

# Assignment

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

Ans

## **1st approach:**

The most ethical approach is to contact the owner of the website and ask for permission to scrape the website .Some website might provide the data.

## **2nd approach:**

Explore data extraction libraries and tools that are designed to handle captcha

## **3rd approach:**

Training a machine learning model to solve it . It can be a time consuming and resource intensive option but it can be effective .

- 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?

Ans

## **Data Collection:**

Start extracting the basic information like Job titles ,industries and education etc

- Different Job titles and industries will have different ranges of salary .Like IT industries will have high salaries compared to Non-IT Industries.
- The level of education also affects the income of individuals .A degree holder might get high income compared to a 10th pass person .

## **Estimating the Salary:**

Some Standard API can be used to estimate the income based on above mentioned data .

## **Machine learning and Data Analysis:**

You can use machine learning model to predict income range based on various features extracted

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

Ans

- Using LinkedIn API we can search for companies and retrieve their details ,including company's LinkedIn URL.
- Write Web scraping script that takes the list of company names and searches LinkedIn profiles.
- LinkedIn Sales Navigator is a paid tool that provides advanced company search .

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

Ans:

**LinkedIn Profile:**

Usually companies prefer all its employees to have linkedIn profile .By searching for employees' profiles with python as a core skill will help to know that a particular company is dealing with python .

**Job Listing:**

While Listing for a job ,companies usually mention the technology they are dealing with in their job description and also search for employees with those specific skills .

**Github:**

Company hosts their open source project on Github . Companies with python project can be segregated and listed.

Some of Companies listed out using this methods are

Netflix ,spotify, Instagram hire employees strong in python

Google hosts their open source project on Github

Facebook employees have python skills in their linkedIn profile

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

Ans

LinkedInAPI do provide “ /v2/messages /” to send linkedin messages to other linkedin users but we need to get permission from LinkedIn for using this API .