In all the following questions, you have to use BeautifulSoup to scrape different websites and collect data as per the requirement of the question.

Every answer to the question should be in form of a python function which should take URL as the parameter. Use Jupyter Notebooks to program, upload it on your GitHub and send the link of the Jupyter notebook to your SME.

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

Answer: from urllib.request import urlopen

from bs4 import BeautifulSoup

html = urlopen('https://en.wikipedia.org/wiki/Main_Page')

bs = BeautifulSoup(html, "html.parser")

titles = bs.find_all(['h1', 'h2','h3','h4','h5','h6'])

print('List all the header tags :', *titles, sep='\n\n')

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.

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 andrating.

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

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

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

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

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

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

   i. Headline

Answer: import requests

from bs4 import BeautifulSoup

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

```
response = requests.get(url)


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

headlines = soup.find('body').find_all('h3')

for x in headlines:

print(x.text.strip())
```

      ii.   Time

      iii.   News Link

6) 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

7) Write a python program to scrape mentioned details from dineout.co.inand make data frame-

      i.   Restaurant name

      ii.   Cuisine

      iii.   Location

      iv.   Ratings

      v.   Image URL