



# Aprendizaje Automatica

Primer Semestre 2025

# Análisis de Serie de tiempo

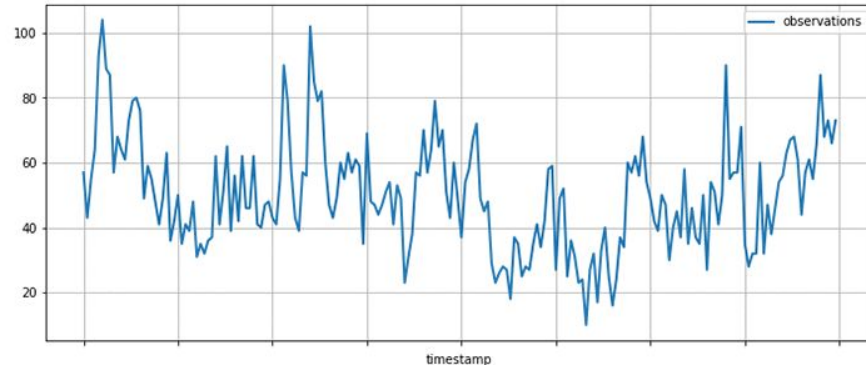
## Introducción

-  ¿Qué es aprendizaje automático?
-  Diferencias entre:
  - Datos **tabulares** (filas independientes)
  - Datos **secuenciales** (orden importa: texto, señales, series de tiempo)
- Ejemplos comunes: precios bursátiles, clima, sensores IoT, consumo eléctrico.

# Utilidad principal

Modelamiento de (y predicción en base a) series de tiempo es útil para, en base a datos históricos:

- Identificar patrones temporales existentes en los datos
- Utilizar esos patrones para predecir lo que ocurrirá a futuro.



# Componentes

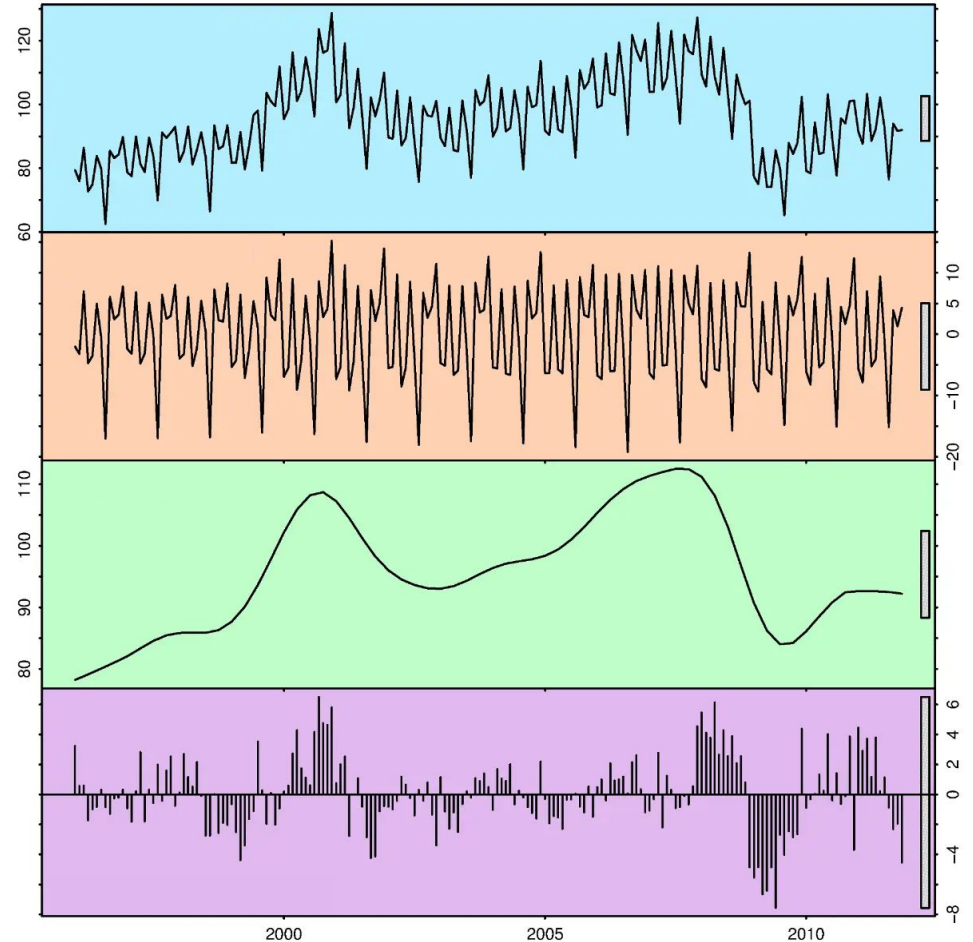
## ¿Qué es una Serie de Tiempo? (10-25 min)

