

## Guía de estudio - Series de Tiempo



¡Hola! Te damos la bienvenida a esta nueva guía de estudio.

### ¿En qué consiste esta guía?

En un mundo en constante movimiento, donde los datos temporales abundan en casi todos los aspectos de nuestra vida, el análisis y la predicción de patrones en series de tiempo se han vuelto esenciales. Desde las ventas mensuales de productos hasta la demanda de servicios, y desde la variabilidad climática hasta el flujo de tráfico en las calles, las series de tiempo ofrecen una ventana única para entender la evolución de los fenómenos a lo largo del tiempo. A medida que nuestra capacidad para recopilar y almacenar datos temporales se ha incrementado exponencialmente, el uso del machine learning en el análisis de series de tiempo ha emergido como una herramienta poderosa para extraer información valiosa y realizar predicciones precisas.

Esta guía es un emocionante viaje a través del mundo de las series de tiempo, con un enfoque particular en su intersección con el machine learning. Nos sumergimos en los conceptos fundamentales de las series de tiempo y exploramos cómo los algoritmos de machine learning pueden abordar los desafíos únicos que presentan. En esta travesía, utilizaremos un escenario práctico.

A lo largo de este artículo, combinaremos teoría y práctica para ilustrar cómo los algoritmos de machine learning pueden descifrar los secretos que se esconden en los datos temporales. Al final de nuestro viaje, habrás adquirido una comprensión sólida de cómo abordar problemas de series de tiempo con herramientas de machine learning y estarás listo para emprender tus propias investigaciones en el emocionante campo de las predicciones temporales.

¡Prepárate para un emocionante recorrido a través del mundo de las series de tiempo y el machine learning!

**¡Vamos con todo!**



## Tabla de contenidos

<b>Guía de estudio - Series de Tiempo</b>	<b>1</b>
¿En qué consiste esta guía?	1
Tabla de contenidos	2
<b>Categoría de datos</b>	<b>2</b>
Cross-sectional	3
Time Series	4
Panel	4
<b>Series de Tiempo</b>	<b>5</b>
Aplicaciones de las series temporales	6
Manejo de datos de tiempo con python	7
Actividad guiada: Manejo de Datos con python	7
<b>Componentes de series de tiempo</b>	<b>8</b>
Actividad guiada: Identificando las componentes en python	9
<b>Modelos de Series de Tiempo</b>	<b>11</b>
ACF y PACF	11
Modelos Autorregresivos (AR)	12
Ventajas y Desventajas	13
Modelos de Medias Móviles (MA)	13
Ventajas y Desventajas	14
Test de Dickey-Fuller	15
ARIMA	16
Actividad guiada: Utilizando los modelos de series de tiempo	17
Preguntas de proceso	17
Referencias bibliográficas	18



**¡Comencemos!**

## Categoría de datos

En el mundo del análisis de datos, diferentes tipos de datos nos brindan información valiosa y revelan patrones únicos en distintos contextos. Vamos a explorar tres categorías fundamentales de datos: **Cross-sectional data**, **Time series data** y **Panel data**.

### Cross-sectional

Los datos cross-sectional, también conocidos como datos de corte transversal o datos de sección transversal, son un tipo de datos en los que se recopila información de múltiples entidades o individuos en un punto específico en el tiempo. En otras palabras, representan una "instantánea" de diferentes unidades de observación en un momento dado.

En un conjunto de datos cross-sectional, cada fila suele representar una entidad o individuo diferente, mientras que las columnas contienen diferentes variables o características medidas para cada entidad. Por ejemplo, un conjunto de datos cross-sectional podría incluir información sobre varios países y sus respectivas poblaciones, ingresos, tasas de crecimiento, etc., todos medidos en un año en particular.

Las características Principales de Cross-sectional Data son las siguientes:

1. **Momentáneo:** Los datos se recopilan en un punto específico en el tiempo.
2. **Múltiples Entidades:** Cada observación corresponde a una entidad distinta.
3. **Variables Múltiples:** Se recopilan datos sobre diferentes variables o características para cada entidad.
4. **Análisis Comparativo:** Se utilizan para comparar características o variables entre diferentes entidades en un mismo punto temporal.

