# Statistical & Machine Learning

## *Group Project Description*

The project is a task with *maximum 3 group members* where one tries to come up with a complete machine learning benchmark experiment.

Purpose

You will work on a given Kaggle data set and you will follow the different phases of a predictive modeling project using the machine learning methods discussed in this course. The purpose is to practice and perfect the predictive modeling process and understand the advantages and drawbacks of each machine learning model.

The output for this project is a note that describes the methodological steps undertaken during the different phases:

- Data preprocessing
- Model building and hyper-parameter selection
- Experimental setup, variable selection and resampling
- Model scoring and model assessment

Submission

- PPT + R script in Jupyter Notebook + Basetable
- Upload the project to GitHub or cloud storage service

Evaluation

The project will be evaluated on:

- The degree of depth of the machine learning pipeline.
- The correct setup of the benchmark experiment pipeline (e.g. model building, hyper-parameter selection, experimental setup, variable selection, resampling, model scoring, etc.).
- The performance of the final model (compared with the Kaggle leaderboard).
- The quality of the R script.

Timeline information

- Section 8, 27 March 2020
- 10min presentation + 5min Q&A per group

Hint

- Follow the checklist in Data Science Presentation Template.
- If the original size of the Kaggle data is too big, take a subsample of 1-5%.