

学号	10212818126
----	-------------

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 胡 佳
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 项目编程技术.....	1
1.1 项目托管平台——GitHub.....	1
1.2 项目开发平台——Jupyter Notebook.....	1
1.3 项目编程语言——Python.....	2
2 程序设计.....	3
2.1 数据分析库.....	3
2.1.1 数据分析库导入.....	3
2.1.2 数据分析库注解.....	3
2.2 导入数据表.....	4
2.3 数据清洗.....	4
2.3.1 数据提取.....	4
2.3.2 数据去重.....	5
2.3.3 数据类型转换.....	5
2.3.4 各字段清洗.....	5
2.4 数据分析.....	6
2.4.1 对城市、省份可选择景点进行统计.....	6
2.4.2 对热门城市、景点的情况统计.....	7
2.4.3 对景点优惠力度分析.....	8
3 问题与解决方案.....	9
3.1 Github 中对文件的处理.....	9
3.2 Pyecharts 版本.....	10
4 总结体会.....	11
5 小组分工明细表.....	12

引言

近几年来，由于经济的迅猛发展及人类不断增加的物质需求，各行各业竞争态势愈演愈烈，人们的生存能力及经济水平日益提升，但生活幸福感指数却逐步下降，生活节奏快、长周期高压下的连续负荷工作导致越来越多的人身体及家庭氛围处于亚健康状态，急需一种有效的生活调节方式以提升生活质量，而旅游不失为最优的释压途径。旅游可以陶冶情操，释放压力、开阔眼界、认知世界，在轻松的环境下陪伴家人、教育孩子。

因此，节假日以家庭为单位的团体旅游成为了大多数人的安排节假日行程的首选。节假日的集中出行极大地带动了旅游业的快速发展，加快了经济发展，为旅游行业带来了巨大的商机，但与此同时，人口的短暂性高强度集中，造成了旅游城市的交通拥堵，景点安全事故频发，景点卫生环境堪忧，部分自然生态环境景点的生态环境可能遭受人为破坏，而景区的安全管理、卫生管理及生态保护管理力量应付常态客流游刃有余，一旦遇到节假日，景区人流急剧增加甚至出现景区内通行拥堵，管理力量就显得力不从心，突发情况下更是措手不及。

为此，如何有效控制节假日旅游高峰期人流量、配合相应人流的管理队伍，是我们此次课设的重点分析任务。根据分析结果制定管理规划、匹配相应的管理制度及任务调度，可以较好的缓解节假日景点的管理压力

1 项目编程技术

1.1 项目托管平台——GitHub

GitHub 是一个面向开源及私有软件项目的托管平台，因为只支持 Git 作为唯一的版本库格式进行托管，故名 GitHub。GitHub 于 2008 年 4 月 10 日正式上线，除了 Git 代码仓库托管及基本的 Web 管理界面以外，还提供了订阅、讨论组、文本渲染、在线文件编辑器、协作图谱（报表）、代码片段分享（Gist）等功能。目前，其注册用户已经超过 350 万，托管版本数量也是非常之多，其中不乏知名开源项目 Ruby on Rails、jQuery、python 等。

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。

1.2 项目开发平台——Jupyter Notebook

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。

1.3 项目编程语言——Python

Python 的设计哲学是“优雅”、“明确”、“简单”。因此，Perl 语言中“总是有多种方法来做同一件事”的理念在 Python 开发者中通常是难以忍受的。Python 开发者的哲学是“用一种方法，最好是只有一种方法来做一件事”。在设计 Python 语言时，如果面临多种选择，Python 开发者一般会拒绝花俏的语法，而选择明确的没有或者很少有歧义的语法。由于这种设计观念的差异，Python 源代码通常被认为比 Perl 具备更好的可读性，并且能够支撑大规模的软件开发。这些准则被称为 Python 格言。在 Python 解释器内运行 `import this` 可以获得完整的列表。

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 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 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 程序设计

2.1 数据分析库

2.1.1 数据分析库导入

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib as mpl
import seaborn as sns
%matplotlib inline