	País	Población (millones)	PIB (\$ billones)
0	EE. UU.	330	21.43
1	China	1440	14.14
2	India	1380	2.87
3	Brasil	211	2.05
4	Rusia	144	1.71

Imagen 1. Cross Sectional Data  
Fuente: Desafío Latam

En resumen, los datos cross-sectional nos proporcionan una instantánea de diferentes entidades en un momento específico, permitiéndonos analizar y comparar características entre ellas. Estos datos son esenciales para comprender las diferencias y similitudes entre grupos y realizar análisis comparativos en una variedad de campos.

## Time Series

Los datos de series temporales son un tipo de datos en el que se recopilan observaciones secuenciales en intervalos de tiempo regulares. Estos intervalos de tiempo pueden ser segundos, minutos, horas, días, meses, años, o cualquier otra unidad de tiempo, dependiendo del contexto. Los datos de series temporales capturan patrones y tendencias que se desarrollan a lo largo del tiempo y son esenciales para el análisis y la predicción en diversas disciplinas.

Las características Principales de la data de Series de Tiempo son las siguientes:

1. **Secuencia Ordenada:** Las observaciones se recopilan en un orden cronológico, formando una secuencia temporal.
2. **Intervalos de Tiempo:** Los datos se registran en intervalos de tiempo regulares.
3. **Tendencias y Patrones:** Los datos a menudo exhiben patrones recurrentes, estacionales o de tendencia.
4. **Predicción:** Se utilizan para predecir valores futuros basados en patrones históricos.

	Fecha	Temperatura (°C)
0	2023-01-01	18
1	2023-01-02	19
2	2023-01-03	17
3	2023-01-04	20
4	2023-01-05	21

Imagen 2. Time Series Data  
Fuente: Desafío Latam

Los datos de series temporales nos permiten explorar y comprender cómo las variables cambian a lo largo del tiempo. Capturan patrones, tendencias y ciclos que son fundamentales para el análisis y la toma de decisiones en diversas industrias. Al comprender las características y las aplicaciones de los datos de series temporales, estamos mejor preparados para abordar problemas complejos y aprovechar las oportunidades de predicción y análisis que ofrecen.

## Panel

La data de panel es un tipo de datos utilizado en estadísticas y análisis económico que combina las características de los datos cross-sectional y los datos de series temporales. En un conjunto de datos de panel, se recopila información de múltiples entidades o individuos a lo largo de varios períodos de tiempo. Esto permite analizar la evolución de estas entidades en función de variables que varían tanto entre las entidades como a lo largo del tiempo.

Las características principales de los paneles son las siguientes:

1. **Entidades y Tiempo:** Los datos de panel contienen observaciones de múltiples entidades en diferentes momentos temporales.
2. **Evolución Temporal:** Permiten analizar cómo las características de las entidades cambian a lo largo del tiempo.
3. **Dimensionalidad:** La combinación de entidades y tiempo puede generar un alto volumen de datos.
4. **Análisis Longitudinal:** Se utilizan para estudiar relaciones a lo largo del tiempo y entre las entidades.

	País	Año	PIB (\$ billones)
0	EE. UU.	2020	21.43
1	EE. UU.	2021	22.67
2	China	2020	14.14
3	China	2021	15.42
4	India	2020	2.87
5	India	2021	3.04

Imagen 3. Panel Data  
Fuente: Desafío Latam

## Series de Tiempo

Las series de tiempo son conjuntos de datos que se recopilan y registran en intervalos regulares a lo largo del tiempo. A diferencia de los datos estáticos, que se recolectan en un solo punto en el tiempo, las series de tiempo capturan la evolución y variación de una variable a medida que pasa el tiempo. Esto las convierte en una herramienta poderosa para el análisis y la predicción en una amplia gama de industrias y campos.

Estos son registros cronológicos de observaciones de una variable o fenómeno a lo largo del tiempo. Estas observaciones pueden ser tomadas en intervalos fijos, como días, meses o años, o incluso en intervalos irregulares según el contexto, donde es necesario realizar agregaciones regulares para poder analizarlas con técnicas de series de tiempo. Cada punto en la serie de tiempo está asociado con un momento específico en el tiempo, lo que permite analizar patrones, tendencias y comportamientos a lo largo de períodos temporales.

