

## 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。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）。
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。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、三居的情况，包含均价、最高价、最低价、中位数等信息。
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。



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。

5. 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？
6. 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？
7. 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？

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。

### **报告要求：**

1. 实验报告中应包含爬虫核心代码、核心数据文件的基本结构、数据处理及数据展示的核心代码。代码及文件中应包含足够的注释。
2. 报告应按照一般实验报告要求，至少包含明确的实验目的、过程、结论。
3. 报告以 pdf 格式提交，总页数不超过 30 页。
4. 报告具体提交方式由助教提供。
5. 报告提交截止时间：2022 年 12 月 31 日 24 点。