

COMPETITIVE PRICING ANALYSIS USING WEB SCRAPING

by Lucila Aldana Quiñonez | Marketing Data Analyst

Business Problem

An e-commerce company specializing in personal care products wants to improve its pricing strategy in relation to direct competitors. Currently, the marketing team tracks prices manually, which results in:

- Delays in decision-making
- Incomplete information
- Difficulty identifying pricing opportunities

The objective of the project is to automate the collection of competitors' product prices and availability in order to perform comparative analyses that enable the company to:

- Adjust prices
- Detect market positioning
- Identify commercial opportunities

To achieve this, the executive team has tasked the data analyst with developing a web scraping pipeline in Python that will:

- Automatically extract product information from competitors' websites
- Store the data in a structured database

Dataset Overview

The dataset will include the following variables:

- Product Name
- Brand
- Price
- Discount
- Availability: In stock / Out of stock
- Category
- Subcategory
- Product URL
- Competitor
- Scraping Date

ANALYSIS AND FINDINGS

This project involves the development of a Python web scraping system for the automated extraction of product information from public e-commerce websites. The objective is to collect key data and store it in a structured database, preparing the dataset for future comparative analyses of pricing and business strategies.

The analysis considers four Argentine e-commerce companies specializing in personal care products as competitors: GPSFarma (Competitor A), Farmaonline (Competitor B), Farmacity (Competitor C), and Farmalife (Competitor D).

For each competitor, the following information is collected and stored in a structured database:

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

Pricing, Discount, and Shipping Strategies

The four competitors analyzed exhibit very similar pricing, discount, and shipping strategies. In general terms, product prices are equivalent across the different platforms, accompanied by frequent promotions that usually range between 10% and 40%, reaching discounts of up to 50% in specific campaigns, and often complemented by bank benefits or cash-back offers with affiliated cards. Likewise, all of them offer free pickup at branches or associated pharmacies, and set minimum purchase thresholds to qualify for free home delivery, which vary slightly between competitors. These homogeneous practices reduce price as a differentiating factor, shifting competition toward other aspects such as the shopping experience, logistical coverage, and product availability.

Ethical and Legal Note

For the development of this project, only websites with publicly available and freely accessible information were used, without requiring login, authentication, or bypassing security measures. The purpose of the web scraping is strictly academic and demonstrative, focused on data analysis and simulating a real business case, while respecting best practices, the sites' terms of use, and the integrity of the information consulted.

Web Scraping Development

Competitor A: GPSFarma

GPSFarma allowed for stable and consistent data extraction from its product catalog. Pricing, discount, and availability information is accessible in a structured format, which enables the creation of a complete and reliable database.

For this reason, GPSFarma was assigned as the main competitor for the comparative analysis, since its data could be collected systematically and reproducibly.

Competitor B: Farmaonline

Farmaonline presents its product catalog dynamically, which prevents direct access to the full product information. Although it was possible to partially explore its structure, a stable and consistent dataset could not be obtained.

Due to these limitations, Farmaonline was excluded from the final dataset to avoid inconsistencies and ensure the quality of the analysis.

Competitor C: Farmacity

Farmacy allowed information extraction only from the first page of results in each category. While it was possible to retrieve some products and validate the general structure of the catalog, access to the rest of the inventory was restricted.

As a result, the data obtained from Farmacity is partial and was used for exploratory purposes, but it is not considered representative of its full offering.

Competitor D: Farmalife

Farmalife uses a fully dynamic product catalog and restricts access to its listings, which prevents systematic data extraction. During the exploratory phase, some products were obtained in isolation, but the results were neither consistent nor reproducible.

For this reason, Farmalife was excluded from the final dataset and the analysis focused on the competitor that allowed stable and reliable data collection.

Technical Limitations of Web Scraping

During the development of this project, an attempt was made to extract product information from four leading online pharmacies in the Argentine market. However, significant differences were identified in the way each site exposes its data.

Only one of the competitors allows complete scraping through static HTML using requests and BeautifulSoup. The remaining sites use JavaScript-based dynamic rendering (mainly platforms such as VTEX), where products, prices, and discounts are loaded asynchronously or through undocumented internal APIs.

Due to these technical limitations, in these cases:

- The initial HTML does not contain the complete product information.
- Calls to internal APIs have restrictions, validations, or inconsistent behavior.
- It is not possible to guarantee stable, reproducible, and scalable extraction without using browser automation tools (e.g., Selenium or Playwright).

For this reason, the project prioritizes data quality and reliability, explicitly documenting the cases in which scraping is not feasible using traditional techniques.