|  |  |
| --- | --- |
| 学号 | 10214518401 |



课 程 设 计

课程名称 python程序设计课程设计

|  |  |
| --- | --- |
| 题 目 | 北上广深租房状况分析 |
| 专 业 | \_\_\_ 软件工程 \_\_\_\_\_\_\_\_ |
| 班 级 | \_\_\_\_\_\_\_ 软件1182班\_\_\_\_\_\_\_\_ |
| 姓 名 | \_\_\_\_\_\_\_\_\_\_\_阮汉成\_\_\_\_\_\_\_\_\_\_\_\_ |
| 成 绩 | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 指 导 老 师 | \_\_\_\_\_\_\_\_\_\_\_徐勇\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

2021 年 1 月 11 日至 2021 年 1 月 22 日

武汉华夏理工学院信息工程学院

**课 程 设 计 任 务 书**

课程名称：python程序设计课程设计 指导教师：徐勇

班级名称： 软件1181-1182 开课院、系：计算机与网络工程系

**一、课程设计目的与任务**

“python程序设计课程设计”是一个综合性的学习实践型实验教学环节，将在“python程序设计”课程的授课基础上，对python的基础语法、python 对文件的操作、python 对数据的操作、数据处理基础、数据可视化等若干个知识点进行综合运用。

python是一种具有天然开源基因的编程语言，了解开源社区和广泛的使用开源工具，也是Python学习实践的重要环节。因此在本次课程设计中，要求掌握jupyter notebook、Git等常用工具，以及github等重要开源社区的使用。

**二、课程设计的内容与基本要求**

数据分析的基本技术和方法在“python程序设计”课程中已经进行完整的讲授，由于数据分析本身是基于业务场景的，因此本次课程设计环节更加偏重于实际的业务场景的实践。通过对近期互联网热点的调查，准备了15个具体的业务场景，用于本次课程设计的具体任务场景。具体包括如下：

1. 北上广深租房状况分析；
2. 蔡某坤粉丝数及转发数据真假状况分析；
3. 地震的数据分析；
4. 英文名字的数据分析；
5. 外籍英文老师收入虚高情况数据分析；
6. 我国城市空气污染和烟花燃放的关系分析；
7. 针对996工作，程序员群体的看法的分析；
8. 吴某凡微博热点的分析；
9. 节假日长假景点人满为患的数据分析；
10. 针对荔枝的品种、销售地等维度，进行价格数据分析；
11. 分析芒果TV《我是大侦探》的观众评论数据；
12. 针对当前儿科医生的缺乏，对相关数据进行分析；
13. 著名网游《绝地求生》的数据分析；
14. 实习岗位状况的数据分析；
15. 电影《流浪地球》的观众评价的数据分析。

每三个学生组建一个课程设计小组，最后的任务输出包括代码每小组一份、课程设计报告每人一份、答辩ppt每小组一份，并进行课程设计成果答辩。小组成员均参与前述工作，但是每个人的侧重点不同。

每个课程设计小组可以从上述15个场景中选取一个作为课程设计的选题，选题中提供了待分析的数据，和现有的分析方法。各小组，通过学习和实践现有的分析方法，理解实战分析的思维过程并锻炼实际动手能力，再此基础上可以扩展更多维度的分析和数据展现形式。

每个班每个选题最多只能被两个小组选中，先选先得。同时，如果各小组发现更有意思的场景，并能够获取到相关待分析的数据，也可以申请作为选题方向。

本次课程设计的目标是培养学生的团队协作能力、对python知识点的综合运用、对实际场景的理解和适应能力、针对答辩的表达能力等。注重过程，期待成果，但不强求结果的尽善尽美。

**三、学时分配进度安排**

|  |  |  |
| --- | --- | --- |
| 序号 | 设计内容 | 所用时间 |
| 1 | 下发任务书，学生查阅资料 | 1天 |
| 2 | 组建团队，并进行选题和团队匹配 | 1天 |
| 3 | 各小组对自己的选题开始进行研究 | 2天 |
| 4 | 各小组开始准备数据并形成初步处理意见 | 1天 |
| 5 | 代码和实现分析的动手实践 | 2天 |
| 6 | 形成初步的报告书和ppt | 1天 |
| 7 | 答辩并完成报告书 | 2天 |
| 合 计 | | 2周 |

**四、课程设计考核及评分标准**

**1.设计报告要求**

课程设计报告要求逻辑清晰、层次分明、书写整洁。课程设计报告为每人一份，同一个小组的各成员的整体报告内容结构一致，但个人侧重点不同，个人着重撰写自己的工作内容，其他人的内容只要体现文档结构的完整性即可。

课程设计考核将综合考虑学生考勤和参与度、团队协作能力，过程管理能力、成果达成情况等。

**2.过程要求**

整个过程要求通过github来进行过程化发布，即阶段性地在github上提交过程结果。

**3.评分标准**

|  |  |
| --- | --- |
| **评分依据** | **评分成绩** |
| 1．团队协作能力 | 25分 |
| 2．python综合运用能力 | 25分 |
| 3．态度认真、刻苦钻研、创新能力 | 10分 |
| 4．过程完成、对工具的使用、对github的运用 | 20分 |
| 5．课程设计答辩逻辑清晰，内容正确 | 10分 |
| 6. 课程设计期间的课堂考勤、遵守纪律 | 10分 |
| 总分 | 100分 |

注：按上述六项分别记分后求和，根据小组成员贡献率综合评定，记载个人最后成绩。

成绩等级：优（90分—100分）、良（80分—89分）、中（70分—79分）、及格（60分—69分）、60分以下为不及格。

**五、指导地点与时间**

本课程设计将安排在第20-21周，采用腾讯会议和QQ群的方式，以在线形式进行。具体安排如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 周次 | 星期一 | 星期二 | 星期三 | 星期四 | 星期五 |
| 第20周 | 第5-8节 | 第5-8节 |  | 第5-8节 |  |
| 第21周 |  | 第5-8节 |  | 第5-8节 |  |

执笔:徐 勇 日期：2021-1-8

审阅:钱小红 日期：2021-1-8

目录

[1. 小组人员 5](#_Toc4475)

[2. 小组选题 5](#_Toc31211)

[3. 环境搭建 5](#_Toc2900)

[4. 各成员任务 5](#_Toc19612)

[5. 选题背景 6](#_Toc26412)

[6.分析内容 6](#_Toc32629)

[7.数据清洗（宁洁） 7](#_Toc13594)

[8.数据处理（郭晓东） 7](#_Toc13661)

[9. 数据分析（阮汉成） 7](#_Toc15086)

[10. 总结（阮汉成） 17](#_Toc6492)

[参考文献 17](#_Toc27172)

[课程设计成绩评定表 17](#_Toc9480)

# 小组人员

郭晓东、宁洁、阮汉成

# 小组选题

北上广深租房状况分析

# 环境搭建

硬件环境：Win10

软件环境：Python3.9.0、MySql 5.7、jupyter notebook、Github

# 各成员任务

宁 洁：负责数据清洗、PPT制作

阮汉成：负责数据分析、任务进度汇总

郭晓东：负责数据处理、答辩

# 选题背景

转眼到了大三，再过不到一年时间我们就要毕业了，到了毕业之后大家肯定都会想要去大城市发展，那么如果说去了北上广深，我们是没有能力直接买一套房子的，所以说，在考虑租房的情况下，租什么样的房既经济又实惠就成了我们的优先考虑了。北上广深四个一线城市，哪个城市的租房更加友好呢？城市各区域的租金是怎么分布的？离地铁口远近、房子大小、是否公寓房、精装简装、集中供暖等，租房时应该优先考虑哪些因素呢？一堆问题困扰着我们。

所以说我们要进行一个租房数据的分析，得出相应的结论，给小伙伴们提供最好的租房建议和房屋各种情况的一个说明。

我们获取了链家网北上广深四个城市的所有租房数据（时间节点：2019年2月25日），数据共有20000条。

对这些数据进行各种层次的分析，以得出我们的结论。

# 6.分析内容

（1）各城市房源数量

（2）各地区数据清洗后的房源数量

（3）单位面积房价

（4）北上广深GDP情况

（5）据地铁口远近对每平米租金的影响

（6）房屋大小对每平米租金的影响

（7）个人房和公寓房

# 7.数据清洗（宁洁）

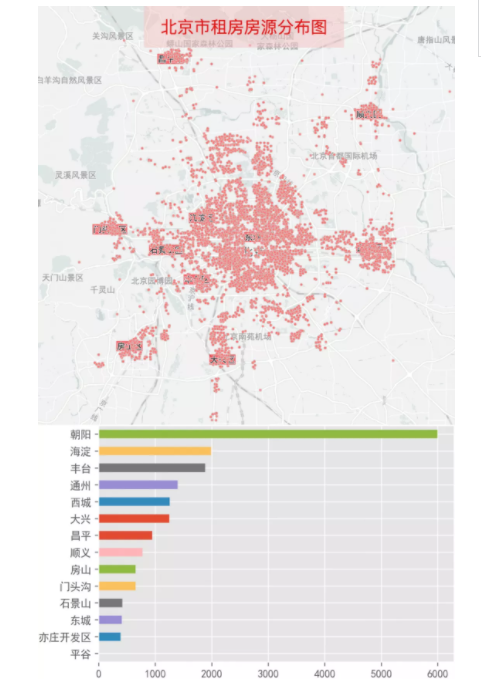
# 8.数据处理（郭晓东）

# 数据分析（阮汉成）

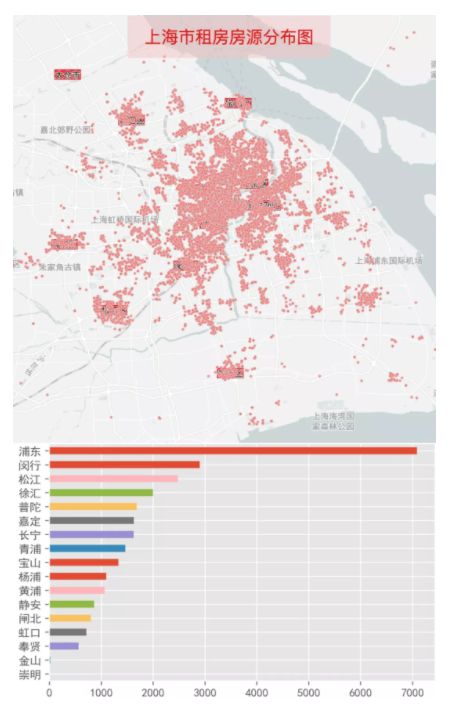
### 9.1各城市的租房房源分布怎么样？

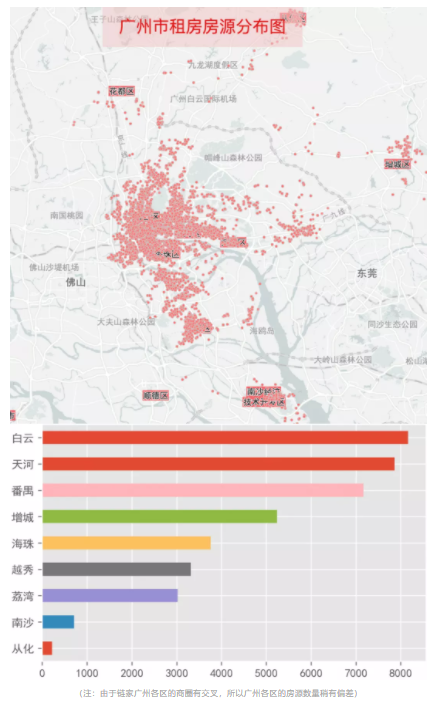
租房分布，也就是租房房源都在城市的哪个区域更多。

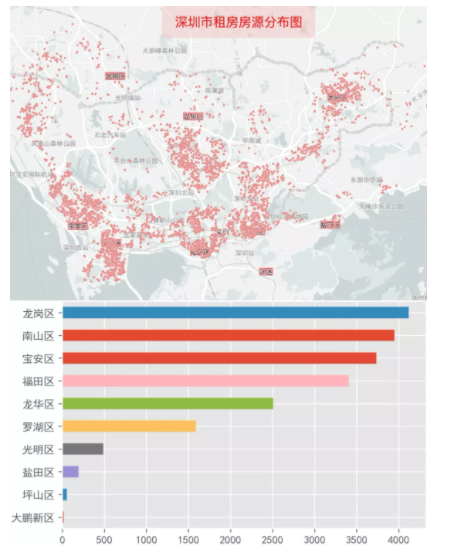
我们把北上广深四个城市的房源都以小点的形式投射在地图上，先来看看北京的。



北京的租房房源有18012条，可以看到除了周边的昌平、顺义、通州、大兴、房山、门头沟等区中心有集中房源外，北京的房子主要集中在二环外四环内，当然这片区域也是最贵的。其中朝阳区的房源最多，占了整个北京房源的1/3，要不咋说朝阳群众666呢。海淀、丰台次之。

  
上海也一样，除了周边的宝山、嘉定、青浦、松江、奉贤等区中心有房源外，房源主要集中在中心城区+浦东（地铁网范围内）。总共27311条房源，浦东就有7000多条，比例超过1/4。

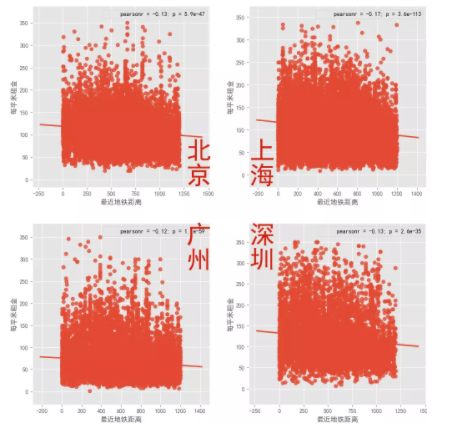
  
广州房源在四个城市中是最多的，共有39457条，主要集中在白云、天河、越秀、荔湾，以及海珠和番禺。其中白云、天河和番禺房源都超过6000条，选择丰富，不过看图也知道，3号线通勤压力巨大。



深圳的房源共有20054条，更集中在各区，除了四个新区零星分布的房源外，主要还是分布在宝安、南山（科技园）、福田、罗湖、龙岗区中心，以及坂田、布吉、3号线沿线，11号线沿线（是市中心的租金太贵了吧？）

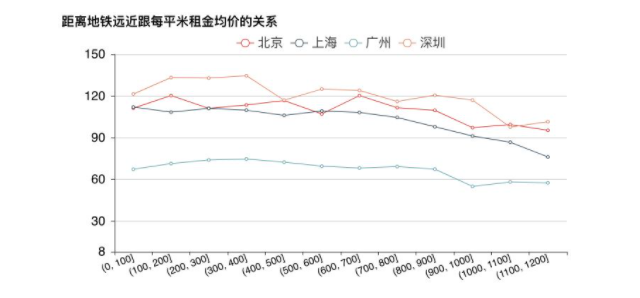
### 9.2各城市房源距离地铁口远近跟租金的关系

租房，其中一个终点考虑的因素就是距离地铁的远近。我们把个城市房源距离最近地铁站的距离跟每平米租金进行了回归分析，并且计算了相关系数。



可以看到，最近地铁距离和每平米租金之间当然是有相关性的，距离地铁越近，租金越贵。北上广深的相关系数分别是-0.13，-0.17，-0.12，-0.13，可见，上海地铁站的距离对租金的影响最大，广州最小。

我们还计算了各城市地铁距离每100米的租金均价。



可见，北京市只要地铁在900以内，每平米租金变化不大。900米和1000米租金差距为12.5元，也就是说，租一个20平的房子，距离地铁站900米和1000米的租金差距是250元。

