

# 第一次作业你的报告题目

Code ▾

雷婷lititia

2024-10-31

- 1 你的主要发现
- 2 数据介绍
- 3 数据概览
- 4 探索性分析
  - 4.1 变量房屋单价的数值描述与图形
  - 4.2 变量房屋总价的数值描述与图形
  - 4.3 变量不同地区房屋数量的数值描述与图形
  - 4.4 探索问题1:房屋单价与房屋面积的关系
  - 4.5 探索问题2:房屋单价与房屋装修的关系
  - 4.6 探索问题3:房屋单价与房型的关系

## 1 你的主要发现

- 1、集中趋势：房屋单价整体基本符合正态分布，集中部分： a、房屋单价在1-2万区间集中 b、在100平左右时也集中 c、简装房屋相对其他几个装修价格更集中，集中在1万3左右
- 2、影响因素：房屋单价会受装修风格、楼型等因素影响出现波动 a、装修风格：精装的房屋单价要高于其他集中装修； b、楼型：板塔结合房屋单价>塔楼房屋单价>板楼房屋单价

## 2 数据介绍

本报告链家数据获取方式如下：

报告人在2023年9月12日获取了链家武汉二手房网站 (https://wh.lianjia.com/ershoufang/)数据。

- 链家二手房网站默认显示100页，每页30套房产，因此本数据包括3000套房产信息；
- 数据包括了页面可见部分的文本信息，具体字段及说明见作业说明。

**说明：**数据仅用于教学；由于不清楚链家数据的展示规则，因此数据可能并不是武汉二手房市场的随机抽样，结论很可能有很大的偏差，甚至可能是错误的。

## 3 数据概览

数据表 (lj)共包括property\_name, property\_region, price\_ttl, price\_sqm, bedrooms, livingrooms, building\_area, directions1, directions2, decoration, property\_t\_height, property\_height, property\_style, followers, near\_subway, if\_2y, has\_key, vr等18个变量,共3000行。表的前10行示例如下：

武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	directions1	directions2	decoration	prop
南湖名都A区   南湖沃尔	237.									

各变量的简短信息：

## Rows: 3,000	
## Columns: 18	
## \$ property_name	<fct> 南湖名都A区, 万科紫悦湾, 东立国际, 新都汇, 保利城一...
## \$ property_region	<fct> 南湖沃尔玛, 光谷东, 二七, 光谷广场, 团结大道, 庙山, ...
## \$ price_ttl	<dbl> 237.0, 127.0, 75.0, 188.0, 182.0, 122.0, 99.0, 193.8...
## \$ price_sqm	<dbl> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163...
## \$ bedrooms	<dbl> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4...
## \$ livingrooms	<dbl> 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2...
## \$ building_area	<dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19, ...
## \$ directions1	<fct> 南, 南, 南, 北, 东南, 南, 南, 南, 南, 南, 南, 南, 东...
## \$ directions2	<fct> 北, NA, NA, 东, NA, 北, NA, 北, 北, 北, 北, NA, 西南...
## \$ decoration	<fct> 精装, 精装, 简装, 精装, 简装, 精装, 简装, 其他, 简装...
## \$ property_t_height	<dbl> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, ...
## \$ property_height	<fct> 中, 中, 低, 高, 中, 低, 低, 中, 低, 低, 高, 高, 中, ...
## \$ property_style	<fct> 塔楼, 板楼, 塔楼, 塔楼, 板塔结合, 板楼, 板楼, 板塔结...
## \$ followers	<dbl> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, ...
## \$ near_subway	<fct> 近地铁, NA, 近地铁, 近地铁, NA, NA, 近地铁, 近地铁, ...
## \$ if_2y	<fct> NA, 房本满两年, NA, 房本满两年, 房本满两年, 房本满两...
## \$ has_key	<fct> 随时看房, 随时看房, 随时看房, 随时看房, 随时看房, 随...
## \$ vr	<fct> NA, VR看装修, NA, NA, VR看装修, NA, VR看装修, NA, NA...

各变量的简短统计：

##	property_name	property_region	price_ttl	price_sqm	
##	东立国际	: 22 白沙洲 : 167	Min. : 10.6	Min. : 1771	
##	保利中央公馆	: 16 盘龙城 : 126	1st Qu.: 95.0	1st Qu.:10799	
##	朗诗里程	: 16 四新 : 116	Median : 137.0	Median :14404	
##	恒大名都	: 15 光谷东 : 112	Mean : 155.9	Mean :15148	
##	阳光100大湖第	: 15 金银湖 : 97	3rd Qu.: 188.0	3rd Qu.:18211	
##	保利城一期	: 13 后湖 : 86	Max. :1380.0	Max. :44656	
##	(Other)	:2903 (Other):2296			
##	bedrooms	livingrooms	building_area	directions1	directions2
##	Min. :1.000	Min. :0.000	Min. : 22.77	南 :2454	北 :1189
##	1st Qu.:2.000	1st Qu.:1.000	1st Qu.: 84.92	东南 : 281	南 : 66
##	Median :3.000	Median :2.000	Median : 95.55	东 : 98	西 : 25
##	Mean :2.695	Mean :1.709	Mean :100.87	北 : 68	东南 : 15
##	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:117.68	西南 : 57	西南 : 12
##	Max. :7.000	Max. :4.000	Max. :588.66	西 : 19	(Other): 21
##				(Other): 23	NA's :1672
##	decoration	property_t_height	property_height	property_style	
##	简装: 634	Min. : 2.00	低 : 816	板楼 :1781	
##	精装:1757	1st Qu.:11.00	高 : 906	板塔结合: 615	
##	毛坯: 436	Median :27.00	中 :1218	平房 : 5	
##	其他: 173	Mean :24.22	NA's: 60	塔楼 : 527	
##		3rd Qu.:33.00		暂无数据: 72	
##		Max. :62.00			
##					
##	followers	near_subway	if_2y	has_key	
##	Min. : 0.000	VR看装修 : 2	房本满两年:1264	随时看房:2525	
##	1st Qu.: 1.000	近地看 : 1	NA's :1736	近地铁 : 7	
##	Median : 3.000	近地铁 :1554		VR看装修: 4	
##	Mean : 6.614	珞狮南 : 1		藏龙岛 : 1	
##	3rd Qu.: 6.000	太子湖1号: 1		关山大道: 1	
##	Max. :262.000	NA's :1441		(Other) : 4	
##				NA's : 458	
##					
##	vr				
##	VR看装修	:2084			
##	VR\7甸城\5\8c:	1			
##	随时看\6\88	: 1			
##	保利拉\8\8f	: 1			
##	育才花\6\al	: 1			
##	(Other)	: 6			
##	NA's	: 906			

可以看到：

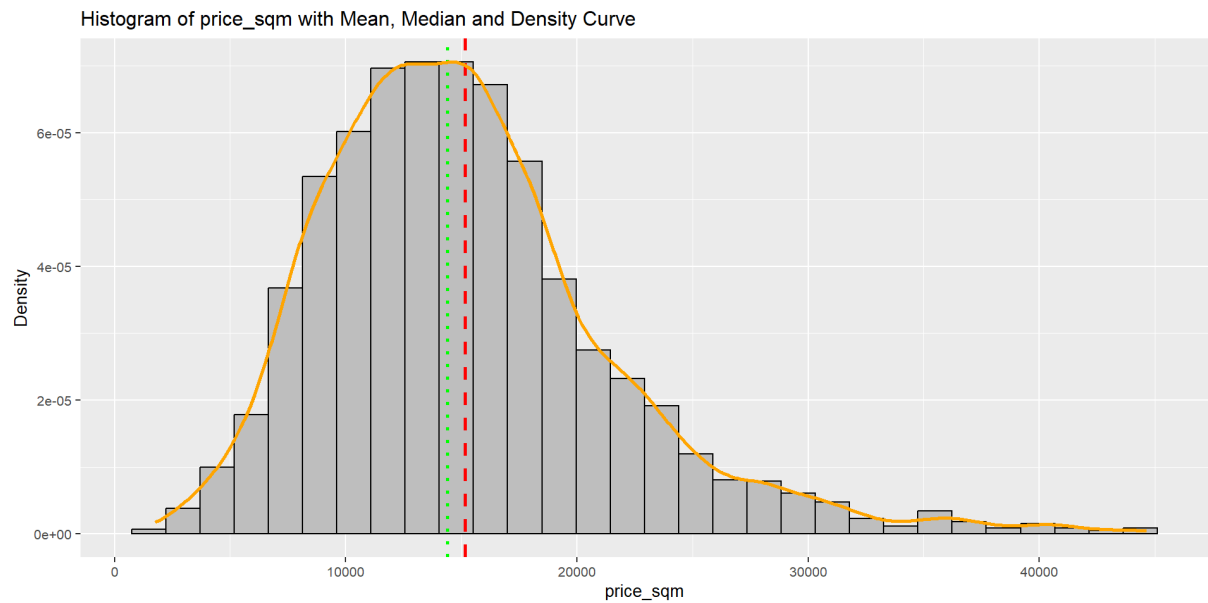
- 直观结论1：房屋单价、房屋总价、房屋面积在1st Qu、Median、Mean、3rd Qu数据处是相对集中的，但极端值相差较大
- 直观结论2：南北朝向房屋数量远高于其他朝向房屋
- 直观结论3：白沙洲、盘龙城、四新这类偏远城区房屋数量较多

