

# Tutoría de Web Scraping con Python



# ¿Qué es el Web Scraping?

01

---

## Acceder

Conectar automáticamente a una página web

02

---

## Leer

Interpretar el contenido HTML de la página

03

---

## Extraer

Obtener información específica y relevante

04

---

## Transformar

Convertir en datos estructurados y utilizables

# Librerías Principales



## Selenium

Controla un navegador web real, carga páginas dinámicas, ejecuta JavaScript y simula la navegación de un usuario



## BeautifulSoup

Analiza HTML y XML, permite navegación y búsqueda sencillas, maneja documentos mal formados y extrae datos rápidamente



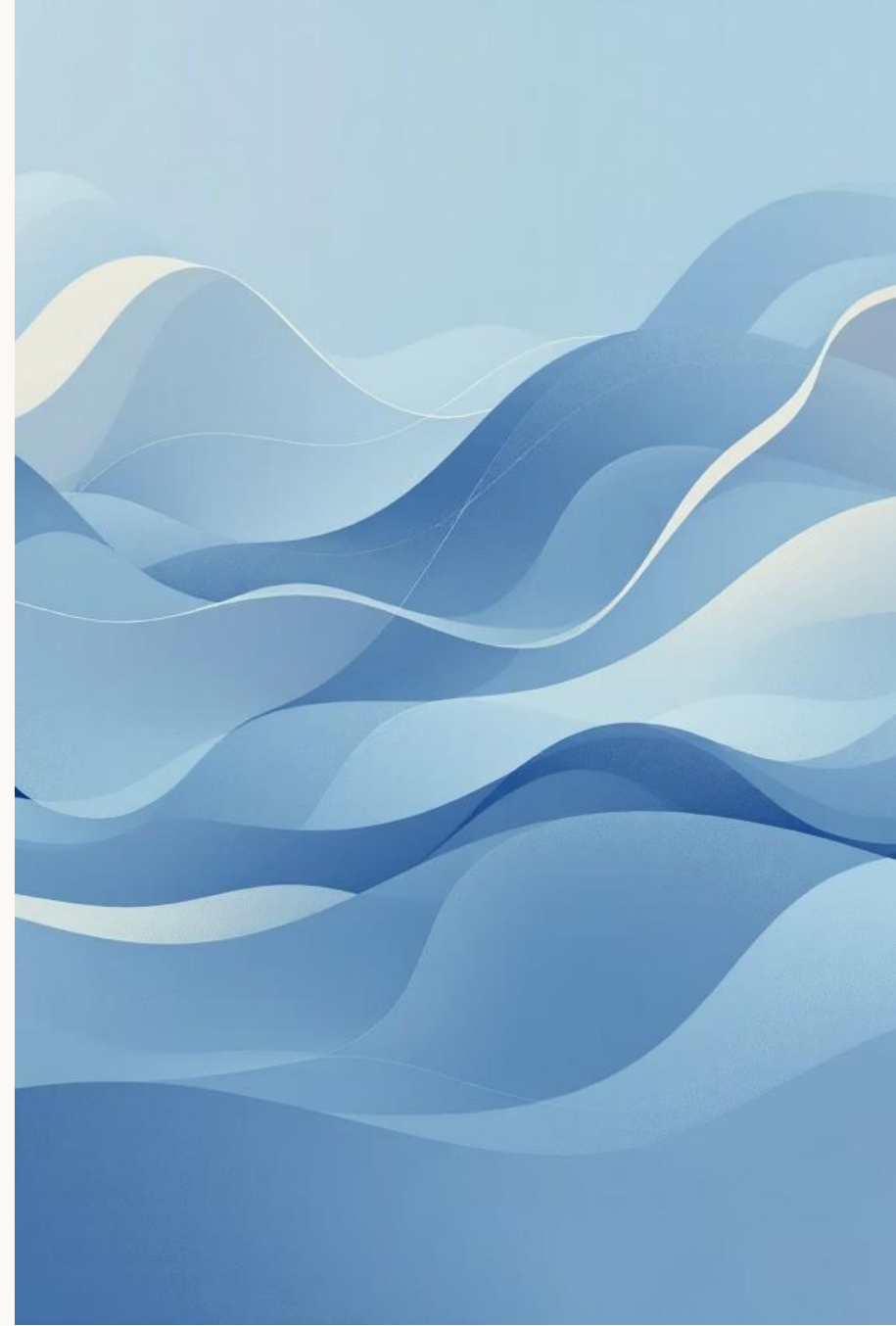
## Pandas

Organiza datos en tablas (DataFrames), permite limpiar, filtrar y actualizar datos, facilita almacenamiento y análisis

# Librerías y paquetes de Apoyo

Herramientas complementarias

- **webdriver-manager:** gestiona automáticamente siendo el driver de Selenium
- **time:** pausas para carga de páginas
- **datetime:** registro de fechas
- **numpy:** gestión de valores nulos



# ¿Cómo Trabajan Juntos?



Selenium

Carga la página web



BeautifulSoup

Extrae los datos



Selenium abre la web, BeautifulSoup la interpreta

# Flujo de Trabajo General

1

Identificar

Qué datos nos interesan

2

Explorar

La web e inspeccionar su estructura

3

Localizar

Los elementos que contienen la información

4

Extraer

Los datos de forma automática con bucles

5

Almacenar

La información en formato utilizable

# Casos de Uso Prácticos



## Crear Datos

Construcción de bases de datos desde una web



## Corregir Datos

Completar valores nulos y actualizar información incompleta



## Enriquecer Datos

Añadir contexto y traducir códigos técnicos a información legible



# 3 Proyectos Reales

El foco estará en entender cómo funciona, cómo aplicarlo y qué posibilidades ofrece



## Base de Datos de Móviles

Crear una base de datos de móviles y precios desde una web en diferentes fechas para analizar variaciones de precios en el tiempo



## Gestión de Valores Nulos

Completar fechas de estreno faltantes en una base de datos de películas



## Enriquecimiento de Datos

Asignar nombres a zonas FAO de pesca a través de su ID usando información externa