|  |  |
| --- | --- |
| 学号 | 10214518329 |



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 | \_\_\_\_\_\_\_\_\_\_\_宁 洁\_\_\_\_\_\_\_\_\_\_\_\_ |
| 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](#_Toc17084)

[2. 小组选题 5](#_Toc3745)

[3. 环境搭建 5](#_Toc16616)

[4. 各成员任务 5](#_Toc2399)

[5. 选题背景 6](#_Toc28294)

[6.分析内容 6](#_Toc11807)

[7.数据清洗（宁洁） 7](#_Toc25112)

[8.数据处理（郭晓东） 14](#_Toc28181)

[9. 数据分析（阮汉成） 14](#_Toc4562)

[10. 总结（宁洁） 14](#_Toc20781)

[参考文献 14](#_Toc20055)

[课程设计成绩评定表 15](#_Toc30687)

# 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）

### 7.1 csv文件信息读取

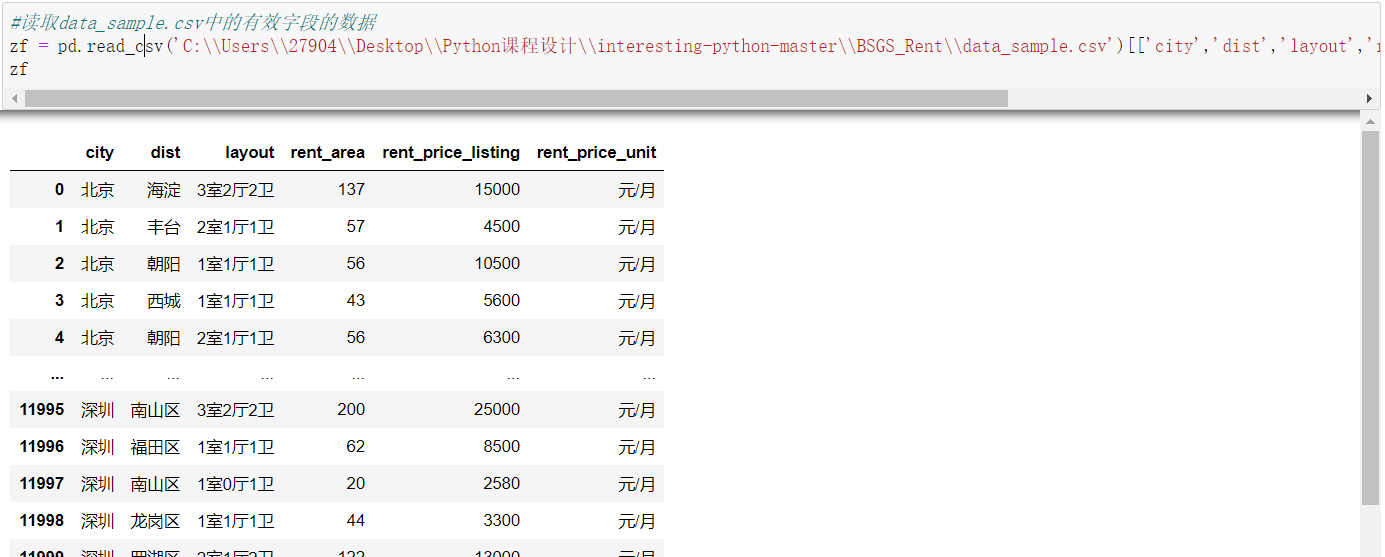
##### 7.1.1 包、库的导入和界面美化



##### 7.1.2读取data\_sample.csv中的有效字段的数据

zf=pd.read\_csv('C:\\Users\\27904\\Desktop\\Python课程设计\\interesting-python-master\\BSGS\_Rent\\data\_sample.csv')[['city','dist','layout','rent\_area','rent\_price\_listing','rent\_price\_unit']]

zf



### 7.2 去重

##### 7.2.1读取原始行数

zf = pd.read\_csv('C:\\Users\\27904\\Desktop\\Python课程设计\\interesting-python-master\\BSGS\_Rent\\data\_sample.csv')[['city','dist','layout','rent\_area','rent\_price\_listing','rent\_price\_unit']]

print("原始行数:",len(zf))



##### 7.2.2去重，并生成新索引删除原索引

zf=zf.drop\_duplicates().reset\_index().drop('index',axis=1)

print('现在数据行数:',len(zf))

zf

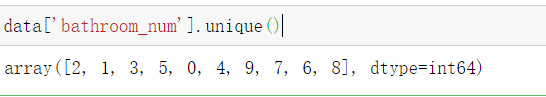


### 7.3 数据专项清洗

##### 7.3.1 浴室信息处理

**1.浴室不同信息集合**

data['bathroom\_num'].unique()



**2.查找异常数据**

data[data['bathroom\_num'].isin(['8','9'])]



**3.异常数据的处理**

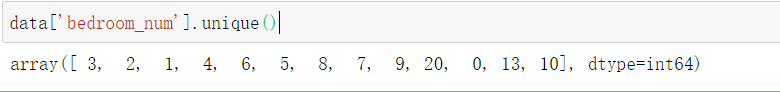
data = data.drop([1206])



##### 7.3.2 卧室信息处理

**1.卧室不同信息集合**

data['bedroom\_num'].unique()



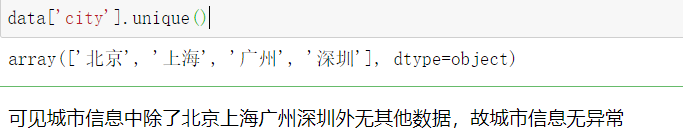
**2.查找异常数据**

data[data['bedroom\_num'].isin(['10','13','20'])]



