Regression project notes

Train\_raw data contains the raw data. It has been preprocessed by the teacher using the Project\_2019\_Preprocessing.Rmd file. This created two files: train\_preprocessed and train\_imputed.

Use train\_preprocessed for your regression model (because data is scaled and all).

Use train\_imputed for your EDA.

You can do other preprocessing stuff if you want but if you do write it in the report.

You can use the log function for the price and then use exponential function to have an idea of your score on Kaggle.

You are very free on what you have to do. You can build a model to predict price (like Kaggle), you can try to explain which variables explain the best the price… etc. Just have to do it well !

Try to come up with your own precise question and try to answer it as well as possible. You cannot do everything in this project and are not asked to do so.