**武汉市租房市场行情分析**

——基于python爬虫爬取的链家网租房信息

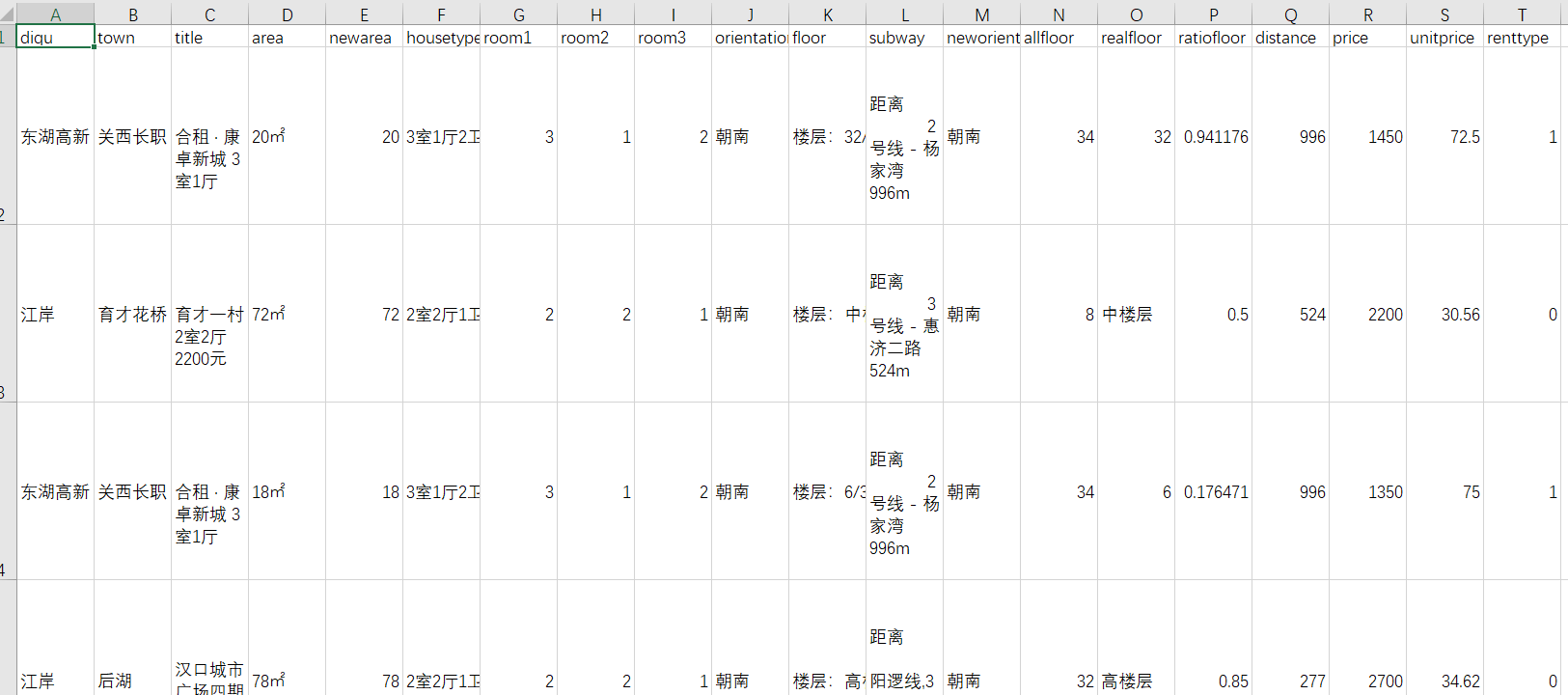
1. 数据清理

将用爬虫爬取的3006条房源信息存入csv文件。



如图，有数据存在大量变量缺失以及存在严重文本数据混杂问题。故先用EXCEL对数据进行初步清洗整理。将楼层信息提取为总楼层（allroof）、房源楼层（realroof）和房源楼层占总楼层的比例（ratioroof），由于大量数据缺乏具体楼层，故用0.15代表低楼层，0.5代表中层，0.85代表高楼层。将价格除以面积得到每平方米单价（unitprice），发现部分相似房源的单价差距较大。分析原因发现房源主要分为整租和合租两种模式，由于这两种方式面积计算方式不同（合租面积通常未考虑客厅和厨卫），导致单价水平也有所不同。故从title中提取变量（renttype）表示房源是整租还是合租。从subway中提取变量distance表示房源到最近地铁站的距离。房型housetype格式为“XX室X厅X卫”，考虑到后续分析需求，拆分为数值变量room1、room2、room3.

基本处理效果如下图所示：



1. 提出问题

从租客视角出发，最关的问题主要有两个：

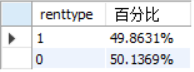
哪里的房子具有较高的性价比？（从而可以根据结论在自己可以接受的最大地理范围内选择性价比更高的房子）

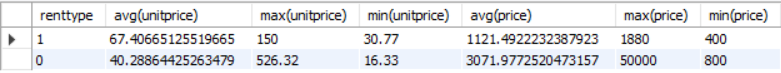
影响房租价格的主要因素有哪些？（了解后，可以根据个人偏好来权衡房租价格和房子条件要求）

1. 描述性统计分析和数据可视化

将清理好的数据导入MySQL进行初步的探索分析，并基于POWER BI 和TABLAEU 进行可视化分析。

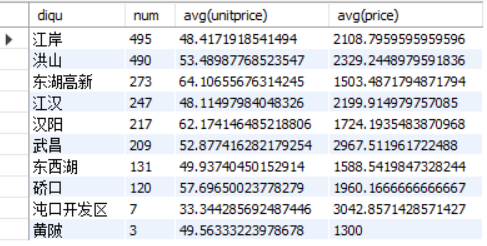
1.出租类型分布及其和房租的关系

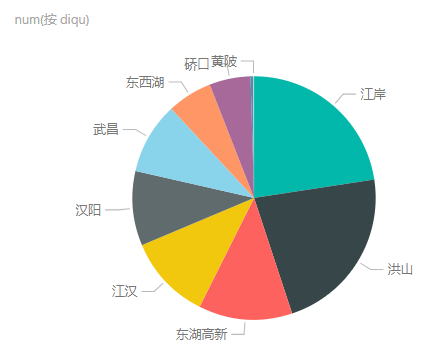
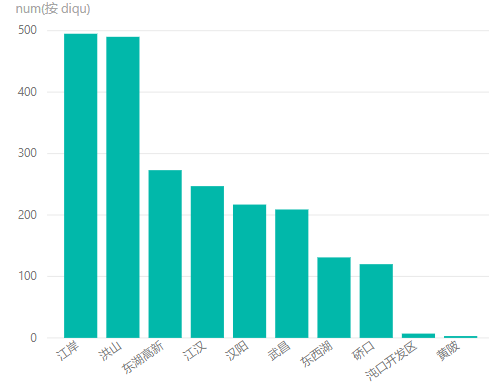


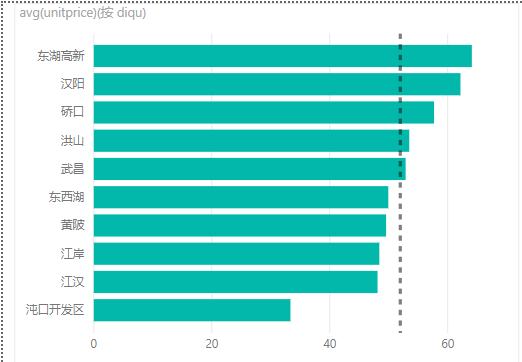
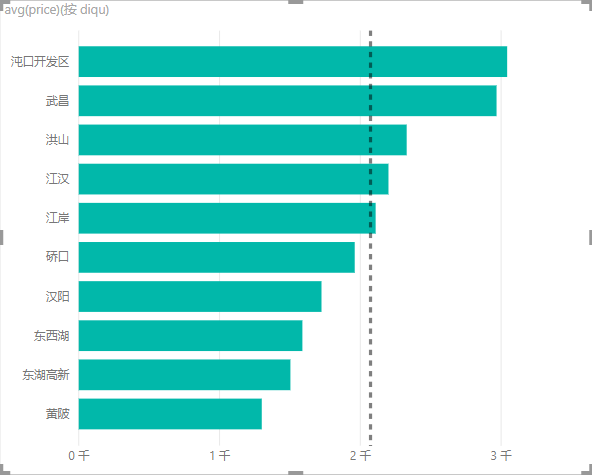


从结果中可见整租和合租类型数量基本持平。但由于面积计算方式的不同，合租模式的单位面积房租总体上高于整租模式。但整租模式的总房租水平显著高于合租，并且波动范围较大。

