

学号	10212818124
----	-------------

武汉华夏理工学院

课 程 设 计

课程名称 python 程序设计课程设计

题 目 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 1181
姓 名 李文娟
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组一份、课程设计

报告每人一份、答辩 ppt 每组一份，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 设计目的.....	1
2 开发环境和开发技术.....	1
2.1 开发平台 jupyter.....	1
2.2 项目托管平台 GitHub.....	2
2.3 编程语言 python.....	2
2.4 绘图模块 pyecharts.....	3
3 需求分析.....	3
4 概要设计.....	4
4.1 分析方向与问题分析.....	4
4.1.1 选择最多的城市与省份.....	4
4.1.2 热门城市与热门景点.....	5
4.1.3 各省份的好评热门景点.....	6
4.1.4 热门但是评分不高的景点.....	7
4.1.5 优惠力度最大的景点.....	7
4.1.6 各级别景区的门票价格.....	8
4.2 总体分析.....	9
5 问题与解决方案.....	9
6 总结.....	10
参考文献.....	11
课程设计成绩评定表.....	12

1 设计目的

近几年来，随着社会的发展和进步，人们的生活也得到了很大的改善，闲暇之余就会有很多人会选择外出放松一下。不过平日里大家都在为了生计而忙碌不已，只有在节假日的时候才有时间和朋友一起出去玩，所以每到节假日的时候我国的众多景点就会出现这样的一幕，没有放假的时候景区几乎杳无人烟，但是一到假期的时候无人问津的地方都挤满了人。

中国游客在过去 10 年如何发生变化——从旗子跟随者到自由自在的行者中国的出境游过去 10 年发生巨变，2010 年中央王国出境游仅 5700 万人次，今年将达 1.8 亿。10 年增加的数字比日本总人口还多。但正如中国游客当初很快发现的那样，一个庞大的游客群并未受到目的地热烈欢迎——尤其当他们齐刷刷从一长串客车上下来时。

对于这样一种普遍存在的社会现象，我们对其原因及其客观规律产生了浓厚的兴趣。为了观察这一现象，我们借助此次课程设计的机会，在两周的时间内，对节假日旅游景点人满为患这一问题做了数据分析。

2 开发环境和开发技术

硬件平台：CPU：Intel(R) Core(TM)，最低要求：64 位 7 代 i3 处理器及以上，内存：16G，最低要求：4G，硬盘空间：SSD-120G，最低要求：机械硬盘-20G

软件平台：jupyter notebook 开发平台、Github 托管平台、Google 浏览器

开发环境：jupyter notebook 开发平台，python3 编程语言

开发技术：pyecharts、pandas、numpy 模块，python3 编程语言

2.1 开发平台 jupyter

Jupyter Notebook（此前被称为 IPython notebook）是一个交互式笔记本，支持运行 40 多种编程语言。

Jupyter Notebook 的本质是一个 Web 应用程序，便于创建和共享文学化程序文档，支持实时代码，数学方程，可视化和 markdown。用途包括：数据清理和转换，数值模拟，统计建模，机器学习等等

用户可以通过电子邮件，Dropbox，GitHub 和 Jupyter Notebook Viewer，将 Jupyter Notebook 分享给其他人。

在 Jupyter Notebook 中，代码可以实时的生成图像，视频，LaTeX 和 JavaScript。

2.2 项目托管平台 GitHub

github 作为开源代码库以及版本控制系统，Github 拥有超过 900 万开发者用户。随着越来越多的应用程序转移到了云上，Github 已经成为了管理软件开发以及发现已有代码的首选方法。

如前所述，作为一个分布式的版本控制系统，在 Git 中并不存在主库这样的概念，每一份复制出的库都可以独立使用，任何两个库之间的不一致之处都可以进行合并。

GitHub 可以托管各种 git 库，并提供一个 web 界面，但它与外国的 SourceForge、Google Code 或中国的 coding 的服务不同，GitHub 的独特卖点在于从另外一个项目进行分支的简易性。为一个项目贡献代码非常简单：首先点击项目站点的“fork”的按钮，然后将代码检出并将修改加入到刚才分出的代码库中，最后通过内建的“pull request”机制向项目负责人申请代码合并。已经有人将 GitHub 称为代码玩家的 MySpace。

在 GitHub 进行分支就像在 Myspace（或 Facebook...）进行交友一样，在社会关系图的节点中不断的连线。

GitHub 项目本身自然而然的也在 GitHub 上进行托管，只不过在一个私有的，公共视图不可见的库中。开源项目可以免费托管，但私有库则并不如此。Chris Wanstrath，GitHub 的开发者之一，肯定了通过付费的私有库来在财务上支持免费库的托管这一计划。

通过与客户的接洽，开发 FamSpam，甚至是开发 GitHub 本身，GitHub 的私有库已经被证明了物有所值。任何希望节省时间并希望和团队其它成员一样远离页面频繁转换之苦的人士都会从 GitHub 中获得他们真正想要的价值。

在 GitHub，用户可以十分轻易地找到海量的开源代码。

2.3 编程语言 python

Python 是完全面向对象的语言。函数、模块、数字、字符串都是对象。并且完全支持继承、重载、派生、多继承，有益于增强源代码的复用性。Python 支持重载运算符和动态类型。相对于 Lisp 这种传统的函数式编程语言，Python 对函数式设计只提供了有限的支持。有两个标准库(functools, itertools)提供了 Haskell 和 Standard ML 中久经考验的函数式程序设计工具。

虽然 Python 可能被粗略地分类为“脚本语言”（script language），但实际上一些大规模软件开发计划例如 Zope、Mnet 及 BitTorrent，Google 也广泛地使用它。Python 的支持者较喜欢称它为一种高级动态编程语言，原因是“脚本语言”泛指仅作简单程序设计任务的语言，如 shellscript、VBScript 等只能处理简单任务的编程语言，并不能与 Python 相提并论。

Python 开发人员尽量避开不成熟或者不重要的优化。一些针对非重要部位的加快运行

速度的补丁通常不会被合并到 Python 内。所以很多人认为 Python 很慢。不过，根据二八定律，大多数程序对速度要求不高。在某些对运行速度要求很高的情况，Python 设计师倾向于使用 JIT 技术，或者用使用 C/C++ 语言改写这部分程序。可用的 JIT 技术是 PyPy。

Python 本身被设计为可扩充的。并非所有的特性和功能都集成到语言核心。Python 提供了丰富的 API 和工具，以便程序员能够轻松地使用 C 语言、C++、Cython 来编写扩充模块。Python 编译器本身也可以被集成到其它需要脚本语言的程序内。因此，很多人还把 Python 作为一种“胶水语言”（glue language）使用。使用 Python 将其他语言编写的程序进行集成和封装。在 Google 内部的很多项目，例如 Google Engine 使用 C++ 编写性能要求极高的部分，然后用 Python 或 Java/Go 调用相应的模块。《Python 技术手册》的作者马特利（Alex Martelli）说：“这很难讲，不过，2004 年，Python 已在 Google 内部使用，Google 招募许多 Python 高手，但在这之前就已决定使用 Python，他们的目的是 Python where we can, C++ where we must，在操控硬件的场合使用 C++，在快速开发时候使用 Python。”

2.4 绘图模块 pyecharts

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。

其特性为：简洁的 API 设计，使用如丝滑般流畅，支持链式调用；囊括了 30+ 种常见图表，应有尽有；支持主流 Notebook 环境，Jupyter Notebook 和 JupyterLab；可轻松集成至 Flask, Django 等主流 Web 框架；高度灵活的配置项，可轻松搭配出精美的图表；详细的文档和示例，帮助开发者更快的上手项目；多达 400+ 地图文件以及原生的百度地图，为地理数据可视化提供强有力的支持。

3 需求分析

本课程设计的主题是对 2019 年各个城市/省份的五一节日长假人满为患的数据进行分析。在具体分析过程中，将从热门景点、打折力度最大的景点、最受青睐的城市、各级别景区的门票价格等方面进行，具体操作时，会将整理出来的数据，运用 pyecharts 模块进行编程绘图更直观的描述出其变化。从对这数据进行重洗、清洁，删除异常点，对整理好的数据针对各问题进行数据分析并绘图几个步骤实施。总的来说，执行过程可分为三大部分：

首先，清洗整理获得的数据，根据问题的要求对问题进行分析，找出相应的解决方法。

其次，根据对问题的分析，收集相关的数据并进行处理。

最后，对各问题数据分析的结果进行总结归纳分析。

4 概要设计

此次课程设计形式为小组设计，每小组由三人组成。本小组由胡佳（组长），李文娟，薛文三人组成。在完成课设的过程中，分工如下：薛文主要负责代码前期的数据清洗与 ppt 前期制作；胡佳主要负责代码中期的问题分析与 Github 托管平台的资源管理，我负责后期的将数据分析的结果进行汇总处理，撰写 ppt 及答辩。

在课程设计过程中，我主要负责后期工作，至此，我将着重描写本人负责的对问题分析的汇总及其产生原因的分析等过程。其分析过程按具体分析的方向依次展现。

4.1 分析方向与问题分析

此次课程设计中，我们小组主要分析方向可分为选择最多的城市与省份、热门城市与热门景点、各省份的好评热门景点、热门但是评分不高的景点、优惠力度最大的景点、各级别景区的门票价格六个问题方向。

