



WEB SCRAPPING

Submitted By:

Rishika Gupta (TEAM ELITE)

INDEX

- ◆ Technologies used
- ◆ Usage of libraries
- ◆ Code for WebScrapping Multiple pages
- ◆ Code
- ◆ Screenshot of the csv file
- ◆ Github Link



TECHNOLOGY USED

- ◆ PYTHON Programming Language
- ◆ Libraries used: Requests, Cloudscraper, BeautifulSoup, Pandas

USAGE OF LIBRARIES

- ◆ Requests: Requests library is one of the integral parts of Python for making HTTP requests to a specified URL
For Installation: `pip install requests`
- ◆ Cloudscraper: It is a web scraping library built exclusively for retrieving data from Cloudflare-protected websites. This ability and its compatibility with popular libraries like Python Requests and BeautifulSoup make it a valuable tool for data extraction. It is used to overcome the 403 forbidden error.
For Installation: `pip install cloudscraper`
- ◆ BeautifulSoup: Beautiful Soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping. For Installation: `pip install beautifulsoup4`
- ◆ Pandas: **Pandas** is an **open-source library** in **Python** that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. Here it is used for creating dataframes and converting the dataframes into csv
For Installation: `pip install cloudscraper`

WEB SCRAPPING MULTIPLE PAGES

In this challenge, The task was to develop a program capable of extracting specific information from multiple web pages and saving data gathered in either Excel or CSV.

For getting data from multiple pages a **for loop** will be used and we will use the next page button url to load the next page as such

```
np=soup.find("a",class_="next page-numbers").get("href")
```

Code

> Users > itsri > OneDrive > Desktop > web scrapping > projct > A.PY > ...

```
1 import requests
2 import cloudscraper
3 from bs4 import BeautifulSoup
4 import pandas as pd
5
6 Names=[]
7 Date=[]
8 Image_url = []
9 Like=[]
10 Blogs=[]
11 img_url_copy=[]
12
13
14 for i in range(1,45):
15     url="https://rategain.com/blog/page/"+str(i)
16     scraper=cloudscraper.create_scraper()
17     r=scraper.get(url)
18     soup=BeautifulSoup(r.text,'xml')
19     np=soup.find("a",class_="next page-numbers").get("href")
20     names=soup.find_all("div",class_="content")
21
22     for i in names:
23         n=i.find("h6").text
24         Names.append(n)
25
26     D=[]
27     date=soup.find_all("div",class_="bd-item")
28     for i in date:
29         sp=i.find("span")
30         s=sp.text
```

> Users > itsri > OneDrive > Desktop > web scrapping > projct > A.PY > ...

```
29         s=sp.text
30
31         D.append(s)
32     for k in range(len(D)):
33         if k%2==0:
34             Date.append(D[k])
35
36     Image=[]
37     for link in soup.find_all("a"):
38         a=link.get("data-bg")
39         Image.append(a)
40     Image_url=[]
41     for val in Image:
42         if val != None :
43             Image_url.append(val)
44     for i in Image_url:
45         l=len(Image_url)
46         if(l<9):
47             Image_url.append("---")
48     for item in Image_url:
49         img_url_copy.append(item)
50
51     l=soup.find_all("a",class_="zilla-likes")
52     for i in l:
53         e=i.find('span').text
54         m=e[:-5]
55         Like.append(m)
56 df=pd.DataFrame({"Blog title":Names,"Blog date":Date,"Blog image URL":img_url_copy,"Blog likes count":Like})
57 df.to_csv("blogSolution.csv")
```


