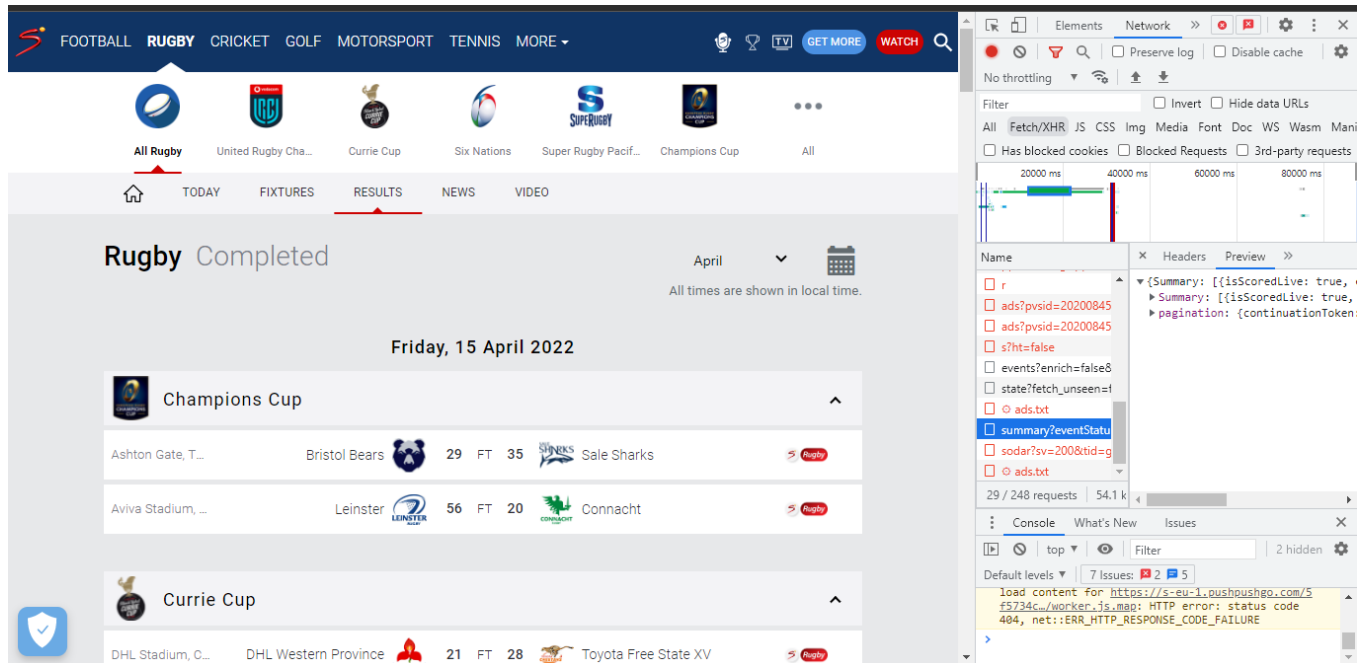


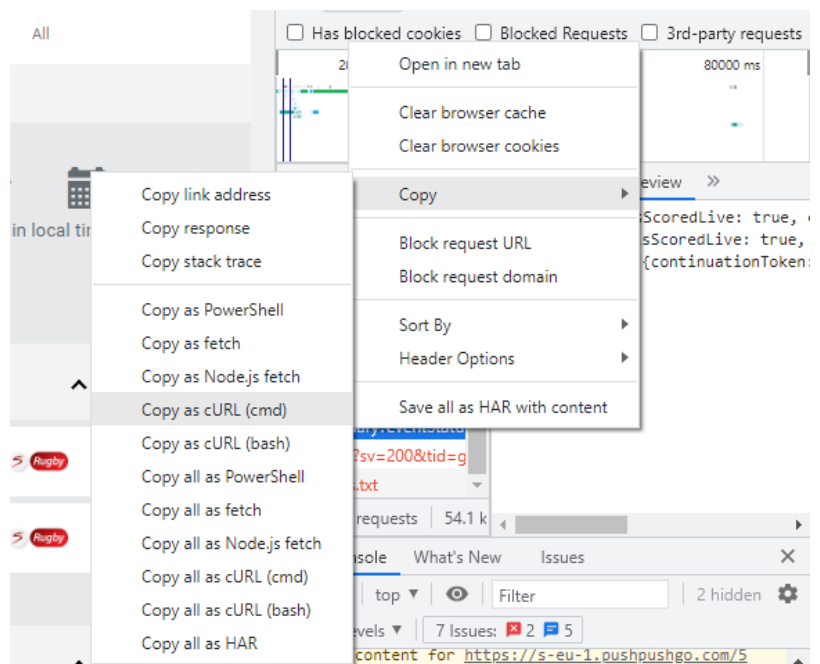
Scraping the Rugby Results from www.supersport.com using Insomnia and Python



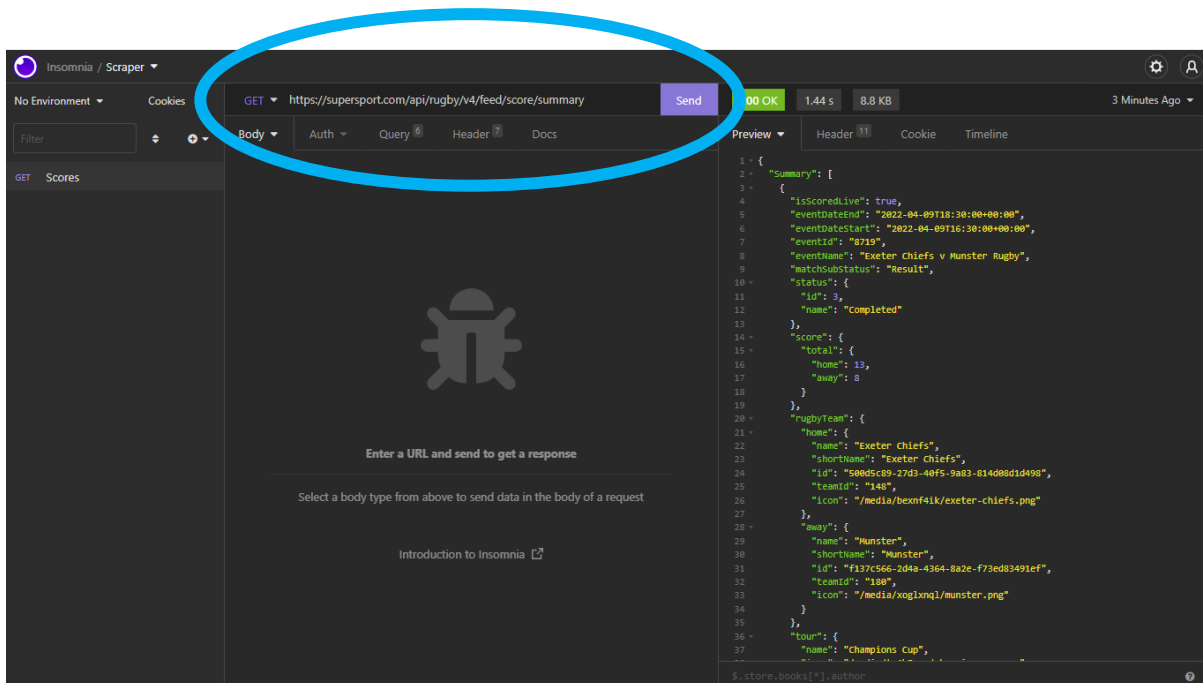
Navigate to www.supersport.com/rugby/results/ and inspect the page.

Under “Network” and “Fetch/XHR” scroll through the options to find the results information in “Preview”.