##	# A tibble: 1 × 35
##	price_ttl_mean price_ttl_median price_ttl_sd price_ttl_min price_ttl_max
##	<dbl> <dbl> <dbl> <dbl> <dbl>
##	1 156. 137 95.5 10.6 1380
##	# 30 more variables: price_sqm_mean <dbl>, price_sqm_median <dbl>,
##	price_sqm_sd <dbl>, price_sqm_min <dbl>, price_sqm_max <dbl>,
##	bedrooms_mean <dbl>, bedrooms_median <dbl>, bedrooms_sd <dbl>,
##	bedrooms_min <dbl>, bedrooms_max <dbl>, livingrooms_mean <dbl>,
##	livingrooms_median <dbl>, livingrooms_sd <dbl>, livingrooms_min <dbl>,
##	livingrooms_max <dbl>, building_area_mean <dbl>,
##	building_area_median <dbl>, building_area_sd <dbl>, ...

##		vars	n	mean	sd	median	trimmed	mad	min
##	property_name*	1	3000	631.53	391.90	622.50	625.14	519.65	1.00
##	property_region*	2	3000	44.70	25.43	48.00	45.03	29.65	1.00
##	price_ttl	3	3000	155.86	95.55	137.00	142.48	66.72	10.60
##	price_sqm	4	3000	15148.49	6323.18	14404.00	14579.25	5465.60	1771.00
##	bedrooms	5	3000	2.69	0.73	3.00	2.68	0.00	1.00
##	livingrooms	6	3000	1.71	0.47	2.00	1.77	0.00	0.00
##	building_area	7	3000	100.87	30.38	95.54	99.59	23.27	22.77
##	directions1*	8	3000	4.78	0.95	5.00	4.93	0.00	1.00
##	directions2*	9	1328	1.45	1.38	1.00	1.01	0.00	1.00
##	decoration*	10	3000	2.05	0.76	2.00	1.99	0.00	1.00
##	property_t_height	11	3000	24.22	12.45	27.00	23.84	11.86	2.00
##	property_height*	12	2940	2.14	0.82	2.00	2.17	1.48	1.00
##	property_style*	13	3000	1.83	1.22	1.00	1.63	0.00	1.00
##	followers	14	3000	6.61	15.22	3.00	3.53	2.97	0.00
##	near_subway*	15	1559	3.00	0.09	3.00	3.00	0.00	1.00
##	if_2y*	16	1264	1.00	0.00	1.00	1.00	0.00	1.00
##	has_key*	17	2542	6.98	0.32	7.00	7.00	0.00	1.00
##	vr*	18	2094	5.01	0.24	5.00	5.00	0.00	1.00
##		max	range	skew	kurtosis	se			
##	property_name*	1345.00	1344.00	0.09	-1.22	7.16			
##	property_region*	87.00	86.00	-0.14	-1.14	0.46			
##	price_ttl	1380.00	1369.40	2.75	16.12	1.74			
##	price_sqm	44656.00	42885.00	1.08	2.03	115.44			
##	bedrooms	7.00	6.00	0.14	1.64	0.01			
##	livingrooms	4.00	4.00	-0.99	-0.18	0.01			
##	building_area	588.66	565.89	2.08	23.64	0.55			
##	directions1*	8.00	7.00	-1.49	7.62	0.02			
##	directions2*	8.00	7.00	3.03	8.15	0.04			
##	decoration*	4.00	3.00	0.69	0.53	0.01			
##	property_t_height	62.00	60.00	0.05	-0.80	0.23			
##	property_height*	3.00	2.00	-0.26	-1.47	0.02			
##	property_style*	5.00	4.00	1.22	-0.01	0.02			
##	followers	262.00	262.00	6.90	68.17	0.28			
##	near_subway*	5.00	4.00	-5.99	394.01	0.00			
##	if_2y*	1.00	0.00	NaN	NaN	0.00			
##	has_key*	9.00	8.00	-15.00	249.61	0.01			
##	vr*	11.00	10.00	11.66	371.99	0.01			