在环境搭建，数据清洗（此处省略该步骤）完成后，我们可将得到的数据进行具体的问题分析并对产生的图表进行数据汇总与透析。分析过程如下。

4.1.1 选择最多的城市与省份

通过数据分析，我们筛选出了旅游选择最多的前 10 个城市，其中三亚遥遥领先于其他城市，其次是杭州、北京、等自然风光，人文风景著名的地点。在前十位排名中，成都居其末位。

通过浏览资料，我们可以发现三亚在地理位置，气候风光等方面都有着得天独厚的优势，其景区整体销售量也处于较高的位置。由此为它旅游业的发展奠定了坚实基础。其数据分析图如下所示：

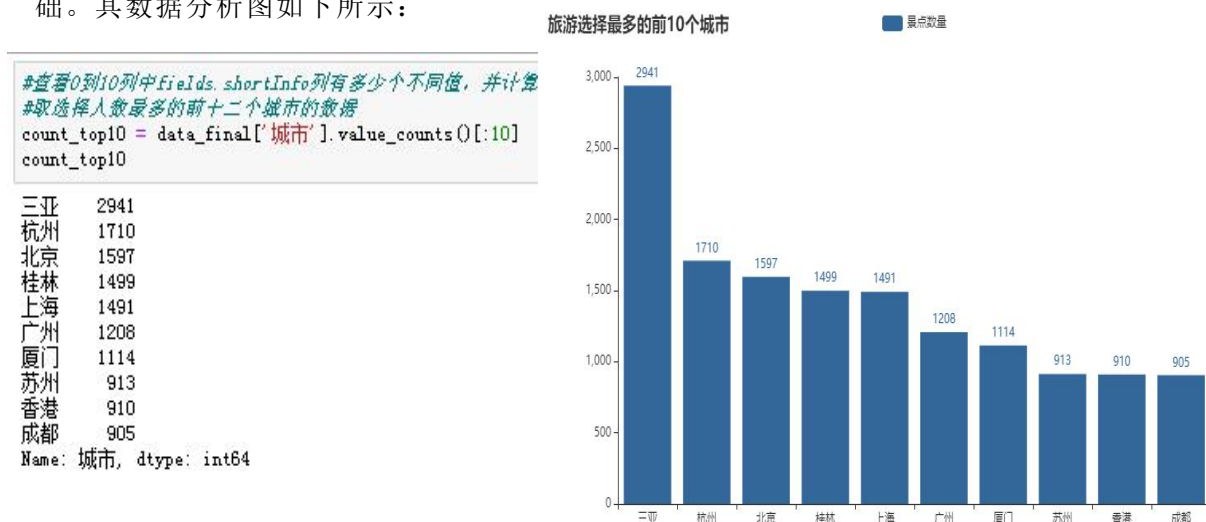


图 4.1 旅游选择最多的前 10 个城市

以省份为单位，我们也计算了每个省份在飞猪上能搜索到的的景点门票项数（分析结果见图 4.2），发现旅游选择最多的省份主要集中在沿海，如广东，其地理位置与香港，澳门接壤，与海南隔海相望。其独特的地理位置为其旅游行业的发展提供了良好的条件。

