ASSIGNMENT-1 WEB SCRAPING

In [9]:
```
# !pip install bs4
# !pip install requests
```

1) Write a python program to display all the header tags from wikipedia.org and make data frame.

In [18]:
```
from bs4 import BeautifulSoup
import pandas as pd
import requests

def get_headers(url):
    response = requests.get(url)
    sp = BeautifulSoup(response.content)
    headers = []
    for header in sp.find_all('span',class_="mw-headline"):
        headers.append(header.text)
    df = pd.DataFrame({'Titles':headers})
    return df

headers = get_headers('https://en.wikipedia.org/wiki/Main_Page')
headers
```

Out[18]:

	Titles
0	Welcome to Wikipedia
1	From today's featured article
2	Did you know ...
3	In the news
4	On this day
5	Today's featured picture
6	Other areas of Wikipedia
7	Wikipedia's sister projects
8	Wikipedia languages

2) Write s python program to display list of respected former presidents of India(i.e. Name , Term ofoffice) from https://presidentofindia.nic.in/former-presidents.htm and make data frame.

In [19]:
```
from bs4 import BeautifulSoup
import pandas as pd
import requests

def get_former_president(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    president_names = []
    president_terms = []
    for i in soup.find_all('div',class_="presidentListing"):
        president_names.append(i.text.split('\n')[1].split('(')[0].strip())
        president_terms.append(i.text.split('\n')[2].split(':')[1].strip())
    df = pd.DataFrame({"Presidents": president_names, "Former": president_terms})
    return df
get_former_president('https://presidentofindia.nic.in/former-presidents.htm')
```

Out[19]:

	Presidents	Former
0	Shri Ram Nath Kovind	25 July, 2017 to 25 July, 2022
1	Shri Pranab Mukherjee	25 July, 2012 to 25 July, 2017
2	Smt Pratibha Devisingh Patil	25 July, 2007 to 25 July, 2012
3	DR. A.P.J. Abdul Kalam	25 July, 2002 to 25 July, 2007
4	Shri K. R. Narayanan	25 July, 1997 to 25 July, 2002
5	Dr Shankar Dayal Sharma	25 July, 1992 to 25 July, 1997
6	Shri R Venkataraman	25 July, 1987 to 25 July, 1992
7	Giani Zail Singh	25 July, 1982 to 25 July, 1987
8	Shri Neelam Sanjiva Reddy	25 July, 1977 to 25 July, 1982
9	Dr. Fakhruddin Ali Ahmed	24 August, 1974 to 11 February, 1977
10	Shri Varahagiri Venkata Giri	3 May, 1969 to 20 July, 1969 and 24 August, 19...
11	Dr. Zakir Husain	13 May, 1967 to 3 May, 1969
12	Dr. Sarvepalli Radhakrishnan	13 May, 1962 to 13 May, 1967
13	Dr. Rajendra Prasad	26 January, 1950 to 13 May, 1962

3) Write a python program to scrape cricket rankings from icc-cricket.com. You have to scrape and make data frame-

a) Top 10 ODI teams in men's cricket along with the records for matches, points and rating.

b) Top 10 ODI Batsmen along with the records of their team and rating.

c) Top 10 ODI bowlers along with the records of their team and rating.

