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/Main_Page"
page=rq.get(a)
code=bs(page.text,'html')
t=code.find('title')
print(t)#removes title tag
```

```
    <title>Wikipedia, the free encyclopedia</title>
```

Extracting a table from a webpage

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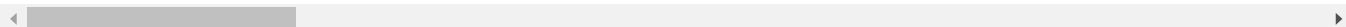
```
    ['Name & India Rank', 'Global Rank', 'Net worth (US$)', 'Company']
```

```
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]
print(l)
```

```
    ['#1 Mukesh Ambani ', '10', '$116.9 B', 'Reliance Industries', '#2 Gautam Adani ', '16', '$86.2 B', 'Adani Group', '#3 Shiv Nadar ',
```

◀ ▬▬▬▬▬▬▬▬   ▶

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

```python
mydf.to_csv("tabledata.csv")
```

```python
import requests as rq
from bs4 import BeautifulSoup as bs
import pandas as pd
a="https://en.m.wikipedia.org/wiki/List_of_tallest_buildings_in_India"
page=rq.get(a)
code=bs(page.text,"html")
t=code.find("table")
h=t.find_all("th")
headings=[i.text for i in h]
df5=pd.DataFrame(columns=headings)
tablerow=t.find_all("td")
data=[i.text for i in tablerow]
st=0
sp=9
print(len(data))
for i in range((len(data)//9)):
  df5.loc[i]=data[st:sp]
  st=+9
  sp+=9
df5
```

```python
import requests as rq
from bs4 import BeautifulSoup as bs
import pandas as pd
a="https://en.m.wikipedia.org/wiki/List_of_tallest_buildings_in_India"
page=rq.get(a)
code=bs(page.text,"html")
t=code.find("table")
h=t.find_all("th")
headings=[i.text for i in h]
print(headings)
```

```
    ['Rank\n', 'Name\n', 'Image\n', 'City\n', 'Height\n', 'Floors\n', 'Year\n', 'Building type\n', 'Reference(s)\n']
```

```python
df5=pd.DataFrame(columns=headings)
tablerow=t.find_all("td")
data=[i.text for i in tablerow]
print(data)
```

```
    ['1\n', 'Palais Royale\n', '\xa0\n', 'Mumbai\n', '320 metres (1,050\xa0ft)\n', '88\n', '2018\n', 'Residential\n', '[22]', '2\n', 'Lo
```
◄   ►

```python
st=0
sp=9
print(len(data))
for i in range(len(data)//9):
  df5.loc[i]=data[st:sp]
  st+=9
  sp+=9
df5
```

```
df5.to_csv("Buildings.csv")
```