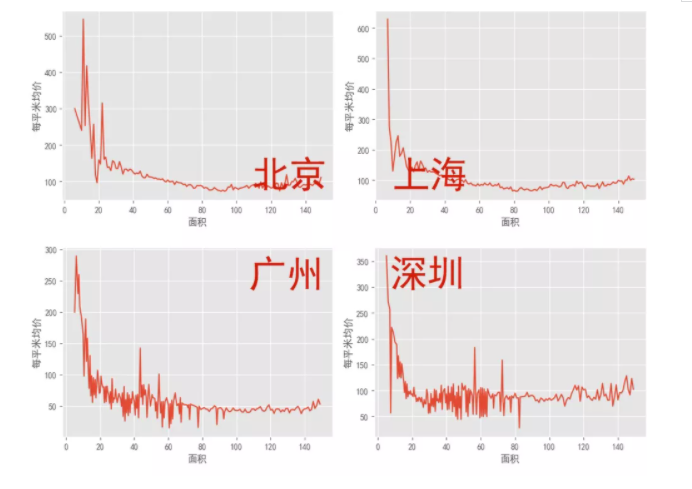
上海市只要地铁在700以内，每平米租金变化不大，700以上租金开始明显下降。

广州市每平米租金明显比其它城市低。同样，只要地铁在900以内，租金变化不大。900米和1000米租金差距为12元。

深圳市只要地铁在400以内，租金变化不大。400米和500米租金差距为17.6元。也就是说，租一个20平的房子，距离地铁站400米和500米的租金差距是352元。

### 9.3房屋大小对每平米的租金影响如何？

先来看这么一张图：

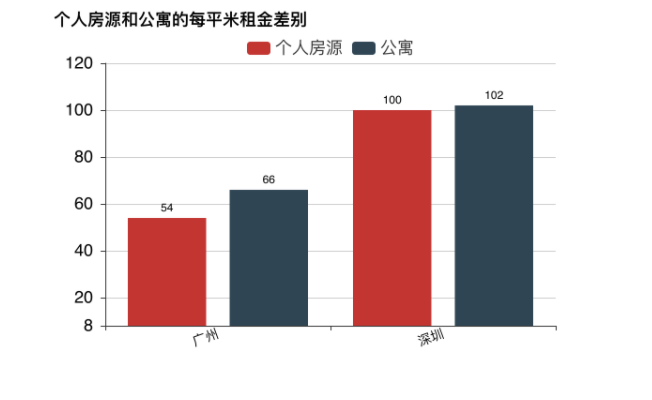


我们把各城市房源的面积和每平米均价绘制出关系图。可以看到，不管是哪个城市，出租面积在15平米以内的房子每平米租金都是最贵的，当然这里不排除租金贵的区域有更多单独出租的小房间这个原因。但是还是可以看出来，如果有认识的好友一起租大房子，不仅每平米的租金更便宜，而且还可以有一个大点的公用客厅。

### 9.4租个人房源好还是公寓好？

现在公寓房越来越多，我们在面临选择的时候，都会想是租个人房源好还是公寓好呢？个人房源质量参差不齐，遇上好的真不容易。而公寓统一装修，风格现代，但是却有各种各样的问题（甲醛、隔音、乱收费等）。

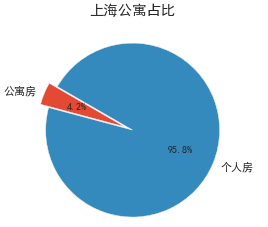
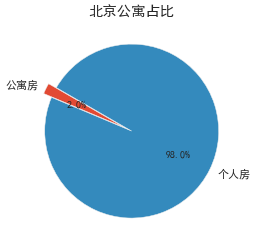
这里给大家提供另一维度的思考：价格。