Las series de tiempo son fundamentales en diversas industrias y campos. En finanzas, se utilizan para predecir precios de acciones y tendencias del mercado. En meteorología, se utilizan para pronosticar el clima. En manufactura, se pueden usar para prever la demanda de productos. En medicina, las series de tiempo pueden ayudar a monitorear la evolución de enfermedades. Estos son solo algunos ejemplos de cómo las series de tiempo pueden brindar información valiosa para la toma de decisiones y la planificación estratégica.

A diferencia de los datos estáticos o transversales, las series de tiempo tienen un componente temporal inherente. Esto significa que los valores de la variable en una serie de

tiempo están vinculados a momentos específicos en el tiempo. Esta característica agrega complejidad y desafíos únicos al análisis de datos. Además, las series de tiempo a menudo exhiben patrones estacionales y tendencias que deben identificarse y modelarse adecuadamente.

Trabajar con datos temporales tiene sus ventajas y desafíos. Las ventajas incluyen la capacidad de identificar patrones a lo largo del tiempo, como estacionalidades y tendencias, que pueden proporcionar información valiosa para la toma de decisiones. Sin embargo, también presenta desafíos, como la variabilidad temporal y la necesidad de manejar adecuadamente los vacíos en los datos.

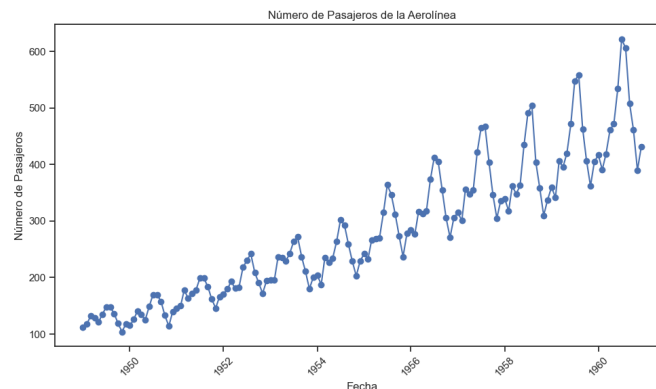


Imagen 4. Series Temporales

Fuente: Desafío Latam

## Aplicaciones de las series temporales

Las series de tiempo se utilizan en diversas áreas:

- **Economía y Finanzas:** Análisis de fluctuaciones en los precios de acciones, tasas de interés y datos económicos.
- **Meteorología:** Pronósticos climáticos y seguimiento de patrones climáticos a lo largo del tiempo.
- **Tecnología:** Monitoreo de tráfico web, uso de aplicaciones y rendimiento de servidores.
- **Salud:** Seguimiento de ritmo cardíaco, mediciones de glucosa y análisis de datos médicos.
- **Planificación Empresarial:** Predicción de ventas, demanda de productos y rendimiento de inventario.

Las series de tiempo son un medio crucial para analizar cómo las variables cambian a lo largo del tiempo. Su estructura secuencial y la información histórica que contienen permiten identificar patrones, tendencias y variaciones estacionales. Al comprender las características y aplicaciones de las series de tiempo, los profesionales pueden tomar decisiones basadas en datos y anticipar cambios futuros en una amplia variedad de campos.

## Manejo de datos de tiempo con python

El manejo de datos de tiempo en Python es una habilidad fundamental para aquellos que deseen analizar y comprender patrones temporales en conjuntos de datos. La capacidad de trabajar con fechas y horarios de manera eficiente y precisa es esencial en diversas disciplinas, desde análisis financiero y científico hasta planificación y pronóstico. Python, con sus bibliotecas como **datetime** y **pandas**, ofrece herramientas robustas que permiten la manipulación, visualización y análisis de datos temporales. En este contexto, explicaremos cómo utilizar estas bibliotecas para crear, transformar y analizar datos de fechas y tiempo, abriendo la puerta a un mundo de insights y descubrimientos en el contexto temporal de los datos.



### Actividad guiada: Manejo de Datos con python

Abre el archivo 02 - Manejo de datos de tiempo para aprender a utilizar los comandos correspondientes.

