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

import requests

from bs4 import BeautifulSoup

import pandas as pd

page = requests.get("https://en.wikipedia.org/wiki/Main\_Page")

soup = BeautifulSoup(page.content, "html.parser")

header\_tags = soup.find\_all(["h1", "h2", "h3", "h4", "h5", "h6"])

header\_text = [tag.get\_text().strip() for tag in header\_tags]

df = pd.DataFrame({"Headers": header\_text})

print(df)

#2) Write a python program to display IMDB’s Top rated 50 movies’ data (i.e. name, rating, year of release) and make data frame.

import requests

from bs4 import BeautifulSoup

import pandas as pd

page = requests.get("https://www.imdb.com/chart/top/")

soup = BeautifulSoup(page.content, "html.parser")

movie\_titles = [title.get\_text().strip() for title in soup.select(".titleColumn a")]

movie\_ratings = [float(rating.get\_text().strip()) for rating in soup.select(".imdbRating strong")]

movie\_years = [int(year.get\_text().strip("()")) for year in soup.select(".titleColumn span.secondaryInfo")]

df = pd.DataFrame({"Title": movie\_titles, "Rating": movie\_ratings, "Year": movie\_years})

print(df.head(50))

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

import requests

from bs4 import BeautifulSoup

import pandas as pd

url = "https://presidentofindia.nic.in/former-presidents.htm"

response = requests.get(url)

soup = BeautifulSoup(response.content, "html.parser")

table = soup.find("table", {"class": "tablepress"})

data = []

for row in table.find\_all("tr"):

cells = row.find\_all("td")

if len(cells) == 2: # Check if the row has 2 cells

name = cells[0].get\_text().strip()

term = cells[1].get\_text().strip()

data.append({"Name": name, "Term of Office": term})

df = pd.DataFrame(data)

print(df)

#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 andrating.

import requests

from bs4 import BeautifulSoup

import pandas as pd

# Scrape top 10 ODI teams

url = "https://www.icc-cricket.com/rankings/mens/team-rankings/odi"

response = requests.get(url)

soup = BeautifulSoup(response.content, "html.parser")

team\_table = soup.find("table", {"class": "table"})

team\_rows = team\_table.tbody.find\_all("tr")

team\_data = []

for row in team\_rows[:10]:

team\_name = row.find("span", {"class": "u-hide-phablet"}).text.strip()

matches = row.find\_all("td")[2].text.strip()

points = row.find\_all("td")[3].text.strip()

rating = row.find\_all("td")[4].text.strip()

team\_data.append([team\_name, matches, points, rating])

team\_df = pd.DataFrame(team\_data, columns=["Team Name", "Matches", "Points", "Rating"])

print("Top 10 ODI Teams")

print(team\_df)

# Scrape top 10 ODI batsmen

url = "https://www.icc-cricket.com/rankings/mens/player-rankings/odi/batting"

response = requests.get(url)

soup = BeautifulSoup(response.content, "html.parser")

batsmen\_table = soup.find("table", {"class": "table"})

batsmen\_rows = batsmen\_table.tbody.find\_all("tr")

batsmen\_data = []

for row in batsmen\_rows[:10]:

batsman\_name = row.find("td", {"class": "table-body\_\_cell name"}).a.text.strip()

batsman\_team = row.find("span", {"class": "table-body\_\_logo-text"}).text.strip()

batsman\_rating = row.find("td", {"class": "table-body\_\_cell u-text-right rating"}).text.strip()

batsmen\_data.append([batsman\_name, batsman\_team, batsman\_rating])

batsmen\_df = pd.DataFrame(batsmen\_data, columns=["Batsman Name", "Team", "Rating"])

print("\nTop 10 ODI Batsmen")

print(batsmen\_df)

# Scrape top 10 ODI bowlers

url = "https://www.icc-cricket.com/rankings/mens/player-rankings/odi/bowling"

response = requests.get(url)

soup = BeautifulSoup(response.content, "html.parser")

bowlers\_table = soup.find("table", {"class": "table"})

bowlers\_rows = bowlers\_table.tbody.find\_all("tr")

bowlers\_data = []

for row in bowlers\_rows[:10]:

bowler\_name = row.find("td", {"class": "table-body\_\_cell name"}).a.text.strip()

bowler\_team = row.find("span", {"class": "table-body\_\_logo-text"}).text.strip()

bowler\_rating = row.find("td", {"class": "table-body\_\_cell u-text-right rating"}).text.strip()

bowlers\_data.append([bowler\_name, bowler\_team, bowler\_rating])

bowlers\_df = pd.DataFrame(bowlers\_data, columns=["Bowler Name", "Team", "Rating"])

print("\nTop 10 ODI Bowlers")

print(bowlers\_df)

#7) 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

import requests

from bs4 import BeautifulSoup

import pandas as pd

url = 'https://www.cnbc.com/world/?region=world'

response = requests.get(url)

soup = BeautifulSoup(response.text, 'html.parser')

articles = soup.find\_all('div', class\_='Card-titleContainer')

headlines = []

times = []

links = []

for article in articles:

headlines.append(article.a.text.strip())

times.append(article.time.text.strip())

links.append(article.a['href'])

data = {'Headline': headlines, 'Time': times, 'News Link': links}

df = pd.DataFrame(data)

print(df)

#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

import requests

from bs4 import BeautifulSoup

import pandas as pd

url = "https://www.journals.elsevier.com/artificial-intelligence/most-downloaded-articles"

response = requests.get(url)

soup = BeautifulSoup(response.content, 'html.parser')

table = soup.find('table', {'class': 'article-table'})

titles = []

authors = []

dates = []

urls = []

for row in table.find\_all('tr')[1:]:

cols = row.find\_all('td')

titles.append(cols[0].text.strip())

authors.append(cols[1].text.strip())

dates.append(cols[2].text.strip())

urls.append(cols[0].find('a')['href'])

df = pd.DataFrame({'Paper Title': titles, 'Authors': authors, 'Published Date': dates, 'Paper URL': urls})

print(df)

#9) Write a python program to scrape mentioned details from dineout.co.in and make data frame- i) Restaurant name

#ii) Cuisine

#iii) location iv) Ratings

# v) Image URL

import requests

from bs4 import BeautifulSoup

import pandas as pd

url = 'https://www.dineout.co.in/delhi-restaurants'

response = requests.get(url)

soup = BeautifulSoup(response.content, 'html.parser')

restaurants = soup.find\_all('div', {'class': 'restnt-info'})

restaurant\_names = []

cuisines = []

locations = []

ratings = []

image\_urls = []

for restaurant in restaurants:

name = restaurant.find('div', {'class': 'restnt-name ellipsis'}).text.strip()

restaurant\_names.append(name)

cuisine = restaurant.find('div', {'class': 'restnt-cuisine ellipsis'}).text.strip()

cuisines.append(cuisine)

location = restaurant.find('div', {'class': 'restnt-loc ellipsis'}).text.strip()

locations.append(location)

rating = restaurant.find('div', {'class': 'restnt-rating'}).text.strip()

ratings.append(rating)

image\_url = restaurant.find('img')['data-original']

image\_urls.append(image\_url)

restaurant\_dict = {

'Restaurant Name': restaurant\_names,

'Cuisine': cuisines,

'Location': locations,

'Ratings': ratings,

'Image URL': image\_urls

}

df = pd.DataFrame(restaurant\_dict)

print(df.head())