**Assignment**

Q1.A developer is assigned a task to scrape 1 lakh website pages from a directory site, while scrapping he is facing such a hcaptcha, which is placed to stop people from scrapping As a project Coordinator suggests ways to solve this problem.

The following techniques and recommendations can assist a project manager in assisting the developer in resolving this issue:

1. Review the terms of service on the website: Make sure that web scraping is not expressly prohibited by the website's robots.txt file or terms of service. If they do, you may need to get the owner of the website's permission or locate another source of the information.
2. If accessible, use official APIs: There are numerous websites that offer official APIs for data access. When obtaining the data without having to deal with CAPTCHAs, using these APIs can frequently offer a more dependable and organized method.
3. Rate Limit Your Requests: Set a rate limit for the website scraping requests you make. A website may get CAPTCHAs if it receives an excessive number of requests in a short period of time.
4. Rotate IP Addresses: To periodically switch your IP address, use a rotating proxy service. This can assist in getting over IP rate limitations and scraping activity detection.
5. Use Headless Browsers: To automate website interactions, use headless browsers like Puppeteer or Selenium. The ability of these programs to mimic human behaviour makes it more difficult for the website to identify automated scraping.
6. Automatically Fix CAPTCHAs: If required, you can automate the CAPTCHA-solving process by using third-party plugins or CAPTCHA-solving libraries or services like 2Captcha and Anti-Captcha.
7. Examine and Model User Conduct: Examine the way an average user navigates the website and attempts to imitate their actions. Before scraping, this can involve scrolling, clicking on links, and completing forms.
8. Disassemble the Task: Divide the effort into smaller, more doable portions rather than scraping 1 lakh pages all at once. This lessens the likelihood of triggering CAPTCHAs and decreases the intensity of your scraping activities.
9. Use Session Persistence: To make the website look more human, try to maintain a persistent session with it. This will allow you to preserve certain session data between queries.
10. Establish an Interval Between Requests: Add erratic intervals of time between your website requests. People don't visit websites consistently or click on links on them.
11. Check for Updates: Keep an eye out for any modifications to the website's architecture or anti-scraping protocols. Be ready to modify your scraping plan as necessary.
12. Legal Considerations: Make sure your scraping operations abide by laws, including those pertaining to copyright and data privacy.
13. Take Ethical Implications Into Account: Consider the moral ramifications of web scraping. Observe the terms of service on the website and refrain from scraping private or sensitive information.

But web scraping can be a complicated activity from a legal and moral standpoint. Always honour the intentions of the website owner, and be ready to locate other data sources in case scraping turns out to be too challenging or presents legal issues.

Q2. Our client has around 10k LinkedIn people profiles, he wants to know the estimated income range of these profiles. Suggest ways to do this?

To make an informed estimate, though, a business analyst can employ a variety of indirect techniques and data sources. Here's a methodical approach:

1. Collection of Data: Gather all of your LinkedIn profiles, preferably in an organized format (such as CSV or Excel) with essential details such as work titles, employers, and locations.
2. Make Use of Public Data: LinkedIn profiles frequently include details about work titles and names of companies. By examining industry- and location-specific wage data from publicly accessible sources like the Bureau of Labor Statistics (BLS), PayScale, Glassdoor, or LinkedIn Salary Insights, you can utilize this information to estimate income indirectly.
3. Segmentation: Divide the profiles into groups according to geography, job title, and industry. Your estimations might be more accurate the more detailed your segmentation is.
4. Remuneration Surveys: If available, make use of pay surveys and information tailored to your industry. The average wage ranges for particular job roles and industries are provided by these sources. Compare the information from these questionnaires to that from your LinkedIn profiles.
5. Insights on LinkedIn Salary (if relevant): With its wage Insights function, LinkedIn offers average wage data for particular job titles and regions. Even though it's not available for every profile, this information can be a useful tool for calculating income.
6. Educational Background: Take into account each person's educational background. Earnings may increase for those with advanced degrees or certificates in in-demand areas. Compare your wage information with the available statistics on educational attainment.
7. Experience Level: Take into account the number of years of experience listed on LinkedIn accounts. Higher salaries are usually earned by experts with more experience. To calculate income based on years of experience in a particular role, use pay data.
8. Location Information: Salaries and cost of living vary across different geographic areas. Utilize the location data from the profiles to calculate an income estimate based on the local cost of living.
9. Machine Learning Models: Using publicly available LinkedIn data and previously obtained wage information from other sources, think about developing a machine learning model to forecast income ranges. Although this can be a more sophisticated method, it might produce more precise outcomes.
10. Benchmarking: Examine your approximations against data that is readily available and industry benchmarks. This can increase accuracy and help you fine-tune your estimations.
11. Certifications and Error Margin: Be open and honest with the client about the limitations and possible margin of error in your income forecasts when presenting your findings. It's possible that LinkedIn profiles don't always fairly depict actual income.
12. Ethical and Privacy Considerations: When using LinkedIn data, make sure to respect ethical and privacy concerns. Steer clear of any behavior that can be interpreted as immoral or intrusive.

There isn't exact income information available on LinkedIn profiles, these estimations are only educated guesses at best and have limits.

Q3. We have a list of 1L company names, need to find LinkedIn company links to these profiles, how to go about this?

We can use both automated techniques and manual checks to discover LinkedIn corporate pages for a list of 1 lakh (100,000) firm names. Here's a methodical approach:

1. Collecting and Preparing Data: Make sure the list of firm names is in an organized file format, like an Excel or CSV file.
2. Search LinkedIn: LinkedIn offers a search function that lets you find company pages. But LinkedIn might restrict your search results, particularly if you don't have a paid account with LinkedIn Sales Navigator. You can still use the company names to conduct a basic search.
3. Export LinkedIn Data: You can export search results, including company profiles if you have access to a LinkedIn Sales Navigator subscription or a premium LinkedIn account. This can be useful for extracting LinkedIn URLs and automating the procedure.
4. Automated Tools for Web Scraping: To automate the process of looking for LinkedIn corporate pages, use web scraping tools such as Beautiful Soup, Selenium, or specialized LinkedIn scraping tools. A script that looks for each firm name extracts the LinkedIn URLs, and saves the results in a spreadsheet or database can be written.
5. Data Enrichment Services: Take into account utilizing services that can link LinkedIn profiles and corporate names. LinkedIn corporate page URLs can be included in company data enrichment services like Clear bit or ZoomInfo.
6. Verification by Hand: You might need to manually confirm that the LinkedIn company pages you found are linked to the correct businesses after the automatic procedure. While it can take some time, accuracy demands this procedure.
7. Enter your company's LinkedIn ID: LinkedIn frequently gives every company page a distinct corporate ID. It may be a more dependable method of locating LinkedIn pages if you can locate a source or service that offers these IDs for your list of company names.
8. If available, the LinkedIn API: You might be able to obtain corporate pages via the LinkedIn API if you have access to it. Nevertheless, adherence to LinkedIn's terms of service and API usage guidelines might be necessary for this.
9. Validation and Cleaning of Data: The collected LinkedIn URLs must be verified and cleaned. Certain URLs might direct users to the profiles of people who share the same name as the business, while others might be inaccurate. Make sure to check that the URLs are accurate.
10. Reporting and Documentation: Keep records of your procedure, including the procedures you took to verify manually and the techniques you used to access the LinkedIn company pages. Create a report using the LinkedIn company page list for your client.
11. Ethical Considerations: Make sure your data-gathering practices adhere to LinkedIn's ethical standards and terms of service. It can be delicate territory to scrape LinkedIn, so exercise caution with regard to legality and privacy.
12. If feasible, Bulk Verification: You might be able to bulk-verify the names of several companies with certain LinkedIn capabilities. Check out the latest changes and features on LinkedIn to see whether any of these alternatives are accessible.

Q4. 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.

List of businesses whose technology stacks are based on Python Working usually entails a mix of the following methods:

1. Job boards and postings: Look through job advertisements on firm employment pages or websites such as Indeed, and LinkedIn. Programming languages and technologies are frequently mentioned by employers in job descriptions. Seek employment posts that state clearly that Python is a necessary or desired skill.
2. Business Pages on LinkedIn: Look for corporate pages on LinkedIn and peruse the "Careers" or "About" sections. Businesses frequently make information about the technologies they employ available. See if the descriptions contain any references to Python.
3. Websites of Companies: Check out the official websites of the businesses that pique your interest. Seek out sections labeled "Our Stack," "About Us," or "Technology Stack." Certain businesses might disclose specifics about the technologies—including Python—that they employ.
4. Technology Articles and News: Examine news and articles on technology that highlight or profile businesses. These articles could go over the technology that businesses employ to provide their goods and services.
5. Business Listings on Expert Networks: Investigate professional networks such as Stack Overflow or GitHub, where businesses frequently exchange details about their open-source projects and tech stack. Look for Python-using corporate profiles and repositories.
6. Technology Forums and Blogs: Look for tech-related forums, blogs, and conversations about certain businesses. Employees and tech aficionados frequently exchange insights about the technologies they use.

These are the names of five businesses that are well-known for utilizing Python within their tech stacks.

1. Use Google: Google is renowned for using Python extensively in a wide range of projects and services, including YouTube and Google Search.
2. Facebook: Python has been used to create Instagram's backend, together with the Django web framework.
3. Showcase: Python is utilized in backend services and data processing, among other areas of Spotify's software stack.
4. Dropbox: Python played a major role in Dropbox's tech stack, especially in the early phases of the business's growth.
5. Reddit: Python is used by the well-known social networking site Reddit for a number of purposes, including as data analysis and web building.

Q5. Need to find an API, through which we can send LinkedIn messages to other LinkedIn users.

We must first establish an application and get an API key in order to utilize the LinkedIn API to send someone a direct message. We can use the API key to make authorized requests to the LinkedIn API once obtained it. The LinkedIn Developer Network website has the documentation for the LinkedIn API, which includes details on how to submit an application and get an API key. After obtaining the API key, we can send a direct message to an individual via the LinkedIn API by submitting a POST request to the relevant endpoint. Please remember that sending spam or unwanted messages over the LinkedIn API is prohibited by the terms of usage. Sending messages to those who have granted us permission should be the only use of the API.

LinkedIn provides its Sales Navigator users with access to the Messaging API. Users of LinkedIn Sales Navigator can combine their messages for lead management and communication with Sales Navigator thanks to this API. Usually available only to Sales Navigator subscribers, access to this API may be subject to additional terms and limitations.

It is imperative that you abide by LinkedIn's terms of use and API usage standards if you are a business analyst or working on behalf of a client and you need to automate messaging on the platform for valid reasons (such as sales prospecting or communication pertaining to a business opportunity). Spam and automated communications are strictly prohibited on LinkedIn, and violators risk having their accounts suspended.

It's important to think about other channels and approaches for communication, like tailoring your outreach, utilizing LinkedIn features like "InMail" within the confines of the platform, or investigating other CRM (Customer Relationship Management) programs that integrate with LinkedIn for lead management and messaging.

Make sure that everything you do on LinkedIn complies with their terms of service, and that your main goal when using the network is to create real, meaningful connections rather than sending out unsolicited or automated messages.

To send a direct message using the LinkedIn API, follow these steps:

1. Make an account as a LinkedIn Developer: If you haven't already, go to the LinkedIn Developer Network website (https://developer.linkedin.com/) and create a developer account.
2. Launch a new application: After obtaining a developer account, launch a new application by selecting the "Create Application" button in the "My Apps" area. Give your application all the information it needs, including a name, description, and website.
3. Acquire API credentials: An API key and secret will be sent to you following the creation of your application. To authenticate your API queries, you'll need these credentials.
4. Authenticate your application: The OAuth 2.0 authorization flow must be used to authenticate your application in order to access the LinkedIn API. In order to do this, users must be sent to the LinkedIn authorization endpoint and provide your application permission to access their LinkedIn data. An access token will be issued to you if authorization is successful.
5. Send a message using an API request: After obtaining an access token, you may use it to send messages via an API request on the authorized user's behalf. You must submit a POST request with the necessary parameters to the Messaging API endpoint in order to send a direct message.