La biblioteca **datetime** de Python facilita la gestión y manipulación de datos de fechas y tiempo. Puedes crear objetos de fecha y hora, formatearlos, realizar operaciones y trabajar con duraciones de tiempo. Además, puedes manejar zonas horarias para adaptar tus datos a diferentes contextos geográficos. Esta capacidad de manejar el tiempo de manera flexible es esencial para diversas aplicaciones, desde análisis de datos hasta desarrollo de aplicaciones.

## Componentes de series de tiempo

En el análisis de series de tiempo, entender y descomponer las distintas componentes que conforman los datos es esencial para comprender los patrones y tendencias subyacentes. En esta guía, explicaremos los componentes clave de las series de tiempo y cómo identificarlos, interpretarlos y visualizarlos utilizando Python.

Las series de tiempo se componen de tres componentes principales:

1. **Tendencia:** Representa la dirección general en la que los datos se están moviendo a lo largo del tiempo. Puede ser ascendente, descendente o incluso constante.
2. **Estacionalidad:** Se refiere a patrones que se repiten en intervalos regulares de tiempo, como estacionalidad mensual o anual. Estos patrones pueden ser causados por factores estacionales o cíclicos.
3. **Residuos (o ruido):** Son las fluctuaciones aleatorias que no pueden atribuirse a ninguna tendencia o estacionalidad específica. Representan la variabilidad no explicada por los otros componentes.

La identificación precisa de estos componentes es crucial para un análisis significativo de las series de tiempo. Algunas técnicas comunes para identificar componentes incluyen:

1. Visualización de la serie de tiempo para detectar tendencias y patrones estacionales.
2. Descomposición de series de tiempo en sus componentes utilizando métodos como el promedio móvil o la descomposición estacional.

La interpretación de los patrones estacionales y de tendencia puede proporcionar información valiosa sobre el comportamiento de los datos a lo largo del tiempo. Por ejemplo, una tendencia ascendente podría indicar un crecimiento constante en los datos, mientras que una estacionalidad anual podría sugerir comportamientos recurrentes en ciertos meses.





## Actividad guiada: Identificando las componentes en python

Para identificar las componentes de las series temporales vamos a utilizar la librería **statsmodels** que nos ayuda a descomponer una serie temporal en las 3 componentes.

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose
import seaborn as sns
sns.set(style="ticks")
```

Para esta etapa vamos a utilizar el dataset “flights” que posee los pasajeros aéreos en el tiempo.

```
# Cargar el conjunto de datos "Air Passengers" desde seaborn
df = sns.load_dataset("flights")

# Convertir la columna 'year' y 'month' a formato de fecha
df['date'] = pd.to_datetime(df['year'].astype(str) + '-' + df['month'].astype(str), errors='coerce')
df = df.drop(['year', 'month'], axis=1)
df = df.set_index('date')
```

Con ayuda del método `seasonal_decompose` de `statsmodels` uno puede descomponer la serie de tiempo en 3 componentes “trend”, “seasonal” y “resid”.

```
# Descomponer la serie de tiempo en componentes
result = seasonal_decompose(df, model='additive')
```

Ahora procedemos a graficar cada una de las componentes por separado

```
# Graficar los componentes
plt.figure(figsize=(10, 6))
plt.subplot(4, 1, 1)
plt.plot(df, label='Serie Original')
plt.legend()

