Project 1

The objective of project 1 is to implement different clustering methods to synthetic and real-world data and validate using external and internal validation techniques

Task 1

Data sets, "Data1.csv", "Data2.csv", "Data3.csv", "Data4.csv", "Data5.csv", "Data6.csv", "Data7.csv", "Data8.csv" contain the data points and their respective class information. For each of the datasets follow the below steps

1. Use K-means and hierarchical clustering methods to generate clusters
2. Evaluate the performance of the clustering algorithm using external validation metrics
3. Plot (2D or 3D) the data points for each dataset and color them according to the original class
4. Plot (2D or 3D) the data points for each dataset and color them according to the class allocated by the clustering algorithm

Task 2

The world indicators dataset compares different countries based on selected attributes.

1. Use K-means and hierarchical clustering methods to group similar countries together
2. Use Internal validation metrics to report the cluster quality
3. Report the best clustering solution. Give a detailed list of all the groups and the countries included within the groups
4. Generate three different scatter plots of your choice and color the data points according to the group. **Example:** "Life expectancy vs GDP", "Infant Mortality vs GDP", etc.

Submission Format

1. Submit all the solutions as a Python Notebook
2. Include texts to explain the solution
3. Include equations for the evaluation metrics in the Python Notebook
4. Use modules in Python that provide functions for different clustering methods and cluster validation
5. Students can also create their own custom functions if necessary
6. This is a group effort
7. Only one member from each group needs to submit the solution
8. Submit the solution by April 7