Data Scientist Task

Input data

InputData.xlsx contains the required input data to complete Task #1 and #2. Candidates can download the input CSV file here (note that by clicking on the previous link, the InoutData.xlsx will be downloaded automatically).

TASK #1

InputData.xlsx, worksheet *Companies* file contains a list of 100 companies: name of the company, website and unique identifier (*client_id*).

The candidate shall create a Python script that extracts the following information from each company website:

- Emails: the candidate shall use regex, external python libraries are not allowed;
- Phone numbers;
- Short description of the company activities in English (e.g. The largest AI-powered database of green startups in Europe);
- URL to:
    - Contact pages;
    - Legal pages (both *Terms and Conditions* and *Privacy Policy*);
    - About us pages (this page can also be under *Mission*, *Who we are* and similar)

Consequently the candidate, using an Excel file, shall create a table structure, mimicking a SQL database, and fill the table(s) with parsed data.

Task #2

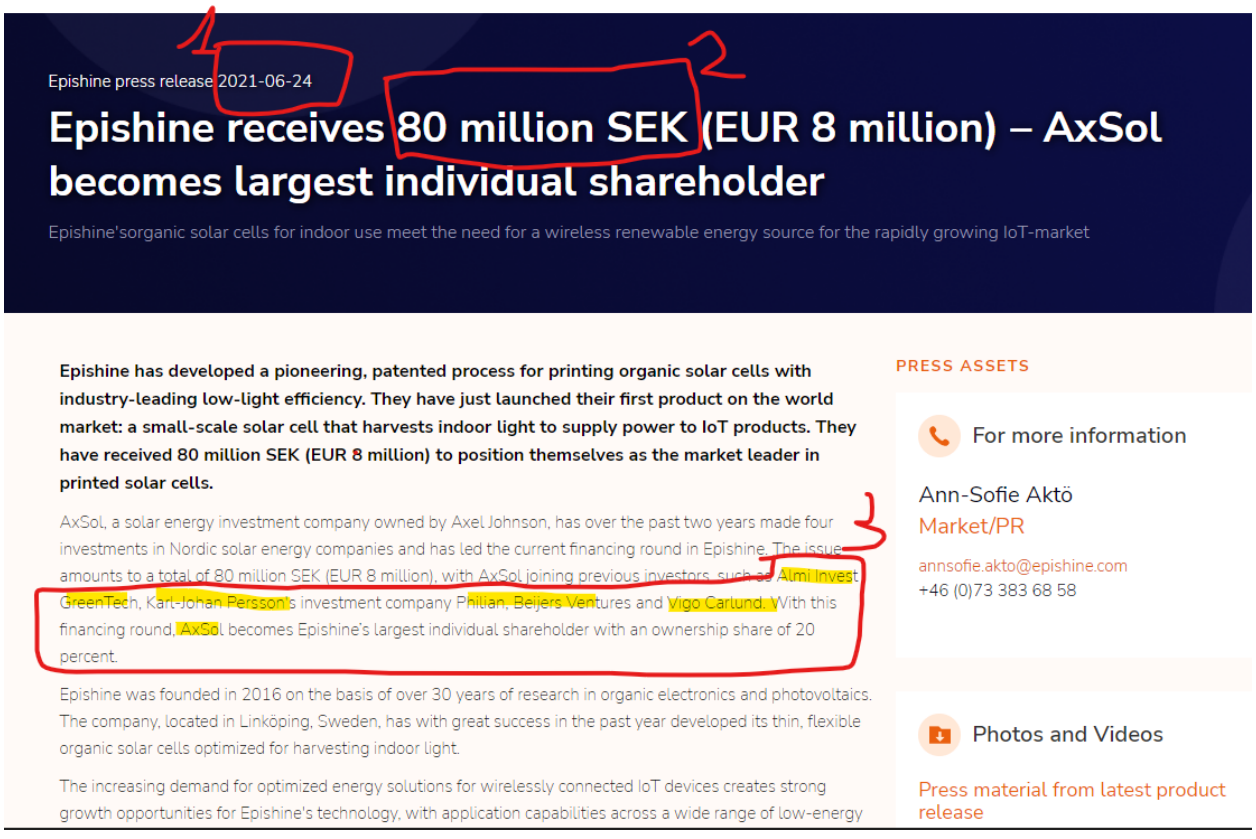
Before describing the second task, a brief introduction about funding rounds is needed.

What are funding rounds?
Funding rounds provide outside investors the opportunity to invest cash in a growing company in exchange for equit. Note that loans/debts are considered funding rounds too. Read more here.

A funding round is a collection of information containing:

1. Date of the investment;
2. Amount invested;
3. Investors taking part in the round.

Please see below the image - the numbers represent a mapping with the list above (link to the full article here).



Epishine press release 2021-06-24

Epishine receives 80 million SEK (EUR 8 million) – AxSol becomes largest individual shareholder

Epishine'sorganic solar cells for indoor use meet the need for a wireless renewable energy source for the rapidly growing IoT-market

Epishine has developed a pioneering, patented process for printing organic solar cells with industry-leading low-light efficiency. They have just launched their first product on the world market: a small-scale solar cell that harvests indoor light to supply power to IoT products. They have received 80 million SEK (EUR 8 million) to position themselves as the market leader in printed solar cells.

AxSol, a solar energy investment company owned by Axel Johnson, has over the past two years made four investments in Nordic solar energy companies and has led the current financing round in Epishine. The issue amounts to a total of 80 million SEK (EUR 8 million), with AxSol joining previous investors, such as Almi Invest GreenTech, Karl-Johan Persson's investment company Philian, Beijers Ventures and Vigo Carlund. With this financing round, AxSol becomes Epishine's largest individual shareholder with an ownership share of 20 percent.

Epishine was founded in 2016 on the basis of over 30 years of research in organic electronics and photovoltaics. The company, located in Linköping, Sweden, has with great success in the past year developed its thin, flexible organic solar cells optimized for harvesting indoor light.

The increasing demand for optimized energy solutions for wirelessly connected IoT devices creates strong growth opportunities for Epishine's technology, with application capabilities across a wide range of low-energy

PRESS ASSETS

For more information

Ann-Sofie Aktö
Market/PR

annsofie.akto@epishine.com
+46 (0)73 383 68 58

Photos and Videos

Press material from latest product release

Note that the Funding Round Date corresponds to the date of the news, i.e. the date when the news was published.

InputData.xlsx, worksheet *Fround* contains a list of 50 articles dealing with the founding round. In the worksheet, the candidate will find the name of the organization, the website, the link to news covering the funding round of the company and the company unique identifier (*client_id*).
The candidate shall create a Python script to to extract the below information for each of the news_url listed:

1. The amount of the round as float (preferably denominated in original currency - in the case above SEK);
2. The date of the round as Timestamp - the article publication date;
3. The investors taking part in the round.

Please note that point 3 is OPTIONAL.

Delivery

1. The candidate must save every variable that takes more than 5 minutes to run (e.g.: data cleaning: if this procedure takes more than 5 minutes, the candidate must save the cleansed database);

2. The code should be tidy and easy to read; please, insert detailed comments to explain what the code is doing.

3. The candidate should deliver a code written in Python along with all the stored variables as explained in point 1.

4. **The candidate shall upload all the relevant documents and scripts in a public GitHub repo (or similar).**