#设置中文标签的显示
plt.rcParams['font.sans-serif'] = ['SimHei'] #设置显示中文字体
```

2.1.2 数据分析库注解

pandas 是 python 的数据分析处理库 NumPy 是 Python 语言的一个扩充程序库。支持高级大量的维度数组与矩阵运算，此外也针对数组运算提供大量的数学函数库。

Numpy 内部解除了 CPython 的 GIL（全局解释器锁），运行效率极好，是大量机器学习框架的基础库！

使用%matplotlib 命令可以将 matplotlib 的图表直接嵌入到 Notebook 之中，或者使用指定的界面库显示图表，它有一个参数指定 matplotlib 图表的显示方式。inline 表示将图表嵌入到 Notebook 中。

%matplotlib inline Seaborn 是一种基于 matplotlib 的图形可视化 python library。它提供了一种高度交互式界面，便于用户能够做出各种有吸引力的统计图表。

Seaborn 其实是在 matplotlib 的基础上进行了更高级的 API 封装，从而使得作图更加容易，在大多数情况下使用 seaborn 就能做出很具有吸引力的图，而使 matplotlib 就能制作具有更多特色的图。

Matplotlib 是一个 Python 的 2D 绘图库，它以各种硬拷贝格式和跨平台的交互式环境生成出版质量级别的图形。通过 Matplotlib，开发者可以仅需要几行代码，便可以生成绘图。一般可绘制折线图、散点图、柱状图、饼图、直方图、子图等等。Matplot 使用 Numpy 进行数组运算，并调用一系列其他的 Python 库来实现硬件交互。

2.2 导入数据表

```
#读取csv中的数据, 以“,”作为分隔符来添加字段名
datas=pd.read_csv('data.csv', delimiter=',', encoding='utf8')
datas.head() #显示顶部5条信息
```

	_id	city	fields.cityName	fields.comment	fields.discountPrice	fields.distance	fields.features	fields.featuresNew	fields.fromName	fields.itemTotalScore
0	5cc7189f04eeab7d62a51c94	石家庄		NaN	NaN	81.2	NaN	[[('key': '一句话描述', 'text': '千奇百怪的珊瑚, 还能亲密触碰浅海动物, ...	NaN	NaN
1	5cc718a004eeab7d62a51c95	石家庄		NaN	NaN	35.9	NaN	[[('key': '一句话描述', 'text': '依据中国古典文学名著《红楼梦》严格设计...	NaN	NaN
2	5cc718a004eeab7d62a51c96	石家庄		NaN	NaN	150.0	NaN	[[('key': '一句话描述', 'text': '玻璃吊桥很刺激, 空中漫步越过峡谷的体验...	NaN	NaN
3	5cc718a004eeab7d62a51c97	石家庄		NaN	NaN	148.0	NaN	[[('key': '一句话描述', 'text': '享受温泉沐浴的同时, 还能畅玩水上乐园, ...	NaN	NaN
4	5cc718a004eeab7d62a51c98	石家庄		NaN	NaN	164.0	NaN	[[('key': '一句话描述', 'text': '参观小动物的同时还体验跳楼机等各种刺激...	NaN	NaN

2.3 数据清洗

2.3.1 数据提取

```
#提取需要的列
col_keep = ['city', 'fields.comment', 'fields.discountPrice', 'fields.features', 'fields.itemId', 'fields.itemTotalScore',
            'fields.price', 'fields.shortInfo', 'fields.sold365', 'fields.soldRecentNum', 'fields.tagList', 'fields.title']

datas=datas[col_keep]
#对提取的列进行重命名
datas.columns = ['city', '评论量', '折扣价', '产品特征', 'ID', '评分', '原价', '城市', '年销量', '月销量', '星级', '项目名']
```

