web scraping on yallakora website

```
import requests
from bs4 import BeautifulSoup
import pandas as pd


def get_valid_date():
    while True:
        try:
            date_input = input("Enter a date in the format DD-MM-YYYY: ")
            day, month, year = map(int, date_input.split('-'))

            if month < 1 or month > 12:
                print("Invalid month. Please enter a month between 1 and 12.")
                continue

            if month == 2:  # Special case for February
                if (year % 4 == 0 and year % 100 != 0) or (year % 400 == 0):
                    max_days = 29  # Leap year
                else:
                    max_days = 28
            elif month in [4, 6, 9, 11]:
                max_days = 30  # April, June, September, November
            else:
                max_days = 31  # January, March, May, July, August, October, December

            if day < 1 or day > max_days:
                print(f"Invalid day. Please enter a day between 1 and {max_days} for month {month}.")
                continue

            print(f"Date entered: {day:02d}-{month:02d}-{year}")
            break

        except ValueError:
            print("Invalid input format. Please enter the date in the format DD-MM-YYYY.")

date_input=get_valid_date()
url = f'https://www.yallakora.com/match-center/?date={date_input}'
                .
```

```
def main(url):

    page = requests.get(url)
    src=page.content
    soup= BeautifulSoup(src,"lxml")
    #print(soup)
    match=soup.find_all("div",{"class": "matchCard"})
    print(len(match))

    def get_info_match(match):
        match_details=[]
        if match:
            for match in match:
                info_match=match.contents[1].find("h2").text.strip()
                print(info_match)
                all_matches_data=match.contents[3].find_all(["div","span"],{"class": ["teamA","score","score","time","teamB"] })
                #print(all_matches_data)
                rows2=[all_matches_data[i:i+5] for i in range(0,len(all_matches_data),5)]
                #print(rows2)
                print(len(rows2))
                #print(rows2)
                matchl=[e.text.strip() for i in rows2 for e in i]
                #print(matchl)
                flattened=[[info_match]+ matchl[i:i+5] for i in range(0,len(matchl),5)]
                #print(flattened)
                match_details.append(flattened)
                #print(match_details)
                #print(len(match_details))
                match_detailsl=[n for i in match_details for n in i]
                df3=pd.DataFrame(match_detailsl,columns =["championship","teamA","score","score","time","teamB"])
                df3['results']=df3.apply(lambda x: str(x['score']),axis=1)

                #df3['results'] =df3['score'].astype(str)+ " / " + df3["score"].astype(str)
                df3=df3.drop(columns=['score','score'])
                df3=df3[["championship","teamA","results","time","teamB"]]
                df3.to_csv("df3.csv")
```

```
    if match:
        #get_info_match(match[i] for i in range(len(match)))

         get_info_match(match)
    else:
         print("No found match")


main(url)
```

	championship	teamA	results	time	teamB
0	تصفيات أمم أفريقيا	بوكسوادا	score 1 score 1 Name: 0, dtype: object	15:00	موريتانيا
1	تصفيات أمم أفريقيا	أوغندا	score 0 score 2 Name: 1, dtype: object	15:00	جنوب أفريقيا
2	تصفيات أمم أفريقيا	زيمبابوي	score 1 score 1 Name: 2, dtype: object	18:00	كينيا
3	تصفيات أمم أفريقيا	إيسواتيني	score 0 score 0 Name: 3, dtype: object	18:00	غينيا بيساو
4	تصفيات أمم أفريقيا	موزمبيق	score 0 score 1 Name: 4, dtype: object	18:00	مالي
5	تصفيات أمم أفريقيا	كاب فيردي	score 1 score 1 Name: 5, dtype: object	18:00	مصر
6	تصفيات أمم أفريقيا	زامبيا	score 1 score 0 Name: 6, dtype: object	18:00	كوت ديفوار
7	تصفيات أمم أفريقيا	أنجـــولا	score - score - Name: 7, dtype: object	21:00	هاتـا
8	تصفيات أمم أفريقيا	الجابون	score - score - Name: 8, dtype: object	21:00	المغرب
9	تصفيات أمم أفريقيا	جامبيا	score - score - Name: 9, dtype: object	21:00	جزر القمر

