

特征选择

```
'id', 'Size_num', 'ConstructionYear',
'Bedroom', 'Washroom', 'Livingroom', 'Kitchen',
'BuildingStructure', 'Elevator',
'Floor_level', 'StairsUnitRatio_num',
'Only_South', 'South_North',
'poi_total_score',
'price_per_meter' # 可以作为目标
]
最终文件保存于"final_house_features.csv"
```

一、POI总显著性特征

特征	类型	说明	计算方法
poi_total_score	数值	房源周边 POI 的加权总显著性	对房源周边 1km 内的 POI 根据类别权重 × 距离衰减累加得到

POI类别权重表

类别	示例	权重建议
amenity	餐馆、银行、邮局	0.9
shop	商场、便利店	0.9
tourism	景点、公园入口	0.6
leisure	健身房、KTV、公园	0.7
healthcare	医院、药店	1.0
public_transport	公交站	1.0
railway	地铁、高铁	0.9
aeroway	机场	-0.5
sport	体育场、游泳馆	0.7
education	学校、幼儿园	1
office	商务楼	0.6

距离衰减函数：

[
 $f(d) = e^{-\alpha d}$, $\alpha=0.5$, d \text{单位: 公里}
]

二、建筑硬件指标

0-1 区间性能评分（仅基于建筑性能）

结构类型	性能评分 (0-1)	等级说明
钢混结构	0.95	性能最优，耐久、抗震、防火等综合最佳
混合结构	0.75	性能良好，次于钢混但优于砖混/砖木
砖混结构	0.45	性能一般，常见于老旧建筑，抗震与耐久较弱
砖木结构	0.25	性能较差，防火与耐久性低，抗震能力弱
未知结构	0	性能未知

电梯：有电梯1，无电梯0，暂无数据0

电梯配置	适用楼层逻辑	评分 (C)
有电梯	适用于 中楼层、高楼层、顶层	1.0
	适用于 低楼层、底层	0.7 (资源过剩，但仍算优点)
	适用于 地下室	0.5 (无实际意义)
无电梯	适用于 底层、低楼层	0.8 (可以接受)
	适用于 中楼层	0.5 (对部分人不便)
	适用于 高楼层、顶层	0.3 (严重缺陷)
	适用于 地下室	0.5 (无影响)
电梯未知	所有楼层	0.5 (保守估计)

朝向组合与评分标准

我们可以根据这两个变量的组合，创建一个统一的“朝向评分”。

Only_South	South_North	说明	朝向评分 (0-1)
0	0	其他朝向 (东、西、北等)	0.40
1	0	仅朝南	0.75

Only_South	South_North	说明	朝向评分 (0-1)
0	1	南北通透	0.95
1	1	逻辑上不可能	需检查数据错误，可设为缺失值或取较高值(0

建筑硬件：

建筑结构'BuildingStructure'，便利性Floor_level+Elevator，朝向'Only_South', 'South_North',梯户比密度StairsUnitRatio_num'。使用相关性和随机森林特征重要性赋权重

三、户型与空间特征：

数值型直接使用

'Size_num', 'Bedroom', 'Washroom', 'Livingroom', 'Kitchen',

poi显著度分数