

In [ ]: *#scrapes Harvey Mudd university from html folder for FALL/Spring semester*

```

In [ ]: from bs4 import BeautifulSoup
import codecs
import re
import pandas as pd
import csv
import os
with open('Harvey_files.csv', mode = 'w', newline = '') as csvfile:
    writer = csv.DictWriter(csvfile, fieldnames=['num_people_enrolled', 'total_class_size', 'class_number', 'term', 'professor'])
    writer.writeheader()

    seasons = ['FALL', 'SPRING']
    #scrapes 2004-2019 for fall and 2011-2019 for spring
    for season in seasons:
        for j in range(4,20):
            if (season is 'SPRING' and j > 10) or season is 'FALL':
                if (j == 6 and season is 'FALL') or (j == 14) or ((j == 13 or j == 17) and season is 'SPRING'):
                    continue
                loc = 'HARVEY HTML/' + season + ' ' + str(j) + '.htm'
                url = codecs.open(loc, 'r', 'utf-8')

                document = BeautifulSoup(url, 'html.parser')
                body = document.find('div', {'class': 'pContent CS'})
                panel_body = body.find('tbody', {'class': 'gbody'})
                each_class = panel_body.find_all('tr')

                t = 0

                for i in each_class:

                    if (t % 4) == 0:
                        each_body = i.find_all('td', {'valign': 'top'})
                        class_name = each_body[1].text
                        class_size_raw = each_body[3].text
                        num_people_enrolled = class_size_raw.split("/",1)[0]
                        total_class_size = class_size_raw.split("/",1)[1]
                        prof_raw = each_body[5]
                        prof_raw1 = prof_raw.find_all('li')

                        professor = []
                        for i in prof_raw1:
                            prof = i.text
                            sub_professor = prof.split("/",1)[0]
                            professor.append(sub_professor)

                        class_info = {
                            'num_people_enrolled' : num_people_enrolled,
                            'total_class_size' : total_class_size ,
                            'class_number' : class_name,
                            'term' : season + ' 20' + str(j),
                            'professor' : professor
                        }

```

```
writer.writerow(class_info)
```

```
t = t + 1
```