

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

→

- Depending on the complexity of the CAPTCHAs, consider using CAPTCHA-solving services such as 2Captcha or Anti-Captcha. These services use human solvers to solve CAPTCHAs and provide an API for automation.
- Schedule your scraping tasks to run during off-peak hours when the website may have fewer users and less stringent anti-scraping measures.
- Continuously monitor the scraping process for changes in the website's defense mechanisms. Be prepared to adapt your scraping strategies accordingly.
- Rotate your IP addresses to avoid being blocked or encountering hCaptcha. This can be achieved using proxy server or VPN.

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?

→

- You can collect publicly available information from these profiles, such as job titles, industries, and locations, and then cross-reference this data with industry-specific salary surveys or government data to estimate income ranges.
- If there is an access to LinkedIn Premium, we may use the "Premium Insights" features to get some information on salary ranges for specific job titles and locations.
- Consider using third-party data providers or APIs that offer income estimates based on various data sources.
- It is beneficial to consult with a data analyst or data scientist who can help develop a more sophisticated and accurate estimation model based on the available data.

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

→

- The most straight forward method is to manually search each company name on LinkedIn and look for the company's official page
- Using LinkedIn's own search functionality to look for company pages.
- An other option is web scraping.

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

→

- Job Posting: Checking job postings on company websites or job boards like LinkedIn or Indeed.
- GitHub: Searching for organization on GitHub and explore their repositories.
- LinkedIn: Using LinkedIn to search for employees or company pages of tech companies, and they may ask python as a skill or else they may have mentioned it on their profile.
- Online Tech Publications: Tech news websites and publications sometimes feature articles about technology stacks of different companies.
- Technology Blogs and Forums: Some companies or their employees may share information about their tech stack on technology blogs, forums, or social media.

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

→

- Authenticate the API request : To do this we'll have to use an access token.
- Obtain API permission: This can be done by creating a LinkedIn developer account & registering the applications.
- Send API Request: After obtaining the access token. We can send a POST request to the endpoint with the recipient's LinkedIn ID and the message text in the request body.