

15 de noviembre de 2024



INSTITUTO POLITÉCNICO NACIONAL



*Practica 5  
Request y beautiful soap*

Instituto Politécnico Nacional  
Escuela Superior de Cómputo  
Licenciatura en ciencia de datos

Desarrollo de aplicaciones para el análisis de datos  
Grupo: 4AV1  
Profesora: Sandra Luz Morales Guitrón  
Alumno: Leonardo Hinostroza Loera

## **Índice**

<b><u>INTRODUCCION .....</u></b>	<b><u>3</u></b>
<b><u>DESARROLLO .....</u></b>	<b><u>3</u></b>
<b><u>LIBRERÍA REQUEST Y PROTOCOLO HTTP .....</u></b>	<b><u>3</u></b>
<b><u>WEB SCRAPING UTILIZANDO REQUEST Y BEAUTIFUL SOAP .....</u></b>	<b><u>7</u></b>
<b><u>CONCLUSIONES.....</u></b>	<b><u>12</u></b>

## Introducción

En esta práctica llevaremos a cabo el web scapping utilizando Python con dos herramientas, el Request y Beautiful Soap. Nuestro objetivo principal a desarrollar dentro de esta práctica es extraer información de un sitio web, aprendiéndolo a navegar mediante un árbol DOM para manipular datos en HTML. En esta práctica se abordaron varios conceptos como el HTTP, GET y manejar los encabezados para poder evadir restricciones dentro de los sitios web más actuales.

## Desarrollo

¿Qué es el Request?

Requests es una biblioteca en Python que se utiliza para realizar solicitudes HTTP a servidores web. Con ella, puedes obtener el contenido de una página web (HTML, JSON, archivos u otros formatos) para posteriormente analizarlo y extraer la información que necesitas.

¿Qué es el beautiful Soap?

Es una biblioteca de Python diseñada para analizar (parsear) y extraer datos de documentos HTML o XML de manera eficiente. Es una herramienta clave en el desarrollo de proyectos de web scraping, ya que permite navegar por la estructura del contenido de una página web y extraer información específica con facilidad.

### Librería Request y protocolo HTTP

- Se importa la librería request para ejecutar los protocolos HTTP
- Se define la URL de Amazon y usar el encabezado user-agent para disfrazarlo como una solicitud común al sitio para evitar bloqueos de la misma página
- Se envía la solicitud GET para verificar si fue exitosa la solicitud
- El res.reason y res.ok igual nos confirma que está todo OK
- Con el with open nos ayuda a generar el código HTML sin necesariamente tener una conexión en tiempo real
- Res.cookies muestra las cookies solicitadas por Amazon y res.request.headers muestra la solicitud de encabezados que se envió

15 de noviembre de 2024

```
1 import requests
2 url = "https://www.amazon.es/dp/B079RC2C8B"
3 h = {"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36 Edg/130.0.0.0"}
4 print(h)
5 res = requests.get(url, headers=h)
6 print(res.status_code)
7 print(res.reason)
8 print(res.ok)
9 with open("codigo_200.html", "w", encoding="utf-8") as f:
10     f.write(res.text)
11 print(res.cookies)
12 print(res.request.headers)
✓ [2] 2s 382ms
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36 Edg/130.0.0.0'}
200
OK
True
<RequestsCookieJar[<Cookie session-id=257-5349059-1777039 for .amazon.es/>, <Cookie session-id-time=20827872811 for .amazon.es/>, <Cookie i18n-prefs=EUR for .amazon.es/>]>
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36 Edg/130.0.0.0', 'Accept-Encoding': 'gzip, deflate', 'Accept': '*/*', 'Connection': 'keep-alive'}
```

## Html generado

15 de noviembre de 2024

## Contenido dentro de la pagina

The screenshot shows a product page on Amazon.es. The main product is a 'DAVIDOFF ESSENTIALS Reporter Bag, Bolso bandolera con cremallera y varios compartimentos...' by DAVIDOFF, priced at 175€. To the left, there's a sidebar for 'Moda > Hombre > Bolsos > Bolsos de hombro' with a grid of six thumbnail images of different bags. The main product image is a large, dark grey backpack with red accents and a logo on the front. Below the main image, a smaller image of a grey laptop bag is shown with the text 'Mochila con USB para Laptop de 15.6 Pulgadas de FANDARE...'. The page includes standard Amazon navigation like 'Electrónica', 'Ofertas', and 'Informática'.

