How it works:

Before performing any data scraping, we first install Docker locally on our computer, because it is required for the library RSelenium to work. After we map the container port, to the host port, we will need to pull and load the image of the browser that we will be using (for example Firefox) through the command prompt of our operating system. Once that is done, we can start the web scraping procedure.

- First of all, we construct our url, based on the country and the term to be searched. For instance:

  https://www.linkedin.com/jobs/search?keywords=%22Data Scientist%22&location=England&position=1&pageNum=0

  This refers to the position of a "Data Scientist" in England and it directs on the first page of the LinkedIn search engine.

- After that is done, we request a connection with the HTTP server by using the link that was already constructed.

- When we get directed to the first page of LinkedIn, we scroll down 5 times, in order to load the maximum number of job advertisements that we can for the country and the search term that we selected above.

- Afterwards, we open the source code and parse the HTML code in order to get it all in a script back on our computer, by using the XML library.

- From that script, we select only the internal href links that contain the job advertisements.

- After constructing this list of internal, anchor links within the site, we access each and every one of them, in order to parse the job descriptions and characteristics of interest and store them in a huge data frame, which is done by using the library rvest. This library allows us to access the HTML, CSS and JavaScript code hierarchically and get to the script that interests us (in this case the job description, as well as Industry and Seniority Level variables).

- Within that loop a pattern-searching vector is applied to the description of the job which tries to detect words of interest (such as R, Python, SAS, etc.), considering lowercase, as well as uppercase terms that could be included.

- After that procedure is done, the algorithm terminates the server connection to the site and returns the data frame that the user asked.