# Statistical Inference and Data Mining, 371-2-1721

## Mid-term Project

## Winter semester 2022-2023

**Submission Date: 25/12/2022 -10 pts for late submission (more than 1 day)**
Team size is 2 students

### Project Description

- Each team must register on the Kaggle website. The competition is on this link
- The competition is a private competition only for this course
- Detailed instructions on the competition, data, form of submission, and model evaluation are in Kaggle.
- **You can only use the multivariate regression model that was presented in class**. Variants such as ridge, lasso, or kernel regression are allowed and encouraged. Using any other method (such as random forest regression, k-NN, neural network) will disqualify your submission.
- Feature engineering, data pre-processing, and data cleaning are welcome and advised!

### Evaluation

70% of the grade will be determined according to your rank on a private test set that will be evaluated **after** the competition is over:

**The bottom 25% get 80, the next 25% get 85, then 90 and 95. The top 3 get 100.**

NOTE: During the competition a **public** test set will be available, and the **leaderboard** will show your current ranking according to that test set. The **private** test set is NOT the public one (but they come from the same distribution), and therefore the final ranking may change.

30% of the grade is a report that you submit according to the following instructions:

- You must present your main results concisely and clearly (the model you submitted, data pre-processing, data exploration). The output should also include graphic visualizations wherever necessary.
- You must address theoretical issues learned in the course that may affect the results and explain their impact on your work (e.g. overfitting.)
- The report should be submitted in PDF format. Jupyter notebooks are not accepted.