CS-150 Worksheet 3 Release date: 22/10/2019

This worksheet is about getting familiar with compression methods and applying them to different pieces of data. **Remember that whitespace characters count in the calculation of the message length!**

☐ Task 3.1

1. Apply Keyword Encoding to the following quote from mentalfloss.com[1]:

> Believe it or not, this sentence is grammatically correct and has meaning: "Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo." First devised by professor William J. Rapaport in 1972, the sentence uses various meanings and parts of speech for the term "buffalo" (and its related proper noun "Buffalo") to make an extremely hard-to-parse sentence.
>
> Although most people know "buffalo" as both a singular and plural term for bison, and "Buffalo" as a city in New York, "buffalo" is also a verb meaning "to bully, confuse, deceive, or intimidate." Using these definitions, Wikipedia suggests the sentence can be read: [Those] (Buffalo buffalo) [whom] (Buffalo buffalo buffalo) buffalo (Buffalo buffalo).
>
> Still too hard to follow for those of us who don't know "buffalo" as a verb. Refine once more: [Those] buffalo(es) from Buffalo [that are intimidated by] buffalo(es) from Buffalo intimidate buffalo(es) from Buffalo.
>
> And once more: Bison from Buffalo, New York who are intimidated by other bison in their community also happen to intimidate other bison in their community.

2. Calculate the compression ratio of the new compressed message.

☐ Task 3.2

1. Apply Run-Length encoding to the following:
   - AAAAAAAAAAaaaAAAABBCCCCDDDdAAAAaEEEEEE
   - 10111011011100000001010111111010000011111001000001
2. Calculate the compression ratio of the new compressed messages above.

☐ Task 3.3

1. Construct a Huffman Tree and encode the following message:
   - the cat in the hat sat on the mat
2. Calculate the compression ratio of the new compressed messages above.

[1] http://mentalfloss.com/article/18209/buffalo-buffalo-buffalo-buffalo-buffalo-buffalo-buffalo-buffalo