DAPR LAB: FINAL EXERCISE
Mandatory (10% mark)

Create a Python notebook (.ipynb) according to the following instructions:

1. Load a database with at least 4 numerical features from one of these data repositories:
- Kaggle.com: https://www.kaggle.com/
- UCI datasets: https://archive.ics.uci.edu/datasets
- Scikit-learn built-in toy datasets: https://scikit-learn.org/stable/datasets/toy_dataset.html
- Seaborn datasets: https://seaborn.pydata.org/generated/seaborn.load_dataset.html

2. Convert data into q pandas dataframe and remove non-numerical features.

3. Handle the potential presence of missing data (dropna, fillna or KNNImputer)

4. Visualize the numerical features using a seaborn pairplot https://seaborn.pydata.org/generated/seaborn.pairplot.html

5. Perform a dimensionality reduction of the dataset using PCA.

6. Cluster the PCA-projected data using a clustering algorithm (k-means, Agglomerative clustering or GMM)

7. Visualize the resulting clusters and comment the results.