# Bluevine Case Study

## The Task

In this assignment, you will explore data from some top-ranked sci-fi and fantasy authors' books.

You must use **`python`** to implement your solution. We strongly suggest using **`requests`** for the data acquisition part and **`pandas`** for the data analysis part of the assignment.

We are interested to see your entire process and not just the end result. So please make sure to document all of the relevant work that you do using comments or any other method of your choosing.

This task is intended to be an opportunity for you to showcase both your skills and your general approach to data analysis and engineering problems. You will need to make some decisions and assumptions as you solve the assignment, make sure to document these as well.

## Allotted Time

The assignment was designed to take no more than 4 hours to complete for people familiar with python and pandas. You should return your final products within 2 days of receiving this assignment.

## The Dataset

You should use the Open Library Books API (no registration required) to fetch data for all books with ISBN numbers listed in the **books-isbns.txt** file. Since not all ISBNs are available in the Open Library archives and some have partial data only, there will be entries missing information and some won't exist at all, which is fine. As fetching the data takes some time, it is strongly recommended that you save the fetched data to your computer so you can reuse it from the local file system while developing/running your solution. In addition, please consider adding a timeout to your API calls so as not to hang for long when no response is received - depending on your internet connection speed, a reasonable timeout could be 1-2 seconds.

ISBNs are essentially book identifiers and different editions of a book will have different ids. So the same book (text) can be represented by more than one ISBN. We consider books with different ISBN identifiers but with the **same title** value to be **the same book**.

## The Questions

1. How many different books are in the list?
2. What is the book with the most number of different ISBNs?
3. How many books don't have a goodreads id?
4. How many books have more than one author?
5. What is the number of books published per publisher?
6. What is the median number of pages for books in this list?
7. What is the month with the most number of published books?
8. What is/are the longest word/s that appear/s either in a book's description or in the first sentence of a book? In which book (title) it appears?
9. What was the last book published in the list?
10. What is the year of the most updated entry in the list?
11. What is the title of the second published book for the author with the highest number of different titles in the list?
12. What is the pair of (publisher, author) with the highest number of books published?

If you can't answer one of the questions, please submit your partial answer and move on to the next one. There will be time to discuss any issues that arose in the followup meeting.

## What to Submit

Please make sure that your code is clear, clean and as readable as possible, following common python best practices.

You should submit a ZIP file containing:

1. A python file called **solution.py**, which can be run and will output the answers to all the questions, printed as logs. Please provide any additional text files or code files written by you if required to run your solution.
2. A text file called **instructions.txt** with instructions on how to run your code.
3. A text file called **answers.txt** file with all the printed logs received when running the **solution.py** file.
4. [Optional] If you worked with a Jupyter/iPython/Google collab notebook and want to share it as well, please feel free to add it to the ZIP file. This is not required and you still must provide all required files previously mentioned (1-3).