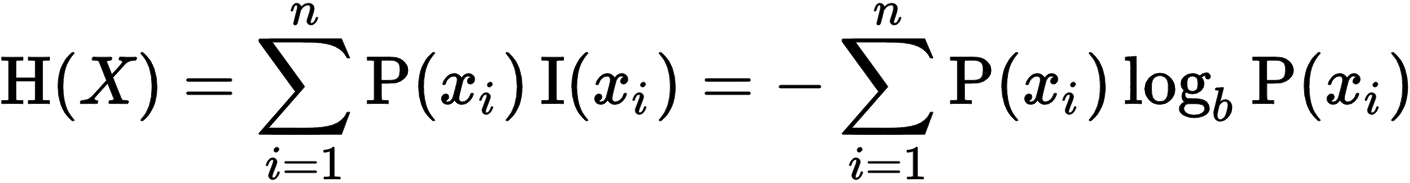
Experiment 3 – exploring the distance to uniform distribution and entropy of the columns of the dataset.

For each column of the dataset calculate the distance to the uniform distribution and the entropy. Plot this data. The next formula is used to calculate the entropy:



,

where b = 2 to calculate the entropy in bits, P(x) is the probability mass function. The probability mass function can be approximated as the frequency of appearances of the letters of the alphabet in the dataset.