

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
  - Contact Administrator/Owner - In some cases, they may provide you with access or an API to collect the data you need without going through CAPTCHA.
  - Use CAPTCHA Solving Services - There are third-party CAPTCHA solving services and APIs available that can automate the process of solving CAPTCHAs.
  - Utilize Headless Browsers - Headless browsers, such as Puppeteer for Chrome or Selenium, can be used to automate browsing and interaction with websites.
  - Machine Learning and Image Recognition - Train a machine learning model to recognize and solve CAPTCHAs automatically.
  - Scraping at Off-Peak Times
  - Manual Captcha entry
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?
  - Job Title and Industry Analysis - Certain job titles or industries tend to have specific income ranges associated with them.
  - Company Size and Reputation - Look at the size and reputation of the companies listed in the profiles. Example, larger, well-known companies often pay higher salaries.
  - Location-Based Estimation Income levels vary significantly by region, so he can consider local cost of living and salary data when estimating income ranges.
  - Years of Experience - Analyze the number of years of experience mentioned in the profiles.
  - Educational Background - Consider the level of education mentioned in the profiles. Higher degrees and prestigious universities may be associated with higher incomes.
  - Machine Learning Models - He can train a machine learning model to predict income based on the available profile data. This would require labelled data for training, which can be challenging to obtain.
  - Data Mining and Web Scraping - If your client's LinkedIn profiles are public, you can scrape the profiles for relevant information.
  - Survey the Individuals - Create a survey and send it to the LinkedIn profiles, asking them to voluntarily provide their income range. This approach relies on people's willingness to respond.

3. We have a list of 1L company names, need to find LinkedIn company links of these profiles, how to go about this?
- Manual Search on LinkedIn - Enter the company name in the LinkedIn search bar and go through the results to find the official company page.
  - LinkedIn API - If you have access to LinkedIn's API (usually via the LinkedIn Marketing Developer Platform), you can automate the search process.
  - Google Search - Enter the company name followed by "LinkedIn" in the search query. This often yields LinkedIn company pages in the search results.
  - LinkedIn Company URL Structure - LinkedIn follows a specific URL structure for company pages. Format:  
<https://www.linkedin.com/company/company-name>
  - LinkedIn Company Directory - LinkedIn has a company directory where you can search for and filter companies by name, industry, location, and more.
  - Web Scraping - If you have a large list of company names, you can use web scraping techniques to automate the process of searching for LinkedIn company pages.
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
- Company Websites and Job Postings - Many companies mention the technologies they use on their websites or in their job postings.
  - LinkedIn and Professional Networks - LinkedIn can be a valuable resource for identifying companies that use Python. Look for python related job postings.
  - Open-Source Projects - Many companies contribute to or maintain open-source projects written in Python. Identifying these projects and their contributors can lead you to companies using Python.
  - Online Technology Forums and Communities - Technology forums like Stack Overflow, GitHub, and Reddit often have discussions about the technologies companies are using.
  - Tech News and Blogs - Tech news sites, blogs, and articles often discuss the technologies used by companies. Keep an eye on tech news related to Python and software development to discover companies that are publicly known for their Python usage.
    - Google
    - LinkedIn
    - Pinterest
    - Instagram
    - Spotify

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

- Messages API:

LinkedIn offers the Messages API allowing members to create messages to one or more first-degree connections or reply to existing conversations. The Compliance Events API provides options to monitor a given member's messaging activities.