



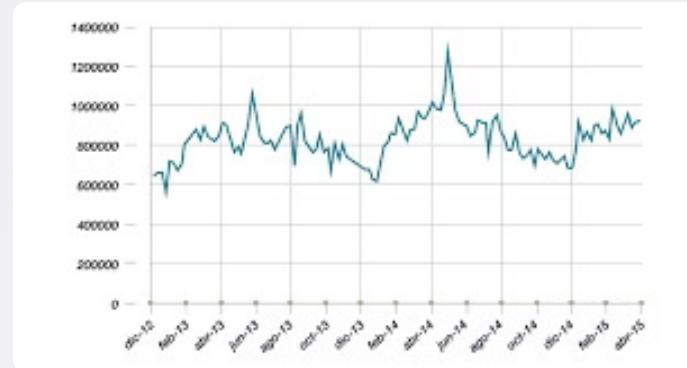
# Series Temporales en Data Science

Fundamentos, Modelos y Forecasting para el Magíster

# La Esencia de los Datos Temporales

¿Qué hace que un dato temporal sea distinto a uno tabular cualquiera? La clave reside en el orden. A diferencia de los datos tabulares aislados, en una serie temporal, lo que ocurrió ayer impacta directamente en lo que sucede hoy.

No son eventos aislados; son parte de una historia continua. Comprender esta interdependencia es fundamental para un análisis preciso y predicciones robustas.



Una serie de tiempo es un conjunto de observaciones tomadas de forma secuencial en el tiempo. Ej. Volumen diario de ventas de un producto, donde la observación es la cantidad de producto vendido por día.



# Adquisición y Reorganización de Datos

El primer paso crítico es preparar la serie temporal para el análisis. Una correcta adquisición y reorganización es vital para asegurar la validez de los modelos futuros.

## Formato de Fecha/Hora

Asegúrese de que el índice o una columna dedicada contenga el formato de fecha y hora correcto. Esto permite una interpretación inequívoca del tiempo.

## Orden Cronológico

Los datos deben estar estrictamente ordenados de manera cronológica ascendente. Un desorden en la secuencia anularía las propiedades temporales de la serie.

## Definición de Frecuencia

Establezca una frecuencia clara (diaria, semanal, mensual, etc.). Esto ayuda al software a entender la temporalidad de sus observaciones y a identificar posibles lagunas.

# Gestión de Datos Faltantes y Consistencia

En las series temporales, los huecos en los datos son comunes. Rellenarlos de forma inteligente es crucial para mantener la continuidad de la serie sin introducir sesgos.