- Definición: colección de observaciones ordenadas en el tiempo.
- Componentes:
  - **Tendencia:** cambio a largo plazo
  - **Estacionalidad:** patrones repetitivos
  - **Ruido:** variaciones aleatorias
- Ejemplo: ventas mensuales de un producto

# Actividad Corta

**Actividad corta:** identifiquen la tendencia o estacionalidad visualmente.

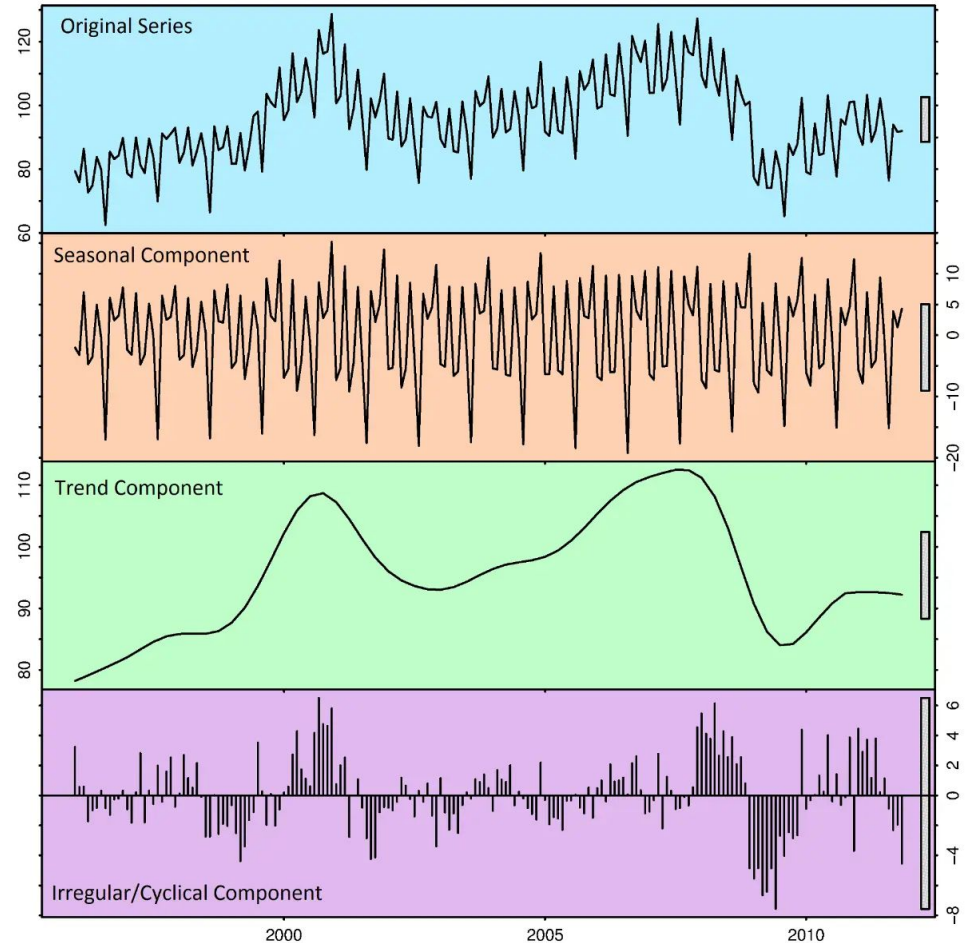
- **Tendencia:** cambio a largo plazo
- **Estacionalidad:** patrones repetitivos
- **Ruido:** variaciones aleatorias



# Actividad Corta

**Actividad corta:** identifiquen la tendencia o estacionalidad visualmente.

- **Tendencia:** cambio a largo plazo
- **Estacionalidad:** patrones repetitivos
- **Ruido:** variaciones aleatorias



# Análisis

## Análisis Exploratorio

- Visualización de datos secuenciales
- Técnicas:
  - Gráficos de línea
  - Promedios móviles:
    - i. El promedio móvil en series de tiempo es un método que calcula medias aritméticas a partir de valores sucesivos en un intervalo de tiempo. Se utiliza para analizar tendencias y patrones en los mercados.
  - Boxplots por período (mes, año, etc.)
- Identificar patrones visuales



# Análisis

## Autocorrelación y Descomposición

- ¿Una observación depende de las anteriores?
- Función de autocorrelación (ACF) y función parcial (PACF)
- Descomposición de series:
  - Adición vs multiplicación
  - Usar `seasonal_decompose` de `statsmodels`





# ¿Qué es la autocorrelación?

**Autocorrelación** significa que los valores pasados influyen en los valores futuros. Por ejemplo:

“Si vendiste mucho este mes, probablemente también venderás bastante el próximo mes”.

Esto rompe el supuesto clásico de independencia entre observaciones (como en regresión lineal).



## Cómo se representa:

- Se calcula el **coeficiente de correlación** entre una serie y una versión desplazada (lag).
- Por ejemplo, correlación entre  $Ventas(t)$  y  $Ventas(t-1)$ ,  $Ventas(t-2)$ ...

# Autocorrelación

## Funciones útiles:

### ◆ `plot_acf`: Autocorrelación (ACF)

- Mide la correlación entre la serie y sus versiones desplazadas.
- Si ves barras significativamente altas (fuera de la banda azul), hay dependencia temporal.

### ◆ `plot_pacf`: Autocorrelación Parcial (PACF)

- Muestra la correlación de cada *lag*, **eliminando la influencia de los lags anteriores**.
- Útil para identificar el número de retardos para un modelo autorregresivo (AR).



# ¿Qué es la descomposición?

La **descomposición de una serie de tiempo** separa sus componentes:

1. **Tendencia:** Componente a largo plazo (¿sube o baja la serie?)
2. **Estacionalidad:** Patrones repetitivos (mensuales, anuales, etc.)
3. **Ruido o residual:** Parte aleatoria que no se puede explicar

## ♦ Tipos de descomposición:

- **Aditiva:** Cuando los componentes se suman:  
$$Y(t) = \text{Tendencia} + \text{Estacionalidad} + \text{Ruido}$$
- **Multiplicativa:** Cuando los efectos se multiplican:  
$$Y(t) = \text{Tendencia} \times \text{Estacionalidad} \times \text{Ruido}$$

Usamos `seasonal_decompose` de `statsmodels` para esto.

# Código

- Análisis serie de tiempo:

[https://drive.google.com/file/d/1Uty5Vt\\_g-ceIEWkqgcUzkA3gaEaToNgw/view?usp=sharing](https://drive.google.com/file/d/1Uty5Vt_g-ceIEWkqgcUzkA3gaEaToNgw/view?usp=sharing)

# Modelo Arima

Un modelo **ARIMA** es útil para hacer predicciones sobre series temporales **no estacionarias**.

- **AR (Autorregresivo)**: usa valores pasados
- **I (Integrado)**: aplica diferenciación para eliminar tendencia
- **MA (Media móvil)**: usa errores pasados para corregir predicciones

La notación **ARIMA(p, d, q)** significa:

- **p**: número de lags autorregresivos
- **d**: número de diferencias para lograr estacionariedad
- **q**: número de lags de los errores (residuos)

En el ejemplo usamos: **ARIMA(1, 1, 1)**

- Toman en cuenta la autocorrelación de la variable a predecir,

consigo misma.

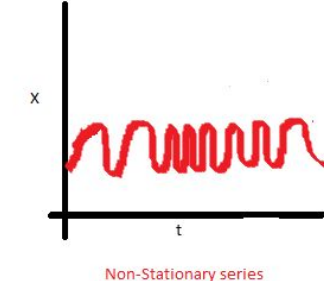
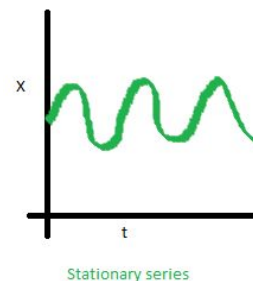
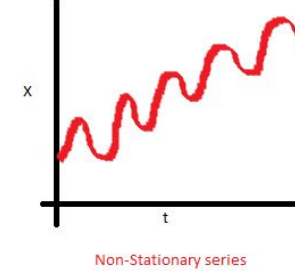
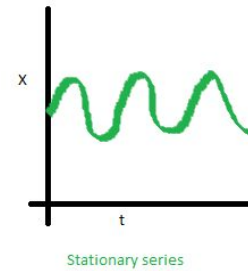
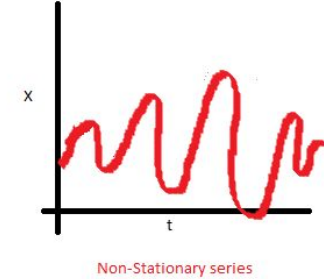
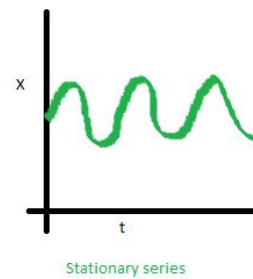
- Para su utilización, es muy conveniente que la serie de tiempo

sea estacionaria.

# Estacionalidad de una TS

Una serie de tiempo estacionaria es una cuyas propiedades no dependen del tiempo en el cual se observa la serie.

- Las series con tendencia o estacionalidad no son estacionarias → estas propiedades afectan los valores de la serie de tiempo a diferentes tiempos.
- Por otro lado, el “ruido blanco” es estacionario → no importa cuándo se le observe, se debería ver relativamente igual.



# Estacionalidad de una TS

¿Cuáles son las condiciones para que una serie de tiempo sea estacionaria?

- La media es constante, la autocovarianza también es constante
- No hay estacionalidad (comportamiento periódico en el tiempo que es predecible)