Out[44]: 旅游选择最多的省份排名

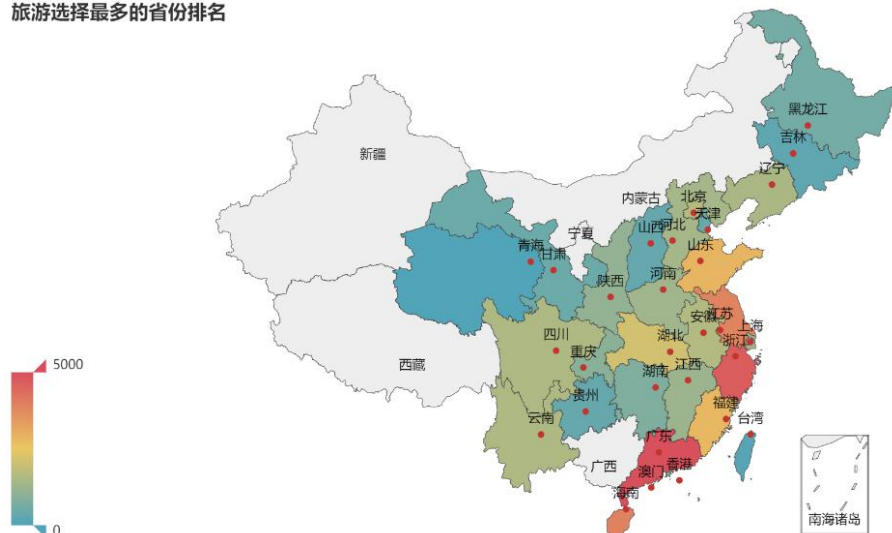


图 4.2 旅游选择较多的省份排名

4.1.2 热门城市与热门景点

根据数据结果，我们可以看出杭州是月售最多的城市，其次是上海北京等地。

分析原因，我们不难发现，图中所示的热门城市皆为经济发展较为良好的地区，可见其城市的经济发展情况与其旅游业的发展也是密切相关的。

最近一个月售出门票Top10城市词云图

```
#将评分大于0评价数大于30的city_qual中的值对应的数据按shortinfo分组并返回其
city_favor_top10 = data_final[(data_final['评分'] > 0) &
                             (data_final['评论量'] > 30) &
                             (data_final['城市'].isin(city_qual))
                             ].groupby('城市')['评分'].mean().nlargest(20)

city_favor_top10
```

城市	评分
武汉	4.948361
海口	4.947619
澳门	4.944660
杭州	4.940244
苏州	4.937500
常州	4.936275
大连	4.931818



图 4.3 旅游热门城市前十排序

根据数据分析结果，我们不难看出灵隐飞来峰是最多人心仪的目标景点，其次为上海迪士尼度假区与故宫博物院。

浏览相关资料，我们可以发现在热门景点中，普遍存在多样性的游乐项目及较多的互动性强的设施。查阅资料，我们发现尤其上海迪士尼等（具有较多的全国首发游乐项目）景点，由此可以得出结论：景点较强的创新性也是吸引游客的重要因素之一。

```
#根据月销量降序，返回前十二条数据的项目名和月销量
data_final.sort_values(by='月销量', ascending=False)[:12][['项目名', '月销量']]
```

	项目名	月销量
44567	上海迪士尼度假区	227123
8162	灵隐飞来峰	119349
8285	【当天可订 官方品质】杭州西湖灵隐飞来峰（不含灵隐寺）门票	117729
46880	香港迪士尼乐园	75092
42479	故宫博物院	50138
25585	广州长隆野生动物世界	41234
29091	珠海长隆海洋王国	34822
14184	九华山	29984
44571	上海野生动物园	25642
25587	广州长隆国际大马戏	25421

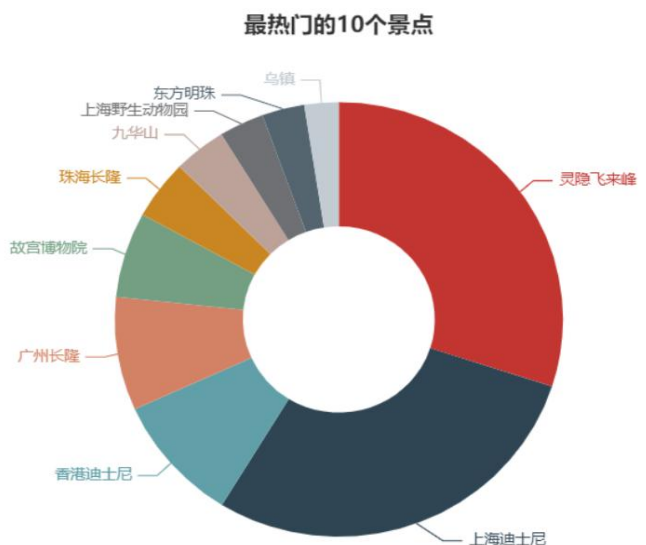


图 4.4 旅游热门景点前十排序

4.1.3 各省份的好评热门景点

我们把这些景点门票数据以省级行政区为单位，筛选出了每个省级行政区最受好评且销量最高的景点。根据分析结果所示，在各省份好评热门景点的数据中，其好评数最多的是杭州西湖灵隐飞来峰，其次是上海迪士尼乐园门票。

其结果显示，在一定程度上怀旧也对游客具有较高的吸引力。也印证了上述游客对创新性较高的景点较为感兴趣这一结论。

```
#根据省份，评分，销量对数据排序，排序结果按省份分组，并将分组好的数据依次传入get_top1后重置标题，销售量和评分的索引：
province_top1 = data_final.sort_values(by=['省份', '评分', '月销量'],
                                       ascending=False).groupby('省份').apply(get_top1)[['项目名', '月销量', '评分']].reset_index()
#删除level列，并使其原数组不变
province_top1.drop(columns='level_1', inplace=True)
#根据“销售数量”列对数据排序
province_top1.sort_values(by='月销量', ascending=False)
```

