Recommended for the projects

- You may focus either on a Data Science or on a Data Engineering project. Projects may contain both Data Science and Data Engineering methods.
- It is possible to implement a Data Science projects in Spark.
- A Data Science project should contain at least **3** dimension reduction methods.
- The dimension reduction methods will not be counted towards the number of ML algorithms in the project. As an example, they may be combined with some ML algorithm for performance measurement (compare it before and after).
- The details for a Data Engineering project will be provided in the second part of this course.

Recommended for the projects

For a Data Science project the technical report may contain the following sections

- Introduction / Context / State-of-the-art / Related work
- Dataset description / Dataset production
- Dataset preparation and exploratory data analysis
- Methods (libraries and models used, experimental setups, evaluation, requirements, limitations, etc.)
- Experiments and their results
- Computational details (hardware, running times, memory usage, parallelization, etc.)
- Discussion (which are the best experimental setups, best models, best tuning parameters?)
- Conclusion

Recommended for the datasets

- The datasets size should be > 1 MB.
- The number of features in the datasets should be at least 10.