



**INSTITUTO POLITÉCNICO NACIONAL**  
**CENTRO DE INVESTIGACIÓN EN CÓMPUTO**  
**DEPARTAMENTO DE DIPLOMADOS Y**  
**EXTENSIÓN PROFESIONAL**



**Programación Python con Aplicaciones en el Ámbito Científico**

**Web Scraping de Productos en la Página Principal de Sanborns Utilizando Selenium**

**PROFESOR**

Alan Badillo Salas

**ALUMNO**

José Vladimir Amador Miranda

[soneroamador@gmail.com](mailto:soneroamador@gmail.com)

Julio 2023

## **Introducción**

El rápido crecimiento del comercio electrónico ha resultado en la creación de vastas cantidades de datos en línea. Esto ha abierto la puerta a la práctica del web scraping, que se refiere a la extracción automática de información de las páginas web.

En este informe, nos centraremos en la extracción de datos de los productos disponibles en la página principal de Sanborns, una famosa cadena de tiendas en México.

Esta práctica nos permitirá adquirir habilidades cruciales en la recopilación de datos, especialmente en lo que respecta a la manipulación de herramientas y bibliotecas de Python para la automatización de procesos de extracción de datos. Además, nos ayudará a entender la importancia de tener una metodología eficiente y eficaz para la recopilación de datos, que es una parte fundamental de cualquier proyecto de análisis de datos o de machine learning.

## **Justificación**

En el ámbito del análisis de datos, tener la habilidad de recolectar y procesar datos directamente de la web es de gran importancia. Específicamente en este caso, la información recolectada podría usarse para analizar las tendencias de los productos, monitorear los cambios de precios, comparar precios con otras tiendas, o hasta construir un sistema de recomendación basado en productos similares.

La importancia de este ejercicio no se limita solo a su aplicación práctica inmediata, sino que también destaca la versatilidad y flexibilidad de las bibliotecas de Python en el tratamiento de datos en línea, alentando una mejor comprensión de la gestión de datos.

El web scraping también tiene un gran potencial para ayudar a las empresas a tomar decisiones informadas basadas en datos recopilados de varias fuentes en línea, lo que en última instancia puede contribuir a mejorar sus estrategias de negocio y rendimiento en el mercado.

## Desarrollo

1. Inicialmente, instalamos las bibliotecas necesarias (selenium, pandas y beautifulsoup4) usando el comando `!pip install`.
2. Luego, creamos un navegador automatizado usando webdriver de Selenium y visitamos la página de Sanborns.
3. Creamos listas vacías para almacenar la información de los productos: nombres, precios e imágenes.
4. Usando las herramientas de Selenium, encontramos los nodos HTML donde se encuentra la información de los productos.
5. Iteramos sobre cada nodo, extraemos la información requerida y la almacenamos en nuestras listas.
6. Creamos un DataFrame de Pandas para estructurar los datos recolectados.
7. Finalmente, guardamos el DataFrame en un archivo CSV y cerramos el navegador automatizado.

```
from selenium import webdriver
from selenium.webdriver.common.by import By
import pandas as pd

# Iniciar el navegador
driver = webdriver.Chrome()

# Visitar la página
driver.get('https://www.sanborns.com.mx/')

# Crear listas vacías para almacenar la información
nombres = []
precios = []
imagenes = []

# Encontrar los nodos con la información de los productos
# Estos selectores pueden variar dependiendo de la estructura de la página
products = driver.find_elements(By.XPATH,
    "//*[@starts-with(@class, 'CardProduct_contDataCard')]")

# Recorrer cada producto y extraer la información
for product in products:
    nombre = product.find_element(By.XPATH, ".//h3").text
    precio = product.find_element(By.XPATH,
```

```

"/p[starts-with(@class, 'CardProduct_precio1')]").text
    imagen = product.find_element(By.XPATH,
"/../picture/img").get_attribute('src')

    # Agregar la información a las listas
    nombres.append(nombre)
    precios.append(precio)
    imagenes.append(imagen)

# Crear un DataFrame con la información
df = pd.DataFrame({
    'nombre': nombres,
    'precio': precios,
    'imagen': imagenes,
})

# Guardar el DataFrame en un archivo CSV
df.to_csv(r'C:\test\sanborns_products.csv', index=False)

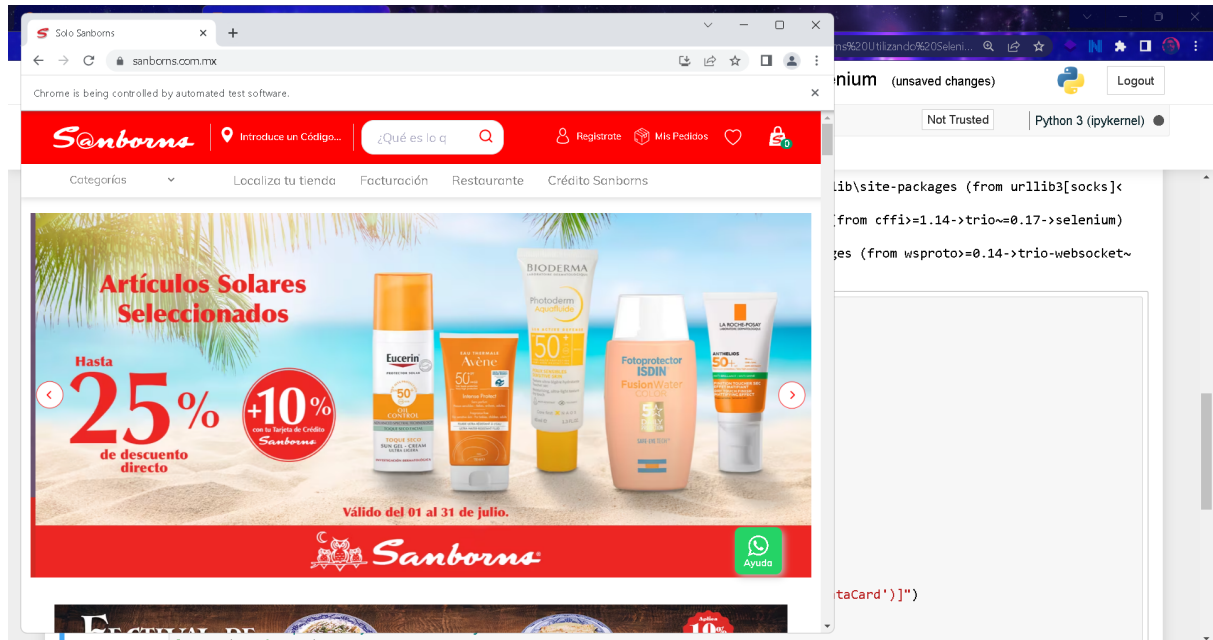
# Cerrar el navegador
driver.quit()