Código de ejemplo propio de otro producto de amazon, de igual manera se llevan acabo los pasos realizados en el pasado

```
> 1 import requests
2 url2 = "https://www.amazon.com.mx/gp/sw/d/B8CX25DTBV/?_encoding=UTF8&pd_rd_plhdr=t&aaxitk=1a00292410de665a5e43f86bd69b458&hsa_cr_id=2987533178001&oid=1731722643&sr=1-1-9&e7e56a-6f64-44f-0281
3 h2 = {"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/139.0.0.0 Safari/537.36 Edg/139.0.0.0"}
4 print(h2)
5 res2 = requests.get(url2, headers=h2)
6 print(res2.status_code)
7 print(res2.reason)
8 print(res2.ok)
9 with open("codigoejemplo.html","w", encoding="utf-8") as f:
10     f.write(res2.text)
11 print(res2.cookies)
12 print(res2.request.headers)
    ✓ [3] 160ms
    {'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/139.0.0.0 Safari/537.36 Edg/139.0.0.0'}
200
OK
True
<RequestsCookieJar[]>
{'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/139.0.0.0 Safari/537.36 Edg/139.0.0.0', 'Accept-Encoding': 'gzip, deflate', 'Accept': '*/*', 'Connection': 'keep-alive'}
```

15 de noviembre de 2024

## Html del ejemplo

```
1  <!DOCTYPE html>
2  <!--[if lt IE 7]> <html lang="es-mx" class="a-no-js a-lt-ie9 a-lt-ie8 a-lt-ie7"> <![endif]-->
3  <!--[if IE 7]>   <html lang="es-mx" class="a-no-js a-lt-ie9 a-lt-ie8"> <![endif]-->
4  <!--[if IE 8]>   <html lang="es-mx" class="a-no-js a-lt-ie9"> <![endif]-->
5  <!--[if gt IE 8]><!-->
6  <html class="a-no-js" lang="es-mx"><!--<![endif]><head>
7      <meta http-equiv="content-type" content="text/html; charset=UTF-8">
8      <meta charset="utf-8">
9      <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
10     <title dir="ltr">Amazon.com.mx</title>
11     <meta name="viewport" content="width=device-width">
12     <link rel="stylesheet" href="https://images-na.ssl-images-amazon.com/images/G/01/AUIClients/AmazonUI-3c913031596ca78a3768f4e934b1cc02ce238101.secure.min._V1_.css"/>
13     <script>
14
15         if (true === true) {
16             var ue_t0 = (+ new Date()),
17                 ue_csm = window,
18                 ue = { t0: ue_t0, d: function() { return (+new Date() - ue_t0); } },
19                 ue_furl = "fls-na.amazon.com",
20                 ue_mid = "A1AM78C64UM0Y8",
21                 ue_sid = (document.cookie.match(/session-id=(\w{8}-\w{4}-\w{4}-\w{4}-\w{12})/)) || [] [1],
22                 ue_sn = "opfcaptcha.amazon.com.mx",
23                 ue_id = '1KAWHYCKDYK68M0Q6E32';
24         }
25     </script>
26 </head>
27 <body>
28
29 <!--
30 Para comentar el acceso automatizado a los datos de Amazon, póngase en contacto con api-services-support@amazon.com.
31 Para obtener información sobre la migración a nuestras API, consulta nuestras API de Marketplace en https://developer.amazonservices.com/ref=rn_c_sv o nuestra P
32
33 To discuss automated access to Amazon data please contact api-services-support@amazon.com.
34 For information about migrating to our APIs refer to our Marketplace APIs at https://developer.amazonservices.com/ref=rn_c_sv, or our Product Advertising API at h
35 -->
36
37 <!--
38 Correios.DoNotSend
39 -->
40
```

## Contenido de la pagina

amazon.com.mx Entrega en Gustavo A. Madero, D.F., 07058 Actualizar ubicación Todas las categorías Buscar Amazon.com.mx

Hola, Identificarse Cuenta y Listas Devoluciones y Pedidos

prime Suscríbete por \$99/mes

Todo Vender Los más vendidos Prime AmazonBasics Ofertas del Buen Fin Lo nuevo Música Outlet Electrónicos Los Más Regalados Hogar y Cocina

Apple iPhone iPad Apple Watch AirPods Mac laptops Mac desktops Apple TV Accesorios Beats by Dr. Dre

Del 15 al 18 de noviembre

Considerar estos artículos disponibles

Apple 2024 MacBook Air (de 13 Pulgadas, Chip M3 de Apple con CPU de 8 núcleos y GPU de 8 núcleos, 8GB Memoria unificada, 256 GB) - Azul Medianoche ⚡ \$32,998<sup>00</sup>

Apple 2024 MacBook Air (de 13 Pulgadas, Chip M3 de Apple con CPU de 8 núcleos y GPU de 10 núcleos, 8GB Memoria unificada, 512 GB) - Azul Medianoche ⚡ \$29,698<sup>10</sup>

Apple 2024 MacBook Air (de 13 Pulgadas, Chip M3 de Apple con CPU de 8 núcleos y GPU de 8 núcleos, 8GB Memoria unificada, 256 GB) - Blanco Estelar ⚡ \$26,771<sup>00</sup>

Electrónicos > Computadoras, Componentes y Accesorios > Laptops > Portátiles Tradicionales

Apple 2024 MacBook Air (de 13 Pulgadas, Chip M3 de Apple con CPU de 8 núcleos y GPU de 8 núcleos, 8GB Memoria unificada, 256 GB) - Azul Medianoche

Vista la tienda de Apple

4.5 ★★★★★ 1 calificaciones

Color: Azul medianoche

Pagos y Seguridad Política de devoluciones

No disponible por el momento.

No sabemos si este producto volverá a estar disponible, ni cuándo.

Se entregará a Gustavo Adolfo Me... 07058. Actualizar ubicación

Agregar a la Wish List



## Web scraping utilizando Request y Beautiful soap

- Se importan igual las librerias
- Se define la URL a utilizar: [https://www.mercadolibre.com.mx/kit-de-construccion-lego-spider-man-daily-bugle-76178-3772-piezas-3/p/MLM18592696?pd\\_filters=deal%3AMLM1156918-1#polycard\\_client=search-nordic&wid=MLM1353725906&sid=search&searchVariation=MLM18592696&position=2&search\\_layout=grid&type=product&tracking\\_id=03f20a03-f556-4f34-9b0e-2dfd1c3a0e26](https://www.mercadolibre.com.mx/kit-de-construccion-lego-spider-man-daily-bugle-76178-3772-piezas-3/p/MLM18592696?pd_filters=deal%3AMLM1156918-1#polycard_client=search-nordic&wid=MLM1353725906&sid=search&searchVariation=MLM18592696&position=2&search_layout=grid&type=product&tracking_id=03f20a03-f556-4f34-9b0e-2dfd1c3a0e26)
- Se envia el GET y se manda el user-agent para llevar acabo la solicitud

```
Web scraping

1 import requests
2 from bs4 import BeautifulSoup
[6]

1 url = "https://www.mercadolibre.com.mx/kit-de-construccion-lego-spider-man-daily-bugle-76178-3772-piezas-3/p/MLM18592696?pd_filters=deal%3AMLM1156918-1#polycard_client=search-nordic&wid=MLM1353725906&sid=search&searchVariation=MLM18592696&position=2&search_layout=grid&type=product&tracking_id=03f20a03-f556-4f34-9b0e-2dfd1c3a0e26"
2 h = {"user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36 Edg/130.0.0.0"}
[7]

1 res = requests.get(url, headers=h)
2 res.status_code
[8]
200
```

Al utilizar BS convierte el HTML obtenido en el arbol DOM comentado en la introducción de la práctica el cual permite ser explorado y analizado con mayor facilidad, es muy extenso

```
> 1 soup = BeautifulSoup(res.text, "html.parser")      soup      res
2 print(soup)
[8] 47ms

<!DOCTYPE html>

<html lang="es-MX">
<head><link href="https://www.google-analytics.com" rel="preconnect"/><link href="https://www.google.com" rel="preconnect"/><link href="https://data.mercadolibre.com" rel="preconnect"/><link href="https://http2.mlstatic.com" rel="preconnect"/><link href="https://stats.g.doubleclick.net" rel="preconnect"/><link href="https://analytics.mercadolibre.com.mx" rel="preconnect"/><link href="https://www.google.com.mx" rel="preconnect"/><script nonce="19zyln7S8noq2DcZkgZ/g=" type="text/javascript">window.NREUM||(NREUM={});NREUM.info={};(NREUM.info={"agent":"","beacon":"bam.nr-data.net","errorBeacon":"bam.nr-data.net","licenseKey":"NRBR-8547a290c864571ffcc","applicationID":"1588198355","agentToken":null,"applicationTime":378.650828,"transactionName":"bgQ0MEcFXXJZBKYNWLd0JBxFFLVCSw98S3J8NU5LRUsKQuokVhFWTSgFT64JTHxLPNJT3hM0z10Vk1qTRlnQpiUUxdaE8Z","queueTime":0,"tttGuid":"ed806ebca49936ba"};(window.NREUM||(NREUM={})).init={privacy:{cookies_enabled:true},ajax:{deny_list:[{"bam.nr-data.net"}]},distributed_tracing:{enabled:true}};(window.NREUM||(NREUM={})).loader_config={agentID:"1834653436",accountID:"89586",trustKey:"1709707",xpId:"XQ40VF5GwTFKF1TDgkEU1E=",licenseKey:"NRBR-8547a290c864571ffcc",applicationID:"1588198355"};;/*! For license information please see nr-loader-spa-1.273.0.min.js.LICENSE.txt */
```

Esta parte del código hace que regrese el title del html y el title.text te da el texto dentro del mismo elemento ya comentado

```
> 1 soup.title      soup
2 soup.title.text
[9] < 10 ms

'Kit De Construcción Lego Spider-Man Daily Bugle 76178 3772 Piezas 3+ | Meses sin interés'
```

## 15 de noviembre de 2024

Aquí visualizamos el html de esa parte y vemos que coincide

```
▼ <div class="ui-pdp-header__title-container"> flex
  <h1 class="ui-pdp-title">Kit De Construcción Lego Spider-Man Daily Bugle 76178 3772 Piezas 3+</h1> = $0
```

Usamos el dir el cual nos muestra todos atributos y funciones del objeto al cual se le aplico el beautiful soap, se utlizara find para ver las etiquetas a utilizar

```
> 1  dir(soup)      soup
  ✓ [10] 10ms
    'reset',
    'select',
    'select_one',
    'self_and_descendants',
    'setup',
    'smooth',
    'string',
    'string_container',
    'string_container_stack',
    'strings',
    'stripped_strings',
    'tagStack',
    'text',
    'unwrap',
    'wrap']
```

Aquí se procede a buscar el elemento h1 con su respectiva clase y nos manda la informacion

```
1  soup.find("h1", class_="ui-pdp-title").text
  ✓ [11] < 10 ms
    'Kit De Construcción Lego Spider-Man Daily Bugle 76178 3772 Piezas 3+'
  ▼ <div class="ui-pdp-header__title-container"> flex
    <h1 class="ui-pdp-title">Kit De Construcción Lego Spider-Man Daily Bugle 76178 3772 Piezas 3+</h1>
```

Aquí se busca el div el cual contiene el precio anterior, luego lo extrae, limpia y lo pasa al precio actual y span termina de eliminar comas y simbolos para dar un dato limpio.

