

Assignment: -

1. A developer is assigned a task to scrape 1 lakh website pages from a directory site, while scrapping he is facing such hcaptcha, which are placed to stop people from scrapping As a project Coordinator suggest ways to solve this problem

**Answer:** If the website has an official API for accessing data, we can use it since that is a legitimate way of doing so. Other than that, we can have some delay between the requests to the website so that it appears as a more human-like behaviour and doesn't trigger anti-scraping mechanisms.

2. Our client has around 10k linkedin people profiles, he wants to know the estimated income range of these profiles. Suggest ways on how to do this?

**Answer:** We can use machine learning models trained on salary data to predict income ranges based on various LinkedIn profile attributes such as job role, company, education etc. One other way is to survey a small sample of willing profile owners and extrapolate the results for the entire data.

3. We have a list of 1L company names, need to find linkedin company links of these profiles, how to go about this?

**Answer:** We can use an automated approach and apply web scraping methods to retrieve LinkedIn page links from search engine results or the websites of the desired companies.

4. How to identify list of companies whose tech stack is built on Python. Give names of 5 companies if possible, by your suggested approach

**Answer:** We can go to different company websites which often show the tech stack they use to build their projects and then list the companies which use Python. Five famously known companies whose tech stacks are built on Python are Google, Instagram, Spotify, Reddit and Dropbox.

5. Need to find an API, through which we can send linkedin messages to other linkedin users

**Answer:** We can use the LinkedIn API to send LinkedIn messages to other LinkedIn users.