```
#查看所有信息
datas.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54675 entries, 0 to 54674
Data columns (total 12 columns):
city          54675 non-null object
评论量        50187 non-null float64
折扣价        54675 non-null float64
产品特征      54675 non-null object
ID            54675 non-null int64
评分          50187 non-null float64
原价          54675 non-null float64
城市          50187 non-null object
年销量        32193 non-null object
月销量        54675 non-null int64
星级          54675 non-null object
项目名        54675 non-null object
dtypes: float64(4), int64(2), object(6)
memory usage: 5.0+ MB
```

	city	评论量	折扣价	产品特征	ID	评分	原价	城市	年销量	月销量	星级	项目名
0	石家庄	NaN	81.2	[[('key': '一句话描述', 'text': '千奇百怪的珊瑚, 还能亲密触碰浅海动物, ...	22189	NaN	81.2	NaN	NaN	104	[[('text': '5A景区', 'icon': None, 'borderColor': ...	石家庄海洋公园
1	石家庄	NaN	35.9	[[('key': '一句话描述', 'text': '依据中国古典文学名著《红楼梦》严格设计...	2091	NaN	35.9	NaN	NaN	336	[[('text': 'AAA', 'icon': None, 'borderColor': ...	荣国府
2	石家庄	NaN	150.0	[[('key': '一句话描述', 'text': '玻璃吊桥很刺激, 空中漫步越过峡谷的体验...	24408	NaN	150.0	NaN	NaN	65	[[('text': '5A景区', 'icon': None, 'borderColor': ...	红崖谷
3	石家庄	NaN	148.0	[[('key': '一句话描述', 'text': '享受温泉沐浴的同时, 还能畅玩水上乐园, ...	1221	NaN	148.0	NaN	NaN	162	[[('text': '4A景区', 'icon': None, 'borderColor': ...	白鹿温泉
4	石家庄	NaN	164.0	[[('key': '一句话描述', 'text': '参观小动物的同时还体验跳楼机等各种刺激...	3995	NaN	164.0	NaN	NaN	246		幸福王国

2.3.2 数据去重

```
#输出去重前的行数
print('去重前数据行数:', len(datas))
```

去重前数据行数: 54675

```
#subset: 列名, 可选, 默认为None; inplace=True表示直接在原来的Data上删除重复项
datas=datas.drop_duplicates(subset='ID')
```

```
#输出去重后的行数
print('去重后数据行数:', len(datas))
```

```
#查看去重后的顶部数据
datas.head()
```

去重后数据行数: 48537

	city	评论 量	折扣 价	产品特征	ID	评分	原价	城市	年销 量	月销 量	星级	项目名
0	石家庄	NaN	81.2	[[{'key': '一句话描述', 'text': '千奇百怪的珊瑚, 还能亲密触碰浅海动物, ...	22189	NaN	81.2	NaN	NaN	104	[[{'text': '5A景区', 'icon': None, 'borderColor': ...	石家庄海洋公园
1	石家庄	NaN	35.9	[[{'key': '一句话描述', 'text': '依据中国古典文学名著《红楼梦》严格设计...	2091	NaN	35.9	NaN	NaN	336	[[{'text': 'AAA', 'icon': None, 'borderColor': ...	荣国府
2	石家庄	NaN	150.0	[[{'key': '一句话描述', 'text': '玻璃吊桥很刺激, 空中漫步越过峡谷的体验...	24408	NaN	150.0	NaN	NaN	65	[[{'text': '5A景区', 'icon': None, 'borderColor': ...	红崖谷
3	石家庄	NaN	148.0	[[{'key': '一句话描述', 'text': '享受温泉沐浴的同时, 还能畅玩水上乐园, ...	1221	NaN	148.0	NaN	NaN	162	[[{'text': '4A景区', 'icon': None, 'borderColor': ...	白泉温泉
4	石家庄	NaN	164.0	[[{'key': '一句话描述', 'text': '参观小动物的同时还体验跳楼机等各种刺激...	3995	NaN	164.0	NaN	NaN	246		幸福王国

2.3.3 数据类型转换

