|  |  |
| --- | --- |
| 学号 | 10214518334 |



课 程 设 计

课程名称 python程序设计课程设计

|  |  |
| --- | --- |
| 题 目 | 数据分析的学习与实践-节假日长假景点人满为患的数据分析 |
| 专 业 | \_\_\_\_ \_信 息 工 程\_\_\_\_\_ \_\_\_ |
| 班 级 | \_\_\_\_\_\_ 软件工程1182\_\_\_\_\_\_\_ \_ |
| 姓 名 | \_\_\_\_\_\_\_\_朱 雪 婷\_\_\_\_\_\_\_\_\_\_\_ |
| 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、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 |

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

[一、课程设计目的与任务 2](#_Toc32744)

[二、课程设计的内容与基本要求 2](#_Toc14896)

[三、学时分配进度安排 3](#_Toc19439)

[四、课程设计考核及评分标准 3](#_Toc10931)

[五、指导地点与时间 4](#_Toc21961)

**[一．设计目的 6](#_Toc19471)**

**[二．开发环境 6](#_Toc17962)**

**[三．总体设计 6](#_Toc16858)**

**[四．分析的问题 11](#_Toc27062)**

**[五、 GitHub运用 19](#_Toc22159)**

[5.1 GitHub下载安装 19](#_Toc812)

[5.2 创建版本库 19](#_Toc653)

[5.3 SSH配置 20](#_Toc6321)

[5.4 添加远程库 20](#_Toc28062)

**[六、 总结 20](#_Toc27705)**

**[课程设计成绩评定表 22](#_Toc28146)**

一．设计目的

项目主要按全国各个城市抓取飞猪“景点门票”栏的景点门票销售数据，并且分析五一哪些景点会人挤人，哪些景点值得一去。

“旅游可以放松自己身心，缓解工作和生活方面的压力，旅行的过程中是不需要有任何的心理负担的，可以使人达到一种完全放松的状况，感受最原始的快乐。旅行的时候不仅可以观赏风景，还可以尝美食，住宾馆，听故事，感受世界各地的饮食文化，增长自己的见识，可以看到更多的人，了解更多的民俗文化，看到更多的景色使自己的内心”。

本研究主要通过分析各项数据来对旅游项目更加完善的优化提出建议，来改善旅游体验。我们都希望自己在有限的时间里面可以获得不同的旅游体验。景点越多的地方，当然旅游选择就越多，可以获得的体验就更具选择性。

二．开发环境

系统环境：windows。

运行环境：python3.6，jupyter。

需要安装的包：

* requests 2.25.1
* pyecharts 0.1.9.4
* pyecharts-javascripthon 0.0.6
* pyecharts-jupyter-installer 0.0.3
* pyecharts-snapshot 0.2.0
* pandas 0.24.1
* numpy 1.19.2
* pymysql 1.0.2
* scrapy

主要的文件为：

* city\_data.csv: 全国城市及所属省份列表
* get\_data.py: 数据爬取代码
* laborday.ipynb:Jupyter notebook代码，对景点门票数据进行分析

上传方式：github

三．总体设计

3.1导入必要的包：

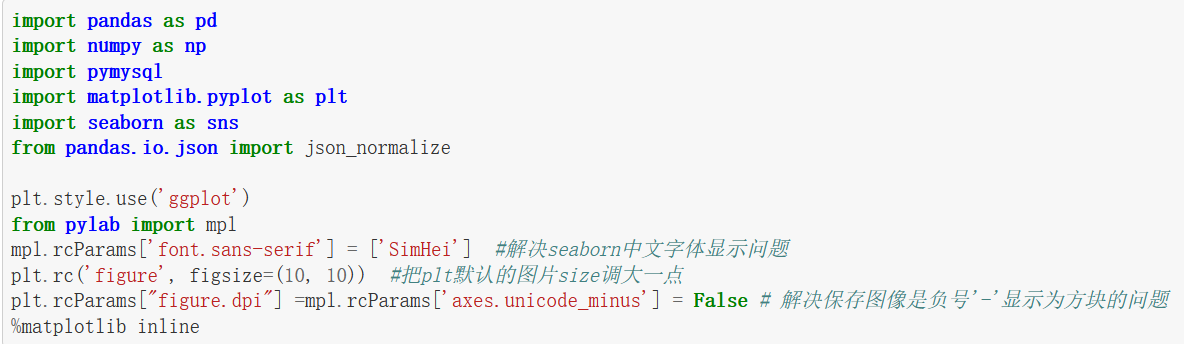
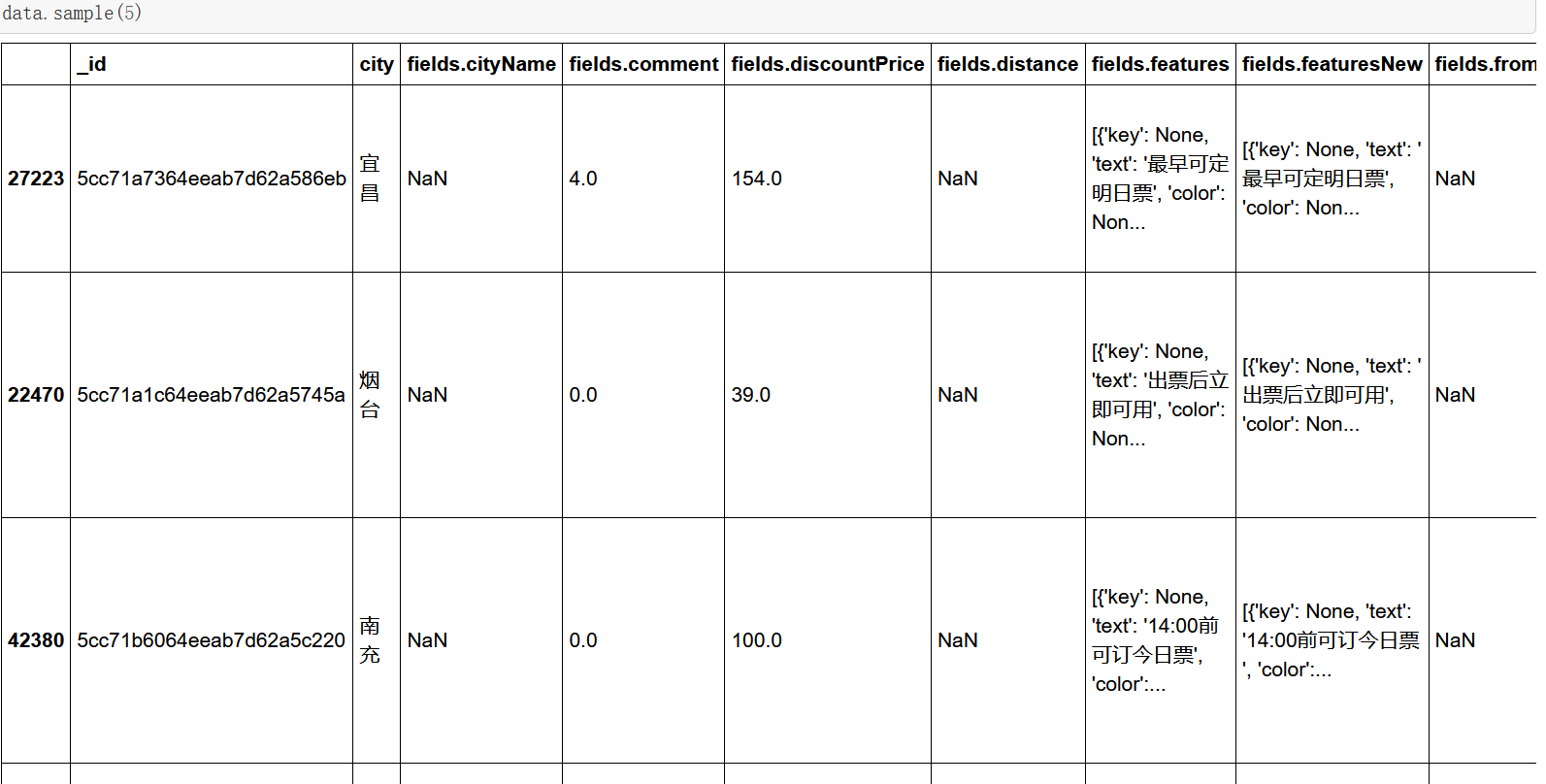


图3.1导入必要的包

3.2读取数据：



图3.2.1读取数据



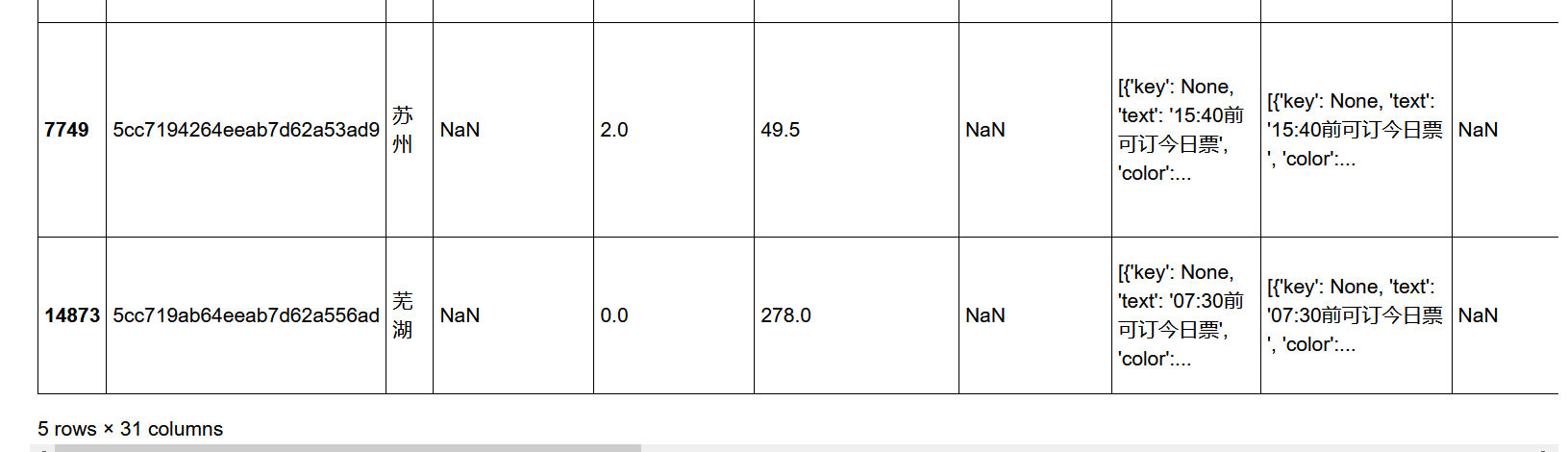
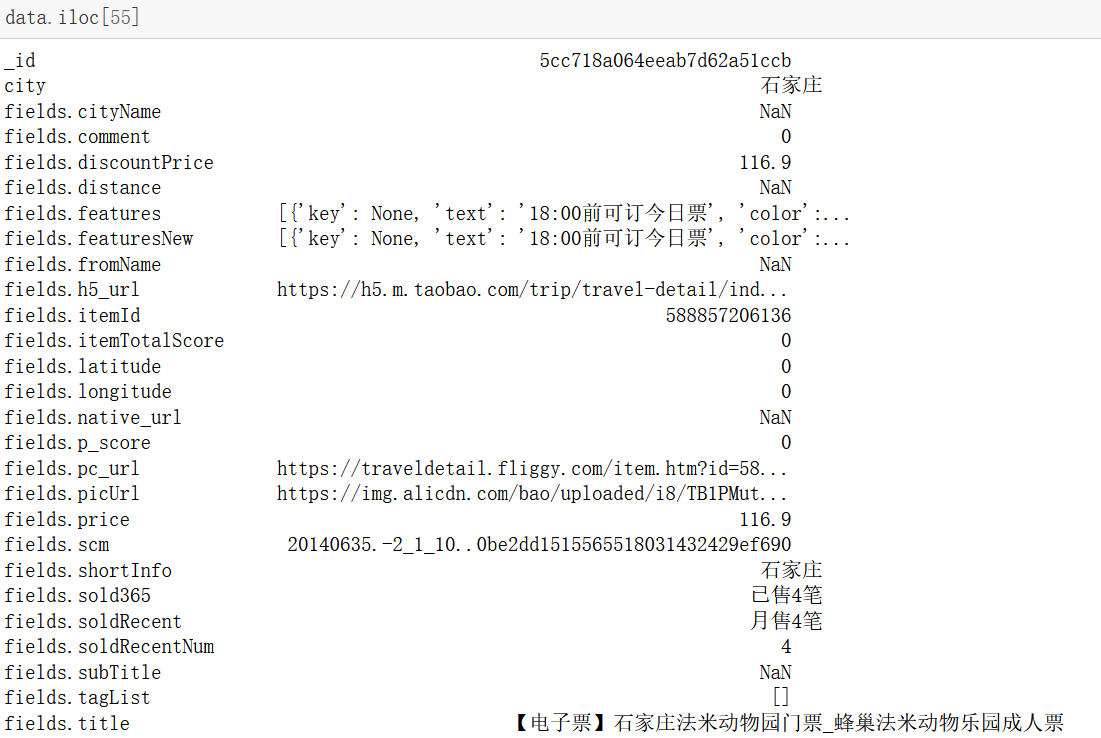


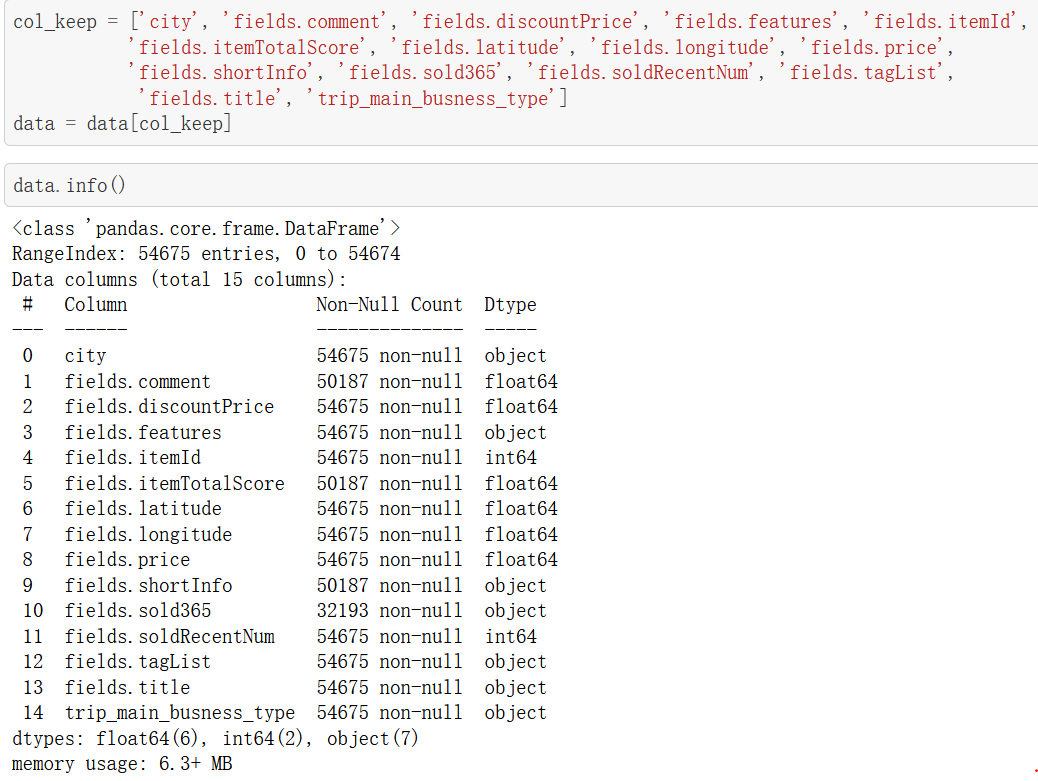
图3.2.2随机抽取5行数据

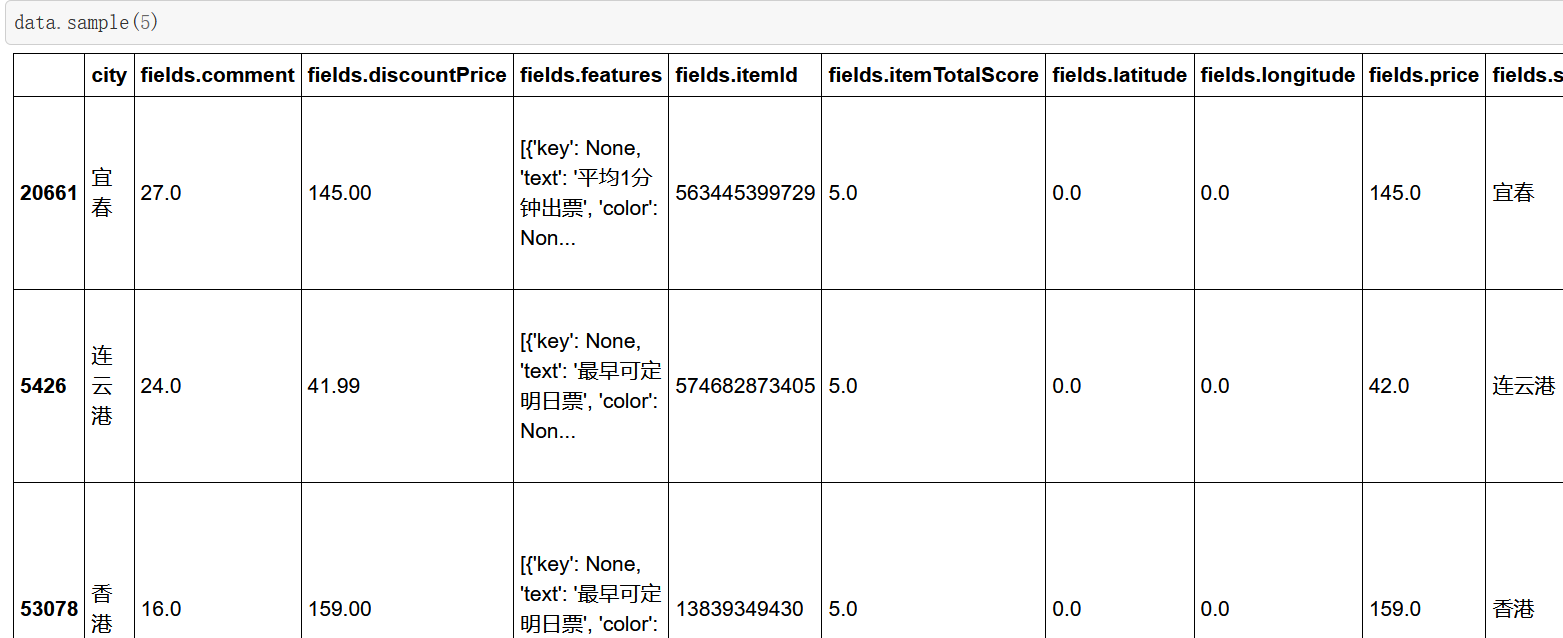
3.3清洗数据

3.3.1 把需要的字段挑出来







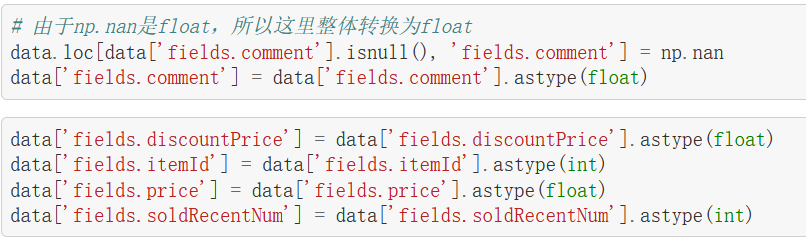




3.3.2 去重

data.drop\_duplicates(subset='fields.itemId', inplace=True)

3.3.3 数据类型转换



3.3.4 各字段清洗

* fields.features
* fields.sold365
* fields.tagList



