Regression assignment

The intention of this assignment is to practice all the techniques that were taught in sessions 1 to 5. In this exercise we will try to predict the price of houses in the Boston area given several characteristics of the house like the lot area, zoning, street type, neighborhood, etc.

The dataset we are using comes from Kaggle. You can have a look to the dataset here: https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview. The Nevertheless, I have uploaded the dataset to campus along with the columns descriptions.

The exercise is oriented to creative feature engineering. You must make many decisions about what variables to drop, how to deal with null values, columns transformations, etc. This exercise is completely open, so I will not provide a Jupyter notebook to start with. However, you can reuse the previous notebooks as a starting point because all the exercises are quite similar. In the same way, you can use any of the models we have seen in class.

Some recommendations:

- This exercise can be completed in one hour or you can work in it for weeks (like the Kaggle contenders do). My advice is starting with the simplest program that works with the data and refine the results from that point.
- Visit the Kaggle code to see what others have done. It is totally fair if you 'borrow' (or even copy) ideas from other Kaggle contenders. That is one of the best methods for learning.

What to deliver: the notebook with the results. You should document clearly what decisions you have made and why. You must print at least the resulting accuracy after cross validation.