



Instituto Politécnico Nacional

Centro de Investigación en Computación

Julio 2023

Nombre del profesor: Alan Badillo Salas

Nombre del alumno: José Emiliano Serrato Rodríguez

Correo electrónico: eserratorodriguez@gmail.com

Introducción:

El presente trabajo es la practica final del curso “Programación Python con aplicaciones en el ámbito científico”, dicha practica consiste en la extracción de datos de la pagina web de *Sanborns* mediante la técnica conocida como “web scraping”.

Dicha actividad será realizada mediante el lenguaje de programación *Python* y usando la librería *Selenium* cuya finalidad es facilitar tareas de automatización de aplicación web y buscadores, ello permite realizar la extracción de cualquier tipo de dato de sitios web de forma más eficiente.

Por último, se buscará guardar los datos obtenidos dentro de un dataframe usando la librería *Pandas* y posteriormente serán guardados dentro de un archivo CSV para su posterior manejo y estudio.

Justificación:

El *web scraping* es una técnica que puede tener varios usos dentro de varias ramas del conocimiento y aplicaciones dentro del ámbito profesional y científico. Entre los beneficios del *web scraping* se pueden encontrar: la recopilación de datos para realizar análisis de mercado y competencia, la investigación de tendencias, realizar diagnósticos de paginas web y automatizar tareas como las anteriormente mencionadas.

Pasos:

```
#primer paso: se instala el paquete de selenium  
!pip install selenium
```

```
#segundo paso: se importan las características necesarias desde selenium  
from selenium import webdriver  
from selenium.webdriver.common.keys import Keys  
from selenium.webdriver.common.by import By  
  
#tercer paso: se selecciona la ruta del navegador y la url de la pagina para ser automatizada  
driver = webdriver.Chrome()  
driver.get("https://www.sanborns.com.mx/")  
  
#cuarto paso: se escriben las rutas y tipos de texto que se desean extraer de la pagina web  
images = driver.find_elements(By.XPATH, "//div[@class='promotionAD']/a/picture/img")  
image_urls = [image.get_attribute("src") for image in images]  
print(image_urls)  
  
precios = driver.find_elements(By.XPATH, "//p[starts-with(@class, 'CardProduct_precio1')]")  
precios_texto = [precio.text for precio in precios]  
print(precios_texto)  
  
nombres = driver.find_elements(By.XPATH, "//p[starts-with(@class, 'CardProduct_precio1')]/../h3")  
nombres_texto = [nombre.text for nombre in nombres]  
print(nombres_texto)  
  
#quinto paso: se crean listas para almacenar los datos obtenidos  
data = {  
    'Nombre': [],  
    'Precio': [],  
    'Imagen': []  
}  
  
#sexto paso: se extraen los datos bajo los parametros de nombre, precio e imagen  
card_products = driver.find_elements(By.XPATH, "//article[starts-with(@class, 'CardProduct_cardProduct')]")  
for card_product in card_products:  
    nombre = card_product.find_element(By.XPATH, "../h3").text  
    precio = card_product.find_element(By.XPATH, "//p[starts-with(@class, 'CardProduct_precio1')]").text  
    imagen = card_product.find_element(By.XPATH, "../img").get_attribute("src")  
  
#se excluyen los datos sin texto  
    if len(nombre) == 0 or len(precio) == 0 :  
        continue  
  
#septimo paso: se llenan las listas creadas anteriormente  
    data['Nombre'].append(nombre)  
    data['Precio'].append(precio)  
    data['Imagen'].append(imagen)  
    print(nombre, precio, imagen)  
  
#se cierra la sesión de selenium  
driver.close()
```

```
#se importa la libreria pandas
import pandas as pd
```

```
#se crea un dataframe con los datos recolectados
df = pd.DataFrame(data)

print(df)
```

```
#se guarda el dataframe en forma de csv
df.to_csv('sanborns.csv', index = False)
```

Resultados:

1.Datos extraídos

Barbie La Película Muñeca de Colección Presidente \$1,699MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735160693.jpg?scale=250&qilty=75>
Barbie Signature Muñeca Los Polinesios Aniversario 10 \$1,591MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735067886.jpg?scale=250&qilty=75>
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>
Patines De Moda Barbie La Película \$529MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/194735162376.jpg?scale=250&qilty=75>
Inflable Infantil en Forma de Piña Bigmouth \$214MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/817742021848.jpg?scale=250&qilty=75>
Cámara Fujifilm Instax Mini 12 blanca \$2,239MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/4547410488999.jpg?scale=250&qilty=75>
Camisa blanca GCandila manga corta 5729FB talla mediana \$749MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7506485820827.jpg?scale=250&qilty=75>
Vestido midi calado blanco \$849MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/679628268064.jpg?scale=250&qilty=75>
NSW Mario Party Superstars \$1,289MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/45496597887.jpg?scale=250&qilty=75>
Free Libre Sensor \$1,439MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/5021791000432.jpg?scale=250&qilty=75>
Samsung Galaxy Watch 4 40 mm Negro \$2,999MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/8806092538290.jpg?scale=250&qilty=75>
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/1200/660685250741.jpg?scale=250&qilty=75>
Gente como yo \$349MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/9786073824194.jpg?scale=250&qilty=75>
Viajar cambiará tu vida \$328MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/9786070748516.jpg?scale=250&qilty=75>
Claudia Sheinbaum: Presidenta \$199MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/9786073835022.jpg?scale=250&qilty=75>
Descubre a los más villanos, arte antiestrés \$199MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/9786072128835.jpg?scale=250&qilty=75>
Samsung Galaxy S23 Ultra 256GB Rosa Telcel R6 \$26,999MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/2005666508721.jpg?scale=250&qilty=75>
Motorola E20 32GB Gris Telcel R9 \$2,499MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/2005676007771.jpg?scale=250&qilty=75>
Xiaomi Redmi Note 9 Pro Gris R9 Telcel \$9,199MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/2005692127569.jpg?scale=250&qilty=75>
iPhone 11 64GB Blanco R4 (Telcel) \$9,999MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/2005681944283.jpg?scale=250&qilty=75>
Internet Security- Multidispositivos 1 Año Kaspersky \$589MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/83832307011.jpg?scale=250&qilty=75>
Laptop Honor Magicbook 15" Ryzen 5 16 RAM 512 GB \$15,299MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/6973316859300.jpg?scale=250&qilty=75>
Microsoft 365 Family \$1,499MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/889842862461.jpg?scale=250&qilty=75>
Mouse wireless pebble m350 rosa \$449MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/97855154811.jpg?scale=250&qilty=75>
Cuadro De Luz \$319MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/39052018527.jpg?scale=250&qilty=75>
Candado De Combinación Números Grandes \$279MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/39052013942.jpg?scale=250&qilty=75>
Irrigador Bucal Sonico Interplak \$599MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/85452000730.jpg?scale=250&qilty=75>
Botellitas Con Masajeador \$279MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/39052019210.jpg?scale=250&qilty=75>
Cámara de vídeo GoPro Hero 11 Black \$7,299MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/818279029079.jpg?scale=250&qilty=75>
Cámara Nikon Z50 16-50 y 50-250 mm F/4 \$29,759MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/18208016327.jpg?scale=250&qilty=75>
CÁMARA CANON T7 EF-S 18-55 IS-7 \$14,959MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/660685207523.jpg?scale=250&qilty=75>
Cámara Sony ILCE-7M3 montura tipo E full frame \$38,261MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/4548736079687.jpg?scale=250&qilty=75>
Figura Hallmark Mickey Mouse paz y amor \$1,299MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/763795764198.jpg?scale=250&qilty=75>
Hallmark taza Harry Potter Hogwarts \$599MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/763795790722.jpg?scale=250&qilty=75>
Barco santa Maria Yijie \$739MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/2003172645664.jpg?scale=250&qilty=75>
Globo negro y verde con base y diámetro de 20 cm \$1,086MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/2003164087182.jpg?scale=250&qilty=75>
Collar Con Dije De Elsa De Frozen I De Disney \$160MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7503029702687.jpg?scale=250&qilty=75>
Pin de Stich playa de Disney \$160MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7503029702427.jpg?scale=250&qilty=75>
Bolsa de Minnie de peluche de Disney BP06 \$306MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7503029702885.jpg?scale=250&qilty=75>
Pulsera Brazalete De Elsa Y Ana De Frozen 11 Con 5 Dijes De Disney \$194MXN <https://resources.sanborns.com.mx/imagenes-sanborns-ii/1200/7503029702663.jpg?scale=250&qilty=75>

