

# 数据预处理和可视化作业-1

## 数据预处理

先按行读入之前保存的 json 文件, 并进行处理, 在处理中 面积 取最小值, 房型 取最小房型, 总价 由 面积 和 均价 相乘得到.

在 csv 文件中写入表头, 按照表头将处理后的数据写入到 csv 文件中.

执行文件 `python process_data.py` 得到处理结果 `new_house.csv`

## 数据可视化

### 散点图

1. 使用 pandas 读入 csv 文件;
2. 取出 均价, 总价 和 类型 三个字段的属性, 并取出 类型 数据的种类用于显示图例和区分颜色;
3. 确定横纵坐标和表名, 绘制散点图 `scatter.png` .

### 直方图

1. 使用 pandas 读入 csv 文件, 并按照行政区分组获得每组 均价 的平均值和 楼盘 的数量

```
grouped_data = data.groupby('行政区').agg({'均价': 'mean', '楼盘名称': 'size'})
```

2. 确定横纵坐标和表名, 绘制直方图 `avg_bar.png` ;
3. 将单价换位总价, 其他步骤相同, 绘制直方图 `total_bar.png` .

### 饼图

将新房均价按照 `(0, 5]`、`(5, 10]`、`(10, 15]`、`(15, 20]` 四个区间进行分类, 绘制饼图 `avg_pie.png` .

1. 使用 pandas 读入 csv 文件, 并将均价按照上述区间进行分段;
2. 绘制饼图