

第一次作业你的报告题目

Code ▾

your name

2024-10-30

- 1 你的主要发现
- 2 数据介绍
- 3 数据概览
- 4 探究哪个区域房屋供应数量最多
- 5 探究哪个区域的房产均价最高
- 6 探究哪个区域的房产平均面积最大
- 7 探索性分析
 - 7.1 整体上来看，房产面积越大，总价越高。
 - 7.2 为什么汉口北、中北路这两者区域差异这么大，但平均房屋面积接近且都为最大呢
- 8 发现总结

1 你的主要发现

1. 白沙洲区域的房子供应数量最多
2. 中北路是武汉平均房价最高的地方
3. 汉口北、中北路房子的平均面积最大

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：

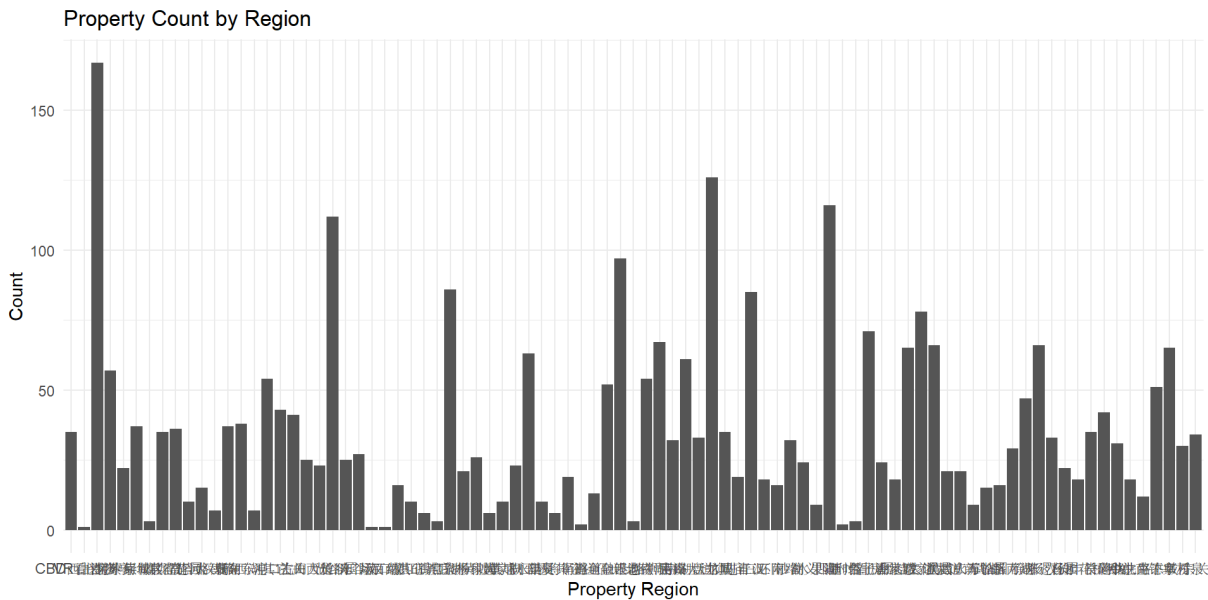
```
## Rows: 3,000
## Columns: 18
## $ property_name      <chr> "南湖名都A区", "万科紫悦湾", "东立国际", "新都汇", "…
## $ property_region    <chr> "南湖沃尔玛", "光谷东", "二七", "光谷广场", "团结大…
## $ price_ttl          <dbl> 237.0, 127.0, 75.0, 188.0, 182.0, 122.0, 99.0, 193.8…
## $ price_sqm          <dbl> 18709, 14613, 15968, 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        <chr> "南", "南", "南", "北", "东南", "南", "南", "南", "…
## $ directions2        <chr> "北", NA, NA, "东", NA, "北", NA, "北", "北", "北", …
## $ decoration          <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简…
## $ property_t_height  <dbl> 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           <dbl> 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
##
##
##
```

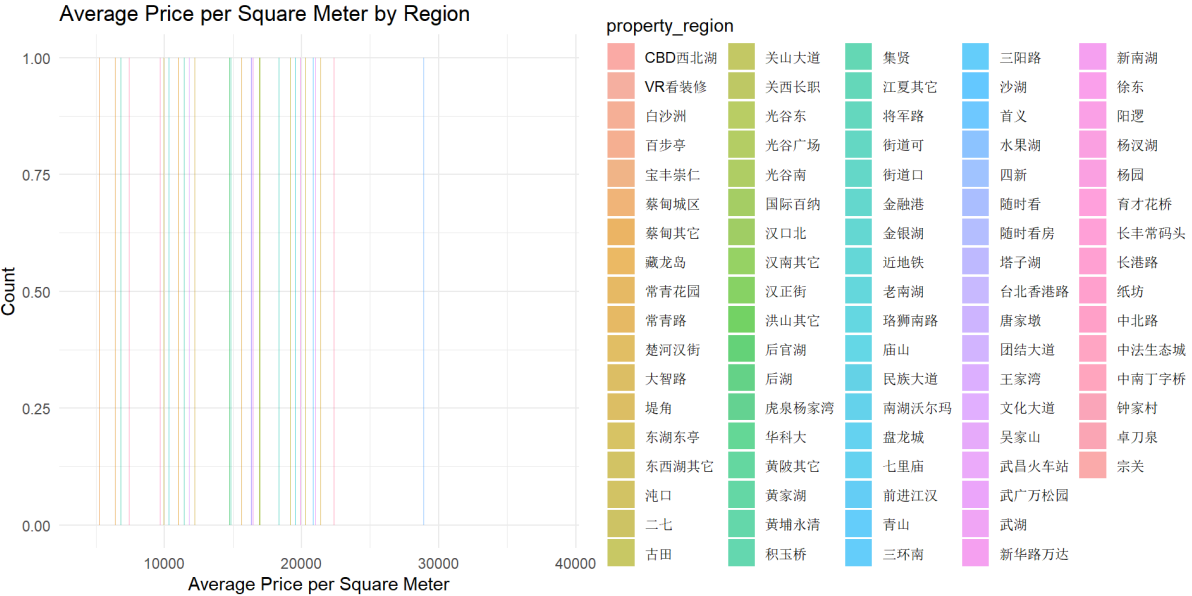
4 探究哪个区域房屋供应数量最多

```
## # A tibble: 87 × 2
##   property_region count
##   <chr>         <int>
## 1 白沙洲         167
## 2 盘龙城         126
## 3 四新          116
## 4 光谷东         112
## 5 金银湖          97
## 6 后湖           86
## 7 青山           85
## 8 王家湾          78
## 9 塔子湖          71
## 10 珞狮南路       67
## # 77 more rows
```



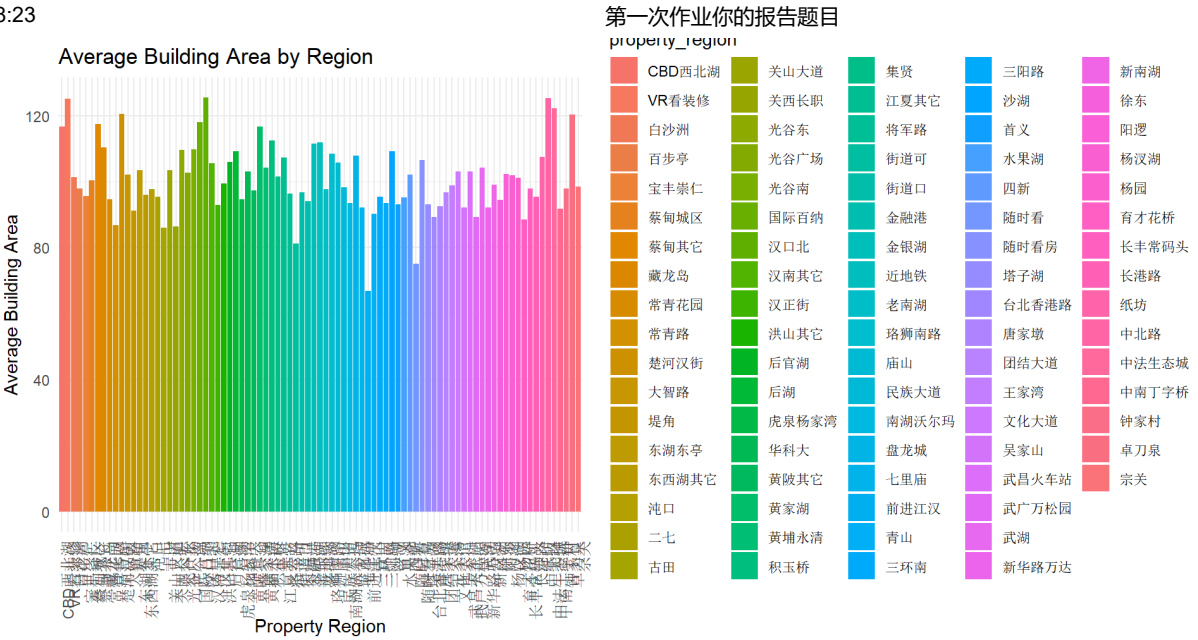
5 探究哪个区域的房产均价最高

```
## # A tibble: 87 × 2
##   property_region avg_price_sqm
##   <chr>          <dbl>
## 1 VR看装修      38351
## 2 中北路        32728.
## 3 水果湖        28562.
## 4 黄浦永清      24957.
## 5 三阳路        24777.
## 6 南湖沃尔玛   24181.
## 7 虎泉杨家湾   23902.
## 8 CBD西北湖    22272.
## 9 楚河汉街     21958.
## 10 关山大道     21480.
## # 77 more rows
```



