

Python Task

Greetings from Credicxo,

Congratulations, You have been shortlisted for the coding round.

The details of the test are mentioned below . Kindly read all the instructions and reply with submissions on this email **ONLY**.

Technologies : *Python latest, BS4/Scrapy/Selenium, requests.*

MAIN TASK

In this task we want you to scrape a minimum hundred URLs.

The URL will be in format of "<https://www.amazon.{country}/dp/{asin}>". The country code and Asin parameters are in the CSV file <https://docs.google.com/spreadsheets/d/1BZSPhk1LDrx8ytywMHWVpCqbm8URT xTJrIRkD7PnGTM/edit?usp=sharing>. The CSV file contains 1000 rows.

Use Selenium or bs4 to Scrape the following details from the page.

1. Product Title
2. Product Image URL
3. Price of the Product
4. Product Details

If any URL throws Error 404 then print the {URL} not available and skip that URL.

After completing each round of hundred URLs **mention the total time it took to complete them in the feedback section of task submission page.**

The output should be in the list of dictionaries, finally represented in JSON.

Upload only your code file, resultant JSON file and a proper readme explaining how you approached the problem.

BONUS 1: If possible, add the google Colab button in your repository. To directly access your code in Google Colab.

BONUS 2: If possible, try connecting some Database like MySQL or PostgreSQL and dump the data in the database. (You still need to create the output JSON file.)

BONUS TASK

This task is not mandatory but if you are able to do it then your chances of getting in would be increased significantly.

Technologies: Python latest, BS4/Scrapy/Selenium, requests.

In this task, you need to write a script to bypass Amazon Captcha,

Example website: <https://www.amazon.com/errors/validateCaptcha>

1. Write a script that solves the captcha and submit the form.
2. Also explain how did you approach this problem to find a solution.

Upload this task in the same repository where you upload your main task but inside the folder Bonus Task.

Note: You can't use the existing captcha solving services API.

You must remember the following points while submitting, otherwise your submission may be rejected.

Submit your task at - <https://forms.gle/3NXcax4TYjb6mTxD7>

1. Upload Project Github link in "Task Link" Section.
2. Add Your Name and contact number in the form.
3. You can add comments (additional information you want us to know in your submission) in the "Feedback/Instructions to use" Section.
4. Please attach your resume.

Deadline to submit is **two days** from the date on which you received this task.

Best Regards,

HR Team

Credicxo Tech Pvt. Ltd.