msd2b

November 22, 2024

[12]:
```
!pip install beautifulsoup4
```

```
Requirement already satisfied: beautifulsoup4 in c:\users\advait
nathe\anaconda3\lib\site-packages (4.12.3)
Requirement already satisfied: soupsieve>1.2 in c:\users\advait
nathe\anaconda3\lib\site-packages (from beautifulsoup4) (2.5)
```

[13]:
```
pip show beautifulsoup4
```

```
Name: beautifulsoup4Note: you may need to restart the kernel to use updated
packages.

Version: 4.12.3
Summary: Screen-scraping library
Home-page:
Author:
Author-email: Leonard Richardson <leonardr@segfault.org>
License: MIT License
Location: C:\Users\ADVAIT NATHE\anaconda3\Lib\site-packages
Requires: soupsieve
Required-by: conda-build, nbconvert
```

[14]:
```
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

[15]:
```
url="https://en.wikipedia.org/wiki/World_population"
response=requests.get(url)
soup=BeautifulSoup(response.content,"html.parser")
print(response)
```

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

[16]:
```
soup.title
```

[16]:
```
<title>World population - Wikipedia</title>
```

[17]:
```
soup.title.text
```

[17]:
```
'World population - Wikipedia'
```

[30]:
```
tables=soup.find_all('table')
dataframe1=[]

for i, table in enumerate (tables):
    rows=table.find_all('tr')[1:]
    data=[]
    for row in rows:
        cols=row.find_all('td')
        cols=[col.text.strip() for col in cols]
        data.append(cols)

    df=pd.DataFrame(data)
    dataframe1.append(df)

dataframe1[2]
```

[30]:

	0	1	2	3
0	Sub-Saharan Africa	1,152 (14.51%)	1,401 (16.46%)	2,094 (21.62%)
1	Northern Africa and Western Asia	549 (6.91%)	617 (7.25%)	771 (7.96%)
2	Central Asia and Southern Asia	2,075 (26.13%)	2,248 (26.41%)	2,575 (26.58%)
3	Eastern Asia and Southeastern Asia	2,342 (29.49%)	2,372 (27.87%)	2,317 (23.92%)
4	Europe and Northern America	1,120 (14.10%)	1,129 (13.26%)	1,125 (11.61%)
5	Latin America and the Caribbean	658 (8.29%)	695 (8.17%)	749 (7.73%)
6	Australia and New Zealand	31 (0.39%)	34 (0.40%)	38 (0.39%)
7	Oceania	14 (0.18%)	15 (0.18%)	20 (0.21%)
8	World	7,942	8,512	9,687

[ ]:
```

```