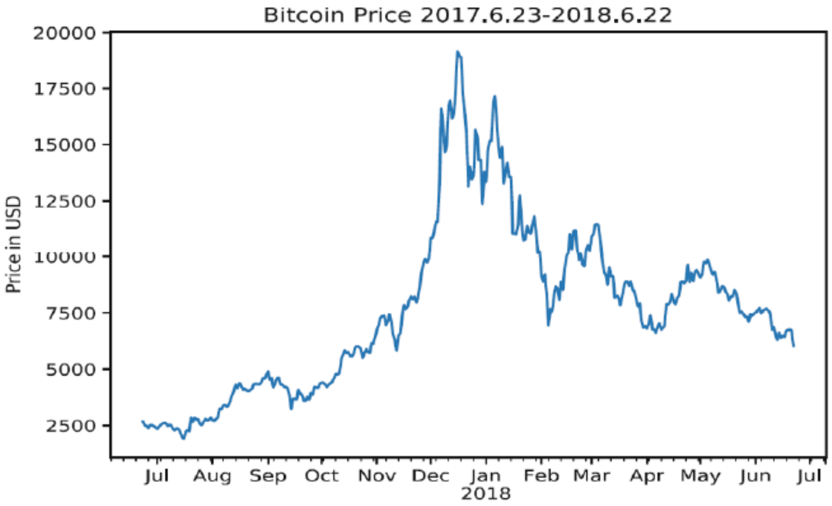
INSTITUTO
Data Science

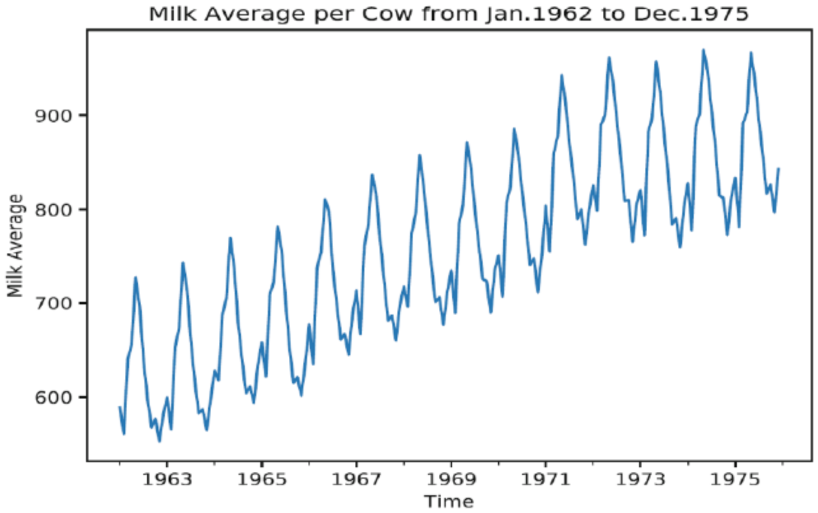
Análisis de Series Temporales con Python

Diplomatura en Data Science con R y Python
Curso complementario

¿Qué son las Series Temporales?

- Son una colección ordenada de registros de cierta magnitud asociada a un fenómeno natural o social, separados por un intervalo de tiempo dado.
- La magnitud medida está afectada por procesos deterministas y aleatorios, por lo tanto, las series temporales son procesos estocásticos y sus valores refieren a una variable aleatoria *X*.



