Assignment 1 c

May 8, 2023

[6]:
```
# import module
import requests
from bs4 import BeautifulSoup

HEADERS = ({'User-Agent':
            'Mozilla/5.0 (Windows NT 10.0; Win64; x64) \
            AppleWebKit/537.36 (KHTML, like Gecko) \
            Chrome/90.0.4430.212 Safari/537.36',
            'Accept-Language': 'en-US, en;q=0.5'})

# user define function
# Scrape the data
def getdata(url):
    r = requests.get(url, headers=HEADERS)
    return r.text

def html_code(url):

    # pass the url
    # into getdata function
    htmldata = getdata(url)
    soup = BeautifulSoup(htmldata, 'html.parser')

    # display html code
    return (soup)

url = "https://www.amazon.in/Sparx-Casual-Stripped-Sneakers-Wh \
       ite/dp/B0B4KF6LFM/ref=sr_1_1?pf_rd_i=27037351031&pf_rd_m=A \
       1VBAL9TL5WCBF&pf_rd_p=2734da86-a990-4242-ae92-ad2e3fa26329 \
       &pf_rd_r=8RE6KRAY8F5H7W6S17KQ&pf_rd_s=merchandised-search- \
      ␣
 ↪9&qid=1683565076&refinements=p_36%3A4516638031&rnid=4516629031&s=shoes&sr=1-1&th=1&psc=1"

soup = html_code(url)
def cus_data(soup):
    # find the Html tag
    # with find()
```

```
    # and convert into string
    data_str = ""
    cus_list = []

    for item in soup.find_all("span", class_="a-profile-name"):
        data_str = data_str + item.get_text()
        cus_list.append(data_str)
        data_str = ""
    return cus_list


cus_res = cus_data(soup)
print(cus_res)
```

```
['Amazon Customer', 'Ritik raj', 'KuLVuNDeR RaNa', 'Reena Patel', 'Md Mohsin
Aziz', 'Md Mohsin Aziz', 'Rajan Singh', 'Faisal siddiqui', 'Value of money and
good quality products.', 'Value of money and good quality products.', 'Jagdish']
```

[ ]: