

```

# -*- coding: utf-8 -*-
"""
Created on Sat Apr 29 19:12:13 2017

@author: Harsh Kevadia
"""
import requests
import re
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt

def run(url):
    jobCounter = 0 #just a counter for questions scraped
    rcounter = 0
    mlCounter = 0
    bdCounter = 0
    pythonCounter = 0
    sasCounter = 0

    pageNum=19 # number of pages to collect

    for n in range(pageNum):
        if n == 0: continue
        pageLink=url + str(n) # make the page url\
        html = None
        print(pageLink)
        for i in range(5): # try 5 times
            try:
                #use the browser to access the url
                response = requests.get(pageLink,headers = { 'User-Agent': 'Mozilla/5.0 (Windows NT
                html = response.content # get the html
                #print(html)
                break # we got the file, break the loop
            except Exception as e:# threw an exception, the attempt to get the response failed
                print ('failed attempt Page URL',i)
                continue
                #time.sleep(2) # wait 2 secs

        if not html:
            continue
        soup = BeautifulSoup(html,'html.parser')
        jobs = soup.find_all('div', {'class':re.compile('jobTitle')})

        for job in jobs:
            for a in job.find_all('a', href=True):
                jobCounter = jobCounter + 1
                data = None
                for j in range(5):
                    try:
                        resp = requests.get(a['href'],headers = { 'User-Agent': 'Mozilla/5.0 (Windc
                        data = resp.content
                        break
                    except Exception as ex:
                        print('failed attempt JOB URL',i)
                        continue
                if not data:
                    continue

```

```

data = str(data).strip('').lower()
if 'python' in data:
    pythonCounter = pythonCounter + 1
if 'big data' in data:
    bdCounter = bdCounter + 1
if 'sas' in data:
    sasFlag = False
    if ' sas,' in data:
        sasFlag = True
    if ',sas ' in data:
        sasFlag = True
    if ' sas ' in data:
        sasFlag = True
    if sasFlag is True:
        sasCounter = sasCounter + 1
if 'r' in data:
    rFlag = False
    if ' r,' in data:
        rFlag = True
    if ',r ' in data:
        rFlag = True
    if ' r ' in data:
        rFlag = True
    if rFlag is True:
        rcounter = rcounter + 1
if 'machine learning' in data:
    mlCounter = mlCounter + 1

return jobCounter,pythonCounter,bdCounter,sasCounter,rcounter,mlCounter
if __name__=='__main__':

    resultFile=open('analysis.txt','w') # input file

    url='https://www.monster.com/jobs/search/?q=data-science&page='

    jobCounter,pythonCounter,bdCounter,sasCounter,rcounter,mlCounter = run(url)

    print('Total Jobs:' + str(jobCounter))
    print('Total Python Skill Jobs:' + str(pythonCounter) + 'Python Percentage: ' + str(pythonCounter/jobCounter*100) + '%')
    print('Total Big Data Skill Jobs:' + str(bdCounter) + 'Big Data Percentage: ' + str(bdCounter/jobCounter*100) + '%')
    print('Total SAS Skill Jobs:' + str(sasCounter) + 'SAS Percentage: ' + str(sasCounter/jobCounter*100) + '%')
    print('Total R Skill Jobs:' + str(rcounter) + 'R Percentage: ' + str(rcounter/jobCounter*100) + '%')
    print('Total Machine Learning Skill Jobs:' + str(mlCounter) + 'Machine Learning Percentage: ' + str(mlCounter/jobCounter*100) + '%')

    resultFile.write('Total Jobs:' + str(jobCounter) + "\n")
    resultFile.write('Total Python Skill Jobs:' + str(pythonCounter) + 'Python Percentage: ' + str(pythonCounter/jobCounter*100) + '%\n')
    resultFile.write('Total Big Data Skill Jobs:' + str(bdCounter) + 'Big Data Percentage: ' + str(bdCounter/jobCounter*100) + '%\n')
    resultFile.write('Total SAS Skill Jobs:' + str(sasCounter) + 'SAS Percentage: ' + str(sasCounter/jobCounter*100) + '%\n')
    resultFile.write('Total R Skill Jobs:' + str(rcounter) + 'R Percentage: ' + str(rcounter/jobCounter*100) + '%\n')
    resultFile.write('Total Machine Learning Skill Jobs:' + str(mlCounter) + 'Machine Learning Percentage: ' + str(mlCounter/jobCounter*100) + '%\n')

    labels = 'Python', 'R', 'Big Data', 'SAS', 'Machine Learning'
    data = [pythonCounter, rcounter, bdCounter, sasCounter, mlCounter]
    colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue', 'lightpink']
    explode = (0, 0, 0, 0, 0) # explode 1st slice

# Plot

```

```
plt.pie(data, explode=explode, labels=labels, colors=colors, autopct='%1.1f%%', shadow=True, st
plt.axis('equal')
plt.show()

resultFile.close()
```