**Word Frequency in Moby Dick**

Use web scraping and NLP to find the most frequent words in Herman Melville's novel, Moby Dick.

#### Project Description

In this Project, you'll scrape the novel Moby Dick from the website Project Gutenberg (which contains a large corpus of books) using the Python package requests. Then you'll extract words from this web data using BeautifulSoup. Finally, we'll dive into analyzing the distribution of words using the Natural Language ToolKit (nltk). The natural language processing tools used here apply to much of the data that data scientists encounter as a vast proportion of the world's data is unstructured data and includes a great deal of text.

To complete this Project, you need to know how to import web data into python and how to work with natural language text. Before starting this project we recommend that you have completed the following courses:

* [Importing Data in Python (Part 2)](https://www.datacamp.com/courses/importing-data-in-python-part-2).
* [Natural Language Processing Fundamentals in Python](https://www.datacamp.com/courses/nlp-fundamentals-in-python).

This Project is based on a live screencast by DataCamp's own [Hugo Bowne-Anderson](https://twitter.com/hugobowne). When you've finished the Project, or if you get stuck, do check out [the screencast with Hugo's solution](https://www.facebook.com/726282547396228/videos/1762301357127670/) (the screencast starts 12 minutes into the video). You can also find Hugo's solution notebook [here](https://github.com/datacamp/datacamp_facebook_live_nlp/blob/master/NLP_FB_live_coding_soln_verbose.ipynb).

#### Project Tasks

* 1 Tools for text processing
* 2 Request Moby Dick
* 3 Get the text from the HTML
* 4 Extract the words
* 5 Make the words lowercase
* 6 Load in stop words
* 7 Remove stop words in Moby Dick
* 8 We have the answer
* 9 The most common word