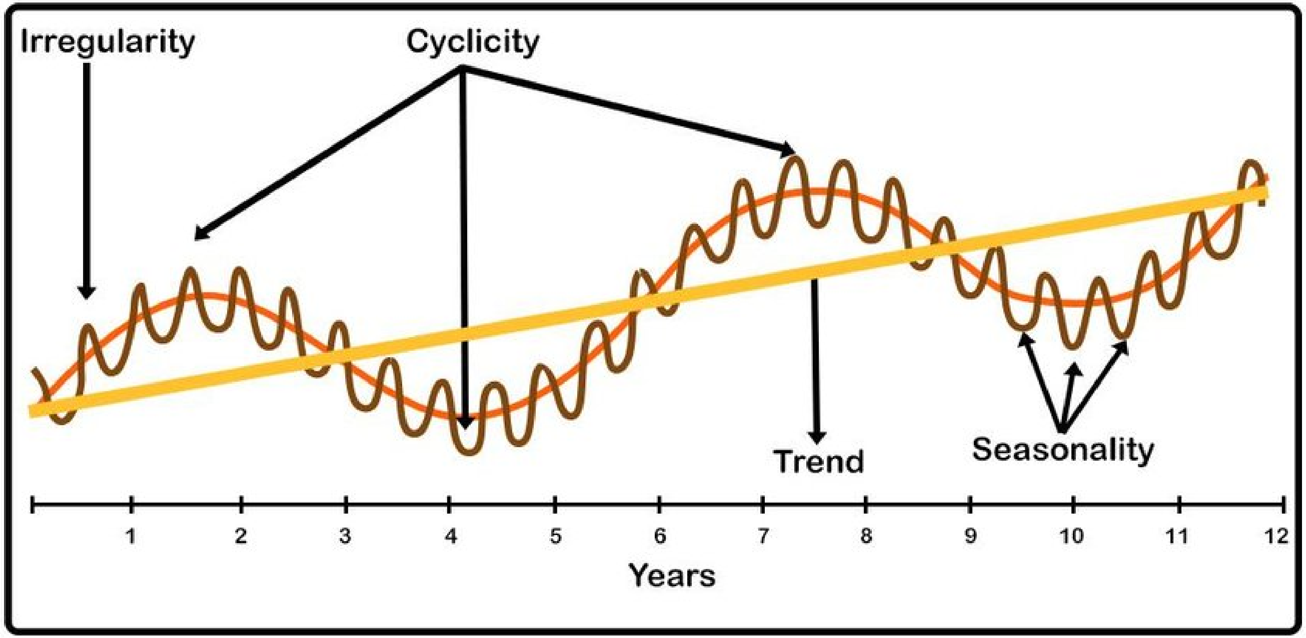


¿Qué son las Series Temporales?

- Modelos para analizar series temporales:
  - ARMA
  - ARIMA
  - SARIMA
  - VAR
  - ARCH
  - REGARMA
  - TBATS
  - Redes neuronales
  - Etc.
- El objetivo de este tipo de tratamiento de datos es analizar la serie temporal en sí misma y predecir comportamientos futuros.

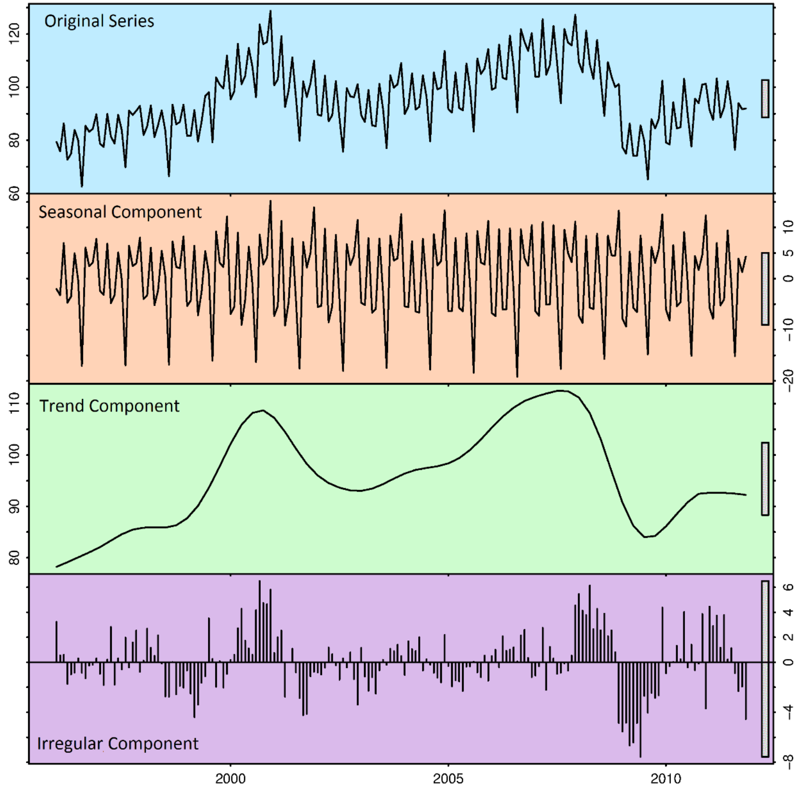
Componentes de una Serie Temporal

- Tendencia (T_t)
- Estacionalidad (S_t)
- Variación aleatoria o residuo (R_t)
- Ciclicidad (C_t)



