

# 商务统计第一次作业

程迪 2023281051004

## 目录

1 你的主要发现	1
2 数据介绍	2
3 数据概览	2
4 探索性分析	5
4.1 不同建筑形式与房屋每平单价的关系 . . . . .	5
4.2 楼栋总层数与房价的影响，根据不同楼层分面 . . . . .	6
4.3 是否近地铁、能否 vr 看装修对关注人数的影响 . . . . .	8
4.4 判断是否满 2 年对二手房总价影响是否显著 . . . . .	9

## 1 你的主要发现

1. 挂售二手房的价格受多种因素影响，包括房屋建筑形式、房屋楼层总数以及分布
2. 区域在售二手房的数量会对该区域的平均单价有影响
3. 在二手房平台房屋关注人数受房屋是否近地铁、能否 vr 看装修影响

## 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 `ncol(lj)` 个变量, 共 `nrow(lj)` 行。表的前 10 行示例如下：

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
```

表 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126
万科紫悦湾	光谷东	127.0	14613	3	2	86
东立国际	二七	75.0	15968	1	1	46
新都汇	光谷广场	188.0	15702	3	2	119
保利城一期	团结大道	182.0	17509	3	2	103
加州橘郡	庙山	122.0	10376	3	2	117
省建筑五公司西区	光谷广场	99.0	12346	2	1	80
保利上城东区	白沙洲	193.8	16336	3	2	118
石化大院	中南丁字桥	325.0	32631	4	1	99
阳光花园	杨汊湖	192.0	17403	3	2	110

##	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			
##	其他: 173	Min. : 2.00	中 :1218	塔楼	: 527		
##	毛坯: 436	1st Qu.:11.00	低 : 816	平房	: 5		
##	简装: 634	Median :27.00	高 : 906	暂无数据:	72		
##	精装:1757	Mean :24.22	NA's: 60	板塔结合:	615		
##		3rd Qu.:33.00		板楼	:1781		
##		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	太子湖1号: 1	NA's :1736	近地铁 : 7			
##	Median : 3.000	珞狮南 : 1		VR看装修: 4			
##	Mean : 6.614	近地看 : 1		世纪花园: 1			

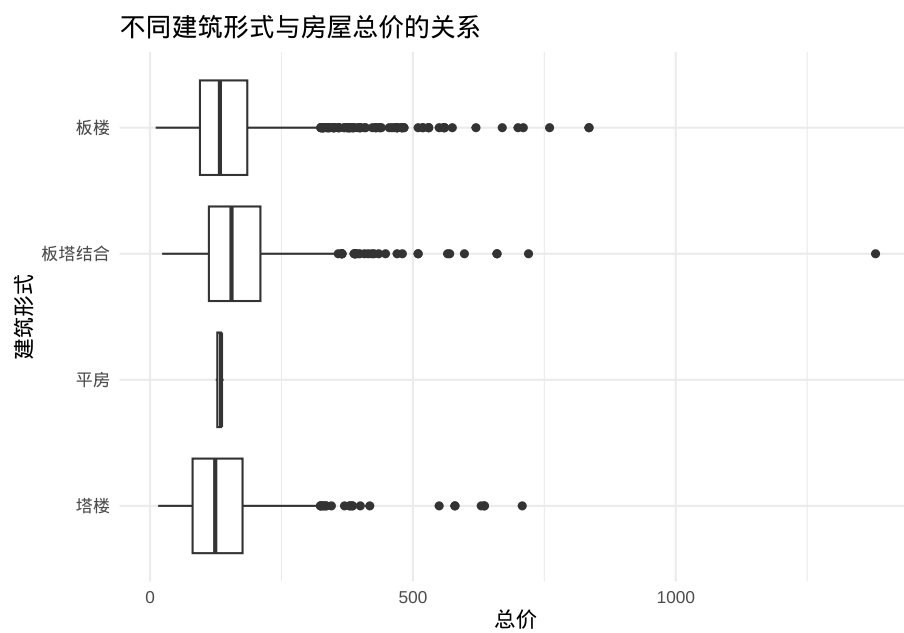
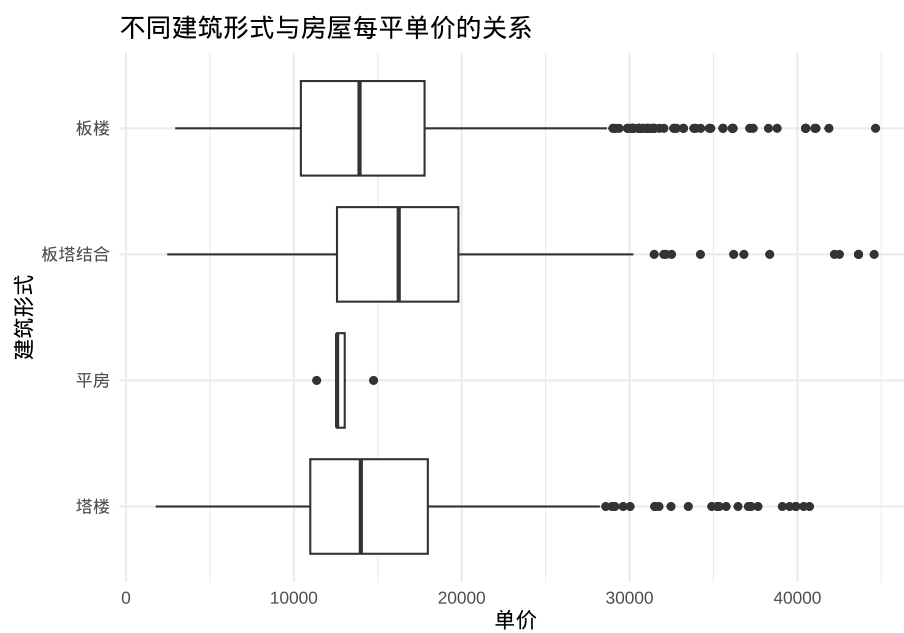
```
## 3rd Qu.: 6.000    近地铁    :1554    仁厚社区: 1
## Max.    :262.000  NA's      :1441    (Other) : 4
##                                     NA's    : 458
##                                     vr
## VR看装修          :2084
## VR\甸城\城: 1
## 保利拉\      : 1
## 塔子湖        : 1
## 江景湾        : 1
## (Other)       : 6
## NA's          : 906
```

可以看到:

