Task 3 Clustering - Machine Learning

- 1. Select any table dataset from <u>https://www.kaggle.com/datasets</u> contains at least 10 attributes and 1000 records. You can reuse the dataset used in the previous tasks, if it's suitable for clustering tasks.
- 2. Visualize the pairwise attribute values distribution for the dataset. Use pairplot/subplots to properly assess the data. Find at least 2 pairs of attributes which can be clustered (i.e. not a single big cluster of values).
- 3. Write a Python program that implements clustering algorithms for the selected pairs of attributes. Use the algorithms listed below. Standardize values if necessary.
 - K-means clustering
 - Agglomerative clustering
 - DBScan clustering.
- 4. Conduct and experiment to find the optimal number of clusters for the selected pairs of attributes. For every result display the clusters' centroids and quality assessment metric (silhouette coefficient or any other metric). Give an explanation for the used quality metric and usability of the algorithm for your data values.
- 5. Choose the best combination of algorithm and hyperparameters for chosen pairs of attributes, write a short report and conclusions for the work.