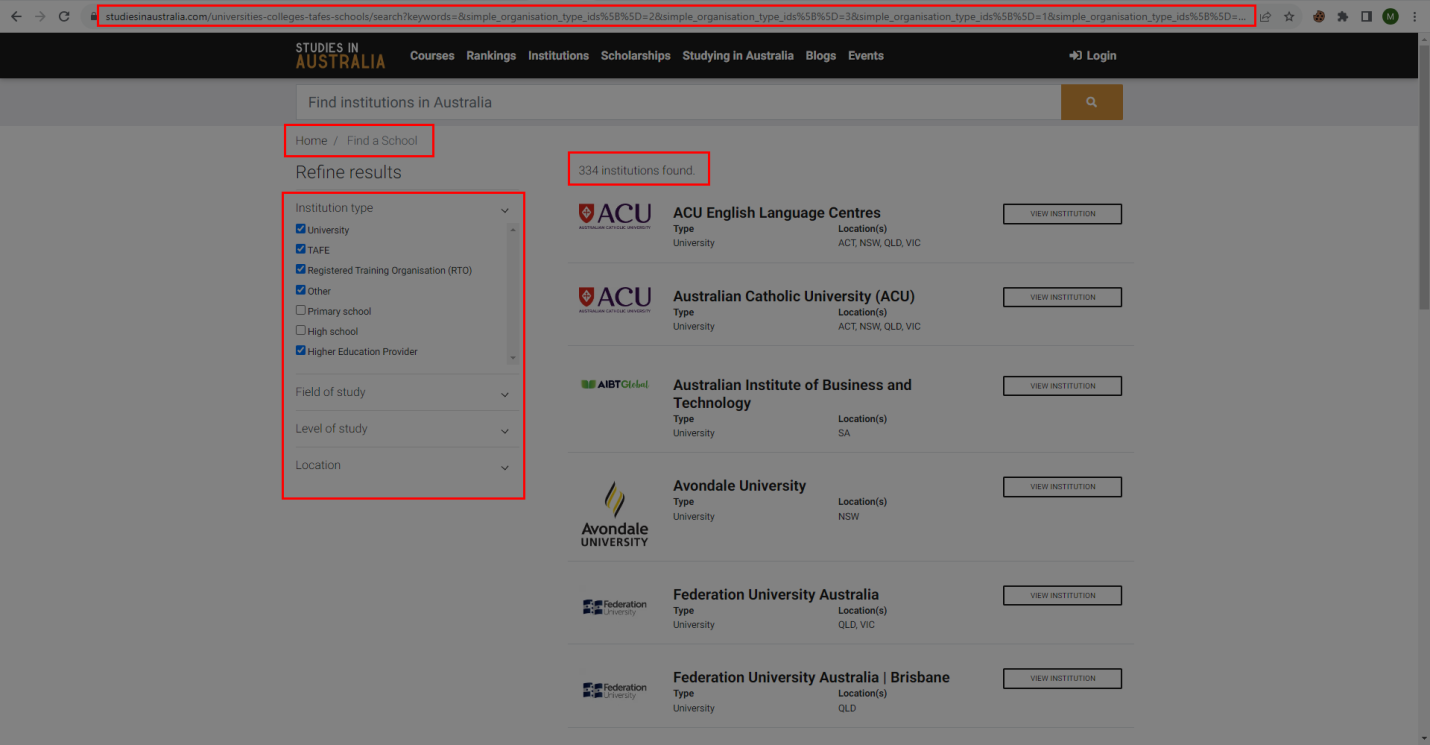
**STUDIES IN AUSTRALIA DOCUMENT**

**Table of contents:**

* Filter page
* Get university or college URL
* Save university or college URL
* Get university or college tab URL
* Save university or college tab URL
* Get campus tab data
* Get rating tab data
* Conclusion

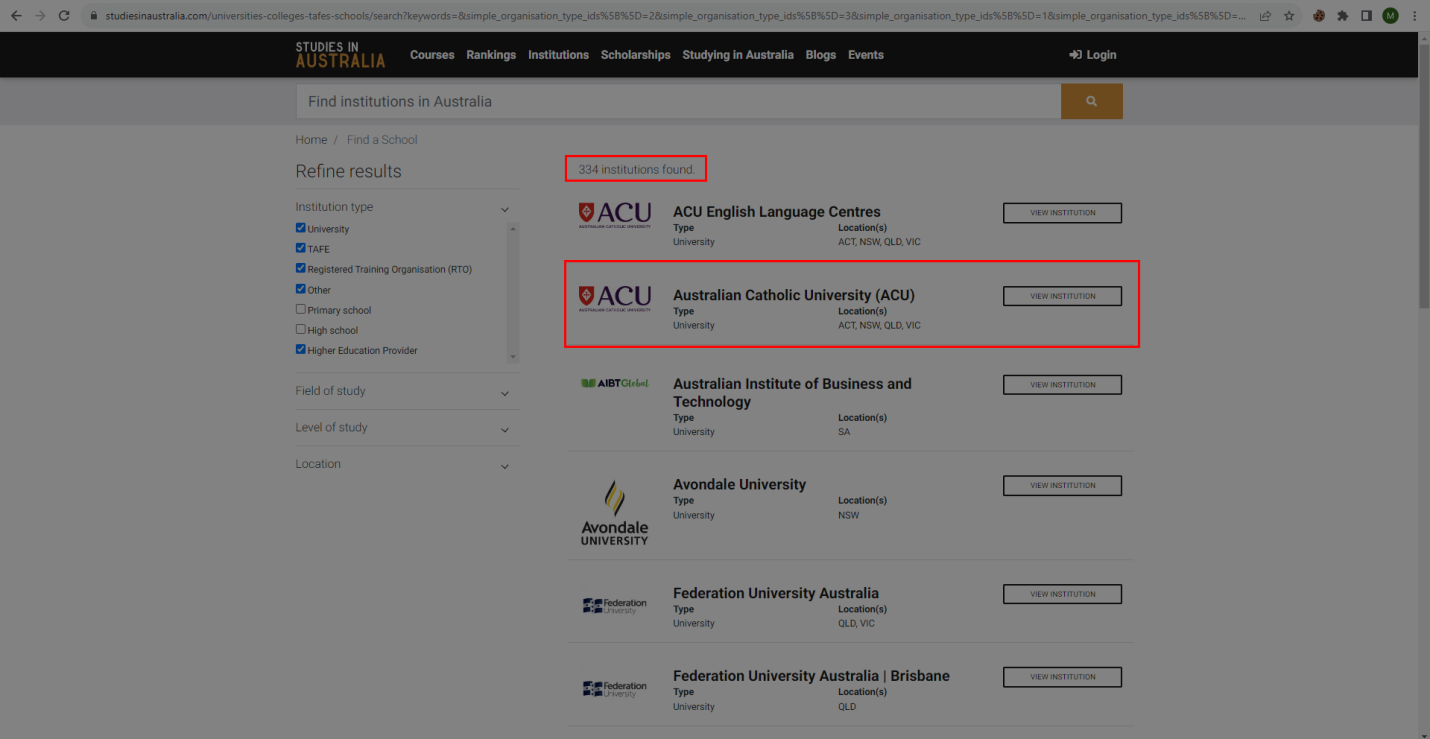
**Filter page**

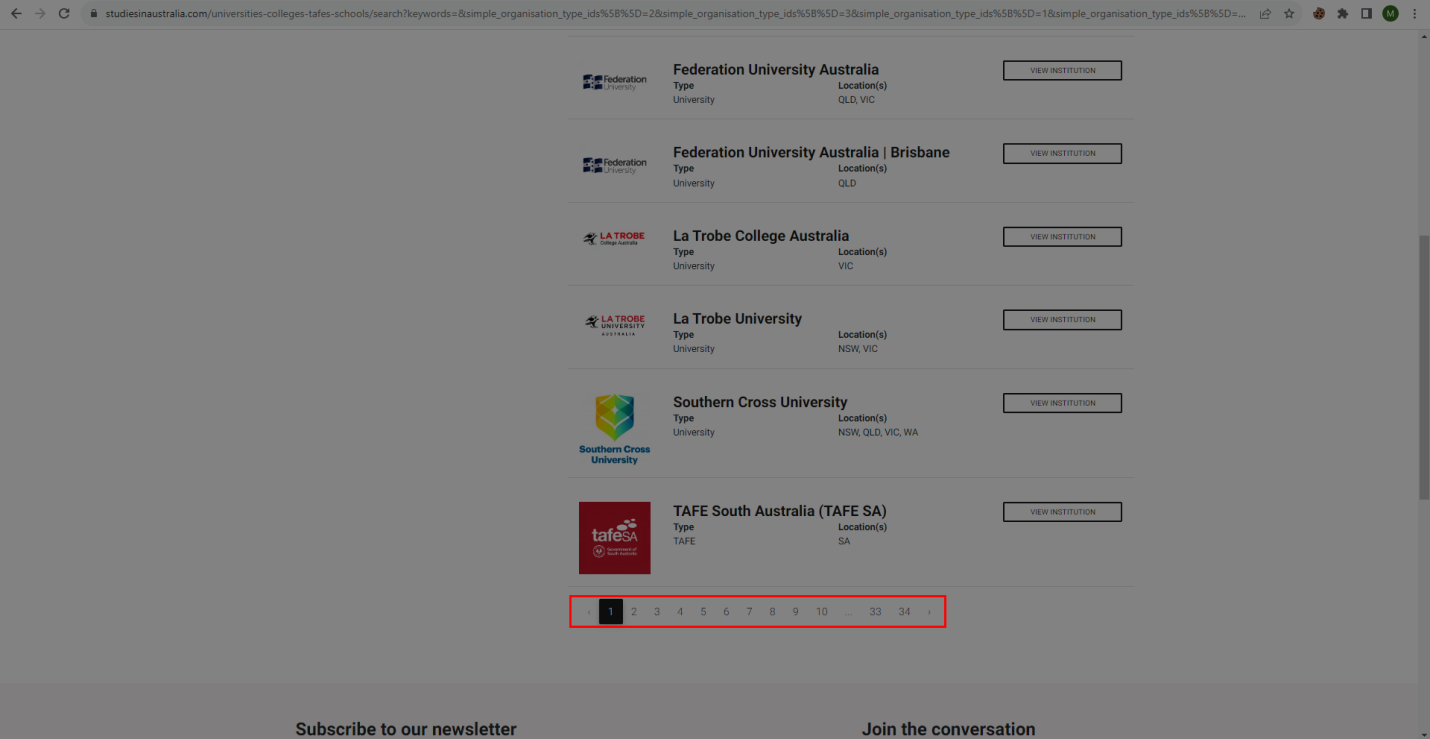
We need **Studies in Australia** website university or college campus and ratings tab data. So we go to this website filter page then institution type filter box select all categories except **primary school and high school** filter. 

**Get university or college URL**

After applied filter Get University or college URL’s from that all pagination pages.

Use **Get\_universities\_urls.py** script fot get all university or college URL’s.