## Regla Práctica: Relleno Sensato

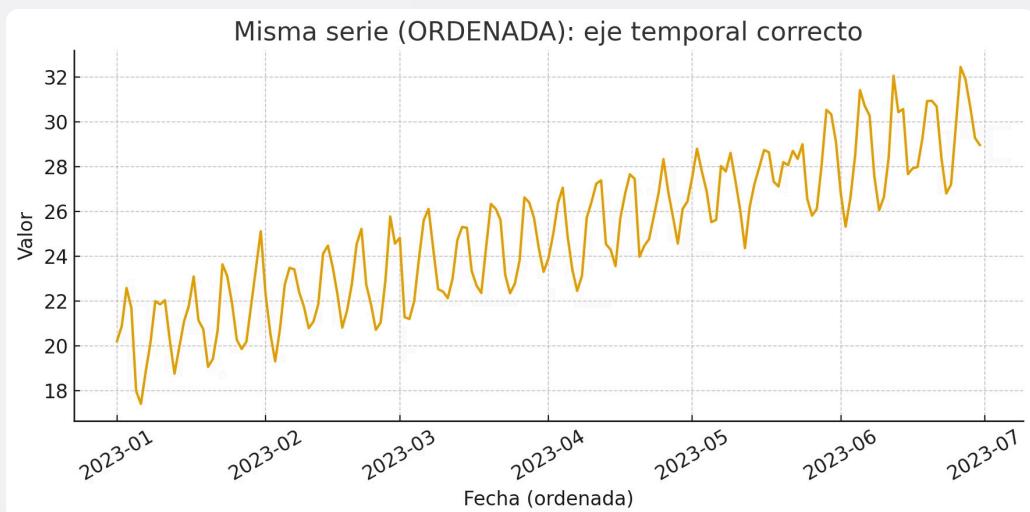
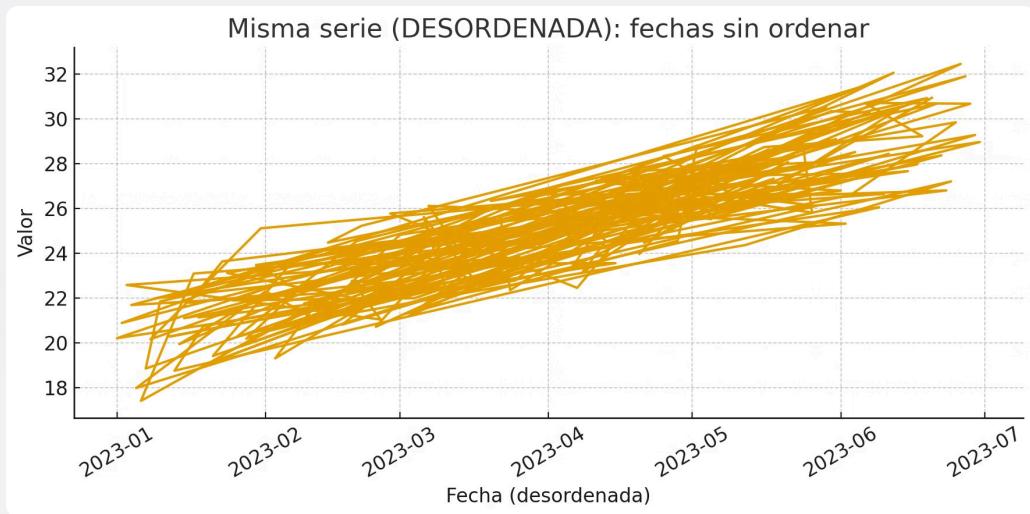
- Utilice una única columna para la fecha/hora.
- Si un valor falta, rellénelo con un método coherente con el contexto de los datos (imputación)
- El **Forward Fill** (usar el último valor conocido) es útil cuando el valor se mantiene constante hasta la próxima observación (ej: precio de una acción).
- Eliminación de valores extremos (depende del problema).

### Ejemplos:

- **Tipo de cambio USD/CLP diario:** Asegure que cada fila sea un día con fecha como índice.
- **Ocupación UCI semanal:** Si un hospital no reporta, decida cómo manejar ese hueco, quizás con interpolación o forward fill.

### ⓘ Mini-actividad

Proyecte dos gráficos de la misma serie: uno desordenado y otro ordenado con eje temporal.  
Pregunte: ¿Cuál está "bien" y por qué?





# Análisis Exploratorio de Datos (EDA) para Series Temporales

La visualización es su mejor aliado para identificar patrones fundamentales en una serie temporal. Debemos entender el comportamiento de la señal, detectar datos faltantes o datos extremos.

Busque estos **componentes** clave:



## Tendencia

Movimiento a largo plazo de la serie, ya sea ascendente, descendente o constante. Indica la dirección general del fenómeno.



## Estacionalidad

Patrones que se repiten en intervalos fijos y predecibles (ej. diaria, semanal, anual). Refleja influencias cíclicas.



## Cambios de Nivel

Variaciones abruptas en el valor medio de la serie, a menudo debido a eventos externos (ej. nueva normativa, crisis económica).

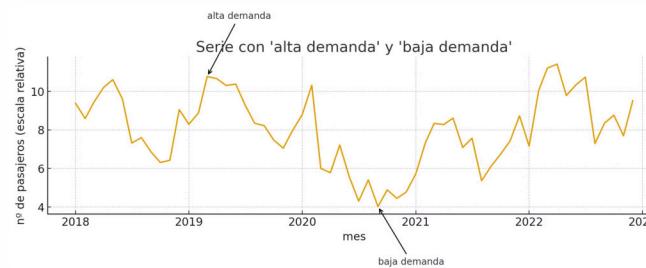


## Valores Atípicos (Outliers)

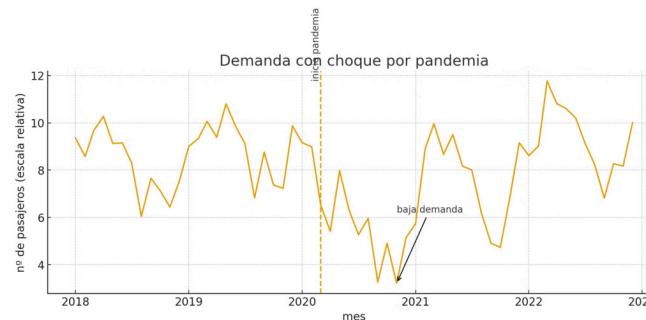
Observaciones que se desvían significativamente del patrón general de la serie. Pueden indicar errores o eventos inusuales.

