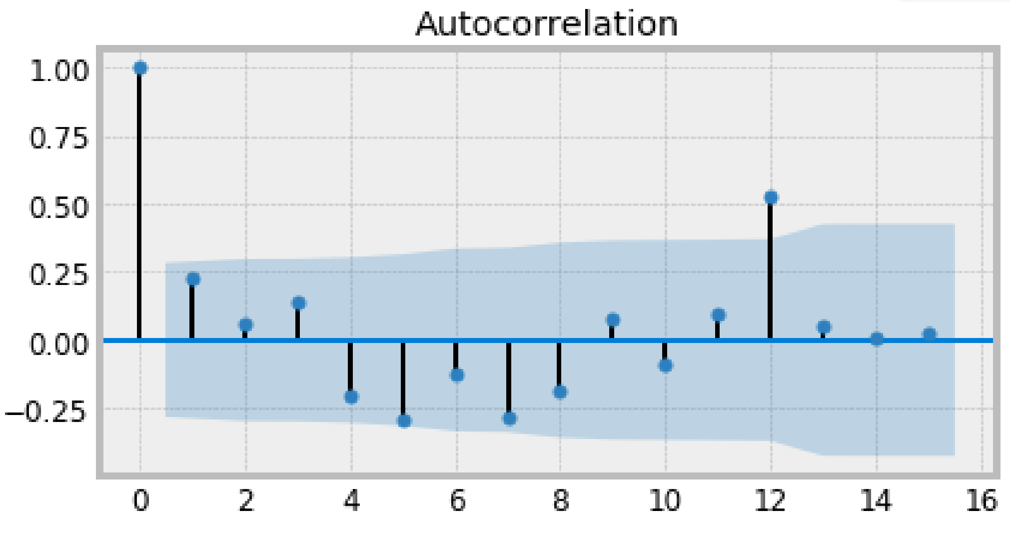
Por ejemplo, se podría medir la correlación entre y(t) con y(t-1).

ACF y PACF plots

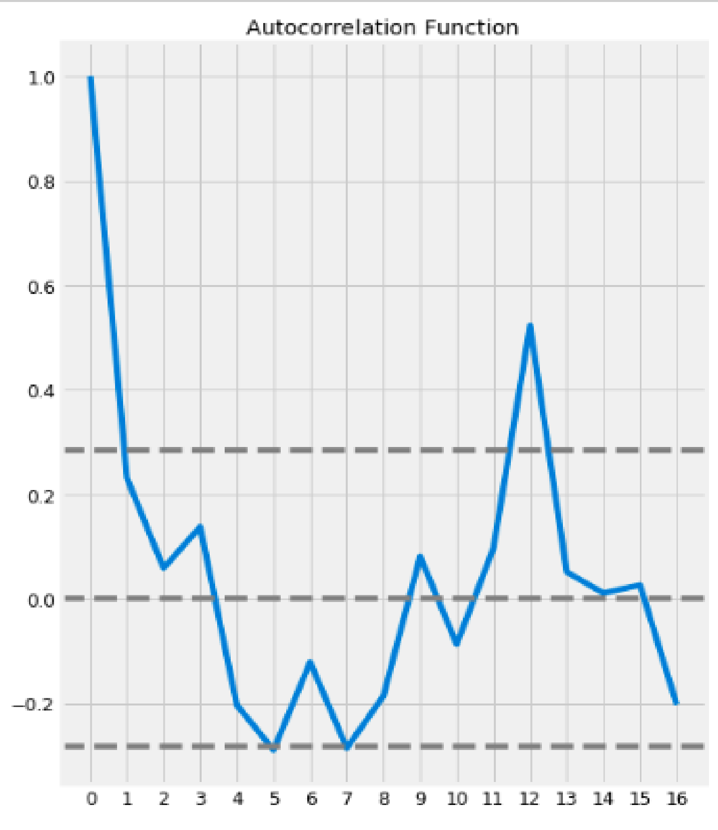


Autocorrelation function (ACF)

q es el valor de retraso donde el gráfico ACF cruza el intervalo de confianza superior por primera vez.



q=1



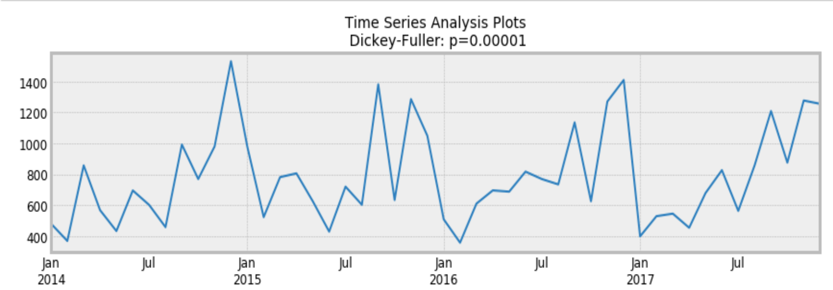
ACF y PACF plots

Parcial Autocorrelation function (PACF)

Encuentra la correlación de los residuos con el siguiente valor de retraso.
Ayuda a determinar el valor del parámetro p.

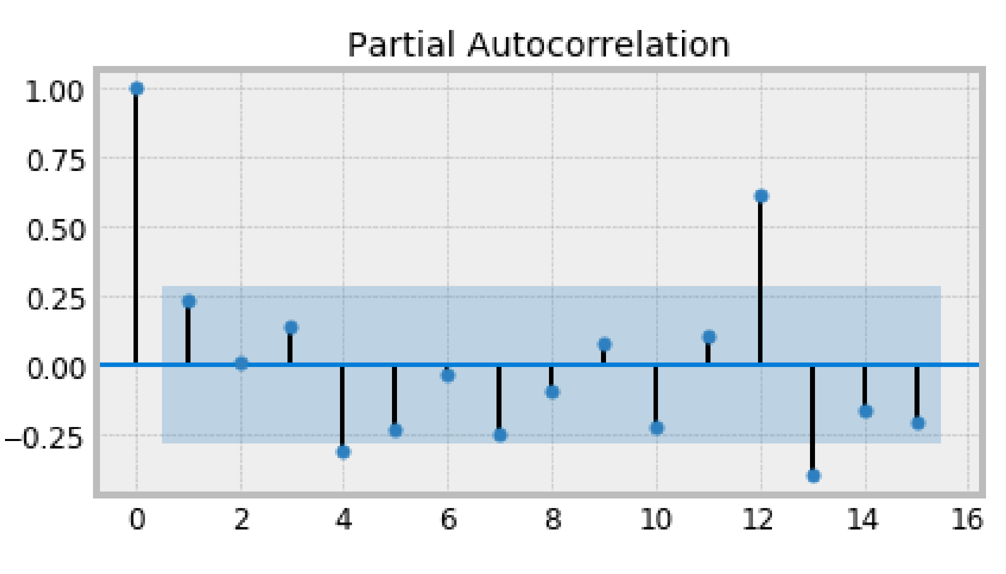
PACF plot remueve todas las componentes que ya fueron explicadas por lags pasados. Por lo tanto, solo tenemos lags que tienen correlación con el residuo, es decir la componente no explicada por lags pasados.

ACF y PACF plots

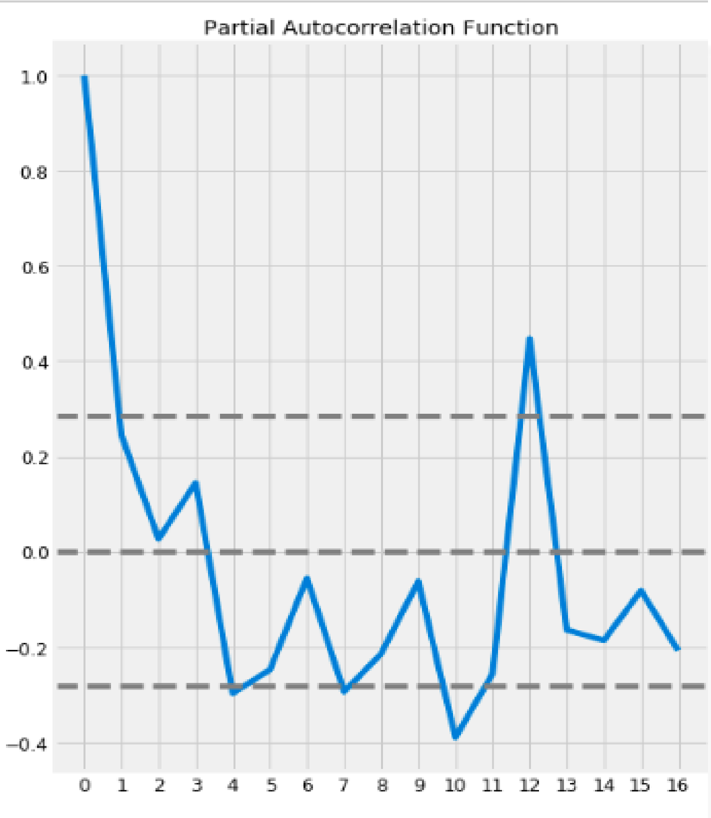


<u>Parcial Autocorrelation function (PACF)</u>

p es el valor de retraso donde el gráfico PACF cruza el intervalo de confianza superior por primera vez.



p=1



AIC (Akaike Information Criterion)

Mide qué tan bien un modelo se ajusta a los datos teniendo en cuenta la complejidad global del modelo.

Un modelo que proporciona un buen ajuste pero usando muchos features recibirá un valor grande de AIC comparado con otro que devuelve el mismo ajuste pero utilizando menos features.

SEASONAL ARIMA (SARIMA)

Algoritmo SARIMA (Seasonal ARIMA)

*Un modelo **SARIMA** pertenece a la clase de modelos ARIMA y es un modelo ARIMA con una componente seasonal.*

$$\boxed{ARIMA(p,d,q)x(P,D,Q)s}$$

Algoritmo SARIMA (Seasonal ARIMA)

ARIMA(p,d,q)x(P,D,Q)s

Componente no seasonal

Componente seasonal

season o periodicidad. Por ejemplo: 12 meses, 52 semanas, 4 para trimestres.

Los parametros P y Q realizan la predicción usando valores pasados y errores en términos anteriores que son múltiplos de s (el alcance del período).

Algoritmo SARIMA (Seasonal ARIMA)

Ejemplo:

Dada una serie temporal mensual (s=12)

- Si P=1 el modelo (autoregresivo de primer orden estacional) predecirá y(t) usando y(t-12).

- Si P=2 el modelo (autoregresivo de segundo orden estacional) predecirá y(t) usando y(t-12) y y(t-24).

HOLT-WINTERS

Holt-Winters y sus predecesores

Simple Exponential Smoothing → Double Exponential Smoothing → Holt-Winters

- **Simple Exponential Smoothing**: Predicción constante a través del tiempo
- **Double Exponential Smoothing**: Tendencia creciente o decreciente
- **Holt-Winters**: Tendencia con comportamiento estacionales

Simple Exponential Smoothing

coeficiente de suavizado del nivel

Predicción en el tiempo **t+1**

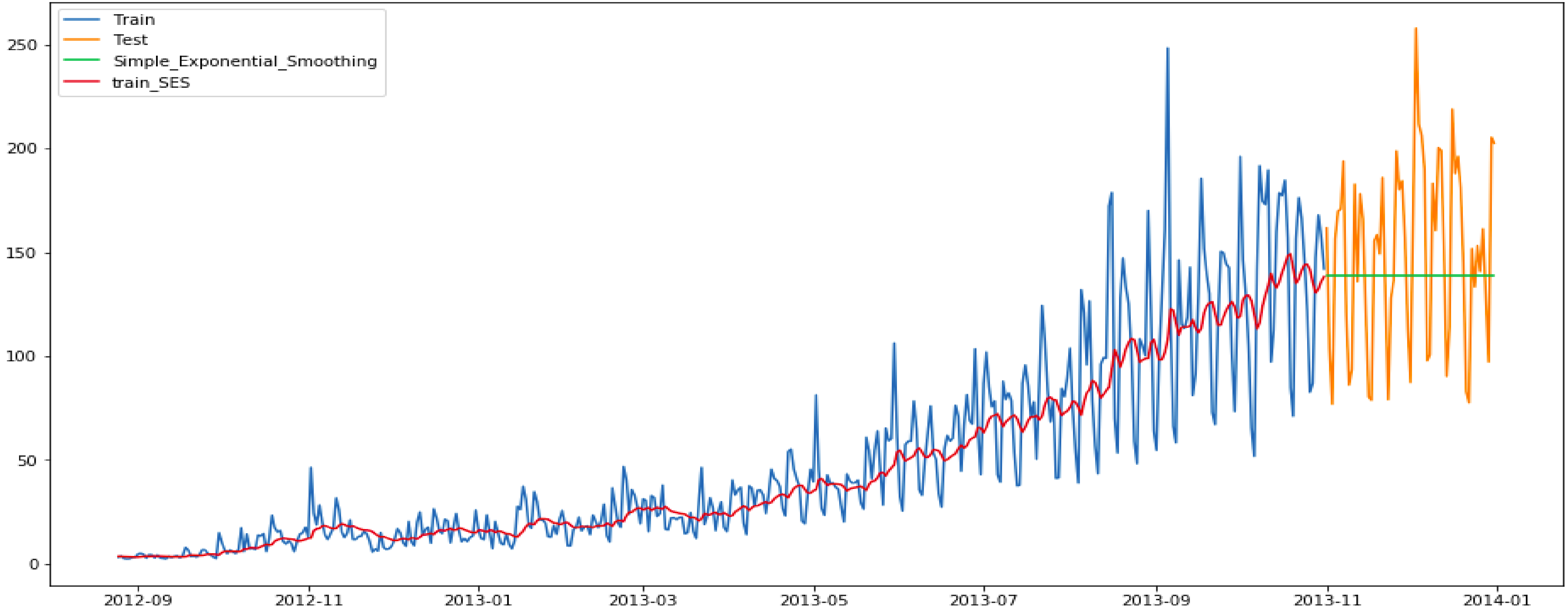
$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_{t-1}$$

Predicción en el tiempo **t-1**

Valor real en el tiempo **t**

Simple Exponential Smoothing

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_{t-1}$$



Double Exponential Smoothing

Simple Exponential Smoothing

Double Exponential Smoothing

Holt-Winters

Double Exponential Smoothing

Caso 1

Nivel a tiempo **t**

Tendencia multiplicativa tiempo **t**

Tendencia multiplicativa tiempo **t-1**

Predicción a tiempo **t+m**

$$\hat{F}_{t+m} = \boxed{\hat{y}_t \cdot T_t^m}$$

Predicción

$$\hat{y}_t = \alpha \cdot y_t + (1-\alpha) \cdot \boxed{(\hat{y}_{t-1} \cdot \hat{T}_{t-1})}$$