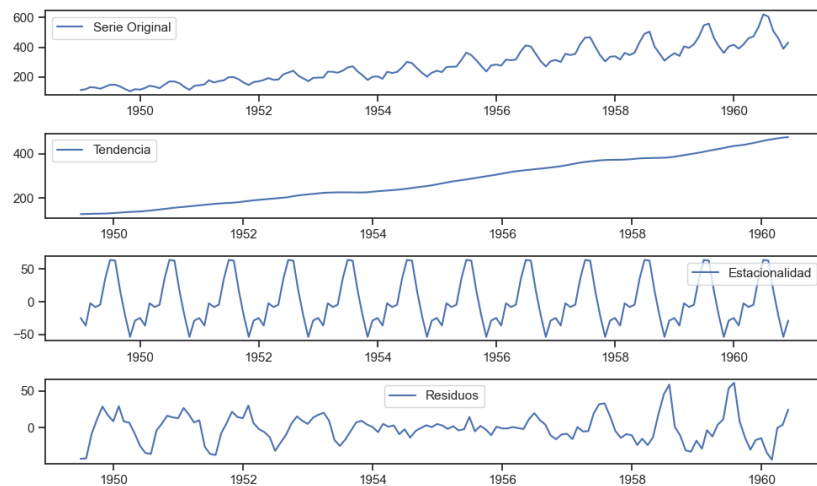
plt.subplot(4, 1, 2)
```

```
plt.plot(result.trend, label='Tendencia')
plt.legend()

plt.subplot(4, 1, 3)
plt.plot(result.seasonal, label='Estacionalidad')
plt.legend()

plt.subplot(4, 1, 4)
plt.plot(result.resid, label='Residuos')
plt.legend()