# Ejemplos

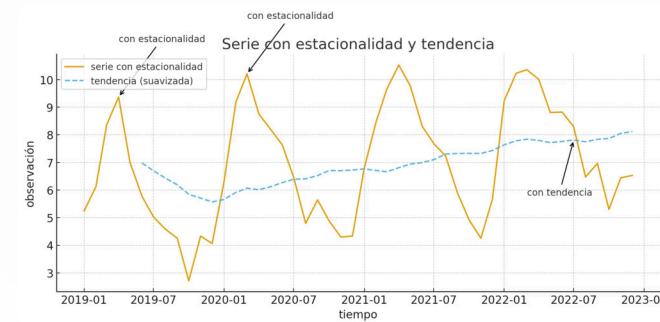
## Serie temporal de pasajeros



Comportamiento: alta y baja demanda,  
¿Hay un contexto?



## Estacionalidad y tendencia



Las series temporales estacionarias es en la cual sus propiedades estadísticas se mantienen sin cambios en el tiempo.

## Datos faltantes



# EDA Avanzado: Descomposición y Transformación

Para entender a fondo una serie temporal, a menudo es necesario descomponerla en sus componentes fundamentales. Esto nos permite aislar y analizar el impacto de cada elemento sobre el comportamiento general de los datos.

## Componentes Clave

Tendencia: Cambio de nivel a largo plazo.

Estacionalidad: Patrón que se repite con una frecuencia fija (diaria, semanal, anual).

1

Ciclo: Fluctuaciones más lentas que la estacionalidad, no necesariamente periódicas.

Eventos: Impactos puntuales (feriados, campañas, shocks).

Ruido: Lo impredecible que queda.

## Detección en la Práctica

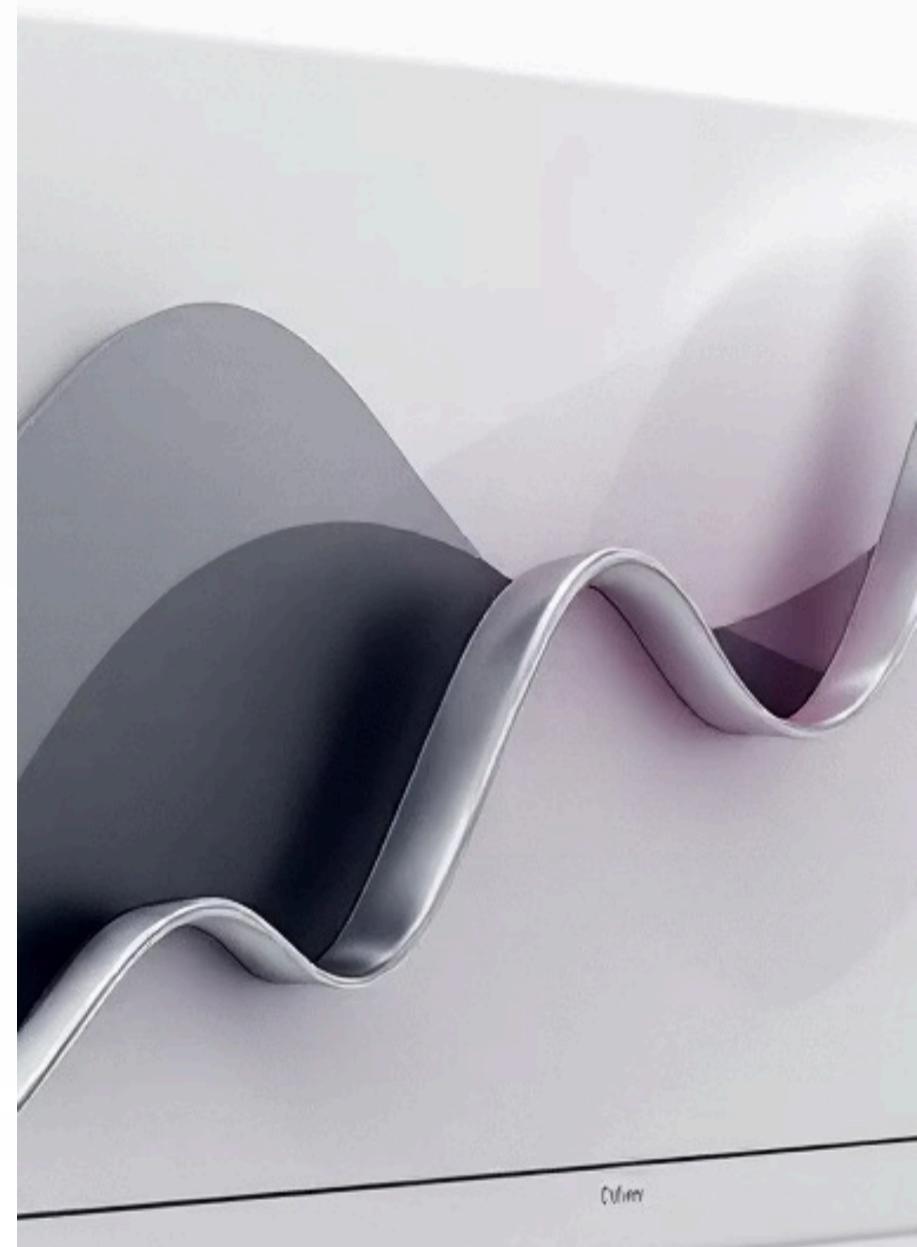
Tendencia: Visualice la serie con una media móvil para observar derivas sostenidas.

2

Estacionalidad: Compare promedios por período (mes, día de semana) o use gráficos de autocorrelación.

Eventos: Identifique picos o caídas en fechas específicas y utilice variables dummy.

Ruido sano: Tras remover los otros componentes, lo que resta debe oscilar aleatoriamente alrededor de cero.



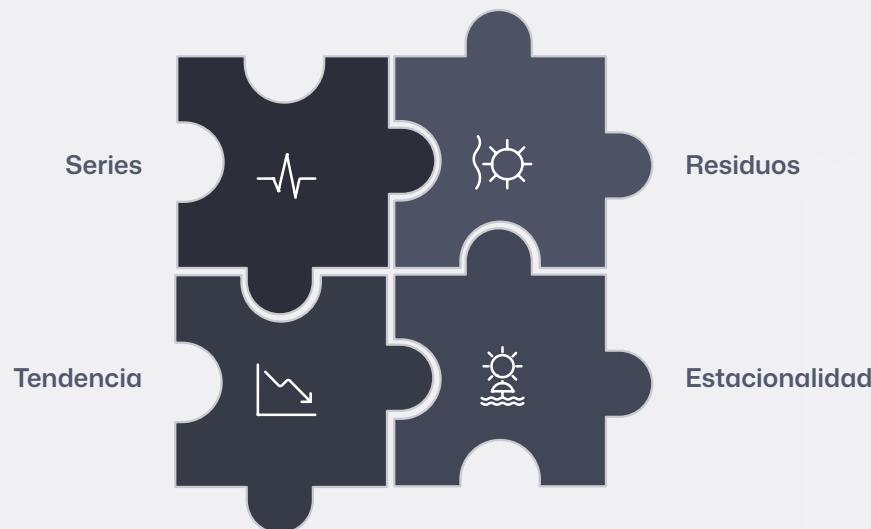
## Estacionariedad: Un Pilar Fundamental

Una serie es "estacionaria" cuando sus propiedades estadísticas (media, varianza y estructura de autocorrelación) permanecen constantes a lo largo del tiempo. Es decir, su comportamiento es predecible y no depende del momento en que se observa.



# EDA Avanzado: Descomposición y Transformación

Para profundizar en el EDA, la descomposición es una herramienta poderosa que permite aislar los componentes de una serie temporal.



Esta técnica separa la serie en:

- **Tendencia:** Línea suave que muestra el movimiento principal.
- **Estacionalidad:** El patrón repetitivo.
- **Residuo:** El "ruido" o la parte no explicada por los dos anteriores.

## Regla Práctica: Graficar Cambios

Primero, visualice la serie en su nivel original para una vista general. Luego, analice los cambios (diferencias de un punto a otro).

Muchos problemas evidentes en el nivel, como la tendencia, se suavizan al observar los cambios, facilitando la detección de otros patrones.

### Ejemplos:

- **Demanda eléctrica:** Observe picos diarios y estacionales (invierno/verano), con una tendencia a largo plazo.
- **Ventas retail:** Identifique patrones anuales (Navidad) y semanales (fines de semana).

