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**

As a project coordinator, I would suggest the following ways to solve the problem of hCAPTCHA when scraping 1 lakh website pages from a directory site:

Use a rotating proxy network. This will change your IP address with each request, making it more difficult for the website to detect that you are a bot.

Use a browser automation tool. These tools can automate the process of solving hCAPTCHA challenges, freeing up your time to focus on other tasks.

Use a cloud-based scraping service. These services have the necessary resources to solve hCAPTCHA challenges at a scale.

Contact the website owner and ask for permission to scrape their data. In some cases, website owners may be willing to grant permission to legitimate scrapers.

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?**

Consider the person's experience level. More experienced professionals typically earn higher salaries than less experienced professionals.

Consider the person's company size. Employees of larger companies typically earn higher salaries than employees of smaller companies.

Consider the person's location. Salaries can vary significantly depending on location. For example, professionals in major metropolitan areas typically earn higher salaries than professionals in rural areas.

Here are some specific steps:

1. Export the list of LinkedIn profiles to a CSV file.

2. Use a spreadsheet program to open the CSV file.

3. Create a new column in the spreadsheet and label it "Estimated Income Range."

4. For each profile, use LinkedIn Salary Insights or a third-party salary estimation tool to generate an estimated salary range.

5. Enter the estimated salary range in the "Estimated Income Range" column.

6. Once you have estimated the income range for each profile, you can use the spreadsheet program to calculate the average income range for the entire group.

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

   1. Manual Search

   2. Use LinkedIn Search tool.

   3. Use web Scrapping tools:

1. Choose a web scraping tool. There are several web scraping tools available, such as Scrapy and Beautiful Soup.

2. Write a script to extract LinkedIn company links from a list of company names. This script will need to use the LinkedIn API to search for each company name and extract the company link from the search results.

3. Run the script on your list of company names. This will generate a list of LinkedIn company links.

4. Save the list of LinkedIn company links to a file.

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**

   1. Search for companies that use Python on their websites.

   2. Look at job postings for python developers.

   3. Use of third-party service to identify companies that use python.

Companies that use python by my suggested approach:

   1. Google

   2. Instagram

   3. Dropbox

   4. Spotify

   5. Reddit

   6. Grass

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

There is no official Linkedin API for sending messages to other LinkedIn users. However, there are a few third-party APIs that you can use to do this. One option is to use the Phantom buster API