Seminar 1 – algorithms comparison, which one is best?

Task: consider the file *results.csv* that contains results of different classification models applied on different data sets. Find a way to represent the results and show which model is better.

The file contains the following columns:

- *id_alg*: integer that identifies the models;
- *param_1, …, param_4*: different model parameters;
- *id_dataset*: integer that identifies the dataset;
- *ind_0, …, ind_9*: metric used to evaluate the model on each folds (10 folds, Stratified k-Folding), the metric is in range [0, 1], higher values indicate better performance;
- *meand_ind*: mean of the metric used over the ten folds;
- *std_ind;* standard deviation of the ten folds.

Some useful references:

- Kaggle kernel showing seaborn plots in python over the *Students performance in exams* dataset (you can also find the dataset in the files section on MS Teams)
- Kaggle kernel showing plotly visualizations