图3.3.4.1fields.features清洗

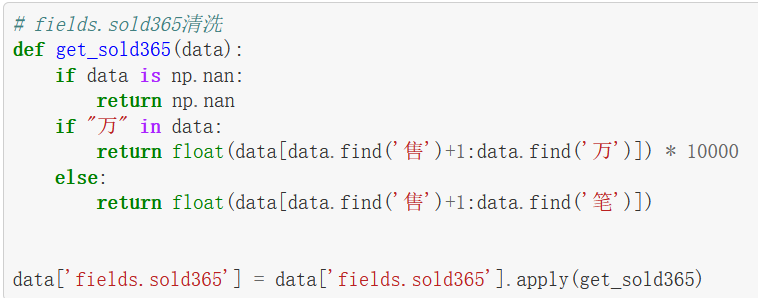
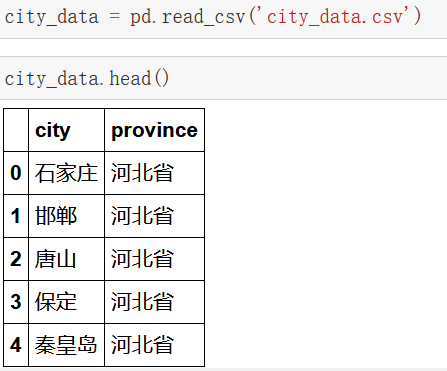


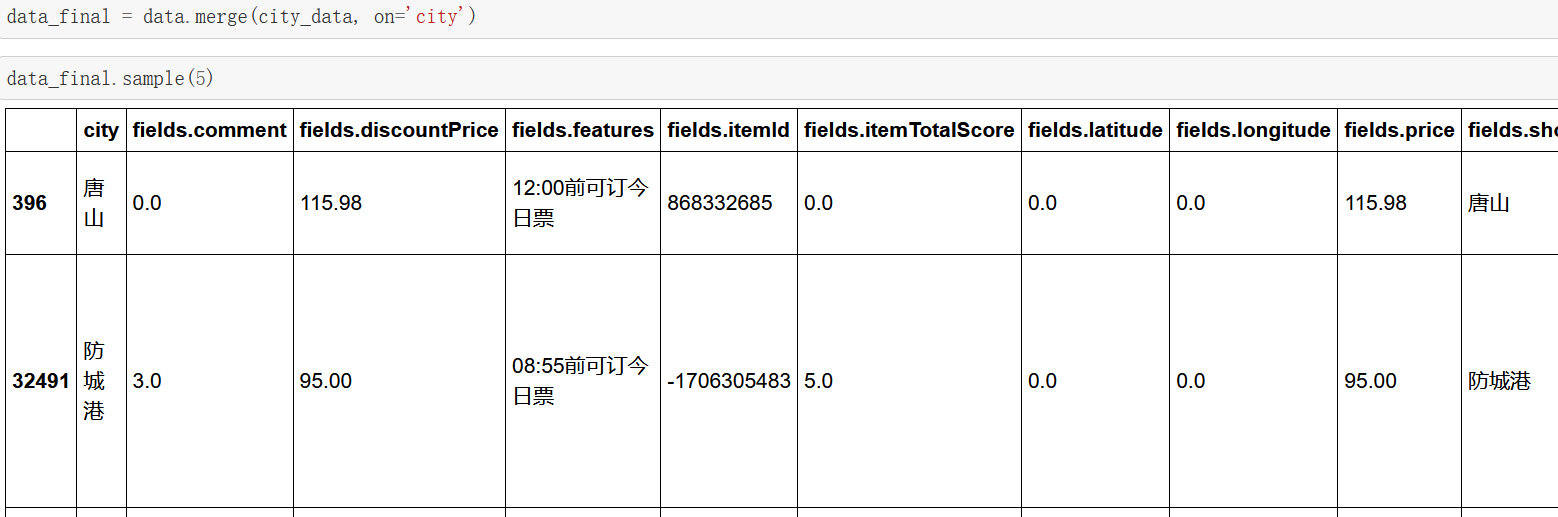
图3.3.4.2fileds.sold365清洗



图3.3.4.3fields.tagList清洗

3.5 跟省份数据合并在一起







四．分析的问题

1. 哪些城市/省份的旅游选择最多？
2. 哪些城市最受游客青睐？
3. 最热门的景点
4. 各省份的好评热门景点
5. 很热门但是评分不好的景点
6. 打折力度最大的景点
7. 各级别景区的门票价格