In [12]:
```
from bs4 import BeautifulSoup
import pandas as pd
import requests

def men_odi_teams_rankings(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    table = soup.find_all('table')[0]
    rows = table.find_all('tr')

    team_data = []
    for tr in rows[1:11]:
        cols = tr.find_all('td')
        team_name = cols[1].find('span',class_="u-hide-phablet").text
        matches = cols[2].text
        points = cols[3].text
        ratings = cols[4].text.strip()
        team_data.append([team_name, matches, points, ratings])

    df = pd.DataFrame(team_data, columns=['Team', 'Matches', 'Points', 'Rating'])
    return df

def men_odi_batsmen_rankings(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    table = soup.find_all('table')[0]
    rows = table.find_all('tr')

    batsman_data = []
    for row in rows[1:11]:
        cols = row.find_all('td')
        batsman = cols[1].text.strip()
        team = cols[2].text.strip()
        rating = cols[3].text.strip()
        batsman_data.append([batsman, team, rating])

    df = pd.DataFrame(batsman_data, columns=['Batsman', 'Team', 'Rating'])
    return df

def men_odi_bowlers_rankings(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    table = soup.find_all('table')[0]
    rows = table.find_all('tr')

    bowler_data = []
    for row in rows[1:11]:
        cols = row.find_all('td')
        bowler = cols[1].text.strip()
        team = cols[2].text.strip()
        rating = cols[3].text.strip()
        bowler_data.append([bowler, team, rating])

    df = pd.DataFrame(bowler_data, columns=['Bowler', 'Team', 'Rating'])
    return df


odi_teams_url = 'https://www.icc-cricket.com/rankings/mens/team-rankings/odi'
odi_batsmen_url = 'https://www.icc-cricket.com/rankings/mens/player-rankings/odi/batting'
odi_bowlers_url = 'https://www.icc-cricket.com/rankings/mens/player-rankings/odi/bowling'

print("Top 10 ODI teams in men's cricket:\n",men_odi_teams_rankings(odi_teams_url))

print("\nTop 10 ODI batsmen:\n", men_odi_batsmen_rankings(odi_batsmen_url) )

print("\nTop 10 ODI bowlers:\n", men_odi_bowlers_rankings(odi_bowlers_url))
```

```
Top 10 ODI teams in men's cricket:
            Team Matches Points Rating
0     Australia      23  2,714    118
1      Pakistan      20  2,316    116
2         India      33  3,807    115
3   New Zealand      27  2,806    104
4       England      24  2,426    101
5  South Africa      19  1,910    101
6    Bangladesh      25  2,451     98
7     Sri Lanka      28  2,378     85
8   Afghanistan      13  1,067     82
9   West Indies      32  2,201     69

Top 10 ODI batsmen:
                  Batsman Team Rating
0             Babar Azam  PAK    886
1  Rassie van der Dussen   SA    777
2           Fakhar Zaman  PAK    755
3            Imam-ul-Haq  PAK    745
4           Shubman Gill  IND    738
5           Harry Tector  IRE    726
6           David Warner  AUS    726
7            Virat Kohli  IND    719
8        Quinton de Kock   SA    718
9           Rohit Sharma  IND    707

Top 10 ODI bowlers:
              Bowler Team Rating
0    Josh Hazlewood  AUS    705
1    Mohammed Siraj  IND    691
2    Mitchell Starc  AUS    686
3        Matt Henry   NZ    667
4       Trent Boult   NZ    660
5        Adam Zampa  AUS    652
6       Rashid Khan  AFG    640
7    Shaheen Afridi  PAK    630
8  Mujeeb Ur Rahman  AFG    630
9     Mohammad Nabi  AFG    626
```

Write a python program to scrape cricket rankings from icc-cricket.com. You have to scrape and make data frame-

a) Top 10 ODI teams in women's cricket along with the records for matches, points and rating.

In [13]:
```
from bs4 import BeautifulSoup
import pandas as pd
import requests

def women_odi_teams_rankings(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    table = soup.find_all('table')[0]
    rows=table.find_all('tr')

    team_data = []
    for tr in rows[1:11]:
        cols = tr.find_all('td')
        team_name = (cols[1].find('span',class_="u-hide-phablet").text)
        matches = (cols[2].text)
        points = (cols[3].text)
        ratings = (cols[4].text.strip())
        team_data.append([team_name, matches, points, ratings])

    df=pd.DataFrame(team_data, columns = ['Team', 'Matches', 'Points', 'Ratings'])
    return df

ODI_ranking = women_odi_teams_rankings('https://www.icc-cricket.com/rankings/womens/team-rankings/odi')

print('Top 10 ODI teams in women's cricket :')
print(ODI_ranking)
```

```
Top 10 ODI teams in women's cricket :
           Team Matches Points Ratings
0     Australia      21  3,603     172
1       England      28  3,342     119
2  South Africa      26  3,098     119
3         India      27  2,820     104
4   New Zealand      28  2,688      96
5   West Indies      29  2,743      95
6    Bangladesh      14    977      70
7     Sri Lanka      12    820      68
8      Thailand      12    806      67
9      Pakistan      27  1,678      62
```

b) Top 10 women's ODI Batting players along with the records of their team and rating.

