# Statistical Learning (MT7049) - Project 4

**Task:** Pick up a data-set from the UCI Machine learning repository: https://archive.ics.uci.edu/ml/index.php
then analyze it using any method from the ESL book. You can use any program package in the analyses. This project should be solved by teamwork with 2 students in each team. Since we may not have even number of students, it is also permitted (but not recommended) if 3 students as a team carry out the project.

**Amount of work:** Project 4 contributes the same weight to the final grade as other projects, so the amount of work for each student should be about the same as in project 1-3. For a team of 2 (or 3) students, the amount of work should therefore be about twice (or 3 times) as in project 1-3.

**Guidelines for the written report:**
1)  Give the student names of the group.
2)  State which data-set is picked and why it is picked
3)  Describe what the data is about and the data properties with suitable exploratory analyses.
4)  Explain which methods in the ESL book are used, and justify why the methods are suitable for the chosen data.
5)  Concisely state how the analysis is performed (e.g. which functions and packages are used, parameter setting, etc.) and the analysis results in terms of graphs, tables, etc. Do your graphs effectively present the message you want to convey?
6)  **Most importantly: Explain and discuss the implications from the analysis results, what you learn from the analysis, what are the possible improvements, etc.**
7)  Give appropriate references.
8)  Your performance of the project will be judged on how clear, logical and organized your report and talk are given, and on whether the suitable methods are employed and correctly implemented for the analyses. Please include enough explanations and discussion and avoid writing a technical manual only on what things are done. Moreover, a report size (not including the appendices) of around 5 pages is preferred.
9)  The written report as a single .pdf file should be submitted to the course page by the day before the oral presentation, i.e., at **1:00pm on Jan 7 or 13, 2025**. Any source code associated with the analyses should be attached as appendices in the report with sufficient documentations. **Members of the same group should submit the same report to the course page.**

**Guideline for the oral presentation:**
1)  The presentation is 10-12min for each group plus ~ 2min for questions.
2)  The talk should cover concisely items 1 to 6 above in the written report.
3)  It is up to you how to distribute the talk time to each member of the group.