Data-X Homework 10:

1. The dataset shown below represents bank customers with 3 features and the label corresponding to each customer identifies whether they are the loan defaulters or not. Description:

    a. **HasJob:** Binary valued, equal to 0 when a person has no job and 1 otherwise.
    b. **HasFamily:** Binary valued, equal to 0 when a person has no family and 1 otherwise.

    c. IsAbove30years: Binary values, equal to 0 when a person's age is 30 or below and 1 otherwise.

    d. Defaulter is also a binary valued label which is equal to 1 if a person is a defaulter and 0 otherwise.

    Using this dataset **identify the best feature to do the first split** in a binary decision tree, so as
    to maximize the information gain in the next split, show your calculations.

HasJob	HasFamily	IsAbove30years	Defaulter
1	1	1	0
1	1	1	0
1	0	1	0
0	1	0	0
0	0	1	1
0	1	0	1
1	0	1	1
1	0	1	1

2. Given a signal of three symbols **S=(A,B,C)** and P(A) =0.7, P(B)=0.2, P(C)=0.1

    What is the entropy of S? What does it mean according to the Source coding Theorem?

    Hint: Go through the lecture slides