In [14]:
```
def women_odi_batsmen_rankings(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    table = soup.find_all('table')[0]
    batsmen_data = []
```

```
    #Top player class_name is different
    top_player_row = table.find('tr',class_="rankings-block__banner")
    top_player_cols = top_player_row.find_all('td')
    player_name = top_player_cols[1].text.strip()
    team = top_player_cols[2].text.strip()
    ratings = top_player_cols[3].text.strip()
    batsmen_data.append({'Player_name': player_name, 'Team' : team, 'Ratings' : ratings})

    rows = table.find_all('tr',class_="table-body")
    for tr in rows[1:11]:
        cols = tr.find_all('td')
        player_name = (cols[1].text.strip())
        team = (cols[2].text.strip())
        ratings = (cols[3].text.strip())
        batsmen_data.append({'Player_name': player_name, 'Team' : team, 'Ratings' : ratings})

    df = pd.DataFrame(batsmen_data)
    return df
batsmen_ranking = women_odi_batsmen_rankings('https://www.icc-cricket.com/rankings/womens/player-rankings/odi/batting')
print('Top 10 ODI women's Batsmen :\n')
print(batsmen_ranking)
```

```
Top 10 ODI women's Batsmen :

            Player_name Team Ratings
0   Chamari Athapaththu   SL     758
1       Laura Wolvaardt   SA     732
2        Natalie Sciver  ENG     731
3           Meg Lanning  AUS     717
4      Harmanpreet Kaur  IND     716
5       Smriti Mandhana  IND     714
6          Ellyse Perry  AUS     626
7       Stafanie Taylor   WI     618
8        Tammy Beaumont  ENG     595
9           Amelia Kerr   NZ     591
10       Marizanne Kapp   SA     585
```

c) Top 10 women's ODI all-rounder along with the records of their team and rating.

In [15]:

```
def women_odi_all_rounfer_ranking(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    table = soup.find_all('table')[0]
    all_rounder_data = []
    #Top player class_name is different
    top_player_row = table.find('tr',class_="rankings-block__banner")
    top_player_cols = top_player_row.find_all('td')
    player_name = top_player_cols[1].text.strip()
    team = top_player_cols[2].text.strip()
    ratings = top_player_cols[3].text.strip()
    allrounders_data.append({'All-Rounder': allrounder, 'Team': team, 'Rating': rating})

    rows = table.find_all('tr',class_="table-body")
    for tr in rows[1:11]:
        cols = tr.find_all('td')
        player_name = (cols[1].text.strip())
        team = (cols[2].text.strip())
        ratings = (cols[3].text.strip())
        allrounders_data.append({'All-Rounder': allrounder, 'Team': team, 'Rating': rating})

    df = pd.DataFrame(allrounders_data)
    return df
batsmen_ranking = women_odi_batsmen_rankings('https://www.icc-cricket.com/rankings/womens/player-rankings/odi/all-rounder')
print('Top 10 women's ODI all-rounder :\n')
print(batsmen_ranking)
```

```
Top 10 women's ODI all-rounder :

            Player_name Team Ratings
0       Hayley Matthews   WI     382
1          Ellyse Perry  AUS     366
2        Marizanne Kapp   SA     349
3           Amelia Kerr   NZ     328
4         Deepti Sharma  IND     322
5      Ashleigh Gardner  AUS     292
6         Jess Jonassen  AUS     250
7         Sophie Devine   NZ     233
8              Nida Dar  PAK     232
9     Sophie Ecclestone  ENG     205
10  Chamari Athapaththu   SL     200
```

