**> Introduction**

The data your department will be receiving has been obtained from scanned documents, which have been processed through some OCR (Optical Character Recognition) software. The software clearly has some issues: the resulting text has inconsistent capitalisation and spacing. It occasionally puts semi-colons between sentences but their use may not be appropriate.

You must write code to format text according to the following rules:

* Sentences should start with a capital letter.
* Sentences should be separated by two spaces.
* Commas should be followed by a single space.
* Semi-colons should be replaced with full stops.
* Correct text should not include duplication of punctuation.

Your code will eventually form part of a data cleansing library, so you should make proper use of functions to organise it. For testing purposes, you may hard code the text samples into your application. Any programmatical solution in Python may be utilised.

In addition, you will need to write a lookup list into your code so that you can check for proper names and ensure that they are capitalised appropriately.

**> Output:**

* Program code in Python
* Evidence of cleansed data[Preview the document](https://canvas.qa.com/courses/350/files/107877/download?wrap=1)

**> Skills required:**

* Python functions
* String operations
* User-defined functions