from bs4 import BeautifulSoup

import requests

import re

import json

headers1 = {"User-agent":"Mozilla/5.0 (X11; Linux x86\_64) AppleWebKit/537.36 (KHTML, like Gecko)"}

src\_url = '<https://health.usnews.com/doctors/city-index/new-jersey>'

src0 = requests.get(src\_url,headers=headers1).text

cities = BeautifulSoup(src0,'lxml')

city\_div = cities.find\_all('div',class\_='Content-o0f126-0 content kteJom')

city\_div = city\_div[2]

for a in range(0,1):

city\_list = city\_div.find\_all('li',class\_='List\_\_ListItem-dl3d8e-1 kfaWAY')

city\_list = city\_list[a]

city\_list\_link = '[https://health.usnews.com](https://health.usnews.com/)' + city\_list.a['href']

#print city\_list\_link

src3 = requests.get(city\_list\_link,headers=headers1).text

spl = BeautifulSoup(src3,'lxml')

sp\_div = spl.find\_all('div',class\_='Content-o0f126-0 content kteJom')

sp\_div = sp\_div[2]

for b in range(0,1):

sp\_list=sp\_div.find\_all('li',class\_='List\_\_ListItem-dl3d8e-1 kfaWAY')

sp\_list = sp\_list[b]

sp\_list\_link = '[https://health.usnews.com](https://health.usnews.com/)' + sp\_list.a['href']

#print sp\_list\_link

url = sp\_list\_link

src = requests.get(url,headers=headers1).text

soup = BeautifulSoup(src,'lxml')

for c in range(0,1):

doc\_list = soup.find\_all('li', class\_ = 'block-normal block-loose-for-large-up')

doc\_list = doc\_list[c]

doc\_link = doc\_list.h3.a['href']

doc\_link1 = '[https://health.usnews.com](https://health.usnews.com/)' + doc\_link

src1 = requests.get(doc\_link1,headers=headers1).text

soup1 = BeautifulSoup(src1,'lxml')

doc\_name\_l = []

doc\_ref = soup1.find('div', class\_ ='flex-row relative' ).h1.text

doc\_name = re.findall('[a-zA-Z0-9]\S\*',str(doc\_ref))

doc\_name = ' '.join(doc\_name)

doc\_name\_l.append(str(doc\_name))

print doc\_name\_l

doc\_ov\_l = []

doc\_ov1 =' '

doc\_ov = soup1.find('div', class\_ ='block-normal clearfix' ).p.text

doc\_ov = re.findall('[a-zA-Z0-9]\S\*',str(doc\_ov))

doc\_ov = doc\_ov1.join(doc\_ov)

doc\_ov\_l.append(doc\_ov)

print doc\_ov\_l

doc\_ex\_l = []

doc\_ex = soup1.find\_all('span', class\_ ='text-large heading-normal-for-small-only right-for-medium-up')

doc\_ex = doc\_ex[1]

doc\_ex = doc\_ex.text.strip()

doc\_ex\_l.append(str(doc\_ex))

print doc\_ex\_l

doc\_lan\_l =[]

doc\_lan = soup1.find('span', class\_ ='text-large heading-normal-for-small-only right-for-medium-up text-right showmore').text.strip()

doc\_lan\_l.append(str(doc\_lan))

print doc\_lan\_l

doc\_loc\_l =[]

doc\_loc1 =' '

doc\_loc = soup1.find('div', class\_ ='flex-small-12 flex-medium-6 flex-large-5').p.span.text.strip()

doc\_loc = re.findall('[a-zA-Z0-9]\S\*',str(doc\_loc))

doc\_loc = doc\_loc1.join(doc\_loc)

doc\_loc\_l.append(doc\_loc)

print doc\_loc\_l

doc\_aff\_l =[]

doc\_aff = soup1.find('section', class\_ ='doctors-hospitals block-loosest').p.text.strip()

doc\_aff\_l.append(str(doc\_aff))

print doc\_aff\_l

doc\_sp\_l =[]

doc\_sp = soup1.find('div', class\_ ='flex-row small-between').a.text.strip()

doc\_sp\_l.append(str(doc\_sp))

doc\_subsp = soup1.find('p', class\_ ='text-large block-tight').text

doc\_sp\_l.append(str(doc\_sp))

print doc\_sp\_l

print ' '

doc\_ed = soup1.find\_all(class\_ ="block-loosest")[5]

doc\_ed1 = doc\_ed.find\_all('li')

doc\_ed5 = ' '

doc\_ed6 = []

for i in doc\_ed1:

doc\_ed2 = str(i.text)

doc\_ed2 = re.findall('[a-zA-Z0-9]\S\*',doc\_ed2)

doc\_ed2 = doc\_ed5.join(doc\_ed2)

doc\_ed6.append(doc\_ed2)

print doc\_ed6

print ' '

doc\_cert2 = []

doc\_cert1 = ' '

doc\_ed3 = soup1.find\_all(class\_ ="block-loosest")[6]

doc\_ed14 = doc\_ed3.find\_all('li')

for i in doc\_ed14:

doc\_cert = str(i.text)

doc\_cert = re.findall('[a-zA-Z0-9]\S\*',doc\_cert)

doc\_cert = doc\_cert1.join(doc\_cert)

doc\_cert2.append(doc\_cert)

print doc\_cert2

c += 1

b += 1

a += 1