

```
In [1]: from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from bs4 import BeautifulSoup
import pandas as pd
import time
```

```
In [5]: driver = webdriver.Chrome(r'C:\Users\DELL\Desktop\WEB_SCRAPING\chromedriver.exe')
driver.get('https://www.thisiswhyimbroke.com/new/')
```

C:\Users\DELL\AppData\Local\Temp\ipykernel_13216\2742691935.py:1: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
driver = webdriver.Chrome(r'C:\Users\DELL\Desktop\WEB_SCRAPING\chromedriver.exe')

```
In [6]: last_height = driver.execute_script('return document.body.scrollHeight')
```

```
In [7]: while True:
        driver.execute_script('window.scrollTo(0, document.body.scrollHeight)')
        time.sleep(4)
        new_height = driver.execute_script('return document.body.scrollHeight')
        if new_height == last_height:
            break
        last_height = new_height

soup = BeautifulSoup(driver.page_source, 'lxml')
product_card = soup.find_all('div', {"class": 'ng-scope', "ng-if": "::post.type == 'post'"})
df=pd.DataFrame({'Product_name':[], 'Product_link':[], 'Product_price':[], 'Product_
for product in product_card:
    try:
        product_name = product.find('a', class_='ng-binding ng-scope').text
        product_link = product.find('a', class_='ng-binding ng-scope').get('href')
        product_price = product.find('div', class_='price ng-binding').text
        product_description = product.find('p', class_='ng-binding').text
        product_image = product.find('img', class_='img-fixed lazy-image--handled').get('s
        product_sales = product.find('a', class_='saved-count ng-binding').text
        df= df.append({'Product_name':product_name, 'Product_link':product_link, 'Product_
    except:
        pass
```

```
-----
KeyboardInterrupt                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_13216\593021277.py in <module>
      1 while True:
      2     driver.execute_script('window.scrollTo(0, document.body.scrollHeight)')
----> 3     time.sleep(4)
      4     new_height = driver.execute_script('return document.body.scrollHeight')
      5     if new_height == last_height:
```

KeyboardInterrupt:

```
In [19]: soup = BeautifulSoup(driver.page_source, 'lxml')
product_card = soup.find_all('div', {"class": 'ng-scope', "ng-if": "::post.type == 'post'"})
df=pd.DataFrame({'Product_name':[], 'Product_link':[], 'Product_price':[], 'Product_
for product in product_card:
    try:
        product_name = product.find('a', class_='ng-binding ng-scope').text
        product_link = product.find('a', class_='ng-binding ng-scope').get('href')
        product_price = product.find('div', class_='price ng-binding').text
        product_description = product.find('p', class_='ng-binding').text
        product_image = product.find('img', class_='img-fixed lazy-image--handled').get('s
```

```
product_sales = product.find('a', class_='saved-count ng-binding').text
df= df.append({'Product_name':product_name, 'Product_link':product_link, 'Product_
except:
    pass

df
```

Out[19]:

	Product_name	Product_link	Product_price	Product_description	Product_image
0					
1	How to Smoke Pot (Properly)	/how-to-smoke-pot-properly/	\$16.66	Get on the fast track to becoming a cannabis c...	https://cdn.thisiswhyimbroke.com/thumb/how-to-..
2	Woolly Mammoth Meatball	/woolly-mammoth-meatball/		Behold the world's first woolly mammoth meatba...	https://cdn.thisiswhyimbroke.com/thumb/woolly-..
3	Elements Range Wood Fired Hot Tub	/koto-design-elements-range-wood-fired-hot-tub/	\$7,712.19	Make cold days something to look forward to wi...	https://cdn.thisiswhyimbroke.com/thumb/koto-de..
4	Disney Castle Playhouse	/disney-castle-playhouse/	\$500,000.00	This Disney castle playhouse will make your li...	https://cdn.thisiswhyimbroke.com/thumb/disney-..
...
1659	OnlyCat Smart Cat Flap	/onlycat-smart-cat-flap/		The OnlyCat smart cat flap will keep unwanted ...	https://cdn.thisiswhyimbroke.com/thumb/onlycat..
1660	Eucalyptus Shower Bunch	/eucalyptus-shower-bunch/	\$12.99	Hang up this eucalyptus bunch and transform yo...	https://cdn.thisiswhyimbroke.com/thumb/eucalyp..
1661	Battle Approved Stomper Electric ATV	/battle-approved-stomper-electric-atv/		Conquer any terrain that comes your way in an ...	https://cdn.thisiswhyimbroke.com/thumb/battle-..
1662	Book In A Bag Kit	/book-in-a-bag-kit/	\$24.99	Book in a bag kit will aid the writer in your ...	https://cdn.thisiswhyimbroke.com/thumb/book-in..
1663	Pyramid PC Case	/pyramid-pc-case/	\$389.99	Set your rig apart from the rest with this bol...	https://cdn.thisiswhyimbroke.com/thumb/pyramid..

1664 rows × 6 columns

```
In [50]: df.to_excel(r'C:\Users\DELL\Desktop\thisiswhyimbroke_new.xlsx')

-----
AttributeError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_13216\4128216356.py in <module>
----> 1 df.to_excel(r'C:\Users\DELL\Desktop\thisiswhyimbroke_new.xlsx')

AttributeError: 'NoneType' object has no attribute 'to_excel'
```

```
In [22]: df = df.dropna()  
df
```

Out[22]:

	Product_name	Product_link	Product_price	Product_description	Product_image
0					
1	How to Smoke Pot (Properly)	/how-to-smoke-pot-properly/	\$16.66	Get on the fast track to becoming a cannabis c...	https://cdn.thisiswhyimbroke.com/thumb/how-to-..
2	Woolly Mammoth Meatball	/woolly-mammoth-meatball/		Behold the world's first woolly mammoth meatba...	https://cdn.thisiswhyimbroke.com/thumb/woolly-..
3	Elements Range Wood Fired Hot Tub	/koto-design-elements-range-wood-fired-hot-tub/	\$7,712.19	Make cold days something to look forward to wi...	https://cdn.thisiswhyimbroke.com/thumb/koto-de..
4	Disney Castle Playhouse	/disney-castle-playhouse/	\$500,000.00	This Disney castle playhouse will make your li...	https://cdn.thisiswhyimbroke.com/thumb/disney-..
...
1659	OnlyCat Smart Cat Flap	/onlycat-smart-cat-flap/		The OnlyCat smart cat flap will keep unwanted ...	https://cdn.thisiswhyimbroke.com/thumb/onlycat..
1660	Eucalyptus Shower Bunch	/eucalyptus-shower-bunch/	\$12.99	Hang up this eucalyptus bunch and transform yo...	https://cdn.thisiswhyimbroke.com/thumb/eucalyp..
1661	Battle Approved Stomper Electric ATV	/battle-approved-stomper-electric-atv/		Conquer any terrain that comes your way in an ...	https://cdn.thisiswhyimbroke.com/thumb/battle-..
1662	Book In A Bag Kit	/book-in-a-bag-kit/	\$24.99	Book in a bag kit will aid the writer in your ...	https://cdn.thisiswhyimbroke.com/thumb/book-in..
1663	Pyramid PC Case	/pyramid-pc-case/	\$389.99	Set your rig apart from the rest with this bol...	https://cdn.thisiswhyimbroke.com/thumb/pyramid..

1664 rows × 6 columns

```
In [47]: #df['Product_price']=df['Product_price'].apply(lambda x: x.strip('$'))  
#df['Product_price'] =df['Product_price'].apply(lambda x: float(x.replace('.', '')))  
df['Product_price']  
  
-----  
TypeError                                Traceback (most recent call last)  
~\AppData\Local\Temp\ipykernel_13216\2349583385.py in <module>  
      1 #df['Product_price']=df['Product_price'].apply(lambda x: x.strip('$'))  
      2 #df['Product_price'] =df['Product_price'].apply(lambda x: float(x.replace('.', '')))  
----> 3 df['Product_price']  
  
TypeError: 'NoneType' object is not subscriptable
```

```
In [49]: df
```

```
In [17]: soup = BeautifulSoup(driver.page_source, 'lxml')
product_card = soup.find_all('div', {"class":'ng-scope',"ng-if":"",":post.type == 'post'"))
#df=pd.DataFrame({'Product_name':[], 'Product_link':[], 'Product_price':[], 'Product_
df=pd.DataFrame({'Product_name':[], 'Product_link':[], 'Product_price':[], 'Product_
for product in product_card:
    try:
        product_name = product.find('a', class_='ng-binding ng-scope').text
        product_link = product.find('a', class_='ng-binding ng-scope').get('href')
        product_price = product.find('div', class_='price ng-binding').text
        product_description = product.find('p', class_='ng-binding').text
        product_image = product.find('img', class_='img-fixed lazy-image--handled').get('s
        product_sales = product.find('a', class_='saved-count ng-binding').text
        df= df.append({'Product_name':product_name, 'Product_link':product_link, 'Product_
    except:
        pass
df
```

Out[17]:

	Product_name	Product_link	Product_price	Product_description	Product_image	Product_sales
0						

```
In [50]: df.to_excel(r'C:\Users\DELL\Desktop\thisiswhyimbroke.xlsx')
```

```
In [14]: soup = BeautifulSoup(driver.page_source, 'lxml')
product_card = soup.find_all('div', {"class":'ng-scope',"ng-if":"",":post.type == 'post'"))
soup.find('img', class_='img-fixed lazy-image--handled').get('src')
```

Out[14]: 'https://cdn.thisiswhyimbroke.com/thumb/how-to-smoke-pot-properly_400x333.jpg'

```
In [51]: import pandas as pd
```

```
In [52]: df = pd.read_excel(r'C:\Users\DELL\Desktop\thisiswhyimbroke_new.xlsx')
df
```

Out[52]:

	Unnamed: 0	Product_name	Product_link	Product_price	Product_description	Pr
0	0	NaN	NaN	NaN	NaN	
1	1	How to Smoke Pot (Properly)	/how-to-smoke-pot-properly/	\$16.66	Get on the fast track to becoming a cannabis c...	https://cdn.thisiswhyimbroke.com/thu
2	2	Woolly Mammoth Meatball	/woolly-mammoth-meatball/		Behold the world's first woolly mammoth meatba...	https://cdn.thisiswhyimbroke.com/thu
3	3	Elements Range Wood Fired Hot Tub	/koto-design-elements-range-wood-fired-hot-tub/	\$7,712.19	Make cold days something to look forward to wi...	https://cdn.thisiswhyimbroke.com/thu
4	4	Disney Castle Playhouse	/disney-castle-playhouse/	\$500,000.00	This Disney castle playhouse will make your li...	https://cdn.thisiswhyimbroke.com/thu

Unnamed: 0	Product_name	Product_link	Product_price	Product_description	Pr
...
1659	OnlyCat Smart Cat Flap	/onlycat-smart-cat-flap/		The OnlyCat smart cat flap will keep unwanted ...	https://cdn.thisiswhyimbroke.com/thu
1660	Eucalyptus Shower Bunch	/eucalyptus-shower-bunch/	\$12.99	Hang up this eucalyptus bunch and transform yo...	https://cdn.thisiswhyimbroke.com/thu
1661	Battle Approved Stomper Electric ATV	/battle-approved-stomper-electric-atv/		Conquer any terrain that comes your way in an ...	https://cdn.thisiswhyimbroke.com/th
1662	Book In A Bag Kit	/book-in-a-bag-kit/	\$24.99	Book in a bag kit will aid the writer in your ...	https://cdn.thisiswhyimbroke.com/thu
1663	Pyramid PC Case	/pyramid-pc-case/	\$389.99	Set your rig apart from the rest with this bol...	https://cdn.thisiswhyimbroke.com/thu

