12/9/2019 parse\_sdsu

```
In [ ]: from bs4 import BeautifulSoup
from requests import get
import re
from selenium import webdriver
import pandas as pd
import csv
with open('SDSU files.csv', mode = 'w', newline = '') as csvfile:
    writer = csv.DictWriter(csvfile, fieldnames = ['num people enrolled', 'tot
al_class_size', 'class_number', 'term', 'professor'])
    writer.writeheader()
    for i in range(9, 20):
        for term in range(2,5,2):
             season = ' Fall'
             if term == 4:
                 season = " Spring"
             combined_temr = str(i) + season
             url = 'https://sunspot.sdsu.edu/schedule/search?mode=search&period
=20'+str(i) + str(term)+'&admin unit=R&abbrev=CS&number=&suffix=&courseTitle=&
scheduleNumber=&units=&instructor=&facility=&space=&meetingType=&startHours=&s
tartMins=&endHours=&endMins=&remaining seats at least='
             if i == 9:
                 url = 'https://sunspot.sdsu.edu/schedule/search?mode=search&pe
riod=2009' + str(term) + '&admin unit=R&abbrev=CS&number=&suffix=&courseTitle=
&scheduleNumber=&units=&instructor=&facility=&space=&meetingType=&startHours=&
startMins=&endHours=&endMins=&remaining seats at least='
             response = BeautifulSoup(get(url).text, 'lxml')
             classes = response.find all('div', {'class':'sectionMeeting'})
             for class block in classes:
                 seats = class_block.find('div',{'class':'sectionFieldSeats'})
                 if seats is not None:
                     professor = class block.find('div', {'class':'sectionField
Instructor column'}).a
                     professor name = None;
                     if professor is not None:
                         professor name = professor.text
                     seats = seats.text.strip()
                     class name = class block.find('div', {'class':'sectionFiel
dCourse column'}).text
                     type = class_block.find('div', {'class':'sectionFieldType
 column'}).text
                     if type == 'Lecture' or type == 'Seminar':
                         enrolled_cap = seats.split('/')
                         class info = {
                             'num_people_enrolled' : enrolled_cap[0],
                             'total_class_size' : enrolled_cap[1],
                             'class number' : class name.strip(),
                             'term' : combined temr,
                             'professor': professor_name
                         writer.writerow(class info)
```