

## **Data Analysis and Data Mining Assignment 2**

Choose a dataset (e.g. from kaggle.com, or seaborn dataset) (excluding iris.csv) and perform data analysis in Python using the following methodology.

1. Load the data and perform necessary data cleaning steps.
2. Analyze the dataset with descriptive statistical indicators and charts. Filter out outliers.
3. Identify correlation relationships between variables.
4. Cluster the data using a selected algorithm, or create a classification model for the dataset using a chosen method, and demonstrate the model's accuracy.

Create a PDF document about the analysis, including: the description and source of the dataset, a brief overview of the applied procedures, the results, and the evaluation of the results. The Python code must be attached to the task, accompanied by explanations (comments).

**Deadline: 8<sup>th</sup> December, 2023**