### PRECIO ANTERIOR

```
1  soup.find("div", class_="ui-pdp-price__second-line")
  ✓ [12] 13ms
    <div class="ui-pdp-price__second-line"><span data-testid="price-part"><span aria-label="5599 pesos" aria-roledescription="Precio" class="andes-money-amount , ui-pdp-price__part andes-money-amount--cents-superscript andes-money-amount--compact" itemprop="offers" itemscope="" itemtype="http://schema.org/Offer" , role="img" style="font-size:36px"><meta content="5599" itemprop="price"><span aria-hidden="true" class="andes-money-amount__currency-symbol" , itemprop="priceCurrency">$</span><span aria-hidden="true" class="andes-money-amount__fraction">5,599</span></span><span , class="ui-pdp-price__second-line__label ui-pdp-color--GREEN ui-pdp-size--MEDIUM ui-pdp-family--REGULAR"><span class="andes-money-amount__discount , ui-pdp-family--REGULAR" style="font-size:18px">35% OFF</span></span></div>
  ?
> 1  precio_anter = soup.find("span", class_="andes-money-amount__fraction").text.replace(",","")      precio_anter      soup
2  precio_anter = float(precio_anter)
3  print(precio_anter)
  ✓ [13] < 10 ms
    8699.0
```

```
▶ <span data-testid="price-part">...</span>
```

## 15 de noviembre de 2024

Aquí se busca el div el cual contiene el precio actual, luego lo extrae, limpia y lo pasa al precio actual y span termina de eliminar comas y simbolos para dar un dato limpio, o sea lo mismo pero para el actual

```
Precio Normal

1 precio = soup.find("span", class_="andes-money-amount ui-pdp-price__part andes-money-amount--cents-superscript andes-money-amount--compact").text.replace(",","");
2 precio.replace("$", "")    precio      soup
3 precio_float = float(precio)
4 print(precio_float)
✓ [14] < 10 ms
5599
5599.0

► <div class="ui-pdp-price__second-line">...</div> flex
► <div class="ui-pdp-price__subtitles">...</div> flex
```

Se busca el div el cual contiene los ranking de los usuarios y se devuelve completo y con el span se extrae la calificación promedio del ranking el cual nos manda 4.9

```
1 soup.find("div" , class_="ui-pdp-header__info")
✓ [16] 32ms
<div class="ui-pdp-header__info"><a class="ui-pdp-review__label ui-pdp-review__label--link" href="/noindex/catalog/reviews/MLM18592695?noIndex=true&access=view_all&modal=true&controlled=true"><span aria-hidden="true" class="ui-pdp-review__rating">4.9</span><span aria-hidden="true" , class="ui-pdp-review__ratings"><svg class="ui-pdp-icon ui-pdp-icon--star-full" fill="none" height="10" viewBox="0 0 10 10" width="10"><use , href="#poly_star_fill"></use></svg><svg class="ui-pdp-icon ui-pdp-icon--star-full" fill="none" height="10" viewBox="0 0 10 10" width="10"><use , href="#poly_star_fill"></use></svg><svg class="ui-pdp-icon ui-pdp-icon--star-full" fill="none" height="10" viewBox="0 0 10 10" width="10"><use , href="#poly_star_fill"></use></svg><svg class="ui-pdp-icon ui-pdp-icon--star-full" fill="none" height="10" viewBox="0 0 10 10" width="10"><use , href="#poly_star_fill"></use></svg><span class="andes-visually-hidden">Calificación 4.9 de 5. 129 opiniones.</span><span aria-hidden="true" , class="ui-pdp-review__amount">(129)</span></a></div>

1 soup.find("span" , class_="ui-pdp-review__rating").text
✓ [15] 20ms
'4.9'
▼ <a href="/noindex/catalog/reviews/MLM18592695?noIndex=true&access=view_all&modal=true&controlled=true" class="ui-pdp-review__label ui-pdp-review__label--link"> (flex)
    <span aria-hidden="true" class="ui-pdp-review__rating">4.9
    </span> == $0
► <span aria-hidden="true" class="ui-pdp-review__ratings">(...)</span>
```

Aquí como se comenta se encuentra la imagen del producto con la clase ui-pdp-image ui-pdp-gallery\_\_figure\_\_image y con el get src se obtiene el link de la imagen del producto

```
1 soup.find(class_ = "ui-pdp-image ui-pdp-gallery__figure__image").attrs
✓ [17] 12ms
{'data-zoom': 'https://http2.mlstatic.com/D_NQ_NP_790191-MLA48251421228_112021-F.webp',
 'data-index': '0',
 'width': '410',
 'height': '500',
 'fetchpriority': 'high',
 'decoding': 'sync',
 'src': 'https://http2.mlstatic.com/D_NQ_NP_790191-MLA48251421228_112021-0.webp',
 'srcset': 'https://http2.mlstatic.com/D_NQ_NP_2X_790191-MLA48251421228_112021-F.webp 2x',
 'class': ['ui-pdp-image', 'ui-pdp-gallery__figure__image'],
 'alt': 'Kit De Construcción Lego Spider-Man Daily Bugle 76178 3772 Piezas 3+'}

1 soup.find(class_ = "ui-pdp-image ui-pdp-gallery__figure__image").attrs.get("src")
✓ [18] < 10 ms
'https://http2.mlstatic.com/D_NQ_NP_790191-MLA48251421228_112021-0.webp'
▶ <label for="gallery-thumbnail-790191-MLA48251421228_112021" class="u
  i-pdp-gallery__label">(...)</label> (flex)
▼ <figure class="ui-pdp-gallery__figure"> (flex)
   == $0
</figure>
</span>
```

