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
Ans :

- Use an API if the target website provides that. It's the most legal way.

- Increase wait time between scraping request, do not send mass request to the server.

- Change/rotate IP frequently.

- Change user agent, browser viewport size and fingerprint.

- Use third party solutions for captcha.

- Resolve the captcha by yourself, check the answer by Thomas Dondorf. Basically you need to wait for the captcha to appear on another browser, solve it from there. Third party solutions does this for you

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?

Ans :

Using Machine Learning to Predict Income Ranges:

Instead of going through profiles one by one, you can use computer programs that can learn from data. These programs, known as machine learning models, can look at information like the job title, industry, location, and experience of LinkedIn users to make educated guesses about their income.

Ways to collect data : web scraping , API access or data enrichment service

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

Ans : as an data scientist i will suggest **Web Scraping**: You can use web scraping tools or services to extract LinkedIn URLs for the company names in your list.

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

ans :

There are few ways to identify list of companies whose tech stack is built on python

One way is stackshare

Stackshare allows us to search for companies by name, industry, or technology stack. Once you have found a company that you are interested in, you can view their tech stack to see if Python is one of the languages that they use.

2)We can directly contact companies to ask if they use python and this can be done by email or reaching them out through social media

3)we can use job listing websites such as

glassdoor , linkedin, indeed etc
Or we can explore industry reports and surveys such as the Stack Overflow Developer Survey and GitHub Octoverse report.
Another way is to investigate the official websites of companies Navigate to their career or jobs section, and look for mentions of Python in job descriptions or technology-related pages.

Here are some companies who use python
1) GOOGLE
2)AMAZON
3)SURVEY MONKEY
4)UBER
5)FACEBOOK

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

Ans :
Sending a direct message to a person using the LinkedIn API requires obtaining the appropriate API permissions and making an API request to the LinkedIn API endpoint. The LinkedIn API provides the /v2/messages endpoint for sending messages.

Here is an example of how you can send a direct message to a person using the LinkedIn API with the /v2/messages endpoint:

- Obtain API permissions: To use the LinkedIn API, you'll need to obtain the appropriate API permissions. This can be done by creating a LinkedIn Developer Account and registering your application.

- Authenticate the API request: To authenticate the API request, you'll need to use an access token. The access token can be obtained by following the authentication process outlined in the LinkedIn API documentation.

- Send the API request: Once you have obtained the access token, you can send a POST request to the /v2/messages endpoint with the recipient's LinkedIn ID and the message text in the request body.

Here is an example of sending a direct message to a person using the LinkedIn API with the /v2/messages endpoint in Python:

- import requests

-
- # Replace with your access token
- access_token = "ACCESS_TOKEN"
-
- # Replace with the recipient's LinkedIn ID
- recipient_id = "RECIPIENT_ID"
-
- # Replace with your message text
- message_text = "Your message text here"

-
- headers = {
- "Authorization": f"Bearer {access_token}",
- "Content-Type": "application/json"
- }
-
- data = {
- "eventCreate": {
- "value": {
- "event": {

- "type": "MESSAGE",
- "message": {
- "text": message_text
- }
- },
- "recipients": {
- "values": [
- {
- "person": {
- "entityUrn": f"urn:li:person:{recipient_id}"

- }
- }
-]
- }
- }
- }
- }
-
- response = requests.post("https://api.linkedin.com/v2/messages", headers=headers, json=data)
-

- if response.status_code == 201:
- print("Message sent successfully")
- else:
- print("Failed to send message")
- print(response.json())

Another answer will be
Reading from the Mailbox

This API resource allows you to access the user's mailbox folder, filtered by type, and also retrieve a specific message from a folder.

GEThttps://api.linkedin.com/v1/people/~/mailbox

Here is the link for more documentation: https://developer.linkedin.com/docs/v1/communications/reading-members-mailbox