In-Class-Exercise-02

Introduction to ParseHub:

ParseHub is a very efficient user-friendly web-scraping tool. This tool has the capability to allow the user to scrape the website without writing the code. This tool is useful for individuals who want to perform research by scraping web pages. It is also useful for businesses that want to gather information to research.

There are both free and paid versions of this tool available. With the paid version there are advanced capabilities and the latest features being offered.

Some Advantages include:

- ParseHub is available in the cloud, which means it can be directly accessed without installation.
- It is very user-friendly as selecting the web elements is done by clicking.
- There is the chance of Previewing the scraped data, modifying, and editing.
- There are advanced techniques for selecting web elements.
- Exporting with ParseHub is a very easy task.
- The paid version of ParseHub also offers API integration.

Steps:

ParseHub follows a simple set of rules.

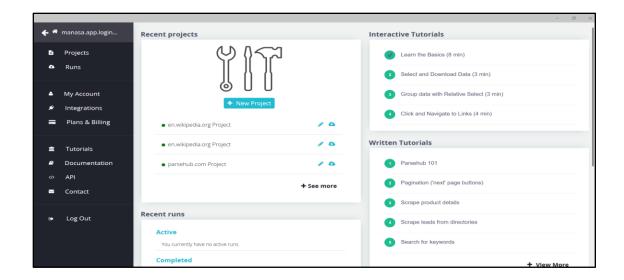
Select- for selecting a web element.

Correct- Confirming selected element.

Command- adding command to selection.

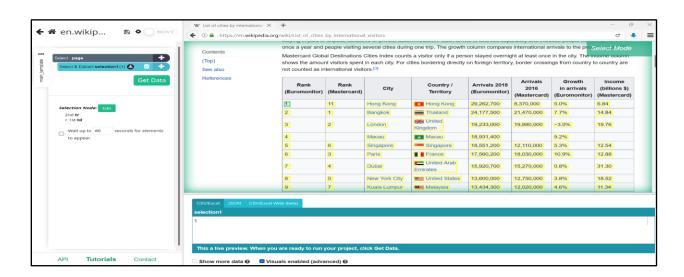
Example Project created using ParseHub:

1. After Installing ParseHub, create an account with your credentials and log in the ParseHub. Select the new project for creating the new scraping project

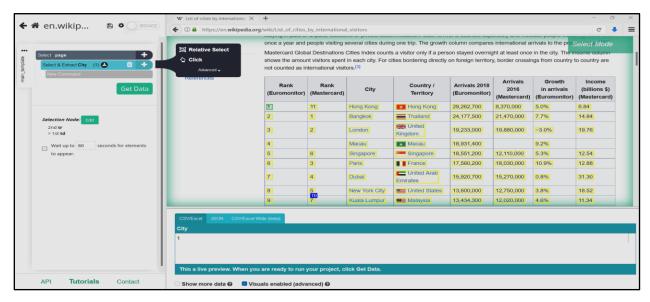


2. After selecting a new project. Provide the website URL from where the data needs to be scrapped. Here, I have selected the **Wikipedia website** that consists of **the "List of cities by international visitors."**

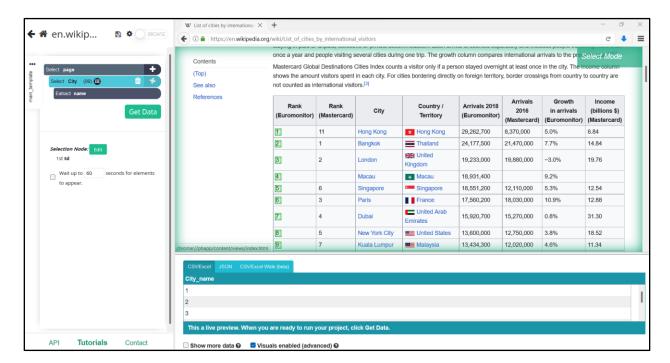




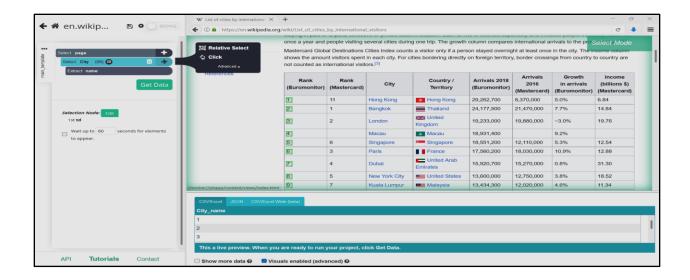
3. Once the URL is pasted, the website page will be loaded. Next, select the web scraping element from the website which you want to extract. Rename the element meaningfully.



