Before you begin:

Chattermill is committed to eliminating bias in our recruitment process and providing equal opportunities to all of our candidates. To help us in this effort, we ask that any files you submit for this task do not include your name or any personal details. Please be vigilant not to include your name in any file/folder names or in any filepath written inside your code.

Thank you for helping us create a diverse team!

The Goal

We would like you to demonstrate your coding knowledge, ability to learn new tools, and skills in simple data manipulation.

This data task uses the Guardian API. Whilst it is not necessary that you have prior experience working with APIs, they are an important aspect of this role, and so we would like to see an attempt to pull data from an API into a CSV. There are many good resources online on how to do this. We like to use Python and the requests library, but you may have your own preference.

Please note: If you feel you can demonstrate the skills required for this task using another free API, please go ahead and do so. However, you will not be marked down for continuing with the Guardian API.

The API:

You can get information about the Guardian API here:
https://open-platform.theguardian.com/documentation/

The API documentation is not perfect, and you may find some parts to be unclear. This is typical of many APIs we work with, so it will be up to you to understand the documentation and experiment with the API.

Please note: For this exercise, please do not use the Client libraries listed at the bottom of the documentation.

Instructions

1. Sign up for a Developer API Key using this link:
   - https://open-platform.theguardian.com/access/
2. Use the Content endpoint to return all articles related to elections or Brexit
3. From the results, pull the following information into a CSV:

- ○ id
- ○ type
- ○ sectionId
- ○ sectionName
- ○ webPublicationDate
- ○ webTitle
- ○ webUrl
- ○ apiUrl
- ○ isHosted
- ○ pillarId
- ○ pillarName
- ○ Wordcount

4. Make sure the results are in descending order based on webPublicationDate
5. Make the following changes to the CSV
   a. Format the dates into day/month/year format and add to a new column named formatted_date
   b. Extract the year from webPublicationDate into a new column name year
   c. Exclude any articles with wordcount < 1000 from the CSV

Expected Output:
You should produce the following in a ZIP archive:
- A CSV file containing the data outlined above
- File(s) containing your code (e.g. a .py or .js file)

Notes:
- The task should take no more than two hours to complete
- Your code should be scalable and work for general queries to the Content endpoint (for example, your code should work for any number of API results)
- Please follow best coding practices to the best of your ability
- Be mindful of API rate limits

Thank you for taking the time to complete these tasks. Your time and effort are greatly appreciated. We're also excited to discuss in more detail your responses during the next stage in the interview process.

If you have any questions about the task, please feel free to reach out to myself at steve@chattermill.io