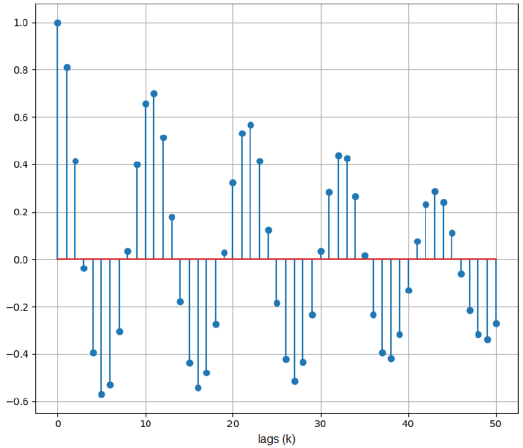
Componentes de una Serie Temporal

- Se puede descomponer a una serie temporal bajo dos modelos:
  - Aditivo: $X_t = T_t + S_t + R_t$
  - Multiplicativo: $X_t = T_t S_t R_t$
- Puede aplicarse también un modelo mixto.



Funciones momento

- Al ser las series temporales una colección de valores de una variable aleatoria, se pueden definir las siguientes funciones momento (sobre la muestra):
  - Media
  - Varianza
  - Autocovarianza
  - Autocorrelación simple
  - Autocorrelación parcial



- Otra función muy útil es la de medias móviles, las cuales pueden ser centradas o asimétricas:

$$MM_C = \frac{X_{t-d} + X_{t-d+1} + \ldots + X_{t+d-1} + X_{t+d}}{2d+1} \qquad MM_A = \frac{X_{t-d+1} + X_{t-d+2} + \ldots + X_t}{d}$$

Estacionariedad

- Es una propiedad que permite aplicar algunos modelos de análisis. Si una serie temporal tiene esta propiedad, entonces se dice que ésta se encuentra en equilibrio estadístico.
- A pesar de que existe una definición matemática de esto, podemos notar si una serie temporal es estacionaria cuando se cumplen estas características:
  - Media y varianza constante
  - No cuenta con efectos de estacionalidad
  - Sus valores no están autocorrelacionados (la fas es cercana a 0 para k>0)
- Existen también pruebas estadísticas para comprobar si una serie es estacionaria: los tests ADF y KPSS son tests de hipótesis complementarios y se recomienda efectuarlos en conjunto para probar o descartar que una serie temporal sea estacionaria.

Estacionariedad

- Si una serie no es estacionaria, se pueden aplicar transformaciones para convertirla en estacionaria.
- La más común de estas transformaciones es la diferenciación y consiste en construir una nueva serie temporal a partir de la diferencia entre valores de la serie original. Existen varias formas de hacerlo:
  - Diferenciación de retraso 1: $Y_t = X_t - X_{t-1}$
  - Diferenciación de retraso *k*: $Y_t = X_t - X_{t-k}$
  - Diferenciación estacional: si *s* es el periodo estacional de la serie, se toma k=s.

Análisis exploratorio

- Explorar la serie temporal con el objetivo de determinar si los intervalos temporales tienen huecos y si hay valores nulos. En caso de que así sea, completar la serie por medio de la interpolación.
- Gráfico de la serie temporal para evaluar la existencia de tendencia o patrón estacional.
- Gráfico de medias móviles y desviación estándar, y cómputo de tests ADF y KPSS para evaluar la estacionariedad de la serie.
- Gráfico de las funciones de autocorrelación, para examinar la estacionariedad de la serie y los patrones estacionales. Además ayudarán en el proceso de construcción del modelo de análisis y predicción de la serie temporal.
- Descomposición de la serie temporal para entender su naturaleza y comportamiento.

Modelo de Medias Móviles MA(q)

- Dada una serie temporal estacionaria, es posible representarla por medio de un modelo de medias móviles de orden *q*:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

Donde ε es una función de ruido blanco y μ y θ_i son los coeficientes a determinar.

