

期末大作业-租房数据分析

1. 要求抓取链家官网北上广深 4 个一线城市，再加上一个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。
2. 应获取每个城市的全部租房数据（一线城市的数据量应该在万的数量级）。
3. 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。采用合适的图或表形式进行展示。
4. 比较 5 个城市一居、二居、三居的情况，包含均价、最高价、最低价、中位数等信息。
5. 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。



例如上图中的“海淀-四季青-五福玲珑居北区”，“四季青”即为板块名称。

6. 比较各个城市不同朝向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？
7. 查询各个城市的人均 GDP，分析并展示其和单位面积租金分布的关系。相对而言，在哪个城市租房的性价比最高？
8. 查询各个城市的平均工资，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？

实验报告需要展示对于作业要求的设计、实现、分析及结论，必要时可以在报告中粘贴爬虫程序的核心代码、数据处理及数据展示的核心代码(应有足够的注释)；可视化图表在报告中展示，报告以 pdf 格式提交到教学云平台，总页数不超过 30 页。提交内容还包括源代码和关键的数据文件，打成一个压缩包，文件名为学号，截止时间 12 月 31 日 24 点。

评分标准

数据抓取	数据处理	房租分析	居室分析	板块分析	朝向分析	人均 GDP 分析	平均工资分析	源程序和说明文档
30	10	10	10	10	10	5	5	10

总计 100，其中数据分析各维度占比较大，也是作业的关键部分。