Growth For Impact Data & Tech Internship Assignment

Thank you for sharing your interest in this role. Here are the key details you would need to complete this assignment.

Link to the job description- here

Data assignment- raw data here

Short quiz (2 questions) & submission link platform- submit here

An advice- AI won't be able to help you do this assignment (100s have tried and failed, so please don't). But if you invest a little time learning web scraping, it will be quite easy for and beyond the internship. Resources to learn web scraping or getting help when stuck- link1, link2

Details about the data assignment-

In the sheet shared above, you will find a few details of ~350 companies. It's from an unclean data source with several details missing. Here's what needs to be done-

1. Create a copy of the spreadsheet shared with you and for the next steps start making changes there.

2. Finding the careers page- using different web scraping libraries and coding skills- find link to their careers page.

3. Find the jobs listings page URL. It could be the same as the careers page, though in several cases the careers page and the actual job listings page are usually different.
   For example-
   Hannah Solar's website domain- https://hannahsolar.com
   Their careers page- https://hannahsolar.com/about-us/careers/
   And their jobs listings page- https://hannahsolar.zohorecruit.com/jobs/Careers

4. Now once you have identified the jobs listings page (not just the careers page), scrape back all the details you can find for the job posts listed there. Key details include- job posting URLs, job titles, job's location, job's category, date of job posting.
   Hint- each jobs listings page could have anywhere between 0 to 200 job postings. For this assignment you only need to return back details of up to 3 job postings (ideally the most recent ones) from any single website.

5. Arrange your findings in an .xlsx file (copy of the original sheet shared above) and submit the form here- submission link . Please note- you need to **fill both** the subsheets/tabs- Data and Methodology for the submission to be considered. If you miss any one of those, your assignment

would be rejected.

6. Please verify the validity of your submission. Do random checks to see if the links you shared are actually working.

Key details-

Remember you don't need to find each and every job across all the websites. The moment you reach a total of 500 job postings, you can submit the assignment.

Here's how the math works-
It's not expected that you will be able to find details for all 350 companies (that number could be anywhere between 200-250). Furthermore, it's possible that only 60-80% of them would have a careers page, and a small percentage of them have posted any job there.

It's likely that most of these websites use job posting services like- Lever, zoho recruit, glassdoor or any other 3rd party provider. On average it's assumed the companies which have a job opening will have between 4-10 job openings. So, at an aggregate level there might be ~1000+ jobs. You don't need to find all of them, just any 500 are sufficient for this exercise.

Hint- Identifying these common job posting service providers (Lever, Zoho recruit...)[A] will help you concentrate your energy and effort.

Bonus hint- using libraries like beautiful soup, scrapy, scrapling could help further

Though strict warning- if you submit your assignment without actually cleaning up the data output given by AI agents, or without sharing methodology, the assignment will be auto-rejected. And yes, it's expected that you wouldn't do the assignment manually.