- 直观结论 1: 市面上挂售二手房房屋, 以精装房、板楼为主, 南北朝向、中间楼层居多; 平均房屋面积约 100 平米, 二居、三居居多, 平均单价 1.5 万元/平; 房产满 2 年房屋过半, 说明以上主流房屋流通较高。
- 直观结论 2: 经中介渠道挂售的二手房, 大部分满足随时看房及 vr 看装修的模式, 客户了解房屋状况的方法多样, 关注人数平均 6 人。
- 直观结论 3: 挂售房屋超过 100 套的区域有白沙洲、盘龙城、四新、光谷东, 这几个区域房屋体量较大

## 4 探索性分析

### 4.1 不同建筑形式与房屋每平单价的关系

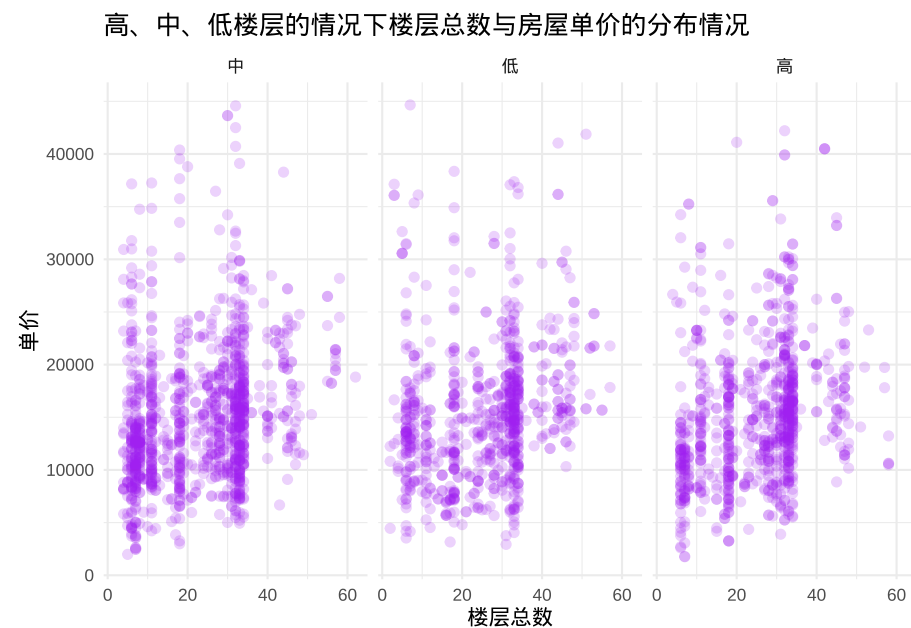


发现：

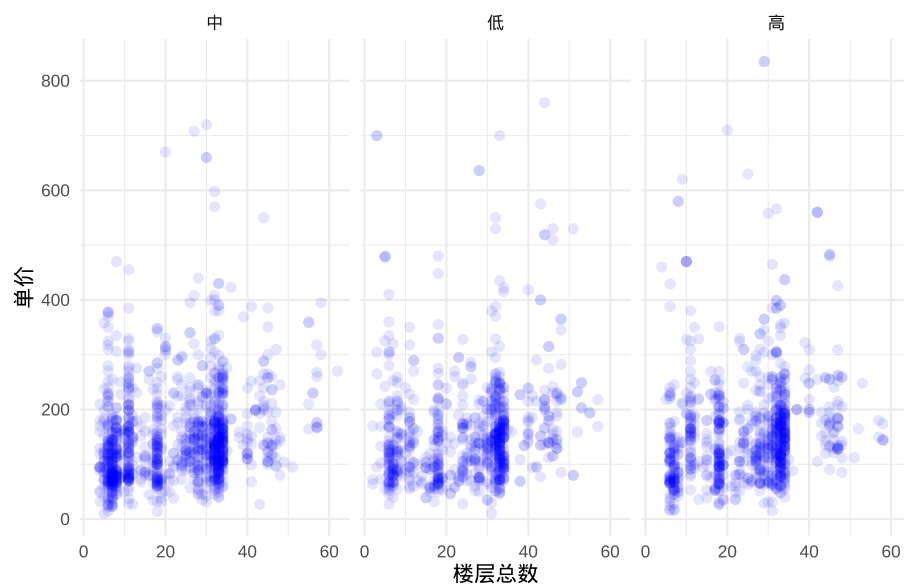
- 发现 1: 板楼结合的建筑房屋单价及房屋总价均超出其他形式的房屋，价格普遍偏高
- 发现 2: 从样本量看，平房样本数量最少，说明该类房屋二手市场上流通较少

## 4.2 楼栋总层数与房价的影响，根据不同楼层分面

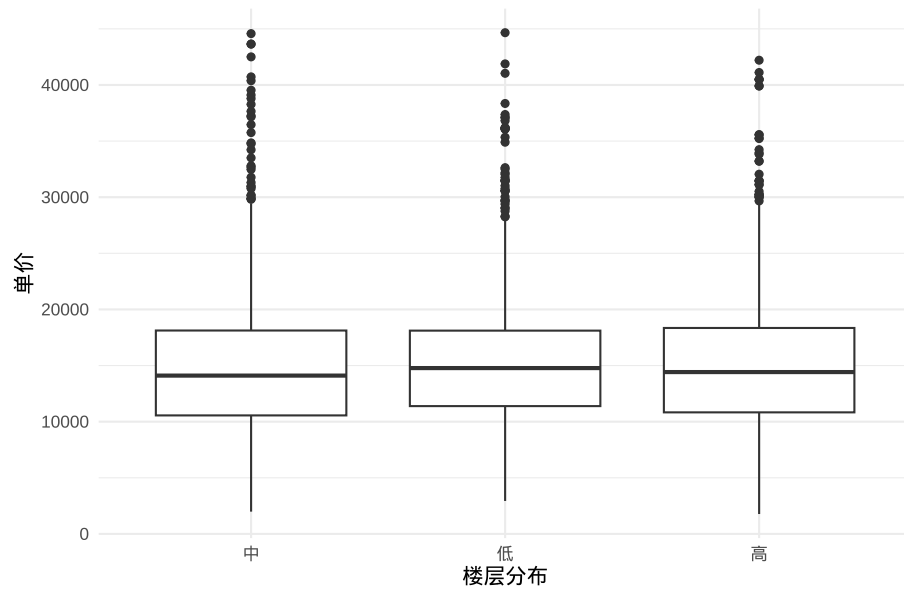
```
## [1] "33" "34" "18" "32" "6"
```



高、中、低楼层的情况下楼层总数与房屋总价的分布情况



高、中、低楼层的情况下楼层分布与房屋单价的分布情况



发现：

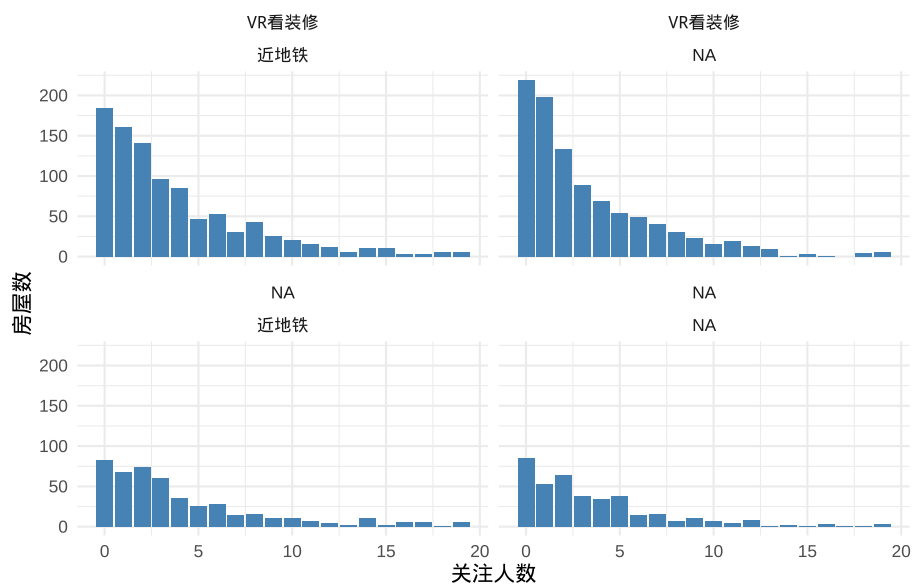
- 发现 1: 市面上在售二手房，总楼层数频率出现最多的前 5 个分别

为”33” “34” “18” “32” “6”。在图像中反应出颜色更深。

- 发现 2: 无论总楼层数是多少, 低楼层的房屋价格区间范围较小, 平方单价略偏高, 说明议价能力较强, 价格波动幅度不大。

### 4.3 是否近地铁、能否 vr 看装修对关注人数的影响

是否近地铁、能否vr看房对关注人数的影响



```
##
## Chi-squared test for given probabilities
##
## data: table(new$near_subway, new$followers)
## X-squared = 10491, df = 80, p-value < 2.2e-16

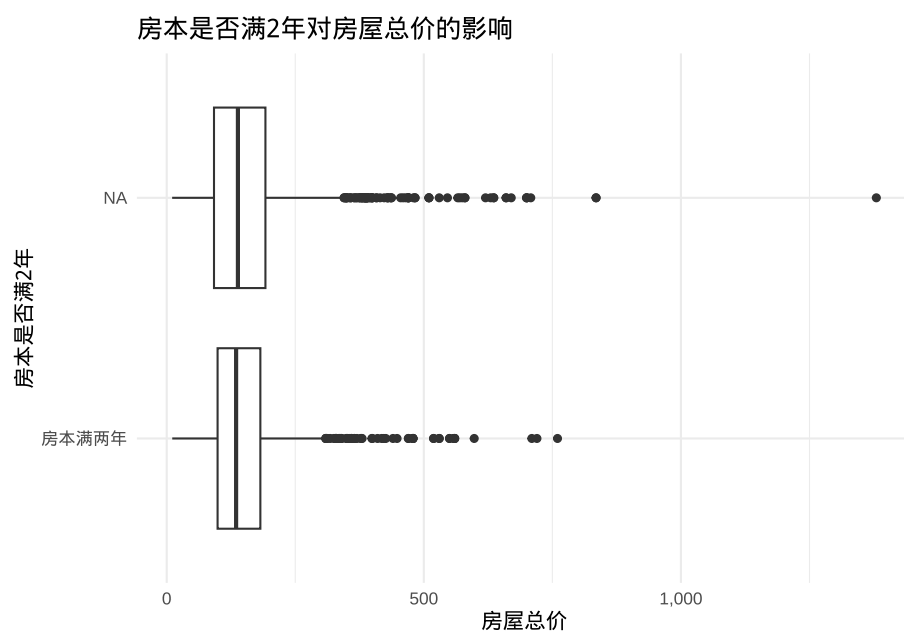
##
## Chi-squared test for given probabilities
##
## data: table(new$vr, new$followers)
## X-squared = 15926, df = 80, p-value < 2.2e-16
```



发现：

- 发现: 从直方图以及卡方检测结果，是否近地铁、能否 `vr` 看装修这类信息对关注人数影响显著

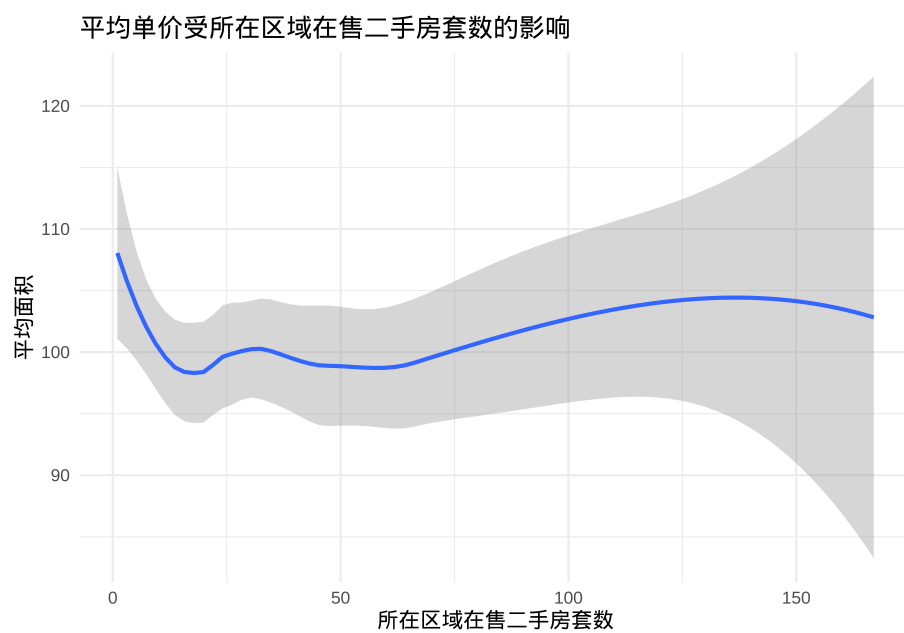
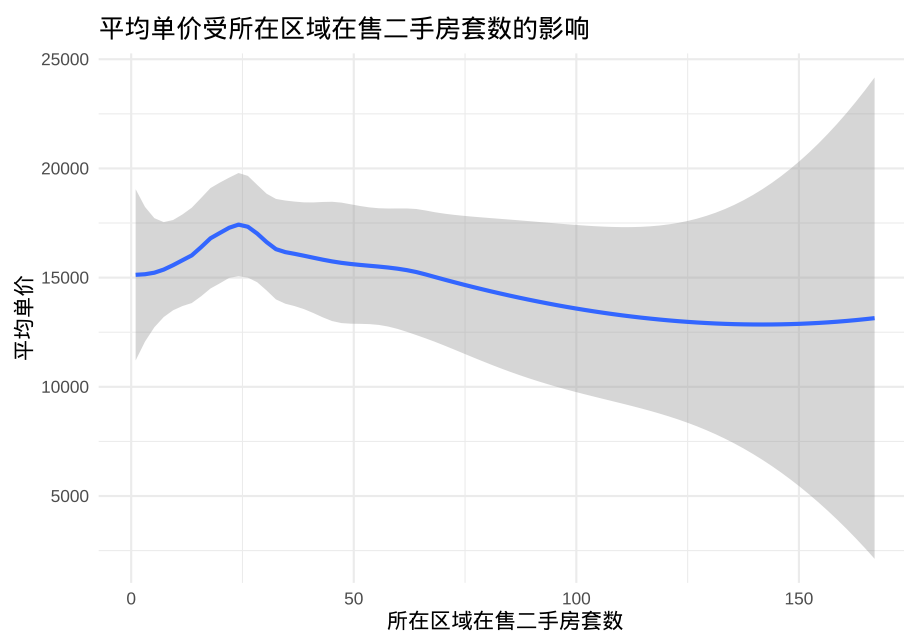
#### 4.4 判断是否满 2 年对二手房总价影响是否显著



发现：

- 发现: 挂售二手房中，半数房屋拿证满 2 年。其中拿证已满 2 年的房屋总价集中度更强一些。

## 根据二手房在各区域分布数量与房屋单价的关系, 分析地区挂牌房屋数量与单价的关系



发现：

- 发现 1: 在售房屋平均单价随着区域套数增长，表现出先上升后平缓下

降的趋势；同时套数多的区域，可选择的不同价位二手房更多。

- 发现 2：在售房屋平均面积随着区域套数增长，表现出先下降后平缓上升的趋势；同时套数多的区域，可选择的面积更多。