```python
import requests
from bs4 import BeautifulSoup
```

```python
response= requests.get("https://books.toscrape.com/")
```

```python
print(response)
```

```
<Response [200]>
```

```python
response.text
```

```
'<!DOCTYPE html>\n<!--[if lt IE 7]>      <html lang="en-us" class="no-js lt-ie9 lt-i
e8 lt-ie7"> <![endif]-->\n<!--[if IE 7]>        <html lang="en-us" class="no-js lt-
ie9 lt-ie8"> <![endif]-->\n<!--[if IE 8]>         <html lang="en-us" class="no-js lt
-ie9"> <![endif]-->\n<!--[if gt IE 8]><!--> <html lang="en-us" class="no-js"> <!--<!
[endif]-->\n    <head>\n        <title>\n    All products | Books to Scrape - Sandbo
x\n</title>\n\n        <meta http-equiv="content-type" content="text/html; charset=U
TF-8" />\n        <meta name="created" content="24th Jun 2016 09:29" />\n        <me
```

```python
response.headers
```

```
{'Date': 'Tue, 19 Dec 2023 14:45:14 GMT', 'Content-Type': 'text/html', 'Content-Length': '51294', 'Connection': 'keep-alive',
'Last-Modified': 'Wed, 08 Feb 2023 21:02:32 GMT', 'ETag': '"63e40de8-c85e"', 'Accept-Ranges': 'bytes', 'Strict-Transport-Security':
'max-age=0; includeSubDomains; preload'}
```

```python
html_data=response.text
```

```python
print(html_data)
```

```
        </body>
    </html>
```

```python
data=BeautifulSoup(html_data, "html.parser")
```

```python
print(data.prettify)
```

```
    <div class="product_price">
    <p class="price_color">Â£45.17</p>
    <p class="instock availability">
    <i class="icon-ok"></i>

            In stock

    </p>
    <form>
    <button class="btn btn-primary btn-block" data-loading-text="Adding..." type="submit">Add to basket</button>
    </form>
    </div>
    </article>
    </li>
    </ol>
    <div>
    <ul class="pager">
    <li class="current">

                    Page 1 of 50

            </li>
    <li class="next"><a href="catalogue/page-2.html">next</a></li>
    </ul>
    </div>
    </div>
    </section>
    </div>
    </div><!-- /row -->
    </div><!-- /page_inner -->
    </div><!-- /container-fluid -->
    <footer class="footer container-fluid">
    </footer>
    <!-- jQuery -->
    <script src="http://ajax.googleapis.com/ajax/libs/jquery/1.9.1/jquery.min.js"></script>
    <script>window.jQuery || document.write('<script src="static/oscar/js/jquery/jquery-1.9.1.min.js"><\/script>')</script>
    <!-- Twitter Bootstrap -->
    <script src="static/oscar/js/bootstrap3/bootstrap.min.js" type="text/javascript"></script>
    <!-- Oscar -->
    <script charset="utf-8" src="static/oscar/js/oscar/ui.js" type="text/javascript"></script>
    <script charset="utf-8" src="static/oscar/js/bootstrap-datetimepicker/bootstrap-datetimepicker.js" type="text/javascript"></scrip
    <script charset="utf-8" src="static/oscar/js/bootstrap-datetimepicker/locales/bootstrap-datetimepicker.all.js" type="text/javascr
    <script type="text/javascript">
                $(function() {


        oscar.init();

        oscar.search.init();

            });
        </script>
    <!-- Version: N/A -->
    </body>
    </html>
    >
```

```python
data.title.string
```

```
    '\n    All products | Books to Scrape - Sandbox\n'
```

```python
data.a
```

```
    <a href="index.html">Books to Scrape</a>
```

```python
data.a.string
```

```
    'Books to Scrape'
```

```python
data.header.a.string
```

```
    'Books to Scrape'
```

```python
b1= data.find(class_="product_pod")
```

```python
b1.h3.a
```

```
    <a href="catalogue/a-light-in-the-attic_1000/index.html" title="A Light in the Attic">A Light in the ...</a>
```

```python
b1.h3.a
```

```
    <a href="catalogue/a-light-in-the-attic_1000/index.html" title="A Light in the Attic">A Light in the ...</a>
```

```python
b1.h3.a['title']
```

```
    'A Light in the Attic'
```

```python
b1.h3.a['href']
```

```
    'catalogue/a-light-in-the-attic_1000/index.html'
```

```python
print('https://books.toscrape.com/'+b1.h3.a['href']) #base url+ book url
```

```
    https://books.toscrape.com/catalogue/a-light-in-the-attic_1000/index.html
```

```python
books_li=data.find_all(class_='product_pod')
```

```python
print(len(books_li))
```

```
    20
```

```python
base_url='https://books.toscrape.com/'
books_urls=[]
for i in books_li:
  books_urls.append(base_url+i.h3.a['href'])
```

```python
books_urls
```

```
    ['https://books.toscrape.com/catalogue/a-light-in-the-attic_1000/index.html',
     'https://books.toscrape.com/catalogue/tipping-the-velvet_999/index.html',
     'https://books.toscrape.com/catalogue/soumission_998/index.html',
     'https://books.toscrape.com/catalogue/sharp-objects_997/index.html',
     'https://books.toscrape.com/catalogue/sapiens-a-brief-history-of-humankind_996/index.html',
     'https://books.toscrape.com/catalogue/the-requiem-red_995/index.html',
     'https://books.toscrape.com/catalogue/the-dirty-little-secrets-of-getting-your-dream-job_994/index.html',
     'https://books.toscrape.com/catalogue/the-coming-woman-a-novel-based-on-the-life-of-the-infamous-feminist-victoria-
    woodhull_993/index.html',
     'https://books.toscrape.com/catalogue/the-boys-in-the-boat-nine-americans-and-their-epic-quest-for-gold-at-the-1936-berlin-
    olympics_992/index.html',
     'https://books.toscrape.com/catalogue/the-black-maria_991/index.html',
     'https://books.toscrape.com/catalogue/starving-hearts-triangular-trade-trilogy-1_990/index.html',
     'https://books.toscrape.com/catalogue/shakespeares-sonnets_989/index.html',
     'https://books.toscrape.com/catalogue/set-me-free_988/index.html',
     'https://books.toscrape.com/catalogue/scott-pilgrims-precious-little-life-scott-pilgrim-1_987/index.html',
     'https://books.toscrape.com/catalogue/rip-it-up-and-start-again_986/index.html',
     'https://books.toscrape.com/catalogue/our-band-could-be-your-life-scenes-from-the-american-indie-underground-1981-
    1991_985/index.html',
     'https://books.toscrape.com/catalogue/olio_984/index.html',
     'https://books.toscrape.com/catalogue/mesaerion-the-best-science-fiction-stories-1800-1849_983/index.html',
     'https://books.toscrape.com/catalogue/libertarianism-for-beginners_982/index.html',
     'https://books.toscrape.com/catalogue/its-only-the-himalayas_981/index.html']
```

```python
all_urls=['https://books.toscrape.com/catalogue/page-1.html']
current_page='https://books.toscrape.com/catalogue/page-1.html'
base_url='https://books.toscrape.com/catalogue/'
response=requests.get(current_page)
```

```python
while response.status_code == 200:
  data=BeautifulSoup(response.text,'html.parser')
  next_page=data.find(class_='next')
  next_page_url=base_url+next_page.a['href']
  print(next_page_url)
```

```
all_urls.append(next_page_url)

current_page= next_page_url
response=requests.get(current_page)
```