**Mini-actividad:** Muestre un gráfico de ventas de helado con tendencia y estacionalidad. Que los alumnos identifiquen visualmente cada componente.

# Estacionariedad: Un Pilar Fundamental

Una serie temporal es **estacionaria** si sus propiedades estadísticas, como la media y la variabilidad, no cambian con el tiempo. El "juego" que describe la serie permanece constante.



## Por Qué Importa

La mayoría de los modelos clásicos de series de tiempo asumen la estacionariedad. Si la serie no es estacionaria, los resultados del modelo no serán confiables ni válidos.



## Regla Práctica

Diferencie la serie lo mínimo necesario para que sea estacionaria. Una sola diferenciación suele ser suficiente para eliminar la tendencia.

## ¿Cómo la sospecho?

- Si la serie "se arrastra" hacia arriba o hacia abajo, no es estacionaria.
- Si los cambios (diferencias) "vibran" alrededor de un valor fijo (como cero), sí lo es.

# Autocorrelación y Correlaciones Espurias

Comprender cómo una serie se relaciona consigo misma en el pasado es crucial, al igual que evitar conclusiones engañosas.

## Autocorrelación

Es la correlación de una serie consigo misma en momentos anteriores. Se mide con el **gráfico de ACF** (Función de Autocorrelación). Es útil para identificar patrones de estacionalidad o tendencias lineales. Una barra alta en el día 7, por ejemplo, sugiere estacionalidad semanal.



## Correlaciones Espurias

Dos series con tendencia a menudo parecen correlacionar, aunque no tengan relación causal alguna. Piense en las ventas de helados y el número de incendios: ambas aumentan en verano, lo que genera una correlación "falsa".



**Regla Práctica:** Nunca correlacione los niveles de dos series con tendencia. En su lugar, compare sus cambios (las diferencias) para inferir relaciones válidas.