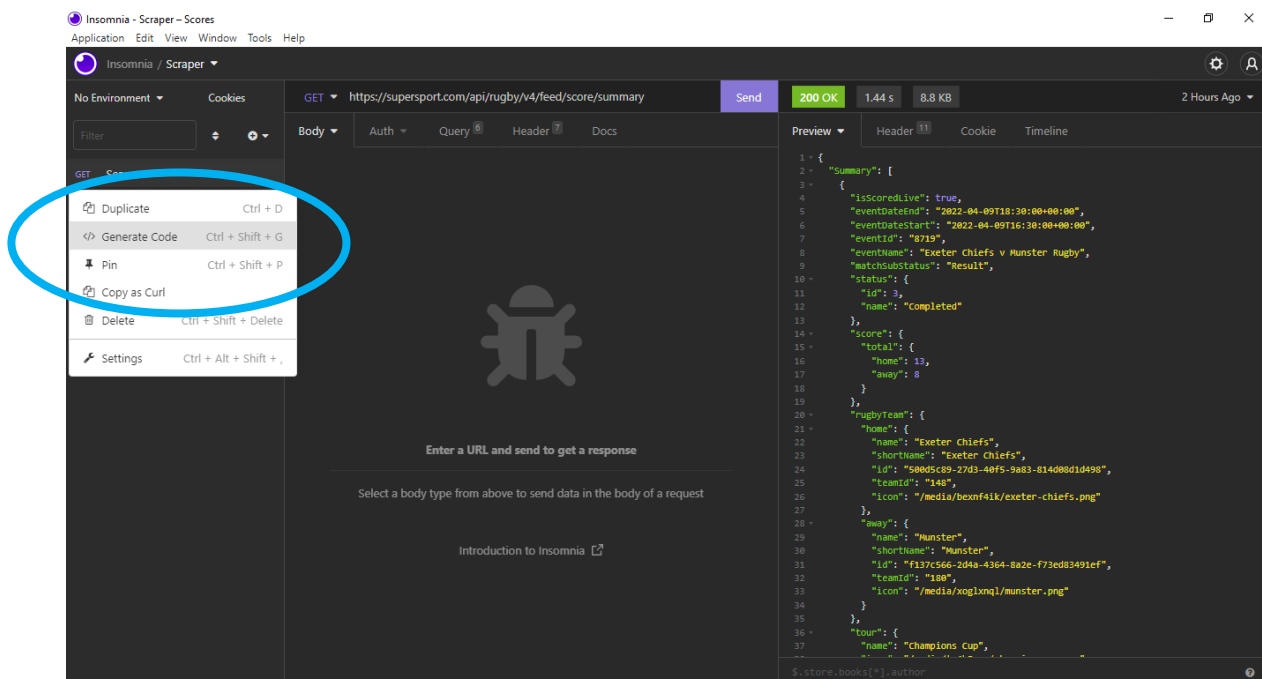
Once that’s found, right click on the link and choose the cURL option.



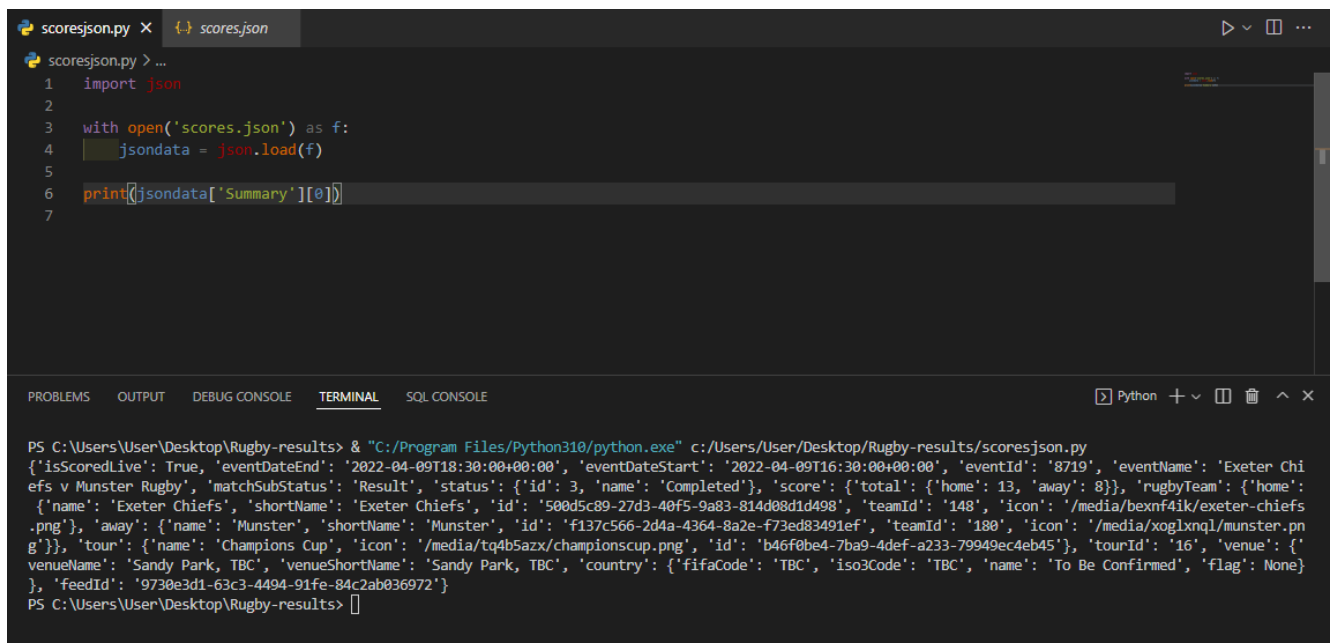
Create a new request in Insomnia and paste the link you copied in the “GET” line and click “Send”. A preview of the data will appear on the right. This contains all the information you will need.



Click the drop down option on your project in the left pane and click “Generate Code”.



/ If you're using Visual Studio Code don't forget to select an interpreter for Python /



The screenshot shows a Visual Studio Code editor with a file named `scoresjson.py` open. The script contains the following code:

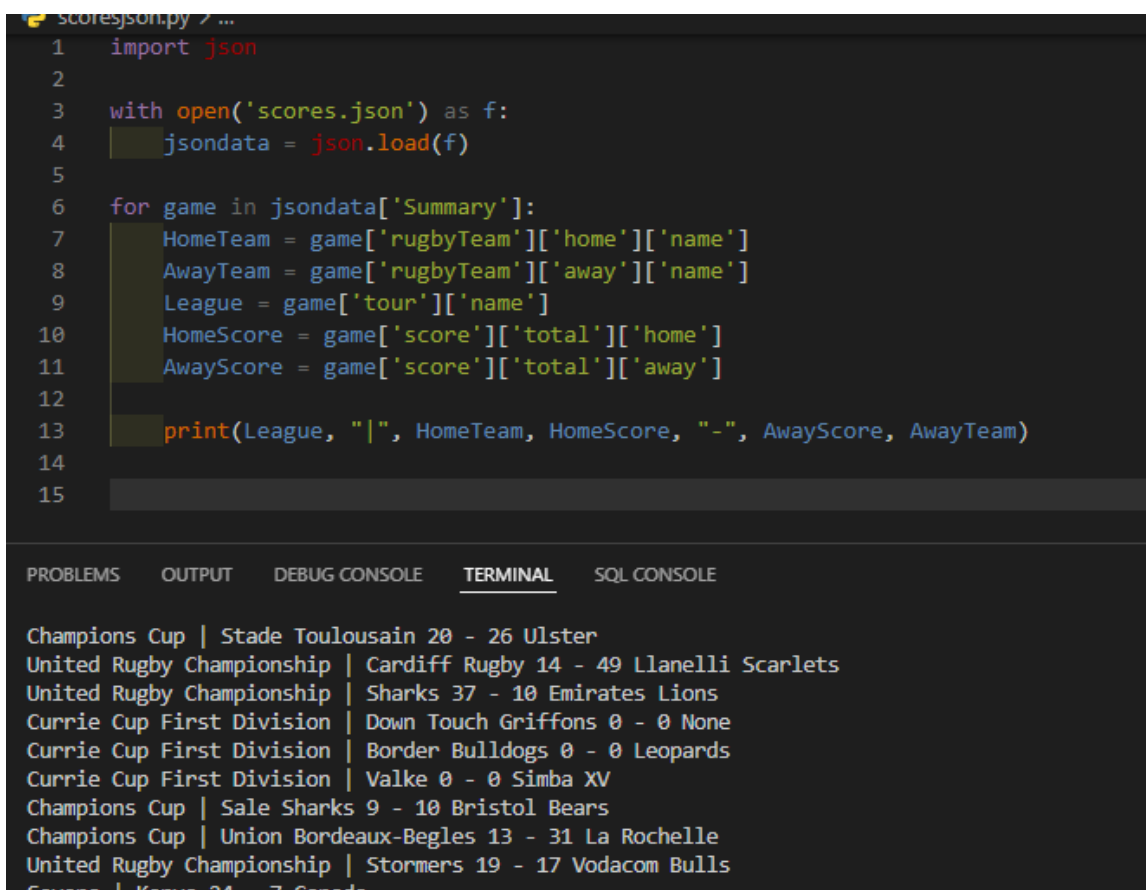
```
1 import json
2
3 with open('scores.json') as f:
4     jsongdata = json.load(f)
5
6 print(jsongdata['Summary'][0])
7
```