	Nombre	Precio	Imagen
0	Barbie La Película Muñeca de Colección Presidente	\$1,699MXN	0 https://resources.sanborns.com.mx/imagenes-san...
1	Barbie Signature Muñeca Los Polinesios Anivers...	\$1,591MXN	1 https://resources.sanborns.com.mx/imagenes-san...
2	Mochila De Dama Barbie X Gorett Backpack Media...	\$951MXN	2 https://resources.sanborns.com.mx/imagenes-san...
3	Patines De Moda Barbie La Película	\$529MXN	3 https://resources.sanborns.com.mx/imagenes-san...
4	Inflable Infantil en Forma de Piña Bigmouth	\$214MXN	4 https://resources.sanborns.com.mx/imagenes-san...
5	Cámara Fujifilm Instax Mini 12 blanca	\$2,239MXN	5 https://resources.sanborns.com.mx/imagenes-san...
6	Camisa blanca GCandila manga corta 5729FB tall...	\$749MXN	6 https://resources.sanborns.com.mx/imagenes-san...
7	Vestido midi calado blanco	\$849MXN	7 https://resources.sanborns.com.mx/imagenes-san...
8	NSW Mario Party Superstars	\$1,289MXN	8 https://resources.sanborns.com.mx/imagenes-san...
9	Free Libre Sensor	\$1,439MXN	9 https://resources.sanborns.com.mx/imagenes-san...
10	Samsung Galaxy Watch 4 40 mm Negro	\$2,999MXN	10 https://resources.sanborns.com.mx/imagenes-san...
11	Cámara Canon kit M50 Mark II EF-M 15-45 mm IS ...	\$13,929MXN	11 https://resources.sanborns.com.mx/imagenes-san...
12	Gente como yo	\$349MXN	12 https://resources.sanborns.com.mx/imagenes-san...
13	Viajar cambiará tu vida	\$328MXN	13 https://resources.sanborns.com.mx/imagenes-san...
14	Claudia Sheinbaum: Presidenta	\$199MXN	14 https://resources.sanborns.com.mx/imagenes-san...
15	Descubre a los más villanos, arte antiestrés	\$199MXN	15 https://resources.sanborns.com.mx/imagenes-san...
16	Samsung Galaxy S23 Ultra 256GB Rosa Telcel R6	\$26,999MXN	16 https://resources.sanborns.com.mx/imagenes-san...
17	Motorola E20 32GB Gris Telcel R9	\$2,499MXN	17 https://resources.sanborns.com.mx/imagenes-san...
18	Xiaomi Redmi Note 9 Pro Gris R9 Telcel	\$9,199MXN	18 https://resources.sanborns.com.mx/imagenes-san...
19	iPhone 11 64GB Blanco R4 (Telcel)	\$9,999MXN	19 https://resources.sanborns.com.mx/imagenes-san...
20	Internet Security- Multidispositivos 1 Año Kas...	\$589MXN	20 https://resources.sanborns.com.mx/imagenes-san...
21	Laptop Honor Magicbook 15" Ryzen 5 16 RAM 512 GB	\$15,299MXN	21 https://resources.sanborns.com.mx/imagenes-san...
22	Microsoft 365 Family	\$1,499MXN	22 https://resources.sanborns.com.mx/imagenes-san...
23	Mouse wireless pebble m350 rosa	\$449MXN	23 https://resources.sanborns.com.mx/imagenes-san...
24	Cuadro De Luz	\$319MXN	24 https://resources.sanborns.com.mx/imagenes-san...
25	Candado De Combinación Números Grandes	\$279MXN	25 https://resources.sanborns.com.mx/imagenes-san...
26	Irrigador Bucal Sonico Interplak	\$599MXN	26 https://resources.sanborns.com.mx/imagenes-san...
27	Botellitas Con Masajeador	\$279MXN	27 https://resources.sanborns.com.mx/imagenes-san...
28	Cámara de video GoPro Hero 11 Black	\$7,299MXN	28 https://resources.sanborns.com.mx/imagenes-san...
29	Cámara Nikon Z50 16-50 y 50-250 mm F/4	\$29,759MXN	29 https://resources.sanborns.com.mx/imagenes-san...
30	CAMARA CANON T7 EF-S 18-55 IS-7	\$14,959MXN	30 https://resources.sanborns.com.mx/imagenes-san...
31	Cámara Sony ILCE-7M3 montura tipo E full frame	\$38,261MXN	31 https://resources.sanborns.com.mx/imagenes-san...
32	Figura Hallmark Mickey Mouse paz y amor	\$1,299MXN	32 https://resources.sanborns.com.mx/imagenes-san...
33	Hallmark taza Harry Potter Hogwarts	\$599MXN	33 https://resources.sanborns.com.mx/imagenes-san...
34	Barco santa Maria Yijie	\$739MXN	34 https://resources.sanborns.com.mx/imagenes-san...
35	Globo negro y verde con base y diámetro de 20 cm	\$1,086MXN	35 https://resources.sanborns.com.mx/imagenes-san...
36	Collar Con Dije De Elsa De Frozen I De Disney	\$160MXN	36 https://resources.sanborns.com.mx/imagenes-san...
37	Pin de Stich playa de Disney	\$160MXN	37 https://resources.sanborns.com.mx/imagenes-san...
38	Bolsa de Minnie de peluche de Disney BP06	\$306MXN	38 https://resources.sanborns.com.mx/imagenes-san...
39	Pulsera Brazaletes De Elsa Y Ana De Frozen 1l C...	\$194MXN	39 https://resources.sanborns.com.mx/imagenes-san...