5) Write a python program to scrape mentioned news details from https://www.cnbc.com/world/?region=world and make data frame-

i) Headline

ii) Time

iii) News Link

In [20]:
```
from bs4 import BeautifulSoup
import pandas as pd
import requests

def scrape_news_details(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)
    news_data = []
    news_articles = soup.find_all('div',class_="RiverPlusCard-cardLeft")
    for article in news_articles:
        headline = article.find('a').text.strip()
        istime = article.find('span', class_="RiverByline-datePublished")
        time = istime.text.strip() if istime else " "
        news_link = article.find('a')['href']
        news_data.append({'Headline': headline, 'Time': time, 'News Link': news_link})
    df = pd.DataFrame(news_data)
    return df
scrape_news_details('https://www.cnbc.com/world/?region=world')
```

Out[20]:

	Headline	Time	News Link
0	Stocks tumble on Friday, notching weekly losse...		https://www.cnbc.com/2023/07/06/stock-market-t...
1			/pro/
2	UK set to ease stock market listing rules		https://www.cnbc.com/2023/07/08/uk-set-to-ease...
3	Ukraine reports advances near eastern city of ...		https://www.cnbc.com/2023/07/07/ukraine-war-li...
4	U.S. payrolls rose by 209,000 in June, less th...		https://www.cnbc.com/2023/07/07/jobs-report-ju...
5	Yellen urges China to support existing institu...		https://www.cnbc.com/2023/07/08/yellen-urges-c...
6			/pro/
7	What Apple's big bet on India means for the te...		https://www.cnbc.com/2023/07/07/what-apples-bi...
8	China hits Alibaba affiliate Ant Group with $9...		https://www.cnbc.com/2023/07/07/china-hits-ali...
9	Fed's Goolsbee sees 'golden path' to lower inf...		https://www.cnbc.com/2023/07/07/feds-goolsbee-...
10	'We are in uncharted territory': World records...		https://www.cnbc.com/2023/07/07/climate-world-...
11	Twitter's desktop app sees surge in outages as...		https://www.cnbc.com/2023/07/07/twitter-deskto...
12	Britain's house prices are slumping as mortgag...		https://www.cnbc.com/2023/07/07/britains-house...
13	A 'nice' workplace culture may be more toxic t...		https://www.cnbc.com/2023/07/07/nice-workplace...
14	'We would not stand idly by': Lagarde pledges ...		https://www.cnbc.com/2023/07/07/we-would-not-s...
15	Yellen says she's 'concerned' about China's ne...		https://www.cnbc.com/2023/07/07/yellen-says-sh...
16	Alibaba launches A.I. tool to generate images ...		https://www.cnbc.com/2023/07/07/alibaba-launch...
17	Countries agree to slash shipping emissions bu...		https://www.cnbc.com/2023/07/07/countries-agre...

Write a python program to scrape the details of most downloaded articles from AI in last 90 days. https://www.journals.elsevier.com/artificial-intelligence/most-downloaded-articles

Scrape below mentioned details and make data frame-

i) Paper Title

ii) Authors

iii) Published Date

iv) Paper URL

In [21]:
```
import requests
from bs4 import BeautifulSoup
import pandas as pd

def scrape_most_downloaded_articles(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.content)

    articles = soup.find_all('li', class_="sc-9zxyh7-1 sc-9zxyh7-2 kOEIEO hvoVxs")

    titles = []
    authors = []
    dates = []
    urls = []

    for article in articles:
        title = article.find('h2', class_="sc-1qrq3sd-1 gRGSUS sc-1nmom32-0 sc-1nmom32-1 btcbYu goSKRg").text.strip()
        author = article.find('span', class_="sc-1w3fpd7-0 dnCnAO").text.strip()
        date = article.find('span', class_="sc-1thf9ly-2 dvggWt").text.strip()
        url = article.find('a', class_="sc-5smygv-0 fIXTHm")['href']

        titles.append(title)
```

```
        authors.append(author)
        dates.append(date)
        urls.append(url)

    data = {'Paper Title': titles,'Authors': authors, 'Published Date': dates, 'Paper URL': urls}

    df = pd.DataFrame(data)
    return df
articles_df = scrape_most_downloaded_articles('https://www.journals.elsevier.com/artificial-intelligence/most-downloaded-articles
articles_df
```

Out[21]:

	Paper Title	Authors	Published Date	Paper URL
0	Reward is enough	David Silver, Satinder Singh, Doina Precup, Ri...	October 2021	https://www.sciencedirect.com/science/article/...
1	Explanation in artificial intelligence: Insigh...	Tim Miller	February 2019	https://www.sciencedirect.com/science/article/...
2	Creativity and artificial intelligence	Margaret A. Boden	August 1998	https://www.sciencedirect.com/science/article/...
3	Conflict-based search for optimal multi-agent ...	Guni Sharon, Roni Stern, Ariel Felner, Nathan ...	February 2015	https://www.sciencedirect.com/science/article/...
4	Knowledge graphs as tools for explainable mach...	Ilaria Tiddi, Stefan Schlobach	January 2022	https://www.sciencedirect.com/science/article/...
5	Law and logic: A review from an argumentation ...	Henry Prakken, Giovanni Sartor	October 2015	https://www.sciencedirect.com/science/article/...
6	Between MDPs and semi-MDPs: A framework for te...	Richard S. Sutton, Doina Precup, Satinder Singh	August 1999	https://www.sciencedirect.com/science/article/...
7	Explaining individual predictions when feature...	Kjersti Aas, Martin Jullum, Anders Løland	September 2021	https://www.sciencedirect.com/science/article/...
8	Multiple object tracking: A literature review	Wenhan Luo, Junliang Xing and 4 more	April 2021	https://www.sciencedirect.com/science/article/...
9	A survey of inverse reinforcement learning: Ch...	Saurabh Arora, Prashant Doshi	August 2021	https://www.sciencedirect.com/science/article/...
10	Evaluating XAI: A comparison of rule-based and...	Jasper van der Waa, Elisabeth Nieuwburg, Anita...	February 2021	https://www.sciencedirect.com/science/article/...
11	Explainable AI tools for legal reasoning about...	Joe Collenette, Katie Atkinson, Trevor Bench-C...	April 2023	https://www.sciencedirect.com/science/article/...
12	Hard choices in artificial intelligence	Roel Dobbe, Thomas Krendl Gilbert, Yonatan Mintz	November 2021	https://www.sciencedirect.com/science/article/...
13	Assessing the communication gap between AI mod...	Oskar Wysocki, Jessica Katharine Davies and 5 ...	March 2023	https://www.sciencedirect.com/science/article/...
14	Explaining black-box classifiers using post-ho...	Eoin M. Kenny, Courtney Ford, Molly Quinn, Mar...	May 2021	https://www.sciencedirect.com/science/article/...
15	The Hanabi challenge: A new frontier for AI re...	Nolan Bard, Jakob N. Foerster and 13 more	March 2020	https://www.sciencedirect.com/science/article/...
16	Wrappers for feature subset selection	Ron Kohavi, George H. John	December 1997	https://www.sciencedirect.com/science/article/...
17	Artificial cognition for social human–robot in...	Séverin Lemaignan, Mathieu Warnier and 3 more	June 2017	https://www.sciencedirect.com/science/article/...
18	A review of possible effects of cognitive bias...	Tomáš Kliegr, Štěpán Bahník, Johannes Fürnkranz	June 2021	https://www.sciencedirect.com/science/article/...
19	The multifaceted impact of Ada Lovelace in the...	Luigia Carlucci Aiello	June 2016	https://www.sciencedirect.com/science/article/...
20	Robot ethics: Mapping the issues for a mechani...	Patrick Lin, Keith Abney, George Bekey	April 2011	https://www.sciencedirect.com/science/article/...
21	Reward (Mis)design for autonomous driving	W. Bradley Knox, Alessandro Allievi and 3 more	March 2023	https://www.sciencedirect.com/science/article/...
22	Planning and acting in partially observable st...	Leslie Pack Kaelbling, Michael L. Littman, Ant...	May 1998	https://www.sciencedirect.com/science/article/...
23	What do we want from Explainable Artificial In...	Markus Langer, Daniel Oster and 6 more	July 2021	https://www.sciencedirect.com/science/article/...

In [ ]: