ASSIGNMENT 1

WEB SCRAPING

In [4]:
```
# 1.PHYTHON PROGRAM TO DISPLAY ALL THE HEADER TAGS FROM WIKIPEDIA.ORG

#import libraries

!pip install bs4
!pip install requests

from bs4 import BeautifulSoup
import requests

#send get requests to the webpage server to get the source code of the page

page = requests.get('https://www.wikipedia.org/')

page

#page content
soup = BeautifulSoup(page.content)

soup

# creating a list of all common heading tags
heading_tags = ['h1','h2','h3']
for tags in soup.find_all(heading_tags):
    print(tags.name + ' -> ' +tags.text.strip())
```

```
Requirement already satisfied: bs4 in c:\users\dell\anaconda3\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\dell\anaconda3\lib\site-packages (from bs4) (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\dell\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)
Requirement already satisfied: requests in c:\users\dell\anaconda3\lib\site-packages (2.28.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2022.9.14)
Requirement already satisfied: idna<4,>=2.5 in c:\users\dell\anaconda3\lib\site-packages (from requests) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\dell\anaconda3\lib\site-packages (from requests) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2.0.4)
h1 -> Wikipedia

The Free Encyclopedia
h2 -> 1 000 000+


articles
h2 -> 100 000+


articles
h2 -> 10 000+


articles
h2 -> 1 000+


articles
h2 -> 100+


articles
```

In [19]:
```
# 2.PROGRAM TO DISPLAY IMBD'S TOP 100 RATED MOVIES

# install libraries
!pip install bs4
!pip install requests


from bs4 import BeautifulSoup
import requests


#send get requests to the webpage server to get the source code of the page

page = requests.get('https://www.imdb.com/list/ls091520106/')
page

# page content

soup=BeautifulSoup(page.content)
soup

# creating a list of top rated imbd's 100 movies

movie_title = soup.find('h3',class_='lister-item-header')
movie_title

movie_title.text

movie_year = soup.find('span',class_='lister-item-year text-muted unbold')
movie_year

movie_year.text

rating = soup.find('span',class_='ipl-rating-star__rating')
rating

rating.text

movies=[]

for i in soup.find_all('h3',class_='lister-item-header'):
    movies.append(i.text)

movie

movie_year=[]

for i in soup.find_all('span',class_='lister-item-year text-muted unbold'):
    movie_year.append(i.text)

movie_year

rating=[]

for i in soup.find_all('div',class_='ipl-rating-star small'):
    rating.append(i.text)

rating

print(len(movie),len(movie_year),len(rating))       #printing length

# making dataframe

import pandas as pd

df=pd.DataFrame({'movie':movie,'movie_year':movie_year,'rating':rating})
df
```

```
Requirement already satisfied: bs4 in c:\users\dell\anaconda3\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\dell\anaconda3\lib\site-packages (from bs4) (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\dell\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)
Requirement already satisfied: requests in c:\users\dell\anaconda3\lib\site-packages (2.28.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\dell\anaconda3\lib\site-packages (from requests) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\dell\anaconda3\lib\site-packages (from requests) (3.3)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2022.9.14)
100 100 100
```

Out[19]:

	movie	movie_year	rating
0	\n1.\nThe Shawshank Redemption\n(1994)\n	(1994)	\n\n\n\n\n\n\n\n\n9.3\n
1	\n2.\nThe Godfather\n(1972)\n	(1972)	\n\n\n\n\n\n\n\n\n9.2\n
2	\n3.\nThe Godfather Part II\n(1974)\n	(1974)	\n\n\n\n\n\n\n\n\n9\n
3	\n4.\nThe Dark Knight\n(2008)\n	(2008)	\n\n\n\n\n\n\n\n\n9\n
4	\n5.\n12 Angry Men\n(1957)\n	(1957)	\n\n\n\n\n\n\n\n\n9\n
...	...	...	...
95	\n96.\nNorth by Northwest\n(1959)\n	(1959)	\n\n\n\n\n\n\n\n\n8.3\n
96	\n97.\nA Clockwork Orange\n(1971)\n	(1971)	\n\n\n\n\n\n\n\n\n8.3\n
97	\n98.\nSnatch\n(2000)\n	(2000)	\n\n\n\n\n\n\n\n\n8.2\n
98	\n99.\nLe fabuleux destin d'Amélie Poulain\n(2...	(2001)	\n\n\n\n\n\n\n\n\n8.3\n
99	\n100.\nThe Kid\n(1921)\n	(1921)	\n\n\n\n\n\n\n\n\n8.3\n

100 rows × 3 columns

In [20]:
```
# 3.PROGRAM TO DISPLAY IMBD'S TOP 100 RATED INDIAN MOVIES

# install libraries
!pip install bs4
!pip install requests

from bs4 import BeautifulSoup
import requests

#send get requests to the webpage server to get the source code of the page

page = requests.get('https://www.imdb.com/list/ls009997493/')
page

# page content

soup=BeautifulSoup(page.content)
soup

# creating a list of top rated imbd's 100 movies

movie_title = soup.find('h3',class_='lister-item-header')
movie_title

movie_title.text

movie_year = soup.find('span',class_='lister-item-year text-muted unbold')
movie_year

movie_year.text

rating = soup.find('span',class_='ipl-rating-star__rating')
rating

rating.text

movies=[]

for i in soup.find_all('h3',class_='lister-item-header'):
    movies.append(i.text)

movie

movie_year=[]

for i in soup.find_all('span',class_='lister-item-year text-muted unbold'):
    movie_year.append(i.text)

movie_year

rating=[]

for i in soup.find_all('div',class_='ipl-rating-star small'):
    rating.append(i.text)

rating

print(len(movie),len(movie_year),len(rating))       #printing length

# making dataframe

import pandas as pd

df=pd.DataFrame({'movie':movie,'movie_year':movie_year,'rating':rating})
df
```

```
Requirement already satisfied: bs4 in c:\users\dell\anaconda3\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\dell\anaconda3\lib\site-packages (from bs4) (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\dell\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)
Requirement already satisfied: requests in c:\users\dell\anaconda3\lib\site-packages (2.28.1)
Requirement already satisfied: idna<4,>=2.5 in c:\users\dell\anaconda3\lib\site-packages (from requests) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\dell\anaconda3\lib\site-packages (from requests) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2022.9.14)
100 100 100
```

Out[20]:

	movie	movie_year	rating
0	\n1.\nRang De Basanti\n(2006)\n	(2006)	\n\n\n\n\n\n\n\n\n8.1\n
1	\n2.\n3 Idiots\n(2009)\n	(2009)	\n\n\n\n\n\n\n\n\n8.4\n
2	\n3.\nTaare Zameen Par\n(2007)\n	(2007)	\n\n\n\n\n\n\n\n\n8.4\n
3	\n4.\nDil Chahta Hai\n(2001)\n	(2001)	\n\n\n\n\n\n\n\n\n8.1\n
4	\n5.\nSwades: We, the People\n(2004)\n	(2004)	\n\n\n\n\n\n\n\n\n8.2\n
...	...	...	...
95	\n96.\nWake Up Sid\n(2009)\n	(2009)	\n\n\n\n\n\n\n\n\n7.6\n
96	\n97.\nRangeela\n(1995)\n	(1995)	\n\n\n\n\n\n\n\n\n7.4\n
97	\n98.\nShatranj Ke Khilarii\n(1977)\n	(1977)	\n\n\n\n\n\n\n\n\n7.5\n
98	\n99.\nPyaar Ka Punchnama\n(2011)\n	(2011)	\n\n\n\n\n\n\n\n\n7.6\n
99	\n100.\nEk Hasina Thi\n(2004)\n	(2004)	\n\n\n\n\n\n\n\n\n7.5\n

100 rows × 3 columns

In [21]:
```
# 4.PROGRAM TO DISPLAY LIST OF RESPECTED FORMER PRESIDENT OF INDIA

#install libraries
!pip install bs4
!pip install requests

from bs4 import BeautifulSoup
import requests

#send get requests to the webpage server to get the source code of the page

page = requests.get('https://presidentofindia.nic.in/former-presidents.htm')
page

#page content

soup=BeautifulSoup(page.content)
soup

#creating a list of respected former president of india

name = soup.find('div',class_='presidentListing')
name

name.text

for i in soup.find_all('div',class_='presidentListing'):
    name.append(i.text)

name
```

```
Requirement already satisfied: bs4 in c:\users\dell\anaconda3\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\dell\anaconda3\lib\site-packages (from bs4) (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\dell\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)
Requirement already satisfied: requests in c:\users\dell\anaconda3\lib\site-packages (2.28.1)
Requirement already satisfied: idna<4,>=2.5 in c:\users\dell\anaconda3\lib\site-packages (from requests) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\dell\anaconda3\lib\site-packages (from requests) (1.26.11)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\dell\anaconda3\lib\site-packages (from requests) (2022.9.14)
```

Out[21]:
```
<div class="presidentListing">
<h3>Shri Ram Nath Kovind (birth - 1945)</h3>
<p><span class="terms">Term of Office:</span> 25 July, 2017 to 25 July, 2022 </p>
<p><a href="https://ramnathkovind.nic.in" target="_blank">https://ramnathkovind.nic.in</a></p>
```

```
Shri Ram Nath Kovind (birth - 1945)
Term of Office: 25 July, 2017 to 25 July, 2022
https://ramnathkovind.nic.in

Shri Pranab Mukherjee (1935-2020)
Term of Office: 25 July, 2012 to 25 July, 2017
http://pranabmukherjee.nic.in

Smt Pratibha Devisingh Patil (birth - 1934)
Term of Office: 25 July, 2007 to 25 July, 2012
http://pratibhapatil.nic.in

DR. A.P.J. Abdul Kalam (1931-2015)
Term of Office: 25 July, 2002 to 25 July, 2007
http://abdulkalam.nic.in

Shri K. R. Narayanan (1920 - 2005)
Term of Office: 25 July, 1997 to 25 July, 2002

Dr Shankar Dayal Sharma (1918-1999)
Term of Office: 25 July, 1992 to 25 July, 1997

Shri R Venkataraman (1910-2009)
Term of Office: 25 July, 1987 to 25 July, 1992

Giani Zail Singh (1916-1994)
Term of Office: 25 July, 1982 to 25 July, 1987

Shri Neelam Sanjiva Reddy (1913-1996)
Term of Office: 25 July, 1977 to 25 July, 1982

Dr. Fakhruddin Ali Ahmed (1905-1977)
Term of Office: 24 August, 1974 to 11 February, 1977

Shri Varahagiri Venkata Giri (1894-1980)
Term of Office: 3 May, 1969 to 20 July, 1969 and 24 August, 1969 to 24 August, 1974

Dr. Zakir Husain (1897-1969)
Term of Office: 13 May, 1967 to 3 May, 1969

Dr. Sarvepalli Radhakrishnan (1888-1975)
Term of Office: 13 May, 1962 to 13 May, 1967

Dr. Rajendra Prasad (1884-1963)
Term of Office: 26 January, 1950 to 13 May, 1962
</div>
```

In [ ]:
```

```