2. Datos agrupados en un dataframe

	Nombre	Precio	Imagen
1	Barbie La Película Mu...	\$1,699MXN	https://resources.san...
2	Barbie Signature Muñ...	\$1,591MXN	https://resources.san...
3	Mochila De Dama Bar...	\$951MXN	https://resources.san...
4	Patines De Moda Bar...	\$529MXN	https://resources.san...
5	Inflable Infantil en For...	\$214MXN	https://resources.san...
6	Cámara Fujifilm Insta...	\$2,239MXN	https://resources.san...
7	Camisa blanca GCan...	\$749MXN	https://resources.san...
8	Vestido midi calado bl...	\$849MXN	https://resources.san...
9	NSW Mario Party Sup...	\$1,289MXN	https://resources.san...
10	Free Libre Sensor	\$1,439MXN	https://resources.san...
11	Samsung Galaxy Wat...	\$2,999MXN	https://resources.san...
12	Cámara Canon kit M5...	\$13,929MXN	https://resources.san...
13	Gente como yo	\$349MXN	https://resources.san...
14	Viajar cambiará tu vida	\$328MXN	https://resources.san...
15	Claudia Sheinbaum: ...	\$199MXN	https://resources.san...
16	Descubre a los más v...	\$199MXN	https://resources.san...
17	Samsung Galaxy S23...	\$26,999MXN	https://resources.san...
18	Motorola E20 32GB ...	\$2,499MXN	https://resources.san...
19	Xiaomi Redmi Note 9 ...	\$9,199MXN	https://resources.san...
20	iPhone 11 64GB Blan...	\$9,999MXN	https://resources.san...
21	Internet Security- Mult...	\$589MXN	https://resources.san...
22	Laptop Honor Magicb...	\$15,299MXN	https://resources.san...
23	Microsoft 365 Family	\$1,499MXN	https://resources.san...
24	Mouse wireless pebbl...	\$449MXN	https://resources.san...
25	Cuadro De Luz	\$319MXN	https://resources.san...
26	Candado De Combin...	\$279MXN	https://resources.san...
27	Irrigador Bucal Sonic...	\$599MXN	https://resources.san...
28	Botellitas Con Masaje...	\$279MXN	https://resources.san...
29	Cámara de video Go...	\$7,299MXN	https://resources.san...
30	Cámara Nikon Z50 16...	\$29,759MXN	https://resources.san...
31	CAMARA CANON T7...	\$14,959MXN	https://resources.san...
32	Cámara Sony ILCE-7...	\$38,261MXN	https://resources.san...

3. Datos guardados en un csv

Conclusiones:

La practica fue concluida de forma exitosa, se extrajeron los nombres, precios e imágenes de 40 productos de la pagina web de *Sanborns*, sin embargo, no se logro extraer la totalidad de datos de interés debido a que estos no estaban guardados dentro de la misma ruta de HTML que los datos extraídos de forma exitosa.

Los obstáculos encontrados dentro de la practica arrojan incertidumbre respecto a la construcción del código de la pagina web de *Sanborns*, ello sienta las bases de un estudio más detallado acerca del mismo, ello para dar con la ubicación de los valores no obtenidos en el presente trabajo.