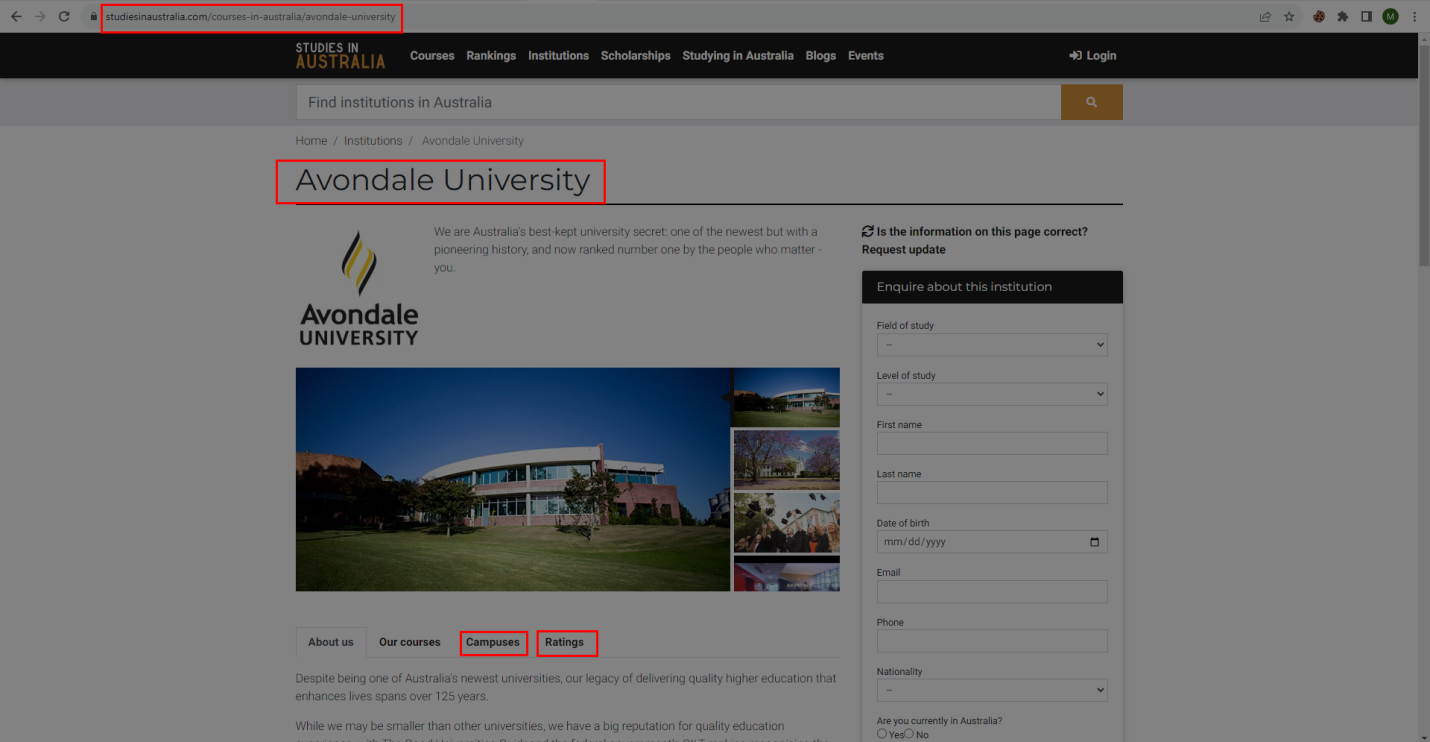


**Save university or college URL’s**

After got university or college URL’s we want to save as a HTML files.

Use **save\_universities\_or\_institution\_urls.py** script for save university or college URL’s.