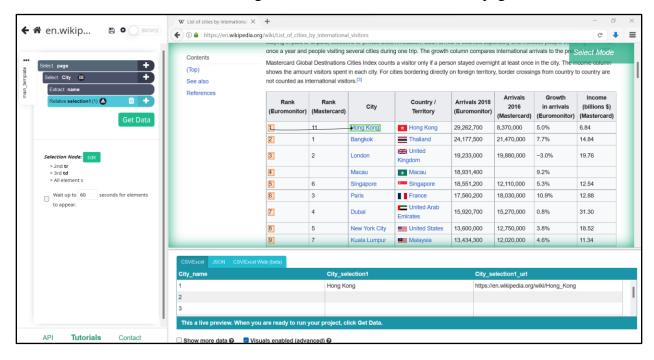
4. Select the second element from the website to select all the similar elements from the webpage.



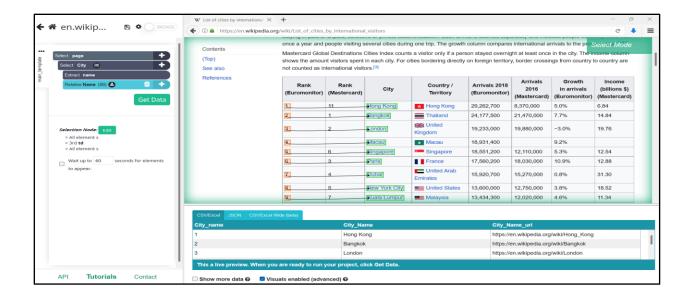
5. Once all the elements are selected. Press the + button beside the selected element tab. And selected relative select to link the other elements with he selected element.



6. Click the first elements and tag their relative elements on the webpage.



7. Select the second element and tag it to its relative element to tag all the elements with their respective relative elements and the element accordingly

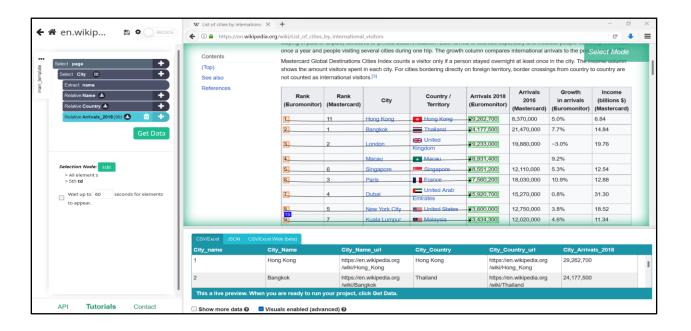


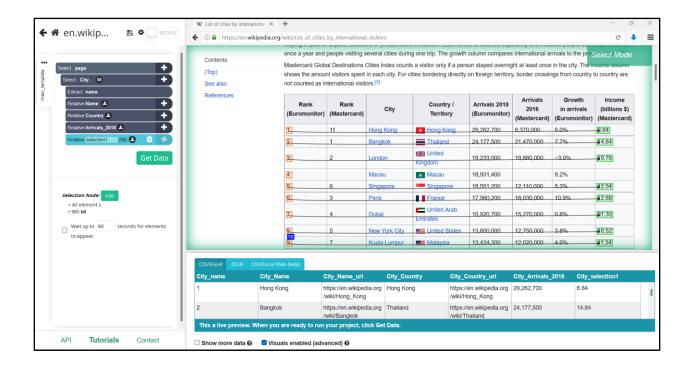
8. Similarly select the element and tag another relative element from the webpage that we want in our result. Name the relative element accordingly





9. Continue the Process with all the elements and all its relative elements. Name the relative elements accordingly

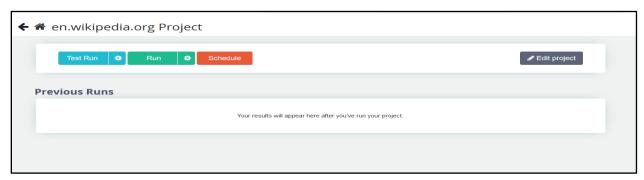


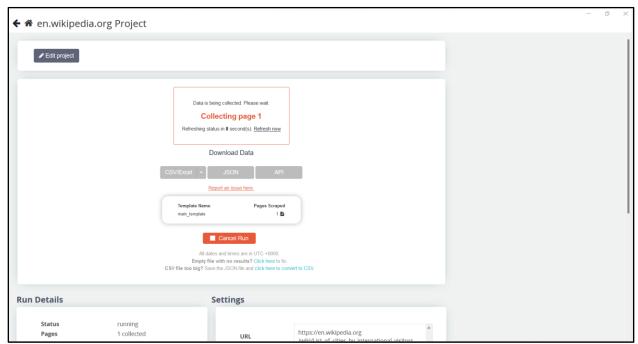


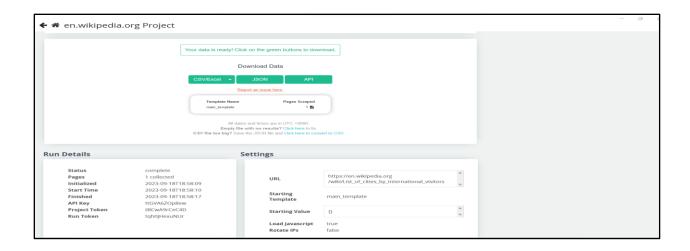
10. Once the all the elements are scraped. Preview the data scraped from the lower right corner from the window. If everything is correct, then press the "Get Data" button to start scraping the data selected.



