

租房数据分析

题目要求：

1. 抓取链家官网北上广深 4 个一线城市，再加上离你家乡最近的一个非一线城市/或者你最感兴趣的一个城市的数据。应尽量获取每个城市的全部租房数据（北上广深 4 个一线城市要求每个城市至少抓取 8000 条有效数据）。
2. 比较 5 个城市的总体房租情况，包含租金的均价、最高价、最低价、中位数等信息，单位面积租金（元/平米）的均价、最高价、最低价、中位数等信息，并采用合适的图或表形式进行展示。
3. 比较 5 个城市一居、二居、三居房源的情况，包含均价、最高价、最低价、中位数等信息，并采用合适的图或表形式进行展示。
4. 计算和分析每个城市不同板块的均价情况，并采用合适的图或表形式进行展示。
5. 比较各个城市不同朝向房源的单位面积租金分布情况，采用合适的图或表形式进行展示。
哪个方向最高，哪个方向最低？各个城市是否一致？如果不一致，你认为原因是什么？
6. 比较各个城市主要租房中介品牌的分布情况，采用合适的图或表形式进行展示。
7. 查询各个城市的平均工资水平，分析并展示其和单位面积租金分布的关系。相对而言，在哪个城市租房的负担最重？
8. 围绕各城市租房问题，结合上述数据及分析，设计自己感兴趣的的数据分析主题，补充查找或爬取相关数据，完成题目设计、数据获取、数据分析及数据展示过程。



数据项位置说明：以上图为例，“马家堡”为板块名称，“贝壳优选”为租房中介品牌。

作业提交方式和要求：

1. 作业在教学云平台-云邮教学空间 (<https://ucloud.bupt.edu.cn/>) 上提交，包括报告和源代码包两部分。(注意：源代码包仅作为在必要时验证程序正确性的备用手段，老师批改作业时不一定会看每一位同学的源代码包。**希望老师批改作业时看到的内容请务必写在报告里**)
2. 报告结构请自行设计，不做统一要求。一般应涵盖任务目标、数据获取与处理说明、数据分析结果、核心代码与实现思路、总结与收获等内容。
3. 除了题目要求的各项数据分析/处理工作的结果，报告中还应包含爬虫核心代码、数据处理及数据展示的核心代码。代码应包含必要的注释或附加的实现思路说明。代码与实现思路说明可以集中呈现，也可以拆分至对应的数据分析/处理工作同步展示。
4. 报告总页数要求不超过 30 页。
5. 文件格式和命名要求：
 - a) 报告要求为 pdf 格式，文件名为“<学号>_<姓名>_大作业”，例如：2021211353_李永琰_大作业.pdf。
 - b) 源代码包要求为 zip 格式，文件名为“<学号>_<姓名>_大作业_代码”，例如：2021211353_李永琰_大作业_代码.zip
6. **作业提交截止时间：2026.01.04(星期日) 24 点**

评分标准参考 (总分：40 分)

代码功能与图表展示								报告整体	
数据爬取	数据预处理	数据分析展示						格式结构	分析论述
		租金	户型	板块	朝向	中介品牌	工资-租金		
10	4	3	3	3	3	2	2	2	6