**Get university or college tab URL’s**

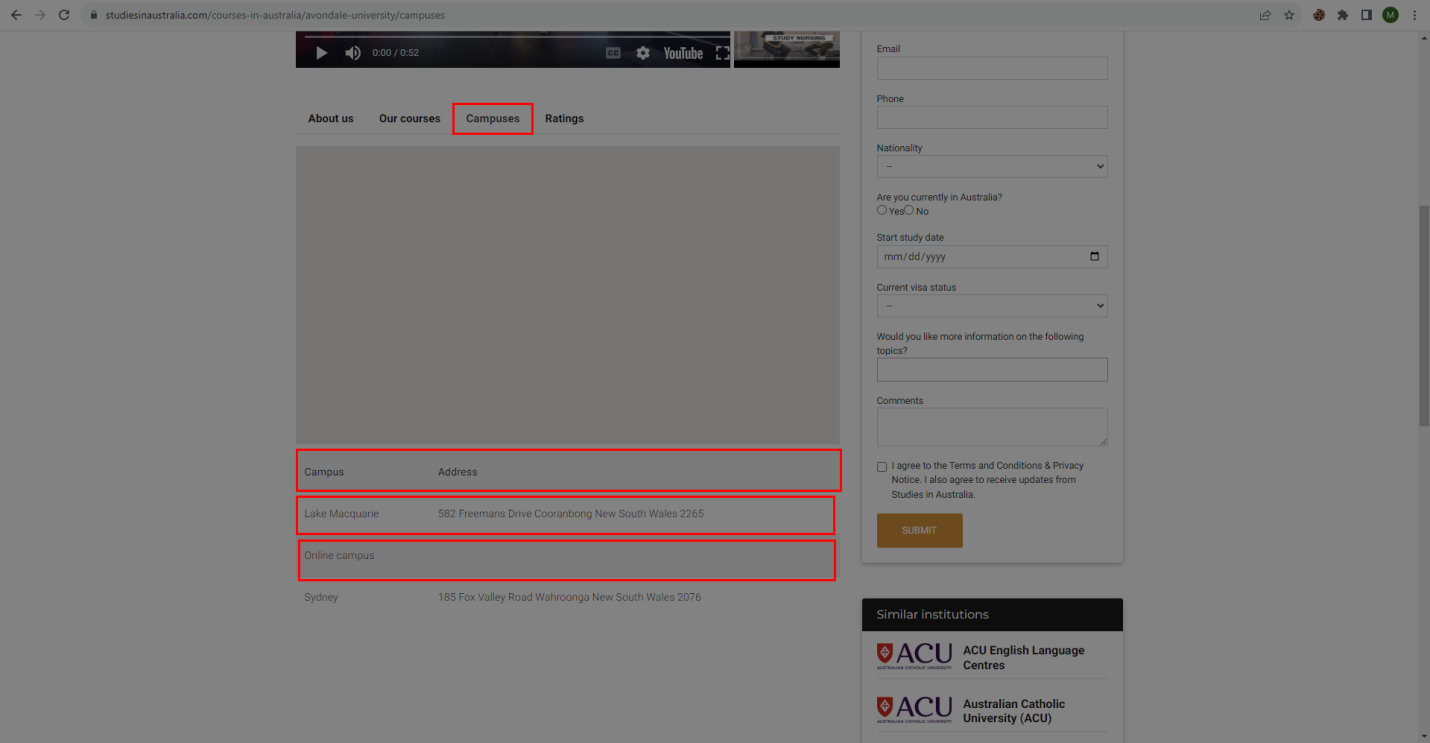
We want campus and ratings tab data. So we need that tab URL’s. After saved university or college URL’s as HTML file we get campus and ratings tab URL’s as a CSV file from already saved HTML files. Use **get\_universites\_tab\_urls.py** script for get campus tab and rating tab URL’s. 

**Save university or college URL’s**

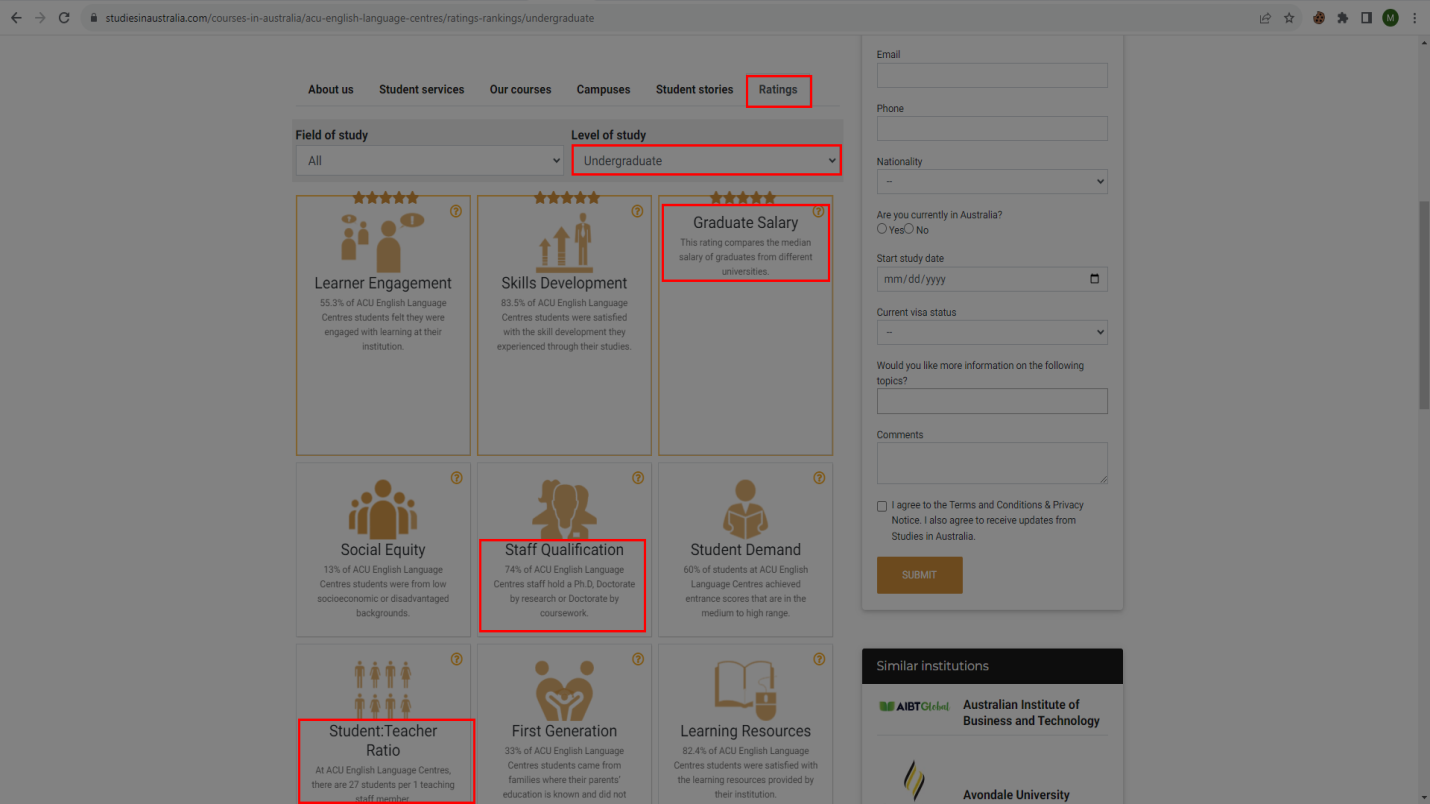
After got university or college campus and ratings tab URL’s we want to save as a HTML files.