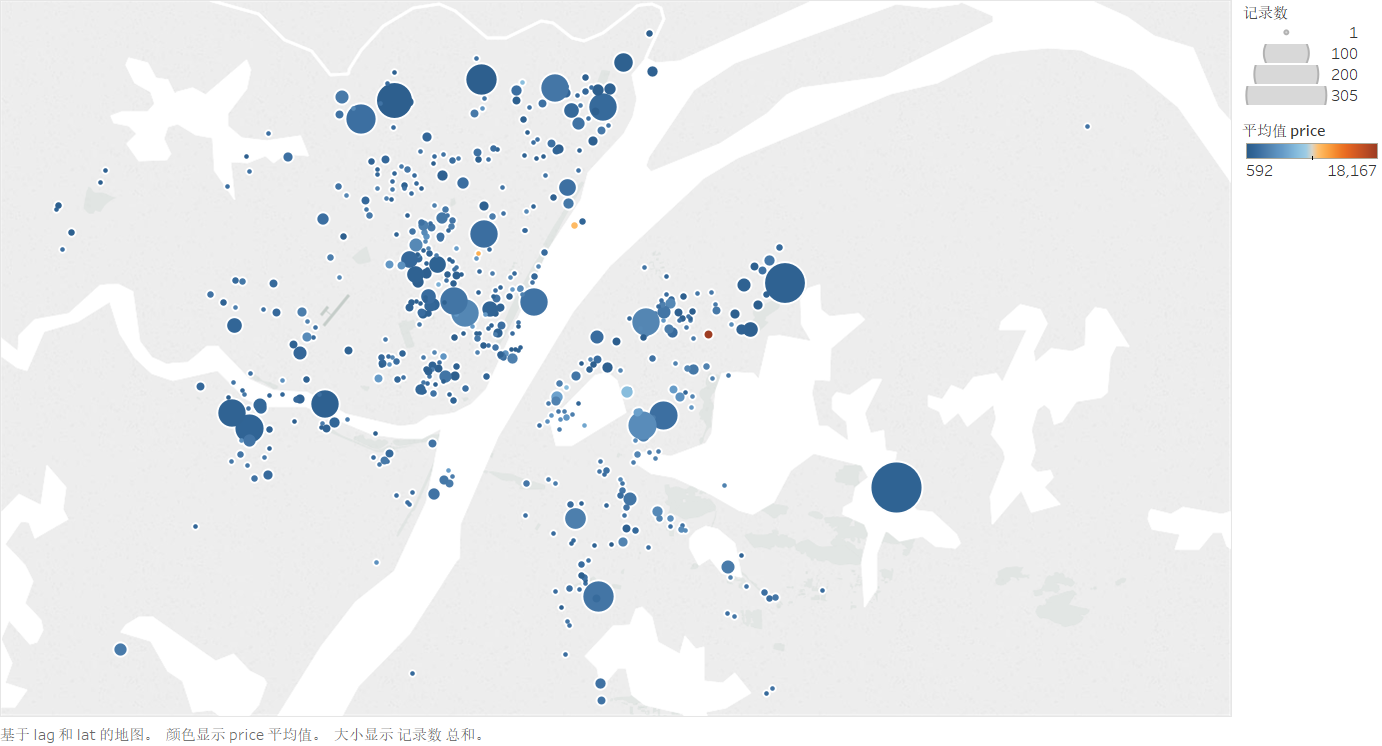
2.不同地区房源分布及地区对房租影响

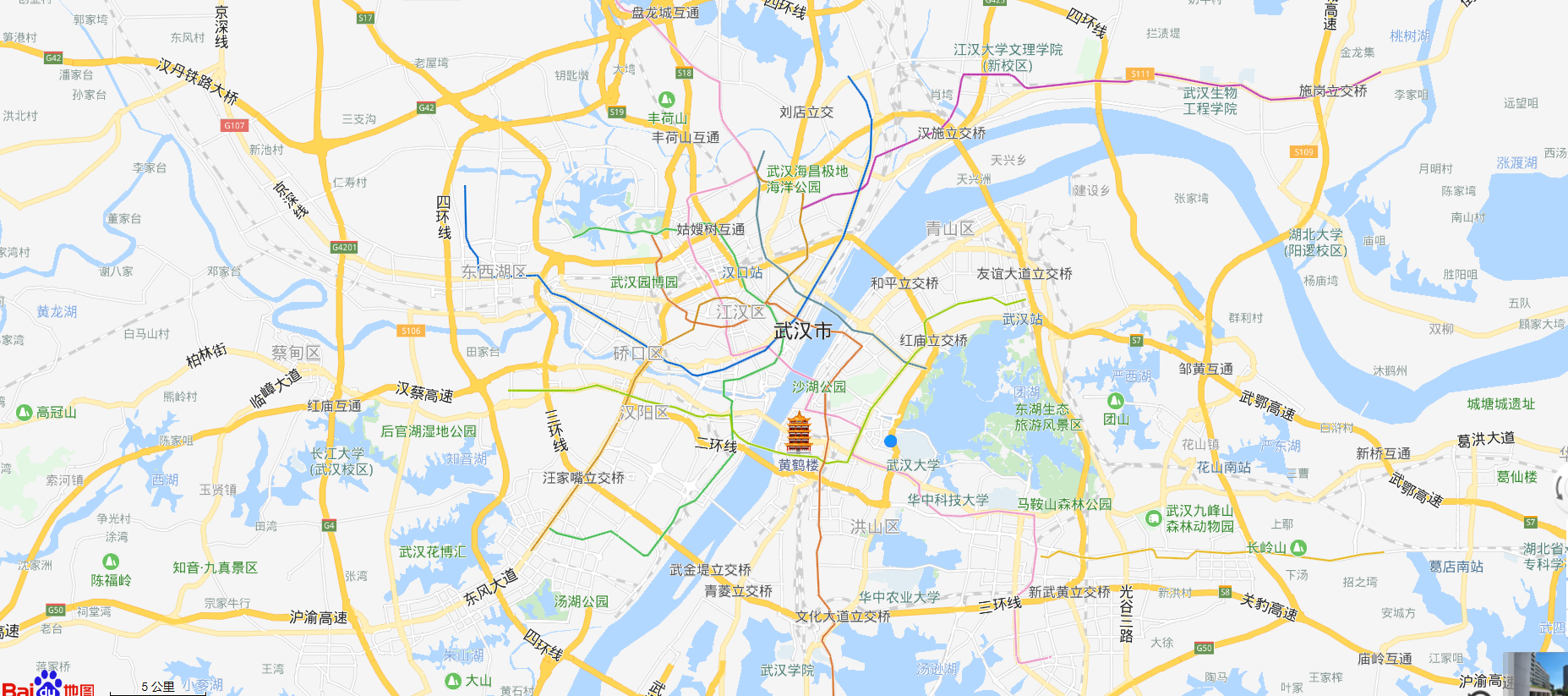






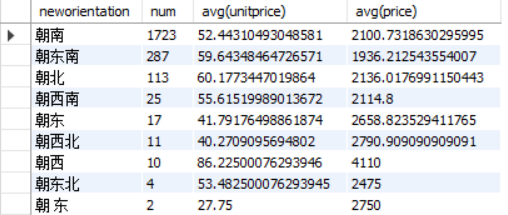
通过对地区房源数量和平均房租的分析绘制出的条形图可以看到，武昌区房源数量偏低而房租显著高于其它地区，充分反映了武昌区房源供应小于需求的状况。而江汉和江岸区房源数量显著高于其它地区，虽然总体房租价格适中，但单位面积房租价格较低，反映出这两个区域房源充足，可能有供大于求得情况。

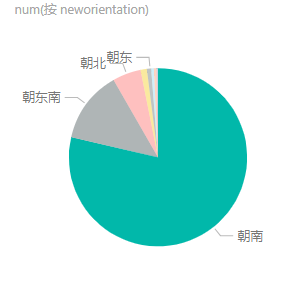
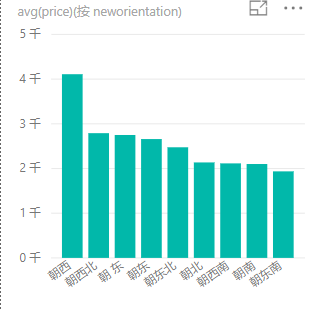




根据房源小区地址在百度地图获取具体经纬度绘制房源及价格分布图。图中圆圈大小代表房源数量，圆圈颜色代表房源价格。从地理位置上看，武汉租房房源分布比较平均分散，郊区房源通常相对少儿而集中，而偏向市中心处的房源分布多而分散，且价格明显高于边缘地区（和郊区房大多整租且面积较大有关系）。

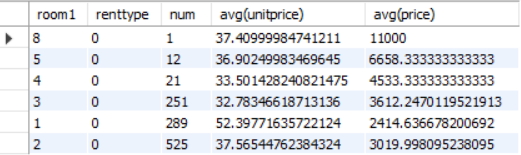
3.不同朝向对房租价格影响

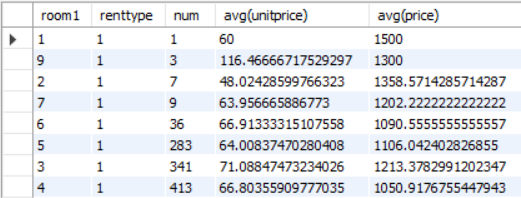


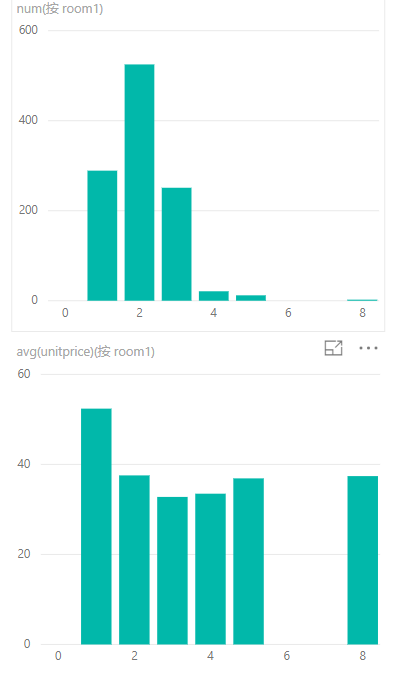
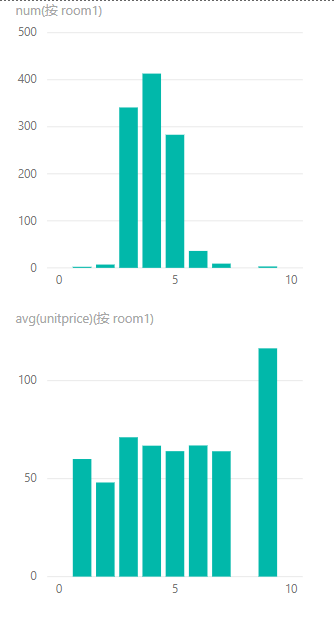
 

如图所示，大部分房源的朝向都是朝南，并且除了朝西方向的房源价格显著高于其它房源，而朝西的房源只有十套，很可能有异常值影响了价格。可以初步假定房间朝向对房租价格没有显著影响。

4．不同房型对房租价格影响



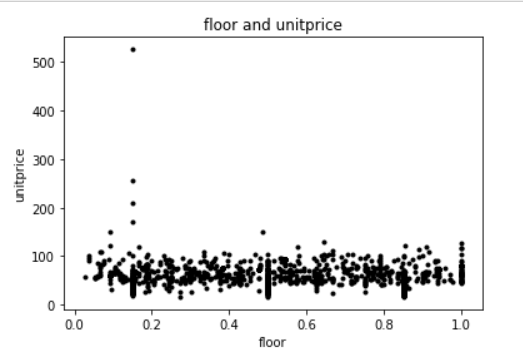
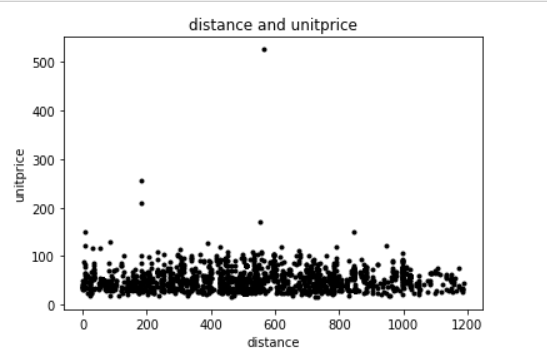


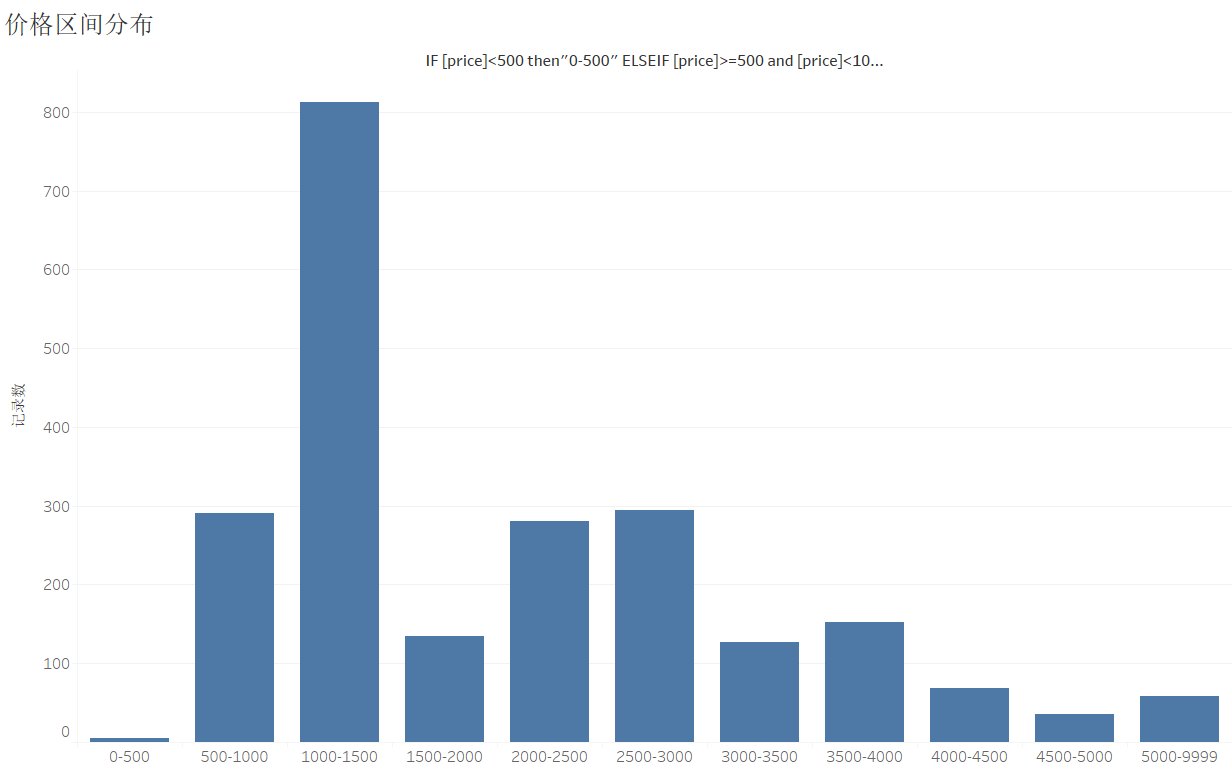
左右图分别是整租和合租类型的房型分布及其均价。可以看出整租房基本集中在1~3室，合租房基本集中在3~5室。排除8、9室异常值的情况下，整租房一室单位面积房租明显高于其它房型，性价比不高。意外的发现，合租房两室单位面积房租最低，性价比相对高。这和平时对于房间数量越多性价比越高的认知相冲突。

5.房租价格和地铁站距离及楼层间关系

如下图所示，房租和房源也地铁站距离没有明显的相关关系。计算两者的相关系数为0.12，可以初步判断距离不明显影响房租价格。同样，房租和楼层间的相关系数为-0.07，可以基本判断为楼层对房租价格没有显著影响。



6.房租价格分布图



根据房租绘制的条形图可见，1000-1500元/月的房源是占比最高并且远高于其它房源的，而这个多见于单间的价格。可见武汉租房主流市场仍然是以平价单间出租居多。

1. 分析结果

通过对可能对房租有影响的因素和房租的相关性的对比分析后，可以发现，楼层和距离地铁站距离以及房间朝向并不明显影响房租价格，因此在选择房源时，这三个因素可以完全根据个人喜好判断。

对于房型，如果选择整租，一室到三室的房源较为充足，且性价比随着房间数量上升而上升，如果选择合租，三室到五室的房源较为充足，但意外的，性价比最高的是两室的房源，而三室到五室之间的性价比差距不大，可以尽量选择三室。

关于地区选择，洪山区和江岸区房源数量远高于其它地区，但洪山区的房租价格在武汉市中偏高，反映了洪山区供求都很大的情况。相对的，武昌区则明显有供大于求得情况。东湖高新区整体房租低，但单位面积房租价格高，反映了其出租房间类型多为一室或合租型。总体上来说，江岸、江汉、黄陂都是租房性价比较高的地区，沌口区房源数量过少，分析结果不具有代表性。

综上，选择房源时，首先考虑房源地区，在可接受范围内选择房源数量大、房租均价低的地区，这是对房租影响最大的因素。其次根据租房类型选择三室或两室的房源可以获得较高性价比。而楼层、地铁距离和房间朝向对房租价格影响不大，可以更多的依赖个人喜好决定。