
数据集

作业 1-3 将会使用同一个数据集，来自于 Kaggle 的房屋价格预测任务。

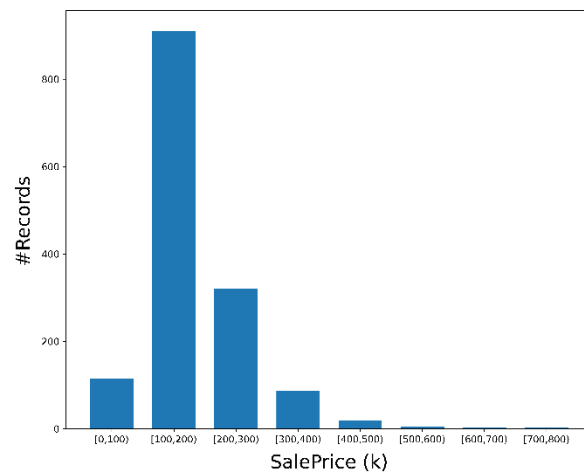
- ✧ 关于数据集每一列属性的描述可以参见 `data_description.txt`
- ✧ 我们的作业只使用 `train.csv` 文件里的数据
- ✧ 数据集下载地址：
<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

作业 1：数据可视化

作业截止日期：北京时间 5 月 4 号 23:59

作业提交内容：画图的脚本文件以及可视化结果

1. 【3 分】如下图所示，根据 train.csv 文件的最后一个属性 SalePrice 的值画柱状图，按照区间大小 10 万来统计画图，即横坐标对应[0,100,000], [100,000,200,000], [200,000,300,000]...等区间。



2. 【3 分】如下图所示，画出 (SalePrice, GarageArea) 的散列图，一套房屋对应一个点，横坐标为该房屋的 SalePrice，纵坐标为对应的 GarageArea。(3 分)



3. 【4 分】数据集的房屋类型（BldgType）共有 5 种，包括 1Fam, 2FmCon, Duplx, TwnhsE, TwnhsI。如下图所示，请为每一种房屋类型提取它所对应的 SalePrice 数据分布，并画出箱线图（boxplot）。