```

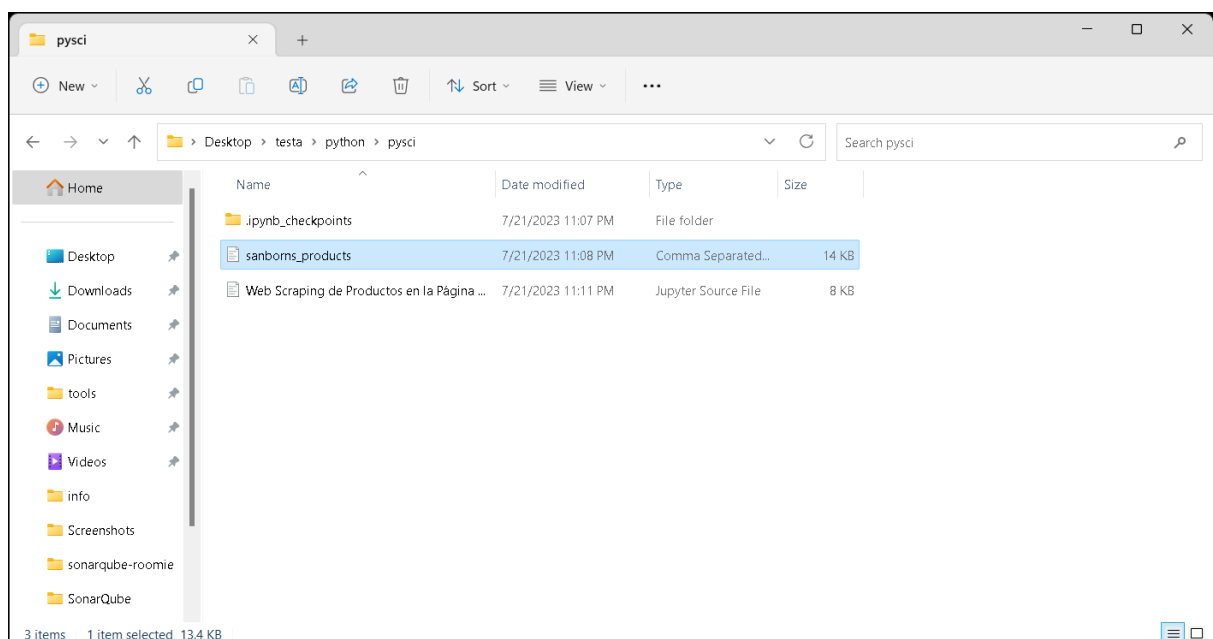
Este proceso de desarrollo no solo nos permite recopilar datos de una página web, sino que también nos da una idea de cómo se puede utilizar la programación para automatizar tareas que, si se realizan manualmente, consumirían un tiempo considerable.