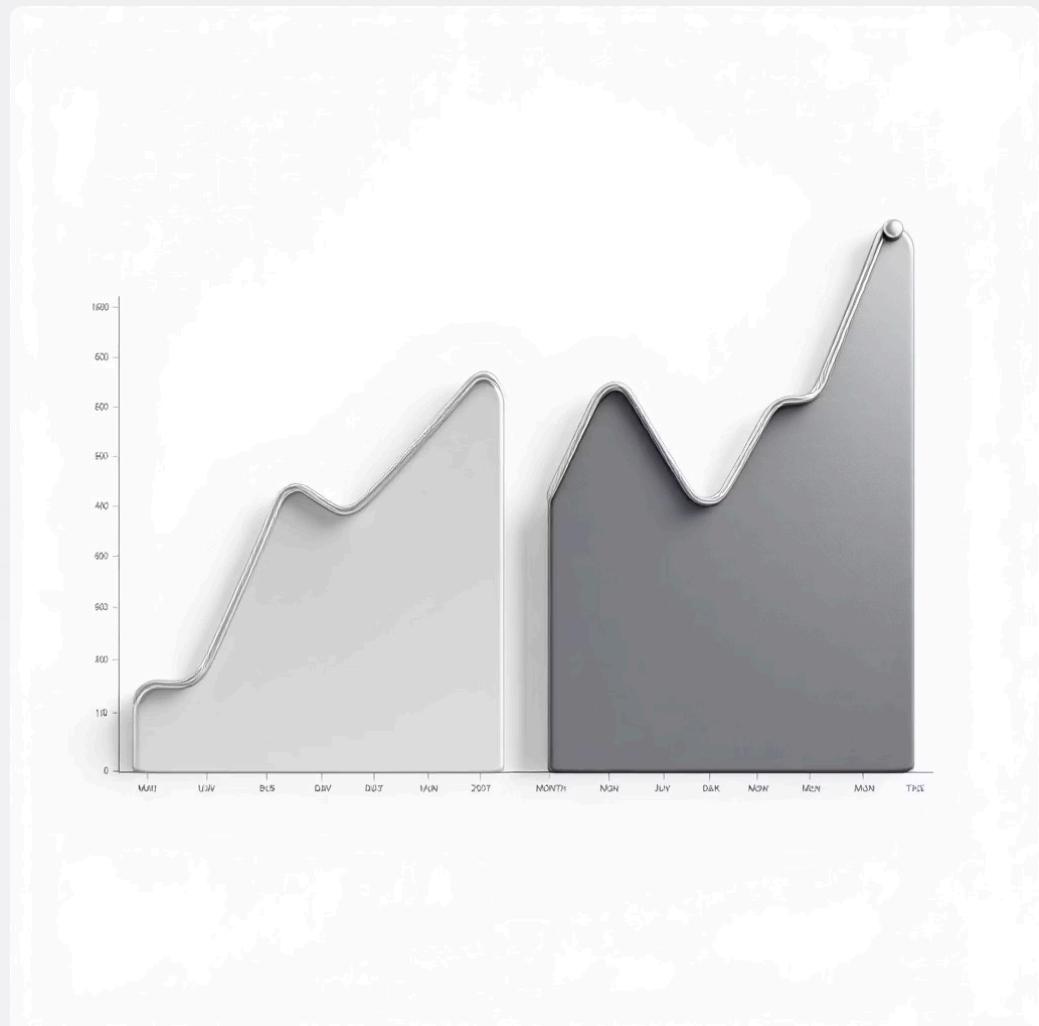
# Evitando Correlaciones Espurias

Aunque dos series temporales muestren tendencias similares, esto **no significa que exista una relación causal entre ellas**. La correlación en los **niveles brutos** puede ser engañosa, llevando a conclusiones erróneas si no se maneja correctamente.

## La Solución: Análisis de Cambios

Para inferir relaciones válidas, es crucial observar los **cambios** o **diferencias** entre un período y otro. Pregúntese: “*¿cuánto subió o bajó esta serie respecto al período anterior?*” Si los cambios de ambas series se mueven juntos de manera consistente, es más probable que exista una relación real.

Considere el ejemplo de las ventas de helados y el nivel del mar: ambas series pueden mostrar una tendencia ascendente a lo largo de los años. Sin embargo, no hay una relación directa o causal. **Si analizamos los cambios mes a mes, es decir, cuánto suben o bajan las ventas de helados y el nivel del mar en cada mes, esa "relación" espuria desaparecerá**, revelando la verdadera independencia entre ellas.



## Pasos Clave para Evitar Engaños

01

### Identifique Tendencias

Si una o ambas series presentan una clara tendencia a lo largo del tiempo, evite correlacionar directamente sus niveles brutos.

02

### Calcule las Diferencias

Transforme las series calculando sus diferencias (por ejemplo, el cambio de un período al siguiente) o retornos porcentuales. Esto ayuda a eliminar la tendencia.

03

### Correlacione los Cambios

Una vez que las series son estacionarias (sin tendencia), correlacione sus cambios. Esto le dará una medida más confiable de su relación.

# Modelos Clásicos de Series Temporales

Estos modelos están diseñados específicamente para capturar la naturaleza temporal de los datos.



## AR (Auto-Regresivo)

"Hoy es una función de lo que pasó en los días recientes".  
Útil cuando la serie tiene una "memoria" a corto plazo.



## ARIMA (Auto-Regresivo Integrado de Promedio Móvil)

La versión más popular. La "I" indica que trabaja sobre las diferencias, permitiendo manejar series con tendencia.  
Versátil y fundamental.



## SARIMA (Seasonal ARIMA)

Extiende ARIMA para incluir un componente estacional ("S").  
Ideal para series con patrones anuales o semanales, como ventas retail.



## VAR (Vector Auto-Regresivo)

Modela varias series juntas, asumiendo que se influyen mutuamente. Útil cuando múltiples variables se mueven interconectadas (ej. demanda eléctrica y temperatura).

## ¿Por qué la Regresión Lineal Clásica Falla?

La regresión lineal asume errores independientes, lo cual no se cumple en series temporales (los errores están autocorrelacionados). Esto invalida los p-values y da una falsa confianza. Modelos que "siguen" la tendencia pueden tener un  $R^2$  alto, pero sin entender la relación real. **Mensaje:** Use modelos que reconozcan el orden temporal o transforme la serie.

**Diagnóstico simple:** Los residuos del modelo (predicción vs. real) deben ser ruido puro. Si aún muestran patrones, el modelo no capturó toda la información.

# Forecasting y Validación Robusta

Para validar un modelo de series temporales, debe simular la realidad: **validar "mirando hacia adelante"**.

01

## Holdout Temporal

El método más simple: entrena con datos hasta una fecha y evalúa con los datos posteriores (ej. entrena hasta enero de 2024, predice febrero-marzo).