https://books.toscrape.com/catalogue/page-2.html
https://books.toscrape.com/catalogue/page-3.html
https://books.toscrape.com/catalogue/page-4.html
https://books.toscrape.com/catalogue/page-5.html
https://books.toscrape.com/catalogue/page-6.html
https://books.toscrape.com/catalogue/page-7.html
https://books.toscrape.com/catalogue/page-8.html
https://books.toscrape.com/catalogue/page-9.html
https://books.toscrape.com/catalogue/page-10.html
https://books.toscrape.com/catalogue/page-11.html
https://books.toscrape.com/catalogue/page-12.html
https://books.toscrape.com/catalogue/page-13.html
https://books.toscrape.com/catalogue/page-14.html
https://books.toscrape.com/catalogue/page-15.html
https://books.toscrape.com/catalogue/page-16.html
https://books.toscrape.com/catalogue/page-17.html
https://books.toscrape.com/catalogue/page-18.html
https://books.toscrape.com/catalogue/page-19.html
https://books.toscrape.com/catalogue/page-20.html
https://books.toscrape.com/catalogue/page-21.html
https://books.toscrape.com/catalogue/page-22.html
https://books.toscrape.com/catalogue/page-23.html
https://books.toscrape.com/catalogue/page-24.html
https://books.toscrape.com/catalogue/page-25.html
https://books.toscrape.com/catalogue/page-26.html
https://books.toscrape.com/catalogue/page-27.html
https://books.toscrape.com/catalogue/page-28.html
https://books.toscrape.com/catalogue/page-29.html
https://books.toscrape.com/catalogue/page-30.html
https://books.toscrape.com/catalogue/page-31.html
https://books.toscrape.com/catalogue/page-32.html
https://books.toscrape.com/catalogue/page-33.html
https://books.toscrape.com/catalogue/page-34.html
https://books.toscrape.com/catalogue/page-35.html
https://books.toscrape.com/catalogue/page-36.html
https://books.toscrape.com/catalogue/page-37.html
https://books.toscrape.com/catalogue/page-38.html
https://books.toscrape.com/catalogue/page-39.html
https://books.toscrape.com/catalogue/page-40.html
https://books.toscrape.com/catalogue/page-41.html
https://books.toscrape.com/catalogue/page-42.html
https://books.toscrape.com/catalogue/page-43.html
https://books.toscrape.com/catalogue/page-44.html
https://books.toscrape.com/catalogue/page-45.html
https://books.toscrape.com/catalogue/page-46.html
https://books.toscrape.com/catalogue/page-47.html
https://books.toscrape.com/catalogue/page-48.html
https://books.toscrape.com/catalogue/page-49.html
https://books.toscrape.com/catalogue/page-50.html

```
---------------------------------------------------------------------
AttributeError                          Traceback (most recent call last)
<ipython-input-24-f447b82a1a26> in <cell line: 6>()
      7    data=BeautifulSoup(response.text,'html.parser')
      8    next_page=data.find(class_='next')
----> 9    next_page_url=base_url+next_page.a['href']
     10    print(next_page_url)
     11    all_urls.append(next_page_url)

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

EXPLAIN ERROR

