

练习题A：二手房房价分析与预测

数据集已经给出

	小区名字	总价	户型	建筑面积	单价	朝向	楼层	装修	区域
0	中天北湾新城	89万	2室2厅1卫	89平米	10000元/平米	南北	低层	毛坯	高新
1	桦林苑	99.8万	3室2厅1卫	143平米	6979元/平米	南北	中层	毛坯	净月
2	嘉柏湾	32万	1室1厅1卫	43.3平米	7390元/平米	南	高层	精装修	经开
3	中环12区	51.5万	2室1厅1卫	57平米	9035元/平米	南北	高层	精装修	南关
4	昊源高格蓝湾	210万	3室2厅2卫	160.8平米	13060元/平米	南北	高层	精装修	二道
5	金色橄榄城	118万	3室1厅1卫	200平米	5900元/平米	南北	高层	简装修	二道

问题描述

一．问题背景

住房一直以来都是人们关心的热门话题，房价也是人们时时刻刻关心的热点。虽然新房子更加上档次，但是二手房有着现房交易，地段较好，配套设施完善，选择面更加广泛等优势，因此二手房越来越受到广大消费者的青睐。根据现有二手房的地段、装修等属性预估该二手房的价格也是买卖双方所关心的问题。因此通过现有数据，分析并且预测二手房的价格是一项有意义的研究课题。

二．解决问题

1.全市二手房数据分析

（一）区域二手房均价分析

根据附件中的数据集，将二手房数据按照“区域”属性进行划分，然后计算每个区域的二手房均价，最后将区域及对应的房屋均价信息通过纵向条形图显示。

（二）区域二手房数据及占比分析

根据附件的数据集，计算各个区域二手房数量占总二手房数量的百分比，并画出饼状图。

（三）二手房装修程度分析

将二手房按照“装修”属性进行划分，然后计算不同装修程度的二手房数量，并用条形图显示不同装修程度二手房的数量。

2.二手房房价预测

(一) 将二手房按照户型进行分组，然后提取前5组最热门的二手房户型（出售数量最多的5组户型），最后计算这5个热门户型的均价并画图显示。

(二) 选择附件中适合的属性，建立模型预测二手房的价格。

解答

预测准确度衡量指标

1. 预测值与实际值的比率:

$$\text{Ratio} = \frac{\text{predicted value}}{\text{actual value}}$$

这个比率可以帮助理解预测值相对于实际值的大小关系，并在代码中被计算为 `df_result3` 和 `df_result4`。

2. 比率的最大值:

$$\text{Max Ratio} = \max\left(\frac{\text{predicted value}}{\text{actual value}}, \frac{\text{actual value}}{\text{predicted value}}\right)$$

这个指标考虑了预测值和实际值之间的两种比率，选择了它们的较大值，并在代码中被计算为 `df_result5`。