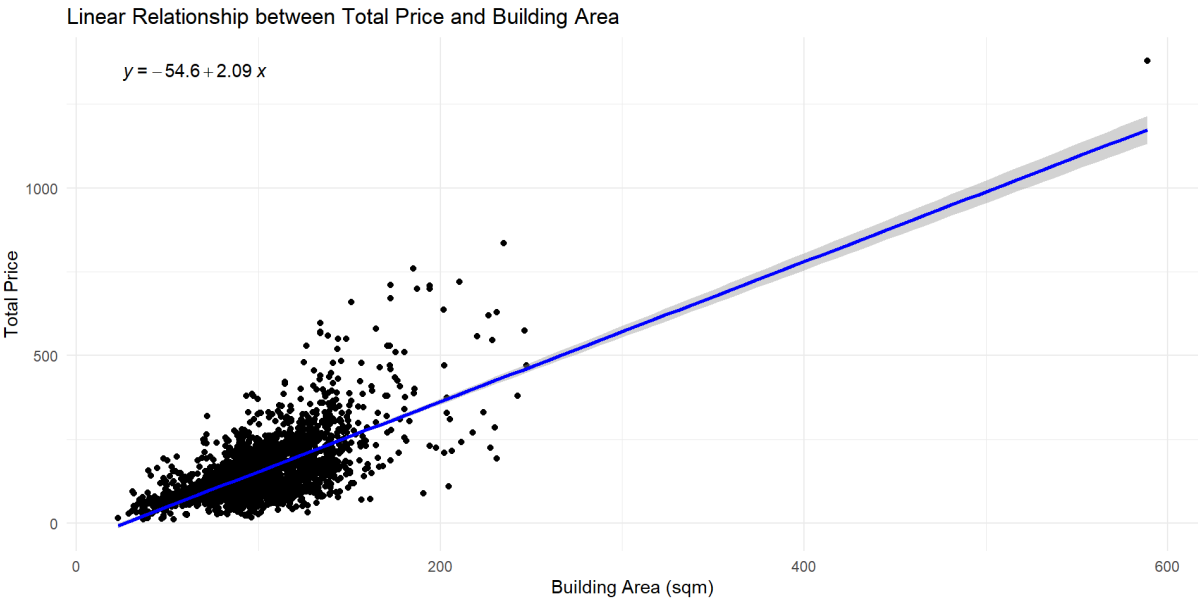
6 探究哪个区域的房产平均面积最大

```
## # A tibble: 87 × 2
##   property_region avg_building_area
##   <chr>          <dbl>
## 1 汉口北        125.
## 2 中北路        125.
## 3 VR看装修      125.
## 4 中法生态城    122.
## 5 楚河汉街     121.
## 6 卓刀泉       120.
## 7 国际百纳     118.
## 8 蔡甸其它     117.
## 9 CBD西北湖    117.
## 10 黄家湖       117.
## # 77 more rows
```



- 可以看到：
- 越繁华的地区均价越高
 - 白沙洲等郊区供应的房产数量比较多
 - 价格最高的区域、最低的区域，平均房屋面积较大。中间则无明显趋势

7 探索性分析



7.1 整体上来看，房产面积越大，总价越高。

- 发现：
- 通过拟合成一元函数直线，可以估算出武汉的平均房价
 - 通过逆合成曲线，可以看出房价相对于面积的变化趋势。上扬则说明面积越大，平均价格越高。

7.2 为什么汉口北、中北路这两者区域差异这么大，但平均房屋面积接近且都为最大呢

- 发现：
- 汉口北土地充裕，且低价便宜，所以普遍面积较大。
 - 中北路房价高，一般为有钱人居住，有钱人注重生活品质，所以房屋面积大

8 发现总结

- 越繁华的地区均价越高。比如中北路最高

- 白沙洲等郊区供应的房产数量比较多。
- 价格最高的区域（中北路）、最低的区域（汉口北），平均房屋面积较大。中间则无明显趋势