```
# 由于np.nan是float, 所以这里整体转换为float
datas.loc[datas['评论量'].isnull(), '评论量'] = np.nan
datas['评论量'] = datas['评论量'].astype(float)
datas['折扣价'] = datas['折扣价'].astype(float)
datas['ID'] = datas['ID'].astype(int)
datas['原价'] = datas['原价'].astype(float)
datas['月销量'] = datas['月销量'].astype(int)
```

2.3.4 各字段清洗

```
def get_fea_text(data):
    if len(data) > 0:
        d=eval(data) #把字符串类型的字典内容转为字典类型
        if d: #如果字典d不为空
            return d[0]['text'] #就返回text值
        else:
            return None
    else:
        return None
#清洗产品特征
datas['产品特征'] = datas['产品特征'].apply(get_fea_text)
#清洗星级数据
datas['星级'] = datas['星级'].apply(get_fea_text)
```

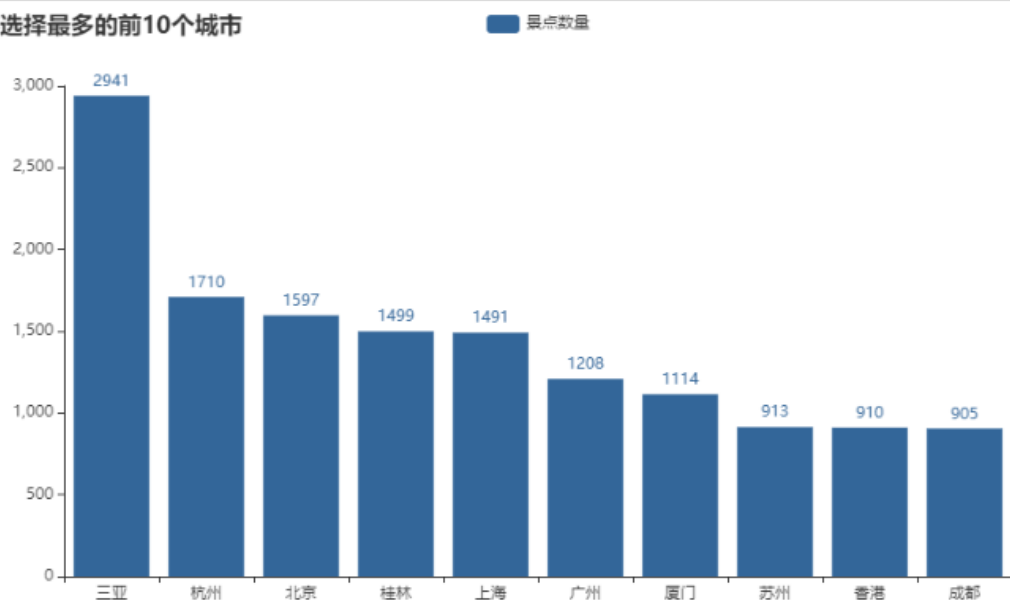
```
# fields sold365清洗
def get_sold365(data):
    if data is np.nan:
        return np.nan
    if '万' in data:
        return float(data[data.find('售')+1:data.find('万')]) * 10000
    else:
        return float(data[data.find('售')+1:data.find('笔')])
#清洗年销量数据
datas['年销量'] = datas['年销量'].apply(get_sold365)
```

2.4 数据分析

2.4.1 对城市、省份可选择景点进行统计

```
from pyecharts.charts import Bar #柱状图
from pyecharts import options as opts
bar = (
    Bar(init_opts=opts.InitOpts(page_title="旅游选择最多的前10个城市"))
    .set_global_opts(title_opts=opts.TitleOpts(title="旅游选择最多的前10个城市"))
    .add_xaxis(count_top10.index.tolist()) #行标签
    .add_yaxis("景点数量", count_top10.values.tolist()) #列标签
    .set_colors("#336699") #设置柱形体颜色
)
bar.render("旅游选择最多的前10个城市.html")
bar.render_notebook()
```

旅游选择最多的前10个城市



