# Data Engineering Evaluation Questions.

As a data engineer, you have been tasked with writing code to analyze a text. The questions below are designed to test your development skills and approach to solving a problem.

Somethings to note:

1. Punctuation such as “ “ ‘ ‘ . , ; \t \n ? – should be excluded from the analysis
2. Analysis should be case insensitive. E.g. “Boy” should be treated the same as “boy”
3. Use text in attached ‘sample.txt’ file to test your code
4. All analyses of your code should describe space and running times
5. For each question, provide code and analysis of the code
6. This task is designed to be accomplished with approximately 2 hours of work. Unlike a real project, this does not have a formal deadline, however, significant delays may disqualify your consideration.

## Question 1: word count

Write code in your favorite programming language to output count of each word found in ‘sample.txt’ file. The count should represent the number of times a word appears in the text. Your code can implement a brute force approach to count the words

* Input: String or file path to ‘sample.txt’
* Output: words and their count

In a few words, give an analysis of your code

## Question 2: improvements

Is there an algorithm and data structure you can implement to improve your code in question 1? If so, write a new code which applies this algorithm and/or data structure to counting words in the provided text.

In a few words, give an analysis of your code and describe how this new algorithm and/or data structure improves your approach in Q1.

## Question 3: highest count

Write code that takes the output from Q1 or Q2 as input and outputs the word with the highest count.

If a few words, give an analysis of your code and describe any algorithm and/or data structure you used.

## Question 4: unit tests

Write unit tests for your code to assert the following:

* Null and empty inputs
* Invalid input
* Valid input with valid output