The terminal output shows the execution of the script, displaying a large JSON object representing a rugby match summary for Exeter Chiefs vs Munster Rugby.

```
PS C:\Users\User\Desktop\Rugby-results> & "C:/Program Files/Python310/python.exe" c:/Users/User/Desktop/Rugby-results/scoresjson.py
{"isScoredLive": True, "eventDateEnd": "2022-04-09T18:30:00+00:00", "eventDateStart": "2022-04-09T16:30:00+00:00", "eventId": "8719", "eventName": "Exeter Chiefs v Munster Rugby", "matchSubStatus": "Result", "status": {"id": 3, "name": "Completed"}, "score": {"total": {"home": 13, "away": 8}}, "rugbyTeam": {"home": {"name": "Exeter Chiefs", "shortName": "Exeter Chiefs", "id": "500d5c89-27d3-40f5-9a83-814d08d1d498", "teamId": "148", "icon": "/media/bexmf4ik/exeter-chiefs.png"}, "away": {"name": "Munster", "shortName": "Munster", "id": "f137c566-2d4a-4364-8a2e-f73ed83491ef", "teamId": "180", "icon": "/media/xoglxnql/munster.png"}}, "tour": {"name": "Champions Cup", "icon": "/media/tq4b5azx/championscup.png", "id": "b46f0be4-7ba9-4def-a233-79949ec4eb45"}, "tourId": "16", "venue": {"venueName": "Sandy Park, TBC", "venueShortName": "Sandy Park, TBC", "country": {"fifaCode": "TBC", "iso3Code": "TBC", "name": "To Be Confirmed", "flag": None}}, "feedId": "9730e3d1-63c3-4494-91fe-84c2ab036972"}
PS C:\Users\User\Desktop\Rugby-results>
```

Here the first item inside the main list is being retrieved to determine what information is useful.

After finding the names of the information you want, create a for loop to scrape the data.



The screenshot shows a Visual Studio Code editor with a file named `scoresjson.py` open. The script contains the following code:

```
1 import json
2
3 with open('scores.json') as f:
4     jsongdata = json.load(f)
5
6 for game in jsongdata['Summary']:
7     HomeTeam = game['rugbyTeam']['home']['name']
8     AwayTeam = game['rugbyTeam']['away']['name']
9     League = game['tour']['name']
10    HomeScore = game['score']['total']['home']
11    AwayScore = game['score']['total']['away']
12
13    print(League, "|", HomeTeam, HomeScore, "-", AwayScore, AwayTeam)
14
15
```

The terminal output shows the execution of the script, displaying a list of rugby matches and their scores.

```
Champions Cup | Stade Toulousain 20 - 26 Ulster
United Rugby Championship | Cardiff Rugby 14 - 49 Llanelli Scarlets
United Rugby Championship | Sharks 37 - 10 Emirates Lions
Currie Cup First Division | Down Touch Griffons 0 - 0 None
Currie Cup First Division | Border Bulldogs 0 - 0 Leopards
Currie Cup First Division | Valke 0 - 0 Simba XV
Champions Cup | Sale Sharks 9 - 10 Bristol Bears
Champions Cup | Union Bordeaux-Begles 13 - 31 La Rochelle
United Rugby Championship | Stormers 19 - 17 Vodacom Bulls
Sevens | Kenya 24 - 7 Canada
```

