Department of Economics **University of Zurich**

Spring term 2020

Part II: Hands-on assignment
Seminar MOEC0482

April 7, 2020

Task Description

In this project, your task is to predict housing prices in Switzerland. In OLAT, You can find training.csv ("price" is the outcome variable) and X_test.csv (only features, without price).

The work is done in the same groups of two students as in the first task. Your output will be price predictions for the test set, the R code of your models as well as a short report. The prediction set only contains two variables: id and price (See Y_test_example.csv). Be free to use any tools you have learnt in the course (or somewhere else) for this task. Make sure you explore methods from at least three topics. The report should explain and discuss your approach (what you do and why you do it) as well as your results. You should also discuss other approaches that you have considered and document why in the end you chose not to use it. The deadline is **May 05 at 23:59**. Please send to lin.xu@econ.uzh.ch.

In the last session on Friday, May 8, 9:15-12:00, the six groups who have not presented in Part I will give short presentations of the results for their preferred specification. The presentation should be no more than 10 minutes. You can directly use your report to present, so writing in the format of presentation or R markdown or any format you prefer (one documentation is enough). We will calculate MSE using the true data and announce the winner that has the smallest MSE at the end.

Some hints

There are missing observations for some variables, in training and test data. We will ask you for predictions for all of our test data and not just those without missing observations. Be careful of dealing with missing variables. (Take a look at R mice package to find good and quick ways to impute the missing variables.) Besides dealing with missing values, there are other data processes that may potentially improve performance, such as robust scale and outlier detection. And do not forget to do the same procedures not only on training set, but also on test set. Play around and have fun!