

Web Scrapping

1. Write a Python program of your Top Ten favorite movies, the year of the movie and the synopsis.

```
In [8]: from requests_html import HTMLSession
session = HTMLSession()
url='https://www.imdb.com/chart/top/'
r=session.get(url)

link=r.html.find('td.titleColumn a[href]')
movie_name =[link[i].text for i in range(len(link))]
url =[list(link[i].absolute_links) for i in range(len(link))]

year =r.html.find('td.titleColumn span')
year =[int(year[i].text.strip('()')) for i in range(len(year))]

from requests_html import AsyncHTMLSession
session=AsyncHTMLSession()
synopse=[]
for i in range(10):
    url_movie=url[i][0]
    r = await session.get(url_movie)
    await r.html.arender(timeout=10,sleep=10)

    syn=r.html.find('div.summary_text')
    synopse.append(syn[0].text)

import pandas as pd

movie_top10=pd.DataFrame()
movie_top10["Movie Name"] = movie_name[0:10]
movie_top10["Year"]=year[0:10]
movie_top10["Synopsis"]=synopse

pd.set_option('display.max_colwidth',-1)
movie_top10

<ipython-input-8-0d846270e096>:32: FutureWarning: Passing a negative integer is deprecated in version 1.0 and will not be supported in future version. Instead, use None to not limit the column width.
pd.set_option('display.max_colwidth',-1)
```

Out[8]:

	Movie Name	Year	Synopse
0	The Shawshank Redemption	1994	Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.
1	The Godfather	1972	An organized crime dynasty's aging patriarch transfers control of his clandestine empire to his reluctant son.
2	The Godfather: Part II	1974	The early life and career of Vito Corleone in 1920s New York City is portrayed, while his son, Michael, expands and tightens his grip on the family crime syndicate.
3	The Dark Knight	2008	When the menace known as the Joker wreaks havoc and chaos on the people of Gotham, Batman must accept one of the greatest psychological and physical tests of his ability to fight injustice.
4	12 Angry Men	1957	A jury holdout attempts to prevent a miscarriage of justice by forcing his colleagues to reconsider the evidence.
5	Schindler's List	1993	In German-occupied Poland during World War II, industrialist Oskar Schindler gradually becomes concerned for his Jewish workforce after witnessing their persecution by the Nazis.
6	The Lord of the Rings: The Return of the King	2003	Gandalf and Aragorn lead the World of Men against Sauron's army to draw his gaze from Frodo and Sam as they approach Mount Doom with the One Ring.
7	Pulp Fiction	1994	The lives of two mob hitmen, a boxer, a gangster and his wife, and a pair of diner bandits intertwine in four tales of violence and redemption.
8	The Good, the Bad and the Ugly	1966	A bounty hunting scam joins two men in an uneasy alliance against a third in a race to find a fortune in gold buried in a remote cemetery.
9	The Lord of the Rings: The Fellowship of the Ring	2001	A meek Hobbit from the Shire and eight companions set out on a journey to destroy the powerful One Ring and save Middle-earth from the Dark Lord Sauron.

2.Write a Python program to find the live weather report (Temperature) in Tokyo, Osaka and Sapporo.

```
In [33]: import requests
import json
API_key='1e1f1b7d456fc67d53e3a50216d65374'
city_list=list(input('list of city names:').split())
for city_name in city_list:
    url =('http://api.openweathermap.org/data/2.5/weather?q={}&appid={}&units=metric'

    response=requests.get(url)

    print("\nCity: {}".format(city_name))
    print("Temp:{} degC".format(response.json()['main']['temp']))
    print("Wind Speed:{} m/s".format(response.json()['wind']['speed']))
    print("Description:{}".format(response.json()['weather'][0]['description']))

list of city names:Tokyo Osaka Sapporo

City: Tokyo
Temp:10.44 degC
Wind Speed:9.26 m/s
Description:clear sky

City: Osaka
Temp:10.32 degC
Wind Speed:3.6 m/s
Description:broken clouds

City: Sapporo
Temp:-1 degC
Wind Speed:9.26 m/s
Description:broken clouds
```

3.Write a Python program to get the number of magnitude 5+ earthquakes detected worldwide by BGS.

modified: write a prog to get the data for recent earthquake world wide

```
In [41]: import pandas as pd

url='https://earthquakes.bgs.ac.uk/earthquakes/recent_world_events.html'
table = pd.read_html(url)
pd.set_option('display.max_colwidth',-1)
table[0]

<ipython-input-41-0c52ac8b8ad7>:5: FutureWarning: Passing a negative integer is deprecated in version 1.0 and will not be supported in future version. Instead, use None to not limit the column width.
pd.set_option('display.max_colwidth',-1)
```

Out[41]:

	Date	Time (UTC)	Lat	Lon	Depth (km)	Mag	Int	Region	Comment
0	2021/01/23	23:36:51.3	-61.825	-55.494	10	6.9	4	SOUTH SHETLAND ISLANDS	NaN
1	2021/01/21	14:27:06.0	34.960	33.750	60	4.9	5	CYPRUS	FELT CYPRUS...
2	2021/01/21	12:23:06.0	5.007	127.517	95	7.0	5	TALAUD,INDONESIA	NaN
3	2021/01/19	02:46:22.0	-31.818	-68.847	20	6.4	6	SAN JUAN,ARGENTINA	NaN
4	2021/01/14	18:28:18.0	-3.005	118.924	18	6.2	9	SULAWESI,INDONSIA	WEST SULAWESI
5	2021/01/11	21:32:58.0	51.240	100.437	10	6.7	5	NORTHERN MONGOLIA	KHUVSGUL PROVINCE
6	2020/12/29	11:19:53.3	45.422	16.255	10	6.4	8	CROATIA	NaN
7	2020/12/27	21:39:15.2	-39.343	-74.990	10	6.7	5	LOS RIOS,CHILE	OFFSHORE LOCATION
8	2020/10/30	11:51:29.1	37.926	26.485	21	7.0	9	EASTERN AEGEAN SEA	NaN
9	2020/10/25	19:35:43.2	46.941	9.342	10	4.6	4	SWITZERLAND	FELT SWITZERLAND...
10	2020/10/19	20:54:37.5	54.289	-160.946	33	7.6	7	ALASKA PENINSULA	SHUMAGIN BANK REGION
11	2020/10/01	01:13:36.8	-19.539	-174.122	28	6.5	5	TONGA ISLANDS REGION	NaN
12	2020/09/18	21:43:47.9	-0.219	-28.260	10	6.9	2	MID-ATLANTIC RIDGE	CENTRAL MID-ATLANTIC
13	2020/09/06	06:51:06.2	6.508	-38.753	10	6.6	2	MID-ATLANTIC RIDGE	CENTRAL MID-ATLANTIC
14	2020/09/01	21:09:19.1	-27.916	-71.370	16	6.5	5	ATACAMA,CHILE	OFFSHORE LOCATION
15	2020/09/01	04:30:03.4	-28.032	-71.266	16	6.3	5	ATACAMA,CHILE	OFFSHORE LOCATION
16	2020/09/01	04:09:28.9	-27.969	-71.306	21	6.8	7	ATACAMA,CHILE	OFFSHORE LOCATION
17	2020/08/30	21:20:38.6	2.485	-30.736	10	6.5	2	MID-ATLANTIC RIDGE	CENTRAL MID-ATLANTIC
18	2020/08/21	04:09:51.3	-6.710	123.465	634	6.9	4	BANDA SEA	NaN
19	2020/08/18	22:29:24.4	-4.207	101.241	26	6.9	4	SUMATRA,INDONESIA	OFF SOUTHERN SUMATRA
20	2020/08/18	22:23:59.4	-4.322	101.135	22	6.8	4	SUMATRA,INDONESIA	OFF SOUTHERN SUMATRA
21	2020/08/18	00:03:50.2	12.026	124.127	10	6.6	7	SAMAR,PHILIPPINES	NaN
22	2020/08/05	12:05:33.9	-16.094	168.065	181	6.5	4	VANUATU	NaN
23	2020/07/22	06:12:46.4	55.097	-157.905	33	7.8	5	ALASKA PENINSULA	SHUMAGIN BANK REGION
24	2020/07/17	02:50:19.6	-7.836	147.770	60	7.0	5	PAPUA NEW GUINEA	NaN

```
In [ ]:
```