|  |  |
| --- | --- |
| 学号 | 10214518333 |



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 | 软件工程\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 班 级 | 软件1182\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 姓 名 | 杨书棋\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 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．选题背景和意义 6](#_Toc62331543)

[1.1选题背景 6](#_Toc62331544)

[1.2意义 6](#_Toc62331545)

[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。 6](#_Toc62331546)

[二．开发环境 6](#_Toc62331547)

[三．程序设计 7](#_Toc62331548)

[3.1导入必要的包： 7](#_Toc62331549)

[3.2读取数据： 7](#_Toc62331550)

[3.3清洗数据 8](#_Toc62331554)

[3.3.1 把需要的字段挑出来 8](#_Toc62331555)

[3.3.2 去重 10](#_Toc62331556)

[3.3.3 数据类型转换 10](#_Toc62331557)

[3.3.4 各字段清洗 10](#_Toc62331558)

[3.5 跟省份数据合并在一起 11](#_Toc62331559)

[四．数据分析 12](#_Toc62331560)

[4.1问题引入 12](#_Toc62331561)

[4.2哪些城市/省份的旅游选择最多？ 12](#_Toc62331562)

[4.3哪些城市最受游客青睐？ 14](#_Toc62331563)

[4.4最热门的景点 16](#_Toc62331564)

[4.5各省份的好评热门景点 17](#_Toc62331565)

[4.6热门但是评分不太高的景点 19](#_Toc62331566)

[4.7打折力度最大的景点 19](#_Toc62331567)

[4.8各级别景区的门票价格 20](#_Toc62331568)

[五．总结 21](#_Toc62331569)

[参考文献 22](#_Toc62331570)

一．选题背景和意义

1.1选题背景

“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”。

1.2意义

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：jupyter。

项目托管平台：GitHub。

编程语言：python。

需要安装的包：

* requests
* pyecharts
* pandas
* numpy
* pymysql
* 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

三．程序设计

3.1导入必要的包：

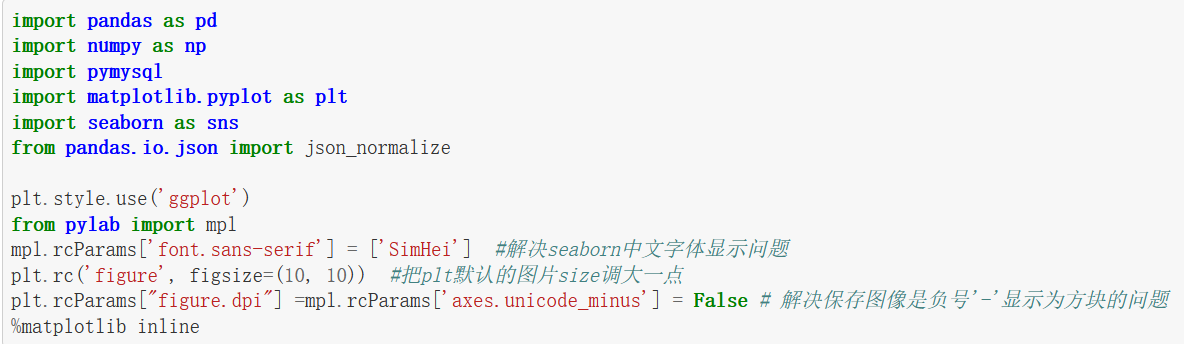
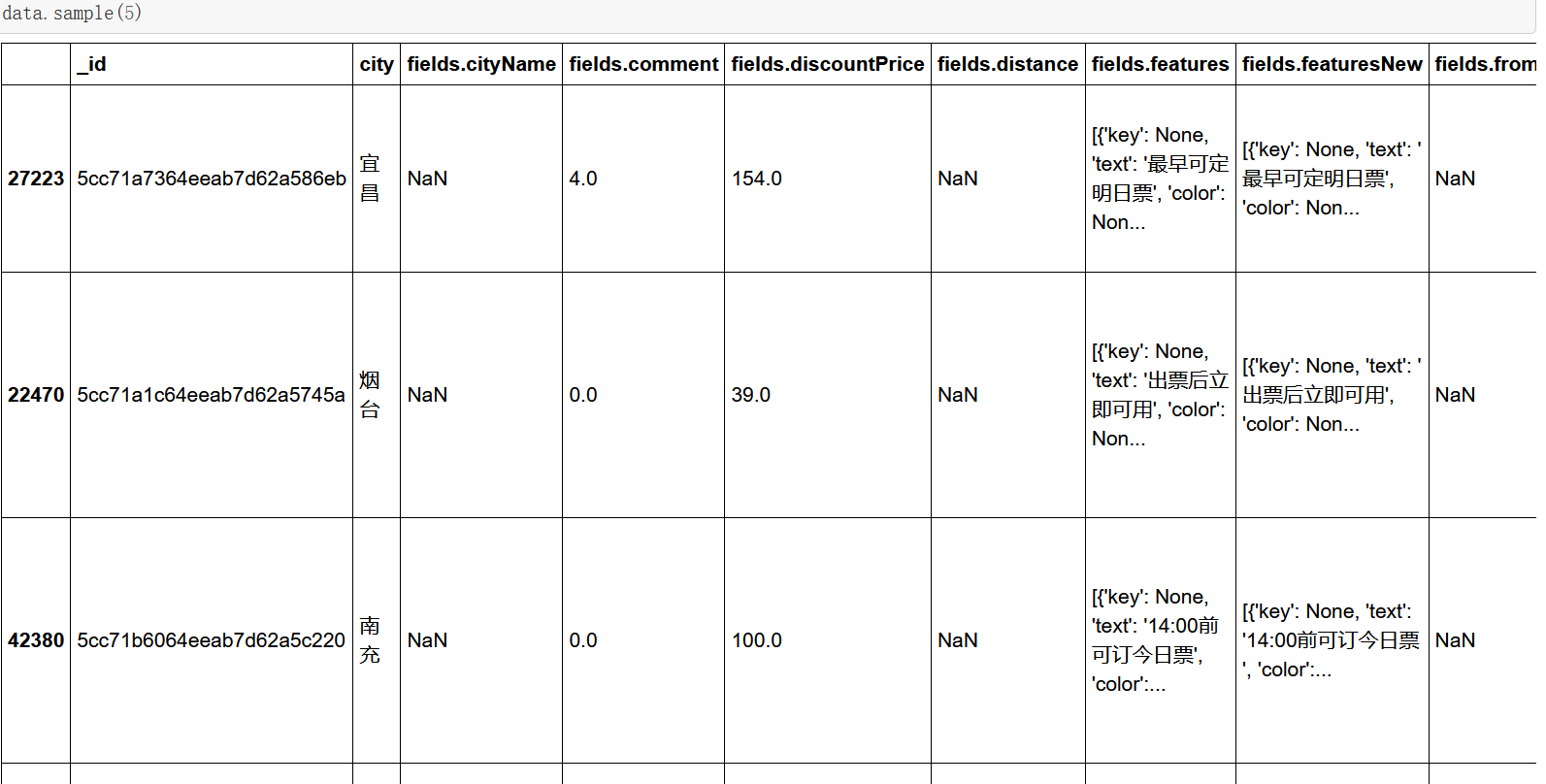


图3.1导入必要的包

3.2读取数据：



图3.2.1读取数据



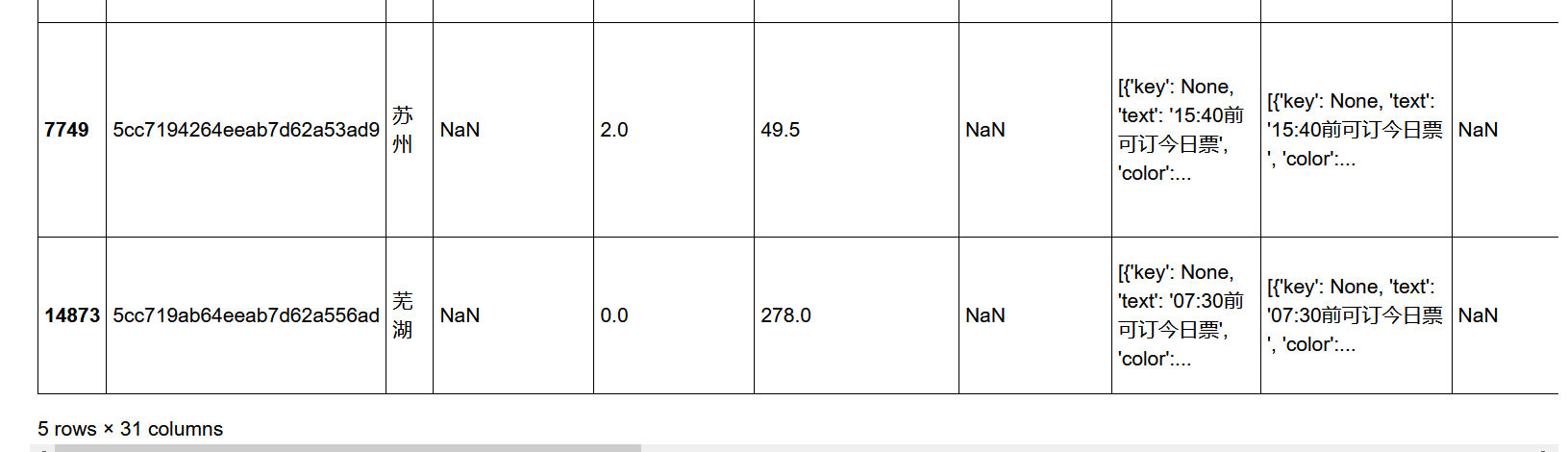
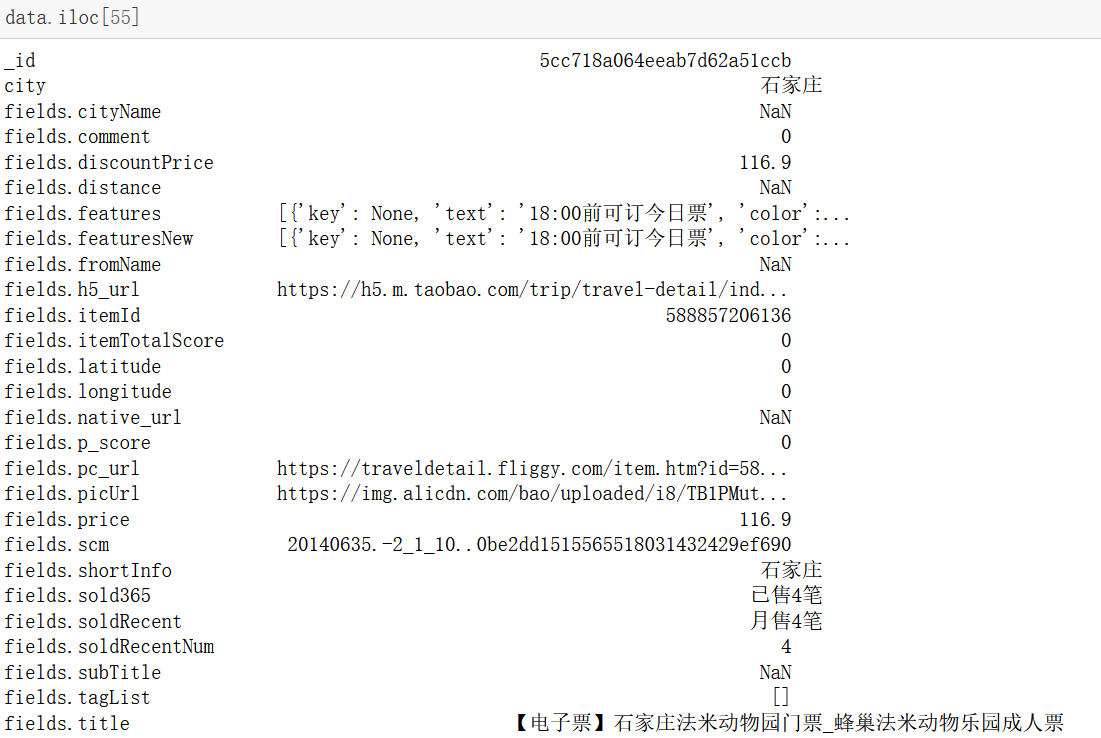


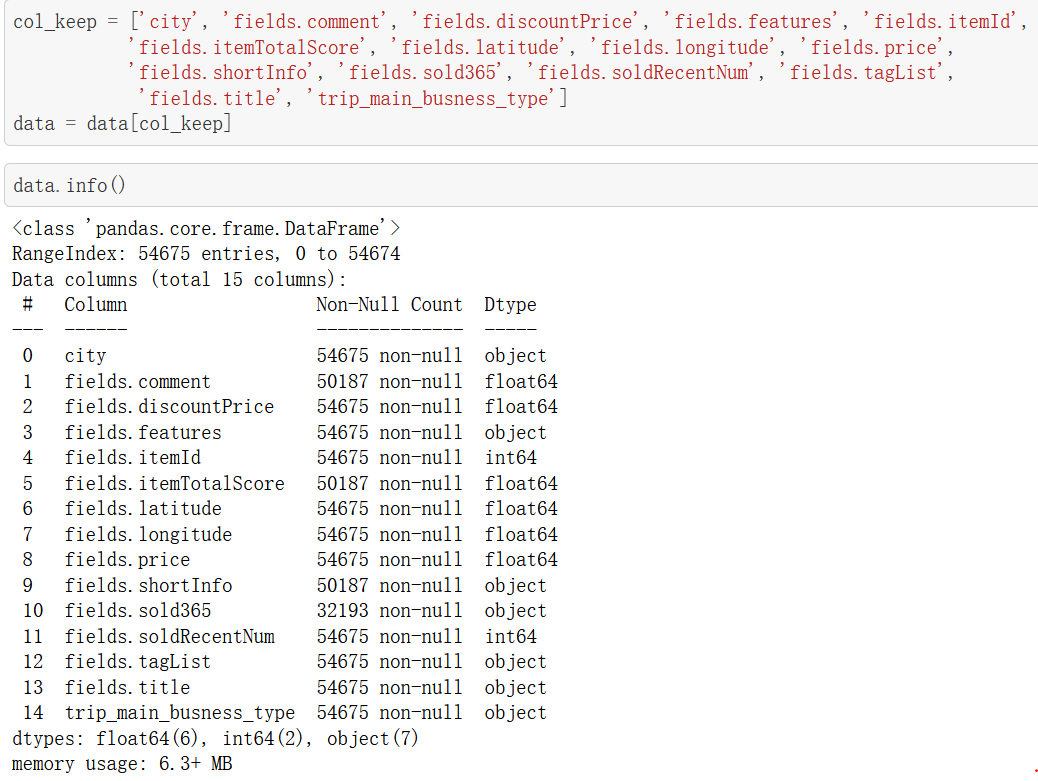
图3.2.2随机抽取5行数据

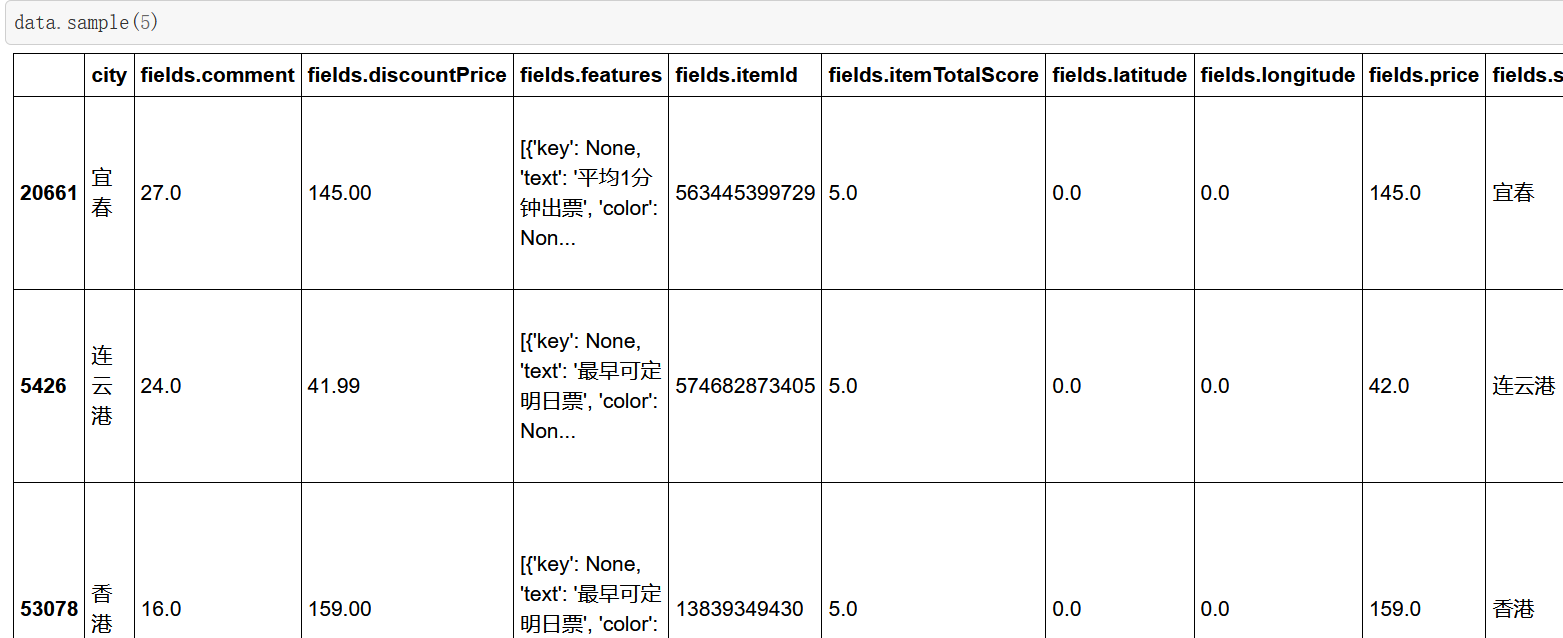
3.3清洗数据

3.3.1 把需要的字段挑出来







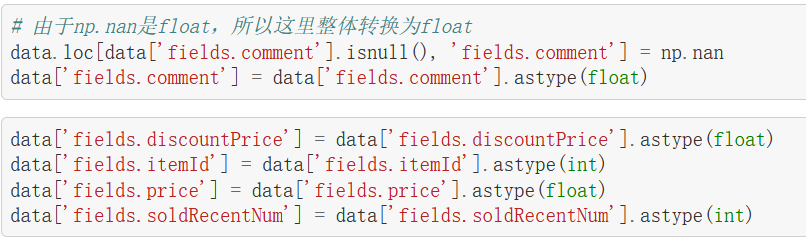




3.3.2 去重

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

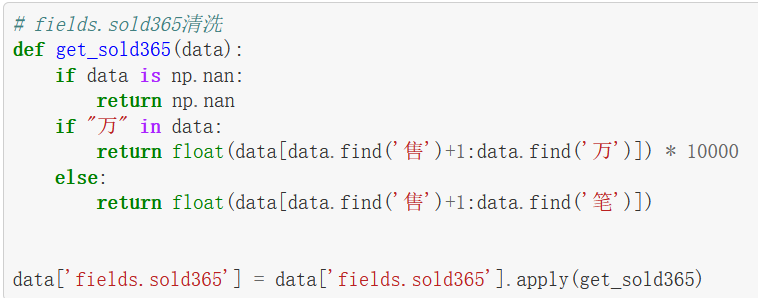
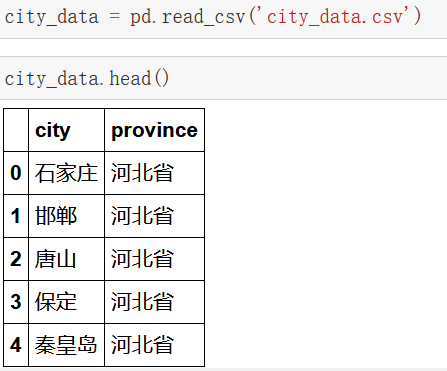


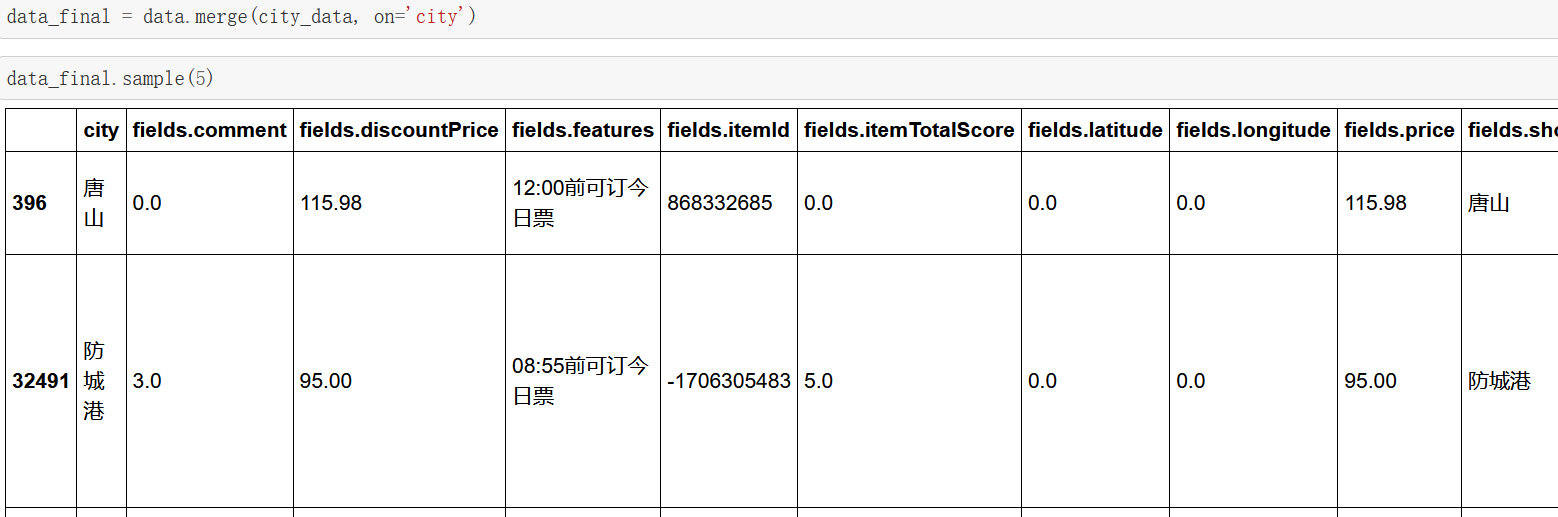
图3.3.4.2fileds.sold365清洗



图3.3.4.3fields.tagList清洗

3.5 跟省份数据合并在一起







四．数据分析

4.1问题引入

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

4.2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：



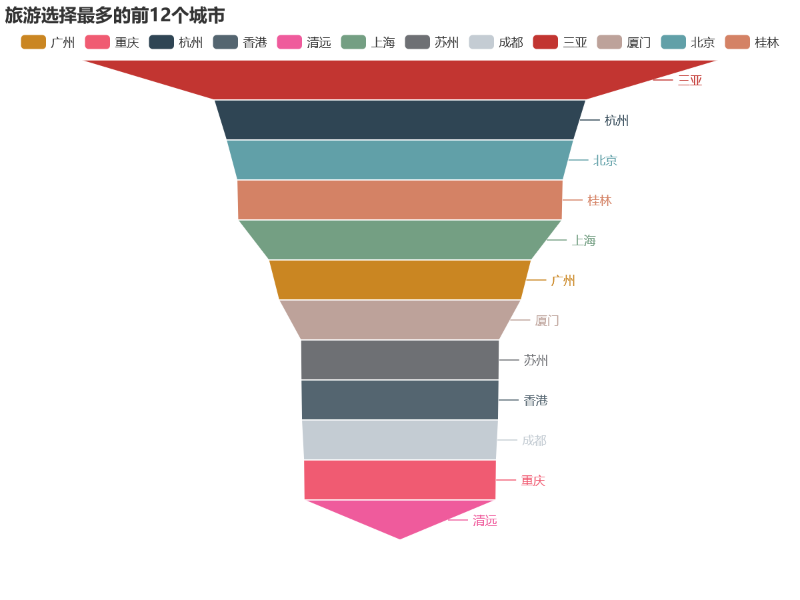


图4.2.1旅游选择最多的前12个城市

使用地图展示各省的旅游选择情况：



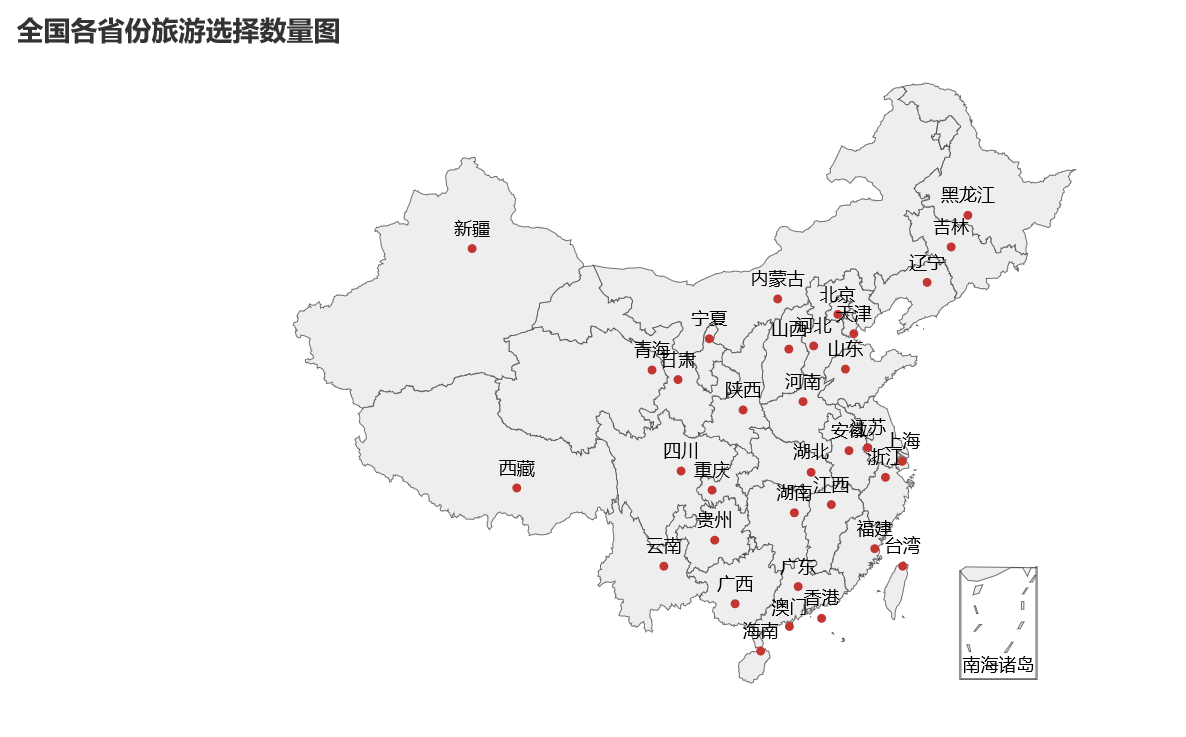


图4.2.2 全国各省分旅游选择数量图

4.3哪些城市最受游客青睐？

先分析一下售票情况



结果如下

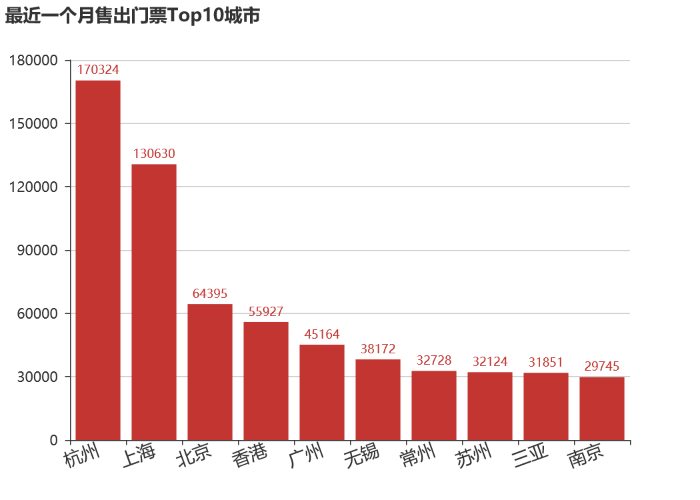
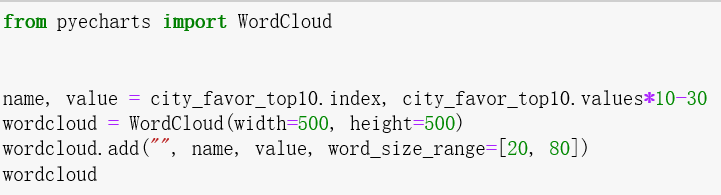


图4.3.1最近一个月出售门票Top10城市

接下来是评分情况



结果如下：分数越高，字体就会越大。

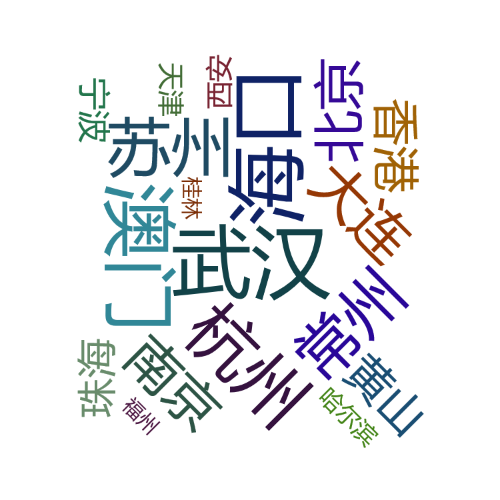
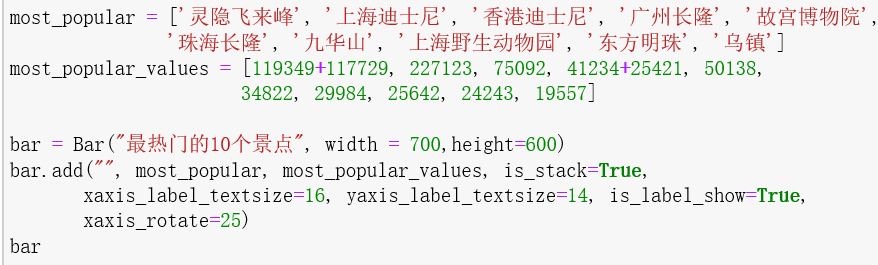


图4.3.2评分情况

4.4最热门的景点



结果如下：

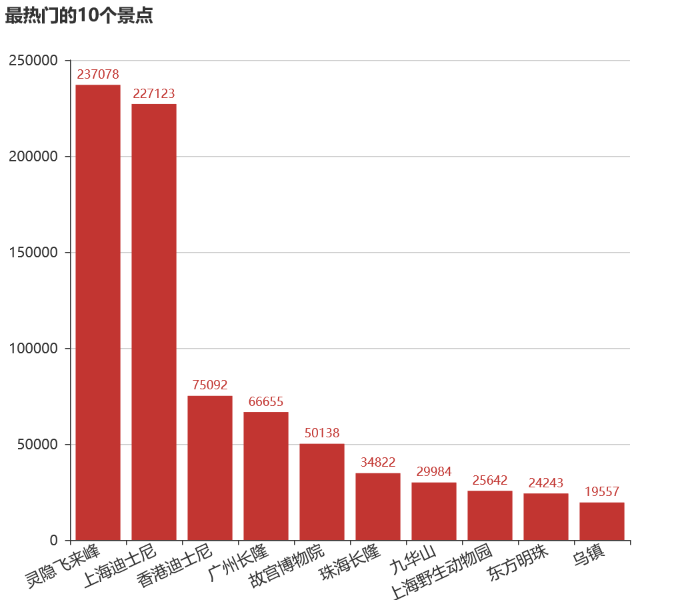


图4.4最热门的10个景点

4.5各省份的好评热门景点



结果如下：



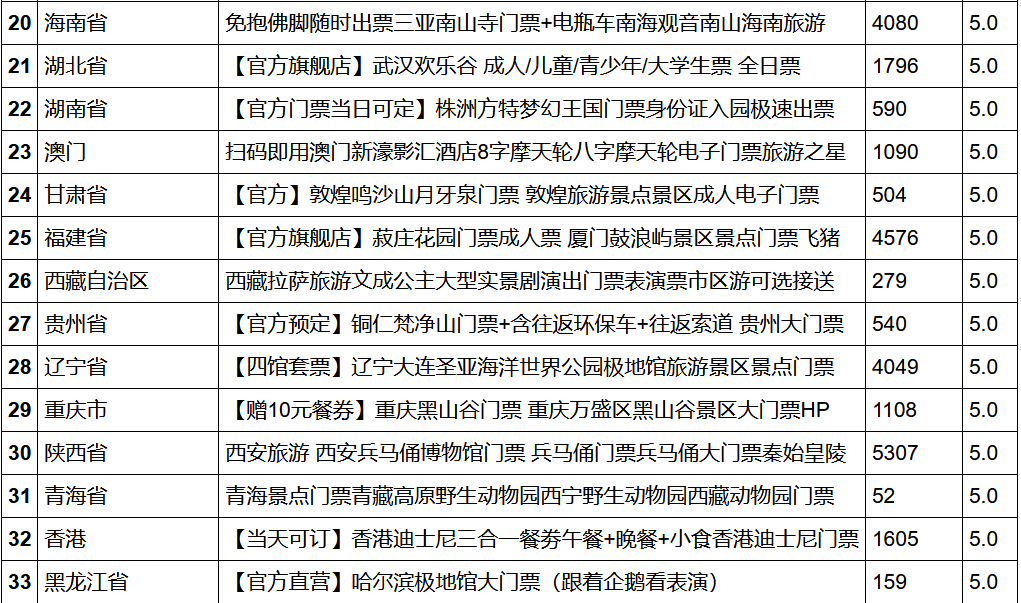
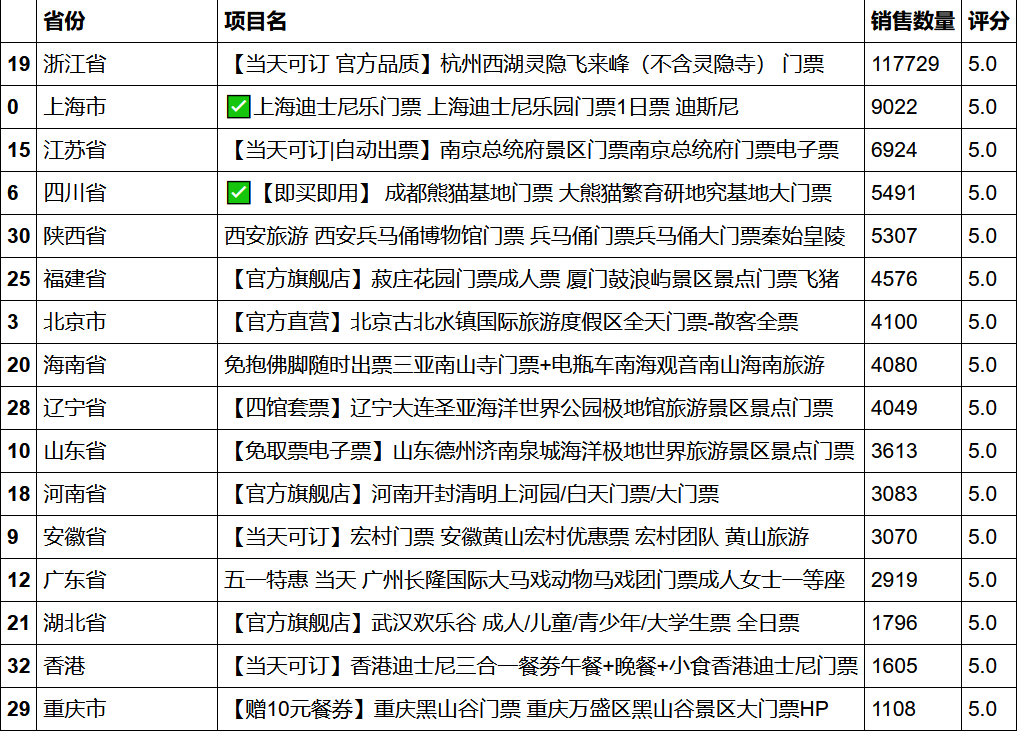


图4.5.1各省份的好评热门景点



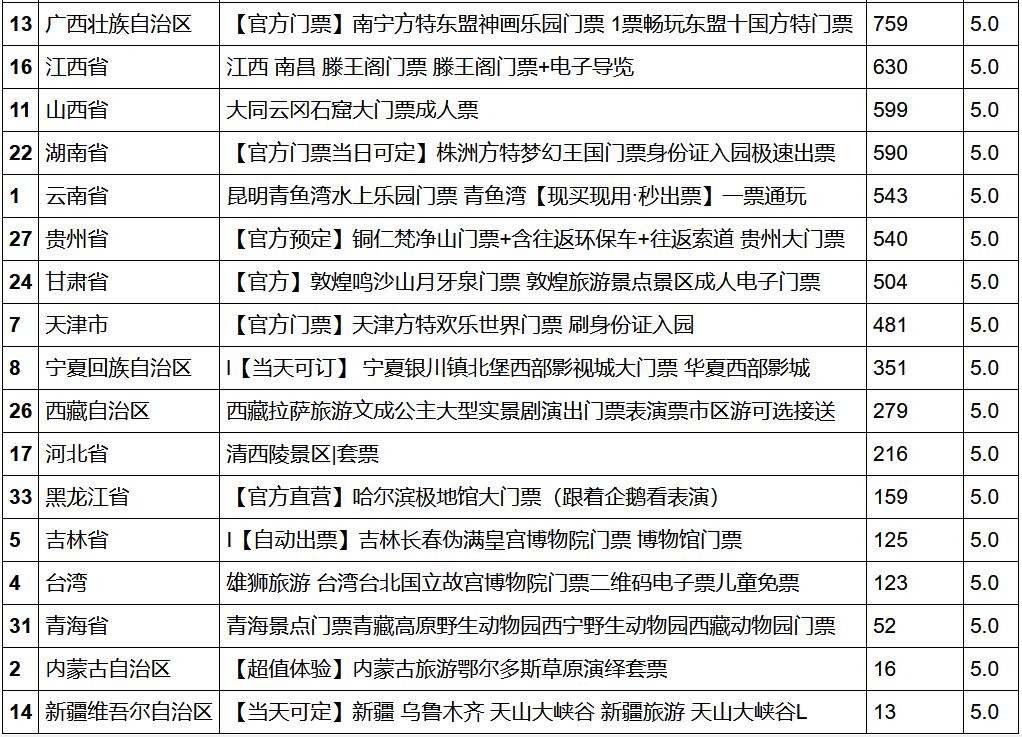


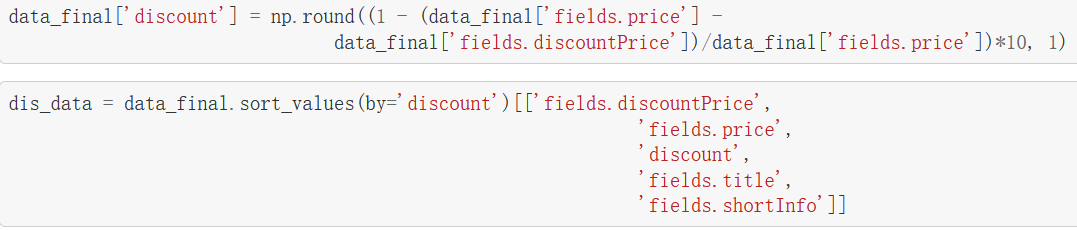
图4.5.2销售数量排名

4.6热门但是评分不太高的景点



图4.6评分不太高的景点

4.7打折力度最大的景点



结果如下：



图4.7打折力度最大的景点

4.8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。

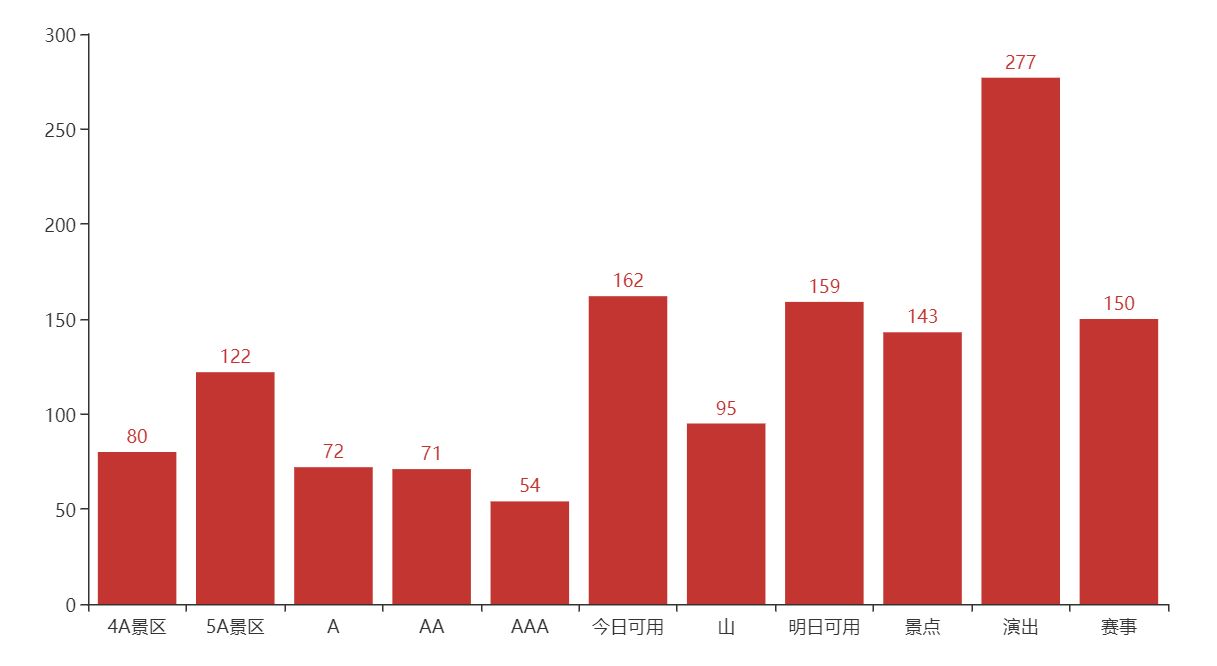


图4.8各级别景区的门票价格

五．总结

在本次Python课程设计中，我学习到了很多对于我来说新的知识。在老师和同学的帮助下，完成了对节假日长假景点人满为患的数据分析。由于本次课程设计是在家中完成，而我因为其他原因，导致本次课程设计能够供我使用的时间并不多。所以只有课设中的一部分由我完成，其他部分由另外两个组员完成。

在本次课设中，我主要负责项目托管和数据分析部分。项目托管平台GitHub也是我第一次接触，根据老师所发的GitHub使用说明和我在网络上查找到的资料，我渐渐对于GitHub熟悉了起来，使用起来也并没有那么困难。而在数据分析的时候，基于老师发的资料，我们也逐渐地将代码完整的写出并运行出来。但也并不是没有遇到难题，我们常常能在一个问题上卡住很久，也是通过查阅资料才得以解决。此次课程设计更让我体会到了小组分工的重要性。由于我没有太多时间的原因，只能和组员在晚上交流，时间上不能很好地配合下来，所以依靠了组员的帮助，才慢慢地将自己所做的部分完善到整个代码当中。

在本次课程设计中，通过老师的指导和教育以及同学对我的答疑解惑，让我感受到学好知识真的很重要，有了好的学习才能在今后利用起来。最后要感谢帮助过我的同学们，希望今后大家能够继续互相学习，互相进步。

参考文献

1.微信推文《五一不看人人人人人人，哪儿耍合适？》

2.python基础与大数据应用.丁辉.北京：人民邮电出版社，2020.

3.董付国.Python可以这样学[M].北京:清华大学出版社，2017.

4.刘宇宙.Python 3.5从零开始学[M].北京:清华大学出版社，2017.

**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.利用pyecharts画图有什么好处？  pyecharts是一个全新的可视化绘图工具，具有良好的交互性，精巧的图表设计。  2.GitHub的作用？  GitHub是一个开源的代码托管平台，使用GitHub可以查看别人的项目、可以建立静态网页、可以管理插件、可以在线编译、可以托管代码等等。  3.groupby()函数的作用？  ① 根据某些条件将数据拆分成组。 ② 对每个组独立应用函数。 ③ 将结果合并到一个数据结构中。 | | | |
| **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 | | | |