plt.tight_layout()
plt.show()
```



En resumen, comprender los componentes de las series de tiempo es esencial para realizar un análisis significativo. La identificación e interpretación de la tendencia, estacionalidad y residuos, junto con la visualización adecuada, nos permiten extraer información valiosa de los datos temporales y tomar decisiones informadas.

## Modelos de Series de Tiempo

En esta exploración de los modelos de series de tiempo, nos enfocaremos en dos enfoques esenciales: los Modelos Autorregresivos (AR) y los Modelos de Medias Móviles (MA). Estos modelos capturan patrones complejos y relaciones inherentes en los datos secuenciales, lo que los convierte en herramientas indispensables para la predicción y el análisis de tendencias. Al comprender cómo funcionan los modelos AR y MA, podrás adentrarte en la esencia de las series de tiempo y desbloquear su potencial para una amplia gama de aplicaciones.

Los Modelos Autorregresivos (AR) son una clase de modelos de series de tiempo que modelan la relación entre una observación actual y un conjunto de observaciones anteriores, también conocido como "lags". Estos modelos capturan las dependencias temporales intrínsecas, permitiéndonos predecir el valor futuro de una serie de tiempo en función de sus valores pasados. Por otro lado, los Modelos de Medias Móviles (MA) se centran en la relación entre una observación y un conjunto de valores pasados de un término de error estocástico. Estos modelos se centran en capturar la relación entre los errores observados en diferentes momentos en el tiempo.

### ACF y PACF

El análisis de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) es fundamental en el modelado de series de tiempo. Estas funciones nos proporcionan información sobre las correlaciones entre valores pasados y presentes de la serie, lo que es crucial para identificar la estructura de un modelo ARIMA.

La Función de Autocorrelación (ACF) muestra la correlación entre un valor en un punto de tiempo y los valores anteriores en intervalos de tiempo diferentes. Ayuda a identificar la cantidad de retrasos pasados que son relevantes para predecir el valor actual. Matemáticamente, la ACF se define como:

$$ACF(k) = \frac{Corr(X_t, X_{t-k})}{Var(X_t)}$$

Donde:

- $X_t$  es el valor en el punto de tiempo  $t$ .
- $X_{t-k}$  es el valor en el punto de tiempo  $t-k$ .

La Función de Autocorrelación Parcial (PACF) muestra la correlación directa entre un valor en un punto de tiempo y los valores anteriores en un solo intervalo de tiempo. La PACF ayuda a identificar la influencia directa de los valores pasados sin la interferencia de otros valores intermedios. Matemáticamente, la PACF se obtiene resolviendo las ecuaciones de regresión para los términos anteriores.

En resumen, el análisis de ACF y PACF son herramientas cruciales para comprender la estructura subyacente de una serie de tiempo y ayudan en la selección de un modelo adecuado para el análisis y la predicción:

- Identificar la naturaleza de la serie de tiempo: Si la ACF decae rápidamente o tiene picos, y si la PACF muestra correlaciones significativas, podemos inferir si la serie es estacionaria o si necesita diferenciación.
- Determinar los valores de p, d y q en un modelo ARIMA: A partir de los patrones en ACF y PACF, podemos obtener pistas sobre los términos de autorregresión (AR) y de media móvil (MA) en el modelo ARIMA (que veremos más adelante)

## Modelos Autorregresivos (AR)

Los modelos autorregresivos (AR, por sus siglas en inglés Autoregressive Models) son una poderosa herramienta en el análisis de series de tiempo. Estos modelos se basan en la idea fundamental de que los valores futuros de una serie de tiempo pueden predecirse mediante una combinación lineal de sus valores pasados, conocidos como "lags". Esta característica distintiva hace que los modelos autorregresivos sean adecuados para capturar patrones temporales y tendencias en los datos a lo largo del tiempo. En esta guía, explicaremos en detalle qué son los modelos autorregresivos, cómo funcionan, cuándo se utilizan, y sus ventajas y desventajas.

Los modelos autorregresivos son una clase de modelos de regresión en la que la variable dependiente se regresa a sí misma en uno o más períodos de tiempo anteriores. Formalmente, un modelo AR de orden "p" se define como:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

Donde:

$y_t$  Es el valor de la serie de tiempo en t.

c es la constante.

$\phi_1, \phi_2, \dots, \phi_p$  son los coeficientes de autorregresión.

$y_{t-1}, y_{t-2}, \dots, y_{t-p}$  son los valores anteriores de la serie de tiempo.

$\epsilon_t$  Es el término de error en el período t.

El funcionamiento de un modelo autorregresivo implica estimar los coeficientes de autoregresión ( $\phi_1, \phi_2, \dots, \phi_p$ ) a partir de los datos de la serie de tiempo. Estos coeficientes capturan cómo los valores anteriores de la serie influyen en el valor actual. Una vez que se han estimado los coeficientes, el modelo se puede utilizar para predecir valores futuros basados en los valores pasados de la serie de tiempo.

## Ventajas y Desventajas

Los modelos autorregresivos encuentran aplicaciones en una amplia gama de campos, como economía, finanzas, meteorología y más. Se utilizan para pronosticar valores futuros de una serie de tiempo, lo que permite tomar decisiones informadas y planificar estratégicamente.