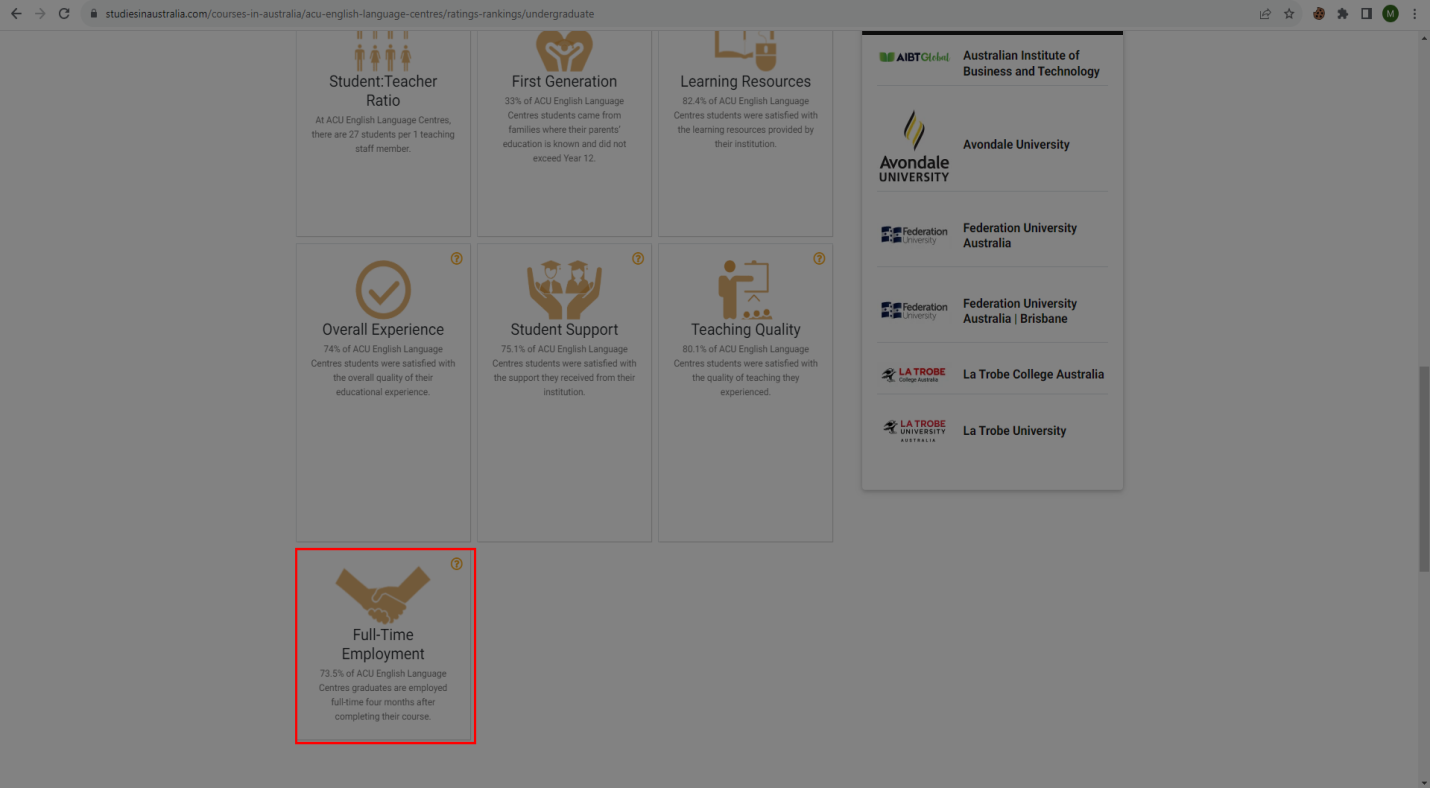
Use **save\_campus\_urls.py** script for save university or college campus and ratings tab URL’s.

**Get campus tab data**

We get campus tab data from already saved campus tab HTML files. We get campus and address data from that HTML files. Then we want latitude and longitude values so we use each address location URL’s. Some address don’t have URL’s and that location URL’s don’t correct location in that situation we skip that URL’s. If we continuously request location URL’s that URL’s are blocked our request. So we use **Scraper API** proxy rotator. After use proxy rotator we get **latitude** and **longitude** values. Use **get\_campus\_data.py** script for get campus data. 

**Get ratings data**

Ratings tab have two types of data that is **undergraduate** and **postgraduate** data. We collect undergraduate and postgraduate URL’s when we get ratings URL’s. We get same data for undergraduate and postgraduate tabs so we use one method in code. We get graduate salary, staff qualification, student teacher ratio and full time employment values. Use **get\_ratings\_data.py** script for get ratings data. 



**Conclusion**

Now we have all CSV files. This is a whole process for get Studies in Australia university or college’s data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| URL Type | Total URL’s | Success URL | Failure URL | Permanent Failure URL |
| Pagination Page | 154 | 154 | 0 | 0 |
| University or college URL’s | 334 | 325 | 9 | 9 |
| Campus URL’s | 324 | 324 | 0 | 0 |
| Academics URL’s | 90 | 90 | 0 | 0 |

**🟋🟋🟋**