In [1]:
```
#wikipedia header files
```

In [4]:
```
from urllib.request import urlopen
from bs4 import BeautifulSoup
html = urlopen('https://en.wikipedia.org/wiki/Main_Page')
bs = BeautifulSoup(html, "html.parser")
titles = bs.find_all(['h1', 'h2','h3','h4','h5','h6'])
print('List all the header tags :', *titles, sep='\n\n')
```

```
List all the header tags :

<h1 class="firstHeading mw-first-heading" id="firstHeading" style="display: none">
<span class="mw-page-title-main">Main Page</span></h1>

<h1><span class="mw-headline" id="Welcome_to_Wikipedia">Welcome to <a href="/wiki/
Wikipedia" title="Wikipedia">Wikipedia</a></span></h1>

<h2 class="mp-h2" id="mp-tfa-h2"><span id="From_today.27s_featured_article"></span
><span class="mw-headline" id="From_today's_featured_article">From today's feature
d article</span></h2>

<h2 class="mp-h2" id="mp-dyk-h2"><span class="mw-headline" id="Did_you_know_...">D
id you know ...</span></h2>

<h2 class="mp-h2" id="mp-itn-h2"><span class="mw-headline" id="In_the_news">In the
news</span></h2>

<h2 class="mp-h2" id="mp-otd-h2"><span class="mw-headline" id="On_this_day">On thi
s day</span></h2>

<h2 class="mp-h2" id="mp-tfp-h2"><span id="Today.27s_featured_picture"></span><spa
n class="mw-headline" id="Today's_featured_picture">Today's featured picture</span
></h2>

<h2 class="mp-h2" id="mp-other"><span class="mw-headline" id="Other_areas_of_Wikip
edia">Other areas of Wikipedia</span></h2>

<h2 class="mp-h2" id="mp-sister"><span id="Wikipedia.27s_sister_projects"></span><
span class="mw-headline" id="Wikipedia's_sister_projects">Wikipedia's sister proje
cts</span></h2>

<h2 class="mp-h2" id="mp-lang"><span class="mw-headline" id="Wikipedia_languages">
Wikipedia languages</span></h2>
```

In [ ]:
```
from bs4 import BeautifulSoup
import requests
page=requests.get('https://presidentofindia.nic.in/former-presidents.htm')
page
soup=BeautifulSoup(page.content)
soup
first_title=soup.find('h3')
first_title
first_title.text

first_title=[]
for i in soup.find_all('h3'):
    first_title.append(i.text)
first_title

Term=soup.find('p')
Term
Term.text
```

```
Term=[]
for i in soup.find_all('p'):
    Term.append(i.text)
Term

Detail=soup.find('div', class_="presidentListing")
Detail
Detail.text

Detail=[]
for i in soup.find_all('div', class_="presidentListing"):
    Details.append(i.text)
Detail
import pandas as pd
df=pd.DataFrame({"PresidentiaL List":Detail})
df
```

In [ ]:

In [12]:
```
#Write a python program to scrape cricket rankings from icc-cricket.com. You have t
# Top 10 ODI Batsmen along with the records of their team andrating.
#       Top 10 ODI bowlers along with the records of their team andrating.
# Write a python program to scrape cricket rankings from icc-cricket.com. You have
# Top 10 women's ODI Batting players along with the records of their team and ratir
# Top 10 women's ODI all-rounder along with the records of their team and rating



from bs4 import BeautifulSoup
import requests
from prettytable import PrettyTable


def Menu():
    print('\n1. Men \n2. Women\n')
    gen=Gender()
    print('\n1. Team Rankings \n2. Player Ranking\n')
    tp=TeamOrPlayer()
    mode=''
    val=''

    if gen=='mens':
        print('\n1. Test\n2. ODI\n3. T20\n')
        mode=Mode()

    if tp=='player-rankings':
        if mode=='':
            print('\n1. ODI\n2. T20\n')
            mode=Mode2()
            print('\n1. Batting\n2. Bowling\n3. All-Rounder\n')
            val=Value()

        return gen,tp,mode,val


def Gender():
    gender=input('Enter your choice:')
    code={'1':'mens','2':'womens'}

    if gender in code:
        return code[gender]
    else:
        print('\nInvalid Input\nTry Again\n')
```

```
        return Gender();


def TeamOrPlayer():
    choice=input('Enter your choice:')
    tp={'1':'team-rankings','2':'player-rankings'}
    if choice in tp:
    return tp[choice]
else:
    print('\nInvalid Input\nTry Again\n')
    return TeamOrPlayer();


def Mode():
    choice=input('Enter your choice:')
    word={'1':'/test','2':'/odi','3':'/t20i'}

    if choice in word:
        return word[choice]
    else:
        print('\nInvalid Input\nTry Again\n')
        return Choice();


def Mode2():
    choice=input('Enter your choice:')
    word={'1':'/odi','2':'/t20i'}

    if choice in word:
        return word[choice]
    else:
        print('\nInvalid Input\nTry Again\n')
        return Choice();


def Value():
    choice=input('Enter your choice:')
    val={'1':'batting','2':'bowling','3':'all-rounder'}

    if choice in val:
        return val[choice]

     else:
        print('\nInvalid Input\nTry Again\n')
        return Value()


def URL():
     gen,tp,mode,val=Menu()
        url='https://www.icc-cricket.com/rankings/'+gen+'/'+tp+mode+'/'+val
        header=gen.upper() +' ' +mode[1:].upper() + ' ' + val.upper()
         print('\n{:<15}  {:<30}\n{:<15}  {:<30}'.format('',tp.upper(),'',header))
        return url,tp


def SOUP(url,tp):
    try:
        res=requests.get(url)
        soup=BeautifulSoup(res.text,'lxml')
        a= soup.find_all('tr',{'class':'table-body'})
        data={}
        for i in a :
        team=[]
        name=''
```

```
        rating=''

        try:
        rank=int(i.contents[1].text)
        except:
        pass

     try:
                name=i.contents[3].text.replace('\n','')
                 name=" ".join(name.split())
            if rank==1 and tp=='player-rankings':
        name=name[0:-3]
        except:
        pass

    try:
        rating=i.contents[9].text
    except:
        if rank==1 :
        rating=i.contents[5].text
    else:
        rating=i.contents[7].text
        team.extend([name,rating])
        data[rank]=team

        return data

     except:
        return SOUP(url,tp)


def Print(data):
     print('\nRANKING \t TEAM\t\t\t\tRATING')
    for i in sorted(data):
    print('{:<10}       '.format(i),end='')
     for j in range(len(data[i])):
            print('{:<26}'.format(data[i][j]),end='     ')
    print()


def main():

    url,tp=URL()
    data=SOUP(url,tp)
    Print(data)


if __name__=='__main__':
    main()
```

```
  File <tokenize>:87
    else:
    ^
IndentationError: unindent does not match any outer indentation level
```

In [ ]:

```
import requests
from bs4 import BeautifulSoup
page=requests.get('https://www.cnbc.com/world/?region=world')
news=BeautifulSoup(page.content)
news
Time=news.find('time')
Time
Time.text
Time=[]
```

```
for i in news.find_all('time'):
    Time.append(i.text)
Time
url = "https://www.cnbc.com/world/?region=world"
webpage = requests.get(url)
trav = BeautifulSoup(webpage.content, "html.parser")
for link in trav.find_all('a'):
    print(type(link), " ", link)
trav.text
```

In [ ]:

```
import requests
from bs4 import BeautifulSoup
page=requests.get('https://www.journals.elsevier.com/artificial-intelligence/most-
page

Books=BeautifulSoup(page.content)
Books
Title=Books.find('h2')
Title
Title.text

Title=[]
for i in Books.find_all('h2'):
    Title.append(i.text)
Title
Author=Books.find('span',class_="sc-1w3fpd7-0 pgLAT")
Author
Author.text
Author=[]
for i in Books.find_all('span',class_="sc-1w3fpd7-0 pgLAT"):
    Author.append(i.text)
Author
Date=Books.find('span',class_="sc-1thf9ly-2 bKddwo")
Date

Date=[]
for i in Books.find_all('span',class_="sc-1thf9ly-2 bKddwo"):
    Date.append(i.text)
Date
import pandas as pd
df=pd.DataFrame({"Title":Title, "Author":Author, "Published Date":Date})
df
```

In [ ]:

```
#
import requests
from bs4 import BeautifulSoup
page=requests.get('https://https://https://www.dineout.co.in/hyderabad-restaurants
page
soup=BeautifulSoup(page.content)
soup
RN=soup.find('div',class_="restnt-info cursor")
RN
RN.text

RN[]
for i in soup.find_all('div',class_="restnt-info cursor"):
    RN.append(i.text)



price=[]
for i in soup.find_all('span',class_="double-line-ellipsis"):
    price.append(i.text)
```

```
price
location=[]
for i in soup.find_all('div',class_="restnt-loc ellipsis"):
    location.append(i.text)
    images=[]
for i in soup.find_all('img',class_="no-img"):
    images.append(i['data-src'])
    Rt=[]
for i in soup.find_all('div',class_="restnt-rating rating-4"):
    Rt.append(i.text)

Rt
import pandas as pd
df=pd.DataFrame({"Restaurant Name":RN, "Cusines":price, "Location":location, "Rati
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: