In [1]:

```
!pip install bs4
!pip install requests
```

```
Requirement already satisfied: bs4 in c:\users\admin\anaconda3\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\users\admin\anaconda3\lib\site-packages (from bs4) (4.11.1)
Requirement already satisfied: soupsieve>1.2 in c:\users\admin\anaconda3\lib\site-packages (from beautifulsoup4->bs4) (2.3.2.post1)
Requirement already satisfied: requests in c:\users\admin\anaconda3\lib\site-packages (2.28.1)
Requirement already satisfied: idna<4,>=2.5 in c:\users\admin\anaconda3\lib\site-packages (from requests) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\admin\anaconda3\lib\site-packages (from requests) (1.26.14)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\admin\anaconda3\lib\site-packages (from requests) (2022.12.7)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\admin\anaconda3\lib\site-packages (from requests) (2.0.4)
```

In [2]:

```
from bs4 import BeautifulSoup
import requests
```

In [3]:

```
page = requests.get('https://en.wikipedia.org/wiki/Main_Page')
```

In [4]:

```
page
```

Out[4]:

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

In [19]:

```
headline = []

for i in page.find_all('span',class_="mw-headline"):
   headline.append(i.text)

headline
```

Out[19]:

```
['Welcome to Wikipedia',
 "From today's featured article",
 'Did you know\xa0...',
 'In the news',
 'On this day',
 "Today's featured picture",
 'Other areas of Wikipedia',
 "Wikipedia's sister projects",
 'Wikipedia languages']
```

In [21]:

```
import pandas as pd
df = pd.DataFrame({'Headline':headline})
df
```

Out[21]:

	Headline
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

In [ ]:

```
page = requests.get('https://www.icc-cricket.com/rankings/womens/player-rankings/odi')
```

In [40]:

```
page
```

Out[40]:

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

In [15]:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

url = "https://www.icc-cricket.com/rankings/mens/team-rankings/odi"
response = requests.get(url)
soup = BeautifulSoup(response.content, "html.parser")

team_data = []
table = soup.find("table", class_="table")
rows = table.find_all("tr")

for row in rows[1:11]:
  cells = row.find_all("td")
  team = cells[1].text.strip()
  matches = cells[2].text.strip()
  points = cells[3].text.strip()
  rating = cells[4].text.strip()
  team_data.append([team, matches, points, rating])

df = pd.DataFrame(team_data, columns=["Team", "Matches", "Points", "Rating"])
print(df)
```

```
              Team Matches Points Rating
0        India\nIND      55  6,640    121
1    Australia\nAUS      42  4,926    117
2  South Africa\nSA      34  3,750    110
3     Pakistan\nPAK      36  3,922    109
4   New Zealand\nNZ      43  4,399    102
5      England\nENG      38  3,777     99
6     Sri Lanka\nSL      47  4,134     88
7   Bangladesh\nBAN      44  3,836     87
8  Afghanistan\nAFG      30  2,533     84
9   West Indies\nWI      38  2,582     68
```

In [16]:

```
url = "https://www.icc-cricket.com/rankings/mens/player-rankings/odi/batting"
response = requests.get(url)
soup = BeautifulSoup(response.content, "html.parser")

batsman_data = []
table = soup.find("table", class_="table")
rows = table.find_all("tr")
for row in rows[1:11]:
  cells = row.find_all("td")
  batsman = cells[1].text.strip()
  team = cells[2].text.strip()
  rating = cells[3].text.strip()
  batsman_data.append([batsman, team, rating])

df = pd.DataFrame(batsman_data, columns=["Batsman", "Team", "Rating"])
print(df)
```

```
                  Batsman Team Rating
0            Shubman Gill  IND    826
1              Babar Azam  PAK    824
2             Virat Kohli  IND    791
3            Rohit Sharma  IND    769
4         Quinton de Kock   SA    760
5          Daryl Mitchell   NZ    750
6            David Warner  AUS    745
7  Rassie van der Dussen   SA    735
8            Harry Tector  IRE    729
9             Dawid Malan  ENG    729
```

In [17]:

```
url = "https://www.icc-cricket.com/rankings/mens/player-rankings/odi/bowling"
response = requests.get(url)
soup = BeautifulSoup(response.content, "html.parser")

bowler_data = []
table = soup.find("table", class_="table")
rows = table.find_all("tr")

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

df = pd.DataFrame(bowler_data, columns=["Bowler", "Team", "Rating"])
print(df)
```

```
            Bowler Team Rating
0  Keshav Maharaj   SA    741
1   Josh Hazlewood  AUS    703
2  Mohammed Siraj  IND    699
3   Jasprit Bumrah  IND    685
4       Adam Zampa  AUS    675
5      Rashid Khan  AFG    667
6    Kuldeep Yadav  IND    667
7      Trent Boult   NZ    663
8  Shaheen Afridi  PAK    650
9  Mohammad Shami  IND    648
```

In [18]:

```
import requests
from bs4 import BeautifulSoup
```

In [19]:

```
page = requests.get('https://www.dineout.co.in/delhi-restaurants/buffet-special')

page
```

Out[19]:

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

In [21]:

```
restaurants = soup.find('div',class_="restnt-info cursor")
restaurants.text

restaurants = []

for i in soup.find_all('div',class_="restnt-info cursor"):

    restaurants.append(i.text)

restaurants
```

Out[21]:

```
['Castle BarbequeConnaught Place, Central Delhi',
 'Cafe KnoshThe Leela Ambience Convention Hotel,Shahdara, East Delhi',
 'India GrillHilton Garden Inn,Saket, South Delhi',
 'The Barbeque CompanyGardens Galleria,Sector 38A, Noida',
 'Delhi BarbequeTaurus Sarovar Portico,Mahipalpur, South Delhi',
 'The Monarch - Bar Be Que VillageIndirapuram Habitat Centre,Indirapuram, Ghaziabad',
 'The Barbeque TimesM2K Corporate Park,Sector 51, Gurgaon']
```

In [22]:

```
cuisine = []

for i in soup.find_all('span',class_="double-line-ellipsis"):

    cuisine.append(i.text)

cuisine
```

Out[22]:

```
['₹ 2,000 for 2 (approx) | Chinese, North Indian',
 '₹ 3,000 for 2 (approx) | Italian, Continental',
 '₹ 2,400 for 2 (approx) | North Indian, Italian',
 '₹ 1,700 for 2 (approx) | North Indian, Chinese',
 '₹ 1,800 for 2 (approx) | North Indian',
 '₹ 1,900 for 2 (approx) | North Indian',
 '₹ 1,500 for 2 (approx) | North Indian, Continental, Chinese, South Indian']
```

In [23]:

```
loc = soup.find('div',class_="restnt-loc ellipsis")
loc

loc.text
```

Out[23]:

```
'Connaught Place, Central Delhi'
```

In [24]:

```
location = []

for i in soup.find_all('div',class_="restnt-loc ellipsis"):
    location.append(i.text)

location
```

Out[24]:

```
['Connaught Place, Central Delhi',
 'The Leela Ambience Convention Hotel,Shahdara, East Delhi',
 'Hilton Garden Inn,Saket, South Delhi',
 'Gardens Galleria,Sector 38A, Noida',
 'Taurus Sarovar Portico,Mahipalpur, South Delhi',
 'Indirapuram Habitat Centre,Indirapuram, Ghaziabad',
 'M2K Corporate Park,Sector 51, Gurgaon']
```

In [25]:

```
images = []

for i in soup.find_all('img',class_="no-img"):
    images.append(i['data-src'])
images
```

Out[25]:

```
['https://im1.dineout.co.in/images/uploads/restaurant/sharpen/8/k/b/p86792-16062953735fbe1f4d3fb7e.jpg?tr=tr:n-medium',
 'https://im1.dineout.co.in/images/uploads/restaurant/sharpen/4/p/m/p406-15438184745c04ccea491bc.jpg?tr=tr:n-medium',
 'https://im1.dineout.co.in/images/uploads/restaurant/sharpen/2/q/t/p2687-16958938576515496lea87c.jpg?tr=tr:n-medium',
 'https://im1.dineout.co.in/images/uploads/restaurant/sharpen/7/p/k/p79307-16051787755fad1597f2bf9.jpg?tr=tr:n-medium',
 'https://im1.dineout.co.in/images/uploads/restaurant/sharpen/5/d/i/p52501-1661855212630de5eceb6d2.jpg?tr=tr:n-medium',
 'https://im1.dineout.co.in/images/uploads/restaurant/sharpen/3/n/o/p34822-15599107305cfa594a13c24.jpg?tr=tr:n-medium',
 'https://im1.dineout.co.in/images/uploads/restaurant/sharpen/1/u/r/p106428-166073786162fcd945925a9.jpg?tr=tr:n-medium']
```

In [26]:

```
ratings = []

for i in soup.find_all('div',class_="restnt-rating rating-4"):

    ratings.append(i.text)
ratings
```

Out[26]:

```
['4', '4.3', '3.9', '3.9', '3.7', '3.8', '4.1']
```

In [29]:

```
import pandas as pd
df = pd.DataFrame({'Restaurants':restaurants,'Location':location,'Cuisine':cuisine,'Images_url':images,'Ratings':rating,})
df
```

Out[29]:

	Restaurants	Location	Cuisine	Images_url	Ratings
0	Castle BarbequeConnaught Place, Central Delhi	Connaught Place, Central Delhi	₹ 2,000 for 2 (approx) | Chinese, North Indian	https://im1.dineout.co.in/images/uploads/resta...	648
1	Cafe KnoshThe Leela Ambience Convention Hotel,...	The Leela Ambience Convention Hotel,Shahdara, ...	₹ 3,000 for 2 (approx) | Italian, Continental	https://im1.dineout.co.in/images/uploads/resta...	648
2	India GrillHilton Garden Inn,Saket, South Delhi	Hilton Garden Inn,Saket, South Delhi	₹ 2,400 for 2 (approx) | North Indian, Italian	https://im1.dineout.co.in/images/uploads/resta...	648
3	The Barbeque CompanyGardens Galleria,Sector 38...	Gardens Galleria,Sector 38A, Noida	₹ 1,700 for 2 (approx) | North Indian, Chinese	https://im1.dineout.co.in/images/uploads/resta...	648
4	Delhi BarbequeTaurus Sarovar Portico,Mahipalpu...	Taurus Sarovar Portico,Mahipalpur, South Delhi	₹ 1,800 for 2 (approx) | North Indian	https://im1.dineout.co.in/images/uploads/resta...	648
5	The Monarch - Bar Be Que VillageIndirapuram Ha...	Indirapuram Habitat Centre,Indirapuram, Ghaziabad	₹ 1,900 for 2 (approx) | North Indian	https://im1.dineout.co.in/images/uploads/resta...	648
6	The Barbeque TimesM2K Corporate Park,Sector 51...	M2K Corporate Park,Sector 51, Gurgaon	₹ 1,500 for 2 (approx) | North Indian, Contine...	https://im1.dineout.co.in/images/uploads/resta...	648

In [45]:

```
import requests
from bs4 import BeautifulSoup

page = requests.get('https://www.cnbc.com/world/?region=world')
page
```

Out[45]:

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

In [31]:

```
headlines = []

for i in soup.find_all('div',class_="RiverPlusCard-container"):
    headlines.append(i.text)

headlines
```

Out[31]:

```
["John Kerry responds to COP28 president's claim there's 'no science' behind fossil fuel phase outan hour ago•Sam Meredith",
 'Here’s what the S&P 500’s summer hot streak can teach us about the late 2023 rallyMichael Santoli',
 'Israel faces growing US calls for restraint amid renewed Gaza fighting',
 'Three reasons a strong Black Friday weekend may not mean a blowout holiday season for retailers4 hours ago•Melissa Repko',
 'A major CEO backer of the remote office model concedes one big work from home failure2 hours ago•Deborah Kong',
 'These tenbagger stocks are well liked by analysts and may rally even more4 hours ago•Lisa Kailai Han',
 "This habit helped Bill Gates change careers after Microsoft—despite his being 'very monomaniacal’3 hours ago•Megan Sauer",
 "'Foreign terrorists' behind deadly Philippine bombing — officials",
 "One dead, two injured after man attacks tourists near Paris' Eiffel Tower",
 "This is how some of 2023's biggest AI winners have performed a year after ChatGPT's launchSamantha Subin",
 "Vice President Harris says 'too many innocent Palestinians have been killed' as fighting intensifies",
 'Signs of a sector rotation — plus 2 more themes to watch in the stock market Zev Fima',
 "U.S. lays out plan at COP 28 to slash climate 'super pollutant' methane from oil and gas",
 'Case for gold fever: NewEdge Wealth sees record rush intensifying Meredith Mutter',
 'Uber shares pop on inclusion in S&P 500Jordan Novet,Ari Levy',
 "China's respiratory illness rise due to known pathogens - official",
 'Inside Apple’s chip lab, home to the most ‘profound change’ at the company in decadesKatie Tarasov',
 'North Korea says interference in its satellites would be declaration of war']
```

In [32]:

```
times = []

for i in soup.find_all('time',class_="LatestNews-timestamp"):
    times.append(i.text)

times
```

Out[32]:

```
['13 Min Ago',
 '1 Hour Ago',
 '1 Hour Ago',
 '2 Hours Ago',
 '2 Hours Ago',
 '3 Hours Ago',
 '3 Hours Ago',
 '3 Hours Ago',
 '3 Hours Ago',
 '3 Hours Ago',
 '4 Hours Ago',
 '4 Hours Ago',
 '4 Hours Ago',
 '9 Hours Ago',
 '21 Hours Ago',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023',
 'December 2, 2023']
```

In [33]:

```
links = []

for i in soup.find_all('a', href="https://www.cnbc.com/2023/12/03/how-reading-helped-bill-gates-change-careers-after-microsoft.html"):

    links.append(i.text)

links
```

Out[33]:

```
['This habit helped Bill Gates change careers after retiring from Microsoft',
 "This habit helped Bill Gates change careers after Microsoft—despite his being 'very monomaniacal’",
 '',
 '',
 'This habit helped Bill Gates change careers after retiring from Microsoft']
```

In [34]:

```
import pandas as pd
df = pd.DataFrame({'Times':times})
df
```

Out[34]:

	Times
0	13 Min Ago
1	1 Hour Ago
2	1 Hour Ago
3	2 Hours Ago
4	2 Hours Ago
5	3 Hours Ago
6	3 Hours Ago
7	3 Hours Ago
8	3 Hours Ago
9	3 Hours Ago
10	4 Hours Ago
11	4 Hours Ago
12	4 Hours Ago
13	9 Hours Ago
14	21 Hours Ago
15	December 2, 2023
16	December 2, 2023
17	December 2, 2023
18	December 2, 2023
19	December 2, 2023
20	December 2, 2023
21	December 2, 2023
22	December 2, 2023
23	December 2, 2023
24	December 2, 2023
25	December 2, 2023
26	December 2, 2023
27	December 2, 2023
28	December 2, 2023
29	December 2, 2023

In [37]:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

In [46]:

```
url = "https://www.journals.elsevier.com/artificial-intelligence/most-downloaded-articles"
response = requests.get(url)
```

In [40]:

```
titles = []

for i in soup.find_all('h2',class_="sc-1qrq3sd-1 gRGSUS sc-1nmom32-0 sc-1nmom32-1 btcbYu goSKRg"):
    titles.append(i.text)

titles
```

Out[40]:

```
['Reward is enough',
 'Explanation in artificial intelligence: Insights from the social sciences',
 'Creativity and artificial intelligence',
 'Conflict-based search for optimal multi-agent pathfinding',
 'Knowledge graphs as tools for explainable machine learning: A survey',
 'Law and logic: A review from an argumentation perspective',
 'Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning',
 'Explaining individual predictions when features are dependent: More accurate approximations to Shapley values',
 'Multiple object tracking: A literature review',
 'A survey of inverse reinforcement learning: Challenges, methods and progress',
 'Evaluating XAI: A comparison of rule-based and example-based explanations',
 'Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights',
 'Hard choices in artificial intelligence',
 'Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making',
 'Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies',
 'The Hanabi challenge: A new frontier for AI research',
 'Wrappers for feature subset selection',
 'Artificial cognition for social human–robot interaction: An implementation',
 'A review of possible effects of cognitive biases on interpretation of rule-based machine learning models',
 'The multifaceted impact of Ada Lovelace in the digital age',
 'Robot ethics: Mapping the issues for a mechanized world',
 'Reward (Mis)design for autonomous driving',
 'Planning and acting in partially observable stochastic domains',
 'What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research']
```

In [41]:

```
authors = []

for i in soup.find_all('span',class_="sc-1w3fpd7-0 dnCnAO"):
    authors.append(i.text)

authors
```

Out[41]:

```
['David Silver, Satinder Singh, Doina Precup, Richard S. Sutton ',
 'Tim Miller ',
 'Margaret A. Boden ',
 'Guni Sharon, Roni Stern, Ariel Felner, Nathan R. Sturtevant ',
 'Ilaria Tiddi, Stefan Schlobach ',
 'Henry Prakken, Giovanni Sartor ',
 'Richard S. Sutton, Doina Precup, Satinder Singh ',
 'Kjersti Aas, Martin Jullum, Anders Løland ',
 'Wenhan Luo, Junliang Xing and 4 more',
 'Saurabh Arora, Prashant Doshi ',
 'Jasper van der Waa, Elisabeth Nieuwburg, Anita Cremers, Mark Neerincx ',
 'Joe Collenette, Katie Atkinson, Trevor Bench-Capon ',
 'Roel Dobbe, Thomas Krendl Gilbert, Yonatan Mintz ',
 'Oskar Wysocki, Jessica Katharine Davies and 5 more',
 'Eoin M. Kenny, Courtney Ford, Molly Quinn, Mark T. Keane ',
 'Nolan Bard, Jakob N. Foerster and 13 more',
 'Ron Kohavi, George H. John ',
 'Séverin Lemaignan, Mathieu Warnier and 3 more',
 'Tomáš Kliegr, Štěpán Bahník, Johannes Fürnkranz ',
 'Luigia Carlucci Aiello ',
 'Patrick Lin, Keith Abney, George Bekey ',
 'W. Bradley Knox, Alessandro Allievi and 3 more',
 'Leslie Pack Kaelbling, Michael L. Littman, Anthony R. Cassandra ',
 'Markus Langer, Daniel Oster and 6 more']
```

In [42]:

```
published_date = []

for i in soup.find_all('span',class_="sc-1thf9ly-2 dvggWt"):
    published_date.append(i.text)

published_date
```

Out[42]:

```
['October 2021',
 'February 2019',
 'August 1998',
 'February 2015',
 'January 2022',
 'October 2015',
 'August 1999',
 'September 2021',
 'April 2021',
 'August 2021',
 'February 2021',
 'April 2023',
 'November 2021',
 'March 2023',
 'May 2021',
 'March 2020',
 'December 1997',
 'June 2017',
 'June 2021',
 'June 2016',
 'April 2011',
 'March 2023',
 'May 1998',
 'July 2021']
```

In [44]:

```
import pandas as pd
df = pd.DataFrame({'Titles':titles,'Authors':authors,'Publishes_date':published_date})
df
```

Out[44]:

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

In [ ]: