



ANDRES ROJO MESA

PROGRAMA: INGENIERIA DE SOFTWARE Y DATOS

ACTIVIDAD: Evidencia de aprendizaje 1. Análisis y herramientas  
de extracción de datos

CURSO: Programación para Análisis de Datos

Docente: ---- Diego Leon Zapata

Universidad Digital de Antioquia

2024

## Introducción

En el contexto actual de comercio electrónico, las plataformas en línea como Mercado Libre se han convertido en herramientas fundamentales tanto para consumidores como para vendedores. Estas plataformas permiten que miles de productos sean comercializados diariamente, con información como precios, descripciones y características detalladas accesibles para los usuarios. Este informe tiene como objetivo analizar cómo obtener datos de productos en Mercado Libre utilizando técnicas de *web scraping*.

En este documento se describe el proceso llevado a cabo para extraer información relevante, como el título y el precio de varios productos, utilizando herramientas de automatización como Selenium, BeautifulSoup y Scrapy. Además, se discuten los beneficios y limitaciones de estas herramientas en la extracción de datos web.

## Descripción de la Página y Artículo a Analizar

Mercado Libre es una de las plataformas de comercio electrónico más grandes en América Latina. Permite a los usuarios comprar y vender productos de diversas categorías, desde tecnología hasta moda y entretenimiento.

El artículo que se seleccionó para este análisis está relacionado con los productos de Tecnología. En este caso, se extraen datos de títulos de productos y precios asociados en Mercado Libre. Los productos incluyen, por ejemplo, laptops, accesorios electrónicos y otros artículos de alta demanda.

## Descripción del Tema de Interés que Deseas Desarrollar en la Primera Práctica

El tema principal de esta práctica es la técnica de *web scraping*, que permite extraer grandes cantidades de datos de páginas web para su posterior análisis. El scraping tiene diversas aplicaciones, desde la obtención de información de precios y disponibilidad de productos en tiendas online, hasta la recopilación de datos para investigaciones de mercado.

En este caso, se emplea *web scraping* para analizar cómo obtener títulos y precios de productos de Mercado Libre, lo cual es de interés para desarrolladores, analistas de datos y empresarios del comercio electrónico, ya que proporciona datos útiles para la comparación de precios y la toma de decisiones comerciales.

## Objetivos

El análisis de este artículo tiene como objetivo responder a las siguientes preguntas:

¿Por qué deseas analizar este artículo y la empresa de comercio?

La razón principal para analizar los productos de Mercado Libre es entender cómo los precios y las descripciones de los artículos se presentan en su plataforma, lo cual es relevante tanto para los vendedores como para los compradores.

A través del análisis de los títulos y precios, se puede obtener una visión general de las tendencias de precios, popularidad y estrategia comercial de los productos vendidos en esta plataforma.

## Objetivo general:

Implementar una solución de scraping para extraer información sobre productos (título y precio) de Mercado Libre, de manera automatizada y eficiente.

Objetivos específicos:

Utilizar Selenium, BeautifulSoup y Scrapy para automatizar la recolección de datos.

Extraer datos de productos de diversas categorías dentro de Mercado Libre.

Presentar los resultados de manera clara y estructurada para su análisis posterior.

### Metodología Empleada de Scraping

La metodología empleada para la extracción de datos se basa en el uso de Selenium, BeautifulSoup y Scrapy:

Configuración inicial:

Se utiliza Selenium WebDriver para automatizar la navegación por la página de Mercado Libre. Selenium permite interactuar con la página de manera dinámica, es decir, haciendo scroll y esperando a que el contenido cargue completamente.

Para facilitar la interacción con la página, se utiliza ChromeDriver junto con la librería webdriver\_manager.

Extracción de datos:

Una vez que la página ha sido completamente cargada mediante el desplazamiento automático (scroll), se utiliza BeautifulSoup para analizar el HTML resultante y extraer los títulos de los productos y los precios asociados.

Se inicia un proyecto nuevo con scrapy startproject extracion\_mercado\_libre, generando una estructura de archivos como esta extracion\_mercado\_libre /

scrapy.cfg

extracion\_mercado\_libre /

\_\_init\_\_.py

items.py

middlewares.py

pipelines.py

settings.py

spiders/

\_\_init\_\_.py

Los productos son seleccionados mediante las clases CSS que identifican los elementos correspondientes: los títulos son extraídos a partir de la clase `poly-component__title`, mientras que los precios son extraídos de la clase `andes-money-amount__fraction`.

Desarrollo del código:

El código implementa un scroll infinito para cargar más productos a medida que se baja por la página.

Los datos extraídos se presentan en un formato legible, donde se muestran el título y el precio de cada producto.

Resultados y Conclusiones

Los resultados obtenidos durante el proceso de scraping incluyen una lista de títulos de productos y sus precios correspondientes. El código logró extraer una muestra representativa de productos dentro de la página de Mercado Libre, aunque en algunos casos, los resultados mostraron productos incompletos debido a la carga dinámica de la página.

Resultados clave:

El scraping permitió obtener títulos y precios de más de 30 productos.

La implementación del scroll y la espera para la carga completa de la página fue efectiva para asegurar que los productos estuvieran completamente cargados antes de su extracción.

Hubo algunas limitaciones relacionadas con la estructura dinámica de la página, que no siempre mostró todos los elementos deseados en las primeras cargas.

Conclusiones:

El uso de Selenium, BeautifulSoup y Scrapy resultó ser efectivo para la extracción de datos de Mercado Libre, permitiendo la recolección de información sobre productos sin necesidad de interactuar manualmente con la plataforma.

Las técnicas de scraping pueden ser muy útiles para la recopilación de datos en tiempo real de sitios de comercio electrónico, aunque se deben considerar aspectos como la carga dinámica de contenido.

Las herramientas utilizadas en este proyecto pueden ampliarse para incluir análisis más complejos, como la comparación de precios entre diferentes productos o la creación de modelos de predicción de precios.