```python
all_urls=['https://books.toscrape.com/catalogue/page-1.html']
current_page='https://books.toscrape.com/catalogue/page-1.html'
base_url='https://books.toscrape.com/catalogue/'
response=requests.get(current_page)

while response.status_code == 200:
  data=BeautifulSoup(response.text,'html.parser')
  next_page=data.find(class_='next')
  if next_page is None:
    break
  next_page_url=base_url+next_page.a['href']
  print(next_page_url)
  all_urls.append(next_page_url)

  current_page= next_page_url
  response=requests.get(current_page)
```

˅ to_csv

```python
import requests
from bs4 import BeautifulSoup
response= requests.get('https://books.toscrape.com/')
data=BeautifulSoup(response.text,'html.parser')
data

b1=data.find(class_="product_pod")
print(b1)

base_url='https://books.toscrape.com/'
b1_url=base_url+b1.h3.a['href']
print(b1_url)

response=requests.get(b1_url)
data=BeautifulSoup(response.text,'html.parser')
data

title=data.h1.string

price=data.find(class_='price_color').string

qty=data.find(class_='instock availability')
qty=qty.contents[-1].strip()

print(title)
print(b1_url)
print(price)
print(qty)

import re
qty=int(re.search('\d+',qty).group())


 price=float(re.search('\d.+', price).group())
```

```
print(title)
print(b1_url)
print(price)
print(qty)

books_details=[]
books_details.append([title, b1_url, price, qty])
print(books_details)

import pandas as pd
df=pd.DataFrame(books_details, columns = ["Title", "Link", "Price", "Quantity In Stock"])
print(df)

df.to_csv("books.details.csv", index=False)
```