

ASSIGNMENT SUBMISSION

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

Ans.

Change Identities: Just like wearing different hats, the computer can pretend to be different people (change its identity) while collecting information.

Wait a Bit: Instead of rushing, the computer can take a little break between asking for information. It's like giving the website a breather.

Use a Special Tool: There are tools that can help solve these puzzles. It's like having a friend who's really good at puzzles to help out.

Be Smart About It: The computer can be clever in how it collects information, so it doesn't look suspicious. It's like being sneaky, but in a good way!

Be Patient: Sometimes, it's best to be patient and not rush. The computer can take its time, just like we do when we're doing something important.

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.

Here's a simple way to estimate the income range of LinkedIn profiles:

Use Available Information: Look at the profiles for any information they've shared about their job titles, industries, and companies they work for. This can give a rough idea.

Employment Sector: Based on the industry they work in, you can make a general estimate. For example, people in finance might have a different income range compared to those in education.

Location Matters: Salaries can vary depending on where a person lives. Consider the location mentioned in their profiles.

Job Titles: Certain job titles tend to have specific salary ranges. For instance, a "Manager" likely earns differently than an "Analyst".

Experience Level: More experienced professionals usually earn higher salaries. Check if they mention how many years of experience they have.

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

Ans.

Automated Search:

- Use a web scraping tool or a specialized program to automatically search for each company name on LinkedIn. This tool can navigate through the search results and extract LinkedIn company link

LinkedIn Search:

- Manually search for each company name on LinkedIn. Look for the official company page and copy the link.
- If possible, use LinkedIn's API (if available) to automate the search and retrieval process. This can save time compared to manual searching.

Outsource the Task:

- Consider outsourcing this task to a third-party service or a freelancer who specializes in data collection. They can use specialized tools to efficiently gather the LinkedIn links.

Check Company Websites:

- Visit the official websites of the companies. They often include social media links, including LinkedIn.

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

Ans:

1. GOOGLE
2. FACEBOOK
3. INSTAGRAM
4. SPOTIFY
5. NETFLIX