##### 7.3.3 城市信息处理

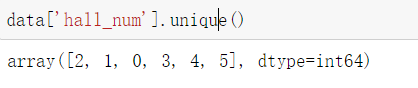
data['city'].unique()



##### 7.3.4玄关信息处理

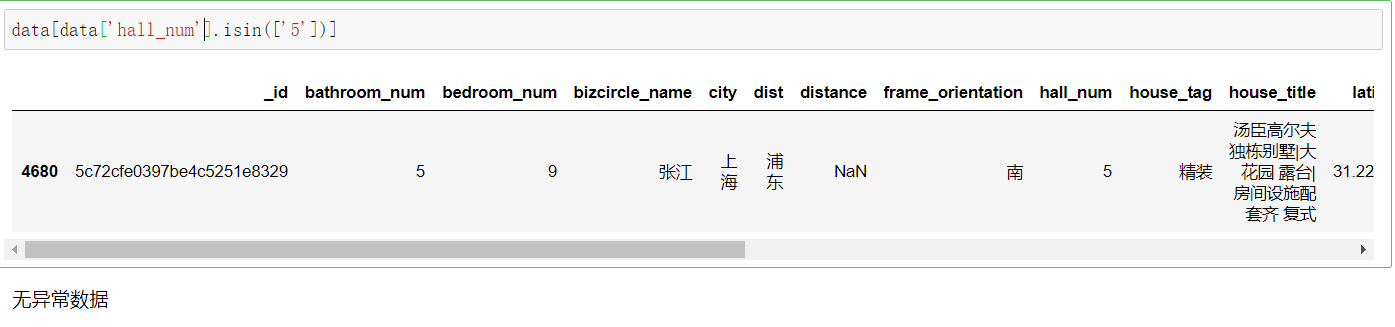
**1.玄关不同信息集合**

data['hall\_num'].unique()



**2.查找异常数据**

data[data['hall\_num'].isin(['5'])]



##### 7.3.5 租房面积信息处理

**1.租房面积类型转换**

#rent\_area字段有些填写的是一个范围，比如23-25平方米，后期转换成“float”类型的时候不好转换，考虑取平均值

def get\_aver(data):

if isinstance(data, str) and '-' in data:

low, high = data.split('-')

return (int(low)+int(high))/2

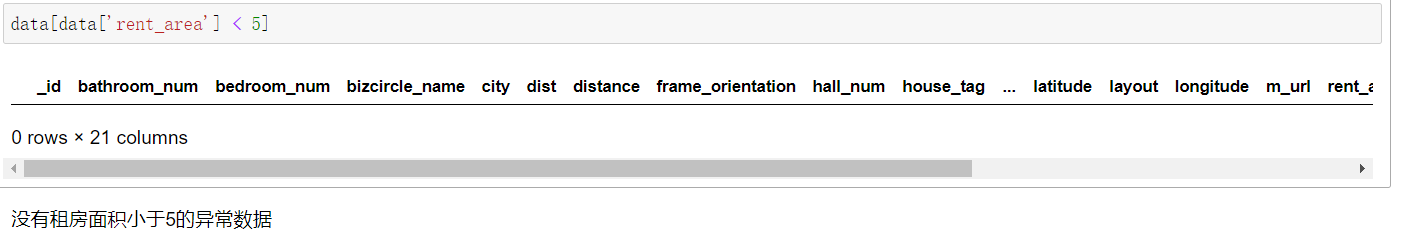
else:

return int(data)

data['rent\_area'] = data['rent\_area'].apply(get\_aver)

**2.查找异常数据**

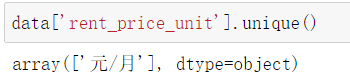
data[data['rent\_area'] < 5]



##### 7.3.6租房价格单位信息处理

**1.查找信息集合**

data['rent\_price\_unit'].unique()



**2.删除多余列**

租金都是以“元/月”计算的，所以这一列没用了，可以删了

data = data.drop(columns='rent\_price\_unit')

##### 7.3.7租房价格信息处理

**1.找出带“-”的价格区间**

data[data['rent\_price\_listing'].str.contains('-')].sample(100)



**2.处理价格区间**

#价格是有区间的，需要按照处理rent\_area一样的方法处理\ndata['rent\_price\_listing'] = data['rent\_price\_listing'].apply(get\_aver)","execution\_count":120,"outputs":[]}]

**3.数据类型转换**

for col in ['bathroom\_num', 'bedroom\_num', 'hall\_num', 'rent\_price\_listing']:

data[col] = data[col].astype(int)

# 8.数据处理（郭晓东）

# 数据分析（阮汉成）

# 总结（宁洁）

经过这次python的课程设计，使我收获很多，却也发现了自己在python学习上存在的一些问题，也解决了一些一直困扰我的问题，以前觉得程序很简单，不放在心上，经过学习才使我明白做程序需要的东西是很多的，你既要懂得程序，而且还要有耐心和恒心，这次课程设计既锻炼了我的动手能力也磨练了我的心理素质。  
这次的课程设计让我们复习巩固了python的基础知识，进一步加深对python的理解，深入了解了jupyter和github的使用方法，利用所学知识把理论和实际结合，在接下来的时间里我要好好学习python这门课程。

# 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. 数据清洗的过程是怎么样的？   答：数据清洗是先找出某一列中所有相同的数据有哪些，再根据数据分析明显不合适的数据进行查找，如果说查找到的数据是异常数据，则对该行数据进行删除。   1. 在数据清洗时遇到哪些困难？   答：数据清洗过程中遇到了范围数据，处理起来比较麻烦，先是定义了一个函数，将最大值最小值加起来除以二再赋值给原始数据，这么解决的。 | | | |
| **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 | | | |