```
from pyecharts.charts import Map #地图
map = (Map(init_opts=opts.InitOpts(page_title="全国各省份旅游选择数量图"))
    .add("", [list(z) for z in zip(province, count_province_top10.values.tolist())], "china")
    .set_global_opts(title_opts=opts.TitleOpts(title="全国各省份旅游选择数量图"), visualmap_opts=opts.VisualMapOpts(max_=5000))
)
map.render("全国各省份旅游选择数量图.html")
map.render_notebook()
```

全国各省份旅游选择数量图

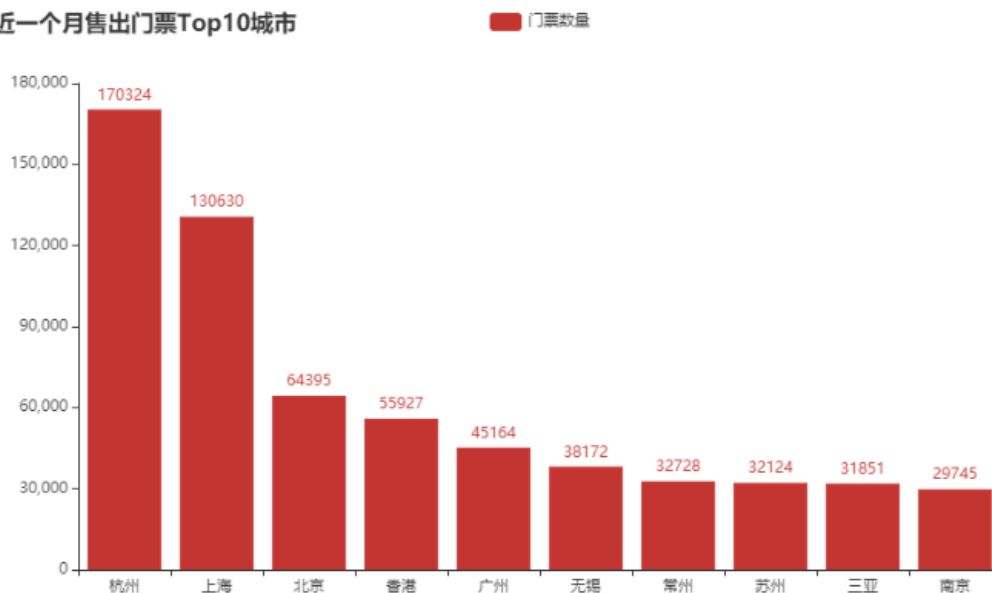


对于旅行，部分旅客希望能在较短的时间内游玩到更多的景点。对于各城市及省份的景区选择数量，我们分别进行了统计。如上图所示，三亚为选择最多的城市，杭州其次。在省份中，澳门，香港等选择最多。此处建议游客选择目标景区时，可考虑上述城市或省份。

2.4.2 对热门城市、景点的情况统计

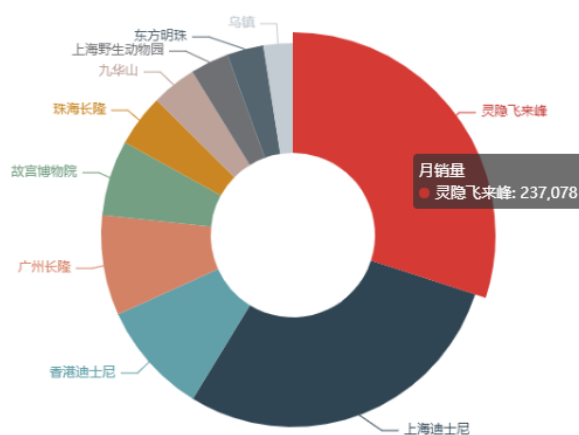
```
from pyecharts.charts import Bar#柱状图
bar=(
    Bar(init_opts=opts.InitOpts(page_title="最近一个月售出门票Top10城市"))
    .set_global_opts(title_opts=opts.TitleOpts(title="最近一个月售出门票Top10城市"))
    .add_xaxis(city_sale_top10.index.tolist())
    .add_yaxis("门票数量",city_sale_top10.values.tolist())
)
bar.render("最近一个月售出门票Top10城市.html")
bar.render_notebook()
```

最近一个月售出门票Top10城市



