

# 《Python 与 R 语言编程》

## 实习报告

# 目录

一、 实习要求 .....	1
二、 数据预处理 .....	1
三、 各区县和乡镇房价特征 .....	2
3.1 整体分析 .....	2
3.2 武汉各地区房价分析 .....	4
3.3 武汉各乡镇房价分析 .....	5
四、 各区县和乡镇 <b>2018-2022</b> 年房价变化 .....	6
4.1 各区县房价变化 .....	6
4.2 各乡镇房价变化 .....	6
五、 高档小区分布特点 .....	7
六、 高档小区与名称的相关分析 .....	9
七、 <b>2022</b> 年的购房建议 .....	11

## 一、实习要求

本次课程结业任务是在老师提供的数据库基础上，利用 python 语言编程实现武汉房价数据分析并可视化，并参考范例完成数据分析报告。

基本要求如下：

1 报告中应有数据分析过程和结果图表，不能仅仅堆放结果，要有文字说明部分；

2 应有以地图的形式呈现的成果；

3 工作主体应该是 PYTHON 完成；如果部分工作用其他软件工具实现比较方便也可以选用；

4 要求报告至少包含以下内容：1 按区县和乡镇统计分析 2022 年各区和乡镇的房价特征；2 按区县和乡镇统计分析 2018-2022 年房价变化；3 根据 2022 年价格分析结果，自定义武汉高档小区标准，并分析高档小区分布特点；4 有人说从小区名称上可以看出小区是否高档，尝试一下，看看是否存在规律；5 假定某人武大遥感院上班，月收入 1 万，结合数据分析和可视化技术给出你在 2022 年的购房建议和理由；

5 可能用到的工具包：数据分析用 pandas，某些计算可能会用到 numpy，可视化建议用 Matplotlib，地理编码使用 geopy 库，地图可以用 folium。

## 二、数据预处理

本次数据为 2018 年底和 2022 年武汉的房价数据。基本包含了小区名称、房屋均价、所处区域、所处乡镇、地铁情况、地址、建造年代、建筑类型、物业管理费、物业公司、开发商、建筑数、房间数、百度经度、百度纬度、WGS1984 经度、WGS1984 纬度多个方面的属性。

属性	类型	说明
小区名称	文本	房产的标识名，无缺失值
均价	数值	分析的重点，无缺失值
区域	文本	房产分类的基准之一，无缺失值
板块	文本	房产分类的基准之一，无缺失值
地铁	文本	大部分是缺失值，可用于筛选临近地铁的房产
地址	文本	地理位置的文本分析基准，无缺失值
建筑年代	数值	少量缺失值，用于相关性分析时应剔除
建筑类型	文本	分析价值较低
物业费用	数值	少量缺失值，用于相关性分析时应剔除
物业公司	文本	可用于房产分类，但总体分析价值较低
开发商	文本	可用于房产分类，但总体分析价值较低
楼栋总数	数值	较独立的变量，可分析房产规模，本作业不讨论
房屋总数	数值	较独立的变量，可分析房产规模，本作业不讨论
经度（百度&WGS-84）	数值	地理坐标
纬度（百度&WGS-84）	数值	地理坐标

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	名称	城市	区县	乡镇	建筑年代	建筑类型	地址	供热	绿化率	面积	总户数	小区类型	供水	得分	月份	环比	价格	wgs84_lng	wgs84_lat	
2	保利新武昌	武汉	洪山	白沙洲	2019	高层	超高/张家湾路8号		2	30	635550	6458	公寓住宅	民用	7.9	12	-0.28	19324	114.2906	30.497237
3	保利上城东	武汉	洪山	白沙洲	2019	高层	超高/烽胜路20号		1	35	803170	2780	公寓住宅	民用	7.8	12	-0.54	16988	114.28586	30.498079
4	东原湖光里	武汉	洪山	白沙洲	2017	多层	小高/黄家湖大道		1	35		2584	公寓住宅	民用	7.4	12	-0.77	15446	114.29138	30.455994
5	烽火紫云	武汉	洪山	白沙洲	2013	超高层	烽胜路17号		1	42		3037	公寓住宅	民用	7.7	12	1.23	14712	114.29025	30.495432
6	安家名校东	武汉	洪山	白沙洲	21-Jul	多层	小高/白沙洲大道	None		35	700000	4348	公寓住宅	民用	7.5	12	0.51	10425	114.25317	30.426286
7	南湖巴黎香	武汉	洪山	白沙洲	2018	多层	高层/文化大道		2	35	250000	2003	公寓住宅	民用	7.5	12	-1.86	16129	114.31043	30.448852
8	佳兆业金域	武汉	洪山	白沙洲	2015	高层	超高/烽胜路36号		1	35	351653	1624	公寓住宅	民用	7.8	12	-0.63	17346	114.28632	30.453405
9	潮龙鑫城	武汉	洪山	白沙洲	2012	高层	白沙洲大道		1	35		3569	公寓住宅	民用	7.5	12	-0.5	11148	114.26144	30.479225
10	华润紫云南	武汉	洪山	白沙洲	2015	小高层	超/白沙洲路5C		1	35	133689	2332	公寓住宅	民用	7.7	12	-0.23	16412	114.2748	30.474649
11	楚天都市悦	武汉	洪山	白沙洲	22-Apr	多层	小高/双沁一路1		2	35	611000	3766	公寓住宅	民用	7.6	12	1.46	14214	114.2846	30.458033
12	世茂云锦玉	武汉	洪山	白沙洲	2018	高层	白沙五路		2	30		1836	公寓住宅	民用	7.7	12	-0.45	16047	114.2589	30.4828
13	清江锦城	武汉	洪山	白沙洲	2022	超高层	白沙洲路55		1	30	235878	2071	公寓住宅	民用	7.8	12	-0.22	17108	114.29225	30.477175
14	华润紫云南	武汉	洪山	白沙洲	2015	小高层	超/白沙洲路51		1	35		1304	公寓住宅	民用	7.7	12	-0.04	16664	114.27377	30.477595
15	佳兆业金域	武汉	洪山	白沙洲	2014	超高层	烽胜路36号		1	35		1935	公寓住宅	民用	7.7	12	-1.06	17765	114.28348	30.453417
16	富强大象	武汉	洪山	白沙洲	2011	小高层	高/烽胜路41号		2	30		1993	公寓住宅	民用	7.8	12	-3.07	13495	114.29061	30.485589
17	福星惠誉东	武汉	洪山	白沙洲	2014	多层	小高/白沙洲大道		2	30	465296	3670	公寓住宅	民用	7.4	12	-1.45	11964	114.26106	30.435131
18	万科金色城	武汉	洪山	白沙洲	2010	高层	超高/长征二路		2	35		2536	公寓住宅	民用	7.6	12	-0.73	15021	114.25537	30.482418
19	万科金色城	武汉	洪山	白沙洲	2010	高层	超高/白沙五路3		2	35		2396	公寓住宅	民用	7.5	12	-1.12	14177	114.25677	30.48486
20	清江锦城	武汉	洪山	白沙洲	2014	超高层	烽胜路55号		1	35		2273	公寓住宅	民用	7.6	12	0.28	16951	114.28955	30.476921
21	万科金色城	武汉	洪山	白沙洲	2010	高层	超高/武金堤路		2	35		1692	公寓住宅	民用	7.6	12	-1.2	16318	114.25192	30.483667
22	万科金色城	武汉	洪山	白沙洲	16-Nov	高层	超高/江州路		2	35		2668	公寓住宅	民用	7.7	12	-1.58	15526	114.26153	30.481073
23	融科花满庭	武汉	洪山	白沙洲	2015	小高层	超/白沙洲大道		2	40		1916	公寓住宅	民用	7.8	12	-0.67	14382	114.25527	30.450452
24	万科金色城	武汉	洪山	白沙洲	2016	高层	超高/江州路		2	35		1540	公寓住宅	民用	7.7	12	-1.76	14964	114.25458	30.486236
25	保利丰园	武汉	洪山	白沙洲	2018	高层	超高/烽胜路88号		1	30		2856	公寓住宅	民用	7.3	12	-0.55	12045	114.29653	30.467559
26	复地悦城	武汉	洪山	白沙洲	2015	高层	超高/烽胜路26号		1	35.5		2532	公寓住宅	民用	7.7	12	-1.38	17787	114.28454	30.493118
27	东原启城	武汉	洪山	白沙洲	2018	超高层	烽胜路66号	None		30		1726	公寓住宅	民用	7.4	12	0.09	16258	114.2868	30.470205

图 1 武汉房价 2022.csv

序号	小区名称	均价	区域	板块	地铁	地址	建筑年代	建筑类型	物业类型	物业公司	开发商	售楼部	房屋总数	经度_百度坐标	纬度_百度坐标	经度_WGS1984坐标	纬度_WGS1984坐标
2	武汉上城	15000	洪山	白沙洲		(洪山)武汉上城上城1号	2009年	高层	公寓住宅	武汉上城物业管理有限公司	武汉上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
3	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
4	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
5	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
6	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
7	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
8	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
9	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
10	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
11	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
12	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
13	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
14	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
15	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
16	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
17	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
18	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
19	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
20	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
21	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
22	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
23	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
24	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
25	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
26	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
27	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
28	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
29	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
30	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
31	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
32	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
33	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
34	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
35	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
36	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
37	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
38	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
39	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
40	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
41	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
42	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
43	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
44	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
45	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
46	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
47	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
48	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
49	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
50	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
51	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
52	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
53	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
54	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
55	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
56	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
57	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
58	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
59	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
60	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
61	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
62	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
63	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
64	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
65	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
66	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
67	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
68	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
69	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层	公寓住宅	保利上城物业管理有限公司	保利上城	售楼部	350户	114.17944	30.590072	114.1642751	30.58672985
70	保利上城	15000	洪山	白沙洲		(洪山)保利上城保利上城1号	2009年	高层									

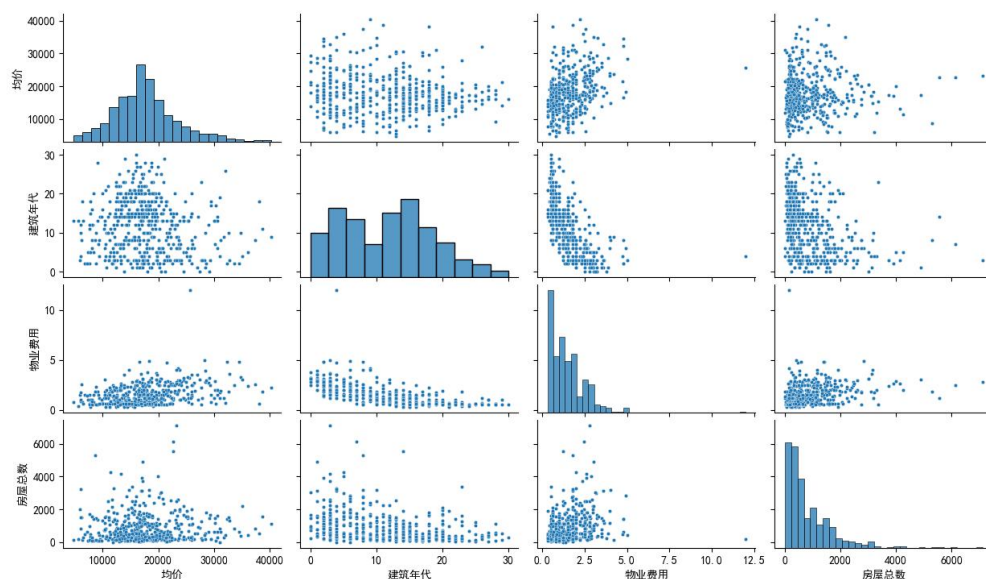
集中在 10000-20000/平方米，2022 年房价 6000-9000/平方米的小区数量明显增多，可能有以下原因：

首先，从政策层面来看，2018 年至今，国家已出台了一系列控制房价的政策，包括加强金融监管、加强土地供应、限制企业购房等。这些政策有利于遏制房价过快上涨，维护了市场稳定，也缓解了部分购房者的压力。

其次，从供需关系来看，武汉整体供应量较大，且近年来还有不少楼盘陆续入市。但与此同时，由于经济环境发生变化，某些行业（如汽车、船舶制造等）可能出现业务不景气、企业裁员等情况，从而导致个人购房意愿降低。

第三，从人口流动角度来看，武汉是以大学生为主体的移民城市，每年都会吸引大量学生前来求学或工作，这也促进了房屋需求。然而，在疫情影响下，某些外来人口可能选择暂时离开武汉，造成了人口流动量下降，也对房价影响不容忽视。

对于数据中的数值变量，可以探究其与房价的相关关系。经验上，楼栋数和房屋数都是比较独立的变量，但它们能反映房地产的规模情况，我们选择将房屋总数也纳入考量。即分析的相关性变量包括均价、建筑年代，物业费用和房屋总数。将这些变量的分布图放在对角线上，并两两之间绘制散点图，可以直观的得到相关图，如下所示：



在相关图上，建筑年代分布比较均匀，但物业费用和房屋总数分布都是正偏分布。从散点图来看，数值变量两两之间的相关性都比较弱。

代码如下：

```
plt.subplot(1,2,1)
plt.hist(df2['均价'].to_list(),bins=100,color='red')
plt.title("武汉2018年房价价格分布图")
plt.xlabel("price")
plt.ylabel("number")
plt.subplot(1,2,2)
plt.hist(df1['价格'].to_list(),bins=100,color='red')
plt.title("武汉2022年房价价格分布图")
plt.xlabel("price")
plt.ylabel("number")
plt.show()
```

3.2 武汉各地区房价分析

地区	均价	最大值	最小值	标准差
东西湖	12204.73768	50000.00000	3800.00000	4407.63403
新洲	6282.78065	19565.00000	3900.00000	1809.41797
武昌	21544.66720	70792.00000	1900.00000	7097.56000
汉南	6687.97884	15344.00000	1400.00000	1853.55785
汉阳	15468.04976	65350.00000	6560.00000	4952.72653
江夏	11020.83936	43016.00000	2000.00000	5318.11504
江岸	20189.62316	71476.00000	3044.00000	8065.15172
江汉	18440.59554	77605.00000	9000.00000	5380.48459
沌口开发区	12705.67111	44176.00000	5142.00000	5562.06151
洪山	17923.12415	67846.00000	2500.00000	6890.92858
蔡甸	16410.95500	34812.00000	6976.00000	3753.95798
硚口	8167.11784	30104.00000	2105.00000	3099.89641
青山	14480.22438	30720.00000	5000.00000	3892.81149
黄陂	8469.87017	31152.00000	2000.00000	3014.54874
	13571.1596	46568.4285	3951.9285	4649.9180

按各地区房价可得以上表格，整体表现为：中心城区（武昌、汉口）房价较高，周围郊区房价相对较低。在房价前五名的区域中，江岸区和汉江区属汉口，武昌、洪山属武昌，在武汉三镇中，仅新洲区房价相对较低。

以均价为标准绘制的区域房价图如下所示：

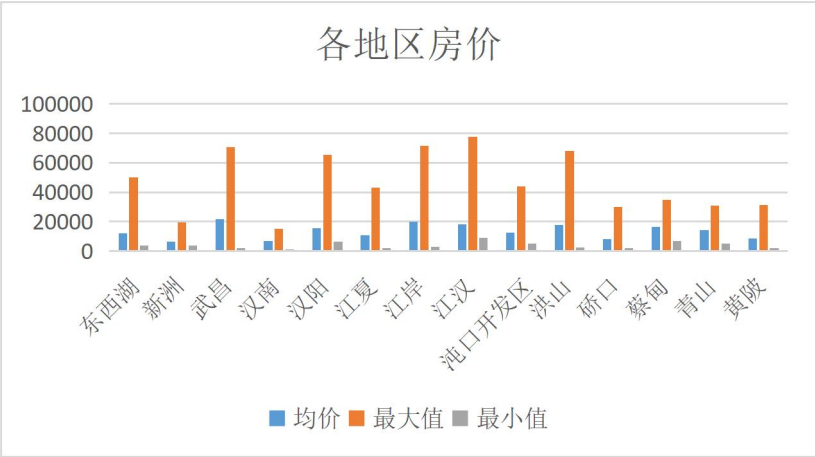


图 6 各地区房价均价、最大值、最小值

江岸区的方差较大，房价不集中，跨度剧烈。相反新洲区和汉南区房价较为集中，方差小。



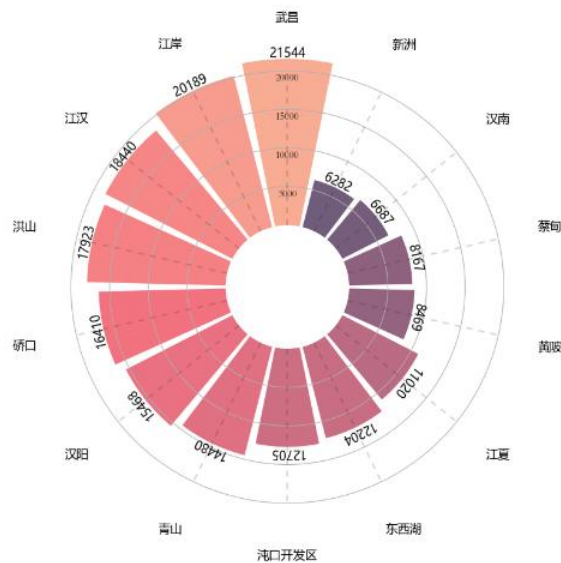
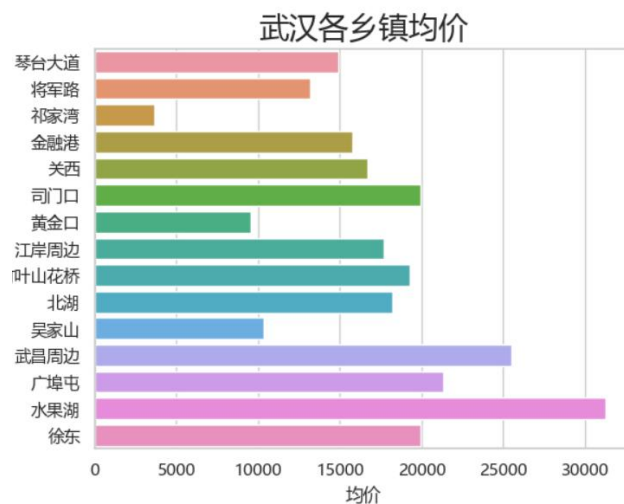


