

Vaidehi Lehekar

9307028831

Web Scraping data from website:

<https://www.houzz.in/professionals/civil-engineers-and-contractors/c/Chennai--Tamil-Nadu>

I chose the website

<https://www.houzz.in/professionals/civil-engineers-and-contractors/c/Chennai--Tamil-Nadu>

Due to its large number of entries with all the information I needed on one page.

I used Python Packages: BeautifulSoup4 and Selenium to extract the information I needed.

Reasons for choosing these packages:

1. BeautifulSoup4 is a very easy to use, open source package I was familiar with as I had used it before for personal projects.
2. However when I started to write the script I noticed that the site I used contained phone numbers of the companies as a hidden element, which was easier to solve with methods in selenium. So, I learnt selenium and extracted all the hidden elements on the site.

I have also included the .ipynb file in which I have written the script for this assignment.