Show 10 per page

1 2 3

Web scraping on books.toscrape website

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
global number_page
number_page=int(input("enter number of page: "))
url =f'https://books.toscrape.com/catalogue/page-{number_page}.html'
def main():
    def scrap_page(url):
        page= requests.get(url)
        src=page.content
        #print(src)
        soup=BeautifulSoup(src,'lxml')
        #print(soup)
        #allproduct=soup.find('div',{'class':'action'}).text.strip()
        #print(allproduct)
        li=soup.find_all(['li'],{'class':['col-xs-6 col-sm-4 col-md-3 col-lg-3']})
        print(len(li))
        book_details=[]
        for i in range(len(li)):
            image1=li[i].find('img')['src']
            title=li[i].find('img')['alt']
            price=li[i].find('p',{'class':'price_color'}).text.strip()
            stock=li[i].find('p',{'class':'instock availability'}).text.strip()
            rating=li[i].find('p',{'class':'star-rating'})['class'][1]
            book_details.append({"image_book":image1,"title_book":title,"price_book":price,"stock_book":stock,"ratings":rating})
        return book_details


    def scrap_website(number_page):
         all_data=[]
         for i in range(1,number_page+1):
              url =f'https://books.toscrape.com/catalogue/page-{number_page}.html'
              page_data=scrap_page(url)

              all_data.extend(page_data)


         return all_data
```

```
    def scrap_website_category(url):
        page= requests.get(url)
        src=page.content
        #print(src)
        soup=BeautifulSoup(src,'lxml')
        search1=soup.find_all('ul',{"class": "nav nav-list"})
        l=[]
        for i in search1:

            l.append(i.text.strip().lower())
        #print(l)
        clean_list=[s.replace("\n","").strip() for s in l]
        #print(clean_list)
        long_string = 'books                                                                travel

        items = [f"{item.strip()}_{index}" for index, item in enumerate(long_string.split(' ' * 60),start=1) if item.strip()]
        url=input(f"enter the item to search books: {items}")
        url=f"https://books.toscrape.com/catalogue/category/{url}/index.html"
        page= requests.get(url)
        src=page.content
        #print(src)
        soup=BeautifulSoup(src,'lxml')
        print(soup)
        li=soup.find_all(['li'],{'class':['col-xs-6 col-sm-4 col-md-3 col-lg-3']})
        print(len(li))
        book_details=[]
        for i in range(len(li)):
            image1=li[i].find('img')['src']
            title=li[i].find('img')['alt']
            price=li[i].find('p',{'class':'price_color'}).text.strip()
            stock=li[i].find('p',{'class':'instock availability'}).text.strip()
            rating=li[i].find('p',{'class':'star-rating'})['class'][1]
            book_details.append({"image_book":image1,"title_book":title,"price_book":price,"stock_book":stock,"ratings":rating})
        return book_details
```

```
    if number_page < 0:
        print("number_page is less from 0 ")
    else:
        data=scrap_website(number_page)
        df1=pd.DataFrame(data)

        df1=df1.to_csv("df1.csv")
    scrap_website_category(url)


main()
```

df1.csv

1 to 10 of 20 entries Filter

	image_book	title_book	price_book	stock_book	ratings
0	../media/cache/2c/da/2cdad67c44b002e7ead0cc35693c0e8b.jpg	A Light in the Attic	£51.77	In stock	Three
1	../media/cache/26/0c/260c6ae16bce31c8f8c95daddd9f4a1c.jpg	Tipping the Velvet	£53.74	In stock	One
2	../media/cache/3e/ef/3eef99c9d9adef34639f510662022830.jpg	Soumission	£50.10	In stock	One
3	../media/cache/32/51/3251cf3a3412f53f339e42cac2134093.jpg	Sharp Objects	£47.82	In stock	Four
4	../media/cache/be/a5/bea5697f2534a2f86a3ef27b5a8c12a6.jpg	Sapiens: A Brief History of Humankind	£54.23	In stock	Five
5	../media/cache/68/33/68339b4c9bc034267e1da611ab3b34f8.jpg	The Requiem Red	£22.65	In stock	One
6	../media/cache/92/27/92274a95b7c251fea59a2b8a78275ab4.jpg	The Dirty Little Secrets of Getting Your Dream Job	£33.34	In stock	Four
7	../media/cache/3d/54/3d54940e57e662c4dd1f3ff00c78cc64.jpg	The Coming Woman: A Novel Based on the Life of the Infamous Feminist, Victoria Woodhull	£17.93	In stock	Three
8	../media/cache/66/88/66883b91f6804b2323c8369331cb7dd1.jpg	The Boys in the Boat: Nine Americans and Their Epic Quest for Gold at the 1936 Berlin Olympics	£22.60	In stock	Four
9	../media/cache/58/46/5846057e28022268153beff6d352b06c.jpg	The Black Maria	£52.15	In stock	One

Show 10 per page

1 2