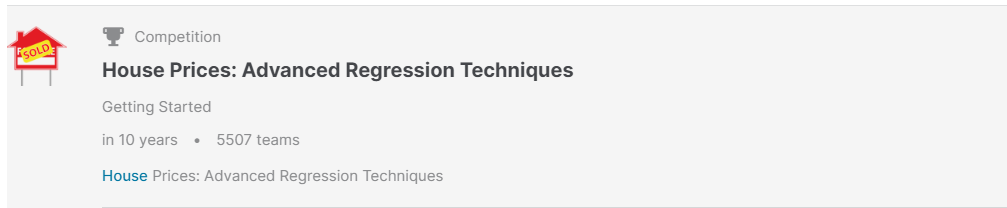
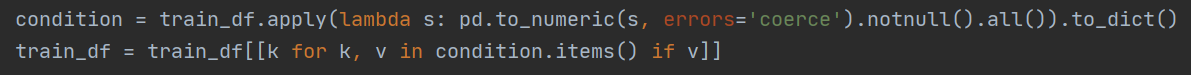
首先从kaggle上下载house\_prices数据集：



使用train.csv进行训练

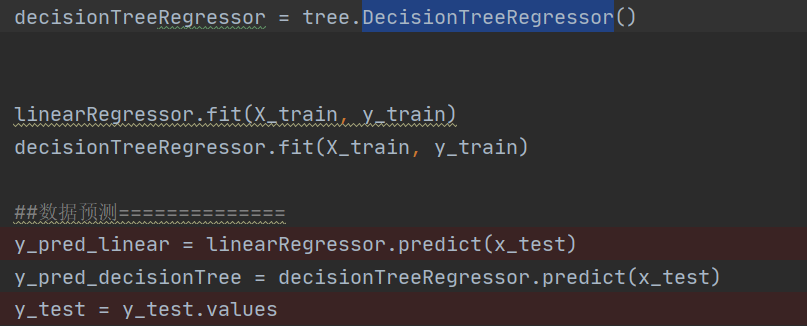


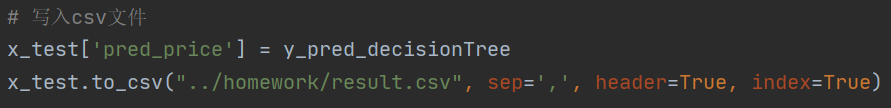
这里我们需要对数据先进行处理，将空值去除：  


将训练数据划分为训练和测试集：

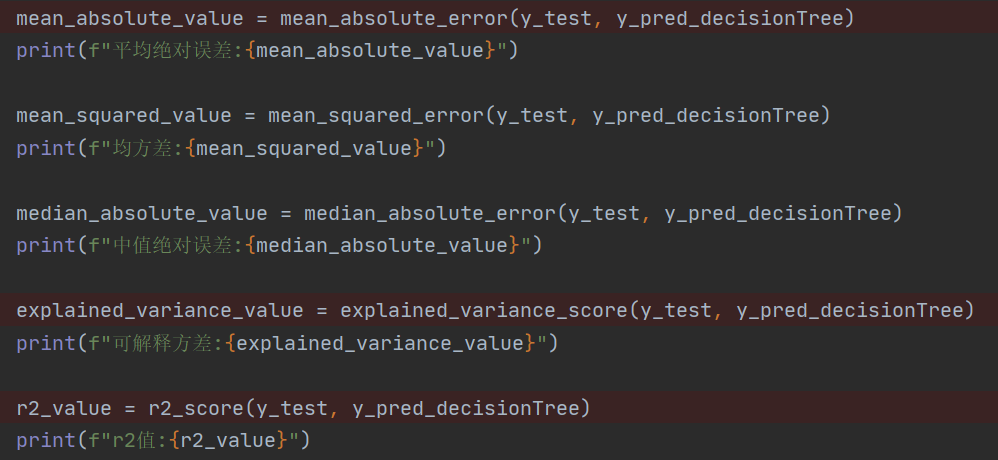


接下来就是训练和预测了，使用的是sklearn.tree. DecisionTreeRegressor这个模型，效果比线性回归好，说明数据并不是线性关系：

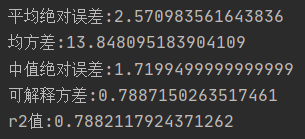


将预测结果写入csv文件：  


评估模型，使用了几个评估度量，包括平均绝对误差，均方差，中值绝对误差，可解释方差，r2值



结果如下：



生成的csv文件会保存在s3中practice/homework/result.csv,手动将其下载到本地。