SCREENSHOT OF THE CSV FILE

A	B	C	D	E
Sno.	Blog title	Blog date	Blog image URL	Blog likes count
1	A Complete Guide to Hotel Revenue Management	22-Nov-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/11/A-Complete-Guide-to-Hotel-Revenue-Management.png	1
2	The Ultimate Guide to Choosing the Right Hotel Booking Engine	22-Nov-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/11/The-Ultimate-Guide-to-Choosing-the-Right-Hotel-Booking-Engine.png	2
3	Hotelâ€™s Exposure with Google AdSenseâ€™s New "Related Search" Feature	20-Oct-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Google-AdSenses-New-Related-Search-Feature-in-Auto-Ads.png	32
4	d Reach & Frequency: Hotels' New Era with Facebook's 'Reservation' Buying	12-Oct-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Transforming-with-Facebooks-Reservation-BUYING-Strategy.png	16
4	Managing Overbookings and Cancellations with Hotel Booking Engines	5-Oct-23	rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Overbookings-Cancellations-Not-a-Problem-with-Modern-Hotel-Booking-Engines.png	11
5	al Distribution System (GDS) vs. Channel Manager: Which is Right for Your Hotel?	1-Oct-23	rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Hotel-Distribution-Done-Right-Choosing-between-GDS-and-Channel-Managers.png	10
6	Jingle All the Way: Europe Christmas Travel Trends	28-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Europe-Christmas-Travel-Trends.jpg	8
7	Why Bing Hotel Ads Should Be in Your Marketing Mix	28-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Bing-Hotel-Ads-Your-Hotels-Secret-Revenue-Booster.png	9
8	Unraveling the Alchemy: Year-End Travel Trends in Singapore	26-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Singapores-Secret-Recipe-for-Year-End-Travel-Success.png	7
9	Year-End U.S. Domestic Travel Insights: A Revenue Manager's Guide to Success	25-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/10/Is-U.S.A.-expecting-a-Domestic-Travel-Boom-this-year-end.png	7
10	Unlocking Top Travel Destinations in Egypt: A Hotelier's Playbook	25-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Unlocking-Top-Travel-Destinations-in-Egypt-A-Hoteliere-Playbook.jpg	7
11	Shaping Success in Egypt's Dynamic Hospitality Landscape: A Roadmap for Hotels	22-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Shaping-Success-in-Egypt's-Dynamic-Hospitality-Landscape.jpg	7
12	Checking In to a Greener Future: Renewable Energy in Hotels	15-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Powering-Hospitality-How-Renewable-Energy-Is-Reshaping-Hotels.png	6
13	Sustaining Communities: The Role of Travel Sellers in Sustainable Tourism	14-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/How-Travel-Sellers-Support-Local-Economies.png	6
14	Flowing Towards Sustainability: Water Conservation in Hotels	11-Sep-23	s://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Drip-by-Drip-How-Hotels-Are-Leading-the-Way-in-Water-Conservation.png	6
15	riendly Practices in Vacation Rentals: A Win-Win for Guests and the Environment	10-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Eco-friendly-practices-in-Vacation-Rentals.png	7
16	Zero-Waste Travel: A Sustainable Path to Explore the World	9-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Trash-to-Treasures-Zero-Waste-Travelers-Handbook.png	6
17	Building a Greener Future: Sustainable Supply Chain Management in Hotels	8-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Hotel-Sustainable-Supply-Chain-Strategies.png	5
18	Airline Sustainability Initiatives: How Airlines are Going Green	7-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Airlines-Sustainable-Innovations.png	4
19	Exploring Eco-Tourism: Where Nature Meets Responsible Travel	6-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Embrace-Eco-Tourism-Your-Passport-to-Sustainable-Wanderlust.png	6
20	Measuring and Reporting Sustainability Impact: Metrics for Hotels and Travel Sellers	6-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Measuring-and-Reporting-Sustainability-Impact.png	4
21	Reducing Carbon Footprint: Sustainable Practices for Tour Operators	5-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Sustainable-Practices-for-Tour-Operators.png	6
22	Preserving Paradise: Biodiversity Conservation in Travel	4-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Wings-and-Wonders-Biodiversity-Conservation-in-Tourism.png	5
23	Green is your Hotel? Do Good and Show it to the World â€“ on Google, for Europe	3-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Green-Hotel-Badge-for-Sustainable-Accommodations.png	7
24	The Digital Evolution of UK Hospitality: A Historic Journey From Key to Click	1-Sep-23	https://rategaincom.wpenginepowered.com/wp-content/uploads/2023/09/Evolution-of-UK-Hospitality.png	7

GITHUB LINK

<https://github.com/Rishika21Gupta/Web-Scrapping-Project.git>

THANK YOU