Nivel

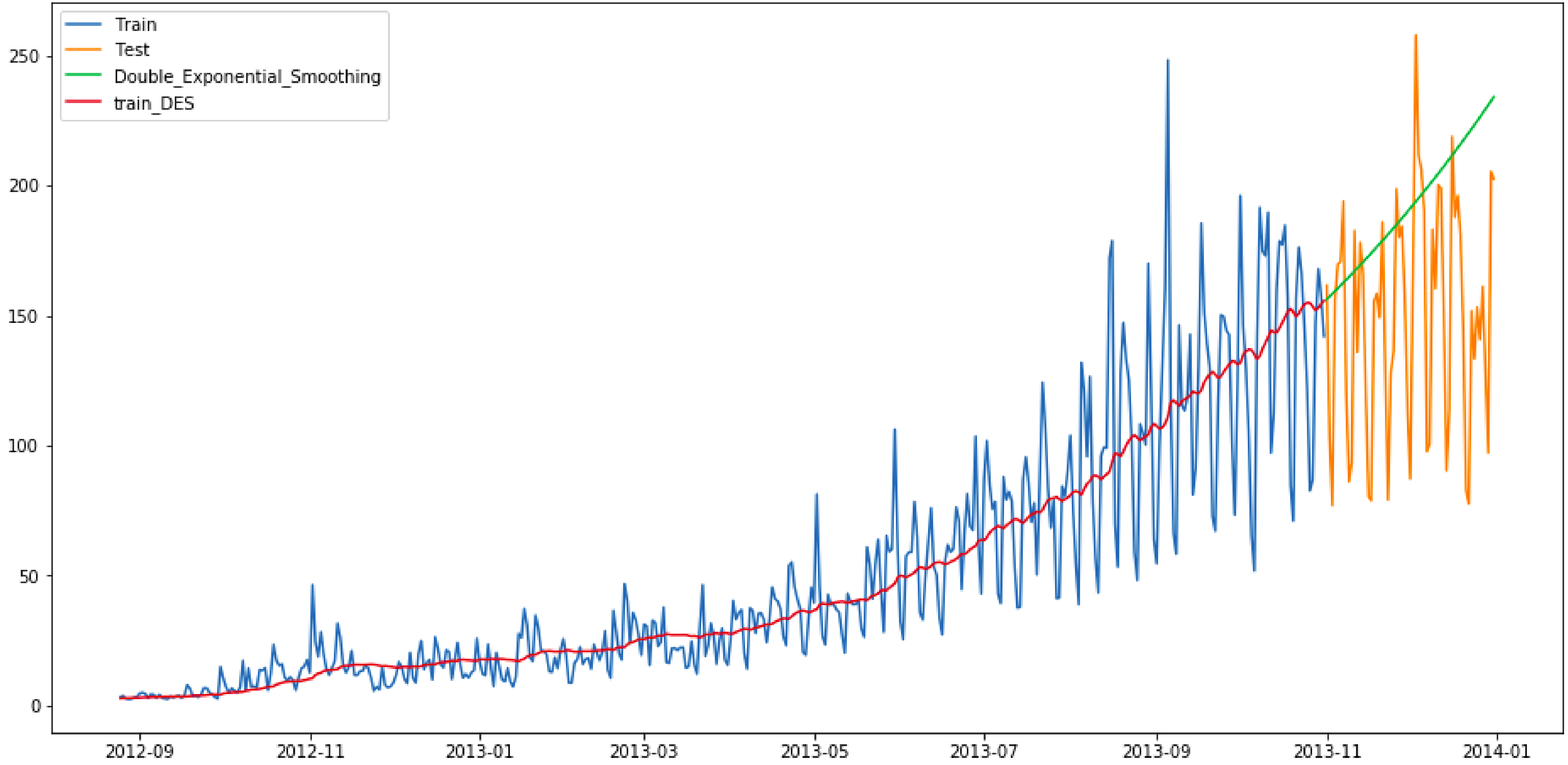
$$\hat{T}_t = \beta \cdot \boxed{(\hat{y}_t / \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$

Tendencia multiplicativa

Coeficiente de suavizado de la tendencia

Double Exponential Smoothing

Caso 1



Double Exponential Smoothing

Caso 2

Nivel a tiempo **t**

Tendencia multiplicativa tiempo **t**

Tendencia multiplicativa tiempo **t-1**

Predicción a tiempo **t+m**

$$\hat{F}_{t+m} = \boxed{\hat{y}_t + m \cdot T_t} \quad \textit{Predicción}$$

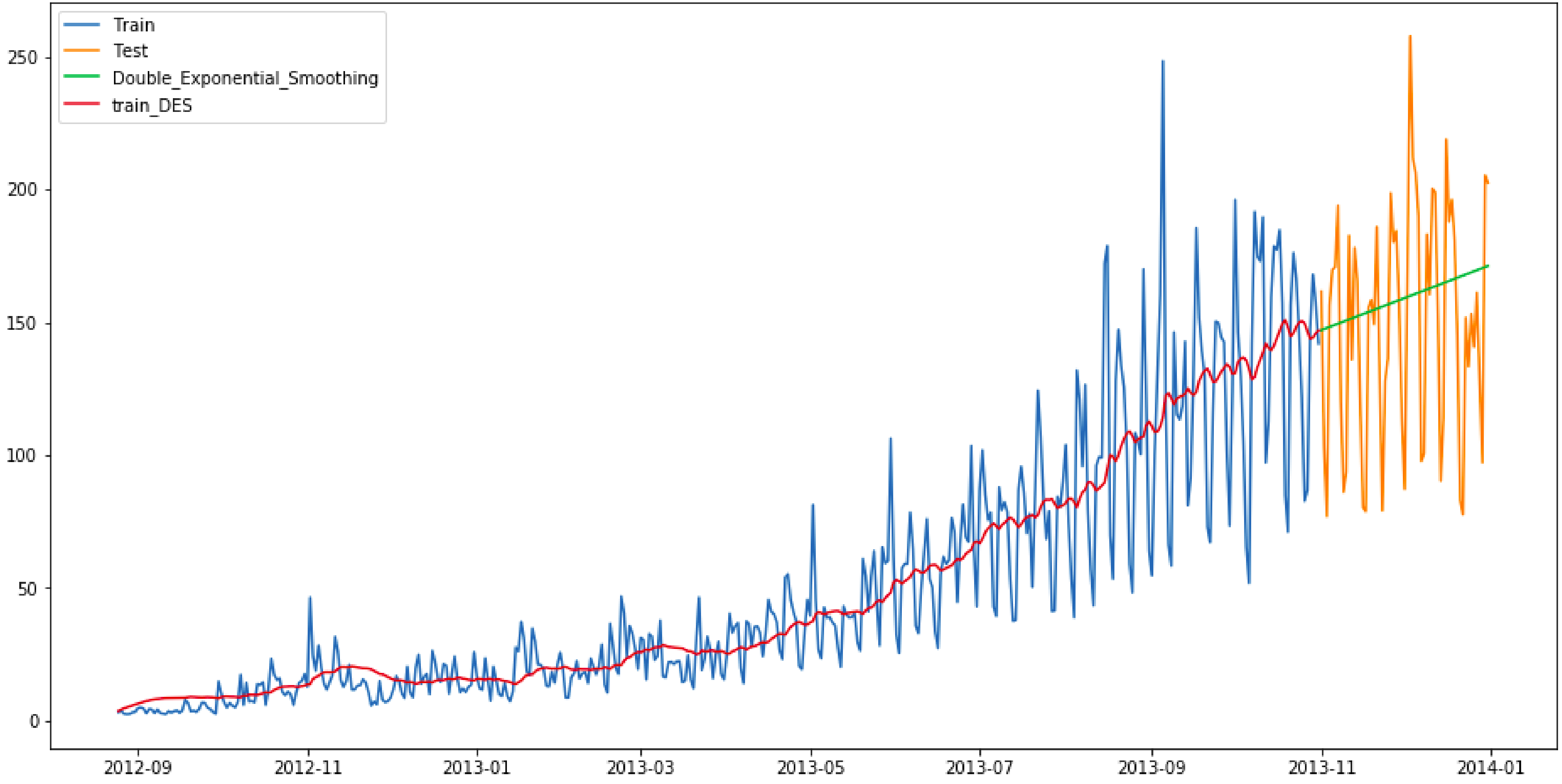
$$\hat{y}_t = \alpha \cdot y_t + (1-\alpha) \cdot \boxed{(\hat{y}_{t-1} + \hat{T}_{t-1})} \quad \textit{Nivel}$$

$$\hat{T}_t = \beta \cdot \boxed{(\hat{y}_t - \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1} \quad \textit{Tendencia aditiva}$$

Coeficiente de suavizado de la tendencia

Double Exponential Smoothing

Caso 2



Double Exponential Smoothing

$$\hat{F}_{t+m} = \boxed{\hat{y}_t \cdot T_t^m}$$

Predicción

$$\hat{y}_t = \alpha \cdot y_t + (1-\alpha) \cdot \boxed{(\hat{y}_{t-1} \cdot \hat{T}_{t-1})}$$

Nivel

$$\hat{T}_t = \beta \cdot \boxed{(\hat{y}_t / \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$

Tendencia multiplicativa

$$\hat{F}_{t+m} = \boxed{\hat{y}_t + m \cdot T_t}$$

Predicción

$$\hat{y}_t = \alpha \cdot y_t + (1-\alpha) \cdot \boxed{(\hat{y}_{t-1} + \hat{T}_{t-1})}$$