1664 rows × 7 columns

```
In [63]: type(df['Product_price'][1])
```

```
Out[63]: str
```

```
In [61]: df['Product_price'] = df['Product_price'].astype(str).apply(lambda x: x.strip('$'))
```

```
Out[61]: 0          nan
1         16.66
2
3        7,712.19
4       500,000.00
...
1659
1660         12.99
1661
1662         24.99
1663        389.99
Name: Product_price, Length: 1664, dtype: object
```

```
In [65]: df['Product_price'] = df['Product_price'].apply(lambda x: x.strip(' \xa0'))
```

```
In [71]: rows_to_drop = df.loc[df.eq('').any(axis=1)]
df = df.drop(rows_to_drop.index)
```

```
In [75]: df.to_excel(r'C:\Users\DELL\Desktop\thisiswhyimbroke_new1.xlsx')
```

```
In [89]: df = pd.read_excel(r'C:\Users\DELL\Desktop\thisiswhyimbroke_new1.xlsx')
df
```

```
Out[89]: Product_name  Product_link  Product_price  Product_description  Product_image
```

	Product_name	Product_link	Product_price	Product_description	Product_image
0	How to Smoke Pot (Properly)	/how-to-smoke-pot-properly/	16.66	Get on the fast track to becoming a cannabis c...	https://cdn.thisiswhyimbroke.com/thumb/how-to-..
1	Elements Range Wood Fired Hot Tub	/koto-design-elements-range-wood-fired-hot-tub/	7,712.19	Make cold days something to look forward to wi...	https://cdn.thisiswhyimbroke.com/thumb/koto-de..
2	Disney Castle Playhouse	/disney-castle-playhouse/	500,000.00	This Disney castle playhouse will make your li...	https://cdn.thisiswhyimbroke.com/thumb/disney-..
3	Mini Desktop Cannon	/mini-desktop-cannon/	15.00	Make your dream of owning a functioning cannon...	https://cdn.thisiswhyimbroke.com/thumb/mini-de..
4	Anker 737 MagGo Charger	/anker-737-maggo-charger/	139.99	Do away with cables and charge your device the...	https://cdn.thisiswhyimbroke.com/thumb/anker-7..
...
1490	The Deck Of Many Things	/the-deck-of-many-things/	15.99	Help your favorite DM torture his players prop...	https://cdn.thisiswhyimbroke.com/thumb/the-dec..
1491	Bandit9 Monaco Electric Sports Car	/bandit9-monaco-electric-sports-car/	149,911.50	Rediscover the joy of driving behind the wheel...	https://cdn.thisiswhyimbroke.com/thumb/bandit9..
1492	Eucalyptus Shower Bunch	/eucalyptus-shower-bunch/	12.99	Hang up this eucalyptus bunch and transform yo...	https://cdn.thisiswhyimbroke.com/thumb/eucalyp..
1493	Book In A Bag Kit	/book-in-a-bag-kit/	24.99	Book in a bag kit will aid the writer in your ...	https://cdn.thisiswhyimbroke.com/thumb/book-in..
1494	Pyramid PC Case	/pyramid-pc-case/	389.99	Set your rig apart from the rest with this bol...	https://cdn.thisiswhyimbroke.com/thumb/pyramid..

1495 rows × 6 columns

```
In [90]: type(df['Product_description'][1000])
```

```
Out[90]: str
```

```
In [93]: df['Product_description'] = df['Product_description'].apply(lambda x: x[:80])
df['Product_description']
```

```
Out[93]: 0      Get on the fast track to becoming a cannabis c...
1      Make cold days something to look forward to wi...
2      This Disney castle playhouse will make your li...
3      Make your dream of owning a functioning cannon...
4      Do away with cables and charge your device the...
...
1490   Help your favorite DM torture his players prop...
```

```
1491     Rediscover the joy of driving behind the wheel...
1492     Hang up this eucalyptus bunch andtransform yo...
1493     Book in a bag kit will aid the writer in your ...
1494     Set your rig apart from the rest with this bol...
Name: Product_description, Length: 1495, dtype: object
```

In []: