此挑战将允许您在真实数据集上测试不同 PRAMA 序列中看到的方法。这是一个回归问题：预测位于华盛顿州金县的房地产销售价格。这个县特别包含西雅图市。所有房地产销售都发生在 2014 年 2 月至 2015 年 5 月之间。

训练数据集包含 17,149 个房地产销售，测试数据集包含 4,287 个。您无法访问测试集中的房地产销售价格，您需要对其进行预测。因此，此挑战的目标是从学习游戏中估计模型，然后在测试游戏中预测商品价格。你有一个示例文件要在 kaggle 平台上提交；它必须包含两列：销售 ID 和价格。

此挑战将以 5 人为一组进行。该组将在整个挑战中保持不变，并且对于即将到来的项目也是如此。

您将需要提交 2 份报告（总结您的方法和获得的结果），以及使用的脚本（R 或 python）：

渲染 n°1：从序列 1、2、3 和 4 中看到的方法，不迟于 03/19

渲染 n°2：从序列 5、6、7、8 和 9 中看到的方法，不迟于 07/05

效果图必须不迟于上述日期发送给 PC 老师。将提供代码（R 或 Python 格式）以及报告（pdf 或 html 格式），名称格式为：

挑战{返回的数量}组{number\_group}

数据集包含 21 个变量，如下所示：

id：销售的ID，

date：销售日期，

prix：售价，

nb\_rooms：房间数量，

nb\_sdb：卫生间数量，数值0.5对应不带淋浴/浴缸的WC，

m2\_interior：房屋居住空间的面积，

m2\_garden：室外空间的面积，

m2\_floor：位于地面以上的房间的面积，

m2\_soussol：位于地面以下的房间的面积，

nb\_floors：楼层数，

vue\_mer：布尔变量，表示是否有海景，

vue\_note：0到4之间的注释表示视图的质量，

state\_note：从 1 到 5 的等级表示房子的状态，

design\_note：从 1 到 13 的等级表示房屋的设计水平，

construction\_year：建设年份，

annee\_renovation：如果有，则为最后的装修年份，否则为0，

m2\_interieur\_15voisins：15 个最近邻居的居住空间的 面积，

m2\_jardin\_15voisins：15 个最近邻居的室外空间面积，以 m2 为单位，

zipcode：住宅地址的邮政编码

lat：纬度，

long: 经度