

Extract the HTML code of a website

1. Import libraries
2. Save the URL
3. Using requests.get(), access the web page
4. Using BeautifulSoup(), access the HTML code

```
from bs4 import BeautifulSoup as bs
import requests as rq
A = "https://en.wikipedia.org/wiki/India"
page=rq.get(A)
code=bs(page.text,"html")
code
t=code.find('title')
print(t.text)
```

India - Wikipedia

```
from bs4 import BeautifulSoup as bs
import requests as rq
z="https://www.forbesindia.com/article/explainers/top-10-richest-people-india/85909/1"
page=rq.get(z)
code=bs(page.text,'html')
t=code.find('table')
h=t.find_all('th')
headings=[i.text for i in h]
print(headings)
```

['Name & India Rank', 'Global Rank', 'Net worth (US\$)', 'Company']

```
import pandas as pd
mydf=pd.DataFrame(columns=headings)
print(mydf)
```

Empty DataFrame
Columns: [Name & India Rank, Global Rank, Net worth (US\$), Company]
Index: []

```
tablerow=t.find_all("td")
l=[i.text for i in tablerow]
l
```

```
['#1 Mukesh Ambani ',
'10',
'$116.9 B',
'Reliance Industries',
'#2 Gautam Adani ',
'16',
'$86.2 B',
'Adani Group',
'#3 Shiv Nadar ',
'37',
'$36.8 B',
'HCL Technologies\xa0\xa0 ',
'#4 Savitri Jindal & family ',
'58',
'$30.9 B',
'JSW Group',
'#5 Dilip Shanghvi ',
'69',
'$25.7 B',
'Sun Pharmaceutical Industries Ltd',
'#6 Cyrus Poonawalla ',
'68',
'$24.2 B',
'Serum Institute of India',
'#7 Kushal Pal Singh',
'98',
'$20.6 B',
'DLF Limited',
'#8 Kumar Birla ',
'97',
'$19.5 B',
'Aditya Birla Group',
```

```
'#9 Ravi Jaipuria',
'100',
'$17.4 B',
'RJ Corp, Varun Beverages ',
'#10 Radhakishan Damani ',
'101',
'$17.2 B',
'DMart, Avenue Supermarts']
```

```
st=0
sp=4
for i in range((len(l)//4)):
    mydf.loc[i]=l[st:sp]
    st+=4
    sp+=4
mydf
```

	Name & India Rank	Global Rank	Net worth (US\$)	Company
0	#1 Mukesh Ambani	10	\$116.9 B	Reliance Industries
1	#2 Gautam Adani	16	\$86.2 B	Adani Group
2	#3 Shiv Nadar	37	\$36.8 B	HCL Technologies
3	#4 Savitri Jindal & family	58	\$30.9 B	JSW Group
4	#5 Dilip Shanghvi	69	\$25.7 B	Sun Pharmaceutical Industries Ltd
5	#6 Cyrus Poonawalla	68	\$24.2 B	Serum Institute of India
6	#7 Kushal Pal Singh	98	\$20.6 B	DLF Limited
7	#8 Kumar Birla	97	\$19.5 B	Aditya Birla Group
8	#9 Ravi Jaipuria	100	\$17.4 B	RJ Corp, Varun Beverages
9	#10 Radhakishan Damani	101	\$17.2 B	DMart, Avenue Supermarts

```
mydf.to_csv("table.csv")
```