(DataCastle) 美国King County房价预测训练赛（新）

链接：

https://www.dcjingsai.com/common/cmpt/美国King%20County房价预测训练赛%EF%BC%88新%EF%BC%89\_竞赛信息.html

## 任务

我们希望学完《数据分析师（入门）》的学员，可以根据课上老师所讲授的知识和回归分析的方法，从给定的房屋基本信息以及房屋销售信息等，建立一个回归模型预测房屋的销售价格。

## 数据

数据主要包括2014年5月至2015年5月美国King County的房屋销售价格以及房屋的基本信息。

数据分为训练数据和测试数据，分别保存在train.csv和test.csv两个文件中。 字段是说明如下：



测试数据不包括房屋销售价格，学员需要通过由训练数据所建立的模型以及所给的测试数据，得出测试数据相应的房屋销售价格预测值。

（注：比赛所用到的数据取自于kaggle datasets，由@harlfoxem提供并分享。我们只选取了其中的子集，并对数据做了一些预处理使数据更加符合回归分析比赛的要求。）

如遇数据下载打开乱码问题： 不要用excel打开,用notepad++或者vs code。文件格式是通用的编码方式utf-8，如果要用excel,请转换为ansl格式或者gbk格式。

注：由于该数据来自其他网站的开源数据，为维护各位小伙伴的学习体验，我们要求：

a 仅限使用该比赛提供的数据，禁止使用其他数据；

b 取得满分或者非常接近满分的选手必须在竞赛圈中开源代码，否则我们将取消其排行榜成绩。

## 评分标准

评分算法

算法通过计算平均预测误差来衡量回归模型的优劣。平均预测误差越小，说明回归模型越好。

参考代码如下：

from sklearn.metrics import mean\_squared\_error

y\_true = [1, 2, 3, 4]

y\_pred = [1.1, 2.2, 3.3, 4.4]

score = mean\_squared\_error(y\_true, y\_pred)