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

**Manual CAPTCHA Solving**

**CAPTCHA Solving Services**

**Optical Character Recognition (OCR)**

**Machine Learning**

**Browser Automation Tools**

.

1. 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**: The job title and industry can give you a rough estimate of someone’s income. You can use salary data from websites like Glassdoor, Payscale.

**Company Size and Location**: The size of the company and the location can also impact income. Larger companies and companies in cities with a high cost of living often pay higher salaries.

**Experience and Education**: The level of experience and education can also be used to estimate income. People with more experience or higher levels of education tend to have higher incomes.

**Use of API**: Some APIs provide estimated income data based on job title, industry, and location. These could be used to automate the process.

**Machine Learning**: If there is a dataset with LinkedIn profile information and known incomes, you can train a machine learning model to predict income based on profile features.

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

This task includes the process of web scrapping, These general approaches can be considered.

**LinkedIn Search**: LinkedIn’s search function could be used to find the company profiles. You could automate this process by using a tool like Selenium to control a web browser and perform the search for each company name.

**Parse Search Results**: Once you have the search results, you’ll need to parse the HTML to extract the URLs of the company profiles. This could be done using a library like BeautifulSoup or newspaper3k in Python.

**Handle Pagination**: If there are many search results for a company name, you’ll need to handle pagination to go through all the results. This can also be done using Selenium.

**Data Storage**: As you extract the URLs, you’ll want to store them in a structured format for later use. This could be a CSV file, a database, or any other form of structured storage.

**Error Handling**: It’s important to include error handling in your script. This will ensure that your script doesn’t crash if it encounters an unexpected situation, like a CAPTCHA or a network error.

1. 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/Blogs**

**Job Postings**

**Tech Conferences and Meetups**

**Networking**

**Online Platforms**:- Stackshare and builtwith

**Business Directories**

**Companies:** Reddit, pinterest, Dropbox, Stripe, Quora

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

LinkedIn has its own APIs, that allows applications to send messages on behalf of users.

Messages can only be senton behalf of a member whose access token you have with the “w\_messages” member permission.

If linkedin messages are to be automatically sent, there are third party tools available can help to automate the process.