	省份	项目名	月销量	评分
19	浙江省	【当天可订 官方品质】杭州西湖灵隐飞来峰（不含灵隐寺）门票	117729	5.0
0	上海市	上海迪士尼乐园门票 上海迪士尼乐园门票1日票 迪斯尼	9022	5.0
15	江苏省	【当天可订]自动出票】南京总统府景区门票南京总统府门票电子票	6924	5.0
6	四川省	【即买即用】成都熊猫基地门票 大熊猫繁育研究基地大门票	5491	5.0
30	陕西省	西安旅游 西安兵马俑博物馆门票 兵马俑门票兵马俑大门票秦始皇陵	5307	5.0
25	福建省	【官方旗舰店】鼓浪屿成人票 厦门鼓浪屿景区景点门票飞猪	4576	5.0
3	北京市	【官方直营】北京古北水镇国际旅游度假区全天门票-散客全票	4100	5.0
20	海南省	免抱佛脚随时出票三亚南山寺门票+电瓶车南海观音南山海南旅游	4080	5.0
28	辽宁省	【四馆套票】辽宁大连圣亚海洋世界公园极地馆旅游景区景点门票	4049	5.0
10	山东省	【免取票电子票】山东德州济南泉城海洋极地世界旅游景区景点门票	3613	5.0
18	河南省	【官方旗舰店】河南开封清明上河园/白天门票/大门票	3083	5.0
9	安徽省	【当天可订】宏村门票 安徽黄山宏村优惠票 宏村团队 黄山旅游	3070	5.0
12	广东省	五一特惠 当天 广州长隆国际大马戏动物马戏团门票成人女士一等座	2919	5.0

图 4.5 各省份的热门景点

4.1.4 热门但是评分不高的景点

很多景点中虽然热门但是去过的游客对这些地点的评价却不是很高,此处我们筛选出了销量大于 100 且评分小于 4.5 的地点。在总数据中,存在以下十条数据满足条件。

结合网上资料分析数据,我们可以推断此类景点存在部分单一性,如其中观赏类景点互动性较差,园馆类景点存在时间限制,地理位置等原因,致使其评分略低。

	城市	项目名	月销量	评分
2545	大连	【领券立减】大连二〇三樱花园大门票旅顺203樱花成人票扫码入园	118	4.3
7124	无锡	【当天可定】无锡善卷洞门票善卷洞含游船联票大门票成人电子票	760	4.2
8317	杭州	浙江杭州临安风之谷景区门票 户外昌化风之谷门票 杭州接送跟车	199	4.3
22782	武汉	【汉口粤汉码头登船】武汉两江游览游船B线(不含餐)夜游长江游	109	4.3
26939	深圳	ZY【当天可定即定即用】深圳观澜山水田园娱乐套票A 含景区大门票	155	3.8
29856	佛山	顺德长鹿农庄门票 休博园 大门票+游乐金+动物园 单人票/双人票	1001	4.4
29877	佛山	长鹿农庄门票休博园大门票 套票动物园2大2小(不含矿洞海洋馆)	210	4.3
40353	西安	陕西 西安 西安城墙景区 大门票+电子导览	221	4.4
42576	北京	北京昌平第七届北京农业嘉年华大门票 农业嘉年华门票 成人票	112	4.4
46919	香港	Dm香港迪士尼乐园门票1日迪斯尼香港旅游自由行景点电子票不刷单	261	4.3

图 4.6 热门但是评分不高的景点

4.1.5 优惠力度最大的景点

分析结果显示,在总数据中,存在下图所示等部分景点存在较大的优惠力度。

通过景点门票打折状况分布图可以看出大多数景点的折扣都在 8 折到全价之间,也有少部分景点门票在 3 折左右。门票折扣价最低的是 2.7 折,是位于广东东莞的“南社明清古村落”。优惠力度最大的是珠海的“海泉湾海洋温泉”,打折前为 1980 元,折后为 589 元。

	折扣价	原价	折扣	项目名	城市
28979	13.5	50.00	2.7	ZM【当天可定】广东 东莞南社明清古村落 成人儿童老人观光大门票	东莞
22713	180.0	630.00	2.9	【大学生3人特惠】武汉海昌极地海洋公园门票 海洋世界含萌宠乐园	武汉
29456	598.0	1980.00	3.0	【598元/10次卡】超值钜惠 珠海海泉湾海洋温泉10次卡 可4人使用	珠海
37355	18.0	53.10	3.4	【双人特价 电子票】成都大熊猫基地门票成都大熊猫繁育研究基地	成都
27530	34.5	100.00	3.4	ZY【提前1小时下单】广东 清远田野绿世界 成人儿童老人大门票	清远
30308	36.0	100.00	3.6	ZM【当天可定】广东 阳江闸坡大角湾景区 成人1日景区大门票	阳江
26931	60.0	166.50	3.6	即买即用 索道+缆车】深圳东部华侨城大侠谷门票 大峡谷亲子家庭	深圳
1004	49.9	135.00	3.7	当日可订 河北 秦皇岛 沙雕海洋乐园门票 沙雕大世界大门票	秦皇岛
30549	20.0	52.99	3.8	【即买即用】南宁园博园门票 第十二届园博会展园博园门票	南宁
30541	21.4	55.00	3.9	【即买即用】南宁园博园门票 第十二届园博会展园博园门票	南宁
30426	40.0	100.00	4.0	ZM【当天可定】提前1小时下单 广东茂名浪漫海岸成人儿童情侣门票	茂名
28929	24.0	60.00	4.0	【当天可定 即定即用】广东 东莞梦幻百花洲 成人儿童老人大门票	东莞
42044	138.0	336.00	4.1	银川天山海世界.黄河明珠五一特惠双人票	银川
6922	35.0	83.00	4.2	【当天可订】江苏无锡鼋头渚门票+客渡船门票(含园内公交车)	无锡
313	69.9	168.00	4.2	【官方旗舰店】石家庄天山海世界69.9元五一特惠票	石家庄

图 4.7 优惠力度较大的景点

结合相关资料分析数据可以发现，部分景点由于天气或旅游淡季或一些突发状况等现象，对不同时段的售票价格及优惠力度采取不同形式。此处建议游客提前浏览目标景点及其城市的相关信息，在能确保拥有良好的旅行体验的情况下，选择目标城市。

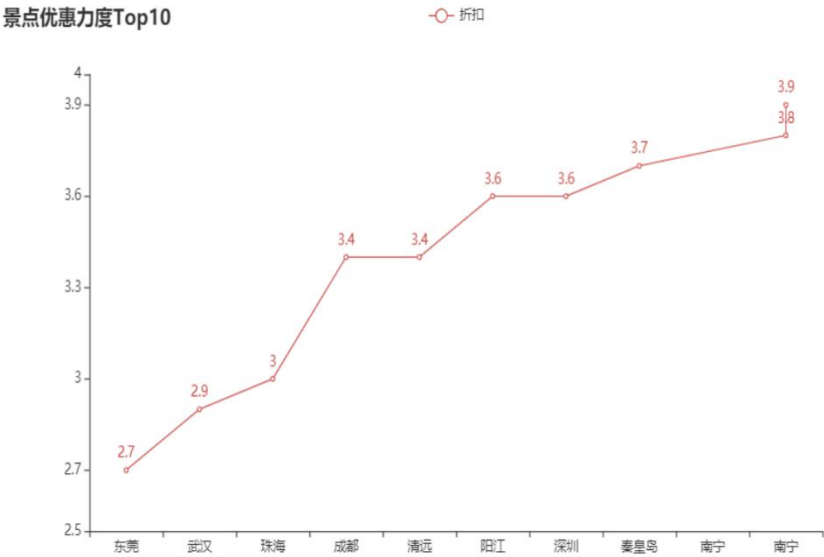


图 4.8 优惠力度较大的景点

4.1.6 各级别景区的门票价格

通过观察数据，我们可以发现，各级别景区票价的整体趋势为：级别越高，景区门票的价格就越高。数据显示：A 级和 AA 级的平均价格差不多，都在 71 元左右。3A 级景区的门票平均价格最低，4A 级景区门票均价在 80 元，5A 级景区门票均价在 122 元，比 4A 的价格高出了 42 元。最高的 5A 级景区门票比最低的 3A 级景区门票高出了整整 68 元。

结合相关资料我们发现，由于历史，人文情怀，景区地理位置，互动性等方面，5A 级景区确实可能拥有较好的旅行体验。但是部分宣传力度较弱的 4A，3A 级景区也存其性价比高于 5A 景区的现象。

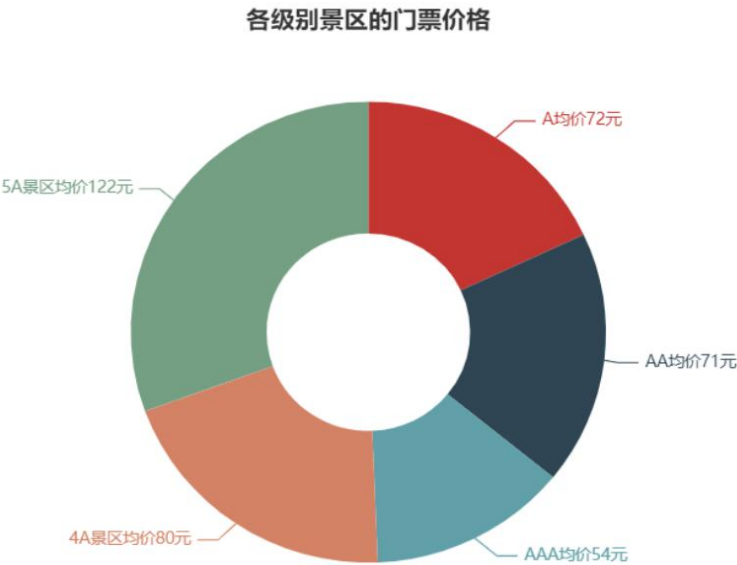


图 4.9 各级别景区的门票价格

4.2 总体分析

根据上述数据分析，我们将主要分析方向分成了三个类别，一是基于选择，满足想要短时间内游玩多地的旅客群体，二是基于热点，面向喜欢热门城市热门景点的旅客，三是基于价格，面向对景点消费价格及优惠力度方面有一定要求的旅客。

对于第一类，基于选择考虑的游客，建议参考图 4.1 与图 4.2。如上图所示，三亚为选择最多的城市，杭州其次。在省份中，澳门，香港，广东等沿海城市选择最多。

对于第二类，基于热点考虑的游客，建议参考图 4.3，图 4.4，图 4.5 及图 4.6。根据上述数据分析结果图所示，在总数据中，武汉，澳门，海口等城市较为热门，灵隐飞来峰，上海迪士尼，香港迪士尼等景区较为热门。查阅相关资料，我们发现此类地区消费水平较高。建议游客可根据其景区整体销售量，参考评论及其景区互动性等特点综合考虑以决定具体目标景点。

对于第三类，基于价格考虑的游客。由图 4.7，图 4.8 及图 4.9 可得，于总数据中 3A 景区平均售价最为划算，东莞，武汉等地区景点优惠力度最大。确定目标景区时，游客可参考具体景点及整体售价特点决定目标景点。

另外，游客对景点的评判大多依赖于游客自己的主观感受，不同游客的关注点也大多有所不同。所以，以上数据，仅供参考。

5 问题与解决方案

在完成课程设计的过程中，我们小组遇到了不少问题，大多数的问题都能及时的得到解决，如软件使用中的问题，模块导入中的问题。但是也有一些问题在问题分析的过程中给我们留下了不可磨灭的印象，如 pyecharts 版本的问题，困扰了我们将近两天的时间。

该问题具体表现在数据分析的绘图中。刚开始接手数据时，使用 pyecharts 绘图一直出错。观察后发现，所有的绘图语句都出现了同一种错误提示。截图如下：

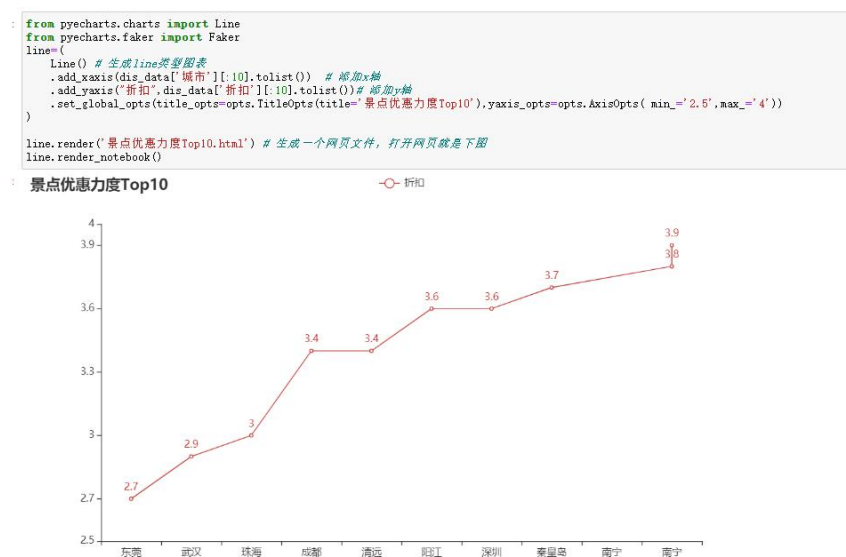
```
TypeError                                Traceback (most recent call last)
<ipython-input-3-aa40cc9f3b9a> in <module>
      1 from pyecharts.charts import Map
----> 2 map = Map("全国各省份旅游选择数量图", width=600, height=600)
      3 map.add(
      4     ..
      5     province,

TypeError: __init__() got an unexpected keyword argument 'width'
```

