import re

import requests

from bs4 import BeautifulSoup

import random

import pandas as pd

def scrape42(url):

\_url = requests.get(url=url,)

soup = BeautifulSoup(\_url.content, 'html.parser')

allLinks = soup.find(id="post-").find\_all("li")

quotes=[]

authors=[]

for link in allLinks:

temp = (str)(link)

temp = temp.replace("<strong>", "")

temp = temp.replace("</strong>", "")

temp = temp.replace("</li>", "")

temp = temp.replace("<li>", "")

temp = temp.replace("<em>", "")

temp = temp.replace("</em>", "")

temp = temp.replace("“", "")

temp = temp.replace("”", "")

if(temp.find("<a") == -1):

quotes.append(temp)

allLinks = soup.find(id="post-").find\_all("a")

for link in allLinks:

temp = (str)(link)

link=BeautifulSoup(temp, "html.parser")

for x in link.find\_all("em"):

x = (str)(x)

x = x.replace("<em>", "")

x = x.replace("</em>", "")

authors.append(x)

df = pd.DataFrame(quotes[:len(authors)], authors)

df.to\_csv("a.csv", header=False, mode='a',)

scrape42('https://www.kdnuggets.com/2017/05/42-essential-quotes-data-science-thought-leaders.html/')

scrape42('https://www.kdnuggets.com/2017/05/42-essential-quotes-data-science-thought-leaders.html/2')

import re

import requests

from bs4 import BeautifulSoup

import random

import pandas as pd

def scrape(url):

\_url = requests.get(url=url,)

soup = BeautifulSoup(\_url.content, 'html.parser')

quotes=[]

authors=[]

allLinks = soup.find(id="hs\_cos\_wrapper\_post\_body").find\_all("h3")

for link in allLinks:

pattern = r'[0-9]'

authors.append(re.sub(pattern, '', (str)(link.text)).replace(". ",""))

allLinks = soup.find(id="hs\_cos\_wrapper\_post\_body").find\_all("p")

for link in allLinks:

temp = (str)(link.text)

if(temp.find("“") != -1):

temp = temp.replace("“", "")

temp = temp.replace("”", "")

quotes.append(temp)

lent = min(len(authors), len(quotes))

df = pd.DataFrame(quotes[:lent], authors[:lent])

df.to\_csv("a.csv", header=False, mode='a',)

scrape('https://www.scuba.io/blog/48-analytics-quotes-experts')