BA Analyst
Assignment

By Riyajul Mansuri

Starts from here →

About me

Hello, I'm Riyajul Mansuri. In 2022, I successfully graduated with a bachelor's in Computer Science, and I've always had a passion for exploring and understanding the world around us. More recently, I successfully completed my data analyst internship, gaining valuable insights into identifying issues that can be addressed through data-driven solutions. The process involves recognizing the problem, gathering, refining, processing, and scrutinizing data to find meaningful solutions.

Some of the key tools I've mastered include :
Microsoft Excel
Tableau
Python (Basic)
SQL
Microsoft PowerPoint (Report And Presentation)

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.

As a project coordinator, overcoming HCAPTCHA challenges in web scraping can be tricky. Here are some tips:
- First, check if the website has a special way (API) to get the data you need. It's usually the best and most ethical way.
- Try using lots of different internet addresses (IPs). This can make it harder for the website to spot scraping.
- Use tools that act like a person on the website, not a machine. This can fool the website's security.
- Some services can solve CAPTCHAs for you. Be careful, though, as this might break the website's rules.
- Slow down your requests. Fast ones often trigger CAPTCHAs.
- Make your web browser look like different ones. Websites use this to tell if you're scraping.
- Act like a regular website user with cookies and stuff.
- Instead of scraping everything, focus on just the parts you want. This reduces CAPTCHAs.
- Always follow the website's rules and the law when scraping. Don't scrape private or sensitive data.
- Try using online scraping tools. They might help with CAPTCHAs.
- If you can, ask the website owner for permission to scrape. They might help you get the data you want without CAPTCHAs.

Remember to be ethical and follow the website's rules. If they don't allow scraping, look for other ways to get the data you need.
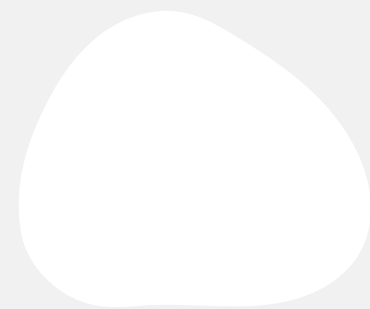
2. Our client has around 10k LinkedIn people profiles, he wants to know the estimated income range of these profiles. Suggest ways on how to do this.

To figure out how much money people from LinkedIn profiles make, we have a plan:

- First, we'll get information from the LinkedIn profiles, like what jobs they have, where they work, and their education.
- We'll also mix in data from other places, like how much money people usually make in certain jobs or areas.
- We'll get help from computers that can learn from the data. They'll guess how much people might make based on their jobs and other information.
- We'll do some math to find patterns in the data. For example, we'll see if having more education means more money or if living in different places changes income.
- If we can, we'll talk to some of the LinkedIn profiles and ask them how much they earn. But not everyone might want to share this.
- We'll always follow the rules and be respectful. We know these are just guesses and may not be exact.

In the end, we'll tell our client a range of how much these people might earn. But remember, these are just estimates, and we can't be 100% sure because we have some limitations with the data and privacy rules. We promise to treat everyone's information with care and follow the law and what's right.

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

To find LinkedIn pages for 100,000 companies, we can do these things:
- We can make computers look for LinkedIn pages by using special tools. These tools can search for each company name on LinkedIn and find the LinkedIn page link.
- LinkedIn has a special way for programs to talk to it, called an API. We can use this API to look for company profiles on LinkedIn. However, using the API has some rules and needs special permission.
- Some services can help us match company names to LinkedIn pages. They have big databases of company info and can give us LinkedIn links.
- If the other ways don't work well, we might need to look for LinkedIn pages one by one. We'll do this by searching for each company name on LinkedIn and writing down the LinkedIn link.
- Before we start, we need to make sure our list of company names is clean and neat. This means we remove any duplicate names, fix spelling mistakes, and use the same naming style.
- Think about how we will handle a big list of 100,000 names. The way we choose should be good for handling lots of names.
- When we get data from LinkedIn or other services, we must follow LinkedIn's rules and the law. We also need to keep people's data safe and follow good ethics.
- Before doing the big search, we'll test our method with a smaller group of names. This way, we make sure it works well and is accurate.
- After we find the LinkedIn links, we need a good system to save and use the data.

This work might need a mix of computer and human effort, depending on how good the data is. We have to make sure our work is fast but also right. And we must always follow LinkedIn's rules and respect people's privacy.

→

4. 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.

We can find companies that use Python in their technology by doing these things:
- We collect information from places like company websites, job ads, and profiles on websites like LinkedIn.
- We look for specific words like "Python developer" or "Python programming" on these websites. These words tell us if a company uses Python.
- We use computer programs to read and understand the information we collect.
- We look for words that tell us a company uses Python tools like Django, Flask, NumPy, or TensorFlow.
- Make sure the data is right. We double-check the information to make sure it's correct.
- We use smart computers that can learn to find companies using Python. These computers learn from data we already know.
- We also use information from other places, like tech job boards and business directories.
- If we can, we ask people to check the information to be more sure it's right.
- Make the List: Finally, we make a list of companies that use Python in their technology.

By My Approach Here are five companies who use Python:
1. Google
2. Facebook
3. Instagram (owned by Meta Platforms, Inc.)
4. Quora
5. Spotify

But remember, the information might not be 100% perfect. It depends on how good our data is, and technology can change. So, we always need to check if it's right.

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

Sending messages on LinkedIn through a special computer program (API) has some rules, and it might not always work for everything. LinkedIn only allowed the use of this special program for specific things, not for regular chatting.

If you want to send LinkedIn messages using a program, here's what you might need to do:

- Go to LinkedIn's special website for developers and see if they changed their rules about messaging. Sometimes they allow more things or new features.
- If LinkedIn lets you use their messaging program, you might have to ask for permission. They look at each request and decide if they'll allow it
- When you use LinkedIn's messaging program, you have to do what LinkedIn says in its rules. This includes keeping people's information private and not sending lots of unwanted or automatic messages.
- LinkedIn might give you special tools to use with their messaging program. These tools make it easier to send messages while following LinkedIn's rules.
- If LinkedIn doesn't let you use their messaging program, you can try different ways to send messages. This might involve using special programs to look at LinkedIn (be careful), using tools that work like a person on a web browser, or trying outside programs (but they might have rules, too).

Just know that LinkedIn's rules and what you can do can change, so always check their special developer website and their rules. And make sure you respect people's privacy and follow the law when sending messages.

Thank You For Giving Me this Opportunity