Modelo Autorregresivo AR(p)

- Dada una serie temporal estacionaria, es posible representarla por medio de un modelo autorregresivo de orden *p*:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t$$

  Donde ε es una función de ruido blanco y φ_i son los coeficientes a determinar.

- Notar que este modelo es una regresión lineal, donde las variables regresoras son los propios *p* valores anteriores de la serie temporal.

Modelo Autorregresivo de Medias Móviles ARMA(p,q)

- Dada una serie temporal estacionaria, es posible representarla por medio de un modelo autorregresivo de medias móviles de órdenes *p q*:

$$X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

  Donde ε es una función de ruido blanco y φ_i θ_i son los coeficientes a determinar.

- Para construir este modelo, se debe primero determinar si la serie es estacionaria, y segundo seleccionar los órdenes *p* y *q*.
- Los órdenes se elijen por medio de las funciones de autocorrelación. El retraso para el cual la fap es casi nula determina *p*; el retraso para el cual la fas es casi nula, determina *q*. Otra manera es elegir el modelo que minimice algún criterio estadístico para el error (AIC, BIC, HQIC).

Modelo Autorregresivo de Medias Móviles Integrado ARIMA(p,d,q)

- Se aplica en los casos donde la serie temporal no sea estacionaria.
- Se debe determinar la cantidad *d* de diferenciaciones (de retraso 1) necesarias para hacer que la serie sea estacionaria.
- Una vez fijado el valor de *d*, el proceso para construir el modelo es análogo al del modelo ARMA(p,q), aplicado a la serie estacionaria.

Modelo Autorregresivo de Medias Móviles Integrado Estacional SARIMA(p,d,q)(P,D,Q)s

- Aplica a casos donde la serie no sea estacionaria y tenga componente estacional, con periodo *s*.
- Es equivalente a realizar una descomposición multiplicativa, en una parte estacional y otra no estacional, donde a cada una se le aplica un modelo ARIMA. Es decir: SARIMA(p,d,q)(P,D,Q)s = ARIMA(p,d,q)ARIMA(P,D,Q).
- Se deben determinar la cantidad *d* de diferenciaciones de retraso 1, y la cantidad *D* de diferenciaciones estacionales con las que se logra convertir a la serie temporal en estacionaria.
- Luego, la determinación de los órdenes p,q,P,Q viene dada por las funciones de autocorrelación y por la minimización de algún criterio estadístico.

Modelo TBATS

- Es un modelo diferente, que se aplica a aquellas series temporales que tengan patrones estacionales complejos, como lo son aquellas que contienen múltiples periodos.
- Sus siglas se corresponden a las estrategias matemáticas utilizadas:
  - Trigonometric seasonality
  - Box-Cox transformation
  - ARMA errors
  - Trend components
  - Seasonal components
- El algoritmo evalúa varias opciones de modelos, con y sin las anteriores estrategias y devuelve el que minimice el criterio estadístico AIC.

Predicción y Evaluación

- División de datos en conjuntos de entrenamiento y prueba (particionar la serie temporal en dos).
- Entrenamiento del modelo elegido con el conjunto de entrenamiento.
- Predicción de valores de entrenamiento ("on-sample prediction").
- Predicción de valores de prueba ("future prediction").
- Evaluación de errores: no debe haber tendencias y deben tener baja desviación estándar. Deben tener una distribución normal.
- Predicción a futuro: si se considera que el modelo es satisfactorio, se puede utilizar para predecir información futura, de la que no se tienen referencias actualmente.

Bibliografía

Huang C., Petukhina, A. *Applied Time Series Analysis and Forecasting with Python.* Springer, 2022.

https://www.kaggle.com/code/prashant111/complete-guide-on-time-series-analysis-in-python/notebook

https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html

https://medium.com/analytics-vidhya/time-series-forecasting-using-tbats-model-ce8c429442a9

Alysha M. De Livera, Rob J. Hyndman and Ralph D. Snyder (2011): *Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing*, Journal of the American Statistical Association, 106:496, 1513-1527. http://dx.doi.org/10.1198/jasa.2011.tm09771