## 15 de noviembre de 2024

Ya para finalizar se extraen las características principales del producto con div y la clase correspondiente con su descripción del mismo

```
Characterísticas del producto

1 soup.find("ul", class_="ui-vpp-highlighted-specs__features-list").text
✓ [19] < 10 ms
'Cantidad de piezas: 3772Construcción recomendada: edificio daily bugle.Figura de Doctor Octopus incluida.El set contiene 3772 piezas de plástico.'

1 result = soup.find("ul", class_="ui-vpp-highlighted-specs__features-list").text
2 formatted_result = result.replace(".", ".\n")
3 print(formatted_result)
✓ [20] < 10 ms
Cantidad de piezas: 3772Construcción recomendada: edificio daily bugle.
Figura de Doctor Octopus incluida.
El set contiene 3772 piezas de plástico.

1 soup.find("div", class_ = "ui-pdp-description pl-24 pr-24 ui-pdp-collapsible--is-collapsed").text
✓ [21] < 10 ms
'DescripciónEl set Lego Marvel Spider-Man Daily Bugle (76178) reúne a 25 personajes clásicos del multiverso arácnido en un sensacional proyecto de , construcción y exposición para adultos.Con sus aproximadamente 82 cm de altura, esta recreación de 3772 piezas del Daily Bugle ofrece un imponente decorado, para las aventuras de un reparto estelar de héroes y villanos Marvel, ¡muchos de ellos novedad en este set! El edificio está repleto de detalles y , fascinantes características, desde un taxi neoyorquino hasta el despacho en el ático del redactor jefe del Bugle. Para los entusiastas de Marvel, la , construcción de la maqueta será a la vez desafiante y muy gratificante. Cuando la terminen, las plantas, azotea y fachadas desmontables dejarán ver el , interior de este tributo a Marvel para que todos lo admiren.-Sumérgete en los sets Lego para fans adultos de la construcción creativa o la cultura de los , cómics.-Contiene minifiguras que dan vida al set como el Doctor Octopus, Spider-Man, Venom, Miles Morales, Spider-Ham, el Duende Verde, Peter Parker, , Spider-Gwen, Mysterio, Sandman, Robbie Robertson y muchas más.-Date el gusto de utilizar tus superpoderes de construcción creativa con esta recreación de , 3772 piezas de las oficinas del Daily Bugle, repleta de detalles auténticos, fascinantes características y la espectacular acción de los superhéroes.-Con , unas medidas aproximadas de 82 cm de altura, 27 cm de anchura y 27 cm de profundidad, la impresionante maqueta está llena hasta la azotea de personajes , clásicos, características de gran realismo y divertidos accesorios.Aviso legal. Recomendado a partir de 3 años.'
  Características > tex
    <h2 class="ui-vpp-text-alignment--left ui-pdp-color--BLACK ui-pdp-size--XSMALL ui-pdp-family--SEMIBOLD highlighted-features-title">Lo que
      tienes que saber de este producto</h2>
    </div>
    ▼<div class="ui-pdp-container__row ui-pdp-container__row--highlighted-features"> (flex)
      ▼<div class="ui-vpp-highlighted-specs__features">
        ▼<ul class="ui-vpp-highlighted-specs__features-list">
          ▼<li class="ui-vpp-highlighted-specs__features-list-item ui-pdp-color--BLACK ui-pdp-size--XSMALL ui-pdp-family--REGULAR">
            ::before
            "Cantidad de piezas: 3772"
          </li>
          ▼<li class="ui-vpp-highlighted-specs__features-list-item ui-pdp-color--BLACK ui-pdp-size--XSMALL ui-pdp-family--REGULAR"> == $0
            ::before
            "Construcción recomendada: edificio daily bugle."
          </li>
```

```
▼ <div class="ui-pdp-collapsible ui-pdp-description-collapse ui-pdp-collapsable--is-collapsed">
  ▼ <div class="ui-pdp-collapsible__container" style="max-height: 400px;">
    ▼ <div class="ui-pdp-description pl-24 pr-24 ui-pdp-collapsible--is-collapsed">
      <h2 class="ui-pdp-description__title">Descripción</h2>
      ▼ <p class="ui-pdp-description__content"> == $0
        "El set Lego Marvel Spider-Man Daily Bugle (76178) reúne a 25 personajes clásicos del multiverso arácnido en un sensacional proyecto de construcción y exposición para adultos."
        <br>
        <br>
        "Con sus aproximadamente 82 cm de altura, esta recreación de 3772 piezas del Daily Bugle ofrece un imponente decorado para las aventuras de un reparto estelar de héroes y villanos Marvel, ¡muchos de ellos novedad en este set! El edificio está repleto de detalles y fascinantes características, desde un taxi neoyorquino hasta el despacho en el ático del redactor jefe del Bugle. Para los entusiastas de Marvel, la construcción de la maqueta será a la vez desafiante y muy gratificante. Cuando la terminen, las plantas, azotea y fachadas desmontables dejarán ver el interior de este tributo a Marvel para que todos lo admiren."
        <br>
        <br>
        "-Sumérgete en los sets Lego para fans adultos de la construcción creativa o la cultura de los cómics."
        <br>
        <br>
```

## Conclusiones

Esta práctica me ayudó a entender cómo usar las herramientas **Requests** y **BeautifulSoup** para hacer web scraping, que básicamente es extraer información de páginas web. Aprendí a manejar el protocolo HTTP, a usar encabezados para evitar bloqueos y a trabajar con el árbol DOM para encontrar y organizar datos.

Lo que más me gustó fue aprender a buscar elementos específicos en una página, como precios, calificaciones e imágenes, y limpiarlos para que los datos queden ordenados. Fue interesante ver cómo se puede extraer información útil de un sitio web de forma estructurada.

En general, fue una actividad bastante laboriosa pero útil para entender cómo se trabaja con datos en formato HTML.