

**Sobre o problema:** Um empresário deseja criar uma startup que auxilia na compra e venda de imóveis, e para isso é necessário realizar a correta precificação do imóvel. Para automatizar essa tarefa, você foi contratado para construir um preditor de preços de imóveis, baseado em suas características.

**Sobre o banco de dados:** o banco de dados possui as estatísticas de 1460 imóveis. Foram levantadas diversas informações, como tipo de rua, condições do sistema elétrico, tipo de piscina, etc. Mas informações podem ser obtidas no arquivo `data_description.txt`. São 81 variáveis ao total, das quais 1 é identificadora (`Id`), 79 são explicativas e 1 é a variável resposta (`SalePrice`). Observe que em algumas variáveis `NA` é uma categoria!

**Sobre a divisão treino/teste:** o cliente exige que 20% das amostras sejam usadas como amostra teste, para uma boa estimativa do RMSE,  $R^2$  e MAE. Utilize seu número de matrícula como `set.seed()`.

**Sobre tratamento dos dados:** qualquer método de pré-processamento pode ser utilizado, mas toda metodologia utilizada deve ser deixada bem claro no relatório final.

**Sobre o treinamento :** deve-se utilizar o Gradiente Boosting. Para definir os valores dos hiperparâmetros, devem ser testados PELO MENOS 6 valores diferentes para profundidade da árvore, 3 para a taxa de aprendizado e 3 para o número de árvores. Alterações no `n.minobsinnode` são opcionais. A utilização das funções `train()` ou `gbm()` ficam à sua escolha. O critério de seleção final do modelo deve ser o com menor RMSE. Utilize seu número de matrícula como `set.seed()`.

**Sobre o relatório final:** deve ser um relatório final, em formato de apresentação (beamer), a ser apresentado ao contratante. Esse relatório deve conter (de maneira explicativa) toda metodologia empregada, resultados dos modelos, modelo final e comportamento do modelo final na amostra teste.

**O que deve ser entregue na data final:** o relatório (.pdf), os códigos utilizados (.txt), os modelos de pré-processamento, o modelo final de regressão e uma função que rode os modelos de pré-processamento e de regressão (.Rdata ou .rda). Os modelos criados devem ser capazes de rodar em uma nova amostra SEMELHANTE A QUE VOCÊS RECEBERAM (a menos da variável `SalePrice`, que NÃO EXISTE). Toda alteração no banco de dados de vocês deve ser feita nessa amostra também. Para fins de registo, os dois integrantes da dupla devem realizar o upload do trabalho. Entre os dias 15 e 17 de setembro teremos as apresentações (horários a definir em breve).

