

Web-Scraping Assignment

About OC Impact:

The Opportunity Collaboration (OC) is a global network of leaders dedicated to building sustainable solutions to poverty and injustice. They annually convene a 3-day collaborative summit that is uniquely designed to help forge alliances and advance organizations' social impact initiatives.

The OC Impact initiative caters to a diverse audience, including grant-makers, impact investors, corporate partners, social entrepreneurs, nonprofits, academic institutions, media and governments.

Delegates in OC Impact are individuals who contribute significantly to the Opportunity Collaboration event, bringing diverse perspectives and experiences to collaborative discussions. OC Fellows, in particular, are recognized for their impactful work and commitment to addressing global challenges related to poverty and injustice.

What you'll be doing:

- Go through the list of OC Impact delegates as mentioned in the given URL- <https://ocimpact.com/delegate-roster/>
- Write code to scrape the information of all delegates listed, including their name, job title, organization, and the answer to each question they have answered.
- Clean the data to remove any HTML tags or irrelevant information, and handle missing values as you deem fit.
- Store the final output in a local instance of DuckDB and share a screenshot of the same before the deadline.

Hint: This is a dynamic website, use the appropriate web-scraping libraries.

Output:

The final data should contain the following columns-

Name	Job title	Organization	Question 1	Question 2	...	Question N
------	-----------	--------------	------------	------------	-----	------------

Note: Replace "Question 1", "Question 2", etc. with the actual questions.

For example, Question 1 can be "What are you working on in the field of social and economic justice?". The values of the column should be the answer that the delegates have given to that question.