

Web scrapping

```
In [1]: 1 #import BeautifulSoup
2 #import request
3 #parcing = one format to another format
4 #preetify = readable
5 # we can use beautifulsoup or celinium
6 # beautifulsoup = scrapping data from one page and it is one of the Library
7 #Labrory = collection of package
8 # bs4 version
9
```

```
In [2]: 1 # use = to get huge information from the webside = use
```

```
In [7]: 1 from bs4 import BeautifulSoup as bs
```

```
In [8]: 1 import requests
```

```
In [9]: 106au-thin-light-laptop/product-reviews/itmddb2d234797b6?pid=COMFZHZFZFFBJBHH&lid=LSTCOMFZHZFZFFBJBHH7WNMPE&marketplace=FLIPKART"
```

```
In [10]: 1 r=requests.get(url) # request module in that get method
2 print(r) # we need responce Like 200 or it might be 400
```

<Response [200]>

```
In [8]: 1 htmlcontent=r.content
2 print(htmlcontent)

<!-- -->
<div class="3qX0zy"><a href="/"><a class="_21ljIi" href="/plus">Explore<!-- --> <spa
n class="_2FVHGh">Plus</span></a></div></div><div class="_1cmsER"><form class="_2M8cLY header-form-search" action="/search" method="GET"><div cla
ss="col-12-12 _2009oE"><div class="_3005Xc"><input type="text" value="" class="_3704LK" title="Search for products, brands an
d more" name="q" autoComplete="off" placeholder="Search for products, brands and more"/></div><button class="L0Z3Pu" type="su
bmit"><svg width="20" height="20" viewBox="0 0 17 18" class="" xmlns="http://www.w3.org/2000/svg"><g fill="#2874F1" fill-rule
="evenodd"><path class="_34RNph" d="m11.618 9.89714.225 4.212c.092.092.101.232.02.3131-1.465 1.46c-.081.081-.221.072-.314-.02
1-4.216-4.203"></path><path class="_34RNph" d="m6.486 10.901c-2.42 0-4.381-1.956-4.381-4.368 0-2.413 1.961-4.369 4.381-4.369
2.42 0 4.381 1.956 4.381 4.369 0 2.413-1.961 4.368-4.381 4.368m0-10.835c-3.582 0-6.486 2.895-6.486 6.467 0 3.572 2.904 6.467
6.486 6.467 3.582 0 6.486-2.895 6.486-6.467 0-3.572-2.904-6.467-6.486-6.467"></path></g></svg></button><input type="hidden" n
ame="otracker" value="search"/><input type="hidden" name="otracker1" value="search"/><input type="hidden" name="marketplace"
value="FLIPKART"/><input type="hidden" name="as-show" value="off"/><input type="hidden" name="as" value="off"/></div><ul clas
s="col-12-12 _1MRYA1"></ul></form></div><div class="go_D0p _2errNR"><div class="_28p97w"><div class="_1psGvi _3BvnxG"><div><a
class="_1_3w1N" href="/account/login?ret=hp-15s-ryzen-3-dual-core-3250u-4-gb-1-tb-hdd-windows-10-home-15s-gr0006au-thin-ligh
t-laptop/product-reviews/itmddb2d234797b6%3Fpid%3DCOMFZHZFZFFBJBHH%26lid%3DLSTCOMFZHZFZFFBJBHH7WNMPE%26marketplace%3DFLIPKAR
T">Login</a></div></div></div><div class="go_D0p"><a href="https://seller.flipkart.com/sell-online/?utm_source=fkwebsit
e&utm_medium=websitedirect" class="_3-PJz-"><span>Become a Seller</span></a></div><div class="go_D0p"><div class="_28p97
w"><div class="_1psGvi _3BvnxG"><div><div class="exehdJ"> <!-- -->More<!-- --> </div></div></div><svg width="4.7" height="8"
-->
```

```
In [12]: 1 soup=bs(htmlcontent,"html.parser") #parser = to get whole content and preetyfy it for furthur
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2728\1339617220.py in <module>
