

Exercise 5 - deadline 23.04.2021 23:59

The dataset for this (and the following exercises) is the Penn World Table dataset.

You can find the latest version of the dataset here:

<https://www.rug.nl/ggdc/productivity/pwt/?lang=en>

The goal for this exercise is to create and analyse a grouping of countries for two different years. Analyse the grouping and the evolution between the two years.

The goals of the exercise:

1. Import the data and filter the two datasets for the chosen two years.
2. Choose the clustering technique and obtain the clusters for the two datasets.
3. Choose one of the years and visualize the results of the clustering. Give as much detail as possible.

The final submission should contain (at least) one script to contain all the goals mentioned before.

Optional goals:

4. Analyse and explain the differences (evolution) between the results for the two years.
5. Import the data directly from the source file (without modification). All processing should be done in the Python script.
6. Redo all the steps using a different clustering method. Analyse the differences between the results.

References:

<https://scikit-learn.org/stable/modules/clustering.html>

Important Note:

The submission should be done individually. Please name your script/notebook accordingly. At the beginning of your script / notebook give the name(s) of all students who worked on the exercise. You may work in small groups (of 2-3) but all students must be confident with the work of others. They should understand and be able to answer detailed questions of the implementation.