

# Web Scraping Automatizado de Precios

Arnau Sastre

[linkedin.com/in/arnausastre](https://www.linkedin.com/in/arnausastre)

August 9, 2025

## Abstract

Este proyecto desarrolla un sistema automatizado para la recolección diaria de precios de productos desde una tienda online. Utilizando técnicas de **web scraping** con **BeautifulSoup** y la automatización con **GitHub Actions**, el sistema almacena un histórico de precios en formato CSV. Es ideal para empresas de ecommerce o análisis competitivo que necesitan realizar seguimiento continuo sin intervención manual.

## 1 Objetivo

Diseñar un sistema capaz de:

- Extraer información de productos (nombre, precio y stock) desde una página web.
- Registrar esa información con marca de tiempo.
- Guardar el histórico en un archivo CSV.
- Automatizar la ejecución diaria del scraping mediante **GitHub Actions**.

## 2 Tecnologías utilizadas

- **Python 3** – lenguaje principal
- **BeautifulSoup** – extracción de datos desde HTML
- **pandas** – manipulación y almacenamiento estructurado
- **GitHub Actions** – automatización programada

## 3 Funcionamiento del sistema

### 1. Scraping de datos

A través de **BeautifulSoup**, se simula la visita a una tienda online, accediendo al HTML público y extrayendo:

- **Product name**
- **Product price**
- **Disponibilidad** (en stock / no disponible)

## 2. Registro estructurado

Los datos se estructuran con **pandas** y se agregan con timestamp a un archivo CSV:

```
data/prices_log.csv
```

## 3. Automatización diaria

El sistema se ejecuta automáticamente todos los días mediante un cron definido en GitHub Actions:

```
on:
  schedule:
    - cron: '0 7 * * *' # Cada día a las 07:00 UTC
```

El workflow lanza el script de scraping y actualiza el CSV en el repositorio o como artefacto del job.

## 4 Aplicaciones reales

- Seguimiento de precios propios o de la competencia
- Comparativa de promociones o descuentos por temporada
- Detección de quiebres de stock o cambios de disponibilidad
- Análisis de pricing en dashboards (Power BI, Looker, etc.)

## 5 Ventajas del sistema

- Fácilmente adaptable a cualquier sitio web
- Bajo coste computacional y sin dependencias externas
- Automatización completa sin intervención humana
- Compatible con pipelines de datos o reporting analítico

## 6 Conclusiones

Se ha implementado un sistema ligero, escalable y útil para automatizar la recolección de precios desde la web. Con la combinación de **BeautifulSoup** y **GitHub Actions**, este sistema es ideal para negocios que requieren análisis de precios de forma diaria y confiable.

## Contacto

Si deseas implementar scraping automatizado en tu empresa para seguimiento de precios, disponibilidad o catálogos, puedes escribirme por **LinkedIn** o **Malt**. También puedes consultar más proyectos en mi **GitHub**.