

ANÁLISIS COMPETITIVO DE PRECIOS MEDIANTE WEB SCRAPING

por Lucila Aldana Quiñonez | Analista de Datos de Marketing

Problema de Negocio

Una empresa de e-commerce de productos de cuidado personal desea mejorar su estrategia de precios frente a competidores directos. Actualmente, el equipo de marketing realiza el seguimiento de precios de forma manual, lo que genera:

- Retrasos en la toma de decisiones
- Información incompleta
- Dificultad para detectar oportunidades de pricing

El objetivo del proyecto es automatizar la recolección de precios y disponibilidad de productos de la competencia para realizar análisis comparativos que permitan:

- Ajustar precios
- Detectar posicionamiento de mercado
- Identificar oportunidades comerciales

Para lograrlo, el equipo directivo ha encomendado al analista de datos la tarea de desarrollar un pipeline de web scraping en Python que permita:

1. Extraer automáticamente información de productos de sitios web de competidores
2. Almacenar los datos en una base estructurada

Conjunto de Datos

El conjunto de datos incluye las siguientes variables:

1. Nombre del Producto
2. Marca
3. Precio
4. Descuento
5. Disponibilidad: En stock / Agotado
6. Categoría
7. Subcategoría
8. URL del Producto
9. Competidor
10. Fecha de Extracción

ANÁLISIS Y HALLAZGOS

Este proyecto consiste en el desarrollo de un sistema de web scraping en Python para la extracción automatizada de información de productos desde sitios web públicos de e-commerce. El objetivo es recolectar datos clave y almacenarlos en una base de datos estructurada, dejando el dataset preparado para futuros análisis comparativos de precios y estrategias comerciales.

El análisis considera como competidores a cuatro empresas argentinas de e-commerce de productos de cuidado personal: GPSFarma (Competidor A), Farmaonline (Competidor B), Farmacity (Competidor C) y Farmalife (Competidor D).

De cada competidor se obtiene la siguiente información, la cual se almacena en una base de datos estructurada:

- product_name
- brand
- price
- discount
- availability
- category
- subcategory
- product_url
- competitor
- scraping_date

Estrategias de Precios, Descuentos y Envíos

Los cuatro competidores analizados presentan estrategias de precios, descuentos y envíos muy similares. En términos generales, los precios de los productos son equivalentes entre las distintas plataformas, acompañados por promociones frecuentes que suelen oscilar entre el 10% y el 40%, alcanzando descuentos de hasta el 50% en campañas puntuales, y complementándose en muchos casos con beneficios bancarios o reintegros con tarjetas adheridas. Asimismo, todos ofrecen retiro gratuito en sucursales o farmacias asociadas, y establecen umbrales de compra mínima para acceder al envío a domicilio sin costo, los cuales varían levemente entre competidores. Estas prácticas homogéneas reducen el precio como factor diferenciador, trasladando la competencia hacia otros aspectos como la experiencia de compra, la cobertura logística y la disponibilidad de productos.

Nota Ética y Legal

Para el desarrollo de este proyecto se utilizaron exclusivamente sitios web con información pública y de libre acceso, sin requerimiento de login, autenticación ni bypass de medidas de seguridad. El objetivo del web scraping es estrictamente académico y demostrativo, enfocado en el análisis de datos y la simulación de un caso real de negocio, respetando las buenas prácticas, los términos de uso de los sitios y la integridad de la información consultada.

Desarrollo del Web Scraping

Competidor A: GPSFarma

GPSFarma permitió una extracción de datos estable y consistente desde su catálogo de productos. La información de precios, descuentos y disponibilidad se encuentra accesible de forma estructurada, lo que permitió construir una base de datos completa y confiable.

Por este motivo, se asignó a GPSFarma como el principal competidor para el análisis comparativo, ya que sus datos pudieron ser recolectados de manera sistemática y reproducible.

Competidor B: Farmaonline

Farmaonline presenta su catálogo de productos de forma dinámica, lo que impide acceder de manera directa a la información completa de los productos. Si bien fue posible explorar parcialmente su estructura, no se logró obtener un conjunto de datos estable y consistente.

Debido a estas limitaciones, Farmaonline fue excluido del dataset final para evitar inconsistencias y asegurar la calidad del análisis.

Competidor C: Farmacity

Farmacity permitió la extracción de información únicamente desde la primera página de resultados de cada categoría. Si bien fue posible recuperar algunos productos y validar la estructura general del catálogo, el acceso al resto del inventario presentó restricciones.

Como resultado, los datos obtenidos de Farmacity son parciales y se utilizaron de manera exploratoria, pero no se consideran representativos del total de su oferta.

Competidor D: Farmalife

Farmalife utiliza un catálogo de productos completamente dinámico y limita el acceso a sus listados, lo que impide una extracción sistemática de información. Durante la etapa exploratoria se lograron obtener algunos productos de forma aislada, pero los resultados no fueron consistentes ni reproducibles.

Por este motivo, Farmalife fue excluido del dataset final y el análisis se centró en el competidor que permitió una recolección de datos estable y confiable.

Limitaciones Técnicas del Web Scraping

Durante el desarrollo de este proyecto se intentó extraer información de productos de cuatro farmacias online relevantes del mercado argentino. Sin embargo, se identificaron diferencias significativas en la forma en que cada sitio expone sus datos.

Solo uno de los competidores permite un scraping completo mediante HTML estático utilizando requests y BeautifulSoup. Los restantes sitios utilizan renderizado dinámico basado en JavaScript (principalmente plataformas como VTEX), donde los productos, precios y descuentos se cargan de forma asíncrona o a través de APIs internas no documentadas.

Debido a estas limitaciones técnicas, en dichos casos:

- El HTML inicial no contiene la información completa de los productos.
- Las llamadas a APIs internas presentan restricciones, validaciones o comportamientos inconsistentes.
- No es posible garantizar una extracción estable, reproducible y escalable sin el uso de herramientas de automatización del navegador (por ejemplo, Selenium o Playwright).

Por este motivo, el proyecto prioriza la calidad y confiabilidad de los datos, documentando explícitamente los casos en los que el scraping no resulta viable mediante técnicas tradicionales.