02

## Backtesting / Rolling Validation

La mejor práctica: entrena con una ventana de historial y predice el siguiente período. Luego, desliza la ventana y repite, simulando una operación real.

03

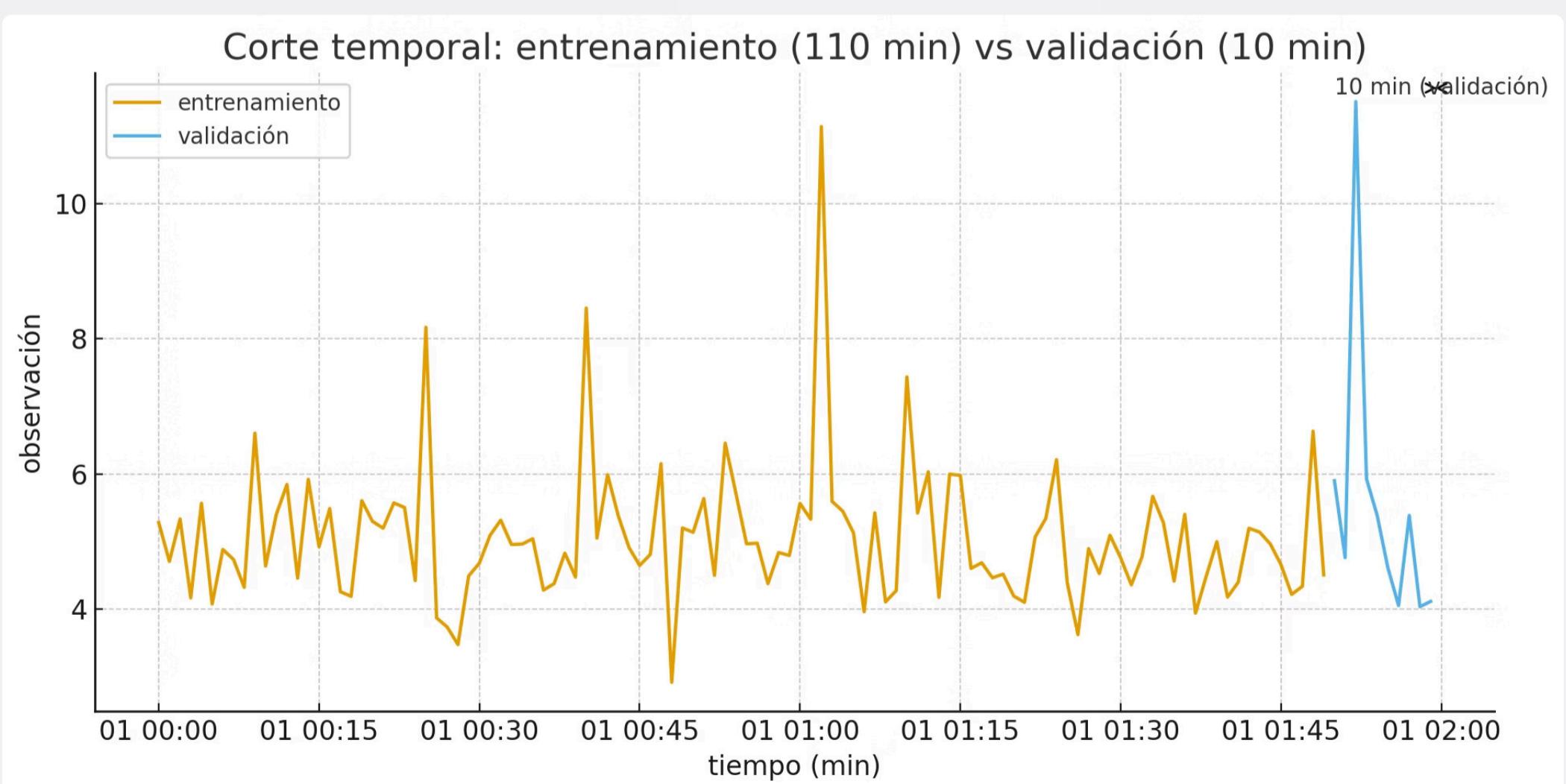
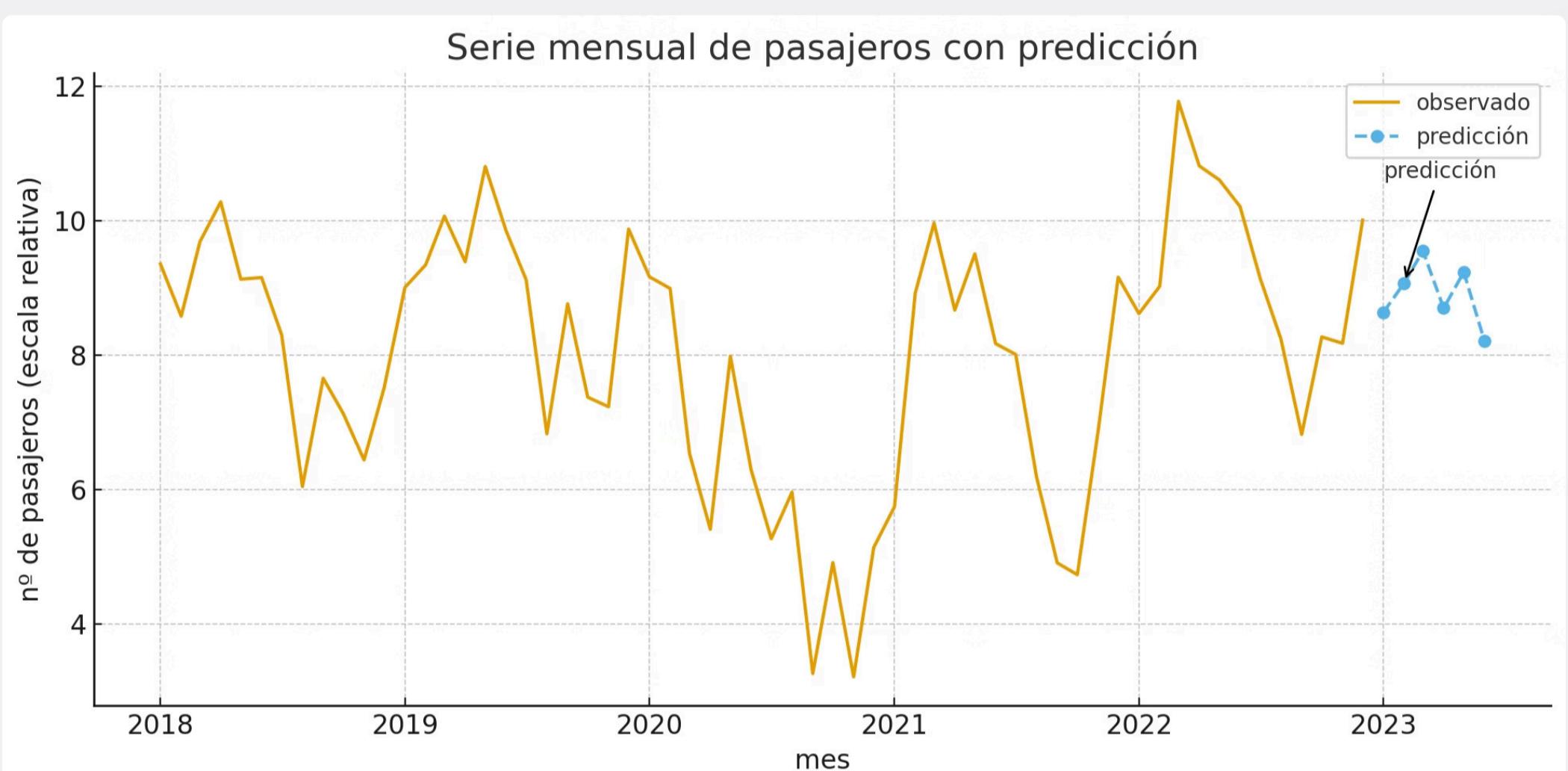
## Métricas de Precisión

MAE y RMSE son comunes. Importante: repórtelas por horizonte de predicción (ej. MAE para día 1, día 7, día 30).

04

## Intervalos de Predicción

Una predicción no es solo un número, es un rango que refleja incertidumbre. Siempre entregue un intervalo (ej. "entre \$100 y \$110 con 95% de confianza").



## Reglas de Oro del Análisis de Series Temporales

- Dibuja antes de modelar. 🖌
- Compara cambios, no niveles, si hay tendencia. 📈
- Valida hacia adelante (rolling/backtesting). 🕒
- Siempre compite contra un modelo "naïve". 🤖
- Entrega intervalos y supuestos. 📊