

链家二手房数据分析报告

张正宇 2023281051035

2023-10-20

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
##
## 载入程辑包: 'scales'
##
##
## The following object is masked from 'package:purrr':
##
##   discard
##
##
## The following object is masked from 'package:readr':
##
##   col_factor
##
##
## 载入程辑包: 'plotly'
##
##
## The following object is masked from 'package:ggplot2':
##
```

```
##      last_plot
##
##
## The following object is masked from 'package:stats':
##
##      filter
##
##
## The following object is masked from 'package:graphics':
##
##      layout
##
##
## 载入需要的程辑包: sysfonts
##
## 载入需要的程辑包: showtextdb
```

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。

各变量的简短信息：

```
## Rows: 3,000
## Columns: 18
## $ property_name      <chr> "南湖名都A区", "万科紫悦湾", "东立国际", "新都汇", "~
## $ property_region    <chr> "南湖沃尔玛", "光谷东", "二七", "光谷广场", "团结大~
## $ price_ttl          <dbl> 237.0, 127.0, 75.0, 188.0, 182.0, 122.0, 99.0, 193.8~
## $ price_sqm          <int> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163~
## $ bedrooms          <int> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4~
## $ livingrooms        <int> 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        <chr> "南", "南", "南", "北", "东南", "南", "南", "南", "~
## $ directions2        <chr> "北", "", "", "东", "", "北", "", "北", "北", "北", ~
```

```
## $ decoration      <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简~
## $ property_t_height <int> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, ~
## $ property_height  <chr> "中", "中", "低", "高", "中", "低", "低", "中", "低"~
## $ property_style   <chr> "塔楼", "板楼", "塔楼", "塔楼", "板塔结合", "板楼", ~
## $ followers        <int> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, ~
## $ near_subway       <chr> "近地铁", NA, "近地铁", "近地铁", NA, NA, "近地铁", ~
## $ if_2y            <chr> NA, "房本满两年", NA, "房本满两年", "房本满两年", "~
## $ has_key          <chr> "随时看房", "随时看房", "随时看房", "随时看房", "随~
## $ vr              <chr> NA, "VR看装修", NA, NA, "VR看装修", NA, "VR看装修", ~
```

各变量的简短统计:

```
##  property_name      property_region      price_ttl      price_sqm
##  Length:3000      Length:3000      Min.   : 10.6      Min.   : 1771
##  Class :character  Class :character  1st Qu.: 95.0      1st Qu.:10799
##  Mode  :character  Mode  :character  Median : 137.0     Median :14404
##                                     Mean  : 155.9     Mean   :15148
##                                     3rd Qu.: 188.0     3rd Qu.:18211
##                                     Max.   :1380.0     Max.   :44656
##      bedrooms      livingrooms      building_area      directions1
##  Min.   :1.000      Min.   :0.000      Min.   : 22.77      Length:3000
##  1st Qu.:2.000      1st Qu.:1.000      1st Qu.: 84.92      Class :character
##  Median :3.000      Median :2.000      Median : 95.55      Mode  :character
##  Mean   :2.695      Mean   :1.709      Mean   :100.87
##  3rd Qu.:3.000      3rd Qu.:2.000      3rd Qu.:117.68
##  Max.   :7.000      Max.   :4.000      Max.   :588.66
##  directions2      decoration      property_t_height  property_height
##  Length:3000      Length:3000      Min.   : 2.00      Length:3000
##  Class :character  Class :character  1st Qu.:11.00      Class :character
##  Mode  :character  Mode  :character  Median :27.00      Mode  :character
##                                     Mean   :24.22
##                                     3rd Qu.:33.00
##                                     Max.   :62.00
##  property_style      followers      near_subway      if_2y
##  Length:3000      Min.   : 0.000      Length:3000      Length:3000
##  Class :character  1st Qu.: 1.000      Class :character  Class :character
##  Mode  :character  Median : 3.000      Mode  :character  Mode  :character
##                                     Mean   : 6.614
##                                     3rd Qu.: 6.000
##                                     Max.   :262.000
##      has_key      vr
```

```
## Length:3000      Length:3000
## Class :character  Class :character
## Mode :character   Mode :character
##
##
##
```

可以看到：

1. 总价（price_ttl）的最小值为 10.6w，最大值为 1380.0w，平均值为 155.9w。
2. 单价（price_sqm）的最小值为 1771，最大值为 44656，平均值为 15148。
3. 卧室数量（bedrooms）的最小值为 1，最大值为 7，平均值为 2.695。
4. 客厅数量（livingrooms）的最小值为 0，最大值为 4，平均值为 1.709。
5. 建筑面积（building_area）的最小值为 22.77，最大值为 588.66，平均值为 100.87
6. 主要朝向（directions1）最多为南
7. 次要朝向（directions2）最多为北
8. 装修状况（decoration）最多为精装
9. 楼栋总层数（property_t_height）最多为 33，其次为 34，最高为 62
10. 建筑形式（property_style）最多板楼
11. 二手房网站的关注人数（followers）最多为 262

探索性分析

数值描述与图形 1：箱线图

绘制一个基于 price_sqm 和 property_region 的箱线图，以比较不同地区的价格分布情况。每个地区的箱线图会以不同的颜色进行区分



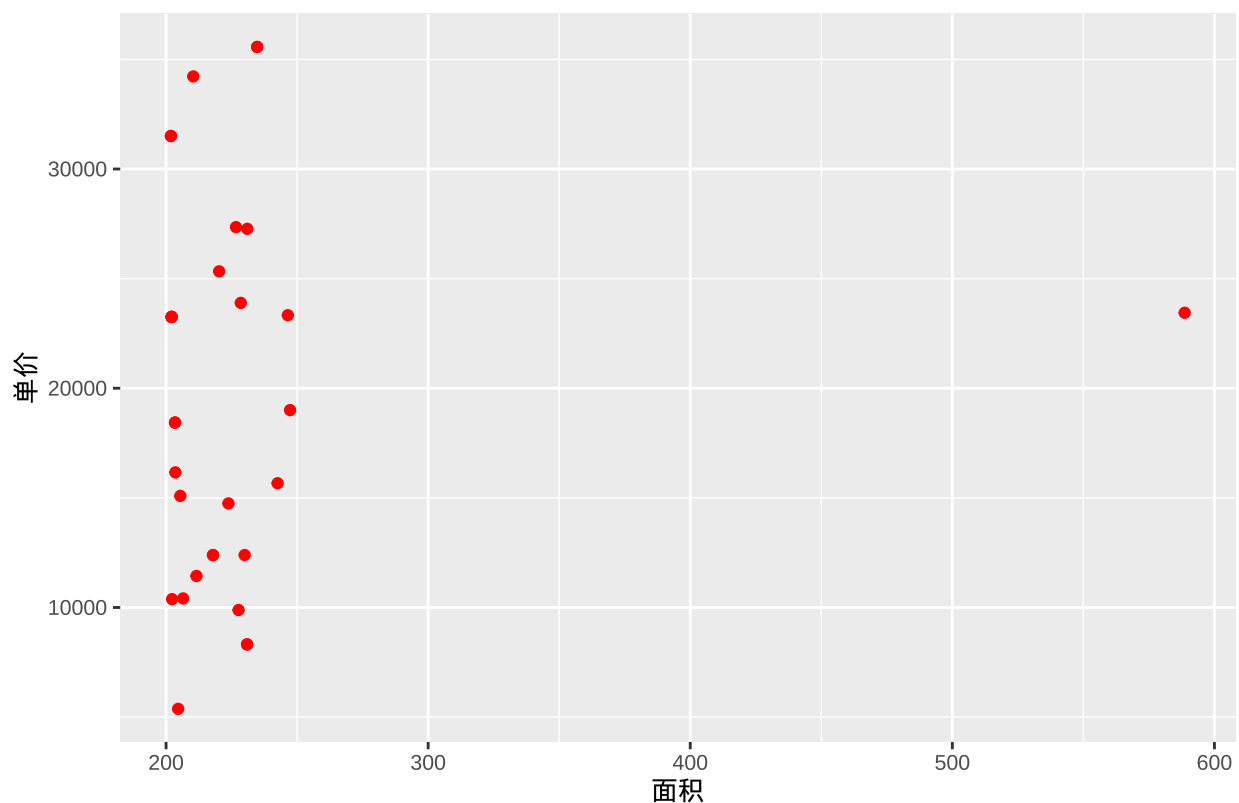
发现：

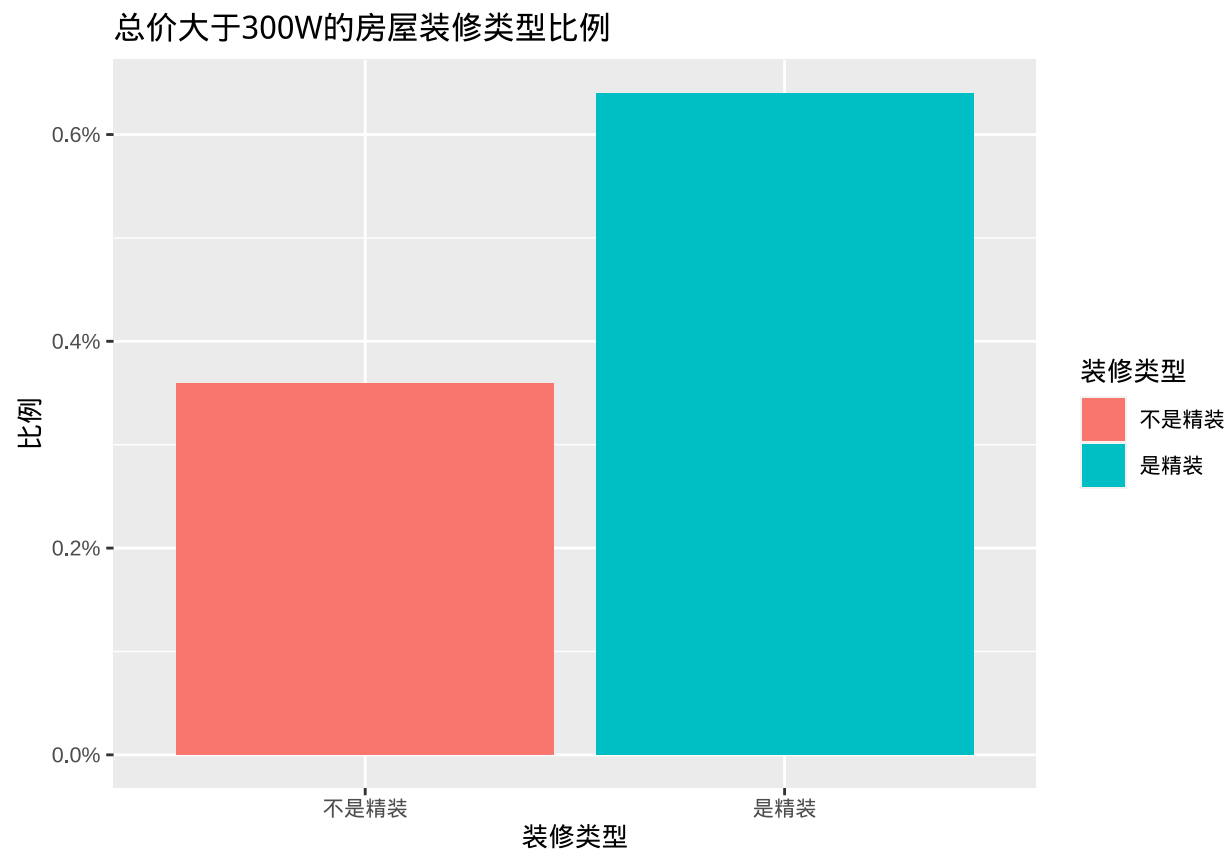
- 中北路的价格中间值和最大值，都是在整个武汉区域最高的

数值描述与图形 2: 散点图

按照 `building_area` , 进行分组, 统计出面积大于 200 的楼盘的单价

建筑面积大于200的每平方米价格散点图

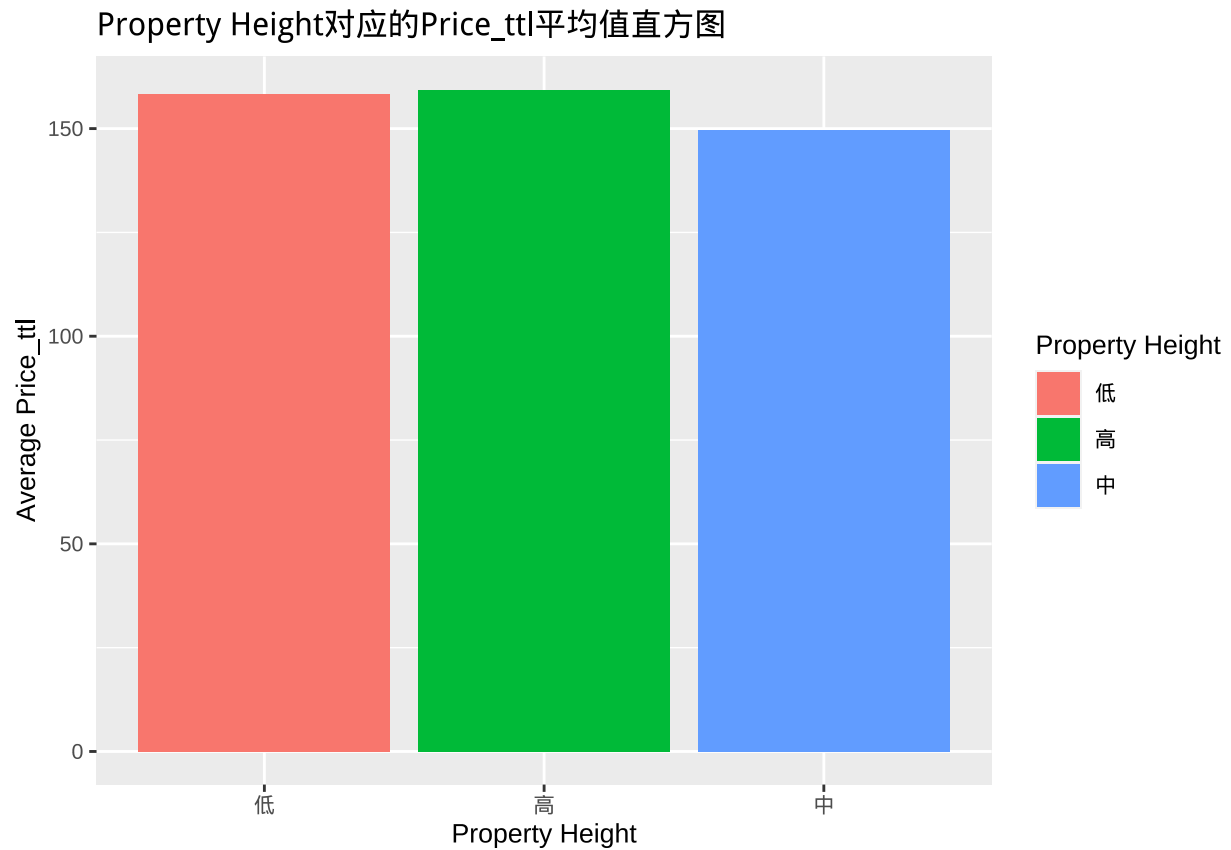




发现：

- 在总价较高的二手房中，很小的比例为非精装，非精装的房子较难卖出高价

探索问题 1 楼层在本楼中的位置对价格的影响

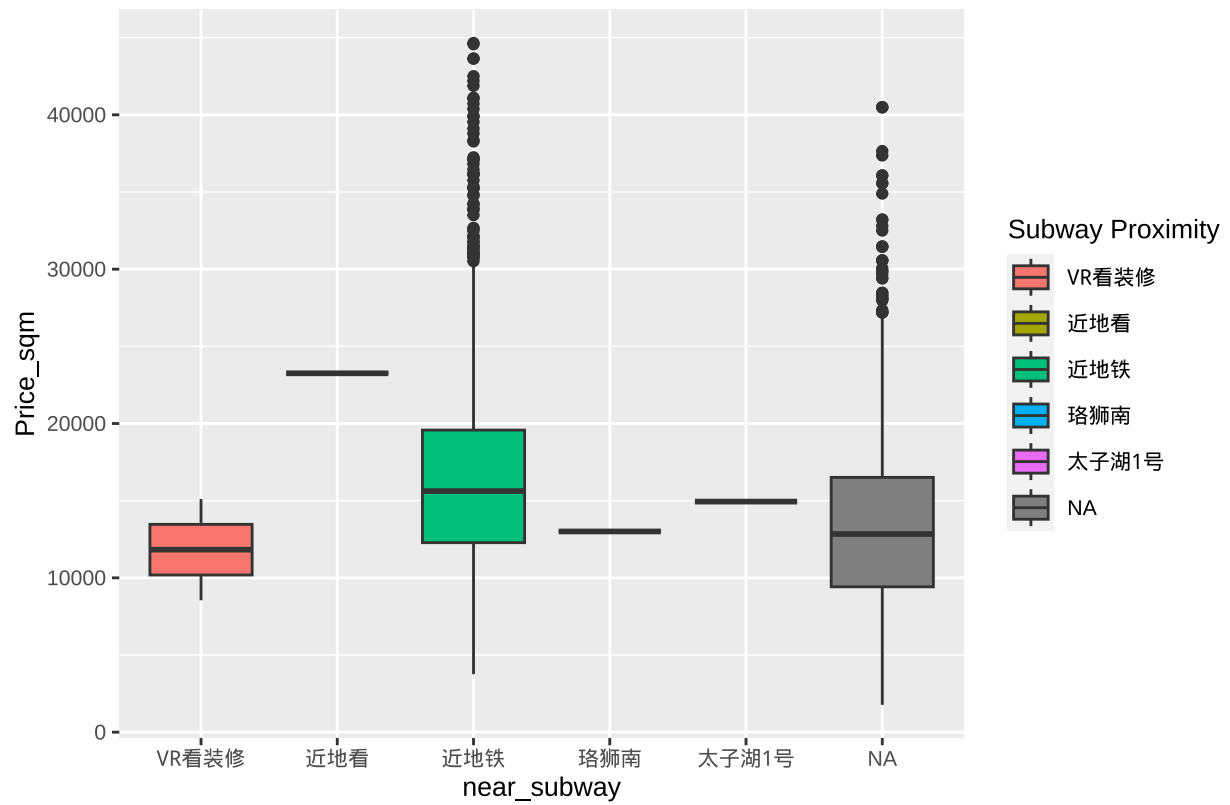


-发现 1 综合来看，中楼层在市场中的接受程度相对较差，总价相对较低 -发现 2 高低楼层之间的价格差距并无想象中的那么大，均价接近

探索问题 2 房屋单价和靠近地铁是否有关系

分别绘制是否靠近地铁房价的箱线图

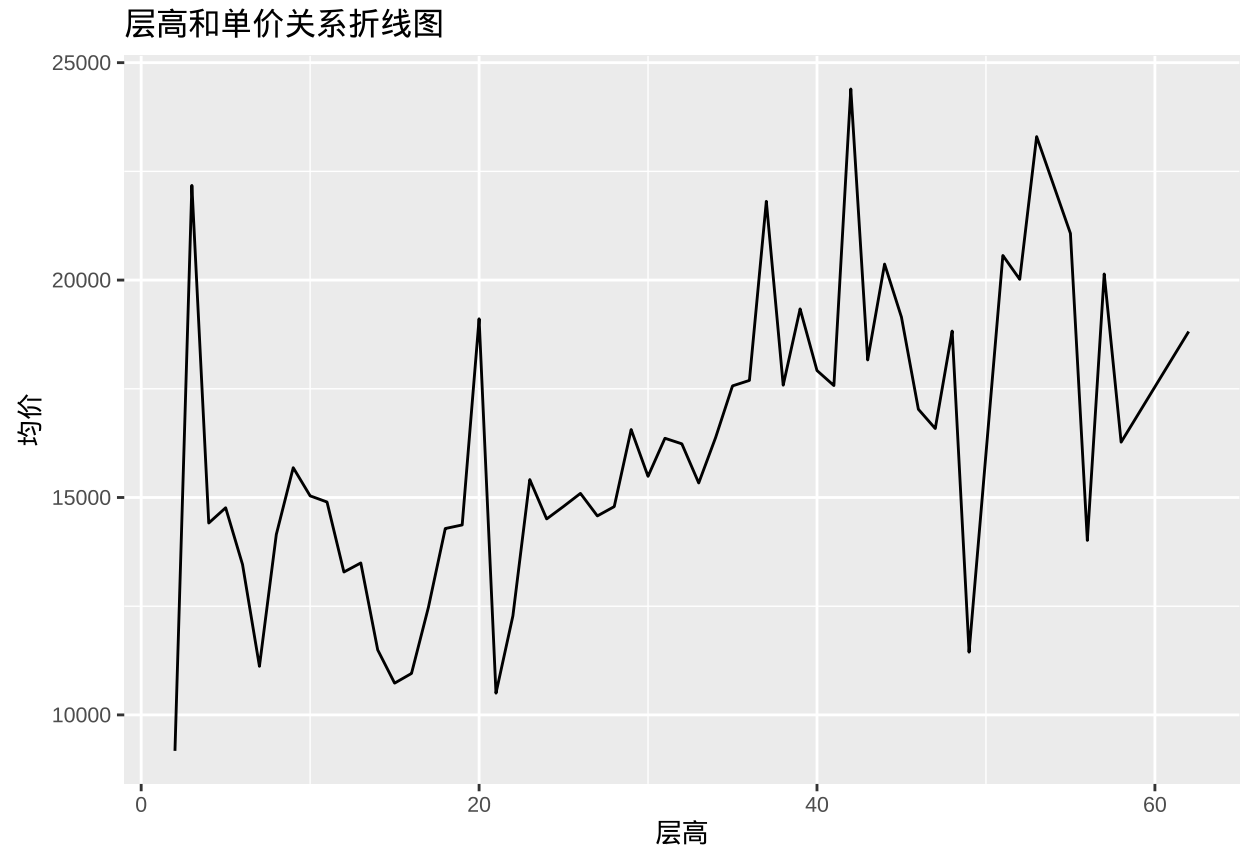
近地铁 vs. 不近地铁房屋单价中位数比较



-发现 1 综合来看，靠近地铁的房屋单价相对更高 -发现 2 近地铁的房屋能够卖出当前二手房市场的最高价，且大部分单价为 3w 以上的房源都靠近地铁

探索问题 3 房屋单价和总层高的关系

绘制折线图



—发现 1 超低楼层房子难以卖出高价 —发现 2 市场上对于超高楼层的接受度还算不错，能够卖出高价

发现总结

- 房屋价格和多个因素有关，例如是否靠近地铁，房屋所在的区域，这些因素有关的房屋升值空间大
- 房屋单价和房屋的面积并没有直接的关系，单价的受制因素非常多，建议多方比对
- 房屋并不是买低楼层就卖不出好价格，主要还得看地理位置和面积大小