图 7 各地区房价均值排名图

### 3.3 武汉各乡镇房价分析

在所有 76 个乡镇中，随机选择 15 个展示其均价的比较：



部分代码如下：

```
chp2022 = HousePrice('武汉房价2022.csv', method='csv')
chp2018 = HousePrice('武汉.xlsx', method='excel')

#regional distribution
df_sorted = chp2022.gePriceStatistics(col='区县', stats=['mean', 'median'], sorted=True)
df_sorted2=chp2022.gePriceStatistics(col='乡镇', stats=['mean'], sorted=True)
drawPolar(df_sorted)
drawBarChart(df_sorted2)

def drawBarChart(dataframe):
    dataframe=dataframe.sample(15)
    sns.set(style='whitegrid')
    sns.barplot(x=dataframe['mean'], y=dataframe.index)
    plt.rcParams['font.sans-serif'] = ['Microsoft YaHei']
    plt.xlabel('均价')
    plt.title('武汉各乡镇均价', fontsize=20)
    plt.show()
```





部分代码如下：

```
# 找出相同的行（即相同的变量值）
common_vars = df1_indexed.merge(df2_indexed, left_index=True, right_index=True)
grouped_region = common_vars.groupby(['乡镇'])
# 计算数值变化，并将结果存储在一个新的DataFrame对象中
diffs = pd.DataFrame()
diffs['variable'] = grouped_region.groups
diffs['num_diff'] = grouped_region['价格'].mean() - grouped_region['均价'].mean()
diffs = diffs.sample(35)
#可视化价格变化
plt.rcParams['font.sans-serif']=['Microsoft YaHei']
plt.figure(figsize=(16,9))
plt.bar(diffs.index, diffs['num_diff'])
plt.title('武汉同一乡镇18-22年房价均值变化')
plt.xlabel('小区名称')
plt.ylabel('价格变化（元/m²）')
plt.xticks(rotation=45)
plt.show()
```

## 五、高档小区分布特点

定义房价在 20000 元/m<sup>2</sup>以上的为高档小区，因为根据第二部分的整体房价布局图可知，大部分房价集中在 10000-20000 元/m<sup>2</sup>，利用代码统计可得房价在 20000 元/m<sup>2</sup>以上的小区占比少于 20%。如果定价过高导致符合条件的小区数少，热力图的视觉显示效果也会减弱，故选择房价在 20000 元/m<sup>2</sup>以上的为高档小区。

以房价为权重，在对应的地理坐标上添加不同深浅的颜色，就可以得到房价的热力图。调用 Folium 的热力图功能，我们可以在地图上直观地发现高房价的集中地区：

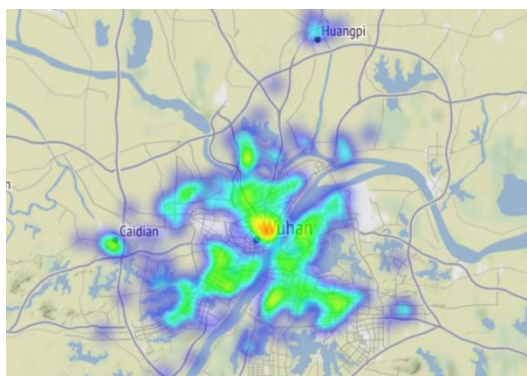


图 10 武汉市总体房价热力图

房价越高的地区颜色越深，房价越低的地区颜色越浅。在此图中，江滩附近的房价较高，呈现出显著的红色。可以很清楚的看到，几乎所有高房价的房源均集中在武昌、洪山、江岸区（即沿江两侧）；而低房价区域则多分布在武汉周边地区。



图 11 15000 元以上的房价热力图（整体）

我们可以得到和上面相同的结论：长江两边的江岸、武昌、洪山区房价普遍较高，而武汉周边地区房价普遍较低。我们也可以对房价进行多样的筛选，在地图上显示需要的信息，例如，选择均价 15000 元以上的房地产信息后，热力图集在城区中：