4 探索性分析

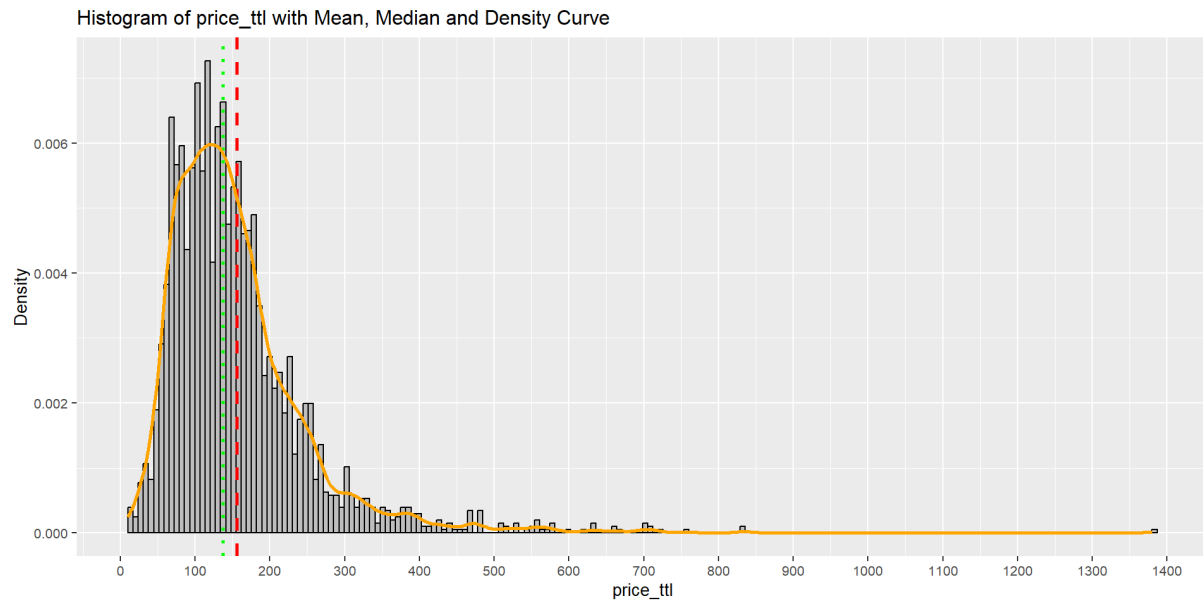
4.1 变量房屋单价的数值描述与图形



发现:

- 发现1:房屋单价基本符合正态分布
- 发现2: 房屋单价在1-2万区间较为集中

## 4.2 变量房屋总价的数值描述与图形



发现：

- 发现1:房屋总价近似正态分布
- 发现2: 房屋总价在100万左右最为集中

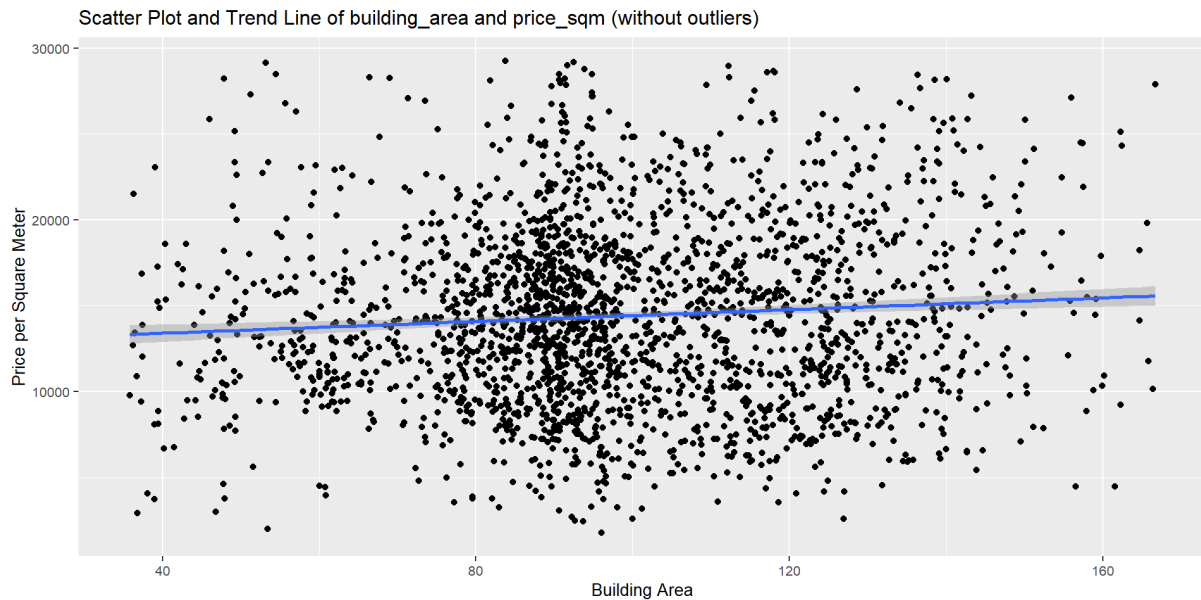
### 4.3 变量不同地区房屋数量的数值描述与图形



发现:

- 发现1：白沙洲、盘龙城、四新这类偏远城区房屋数量较多

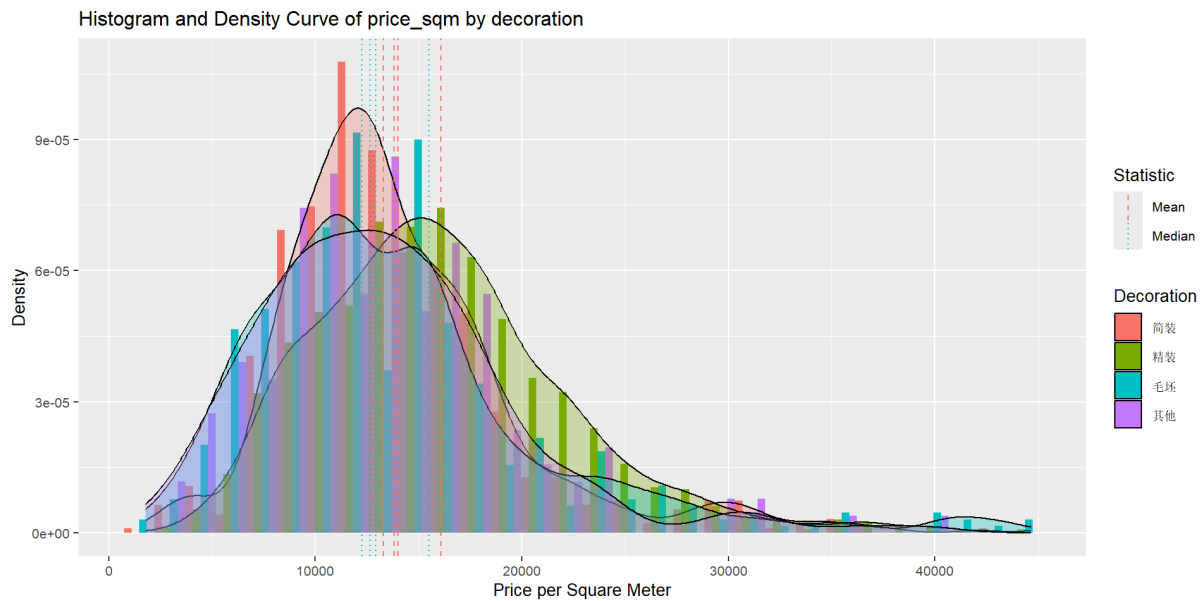
#### 4.4 探索问题1:房屋单价与房屋面积的关系



发现:

- 发现1: 房屋单价在建筑面积较大或较小时较为离散, 在100平左右时集中
- 发现2: 不论面积, 房屋都基本在1-2万区间集中

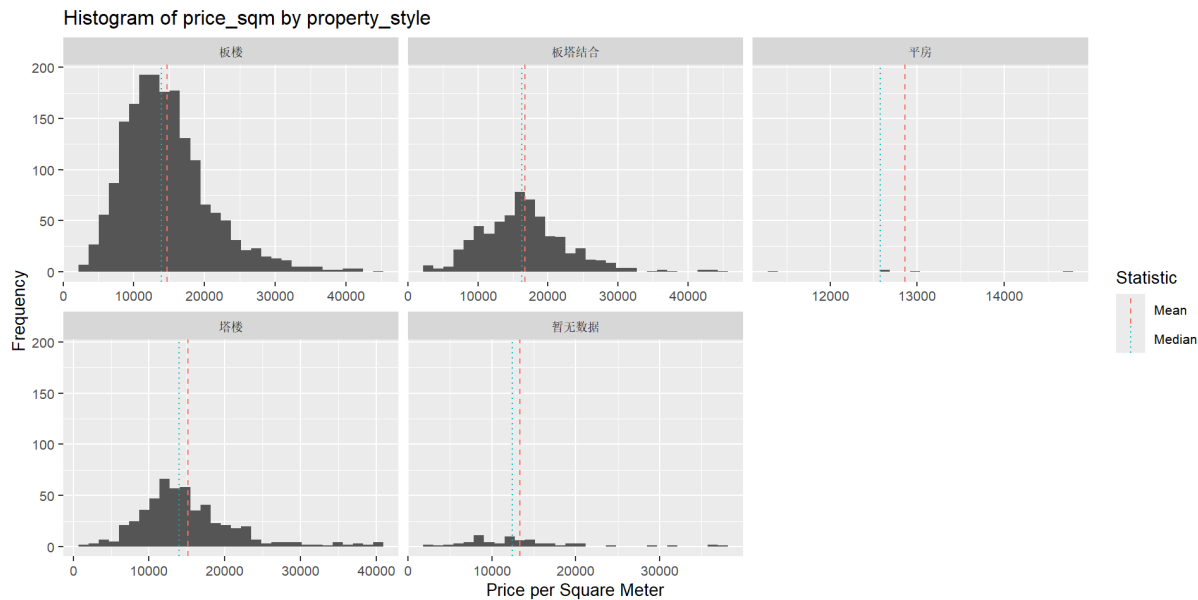
#### 4.5 探索问题2:房屋单价与房屋装修的关系



发现:

- 发现1: 精装的房屋单价要高于其他集中装修
- 发现2: 简装房屋相对其他几个装修价格更集中, 集中在1万3左右

4.6 探索问题3:房屋单价与房型的关系



发现:

- 发现1: 各楼型中房屋单价都是均值略高于中位数, 即都存在极端值拉高均价的情况
- 发现2: 房屋单价: 板塔结合>塔楼>板楼

发现总结

-结合2023年9月12日获取了[链家武汉二手房网站]数据分析武汉二手房信息发现, 价格上: 武汉二手房在房屋单价、房屋总价上有明显的集中性, 单价集中在1-2万区间, 总价集中在100万左右; 朝向上: 房屋主朝南、次朝北的房屋呈超峰态分布; 楼型上: 板楼呈峰态分布; 地域上: 白沙洲、盘龙城、四新这类偏远城区房屋频率高于中心城区。总体来说武汉二手房呈现以南北朝向、板楼房型、远程区、单价1-2万、总价100万左右的特征的房源类型流通。-分析武汉二手房单价与房屋面积、房屋装修和楼型相关性发现, 房屋单价受建筑面积影响不大, 主要在1-2万单价、面积100平左右时集中; 但在装修上, 不同装修对价格有一定影响, 精装的房屋单价要高于其他集中装修, 简装房屋相对其他几个装修价格峰态更明显, 集中在1万3的单价左右; 在建筑楼型上房屋单价: 板塔结合>塔楼>板楼。总体来说武汉二手房单价受装修和楼型影响波动, 受面积影响较小。

...