```
from pyecharts.charts import Pie#饼状图
x_data=['灵隐飞来峰','上海迪士尼','香港迪士尼','广州长隆','故宫博物院','珠海长隆','九华山','上海野生动物园','东方明珠','乌镇']
y_data=[119349+117729, 227123, 75092, 41234+25421, 50138, 34822, 29984, 25642, 24243, 19557]
data=list(zip(x_data,y_data))
pie=(
    #创建饼图对象Pie()
    Pie()
    .add(series_name="月销量",data_pair=data,radius=["30%","70%"])
    .set_global_opts(title_opts=opts.TitleOpts(title="最热门的10个景点",pos_left="center"),
        legend_opts=opts.LegendOpts(is_show=False))
)
pie.render("最热门的10个景点.html")
pie.render_notebook()
```

最热门的10个景点



根据上述数据分析结果图，我们可以看出，在总数据中杭州、上海等城市较为热门，灵隐飞来峰，上海迪士尼，香港迪士尼等景区较为热门。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决定具体目标景点。

2.4.3 对景点优惠力度分析

```
from pyecharts.charts import Line
from pyecharts.faker import Faker
line = (
    Line() # 生成line类型图表
    .add_xaxis(dis_data['城市'][:10].tolist()) # 添加x轴
    .add_yaxis("折扣", dis_data['折扣'][:10].tolist()) # 添加y轴
    .set_global_opts(title_opts=opts.TitleOpts(title="景点打折力度Top10"), yaxis_opts=opts.AxisOpts(min_='2.5', max_='4'))
)

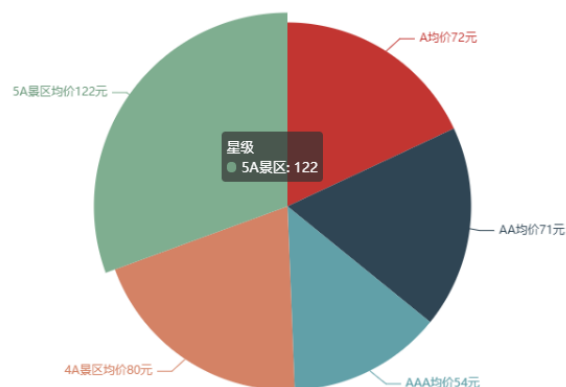
line.render('景点打折力度Top10.html') # 生成一个网页文件, 打开网页就是下图
line.render_notebook()
```

景点打折力度Top10



```
from pyecharts.charts import Pie # 饼图
x_data=level_data.index.tolist()
y_data=np.round(level_data.values,0).tolist()
data=list(zip(x_data,y_data))
pie = (
    # 创建饼图对象Pie()
    Pie()
    .add(series_name='星级', data_pair=data)
    .set_global_opts(title_opts=opts.TitleOpts(title="各级别景区的门票价格", pos_left="center"),
        legend_opts=opts.LegendOpts(is_show=False))
    .set_series_opts(label_opts=opts.LabelOpts(formatter='{b}均价 {c}元'))
)
pie.render("各级别景区的门票价格.html")
pie.render_notebook()
```

各级别景区的门票价格



由上图可得, 总数据中景区等级越高门票价格也就相应升高, 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。

3 问题与解决方案

3.1 Github 中对文件的处理

在 git bash 中键入 `$ git push origin master` 进行提交的时候出现如下错误:

```
MINGW64:/f/git
LOVE@DESKTOP-TGML049 MINGW64 /f/git (master)
$ git push origin master
To github.com:scholar-gif/Python.git
! [rejected]        master -> master (fetch first)
error: failed to push some refs to 'github.com:scholar-gif/Python.git'
hint: Updates were rejected because the remote contains work that you do
hint: not have locally. This is usually caused by another repository pushing
hint: to the same ref. You may want to first integrate the remote changes
hint: (e.g., 'git pull ...') before pushing again.
hint: See the 'Note about fast-forwards' in 'git push --help' for details.
```

问题原因: 远程库与本地库不一致造成的, 先把远程库同步到本地库就可以了。

解决办法: 使用命令行: `git pull --rebase origin master`

该命令的意思是把远程库中的更新合并到 (pull=fetch+merge) 本地库中, `--rebase` 的作用是取消掉本地库中刚刚的 commit, 并把他们接到更新后的版本库之中。执行 pull 执行成功后, 可以成功执行 `git push origin master` 操作。

```
MINGW64:/f/git
LOVE@DESKTOP-TGML049 MINGW64 /f/git (master)
$ git pull --rebase origin master
remote: Enumerating objects: 3, done.
remote: Counting objects: 100% (3/3), done.
remote: Compressing objects: 100% (2/2), done.
remote: Total 2 (delta 1), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (2/2), 614 bytes | 10.00 KiB/s, done.
From github.com:scholar-gif/Python
 * branch                master      -> FETCH_HEAD
 15ab048..68b34cf master      -> origin/master
Updating 15ab048..68b34cf
Fast-forward
 "\346\226\207\344\273\266.txt" | 11 -----
 1 file changed, 11 deletions(-)
 delete mode 100644 "\346\226\207\344\273\266.txt"
```

```
MINGW64:/f/git
LOVE@DESKTOP-TGML049 MINGW64 /f/git (master)
$ git push origin master
Enumerating objects: 7, done.
Counting objects: 100% (7/7), done.
Delta compression using up to 4 threads
Compressing objects: 100% (5/5), done.
Writing objects: 100% (5/5), 664 bytes | 221.00 KiB/s, done.
Total 5 (delta 2), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (2/2), completed with 2 local objects.
To github.com:scholar-gif/Python.git
 d862e83..15ab048 master -> master
```

3.2 Pyecharts 版本

在使用 pyecharts 的绘制图形时 mac 一直出错，然后网上查了很久没有查到，最后发现可能是版本问题。

```
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'
```

pyecharts 是一个用于生成 echart（百度开源的数据可视化 javascript 库）图表的类库，分为 v0.5.x 和 v1.x 两个大版本，而新版本是不向下兼容的，查了一下自己的版本是 v0.5.10 版本的。

❏ 选择 C:\Windows\system32\cmd.exe

```
prometheus-client      0.9.0
prompt-toolkit          3.0.8
pycparser               2.20
pyecharts               0.5.10
pyecharts-javascripton  0.0.6
pyecharts-jupyter-installer 0.0.3
Pygments               2.7.2
```

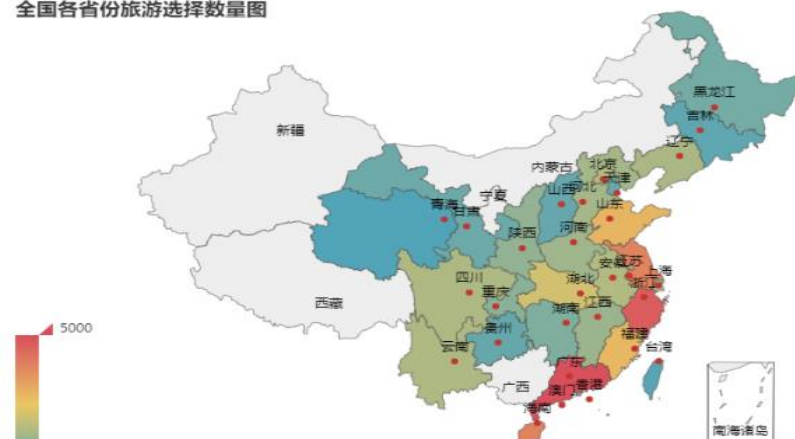
解决方法：装新版本 1.9.0

```
C:\Windows\system32\cmd.exe
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/a094c0d11c3c87cf407cbd9dc069882e9bba08374df1ad81dfb39de1a8d3/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: wcwidth in c:\users\love\appdata\local\programs\python\python37\lib\site-packages (from prettytable->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 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.
```

然后问题就解决了，所有图片都能正常出来。

```
from pyecharts.charts import Map
map = Map(init_opts=opts.InitOpts(page_title="全国各省份旅游选择数量图"))
map.add("", [list(z) for z in zip(province, count_province_top10.values.tolist())], "china")
map.set_global_opts(title_opts=opts.TitleOpts(title="全国各省份旅游选择数量图"), visualmap_opts=opts.VisualMapOpts(max_=5000))
map.render("全国各省份旅游选择数量图.html")
map.render_notebook()
```

全国各省份旅游选择数量图



4 总结体会

经过这次课设,给予我们实践的机会使我积累了必须的经验,为以后的学习打下了基础,也为毕业设计铺平了道路。经过这次的课程设计,使我学会了很多东西,包括书本上的理论知识及书本上学不到的知识。具体的收获与体会如下:

一、经过课程设计,培养了我耐心、仔细、谨慎的工作态度。虽然刚开始的时候很懵,但在通过各种途径慢慢去了解 jupyter notebook、git、github 的过程中,发现没有我想象中那么复杂,只是需要多花些心思,所有恐惧都来源于未知。后来发现老师给的资料里面有已经爬下来的数据以及写好的代码时,突然觉得事情变得好简单,但在实践过程中又发现没有那么简单,各种错误随之而来,我们又只好一条条修复 bug。终于最后还是被我们整体运行出来,又想着创新,却发现没有多少时间了,最后只能换几个图形类型。整个过程是对知识的探索,每天都会有些不同的收获,但由于路上走的太慢,以至于走不了太远,但没关系,只要持之以恒,总有走到终点的一天。

二、此次课程设计中,我们也存在部分不足,如前期准备时间过长,导致后期时间不足,无法尝试更多的内容,如将数据转为数据库文件等。在往后的此类课设中,我们希望能继续尝试这方面的内容,在此次的基础上,学习到更多的知识,得到更多的尝试。

三、经过这次课程设计,使我更充分认识了团队合作的重要性。由于这次课设是以小组为单位,每个人都有分工。在设计的过程中我们也遇到了较多的问题,小组中的每一个成员都没有产生过放弃的念头,而这些问题最终也都得到了解决。此次课程设计也对我们团队精神进行了考察,让我们更加的配合默契,沟通也得到了提升,在成功后一起体会喜悦的心情。

最终,我对我们小组的成员同学表示衷心的感谢,因为我们是靠互帮互助才能完成课程设计的,我想其他同学也跟我有一样感受,同时也要感激我们的指导教师,教师严谨细致、一丝不苟的作风一向是我工作、学习中的榜样;教师循循善诱的教导和不拘一格的思路给予我无尽的启迪。

5 小组分工明细表

项目名称	节假日长假景点人满为患的数据分析		
任务名称	主负责人	协助人	协助人
小组成员	胡佳	李文娟	薛文
选题	胡佳	李文娟	薛文
小组讨论分析	胡佳	李文娟	薛文
建立 github	胡佳	薛文	无
环境搭建	胡佳	李文娟	无
数据清洗	薛文	无	无
数据分析	胡佳	李文娟	无
注释文档编写	薛文	李文娟	无
PPT 报告编写	李文娟	薛文	无
Github 管理	胡佳	无	无
项目汇报	李文娟	无	无

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>问题 1：什么是 pyecharts？</p> <p>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>问题 2：简述 pandas 中的 apply 函数？</p> <p>apply()函数的基本形式为：DataFrame.apply(func, axis=0, broadcast=False, raw=False, reduce=None, args=(), **kwds)其中，func 是你想要这个 DataFrame 应用的函数，可以自己编写，也可以是已经存在的，相当于 C/C++的函数指针。如果这个参数是自己实现的话，函数的传入参数根据 axis 来定，比如 axis = 1，就会把一行数据作为 Series 的数据结构传入给自己实现的函数中，我们在函数中实现对 Series 不同属性之间的计算，返回一个结果，则 apply 函数会自动遍历每一行 DataFrame 的数据，最后将所有结果组合成一个 Series 数据结构并返回。</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>		