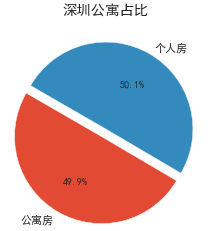
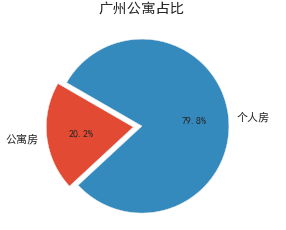


广州和深圳公寓数量占总房源数量的20%和51%（深圳怎么那么多公寓？）。可见广州和深圳公寓都要比个人房源贵一些，广州平均每平米贵12元，深圳贵2元。

这看起来也贵不了多少啊。真的是这样吗？

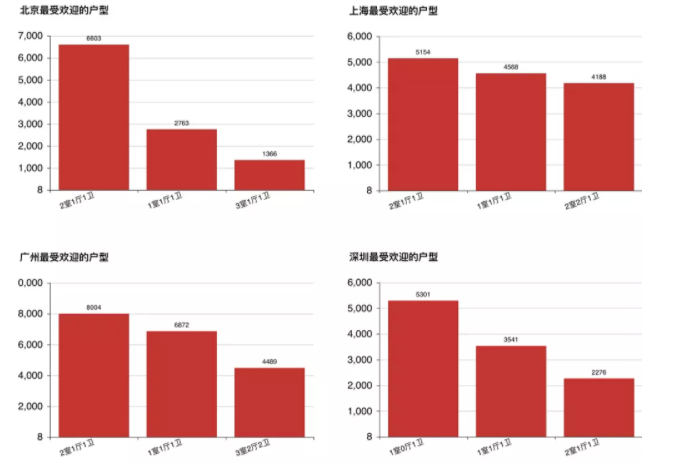
各地所爬数据公寓房占比：





### 9.5 北上广深租房时都看重什么？

我们统计了北上广深各城市房源最多的3种户型。



可见，北上广都偏爱2室1厅1卫的房子，而深圳却是1室0厅1卫的房子最多，看来，在深圳奋斗的人，能住带厅的房子已经不容易了。不过四个城市都是小户型出租的居多。

# 总结（阮汉成）

在开始学习Python这门课程之前，我对用Python进行数据分析有一个误解：我必须不得不对Python编程特别精通才行。

 经过这次的课程设计之后，更是深有体会，虽然在这其中也遇到了很多的困难，毕竟是第一次进行数据分析，但是和小组伙伴之间的交流和网上的资料查阅还是解决了不少的困难疑惑，也让我体现到了其中带过的快乐，让我懂得了不过是现在还是以后在工作上，你想变得更加优秀，做出更好的东西，都得有一个好的团队，俗话说的好人多就是厉害，的确这也很好的体现了。

最后我也是非常感谢我的组员，有什么困难都可以一起去解决，互相沟通，也是很感谢老师和同学在学习中给我带来的帮助。

# 参考文献

[1]Python基础与大数据应用.人民邮电出版社，2017年1月

[2]百度文库

[3]CSDN

[4]百度百科

**课程设计成绩评定表**

|  |  |  |  |
| --- | --- | --- | --- |
| 课程设计题目 | 数据分析的学习与实践 | | |
| 课程设计学生答辩或质疑记录：   1. 数据分析基于什么前提？   答：基于数据的前提，因为有数据整理汇总我们才能够进行数据分析，从而得出结论，若是没有数据为基础所有的分析都只是空谈，所以我认为数据分析是用数据来说话，比真正的空谈更为实际 | | | |
| **评 分 依 据** | | **分 值** | **评分成绩** |
| 1．团队协作能力 | | 25分 |  |
| 2．python综合运用能力 | | 25分 |  |
| 3．态度认真、刻苦钻研、遵守纪律 | | 10分 |  |
| 4．过程完成、对工具的使用、对github的运用 | | 20分 |  |
| 5．课程设计答辩逻辑清晰，内容正确 | | 10分 |  |
| 6. 课程设计期间的课堂考勤、创新能力 | | 10分 |  |
| 总 分 | | 100分 |  |
| 最终评定等级为：  指导老师签字：  2021 年 1 月 15日 | | | |