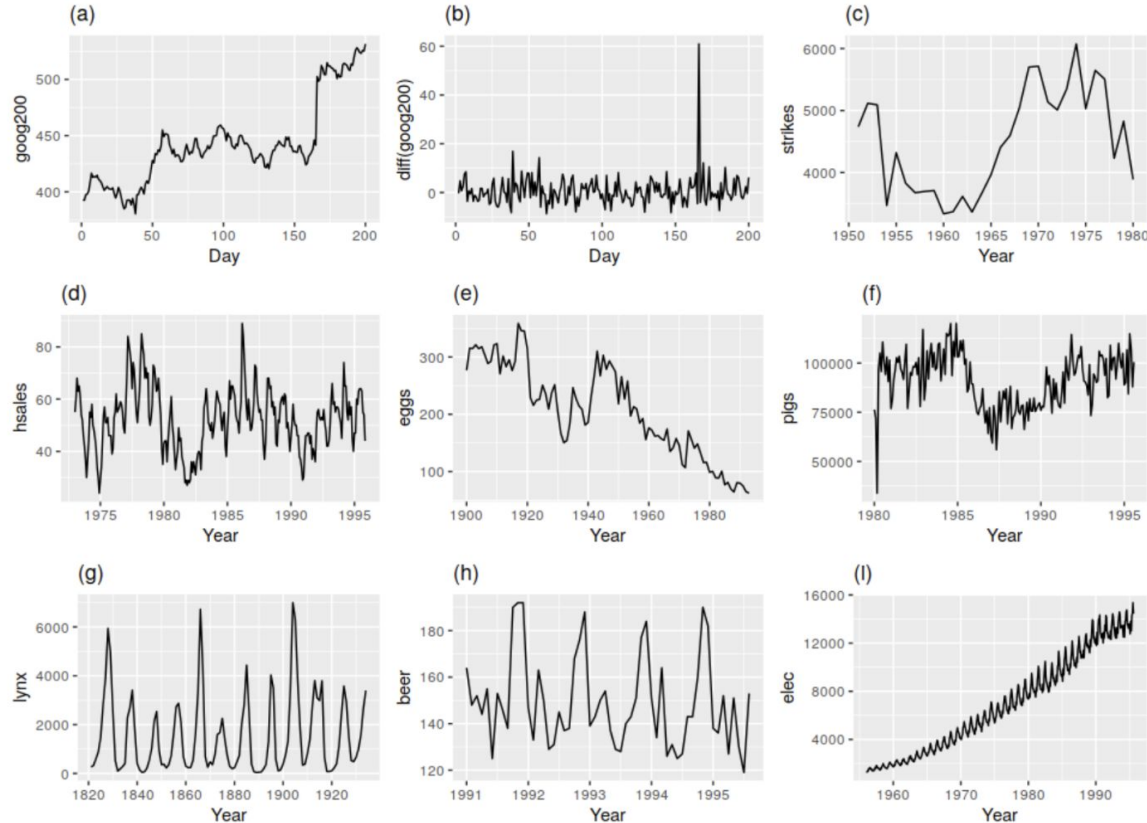
• ¿Cómo chequear que una serie de tiempo es estacionaria?

- Visualmente
- Test global vs local (media y desviación estándar)
- Usar el ADF test (Augmented Dickey-Fuller test)

• ¿Cómo hacer una serie de tiempo estacionaria?

- Diferencia: nueva serie de tiempo donde cada valor  $z_t = y_t - y_{t-1}$

# Estacionalidad de una TS



(d), (h) e (i) tienen estacionalidades

(a), (c), (f) e (i) tienen tendencia y cambio de nivel

(i) también tiene un aumento en la varianza

Nos quedan (b) y (g) como series estacionarias



# Código

- Arima:

[https://drive.google.com/file/d/1Jvv1vG6HiGn8aEV0lC-xy\\_rZvLfHibLB/view?usp=sharing](https://drive.google.com/file/d/1Jvv1vG6HiGn8aEV0lC-xy_rZvLfHibLB/view?usp=sharing)

# Laboratorio

Utilizando el dataset El sistema OGLE-LMC-CEP0227 realiza evaluaciones e intenta aplicar Arima, debes considerar la columna tiempo en calendario Juliano.



Nube de Magallanes

## 3. Análisis preprocesamiento de datos

### ■ Tipo de Datos:

- Series de tiempo astronómicas - Estrellas variables cefeidas.
- Son datos fotométricos (Tiempo = HDJ, Julian Day Heliocentric).
- Magnitud fotométrica en la banda I.

### ■ Dimensión de los Datos:

- 3345 Series de tiempo.
- Cada serie tiene 3 columnas, el tiempo, la magnitud y un Error que en este caso es despreciado.
- Observación/ ejemplo: 3 de abril de 1995 a las 9h 30m T.U. será el 2.449.811,3958 en Día Juliano.

# Código

<https://colab.research.google.com/drive/1miDD5GKmnja13W4YOWzarh6al2Eg7vy1#scrollTo=2c57a20f>