Las ventajas de los modelos autorregresivos incluyen:

1. **Captura de Tendencias Temporales:** Los AR son efectivos para capturar patrones y tendencias temporales en los datos, lo que los hace útiles para pronósticos a corto y mediano plazo.
2. **Simplicidad:** La interpretación de los coeficientes es directa y proporciona información sobre cómo los valores pasados afectan los valores futuros.
3. **Flexibilidad:** Los modelos AR pueden ser utilizados en combinación con otros modelos de series de tiempo para lograr pronósticos más precisos.

A pesar de sus ventajas, los modelos autorregresivos también presentan algunas desventajas:

1. **Sensibles a Outliers:** Los AR son sensibles a valores atípicos (outliers) en los datos, lo que puede afectar la calidad de las predicciones.
2. **Limitaciones en Tendencias no Lineales:** Los AR asumen relaciones lineales entre los valores pasados y futuros, lo que puede ser limitante en casos de tendencias no lineales.

En resumen, los modelos autorregresivos son una herramienta esencial para el análisis y la predicción de series de tiempo. Al comprender cómo funcionan, sus aplicaciones y sus pros y contras, estarás mejor preparado para aplicarlos de manera efectiva en situaciones del mundo real y tomar decisiones basadas en datos sólidos y proyecciones precisas.

## Modelos de Medias Móviles (MA)

Los modelos de medias móviles (MA, por sus siglas en inglés Moving Average Models) son una parte esencial del análisis de series de tiempo. Estos modelos son especialmente útiles para capturar patrones temporales y fluctuaciones aleatorias en los datos. En esta guía, explicaremos en detalle qué son los modelos de medias móviles, cómo funcionan, cuándo se utilizan, y sus ventajas y desventajas.

Los modelos de medias móviles son una clase de modelos de regresión en la que la variable dependiente se relaciona con los errores pasados de predicción y los términos de error actuales. Un modelo MA de orden "q" se define como:

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Donde:

$y_t$  Es el valor de la serie de tiempo en t.

c es la constante.

$\theta_1, \theta_2, \dots, \theta_q$  son los coeficientes de autorregresión.

$\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$  son los valores anteriores de la serie de tiempo.

$\epsilon_t$  Es el término de error en el período t.

Los modelos MA se basan en la idea de que los errores pasados de predicción pueden influir en el valor actual de la serie de tiempo. Estos modelos estiman los coeficientes de la media móvil ( $\theta_1, \theta_2, \dots, \theta_q$ ) a partir de los errores pasados y presentes de la serie. Una vez que se han estimado los coeficientes, el modelo MA se utiliza para hacer predicciones futuras.

## Ventajas y Desventajas

Los modelos de medias móviles encuentran aplicaciones en diversas áreas, como economía, finanzas, demanda de productos y más. Se utilizan para pronosticar valores futuros de una serie de tiempo, especialmente cuando se sospecha que los valores pasados de errores pueden influir en los valores futuros.

Las ventajas de los modelos MA incluyen:

1. **Captura de Fluctuaciones Aleatorias:** Los MA son efectivos para capturar las fluctuaciones aleatorias o ruido en los datos de la serie de tiempo.
2. **Simplicidad e Interpretación:** La interpretación de los coeficientes es directa y proporciona información sobre cómo los errores pasados afectan los valores futuros.

3. **Incorporación de Información Pasada:** Los modelos MA incorporan información pasada en la predicción, lo que puede mejorar la precisión en comparación con simples promedios móviles.

A pesar de sus ventajas, los modelos de medias móviles también presentan algunas desventajas:

1. **No Captura Tendencias Temporales:** Los MA no son adecuados para capturar tendencias a largo plazo en los datos, ya que se centran principalmente en las fluctuaciones aleatorias.
2. **Sensibilidad a Valores Atípicos:** Los valores atípicos pueden afectar significativamente las predicciones de los modelos MA.

En resumen, los modelos de medias móviles son una herramienta valiosa en el análisis y pronóstico de series de tiempo. Al comprender cómo funcionan, sus aplicaciones y sus ventajas y desventajas, estarás mejor preparado para utilizarlos de manera efectiva en situaciones del mundo real y obtener proyecciones más precisas y confiables.

## Test de Dickey-Fuller

El Test de Dickey-Fuller es una herramienta estadística utilizada para determinar si una serie de tiempo es estacionaria o no. Fue desarrollado por los economistas David Dickey y Wayne Fuller. La estacionaridad es un concepto clave en el análisis de series de tiempo, ya que muchas técnicas y modelos asumen que los datos son estacionarios o se pueden transformar en estacionarios.

La idea detrás del Test de Dickey-Fuller se basa en la hipótesis nula de que la serie de tiempo tiene una raíz unitaria, lo que significa que tiene alguna forma de dependencia temporal y no es estacionaria. La hipótesis alternativa es que la serie de tiempo no tiene raíz unitaria y, por lo tanto, es estacionaria.

El test se realiza calculando una estadística conocida como el "ADF Statistic" (Augmented Dickey-Fuller Statistic) a partir de la serie de tiempo. Esta estadística compara la magnitud de la raíz unitaria con la variabilidad de la serie de tiempo. Cuanto más negativa sea la estadística, más fuerte es la evidencia en contra de la presencia de una raíz unitaria y, por lo tanto, en favor de la estacionaridad.

