

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.

Answer:-

We can solve the issue by

- We can probably use the API's that maybe provided by the official site or you can use other API's that can help us get us through the scrapping without any captcha, this is a straight forward ethical way to approach a captcha problem such as this.

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?

Answer:-

There are few ways to conclude this task

One way is to do normal machine learning

- i) Prepare data – This data can be sourced by the external sources for reference
- ii) Clean the data
- iii) Create a machine learning model and use it on the cloned labelled data that can predict income ranges of each job from already trained data.
- iv) Validation of the model in terms of accuracy of the model
- v) Models like regression, decision trees, random forests, or neural networks can be used.

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

Answer:-

If there are 100000 company names web scrapping is the best way to get the data quickly, I have done some research and found out that selenium libraries can be a great way to do it and I am providing the code that I have found in the internet for this. I will upload the respected code in the GitHub.

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

Answer:-

We can identify the companies that use python as tech stack by few ways according to me such as

- **Visiting company websites** - going through the careers they are providing if they provide jobs such as data science, python developer, data analysis, business analyst and AI Engineer those companies must be using python as their tech stack.
- **Online technology articles and News** – These articles can literally provide which companies have been using python as tech stack and give us an idea on it.
- **Job Postings** – visit linkedin, indeed etc., and research about the company if they have posting such as python development then the company must be using python as their tech stack.

The 5 companies that I have found in my research are:-

1. Netflix
2. Google
3. Dropbox
4. Goldman Sachs
5. Paypal

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

Answer:-

As per my research I have found that **PhantomBuster** is a best Automated API tool that can be used to send LinkedIn messages.

