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.**

Web scraping can be a challenging task, especially when websites have implemented measures to prevent scraping. To overcome this:

a. Respect Robots.txt: Check if the website's robots.txt file allows scraping of the content we need. If it does, ensure we are following the guidelines mentioned in the file.

b. Use Headless Browsers: Use headless browsers like Puppeteer (for Chrome) or Selenium to scrape the content. These tools can simulate user interaction, making it harder for websites to detect automated scraping.

c. User Agents and Proxies: Rotate user agents and use a pool of proxies to avoid IP bans. This helps in making your scraping requests appear more like regular user traffic.

d. Delay Between Requests: Add delays between your scraping requests to mimic human behaviour.

e. CAPTCHA Solving: Some websites may use CAPTCHAs to prevent scraping. Consider using CAPTCHA solving services or employing machine learning models to automatically solve them.

f. Session Management: Maintain a persistent session to handle cookies and login credentials, especially if you need to access authenticated content.

g. Monitor for Changes: Periodically check the website's structure and update your scraping script if the site changes.

2. **Our client has around 10k linkedin people profies, he wants to know the estimated income range of these profiles .Suggest ways on how to do this?**

Estimating income ranges from LinkedIn profiles can be challenging, as income information is typically not publicly available. However, you can make educated guesses based on the job titles, industries, and locations. Here's a basic approach:

a. Job Title and Industry Analysis: Analyze the job titles and industries of the LinkedIn profiles. Research industry-specific salary data to estimate income ranges for those roles.

b. Location Data: Consider the location of the profiles. In some regions, salaries may vary significantly. Use cost-of-living data for more accurate estimations.

c. Public Information: Some profiles may provide salary information. We can use this data if available, but it's often limited.

d. Machine Learning: Train a machine learning model using a labeled dataset of LinkedIn profiles with known income ranges. This can help predict income based on profile attributes

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

1.Use LinkedIn Search:

Start by manually searching for each company name on LinkedIn. This method can be time-consuming but is reliable. Here's how:

a. Log in to your LinkedIn account.

b. In the LinkedIn search bar, enter the company name and press "Enter."

c. Review the search results to find the official LinkedIn company page for each company.

2.LinkedIn Company URL Structure:

LinkedIn often follows a predictable URL structure for company pages. The URL format typically includes "linkedin.com/company/company-name." You can construct

these URLs based on the company names and see if they lead to the respective LinkedIn pages.

3.Use LinkedIn Company API (If Available):

LinkedIn provides APIs for company data, but access to these APIs may require permission and compliance with LinkedIn's terms of use. If you have access to the LinkedIn API, you can use it to search for company profiles based on their names.

4.Data Enrichment Services:

Consider using data enrichment services or data providers that offer company data, including LinkedIn company pages. These services can automate the process of finding company pages based on company names.

5.Scraping (Exercise Caution):

Web scraping can be an option, but it should be used with caution and in compliance with LinkedIn's terms of service. Automated scraping can lead to your LinkedIn account being restricted, so use this method with care.

6.LinkedIn Premium or Sales Navigator:

If you have a LinkedIn Premium or Sales Navigator account, you may have access to more advanced search features that can help you find company pages more efficiently.

7.Crowdsourcing or Outsourcing:

Consider crowdsourcing the task by hiring temporary workers or using crowdsourcing platforms to find LinkedIn company pages for your list of company names.

8.Cross-Reference with Other Data Sources:

Cross-reference the list of company names with other professional databases, such as Dun & Bradstreet, which may provide LinkedIn company page links.

When handling a list of 100,000 company names, automation and data enrichment services can be helpful. However, always ensure that you are compliant with LinkedIn's terms of use, privacy policies, and legal requirements when collecting data

from the platform. It's also important to respect the privacy and rights of the companies and individuals involved.

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

1.Job Listings and Company Career Pages:

Companies often list the technologies they use in their job listings and career pages. Look for job postings that mention Python or related technologies. You can explore the career pages of companies in the tech industry or companies known for Python development.

2.LinkedIn Profiles:

Examine LinkedIn profiles of employees working at various companies. Check for Python-related skills and job descriptions that mention Python. Companies with a significant number of employees with Python skills are likely Python users.

3.GitHub Repositories:

Search on GitHub for repositories associated with the company. If the company has open-source projects or contributions to Python-related projects, this is a strong indicator that Python is part of their tech stack.

4.Company Blogs and Technical Publications:

Companies often write about their technology choices and development processes in technical blogs or publications. Search for articles or blog posts where they discuss Python in their tech stack.

5.Technology Stacks on Company Websites:

Check the "Technology Stack" or "Our Tools" sections on the company's official website. Some companies list the technologies they use, including Python.

6.Tech News and Awards:

Look for news articles, awards, or recognitions that highlight a company's use of Python or Python-related technologies. Tech news outlets or industry-specific awards might provide this information.

7.Tech Conferences and Meetups:

Research if employees from the company have presented at Python-related conferences or if the company sponsors or participates in Python meetups or events.

Now, here are five well-known companies that are associated with Python based on the approaches mentioned:

1.Google: Google is known for its extensive use of Python in various projects, including Python libraries and tools like TensorFlow and Colab.

2.Facebook (Meta Platforms): Facebook has a strong Python presence, using it in backend development, data analysis, and machine learning.

3.Dropbox: Dropbox originally used Python for its server infrastructure and client applications.

4.Instagram (owned by Meta Platforms): Instagram's backend is known to be heavily powered by Python and the Django web framework.

5.Spotify: Spotify uses Python for various purposes, including backend services, data analysis, and machine learning.

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

LinkedIn does not officially provide a public API that allows you to send messages to other LinkedIn users programmatically. In the past, LinkedIn had an API that provided messaging capabilities, but it was limited in functionality and is no longer available for new developers.

LinkedIn has stringent policies when it comes to messaging to prevent spam and abuse. If you attempt to automate or send messages at scale without user consent, you risk having your LinkedIn account restricted or banned.

LinkedIn does offer a messaging feature for user-to-user communication directly through the platform. To send messages on LinkedIn, you can follow these steps:

Log in to your LinkedIn account.

Visit the profile of the person you want to message.

Click the "Message" button on their profile.

Compose your message and send it.

When sending messages on LinkedIn, ensure that you are sending relevant and personalized messages to connections or users with whom you have a legitimate reason to communicate. Sending unsolicited or spammy messages can harm your LinkedIn reputation.

If you're looking to send messages at scale for legitimate purposes, such as recruitment or sales outreach, LinkedIn offers the LinkedIn Sales Navigator platform. This premium service provides enhanced messaging and outreach capabilities. However, it's important to use Sales Navigator within LinkedIn's terms of service and best practices.