后来我们不断查阅相关资料，终于知道了问题出在 pyecharts 版本的兼容上。根据网友的介绍，我们查出，目前使用的是 v0.5 版，需要更改到 v1 以上的版本，才能在 jupyter notebook 中使用 pyecharts 的绘图功能。于是，我们选择安装了 1.9.0 版。截图如下：

```
C:\Users\LOVE>pip install pyecharts==1.9.0 -i http://pypi.douban.com/simple --trusted-host pypi.douban.com
Looking in indexes: http://pypi.douban.com/simple
Collecting pyecharts==1.9.0
  Downloading http://pypi.douban.com/packages/d2/1a/a094c0d11c3c87cf407cbd9dc069882e9bba08374df1ad81dfb39dela8d3/pyecharts-1.9.0-py3-none-any.whl (143kB)
    100% |#####| 143kB 885kB/s
Requirement already satisfied: simplejson in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from pyecharts==1.9.0)
Requirement already satisfied: prettytable in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from pyecharts==1.9.0)
Requirement already satisfied: jinja2 in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from pyecharts==1.9.0)
Requirement already satisfied: MarkupSafe in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from jinja2->pyecharts==1.9.0)
Requirement already satisfied: setuptools in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from prettytable->pyecharts==1.9.0)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from jinja2->pyecharts==1.9.0)
Installing collected packages: pyecharts
  Found existing installation: pyecharts 0.5.10
  Uninstalling pyecharts-0.5.10:
    Successfully uninstalled pyecharts-0.5.10
  Successfully installed pyecharts-1.9.0
You are using pip version 10.0.1, however version 20.3.3 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

安装新的版本以后，我们迫不及待的进入 jupyter notebook 中进行试验，发现图片果然可以出来了。截图如下：



不过，随之而来也有一些语句的小错误。由于这些错误在较短的时间内都得到了解决，此处就不一一举例了。



6 总结

本次课程设计使我对 python 语言的使用与项目开发过程有了更深刻的认识，对 python 的编程环境更加熟练。在此次课设中，我们用到了很多上课与实验时学到的知识，可以说是将理论与实践相结合，锻炼了自己的动手能力，体会到了其中的奥秘和乐趣，获益匪浅。

在这次课程设计中，我们小组完成的题目是对节假日旅游景点人满为患的数据分析与实践。初时听到这个题目时，我觉得很有压力，甚至有些不知所措。幸而在课堂中，老师带着我们一起缕清脉络，让我们对它的整体框架有了一定的思路。在自己动手进行设计时，我一边学习老师的语句与课本上的实践步骤，一边翻阅相关网站了解相关的实践记录。此次课程设计使我在复习课堂知识的同时，又增加了不少实践经验，收获了很多新的知识。这次课程设计让我意识到，平时的课堂上的学习是比较有限的，我们就应该放开思维，自己多想多查多问多动手。这样才能使我们的学习更加全面化，使我们能够更加熟练的掌握各项实用语言及各种开发工具。

最后，在此次实验中，我认识到，完成一项任务，最重要的一点是要不断的沟通和学习。整个课程设计的过程中，我们小组一直保持着有效的联系，每天下午一点，我们小组长胡佳就发起腾讯会议或 QQ 电话，保证每个小组成员时刻了解小组进度并明确各自的分工。在这里也非常感谢她的督促，使我们在家也能保持与在学校一致学习的心态。希望下次还能有幸参与此类程序设计的学习。

参考文献

- [1] 刘宇宙.python3.5 从零开始学【M】. 北京：清华大学出版社，2017.
- [2] 丁辉.Python 基础与大数据应用. 人民邮电出版社
- [3] 刘宇宙.python3.5 从零开始学[M]. 北京：清华大学出版社，2017.
- [4]  <https://mp.weixin.qq.com/s/iuCNreCuKzrggdXtvurpkQ>
- [5]  <https://www.baidu.com>

课程设计成绩评定表

课程设计题目	节假日长假景点人满为患的数据分析	
<p>课程设计学生答辩或质疑记录：</p> <p>1. 简述 pyecharts 的用途。</p> <p>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，主要用于数据可视化。pyecharts 是一个用于生成 Echarts 图表的类库。实际上就是 Echarts 与 Python 的对接。使用 pyecharts 可以生成独立的网页，也可以在 flask , Django 中集成使用。</p> <p>2. 简述数据分析的目的。</p> <p>答：数据分析是指用适当的统计分析方法对收集来的大量数据进行分析，提取有用信息和形成结论而对数据加以详细研究和概况总结的过程，是为了寻求问题的答案而实施的有计划、有步骤的行为。</p> <p>3. 数据清洗的目的是什么。</p> <p>答：数据清洗的目的有两个，第一是通过清洗让数据可用。第二是让数据变的更适合进行后续的分析工作。换句话说就是有”脏”数据要洗，干净的数据也要洗。</p>		
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	
<p>最终评定等级为：</p> <p style="text-align: right;">指导老师签字：_____</p> <p style="text-align: right;">2021 年 1 月 15 日</p>		