UNIVERSITY OF
EASTERN FINLAND

Please carefully read and follow the general instructions regarding Python notebooks. Failing to meet the requirements might lead to penalties. `https://moodle.uef.fi/mod/page/view.php?id=2775059`
If you suspect that something is wrong with some task instructions, please contact the lecturer.
If you face persistent issues while working on a task, do ask for help, e.g. during a course meeting or by contacting the lecturer via email.

Attached material.

`clustering_beans.ipynb` Python notebook

`Dry_Bean_Dataset_small.csv` small variant of the beans dataset from `https://archive.ics.uci.edu/ml/datasets/dry+bean+dataset`.

Task 1 (Basics of clustering on the beans dataset).
The Python notebook `clustering_beans.ipynb` provides an example of applying various clustering algorithms to a small variant of the beans dataset from `https://archive.ics.uci.edu/ml/datasets/dry+bean+dataset`.
    Read and run through it.
    This notebook shows an example of how to prepare the dataset, how to run different clustering algorithms from the popular `sklearn` library on the data, and how to compare and evaluate the results. You should add your observations, about the dataset, the mining process and the obtained results, as relevant, throughout the notebook. In particular, replace "[... add your observations here and throughout, as relevant]" at the end of the notebook.

Task 2 (Clustering on other datasets).
Carry out a similar analysis on some other datasets, such as

miceprotein from `https://archive.ics.uci.edu/ml/datasets/mice+protein+expression`

realestate from `https://archive.ics.uci.edu/ml/datasets/real+estate+valuation+data+set`

tennis from `https://archive.ics.uci.edu/ml/datasets/tennis+major+tournament+match+statistics`

wholesale from `https://archive.ics.uci.edu/ml/datasets/wholesale+customers`

    You can start from the provided example notebook, but you must update the explanations and observations according to the dataset considered and the results you obtain.
    In other words, for each dataset, you should explain what the dataset represents, actually look at the results and comment on how to interpret them.
    You should not analyze all of the datasets listed, but consider at least one. You might consider datasets of your own choosing. It is better to carry out a careful and thorough analysis of few of the datasets, than to cursorily work on all of them.