Title- Business Analyst Internship Assignment

Analystt.AI

Name- Vaishnavi Sanjay Jadhav

Contact- 7066954736

Mail-[vaishnavij77@outlook.com](mailto:vaishnavij77@outlook.com)

LinkedIn- <https://www.linkedin.com/in/vaishnavi-jadhav-bba28a272>

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 Investigating third-party services or tools that offer LinkedIn messaging capabilities while adhering to LinkedIn's terms and rules. Scrapping as a project Coordinator suggest ways to solve this problem

* When facing hcaptcha challenges while scraping the data, the goal is to do the work smoothly while staying in the boundaries,

1. I'd use a headless browser to mimic the way a person interacts with a website, making the scraping process appear more natural and less likely to trigger hcaptcha.
2. Polite scraping might also work as we are introducing delays between request, it will prevent overloading the website and potentially triggering the hcaptcha prompts.
3. 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?

* 1. Utilizing LinkedIn's advanced search filters, like LinkedIn Premium or LinkedIn Sales Navigator, to narrow down profiles by industry, and job titles to make more informed income estimates.

If we got our hands on LinkedIn Premium subscription, it provides enhanced search filters and access to more data, making it easier to find income-related information.

LinkedIn's own salary insights can provide salary estimates based on job title, location, and industry.

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

* 1. Exploring the availability of a LinkedIn Company API for automated data retrieval. If that’s not possible then carefully considering web scraping while respecting LinkedIn's terms of service.

2.Or we can simply go with manual search to identify company names but this way it might consume a lot of time.

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

* 1. The best way to that is to use LinkedIn job postings and simply look for those who have python mentioned in their skills, this way it will be a lot easier to identify companies which uses python.

2.Or another way we can use GitHub public repositories that include python related projects.

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

* 1. First we must check if LinkedIn’s provides APIs but we must be aware of restrictions and follow LinkedIn’s guidelines and policies

2. We can Invest in third-party services or tools that offer LinkedIn messaging capabilities while adhering to LinkedIn's terms and rules.

3.Or we can use automation tools