Nivel

$$\hat{T}_t = \beta \cdot \boxed{(\hat{y}_t - \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$

Tendencia multiplicativa

Holt-Winters

Simple ExponentialSmoothing → Double Exponential Smoothing → Holt-Winters

Holt-Winters

Caso 1

Periodicidad a tiempo **t+m** y menor que **L**

$$\hat{F}_{t+m} = \boxed{\boxed{(\hat{y}_t + m \cdot \hat{T}_t)} + s_{t-L+1+(m-1) mod L}}$$

$$\hat{y}_t = \alpha \cdot \boxed{(y_t - s_{t-L})} + (1-\alpha) \cdot \boxed{(\hat{y}_{t-1} + \hat{T}_{t-1})}$$

$$\hat{T}_t = \beta \cdot \boxed{(y_t - \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$

$$s_t = \gamma \cdot \boxed{(y_t - \boxed{\hat{y}_{t-1} - \hat{T}_{t-1})}} + (1-\gamma) \cdot s_{t-L}$$

Predicción

Nivel

Tendencia aditiva

Periodicidad aditiva

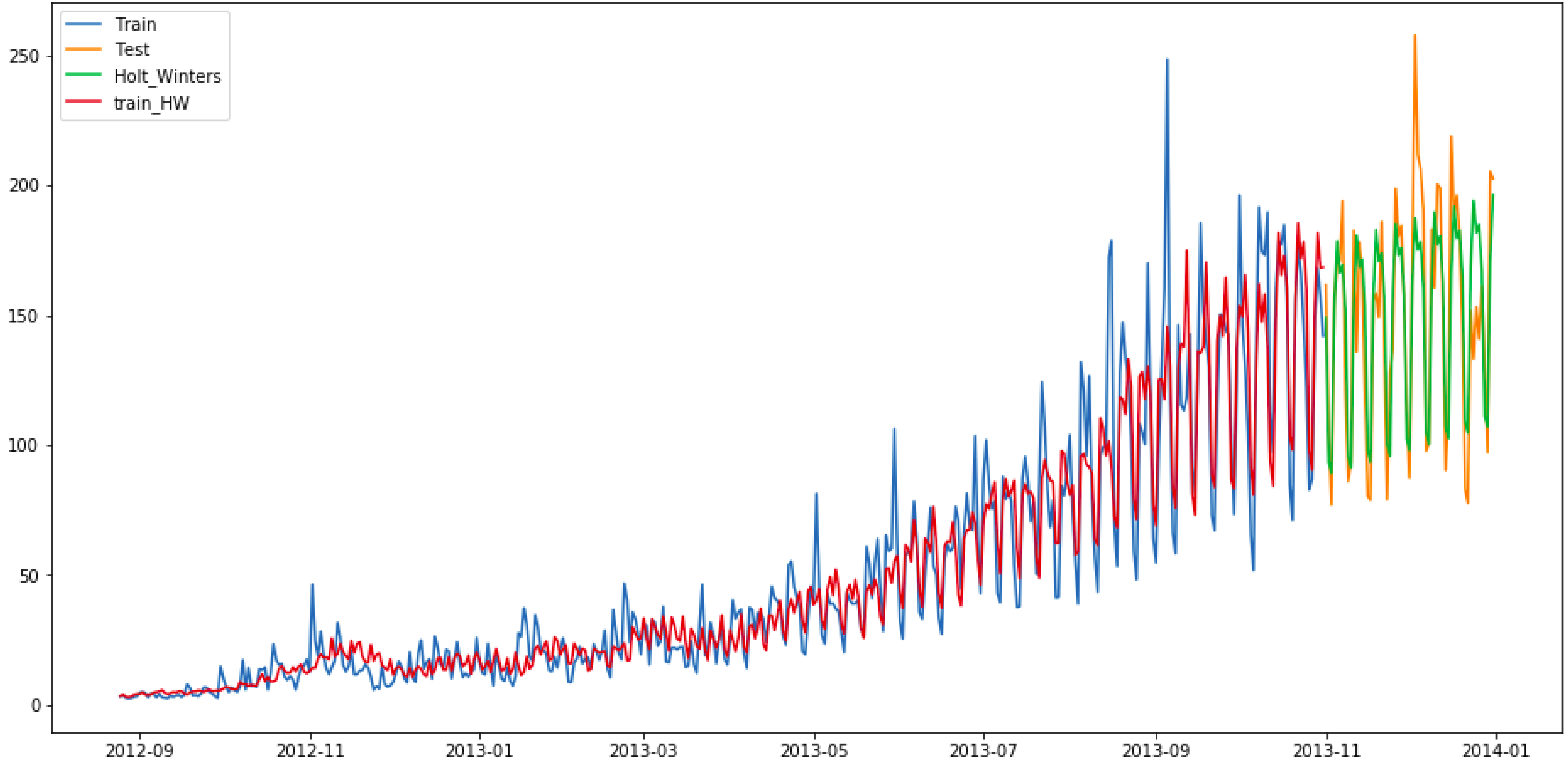
Periodicidad a tiempo **t**

Coeficiente de suavizado estacional

período

Holt-Winters

Caso 1



Holt-Winters

Caso 2

Periodicidad a tiempo **t+m** y menor que **L**

$$\hat{F}_{t+m} = \boxed{\boxed{(\hat{y}_t + m \cdot \hat{T}_t)} \cdot s_{t-L+1+(m-1) \bmod L}}$$ *Predicción*

$$\hat{y}_t = \alpha \cdot \boxed{(y_t / s_{t-L})} + (1-\alpha) \cdot \boxed{(\hat{y}_{t-1} + \hat{T}_{t-1})}$$ *Nivel*

$$\hat{T}_t = \beta \cdot \boxed{(y_t - \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$ *Tendencia aditiva*

$$s_t = \gamma \cdot \boxed{(y_t / \boxed{(\hat{y}_{t-1} + \hat{T}_{t-1})})} + (1-\gamma) \cdot s_{t-L}$$ *Periodicidad multiplicativa*

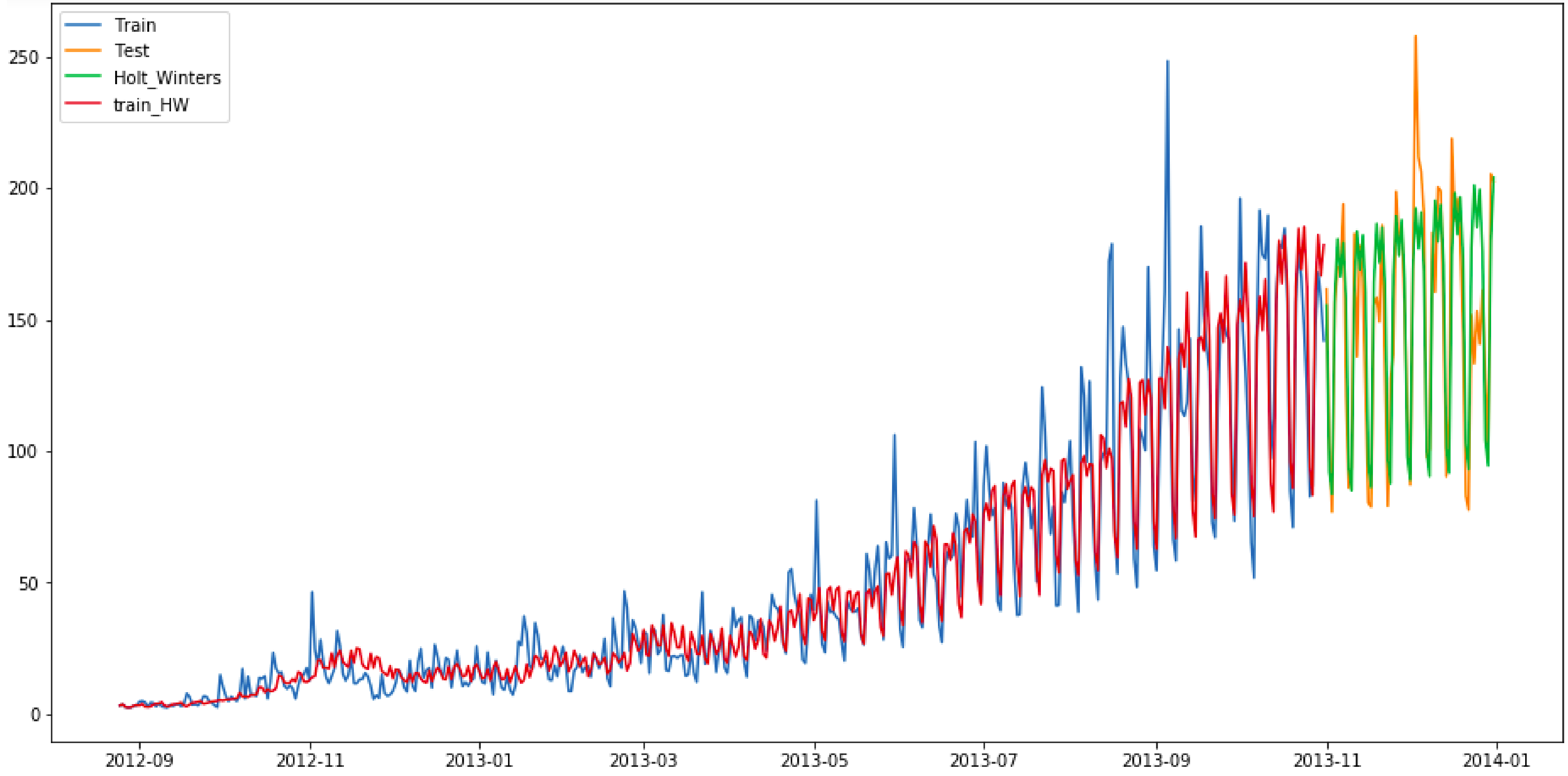
Periodicidad a tiempo **t**

Coeficiente de suavizado estacional

período

Holt-Winters

Caso 2



Holt-Winters

Caso 3

Periodicidad a tiempo **t+m** y menor que **L**

$$\hat{F}_{t+m} = \boxed{\boxed{(\hat{y}_t \cdot \hat{T}_t^{m})} + s_{t-L+1+(m-1) \bmod L}}$$

$$\hat{y}_t = \alpha \cdot \boxed{(y_t - s_{t-L})} + (1-\alpha) \boxed{\cdot\, \hat{y}_{t-1} \cdot \hat{T}_{t-1}}$$

$$\hat{T}_t = \beta \cdot \boxed{(y_t / \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$

$$s_t = \gamma \cdot \boxed{(y_t - \boxed{\hat{y}_t \cdot \hat{T}_{t-1})}} + (1-\gamma) \cdot s_{t-L}$$

Predicción
Nivel
Tendencia multiplicativa
Periodicidad aditiva

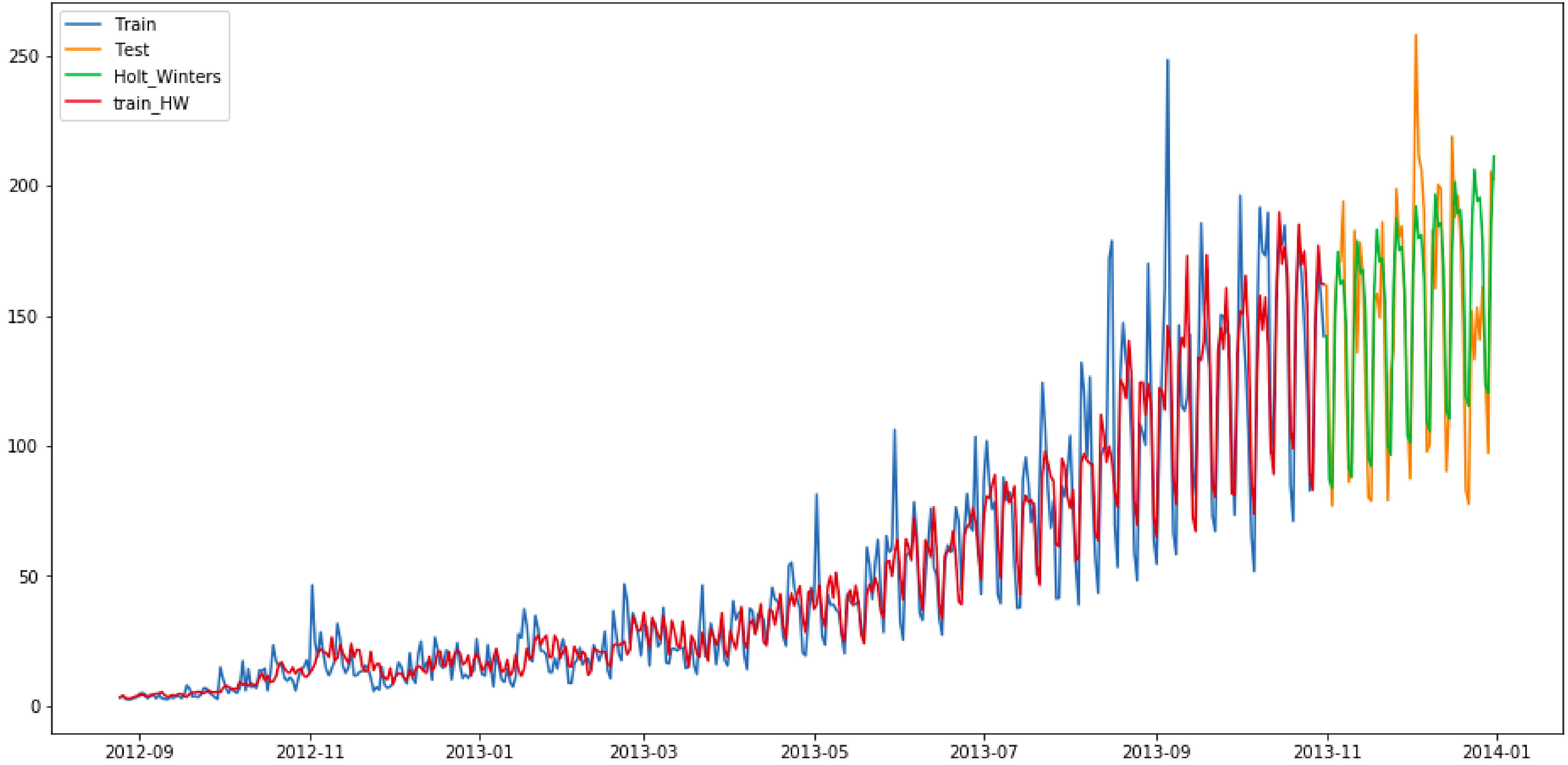
Periodicidad a tiempo **t**

Coeficiente de suavisado estacional

período

Holt-Winters

Caso 3



Holt-Winters

Caso 4

Periodicidad a tiempo
t+m y menor que **L**

$$\hat{F}_{t+m} = \boxed{\hat{y}_t \cdot \hat{T}_t^m} \cdot s_{t-L+1+(m-1) \bmod L}$$ *Predicción*

$$\hat{y}_t = \alpha \cdot \boxed{(y_t / s_{t-L})} + (1-\alpha) \cdot \boxed{\hat{y}_{t-1} \cdot \hat{T}_{t-1}}$$ *Nivel*

$$\hat{T}_t = \beta \cdot \boxed{(y_t / \hat{y}_{t-1})} + (1-\beta) \cdot \hat{T}_{t-1}$$ *Tendencia multiplicativa*

$$s_t = \gamma \cdot \boxed{(y_t / \boxed{(\hat{y}_t \cdot \hat{T}_{t-1})})} + (1-\gamma) \cdot s_{t-L}$$ *Periodicida multiplicativa*

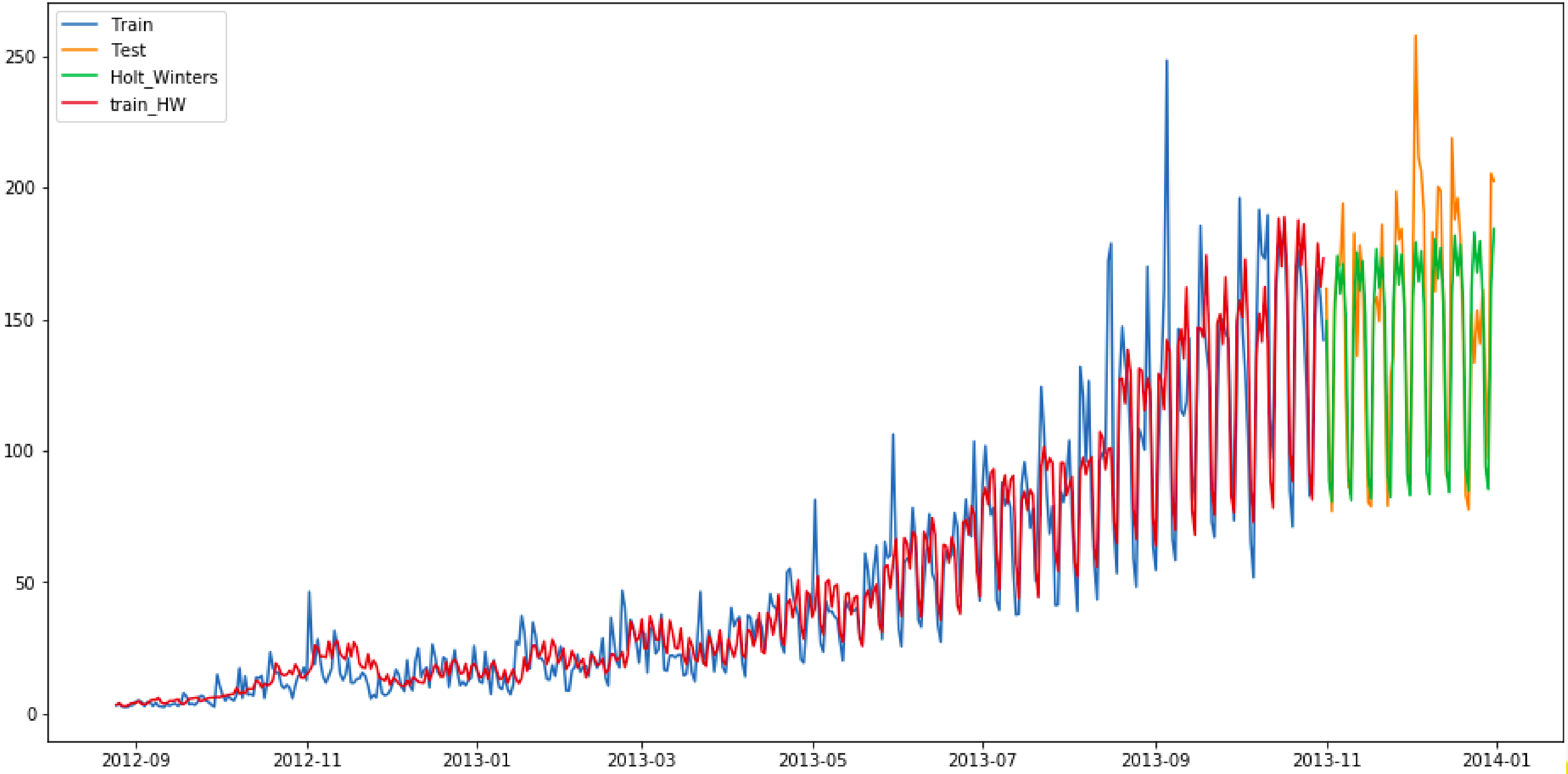
Periodicidad a tiempo **t**

Coeficiente de suavizado estacional

período

Holt-Winters

Caso 4



Holt-Winters

Periodicidad

Tendencia

Tendencia Aditiva – Periodicidad Aditiva

$$\hat{F}_{t+m} = (\hat{y}_t + m \cdot \hat{T}_t) + s_{t-L+1+(m-1) mod L}$$

$$\hat{y}_t = \alpha \cdot (y_t - s_{t-L}) + (1-\alpha) \cdot (\hat{y}_{t-1} + \hat{T}_{t-1})$$

$$\hat{T}_t = \beta \cdot (y_t - \hat{y}_{t-1}) + (1-\beta) \cdot \hat{T}_{t-1}$$

$$s_t = \gamma \cdot (y_t - \hat{y}_{t-1} - \hat{T}_{t-1}) + (1-\gamma) \cdot s_{t-L}$$

Tendencia Aditiva – Periodicidad Multiplicativa

$$\hat{F}_{t+m} = (\hat{y}_t + m \cdot \hat{T}_t) \cdot s_{t-L+1+(m-1) mod L}$$

$$\hat{y}_t = \alpha \cdot (y_t / s_{t-L}) + (1-\alpha) \cdot (\hat{y}_{t-1} + \hat{T}_{t-1})$$

$$\hat{T}_t = \beta \cdot (y_t - \hat{y}_{t-1}) + (1-\beta) \cdot \hat{T}_{t-1}$$

$$s_t = \gamma \cdot (y_t / (\hat{y}_{t-1} + \hat{T}_{t-1})) + (1-\gamma) \cdot s_{t-L}$$

Tendencia Multiplicativa – Periodicidad Aditiva

$$\hat{F}_{t+m} = (\hat{y}_t \cdot \hat{T}_t^m) + s_{t-L+1+(m-1) mod L}$$

$$\hat{y}_t = \alpha \cdot (y_t - s_{t-L}) + (1-\alpha) \cdot \hat{y}_{t-1} \cdot \hat{T}_{t-1}$$

$$\hat{T}_t = \beta \cdot (y_t / \hat{y}_{t-1}) + (1-\beta) \cdot \hat{T}_{t-1}$$

$$s_t = \gamma \cdot (y_t - \hat{y}_t \cdot \hat{T}_{t-1}) + (1-\gamma) \cdot s_{t-L}$$

Tendencia Multiplicativa – Periodicidad Multiplicativa

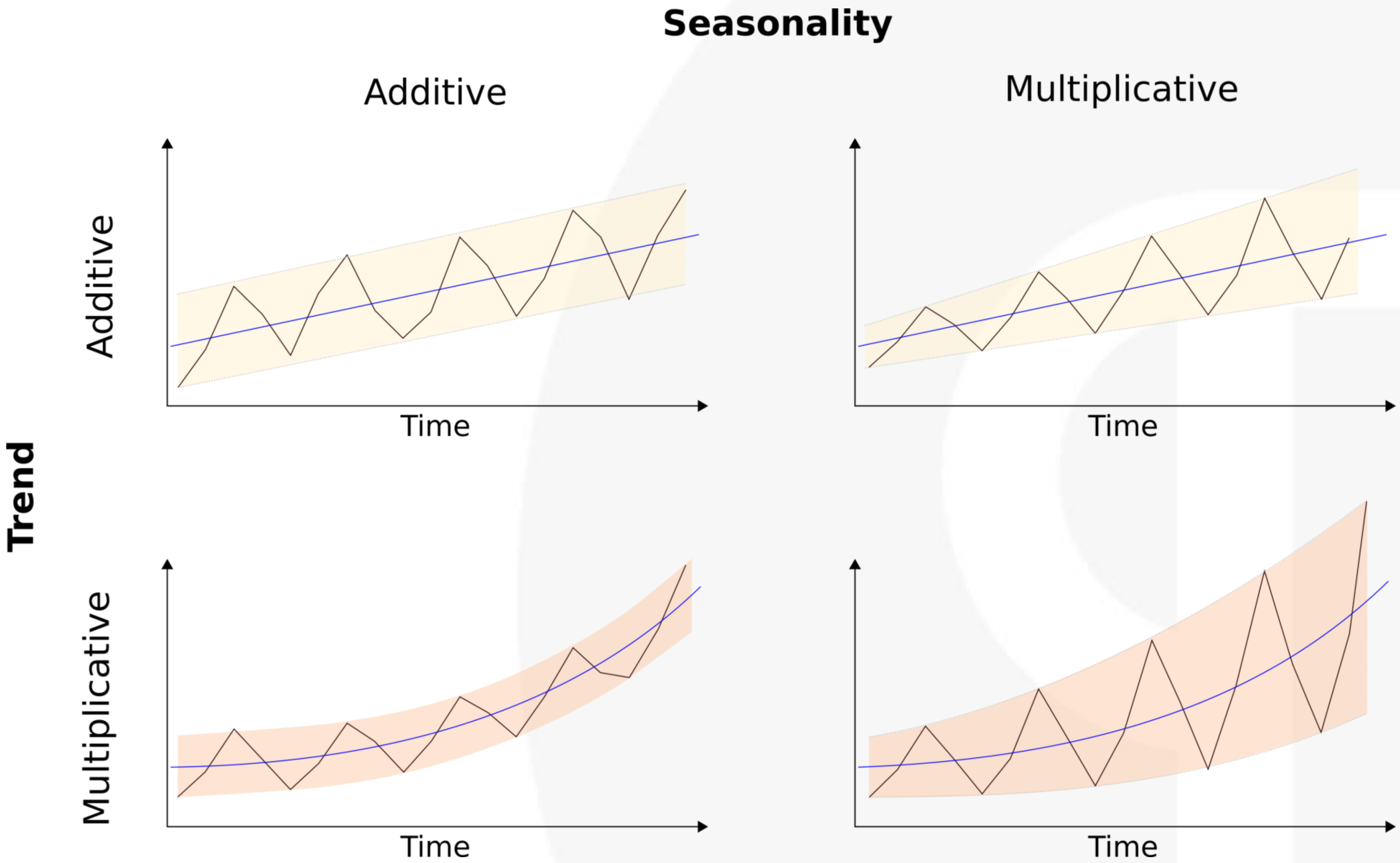
$$\hat{F}_{t+m} = \hat{y}_t \cdot \hat{T}_t^m \cdot s_{t-L+1+(m-1) mod L}$$

$$\hat{y}_t = \alpha \cdot (y_t / s_{t-L}) + (1-\alpha) \cdot \hat{y}_{t-1} \cdot \hat{T}_{t-1}$$

$$\hat{T}_t = \beta \cdot (y_t / \hat{y}_{t-1}) + (1-\beta) \cdot \hat{T}_{t-1}$$

$$s_t = \gamma \cdot (y_t / (\hat{y}_t \cdot \hat{T}_{t-1})) + (1-\gamma) \cdot s_{t-L}$$

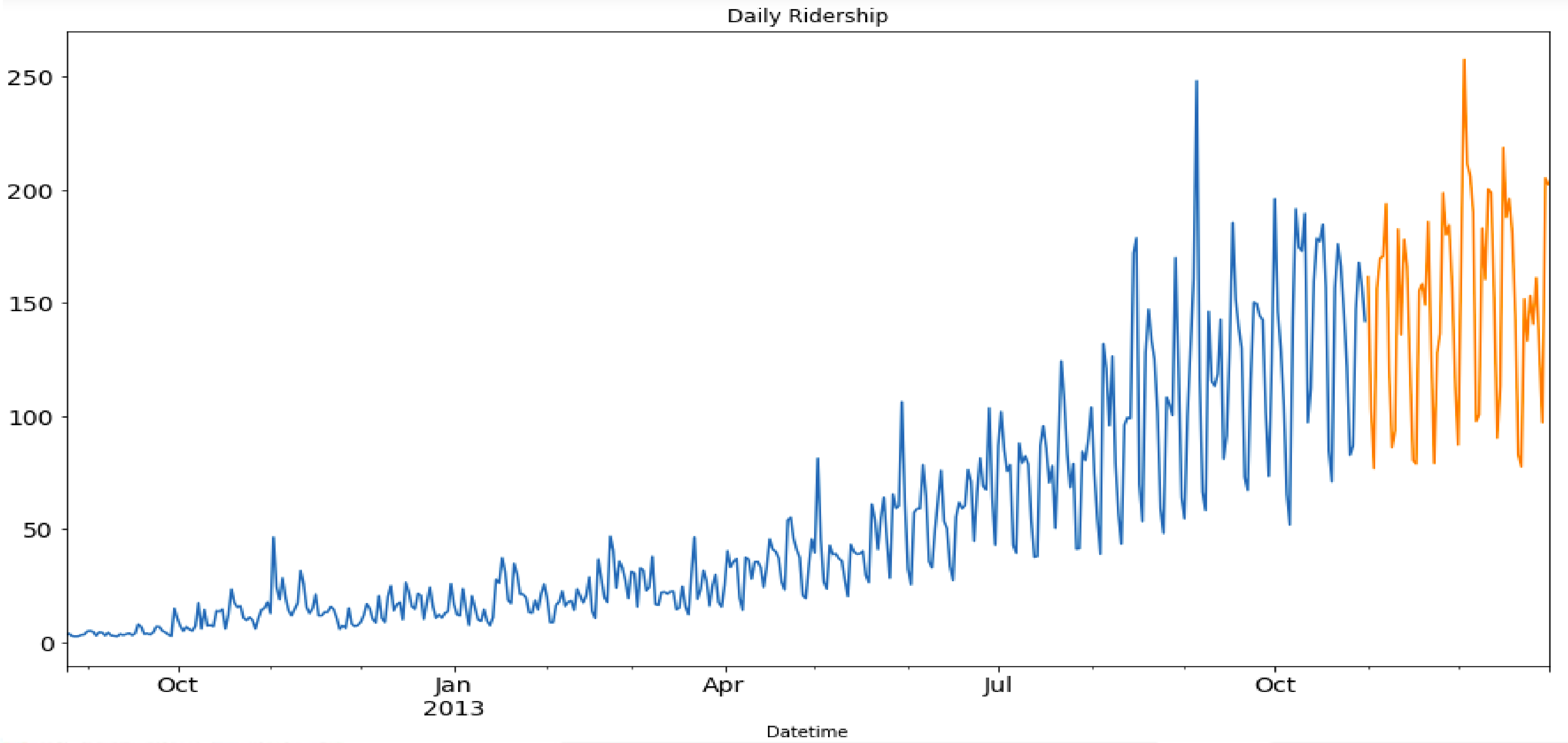
¿Cómo elegir cuál modelo es el ideal?



https://medium.com/better-programming/exponential-smoothing-methods-for-time-series-forecasting-d571005cdf80

¿Cómo elegir cuál modelo es el ideal?

Podemos ver que la tendencia es multiplicativa al igual que la periodicidad



XGBOOST

Proceso de boosting

ID	Married	Gender	Current City	Monthly Income	Age (target)
1	Y	M	A	51,000	35
2	N	F	B	25,000	24
3	Y	M	A	74,000	38
4	N	F	A	29,000	30
5	N	F	B	37,000	33

ID	Married	Gender	Current City	Monthly Income	Age (target)	Mean Age (prediction 1)	Residual 1
1	Y	M	A	51,000	35	32	3
2	N	F	B	25,000	24	32	-8
3	Y	M	A	74,000	38	32	6
4	N	F	A	29,000	30	32	-2
5	N	F	B	37,000	33	32	1

Error de la primera predicción

Esto es lo que queremos predecir

Primera predicción (promedio)

https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

Proceso de boosting

ID	Married	Gender	Current City	Monthly Income	Age (target)
1	Y	M	A	51,000	35
2	N	F	B	25,000	24
3	Y	M	A	74,000	38
4	N	F	A	29,000	30
5	N	F	B	37,000	33

ID	Married	Gender	Current City	Monthly Income	Age (target)	Mean Age (prediction 1)	Residual 1
1	Y	M	A	51,000	35	32	3
2	N	F	B	25,000	24	32	-8
3	Y	M	A	74,000	38	32	6
4	N	F	A	29,000	30	32	-2
5	N	F	B	37,000	33	32	1

3, -8, 6, -2, 1

X<=0
-5

X>0
3

ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)
1	35	32	3	3	35
2	24	32	-8	-5	27
3	38	32	6	3	35
4	30	32	-2	-5	27
5	33	32	1	3	35

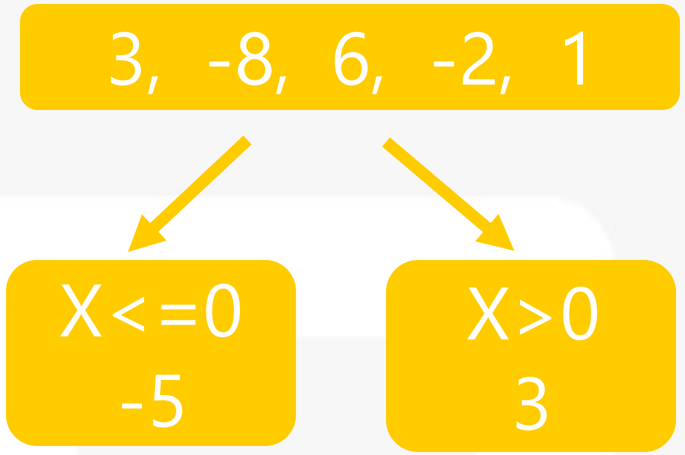
Predicción del árbol que minimiza el error

https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

Proceso de boosting

ID	Married	Gender	Current City	Monthly Income	Age (target)
1	Y	M	A	51,000	35
2	N	F	B	25,000	24
3	Y	M	A	74,000	38
4	N	F	A	29,000	30
5	N	F	B	37,000	33

ID	Married	Gender	Current City	Monthly Income	Age (target)	Mean Age (prediction 1)	Residual 1
1	Y	M	A	51,000	35	32	3
2	N	F	B	25,000	24	32	-8
3	Y	M	A	74,000	38	32	6
4	N	F	A	29,000	30	32	-2
5	N	F	B	37,000	33	32	1



Error del último modelo entrenado

Predicción del modelo

ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)
1	35	32	3	3	35
2	24	32	-8	-5	27
3	38	32	6	3	35
4	30	32	-2	-5	27
5	33	32	1	3	35

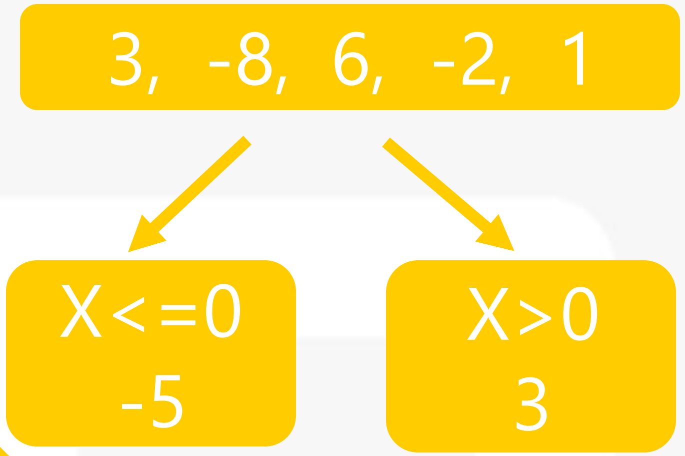
ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)	Residual 2 (latest target)
1	35	32	3	3	35	0
2	24	32	-8	-5	27	-3
3	38	32	6	3	35	-3
4	30	32	-2	-5	27	3
5	33	32	1	3	35	-2

https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

Proceso de boosting

ID	Married	Gender	Current City	Monthly Income	Age (target)
1	Y	M	A	51,000	35
2	N	F	B	25,000	24
3	Y	M	A	74,000	38
4	N	F	A	29,000	30
5	N	F	B	37,000	33

ID	Married	Gender	Current City	Monthly Income	Age (target)	Mean Age (prediction 1)	Residual 1
1	Y	M	A	51,000	35	32	3
2	N	F	B	25,000	24	32	-8
3	Y	M	A	74,000	38	32	6
4	N	F	A	29,000	30	32	-2
5	N	F	B	37,000	33	32	1



Remplazar y volver a empezar

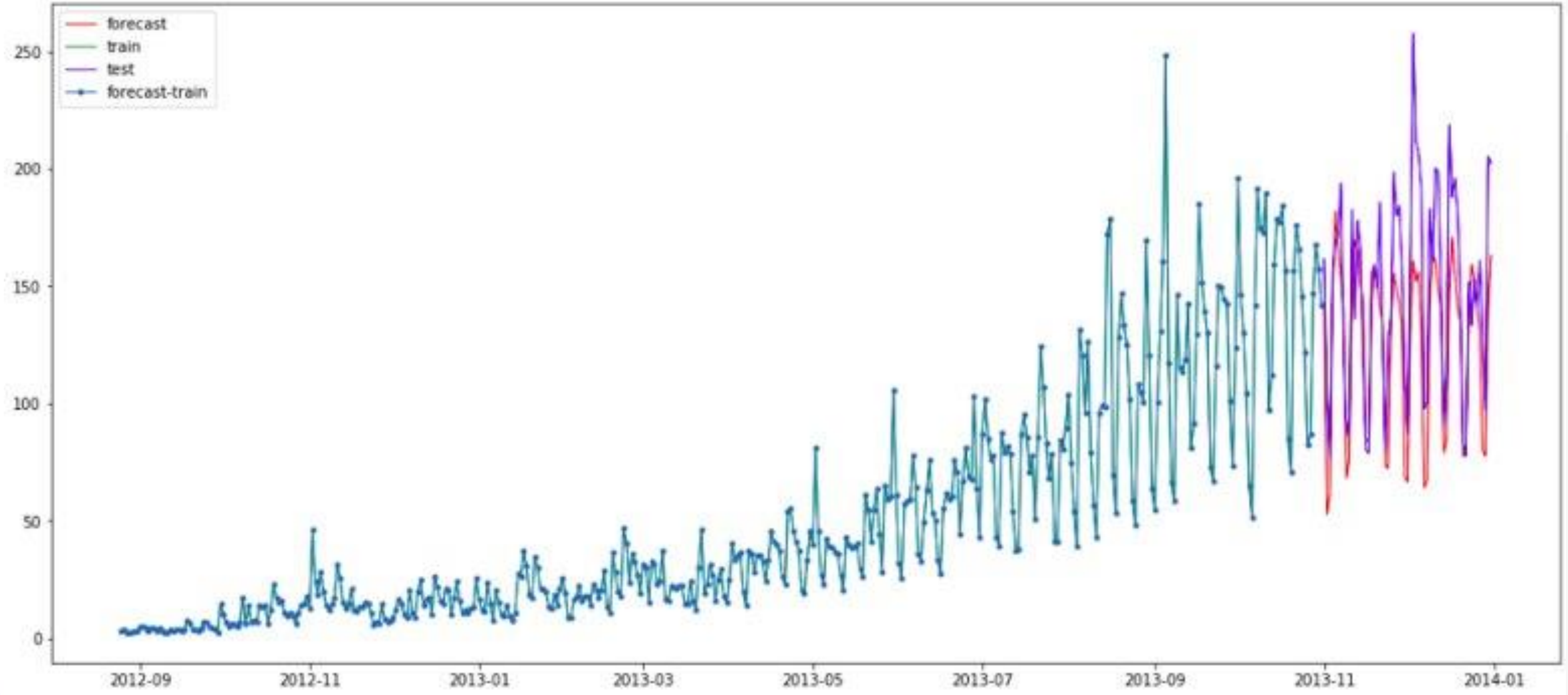
ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)
1	35	32	3	3	35
2	24	32	-8	-5	27
3	38	32	6	3	35
4	30	32	-2	-5	27
5	33	32	1	3	35

ID	Age (target)	Mean Age (prediction 1)	Residual 1 (new target)	Prediction 2	Combine (mean+pred2)	Residual 2 (latest target)
1	35	32	3	3	35	0
2	24	32	-8	-5	27	-3
3	38	32	6	3	35	-3
4	30	32	-2	-5	27	3
5	33	32	1	3	35	-2

https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/

Ejemplo de XGBoost

Importante: XGBoost no usa variables temporales, por lo tanto, hay que cambiarlas a variables numéricas



Conclusiones

Holt-Winters

- *Datos con nivel, tendencia y estacionalidad.*
- *No supone estacionariedad.*
- *No es sensible a outliers.*

XGBoost

- *Se recomienda usarlo como último recurso.*
- *En general proporciona buenos resultados.*

ARIMA/SARIMA

- *Datos estacionarios.*
- *Predicciones a corto plazo.*
- *Sensible a outliers.*
- *Mínimo de 38-40 datos históricos.*
- *Componentes estacionales y no estacionales.*

MUCHAS GRACIAS!!