4.1哪些城市/省份的旅游选择最多？

使用漏斗图展示个城市的旅游选择情况：



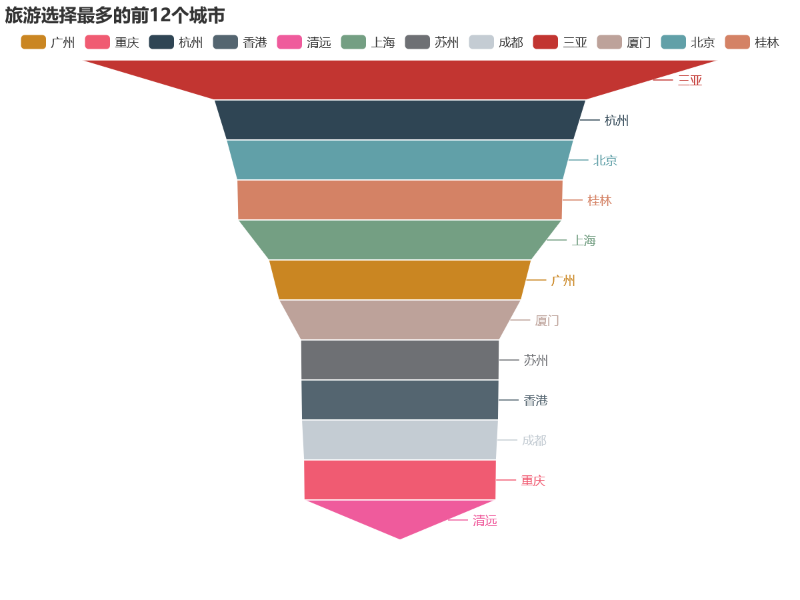


图4.1.2旅游选择最多的前12个城市

使用地图展示各省的旅游选择情况：



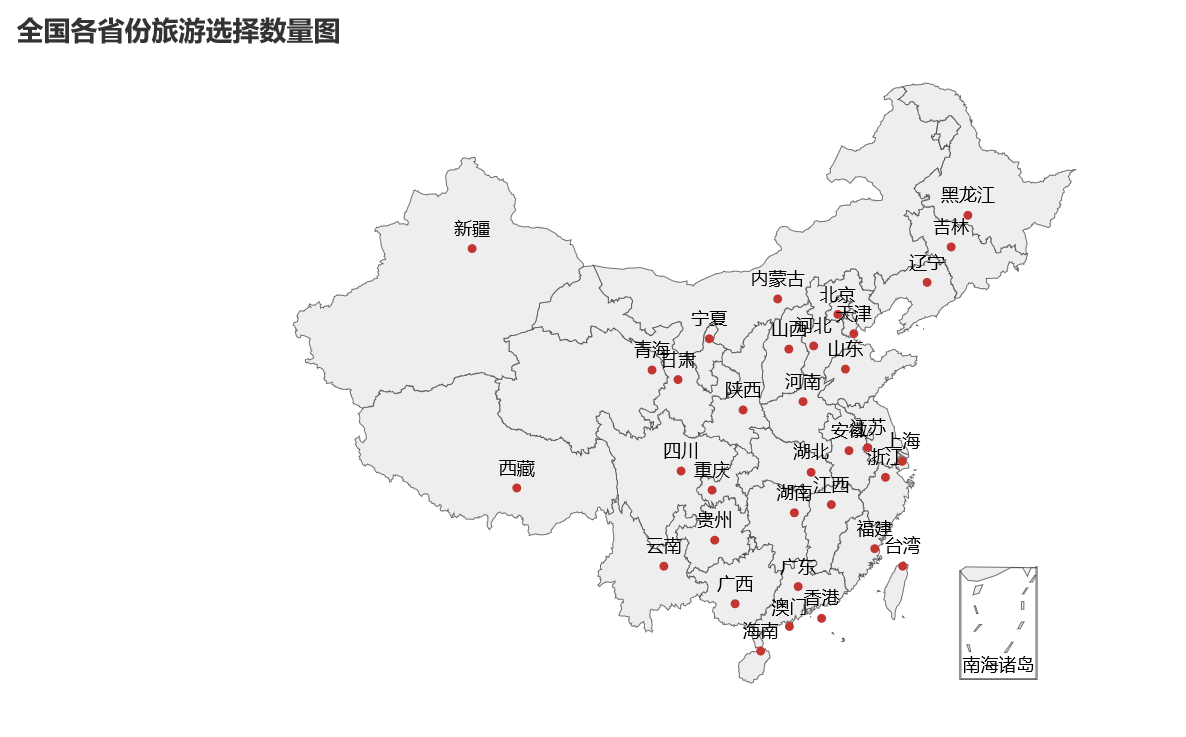


图4.1.2 全国各省分旅游选择数量图

4.2哪些城市最受游客青睐？

先分析一下售票情况



结果如下

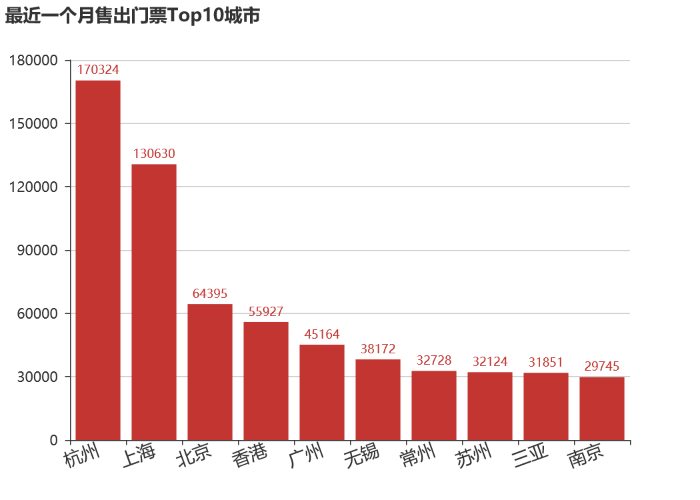


图4.2.1最近一个月出售门票Top10城市

接下来是评分情况



结果如下：分数越高，字体就会越大。

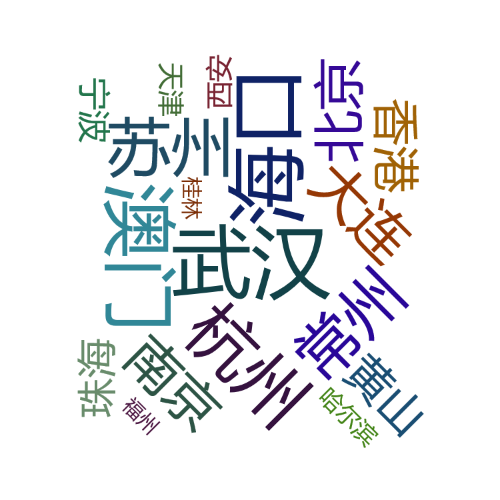
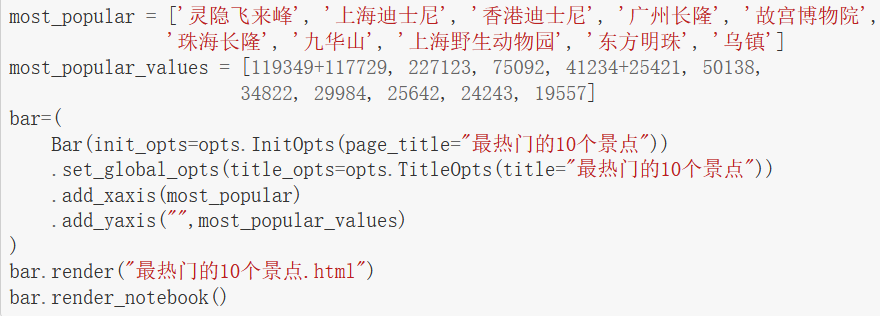


图4.2.2评分情况

4.3最热门的景点



结果如下：

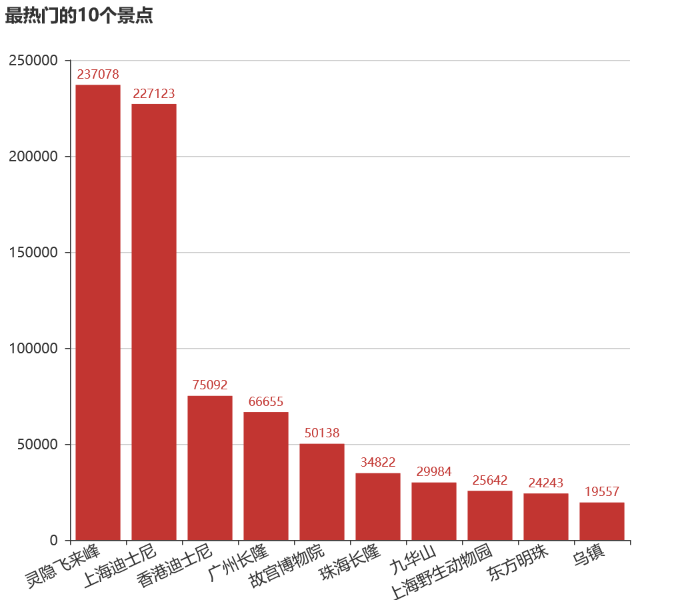


图4.3最热门的10个景点

4.4各省份的好评热门景点



结果如下：



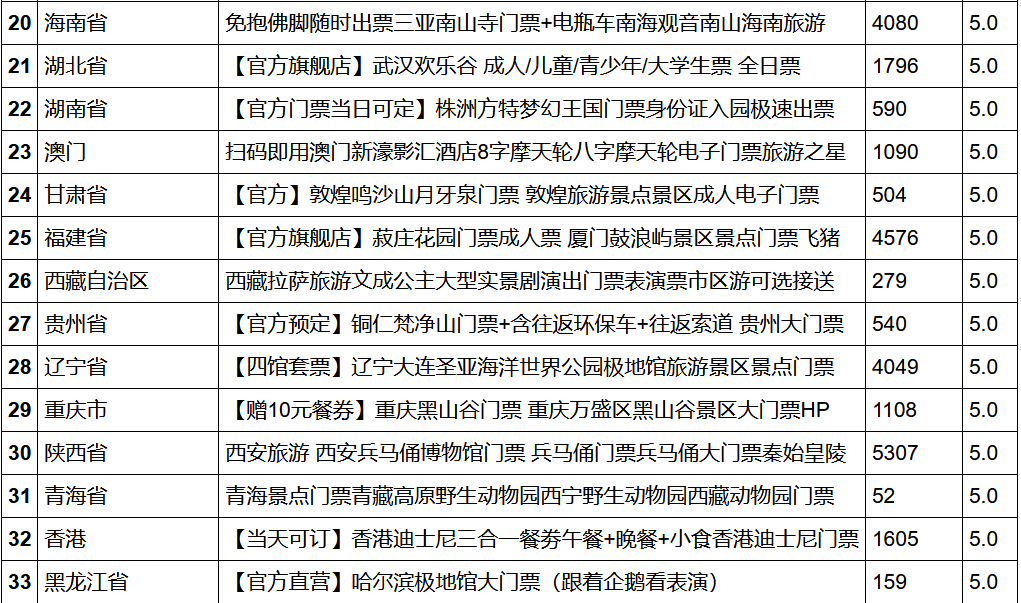
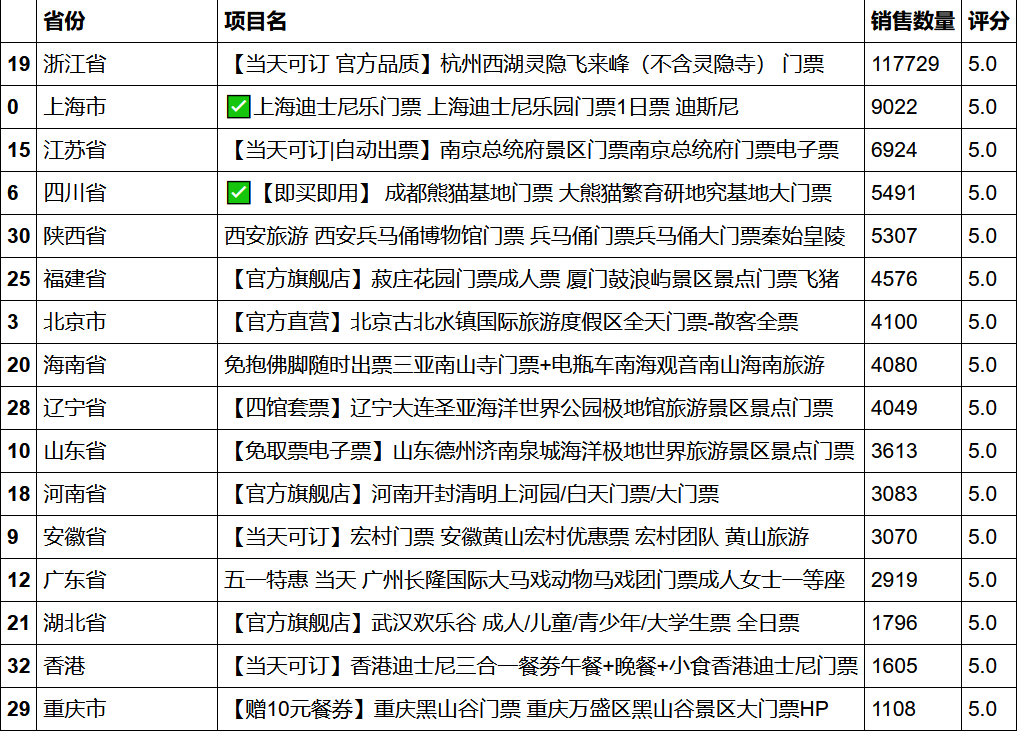


图4.4.1各省份的好评热门景点



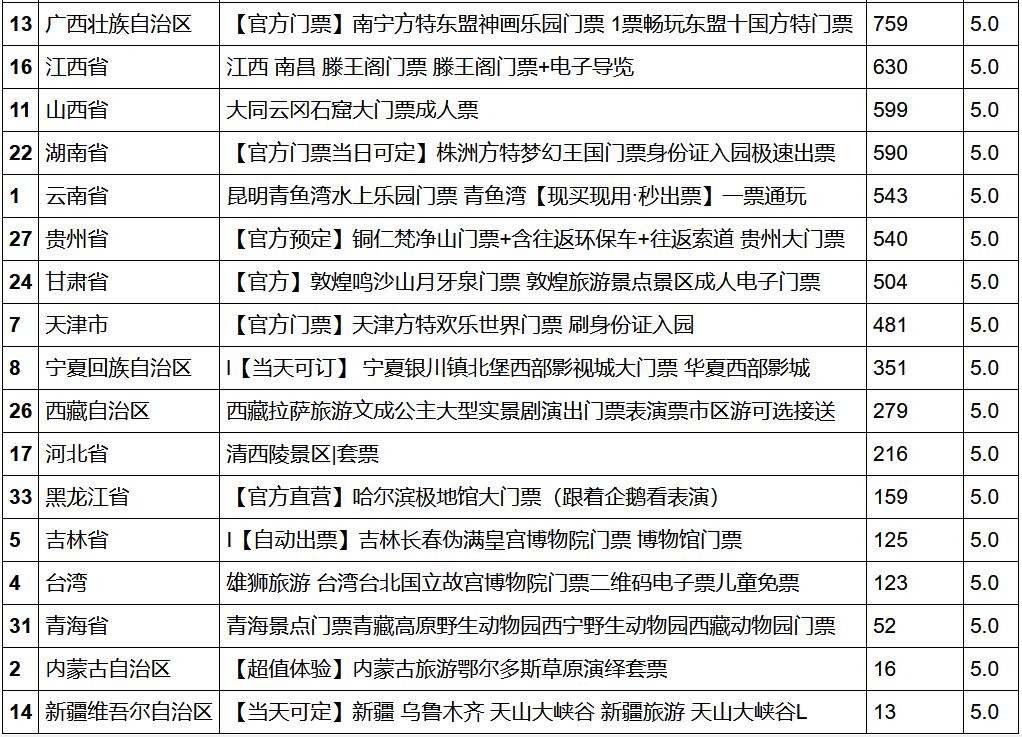


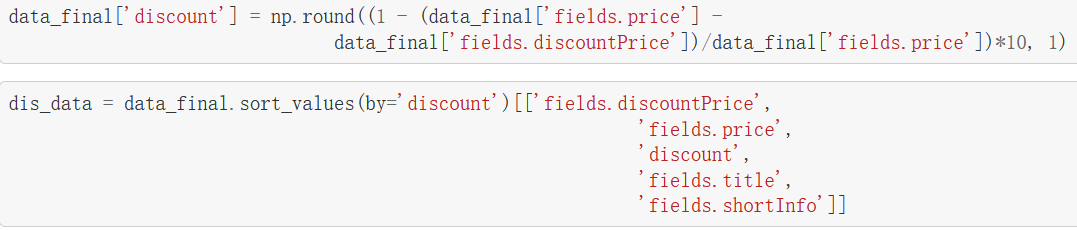
图4.4.2销售数量排名

4.5热门但是评分不太高的景点



图4.5评分不太高的景点

4.6打折力度最大的景点



结果如下：



图4.6打折力度最大的景点

4.7各级别景区的门票价格



结果如下：由结果可见，星级越高，则平均门票价钱也越高。

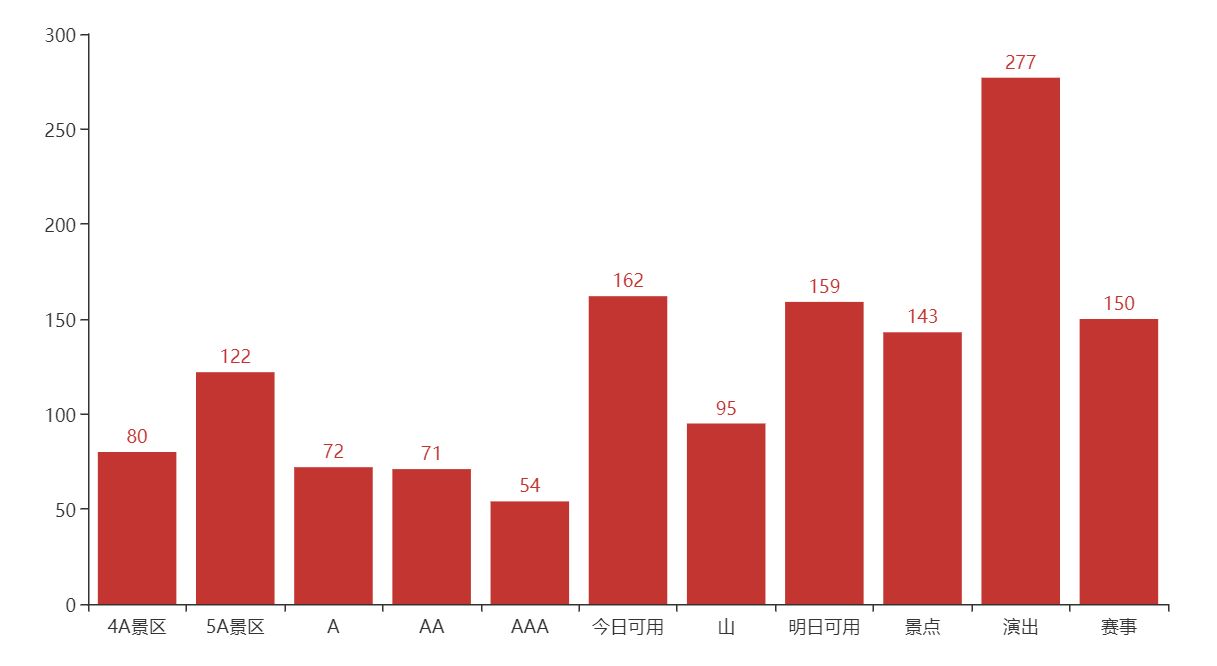


图4.7各级别景区的门票价格

1. GitHub运用

5.1 GitHub下载安装

直接在官网下载，然后按默认选项安装即可，安装完成后，在开始菜单里找到“Git”->“Git Bash”，蹦出一个类似命令行窗口的东西，就说明Git安装成功。

安装完成后，还需要最后一步设置，在命令行输入：

$ git config --global user.name "Your Name"

$ git config --global user.email "email@example.com"

5.2 创建版本库

创建一个空目录：

$ mkdir learngit

$ cd learngit

$ pwd

/c/Users/dell/learngit

通过git init命令把这个目录变成Git可以管理的仓库：

$ git init

Initialized empty Git repository in/c/Users/dell/learngit/.git/

5.3 SSH配置

获取SSH Key

cd ~

rm -r .ssh

sh-keygen -t rsa -C 邮箱

cd .ssh

cat id\_rsa.pub

复制下来然后打开github点击左上角头像然后settings

进入ssh and gpg keys，然后ssh keys，点击new ssh key ，把复制的key粘在下面

git init

命令执行后，文件夹中会多一个.git 文件夹

git config --global user.name "<username>"

git config --global user.email "<email>"

git add .

git status

git commit -m "<message>"

git log

在 Repositories 标签下点击 New 按钮创建

git remote add origin git@github.com:<username>/<directory name>

git push origin master

5.4 添加远程库

登陆GitHub，然后，在右上角找到“Create a new repo”按钮，创建一个新的仓库：

在Repository name填入learngit，其他保持默认设置，点击“Create repository”按钮，就成功地创建了一个新的Git仓库：

在本地的learngit仓库下运行命令：

$ git remote add origin git@github.com:michaelliao/learngit.git

把本地库的所有内容推送到远程库上：

$ git push -u origin master

1. 总结

这次课程设计的完成，标志着python课设的结束，也是我对于勇于挑战自己的一个见证，我认真对待这次的课程实践，把这当作一次对于自己的考核，我在这个过程中，不仅仅学到了很多硬性的知识，也深切感受到自己用于挑战一件事带来的美好和成就感，以后我会更加有信心面对所有的挑战，不断锻炼自己获得成长。

这次python课程设计对节假日长假景点人满为患的数据进行分析，更加巩固了python的语法，而且这次课设了解到了jupyter notebook和GitHub的用法，还锻炼了与同学之间的合作能力。相比对结果的分析，这次课设复习了数据清洗和有关数据的图片分析。对数据去重、数据类型转换的用法更加熟悉了。本次在完成项目的过程中有一些失误，比如目标不够清晰，本末倒置，在最先开始的一周更多的时间花在了GitHub和jupyter的使用上；实现过程中做出的东西虽然功能完成的不错，但是没有亮点。

最后在答辩后，听取老师的建议后和对比同学的成果后，我们意识到，我们的不足和缺陷，我们的答辩和设计缺少一定的创新。

在做课设的时候,我一方面积累自己开发设计的经验和学习能力，另一方面也增加我对软件开发的理解和体会，我体会到完成功能是最基本的，也是带我兴奋的动力，而代码的健壮，系统的合理才是我们的追求和努力的方向。我在这次课程设计后，都体会到了﹐在认真做一件事后，自己得到的成长，以后自己也会认真的去做每件事。

**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. 为什么会用到pyecharts？   答：pyecharts 是一个用于生成 Echarts 图表的类库。Echarts 是百度开源的一个数据可视化 JS 库。用 Echarts 生成的图可视化效果非常棒，为了与 Python 进行对接，方便在 Python 中直接使用数据生成图。pyecharts最开始因为版本问题无法得到运行结果，最终使用的版本号为pyecharts 0.1.9.4。   1. python如何读取csv文件某列时保留空值？   答：取了一列带空值的dataFrame，统计元素个数的时候用这个dict(data['Gender'].value\_counts()) 就是默认去掉空值，那画图也就没有空值了，  可以用这个：dict(data['Gender'].value\_counts(dropna=False))，就会保留空值  dict(data['Gender'].value\_counts())  Out[101]: {1.0: 25, 2.0: 13}  dict(data['Gender'].value\_counts(dropna=False))  Out[102]: {nan: 40, 1.0: 25, 2.0: 13}   1. python去重方法？   答：1. 使用set的特型，python的set和其他语言类似, 是一个无序不重复元素集。   1. 使用keys()方法。 2. 循环遍历法。   按照索引再次排序。 | | | |
| **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、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 |  |
| 最终评定等级为：  指导老师签字：  2021 年 1 月 15日 | | | |