El Test de Dickey-Fuller también proporciona un "p-value", que se utiliza para evaluar la significancia estadística del resultado. Si el "p-value" es menor que un umbral predefinido (como 0.05), se rechaza la hipótesis nula de que la serie de tiempo es no estacionaria.

El proceso del Test de Dickey-Fuller implica los siguientes pasos:

1. Calcular la estadística de prueba: Se calcula el ADF Statistic utilizando una fórmula que involucra los coeficientes de regresión en una ecuación que modela la serie de tiempo.
2. Comparación con valores críticos: La estadística de prueba se compara con valores críticos de la distribución para determinar si es significativa.
3. Cálculo del p-value: Se calcula el "p-value" asociado con la estadística de prueba.
4. Toma de decisión: Si el "p-value" es menor que el umbral predefinido, se rechaza la hipótesis nula, lo que sugiere que la serie de tiempo es estacionaria. Si el "p-value" es mayor, no se rechaza la hipótesis nula y la serie de tiempo se considera no estacionaria.

En resumen, el Test de Dickey-Fuller es una herramienta fundamental para evaluar si una serie de tiempo es estacionaria o no. La estacionaridad es un concepto crucial en el análisis de series de tiempo y es esencial para elegir y aplicar modelos adecuados.

## ARIMA

El modelo ARIMA (Autoregressive Integrated Moving Average) es una técnica ampliamente utilizada en el análisis de series de tiempo para modelar y predecir datos secuenciales. ARIMA combina los componentes de autorregresión (AR) y promedio móvil (MA) junto con la diferenciación (I) para manejar patrones estacionales y de tendencia en los datos.

La teoría detrás de ARIMA se basa en tres componentes principales:

1. Autorregresión (AR): Este componente captura la relación entre un valor actual y sus valores anteriores. Un modelo AR(p) utiliza los p valores anteriores para predecir el valor actual.
2. Media Móvil (MA): Este componente modela la relación entre el valor actual y los errores pasados. Un modelo MA(q) utiliza los errores anteriores para predecir el valor actual.
3. Diferenciación (I): La diferenciación se aplica para hacer que la serie de tiempo sea estacionaria, es decir, para eliminar patrones de tendencia y estacionalidad. Una diferenciación de grado d implica restar la serie de tiempo actual de la serie de tiempo retrasada d veces.

La selección de los parámetros p, d y q en un modelo ARIMA se basa en el análisis de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) de la serie de tiempo. La ACF y la PACF proporcionan pistas sobre los valores de p y q respectivamente,



mientras que  $d$  se determina observando la necesidad de diferenciación para hacer la serie estacionaria.

En resumen, ARIMA es un modelo poderoso para capturar patrones complejos en series de tiempo. Combina autorregresión, promedio móvil y diferenciación para modelar tendencias, estacionalidades y otros patrones en los datos. La selección de los parámetros  $p$ ,  $d$  y  $q$  se basa en el análisis de ACF y PACF, y estos parámetros ayudan a definir la estructura del modelo ARIMA.



## Actividad guiada: Utilizando los modelos de series de tiempo

Realizaremos un análisis completo de series de tiempo, desde un dataset de pasajeros, para estimar la cantidad de pasajeros de los próximos periodos. Para este ejercicio vamos a utilizar la librería `pmdarima` y vamos a utilizar el método `auto_arima`.

Analiza el desarrollo propuesto en el archivo 02 - Aplicación

Prueba con otro set de datos que te interese.



**¡Lo lograste! / ¡Felicitaciones!**

Para hacer énfasis en el logro de un aprendizaje u objetivo planteado en la guía

## Preguntas de proceso

### Reflexiona:

- ¿Qué es una serie de tiempo y cómo se diferencia de otros tipos de datos?
- ¿Cuáles son las tres componentes principales de una serie de tiempo y cómo se identifican?
- ¿Cuál es la diferencia entre una serie de tiempo estacionaria y no estacionaria?
- ¿Qué significa que una serie de tiempo tenga una tendencia? ¿Cómo se puede tratar la tendencia en el análisis?
- ¿Qué es un modelo AR (Autorregresivo)? ¿Cómo se construye un modelo AR y qué tipo de patrones captura?
- ¿Cuál es la principal característica de un modelo MA (Media Móvil)? ¿Cómo se diferencia de un modelo AR?
- ¿Qué es un modelo ARIMA y cuándo se utiliza en el análisis de series de tiempo?
- ¿Cómo se seleccionan los valores de p, q y d en un modelo ARIMA?
- ¿Cuáles son las ventajas y desventajas de los modelos ARIMA en comparación con otros enfoques de series de tiempo?



## Referencias bibliográficas

1. Theodore Petrou - Pandas Cookbook\_ Recipes for Scientific Computing, Time Series Analysis and Data Visualization using Python-Packt Publishing (2017)
2. Avishek Pal, PKS Prakash - Practical time series analysis\_ master time series data processing, visualization, and modeling using Python-Packt Publishing (2017)
3. Python Data Science Handbook:  
<https://jakevdp.github.io/PythonDataScienceHandbook>
4. [https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal\\_decompose.html](https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html)
5. <https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima.model.ARIMA.html>



**¡Continúa aprendiendo y practicando!**