Software Developer

Case Study

Hello,

Dear candidate, in this project you will build a simple scraper for one of our customer.

Description of Assignment

As you know Analyticahouse is a consultant on digital marketing area. We have bunch of customers taking consulting services from us. (which are leading brands in their sector 😎)

Lets say one of our customers demanded a Product Report about it's products. Your mission is to build a system that do scraping, writing and e-mailing.
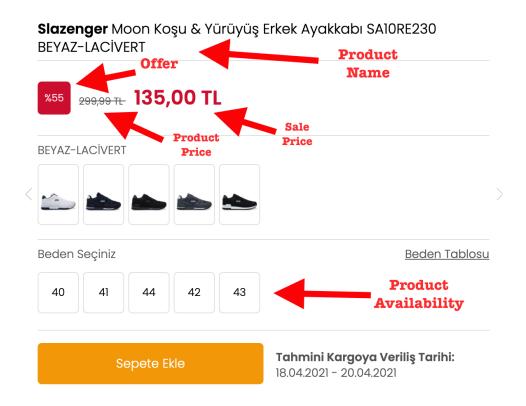
This mission has three phase:

1. Scrape necessary metrics from product pages
2. Write this data to Google Sheets
3. Using Google Apps Script : Sort the products by availability value send this data with an e-mail to "okan@analyticahouse.com" and write your name to the subject area of e-mail.
4. Bonus

1. Scrape necessary metrics from product pages

Your mission is build a crawling system which will go to those URLs (There is an xlsx file at the attachment), find out which urls are pointing product page(such as this https://www.markastok.com/buratti-suya-ve-ruzgara-dayanikli-kapusonlu-softshell-erkek-mont-2011055-indigo) , scrape necessary metrics and write to Excel as separate columns. At the below picture all metrics can be seen. Note: You should append **www.markastok.com.tr** to every path for reaching page.
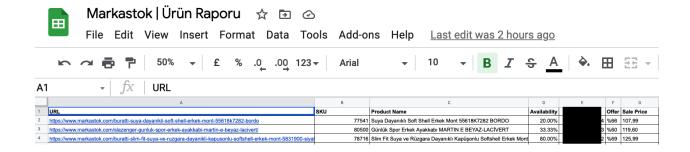
2. Write this data to Google Sheets

For the second mission you should write this data to Google Sheet programmatically with your gmail user.

Google has comprehensive documentation about it therefore you should look here: https://developers.google.com/sheets/api

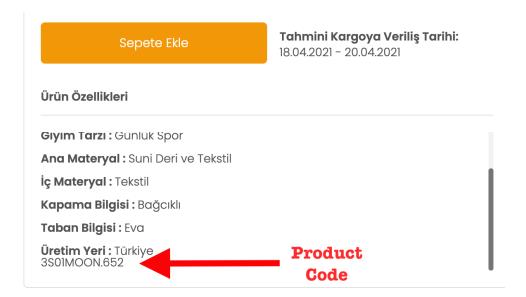At the end we expect you to create a report like this :

3. Sorting and Sending data with an Email

For the final mission, you should sort the products by availability value and send this data to "okan@analyticahouse.com" using Google Apps Script. You should not use the UI of Sheets, all the work should be done by Google Apps Script and of course with plain javascript. You should upload your Javascript code to Github as well.

Again Google has dedicated documentation about it so you should check it! https://developers.google.com/apps-script

4. Bonus

Every product has product code in it's page. If you manage to scrape this value and put it as a separate column, you will gain additional points. Below there is a picture indicating Product Code.

Additional Questions

1. If I'd have 10.000 urls that I should visit, then it takes hours to finish. What can we make to fasten this process?
2. What can we make or use to automate this process to run once a day?Write your recommendations.
3. Please briefly explain what an API is and how it works.

Sending the project

To demonstrate your work, you should follow these steps.

1. Use git and upload your source code(Python, Javascript) onto your GitHub page as a public repository. It would be nice if we don't see `Add files via upload` as a commit description :)
2. Add a README.md file with the project description into your repo. Project description should include following items.
   a. URL of the Report.
   b. Used tech-stack (library/framework etc.)
   c. A brief description of the challenges you face.
   d. What did you learn from this project?
   e. Answers of the additional questions.

That's it. Good Luck and Take Care! 🍀