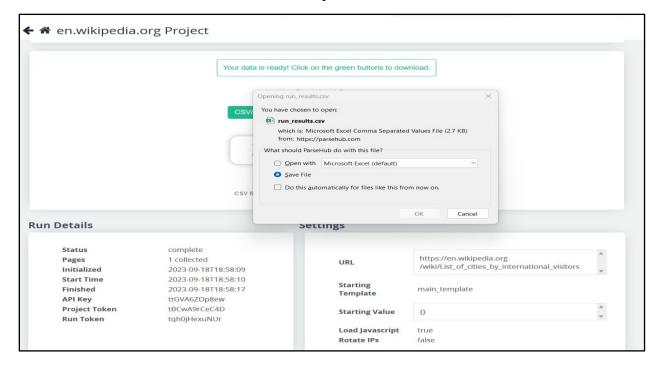
11. Press the "Run" button to start the scraping from the website.







12. When everything is ready. Save the file in the format required. For example I have chosen the CSV/Excel file to store the data scraped.



Output:

The sample of the output scraped looks like the one below. It consists of 100 rows. The link to output data

https://github.com/ManasaCherukupally1/Manasa_INFO5731_Spring2023/blob/main/Cherukupally_Manasa_wikipedia_results.csv

The columns include

- a. City_Rank-Rank of city
- b. City Name-Name of City
- c. City_Country-Country of city

- d. City_Arrivals- Number of arrivals in the city in 2018
- e. **City_income**-Income value of the city
- f. City_Name_Url- Url link for city description
- g. City_country_Url-Url link for description of country of city

	Α	В	C	D	E	F	G	Н
1	City_Rank	City_Name	City_Name	City_Coun	City_Country_url	City_Arrivals_2018	City_Income	_In_Billions
2	1	Hong Kong	https://en.	Hong Kong	https://en.wikipedia.org/wiki/Hong_Kong	29,262,700	6.84	
3	2	Bangkok	https://en.	Thailand	https://en.wikipedia.org/wiki/Thailand	24,177,500	14.84	
4	3	London	https://en.	United Kin	https://en.wikipedia.org/wiki/United_Kingdom	19,233,000	19.76	
5	4	Macau	https://en.	Macau	https://en.wikipedia.org/wiki/Macau	18,931,400		
6	5	Singapore	https://en.	Singapore	https://en.wikipedia.org/wiki/Singapore	18,551,200	12.54	
7	6	Paris	https://en.	France	https://en.wikipedia.org/wiki/France	17,560,200	12.88	
8	7	Dubai	https://en.	United Ara	https://en.wikipedia.org/wiki/United_Arab_Emirates	15,920,700	31.3	
9	8	New York	https://en.	United Sta	https://en.wikipedia.org/wiki/United_States	13,600,000	18.52	
10	9	Kuala Lum	https://en.	Malaysia	https://en.wikipedia.org/wiki/Malaysia	13,434,300	11.34	
11	10	Istanbul	https://en.	Turkey	https://en.wikipedia.org/wiki/Turkey	13,433,000	7.54	
12	11	Delhi	https://en.	India	https://en.wikipedia.org/wiki/India	12,645,300	1.91	
13	12	Antalya	https://en.	Turkey	https://en.wikipedia.org/wiki/Turkey	12,438,800		
14	13	Shenzhen	https://en.	China	https://en.wikipedia.org/wiki/China	12,202,100	0.83	
15	14	Mumbai	https://en.	India	https://en.wikipedia.org/wiki/India	10,590,100	3.6	
16	16	Phuket	https://en.	Thailand	https://en.wikipedia.org/wiki/Thailand	10,550,700		
17	17	Rome	https://en.	Italy	https://en.wikipedia.org/wiki/Italy	10,065,400	4.47	
18	18	Tokyo	https://en.	Japan	https://en.wikipedia.org/wiki/Japan	9,985,100	13.48	
19	19	Pattaya	https://en.	Thailand	https://en.wikipedia.org/wiki/Thailand	9,606,400		
20	20	Taipei	https://en.	Taiwan	https://en.wikipedia.org/wiki/Taiwan	9,597,800	9.6	
21	21	Mecca	https://en.	Saudi Arab	https://en.wikipedia.org/wiki/Saudi_Arabia	9,565,200		
22	22	Guangzho	https://en.	China	https://en.wikipedia.org/wiki/China	9,004,800	1.66	
23	23	Prague	https://en.	Czech Rep	https://en.wikipedia.org/wiki/Czech_Republic	8,948,600	2.7	
24	24	Medina	https://en.	Saudi Arab	https://en.wikipedia.org/wiki/Saudi_Arabia	8,547,200		
25	25	Seoul	https://en.	South Kore	https://en.wikipedia.org/wiki/South_Korea	8,431,400	11.3	
26	~		//		1 // 21 12 / 21/61 1 1 1	0.354.300	4.3	