

学号	10212818108
----	-------------

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	<u>数据分析的学习与实践-地震的数据分析</u>
专 业	<u>软件工程</u>
班 级	<u>软件 1181</u>
姓 名	<u>陈廷锋</u>
成 绩	<u></u>
指 导 老 师	<u>徐勇</u>

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

1 项目选题 .....	6
2 项目介绍 .....	6
3 环境介绍 .....	6
3.1 Jupyter Notebook .....	6
3.2 GitHub .....	6
4 小组介绍 .....	7
4.1 小组成员 .....	7
4.2 任务分配 .....	7
5 数据分析 .....	7
5.1 数据分析方向 .....	7
6 总结 .....	11
参考文献 .....	11

# 1 项目选题

地震的数据分析

## 2 项目介绍

地震（英文名称:earthquake）又称地动、地振动，是地壳快速释放能量过程中造成的振动，期间会产生地震波的一种自然现象。地球上板块与板块之间相互挤压碰撞，造成板块边沿及板块内部产生错动和破裂，是引起地震的主要原因。地震开始发生的地点称为震源，震源正上方的地面称为震中。破坏性地震的地面振动最烈处称为极震区，极震区往往也就是震中所在的地区。地震常常造成严重人员伤亡，能引起火灾、水灾、有毒气体泄漏、细菌及放射性物质扩散，还可能造成海啸、滑坡、崩塌、地裂缝等次生灾害。

地震成因:地球表层的岩石圈。地壳岩层受力后快速破裂错动引起地表振动或破坏就叫地震。由于地质构造活动引发的地震叫构造地震；由于火山活动造成的地震叫火山地震；固岩层（特别是石灰岩）塌陷引起的地震叫塌陷地震。

## 3 环境介绍

### 3.1 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。

### 3.2 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。

2018 年 6 月 4 日，微软宣布，通过 75 亿美元的股票交易收购代码托管平台 GitHub。

## 4 小组介绍

### 4.1 小组成员

陈廷锋、黄婷、李思非

### 4.2 任务分配

陈廷锋：答辩、调试代码

黄婷：PPT 制作、报告框架、数据采集

李思非：github 上传及管理

三个人都参与代码调试。

## 5 数据分析

### 5.1 数据分析方向

#### 1、引发地震的因素有哪些？

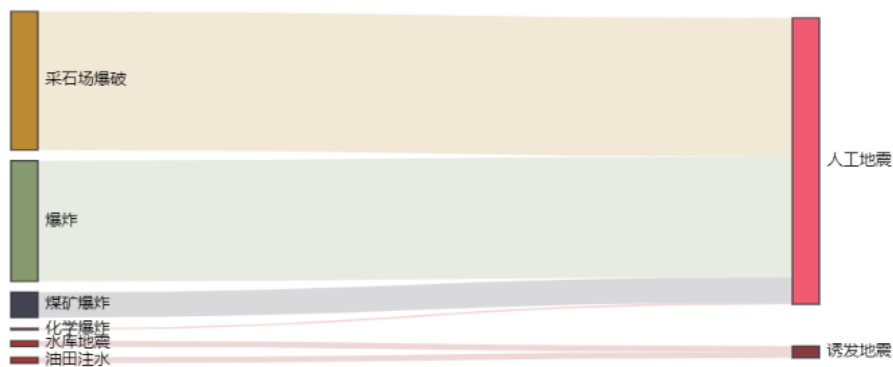
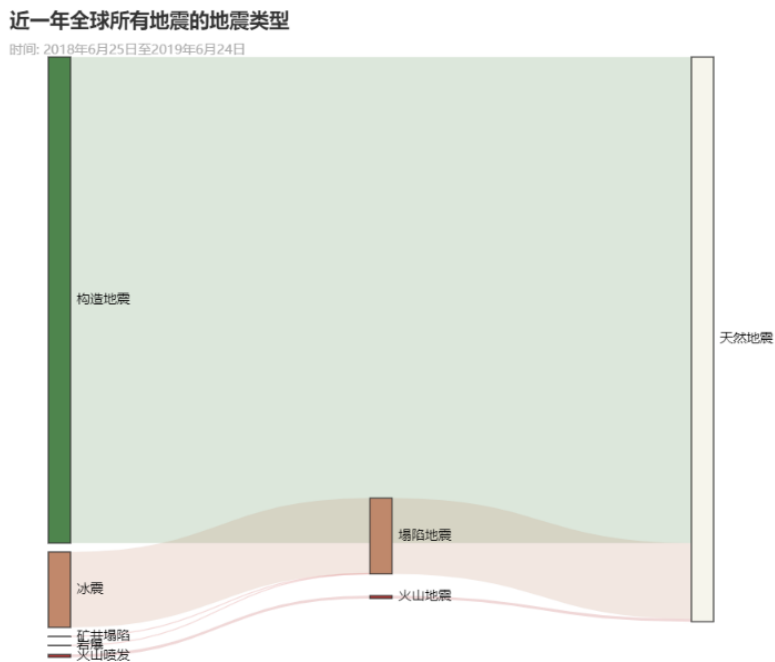
通过代码 `eq_type = eq1_year['type'].value_counts()` 对数据类型进行统计，得出地震引发的因素发生的次数。如图：5.1 所示

```
In [9]: eq_type = eq1_year['type'].value_counts()
eq_type

Out[9]: earthquake      158194
quarry blast           1173
explosion               1023
ice quake              776
mining explosion       214
other event            136
volcanic eruption      28
chemical explosion     15
rock burst              2
mine collapse           2
sonic boom              2
Name: type, dtype: int64
```

图 5.1 类型统计代码

地震分天然地震(图 5.2)和人工地震(图 5.3)两大类。天然地震主要是构造地震,它是由于地下深处岩石破裂、错动把长期积累起来的能量急剧释放出来,以地震波的形式向四面八方传播出去,到地面引起的房摇地动。



由上图可以知道:构造地震约占地震总数的 90%以上。其次是由塌陷发引起的地震,称为塌陷地震,约占地震总数的 7%。此外,某些特殊情况下了也会产生地震,如火山爆发(火山地震)等。人工地震是由人为活动引起的地震。

2、全世界地震频发的地区有哪些？

我们对大于 6 级以上的数据进行分析，以地区进行次数统计得到如图 5.4 所示：



```

In [22]: eq20_years['region'] = region

In [23]: region_top10 = eq20_years['region'].value_counts().nlargest(11)
region_top10

Out[23]: Indonesia      15856
Japan      9701
Papua New Guinea      7969
Chile      5309
Philippines      4461
Tonga      4230
Vanuatu      3562
New Zealand      2999
Solomon Islands      2970
Alaska      2745
Japan region      2577
Name: region, dtype: int64

```

图 5.4 地区发生次数

可以看到，印尼近二十年 6 级以上的地震高达 15000 次，其次是日本高达 12000 次，如图 5.5 所示：

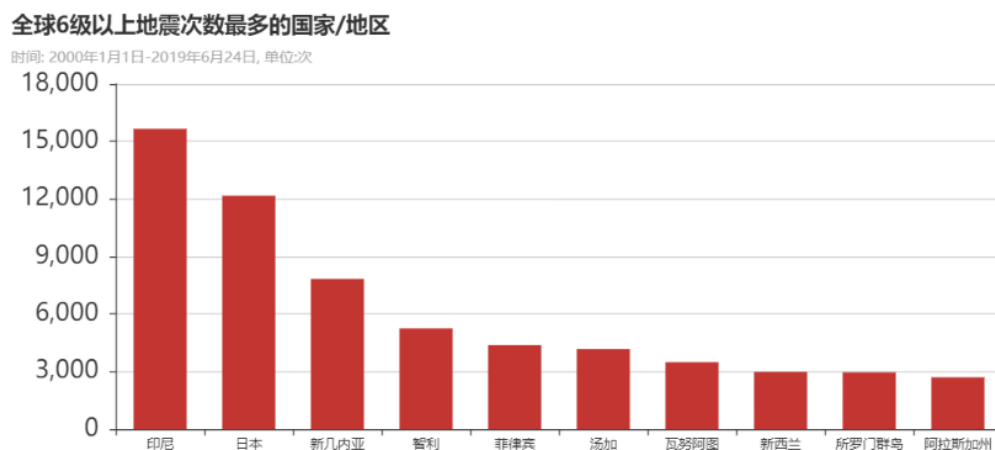


图 5.5 全球 6 级以上地震次数最多的国家/地区

### 3、有哪些引发全世界舆情关注的大地震？

我们对 6 级以上的数据进行分析，如图 5.6 所示代码可知：6 级到 7 级以上的次数为 2098 起，7 级到 8 级以上的次数为 288 起，8 级到 9 级以上的次数为 22 起，9 级以上的为 2 起。图 5.7 部分大地震数据。

```

In [41]: eq_gt_6 = eq20_years[eq20_years['mag'] > 6]

In [42]: eq_gt_6[eq_gt_6['mag'] < 7].shape[0]

Out[42]: 2098

In [43]: eq_gt_6[(eq_gt_6['mag'] >= 7) & (eq_gt_6['mag'] < 8)].shape[0]

Out[43]: 288

In [44]: eq_gt_6[(eq_gt_6['mag'] >= 8) & (eq_gt_6['mag'] < 9)].shape[0]

Out[44]: 22

In [45]: eq_gt_6[eq_gt_6['mag'] >= 9].shape[0]

Out[45]: 2

```

图 5.6 地震次数代码统计图

```

In [44]: eq_gt_7.loc[eq_gt_7.sort_values(by=['mag', 'depth'], ascending=[False, True]).index].head(30)

```

	time	latitude	longitude	depth	mag	magType	id	place	type	status	region	influence
110617	2011-03-11 05:46:24.120000+00:00	38.2970	142.3730	29.00	9.1	mww	official20110311054624120_30	2011 Great Tohoku Earthquake, Japan	earthquake	reviewed	Japan	9
103381	2004-12-26 00:58:53.450000+00:00	3.2950	95.9820	30.00	9.1	mw	official20041226005853450_30	2004 Sumatra - Andaman Islands Earthquake	earthquake	reviewed	Sumatra - Andaman Islands	9
8669	2010-02-27 06:34:11.530000+00:00	-36.1220	-72.8980	22.90	8.8	mww	official20100227063411530_30	offshore Bio-Bio, Chile	earthquake	reviewed	Chile	8
117108	2012-04-11 08:38:36.720000+00:00	2.3270	93.0630	20.00	8.6	mw	official20120411083836720_20	off the west coast of northern Sumatra	earthquake	reviewed	off the west coast of northern Sumatra	8
100579	2005-03-28 00:00:00.000000+00:00	2.0850	97.1080	30.00	8.6	mww	official20050328160936530_30	northern Sumatra	earthquake	reviewed	Indonesia	8

图 5.7 部分大地震数据

#### 4、为什么四川省及其周围好像从 08 年开始就地震频发？

通过对数据文件的分析可以看到四川省及其周围从 08 年开始就地震频发。如图 5.8 所示；

近40年四川省及其周围4.5级以上地震次数

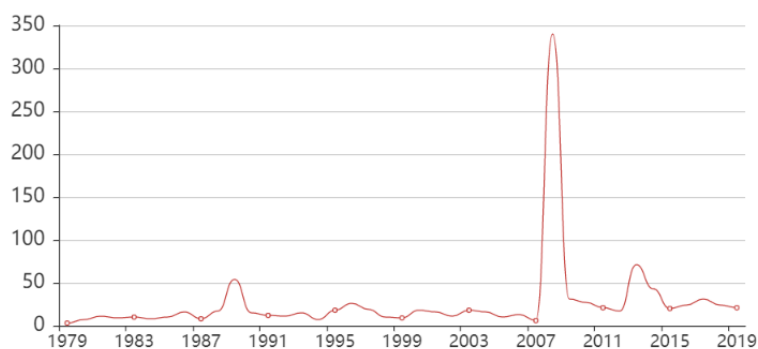


图 5.8 近 40 年四川省及其周围 4.5 级以上地震次数

## 6 总结

当前的科技水平尚无法预测地震的到来，未来相当长的一段时间内，地震也是无法预测的。所谓成功预测地震的例子，基本都是巧合。对于地震，我们更应该做的是提高建筑抗震等级、做好防御，而不是预测地震。

地震是一种无法干预、目前还不能完全预测的自然现象和灾害。它告诉我们要敬畏自然、理解自然，并且要通过学习掌握自然规律，来减少灾害造成的影响。在地震来临时不信谣、不传谣，做好相关的应对措施，正确应对。

## 参考文献

- [1] Python 基础与大数据应用. 丁辉. 人民邮电出版社, 2020. 1
- [2] CSDN: <https://blog.csdn.net/xvsong5142540/article/details/106849168>
- [3] 博客园: <https://www.cnblogs.com/beeblog72/p/13080727.html>

## 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	
<p>课程设计学生答辩或质疑记录：</p> <p>1. 编码过程中遇到什么问题？</p> <p>答：刚开始跑代码的时候，图表出不来，一直报错，后面发现是 pyecharts 的版本太低，我们换成了 0.5.5 的版本就可以正常运行。</p> <p>2. 可视化扩展库 matplotlib 的模块 pyplot 中哪个函数可以用来设置同一个画布中多个子图之间的水平间距和垂直间距？</p> <p>答：subplots_adjust() 函数。</p> <p>3. %matplotlib 是干嘛的？</p> <p>答：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。</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>		