----> 1 soup=bs(htmlcontent,"html.parser") #parser = to get whole content and preetyfy it for furthur

NameError: name 'htmlcontent' is not defined
```

```
In [11]: 1 print(soup.prettify)
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2728\1128155920.py in <module>
----> 1 print(soup.prettify)

NameError: name 'soup' is not defined
```

```
In [11]: 1 # to find title of the product
2
3 title=soup.title
4 print(title)
5 print(type(title))
```

```
<title>Hp 15s Ryzen 3 Dual Core 3250u 4 Gb 1 Tb Hdd Windows 10 Home 15s Gr0006au Thin Light Laptop Reviews: Latest Review of Hp
15s Ryzen 3 Dual Core 3250u 4 Gb 1 Tb Hdd Windows 10 Home 15s Gr0006au Thin Light Laptop | Price in India | Flipkart.com</title>
<class 'bs4.element.Tag'>
```

```
In [12]: 1 print(title.string)
```

```
Hp 15s Ryzen 3 Dual Core 3250u 4 Gb 1 Tb Hdd Windows 10 Home 15s Gr0006au Thin Light Laptop Reviews: Latest Review of Hp 15s Ry
zen 3 Dual Core 3250u 4 Gb 1 Tb Hdd Windows 10 Home 15s Gr0006au Thin Light Laptop | Price in India | Flipkart.com
```

```
In [15]: 1 price=soup.find_all("div",class_="_30jeq3")
2 print(price)
```

```
[<div class="_30jeq3">₹32,990</div>]
```

```
In [16]: 1 print(price[0].string)
```

```
₹32,990
```

```
In [18]: 1 prod_price=[]
2 for i in range(0,len(price)):
3     prod_price.append(price[i].get_text())
4 print(prod_price)
```

```
['₹32,990']
```

```
In [19]: 1 cust_name=soup.find_all("p",class_="_2sc7ZR _2V5EHH")
2 print(cust_name)
```

```
[<p class="_2sc7ZR _2V5EHH">Vaibhav Bharati</p>, <p class="_2sc7ZR _2V5EHH">MD MATHIN ALI</p>, <p class="_2sc7ZR _2V5EHH">Sayan
i Banerjee </p>, <p class="_2sc7ZR _2V5EHH">Satyajit Swain</p>, <p class="_2sc7ZR _2V5EHH">Vijay Singh Shekhawat</p>, <p clas
s="_2sc7ZR _2V5EHH">Anoop Mathew</p>, <p class="_2sc7ZR _2V5EHH">Vivek Raturi</p>, <p class="_2sc7ZR _2V5EHH">Anand Amirtharaj
</p>, <p class="_2sc7ZR _2V5EHH">Flipkart Customer</p>, <p class="_2sc7ZR _2V5EHH">Shivani Pankaj</p>]
```

```
In [20]: 1 name=[]
2 for i in range(0,len(cust_name)):
3     name.append(cust_name[i].get_text())
4 print(name)
```

```
['Vaibhav Bharati', 'MD MATHIN ALI', 'Sayani Banerjee ', 'Satyajit Swain', 'Vijay Singh Shekhawat', 'Anoop Mathew', 'Vivek Ra
turi', 'Anand Amirtharaj', 'Flipkart Customer', 'Shivani Pankaj']
```

```
In [21]: 1 comment=soup.find_all("p",class_="_2-N8zT")
2 print(comment)
```

```
[<p class="_2-N8zT">Very Good</p>, <p class="_2-N8zT">Terrific</p>, <p class="_2-N8zT">Worth every penny</p>, <p class="_2-N8z
T">Fair</p>, <p class="_2-N8zT">Fair</p>, <p class="_2-N8zT">Just wow!</p>, <p class="_2-N8zT">Good</p>, <p class="_2-N8zT">Won
derful</p>, <p class="_2-N8zT">Very Good</p>, <p class="_2-N8zT">Nice product</p>]
```

```
In [22]: 1 comment_name=[]
2 for i in range(0,len(comment)):
3     comment_name.append(comment[i].get_text())
4 print(comment_name)
```

```
['Very Good', 'Terrific', 'Worth every penny', 'Fair', 'Fair', 'Just wow!', 'Good', 'Wonderful', 'Very Good', 'Nice product']
```

```
In [23]: 1 comment_name
```

```
Out[23]: ['Very Good',
'Terrific',
'Worth every penny',
'Fair',
'Fair',
'Just wow!',
'Good',
'Wonderful',
'Very Good',
'Nice product']
```

check how much stars gives customer to this product

```
In [29]: 1 in_comment=soup.find_all("div",class_="t-ZTKy")
2         print(in_comment)

[<div class="t-ZTKy"><div><div class="">Read carefully.. My Experience was complicated just because of boot time i.e. on/off time was too long until I didn't go for updates but once I have done all updates... I am telling you guys...Maska..hai.. Maska.<br/><br/>Superb performance.. and look wise i liked it very much. I use to play GT racing without any lag.<br/><br/>My daughter is also happy for her online studies.<br/><br/>My suggestion is that don't think twice just go for it with Flipkart.<br/><br/>Thanks flipkart for very safe and on time delivery.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">#Price up to mark<br/>2)Its wonderful product for those ,who looking for student and domestic work<br/><br/>1)Its just my dream lappi,(laptop)<br/>2)just go for laptop,its really awesome<br/>3)I just bought it for my project and bussiness work it's awesome<br/>4)one of the best product from HP<br/>5)just price is little high ,it have to be reduce d<br/>6) I loved it<br/>7) I recommend it for u,👍❤️</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">Nice laptop and Ryzen 3 is a good processor for multi tasking... Design and color both are good. It's better to go with this HP laptop in this price segment...</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">This laptop needs to be upgraded to at least 8 gb memory for better performance as most of HP bloatware that comes with the windows build use almost 80% of ram all the time. Besides the built in vega gpu shares Ram as it does not have dedicated graphics memory. An SSD upgrade and booting the OS in there should do wonders to the disk I/O and overall performance. Out of the box, laptop is sluggish so 3 stars.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">HP Product is Good/Ok, but personally i will prefer Dell Laptop. Dell Laptop is best & friendly for smooth process<br/>& I thing i3 is better comparison to Rayzen 3.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">Awesome and nice for students and all. The favorite things is its comes with Windows and office home and student precious.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">Do not expect much :<br/>1)Very slow takes a lot of time just open a browser.<br/>2)Design is good <br/>3)Battery backup ok.<br/>if you want to do basic editing and some gaming then go for atleast 8 GB ram and 4 core processor .</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">Compare to i3, ryzen 3 is far better in performance with 4gb ram, full Battery charged still works more than 6 hrs... No heating problem.. Display only not bad quality, rest of them are worth for this amount</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">Very nice product finishing is very good.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div><div class="">This laptop is very light weight and I am happy!! 😊😊</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>]
```

```
In [31]: 1 review=[]
2         for i in range(0,len(in_comment)):
3             review.append(in_comment[i].get_text())
4         (review)
```

```
Out[31]: ["Read carefully.. My Experience was complicated just because of boot time i.e. on/off time was too long until I didn't go for
updates but once I have done all updates... I am telling you guys...Maska..hai.. Maska..Superb performance.. and look wise i li
ked it very much. I use to play GT racing without any lag..My daughter is also happy for her online studies.My suggestion is th
at don't think twice just go for it with Flipkart.Thanks flipkart for very safe and on time delivery.READ MORE",
"#Price up to markIts wonderful product for those ,who looking for student and domestic work1)Its just my dream lappi,(lapto
p)2)just go for laptop,its really awesome3)I just bought it for my project and bussiness work it's awesome4)one of the best pro
duct from HP5)just price is little high ,it have to be reduced6) I loved it7) I recommend it for u,💯💯❤️READ MORE",
"Nice laptop and Ryzen 3 is a good processor for multi tasking... Design and color both are good. It's better to go with this
HP laptop in this price segment...READ MORE",
'This laptop needs to be upgraded to at least 8 gb memory for better performance as most of HP bloatware that comes with the w
indows build use almost 80% of ram all the time. Besides the built in vega gpu shares Ram as it does not have dedicated graphic
s memory. An SSD upgrade and booting the OS in there should do wonders to the disk I/O and overall performance. Out of the box,
laptop is sluggish so 3 stars.READ MORE',
'HP Product is Good/Ok, but personally i will prefer Dell Laptop. Dell Laptop is best & friendly for smooth process& I thing i
3 is better comparison to Rayzen 3.READ MORE',
'Awesome and nice for students and all. The favorite things is its comes with Windows and office home and student precious.REA
D MORE',
'Do not expect much :1)Very slow takes a lot of time just open a browser.2)Design is good 3)Battery backup ok.if you want to
do basic editing and some gaming then go for atleast 8 GB ram and 4 core processor .READ MORE',
'Compare to i3, ryzen 3 is far better in performance with 4gb ram, full Battery charged still works more than 6 hrs... No heat
ing problem.. Display only not bad quality, rest of them are worth for this amountREAD MORE',
'Very nice product finishing is very good.READ MORE',
'This laptop is very light weight and I am happy!! 😊😊READ MORE']
```

Applying Tabular data structure

```
In [5]: 1 import pandas as pd
```

in pandas data stored in frames

```
In [6]: 1 df=pd.DataFrame()
2         df["cust_name"]=name
3         df["comment"]=comment_name
4         df["stars"]=cust_stars
5         df["in_comment"]=review
6
7
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2728\3555515426.py in <module>
      1 df=pd.DataFrame()
----> 2 df["cust_name"]=name
      3 df["comment"]=comment_name
      4 df["stars"]=cust_stars
      5 df["in_comment"]=review

NameError: name 'name' is not defined
```

```
In [2]: 1 df
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2728\1396537375.py in <module>
----> 1 df

NameError: name 'df' is not defined
```

```
In [ ]: 1
```