## Resultados

*El programa abre un navegador para ejecutar el raspado.*



*Se crea un archivo llamado sanborns\_products.csv con los datos recolectados.*



*Visualización del archivo CSV.*

```
File Edit Selection View Go Run Terminal Help sanborns_products.csv - Visual Studio Code
sanborns_products.csv x
C:\Users\josea> Desktop > tests > python2 pysd > sanborns_products.csv
1 nombre,precio,imagen
2 Barbie La Pelicula Muñeca de Colección Presidente,"$1,699MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735160693.jpg?scale=250&qilty=75
3 Barbie Signature Muñeca Los Polinesios Aniversario 10,"$1,591MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735067886.jpg?scale=250&qilty=75
4 Mochila De Dama Barbie X Gorett Backpack Mediana Color Camel Modelo BR22067,K,$951MXN,https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7509048397224.jpg?scale=250&qilty=75
5 Patines De Moda Barbie La Pelicula,$5299MXN,https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735162376.jpg?scale=250&qilty=75
6 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7509048400689.jpg?scale=250&qilty=75
7 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735160747.jpg?scale=250&qilty=75
8 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7509048404403.jpg?scale=250&qilty=75
9 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7509048404397.jpg?scale=250&qilty=75
10 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7509048404373.jpg?scale=250&qilty=75
11 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/887961918717.jpg?scale=250&qilty=75
12 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7509048404748.jpg?scale=250&qilty=75
13 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735160822.jpg?scale=250&qilty=75
14 Mochila De Dama Barbie X Gorett Backpack Mediana Color Camel Modelo BR22067,K,$951MXN,https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/817742021848.jpg?scale=250&qilty=75
15 Cámara Fujifilm Instax Mini Z1 blanca,"$2,329MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/454710488999.jpg?scale=250&qilty=75
16 Camisa blanca Candillo manga corta 5729F8 talla mediana,$749MXN,https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/750648520827.jpg?scale=250&qilty=75
17 Vestido midi calado blanco,$849MXN,https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/679628268064.jpg?scale=250&qilty=75
18 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/8410412000437.jpg?scale=250&qilty=75
19 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7506235591021.jpg?scale=250&qilty=75
20 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/4547410488968.jpg?scale=250&qilty=75
21 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/817742020834.jpg?scale=250&qilty=75
22 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/8410412490092.jpg?scale=250&qilty=75
23 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/8425402159552.jpg?scale=250&qilty=75
24 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/677273840616.jpg?scale=250&qilty=75
25 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/5019993648719.jpg?scale=250&qilty=75
26 MSN Mario Party Superstars,"$1,299MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/45496597887.jpg?scale=250&qilty=75
27 Ring fiberoptico,"$1,439MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/6021761000432.jpg?scale=250&qilty=75
28 Samsung Galaxy Watch 4 40 mm Negro,"$2,999MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/880692532820.jpg?scale=250&qilty=75
29 Cámara Canon kit M50 Mark II EF M 15-45 mm IS STM + maleta 100S + SD 32GB+ curso online ABC Vlogger,"$13,929MXN",https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/880692532820.jpg?scale=250&qilty=75
30 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7503036888839.jpg?scale=250&qilty=75
31 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/880692044425.jpg?scale=250&qilty=75
32 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/17817832021.jpg?scale=250&qilty=75
33 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/5025155046883.jpg?scale=250&qilty=75
34 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/18208016327.jpg?scale=250&qilty=75
35 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/195477875791.jpg?scale=250&qilty=75
36 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/880692766426.jpg?scale=250&qilty=75
37 https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/880691765079.jpg?scale=250&qilty=75
```

## Conclusión

El proceso de web scraping presentado en este informe demostró ser una herramienta eficiente para la recopilación de datos de la página web de Sanborns. Sin embargo, la ejecución del mismo no estuvo exenta de desafíos. El principal obstáculo fue la selección de los nodos de los productos en la página web de Sanborns. Tras intentar utilizar selectores de clase CSS y enfrentar inconsistencias, finalmente se decantó por XPath para una identificación más precisa.

Un problema adicional que podría presentarse es la ejecución de este código en Google Colab, dado que requiere de la interacción con un navegador web en tiempo real. Google Colab es un entorno de ejecución basado en la nube y, por lo tanto, no tiene un navegador gráfico.

En términos generales, este ejercicio ha sido una excelente oportunidad para profundizar en el manejo de herramientas de web scraping y reforzar las habilidades de programación y gestión de datos en Python. Sin embargo, la experiencia también ha subrayado la necesidad de adaptabilidad y flexibilidad en este campo, ya que el web scraping a menudo requiere ajustes y refinamientos para superar los desafíos que surgen debido a la naturaleza dinámica y en constante cambio de las páginas web.