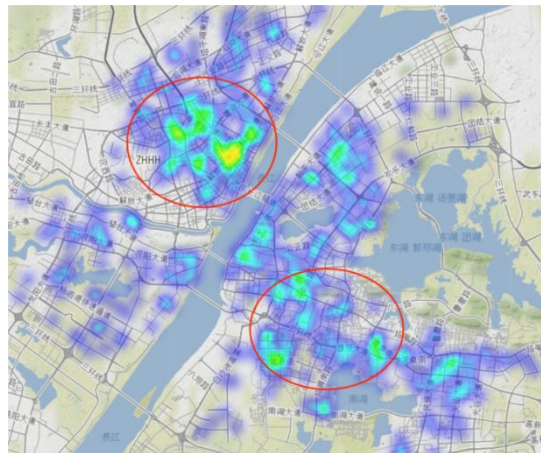


图 12 15000 元以上的房价热力图（局部）

将房价前 30 的地产在地图上显示：

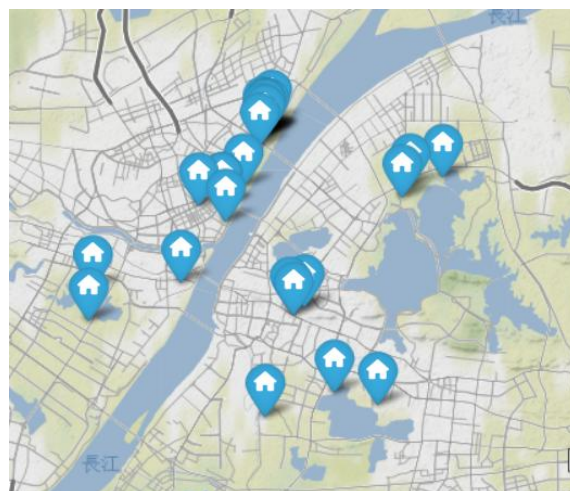


图 13 2022 年武汉房价前 30 的地产

热力图放大后的结果，可以发现高房价地区的多个中心。其中最主要的两个分别位于江滩附近和洪山区中心。热力图能有效地反应房价随地区变化的梯度信息，因此，在宏观角度看效果比较好。但对于本身的价格情况，热力图则无能为力，过度放大热力图，反而不能发现有效地信息，因为对于单独的房价，其梯度为 0，在热力图上都是同样的点。

部分代码截图：

```
#High-end community heat map
geodata = chp2022.getGeodata(price_filter=price_filter,区县=['洪山','武昌','江岸','江滩','汉阳','东西湖','蔡甸','青山'])
drawHeatMap(geodata)

#extrema locations
price_extrema=chp2022.getPriceExtrema(ascending=False,number=30)
drawLocation(price_extrema)
```

```
def drawHeatMap(geodata, filepath='HeatMap.html', center=[30.5, 114], zoom=10):
    map_osm = folium.Map(location=center, zoom_start=zoom, tiles='Stamen Terrain')
    print(map_osm)
    HeatMap(geodata, radius=10).add_to(map_osm)
    map_osm.save(filepath)
    webbrowser.open(filepath)

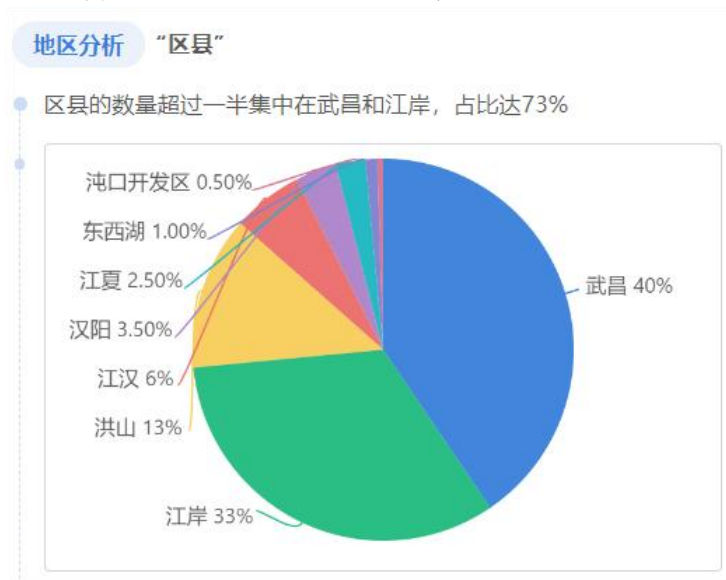
def drawLocation(price_extrema):
    map_osm = folium.Map(location=[30.5, 114], zoom_start=10, tiles='Stamen Terrain')
    print(map_osm)
    tooltip='click for info'
    lat=price_extrema['wgs84_lat'].values
    lon=price_extrema['wgs84_lng'].values
    names=price_extrema['名称'].values
    prices=price_extrema['价格'].values
    for i in range(len(lat)):
        msg=names[i]+'\\n'+price[i].format(prices[i])
        if names[i] in ['宏盛公寓', '广埠屯小区', '珞珈山公寓', '卓刀泉科技公寓']:
            folium.Marker([lat[i], lon[i]], popup=msg, tooltip=tooltip, icon=folium.Icon(color='red', icon='home', prefix='fa')).add_to(m
        else:
            folium.Marker([lat[i], lon[i]], popup=msg, tooltip=tooltip, icon=folium.Icon(icon='home', prefix='fa')).add_to(map_osm)
    map_osm.save('location.html')
    webbrowser.open('location.html')
```

## 六、高档小区与名称的相关分析

重新定义房价全市排名前 200 的为高档小区。

影响房价的因素有很多，哪些因素会与房价有着比较强烈的相关性是我们需要探讨的问题。这里根据经验，影响房价的因素主要是：地理位置、房屋结构、管理费、建造年代等等。本次分析从小区名称下手，分析其与高档小区房价的相关性。

也可以利用机器自动判读文本信息，利用 jieba 库的中文分词功能，统计高价房产的地址信息，得到词频表，一些有用的信息如下：



关键词	计数	关键词	计数
武昌	81	江汉	12
江岸	66	汉阳	7
洪山	26	江夏	5

表 1 武汉房价前 200 房产的区县关键词

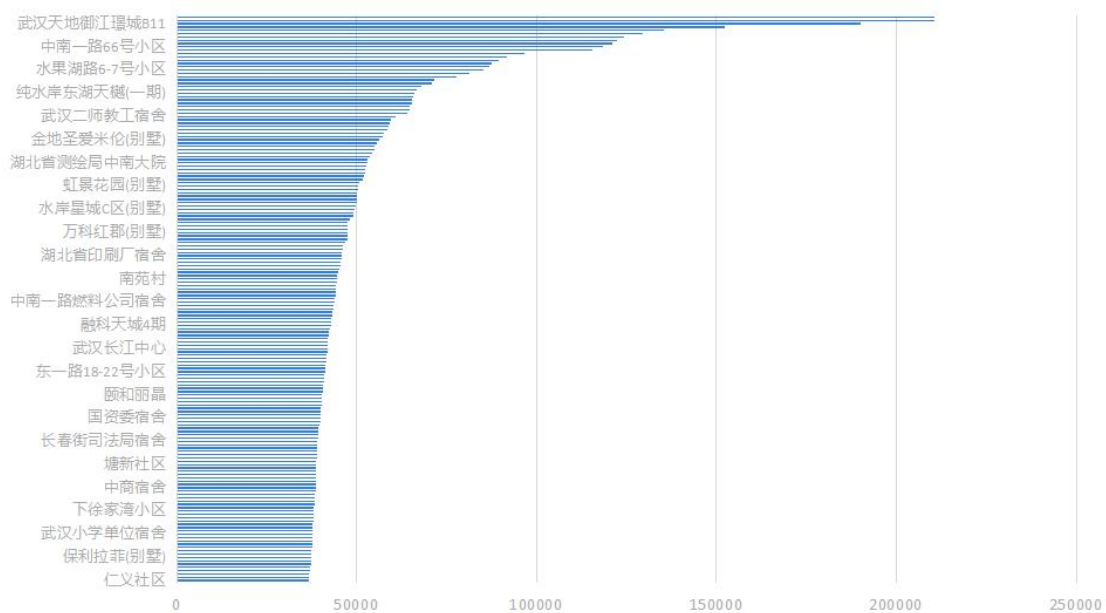


图 14 房价排名前 200 小区价格累计图

小区	193	新村	12	天运	12
花园	57	宿舍	12	尚街	12
社区	24	朝城	12	奥科苑	12
新洲	22	兴铁里	12	黄茂里	12
自建房	20	丰兆里	12	益华	12
新光	16	金寓	12	实轩	12
常乐	16	汉泰苑	12	中庭	12
新城	14	木天署	12	安居	12
永平	13	良兴小苑	12	永安	12
新家园	13	资生里	12	百佳	11
长丰	13	湖雅苑	12	书香	11
阳光	13	岳府	12	新风	11
				勤政	11

统计得到的词频表和我们先前的分析也是类似的，武昌、江岸、洪山区的房价较高；江岸、江汉、东湖等高频词说明江景、湖景房价格较高。词频表还可以给出一些具体的信息，如阳光、勤政、永安、常乐带有积极向上倾向的小区更受人们青睐等。

部分代码如下：

```
# 加载小区数据
df = pd.read_csv("武汉房价2022.csv")
sort_price = df.sort_values('价格')[:200]
# 将小区名称列转换为列表
name_list = sort_price['名称'].tolist()
# 分词并提取关键词
keywords_dict = {}
for name in name_list:
    keywords = jieba.analyse.extract_tags(name, topK=10, withWeight=True)
    for keyword, weight in keywords:
        if keyword in keywords_dict:
            keywords_dict[keyword] += weight
        else:
            keywords_dict[keyword] = weight
# 按关键词权重排序并输出前100名
top_keywords = sorted(keywords_dict.items(), key=lambda x:x[1], reverse=True)[:100]
for keyword, count in top_keywords:
    print(f"{keyword} {round(count)} ")
```



```

# 加载小区数据
df = pd.read_excel("wuhan.xlsx")
sort_price = df.sort_values('价格')[:100]
# 定义高档小区的名词、形容词词性列表
high_grade_pos = ['n', 'ns', 'nt', 'nz', 'a']
# 创建正则表达式匹配规则，过滤掉一些描述非小区本身属性的用词（如“城市花园”、“xx天香”等）
pattern = re.compile(r"((花园)|(景观)|(^万)|(^中心)|(^明珠)|(^翡翠)|(^瑞)|(^招商)).*")
# 预处理小区名称
def process_name(name):
    # 去除特殊符号、空格和数字
    name = re.sub(r'[\W_]+', '', name)
    # 分词并获取每个词的词性
    words = pseg.cut(name)
    # 仅保留名词和形容词，并返回单个字符串
    return '/'.join([word.word for word in words if word.flag in high_grade_pos])

# 对每个小区名称进行预处理，并进行词频统计
name_freq = sort_price['名称'].apply(lambda x: process_name(str(x))).value_counts()
total_count = len(sort_price)
print(f"{'名字':25} 数量 百分比")
for i, (name, count) in enumerate(name_freq.iteritems()):
    if i >= 20:
        break
    proportion = count / total_count
    print(f"{'name':25} {'count':6} {'proportion:.2%}")

```

## 七、2022 年的购房建议

考虑到一万元的月收入，我们认为价格低于 15000 元的房子比较适合。另外，为满足上班的需求，位置不应离武汉大学过远或者有直达地铁。因此，对房产的筛选可以按价格和距离两个维度同时进行，距离由经纬度之差得到，因此只是直线距离，没有考虑实际路径的问题。如果要进行更细致的分析，需要交通和道路信息，这里不做讨论。

我们以 15000 元为价格阈值上限，以 5km 为半径或者在二号线地铁有站点，筛选了一些符合要求的房产。考虑到表格中的文本信息，我们尽可能选择接近地铁 2 号线的并排除小区名称中带有“某单位宿舍”字眼的，剩下的约 13 处房产分布如下：

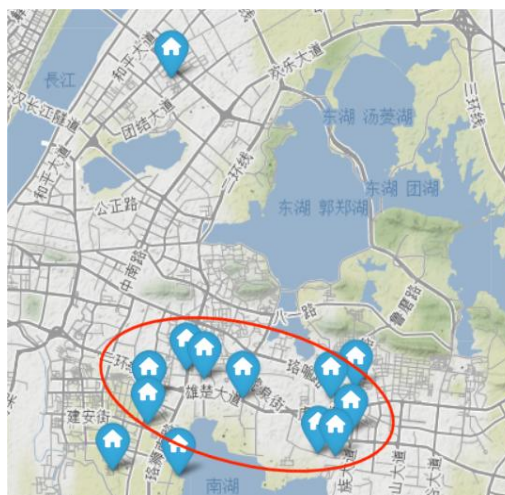


图 15 推荐房产

部分代码如下：

```

#select houses
selected=chp2022.filter(15000,30.541093,114.3607347)
drawLocation(selected)

def filter(self,price_threshold,lat,lon):
    selected=self.df[self.df['价格']<price_threshold]
    selected=selected.drop(selected[selected['名称'].str.contains(pat='宿舍',regex=False)].index)
    selected=selected.drop(selected[selected['名称'].str.contains(pat='酒店',regex=False)].index)
    selected=selected.drop(selected[selected['名称'].str.contains(pat='国际',regex=False)].index)
    selected=selected[self.df['wgs84_lng']-lon<0.05]
    selected=selected[self.df['wgs84_lng']-lon>-0.05]
    selected=selected[self.df['wgs84_lat']-lat<0.05]
    selected=selected[self.df['wgs84_lat']-lat>-0.05]
    selected=selected.dropna()
    #selected=selected[selected['地铁'].str.contains('2')]
    return selected

```