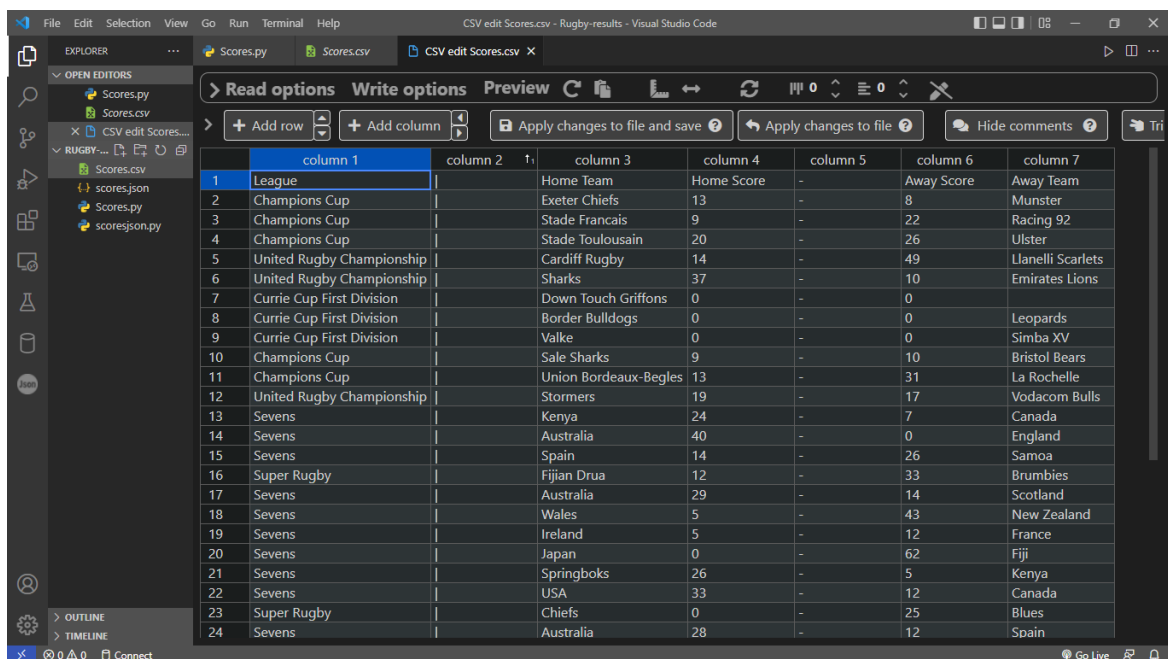
Once you have successfully scraped the data from the local file, copy the for loop into the file you created earlier with the code generated by Insomnia.

Here you will need to import writer from csv to export the data to a csv file.

* Note the change `json.load` to `json.loads` in this file. This is because it is loading the strings externally from the server.

