webscrapping

April 28, 2025

[3]:
```
#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'})

# Scrape the data
def getdata(site_url):
        res = requests.get(site_url, headers=HEADERS)
        return res.text


def gethtml(site_url):

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

        # display html code
        return (soup)


site_url = "https://www.amazon.in/
 ↪StonKraft-Collectible-Marble-Malachite-Delivery/dp/B00JY9ES7Y/ref=sr_1_10?
 ↪crid=1NNTQEXVFJL6U&dib=eyJ2IjoiMSJ9.
 ↪rY9aE6NKWZCF3Flulk4dgrKmGhxfwyhXQLiv2tDyYnMdYd48pRRpA5cvr4g0ynJdhla7z9jX90EsPe0bfU3xa74n7ct
 ↪33d04m321zfkRuMsR6jDvWpX3h5Dki2Pxm15w5Iuqy4&dib_tag=se&keywords=square+off+chess+board&qid=

soup = gethtml(site_url)
#print(soup)

def getCustomerName(soup):
```

```
    # find the Html tag
    # with find()
    # and convert into string
    data_string = ""
    custumer_list = []

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

def getRating(soup):
        data_string = ""
        rating_list = []

        for item in soup.find_all("span", class_="a-icon-alt"):
                        data_string = data_string + item.get_text()
                        rating_list.append(data_string)
                        data_string = ""
        return rating_list

def getReviewDate(soup):
        data_string = ""
        date_list = []

        for item in soup.find_all("span", class_="review-date"):
                        data_string = data_string + item.get_text()
                        date_list.append(data_string)
                        data_string = ""
        return date_list


custumer_res = getCustomerName(soup)
print(custumer_res)

rating = getRating(soup)
print(rating)

date = getReviewDate(soup)
print(date)
```

```
['Amazon Customer', 'Amazon Customer', 'Sandeep', 'Sandeep']
['5.0 out of 5 stars', 'Previous page', 'Next page', 'Previous page', 'Next
page', '5 out of 5 stars.', '5 out of 5 stars.', '5 out of 5 stars.', '5.0 out
of 5 stars', '5.0 out of 5 stars', '4.7 out of 5 stars', '4.4 out of 5 stars',
'4.6 out of 5 stars', '3.6 out of 5 stars', '4.0 out of 5 stars', 'Previous
```

page', '4.3 out of 5 stars', '4.4 out of 5 stars', '3.0 out of 5 stars', '4.0 out of 5 stars', '4.4 out of 5 stars', 'Next page', '5 out of 5 stars', '5.0 out of 5 stars', 'Previous page', 'Next page', '5 Stars', '5.0 out of 5 stars', '5.0 out of 5 stars']
['Reviewed in India on 17 February 2021']