Workshop Systems Analysis

Juan Diego Lozada Gonzalez

COD: 2022020014

Universidad Distrital Francisco Jose de Caldas

Systems Analysis

Carlos Andres Sierra

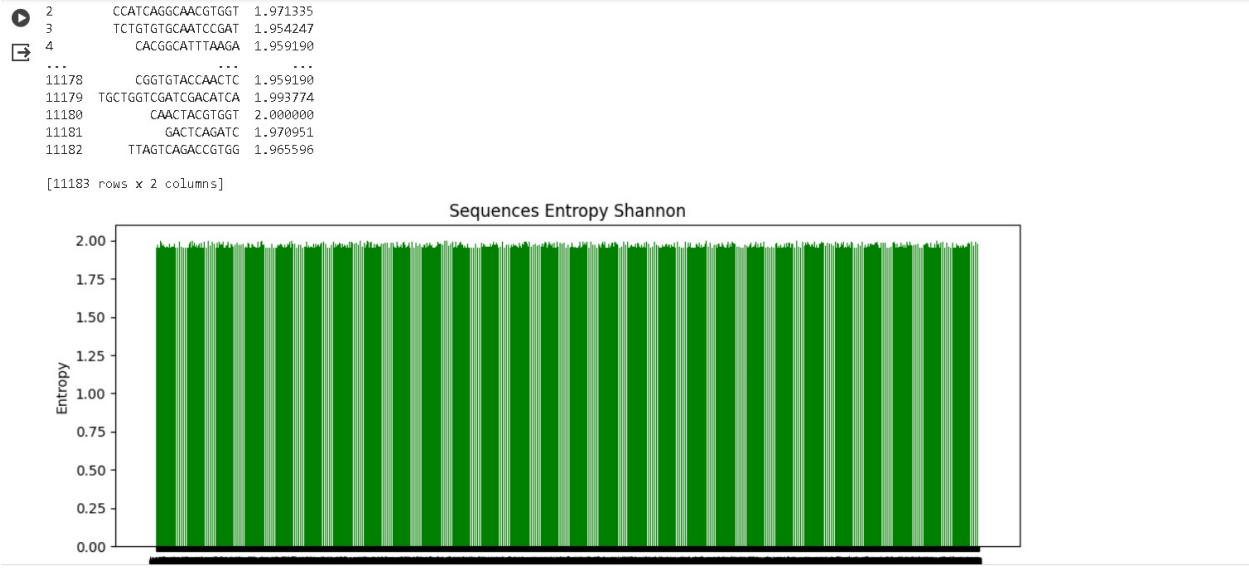
March 14, 2024

Workshop 1

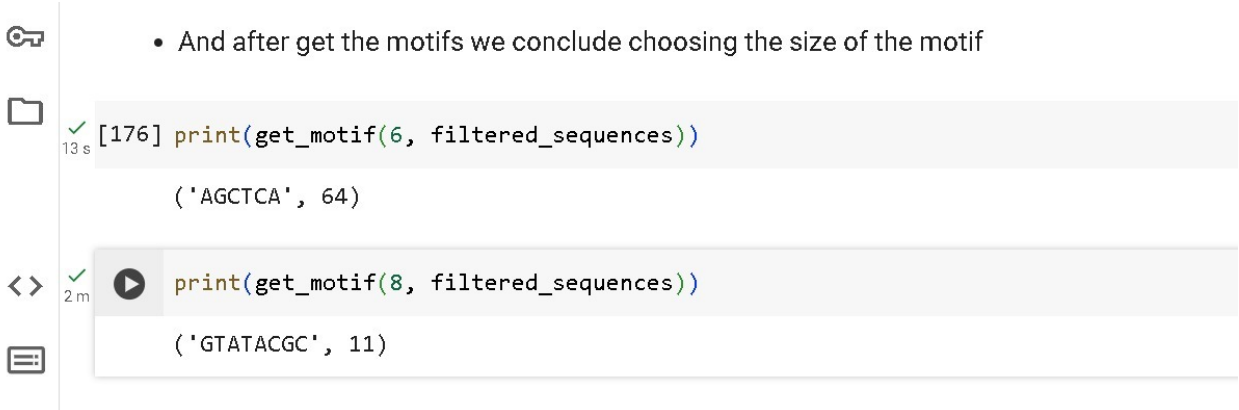
In this workshop we calculate the entropy from a sequence of nitrogen bases (A,G,T,C) with the goal to analyze why is important the entropy in the area of data, the Shannon Entropy have many uses but we are using to measure the information content in a set of symbols

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{1}$$

The equation (1) is the formula. After some results we conclude that is we have 50,000 of database size around the 20% continue after the filter, in my case i use 1.95 of filter. And we conclude that between more near the number of 2, the size will decreases.



And at the moment of get the motifs 6 and 8, i realize that if we choose a motif more big the combinations will decrease as you can see



In conclusion, the Shannon Entropy help us to find patrons and in this case those patrons represent genetics diseases and the Shannon Entropy can make the difference depending the treshold you choose