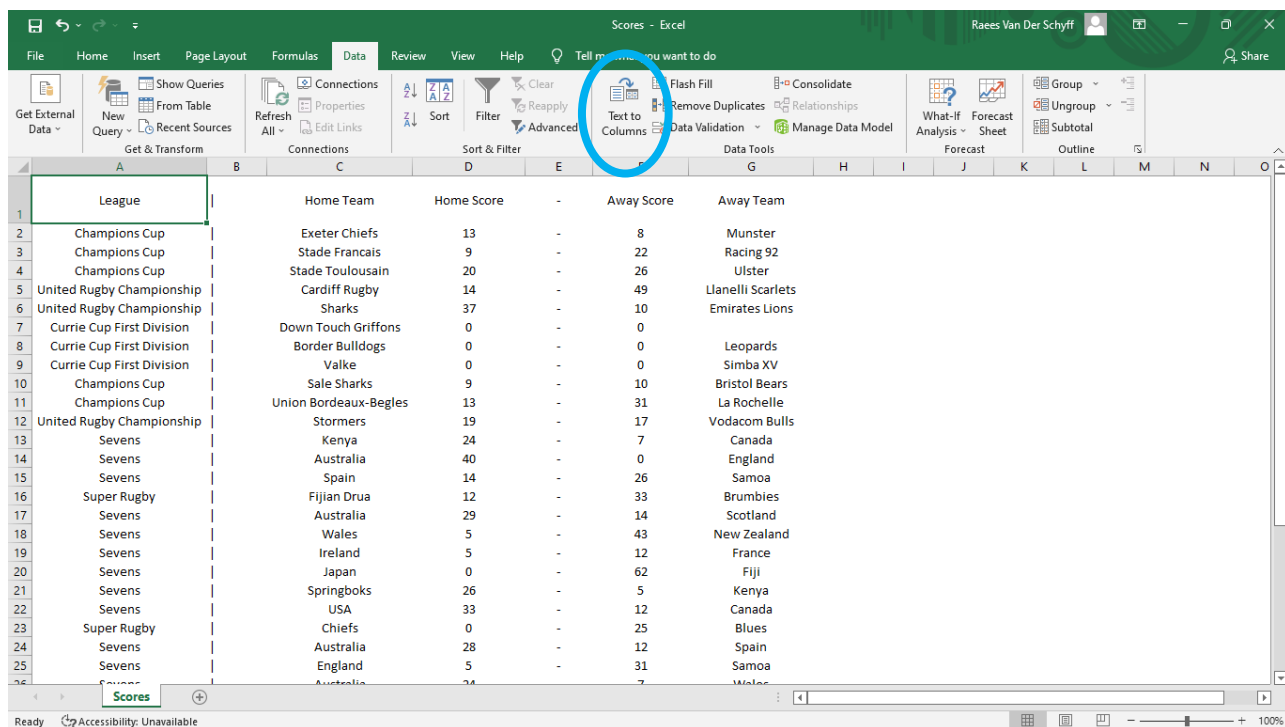
```
10 headers = {
11     "authority": "supersport.com",
12     "accept": "application/json",
13     "accept-language": "en-GB,en-US;q=0.9,en;q=0.8",
14     "cookie": "_gid=GA1.2.1598209652.1650061949; _gcl_au=1.1.1606404431.1650061950; cookie_consent_performance=1; cookie_
15     "dnt": "1",
16     "referer": "https://supersport.com/rugby/results/",
17     "sec-ch-ua": "^\\^"
18 }
19
20 response = requests.request("GET", url, data=payload, headers=headers, params=querystring)
21
22 jsondata = json.loads(response.text)
23
24 with open('Scores.csv', 'w', encoding='utf=8', newline='') as f:
25     thewriter = writer(f)
26     header = ['League', '|', 'Home Team', 'Home Score', '-', 'Away Score', 'Away Team']
27     thewriter.writerow(header)
28
29     for game in jsondata['Summary']:
30         HomeTeam = game['rugbyTeam']['home']['name']
31         AwayTeam = game['rugbyTeam']['away']['name']
32         League = game['tour']['name']
33         HomeScore = game['score']['total']['home']
34         AwayScore = game['score']['total']['away']
35
36         info = [League, "|", HomeTeam, HomeScore, "-", AwayScore, AwayTeam]
37         thewriter.writerow(info)
```

After running this, a file named Scores.csv will be created in the current folder. Using the Edit CSV extension, you can open the file directly in VS Code.



	column 1	column 2	column 3	column 4	column 5	column 6	column 7
1	League		Home Team	Home Score	-	Away Score	Away Team
2	Champions Cup		Exeter Chiefs	13	-	8	Munster
3	Champions Cup		Stade Francais	9	-	22	Racing 92
4	Champions Cup		Stade Toulousain	20	-	26	Ulster
5	United Rugby Championship		Cardiff Rugby	14	-	49	Llanelli Scarlets
6	United Rugby Championship		Sharks	37	-	10	Emirates Lions
7	Currie Cup First Division		Down Touch Griffons	0	-	0	
8	Currie Cup First Division		Border Bulldogs	0	-	0	Leopards
9	Currie Cup First Division		Valke	0	-	0	Simba XV
10	Champions Cup		Sale Sharks	9	-	10	Bristol Bears
11	Champions Cup		Union Bordeaux-Begles	13	-	31	La Rochelle
12	United Rugby Championship		Stormers	19	-	17	Vodacom Bulls
13	Sevens		Kenya	24	-	7	Canada
14	Sevens		Australia	40	-	0	England
15	Sevens		Spain	14	-	26	Samoa
16	Super Rugby		Fijian Drua	12	-	33	Brumbies
17	Sevens		Australia	29	-	14	Scotland
18	Sevens		Wales	5	-	43	New Zealand
19	Sevens		Ireland	5	-	12	France
20	Sevens		Japan	0	-	62	Fiji
21	Sevens		Springboks	26	-	5	Kenya
22	Sevens		USA	33	-	12	Canada
23	Super Rugby		Chiefs	0	-	25	Blues
24	Sevens		Australia	28	-	12	Spain

Open the file with Excel and use the Text to Columns options in Data to separate the information into cells.



If you followed all these steps successfully then well done on scraping your first website!

Thanks for reading! And happy scraping (: