|  |  |
| --- | --- |
| 学号 | 10214518323 |



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- 7）针对996工作，程序员群体的看法的分析 |
| 专 业 | \_\_\_\_\_\_\_\_\_\_软件工程\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 班 级 | \_ \_\_\_\_\_\_\_\_软件1182\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 姓 名 | \_\_\_\_\_\_\_\_\_\_\_黄鸿宇\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 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.设计目标 7](#_Toc62229416)

[2.设计内容 8](#_Toc62229417)

[3 .996icu项目 8](#_Toc62229418)

[3.1 项目背景 8](#_Toc62229419)

[3.2 用户群体 9](#_Toc62229420)

[3.3 使用技术说明 9](#_Toc62229421)

[3.4 项目需求概述 10](#_Toc62229422)

[3.5 总体设计 11](#_Toc62229423)

[3.5.1 数据流程图 11](#_Toc62229424)

[4.具体项目及代码展示 11](#_Toc62229425)

[4.1 关注996的程序员主要来自哪些公司的可视化 12](#_Toc62229426)

[4.1.1 效果图 12](#_Toc62229427)

[4.1.2 代码展示 12](#_Toc62229428)

[4.2 关注996的程序员的关注的问题的可视化 14](#_Toc62229429)

[4.2.1 效果图 14](#_Toc62229430)

[4.2.2 代码展示 15](#_Toc62229431)

[5.心得与体会 16](#_Toc62229432)

[8.参考文献 17](#_Toc62229433)

# 1.设计目标

* Python 课程设计，完成所有课程开发题目的需求分析、概要设计工作，制定具体的开发计划，并交付相关的文档;
* “面向对象”课程设计，根据以完成的文档和开发计划，实现自选的开发题目，输出代码和相关文档；
* “数据库原理与应用”课程设计，根据以完成的文档和开发计划，实现数据库应用相关自选的开发题目，输出代码和相关文档;

# 2.设计内容

* 使用 Python 语言完成“关注996的程序员主要来自哪些公司、高校、城市”分析；
  + 使用 Python 语言完成“关注996的程序员的画像”分析；
  + 用 Python 语言完成“关注996的程序员的简介词云图”分析；

# 3 .996icu项目

## 3.1 项目背景

在国内互联网圈，“996”工作制早就是一个公开的秘密，近日终于被摆到了台面上。马云、刘强东先后表态，《人民日报》发文评论，使这个现象获得了前所未有的关注。

但当我们在公共话语空间内讨论996时，往往忽视了这一概念下实际存在着各种各样的具体情景。不同阶段、不同规模的企业，可能以不同的形式执行高强度工作模式，而身处不同位置的员工，对这种工作状态的感受也截然不同。

996.ICU指工作996、生病ICU，也就是工作从早上9点上班到晚上9点下班，每周工作6天，生病了就住进ICU。2019年3月27日，一个名为996ICU的项目在GitHub上传开。程序员们揭露“996ICU”互联网公司，抵制互联网公司的996工作制度。在这个项目指向的域名页面上，发起人这样写到：“什么是996.ICU？工作996，生病ICU”。他将996工作制下最低72个工时与《劳动合同法》等条文对比，并呼吁“程序员生命为重。“996ICU”成为GitHub上最受欢迎的项目，越来越多的人给这个项目“加星”，这是程序员们对一个开源项目表示肯定或支持的方式，类似于“点赞”。仅仅三四天后，“996ICU”星数已超过10万，仅15天后，“996ICU”星数达到20万，GitHub排行榜第二 [7] ，而且还在继续增长。作为对比，在iOS这个话题下的一些热门项目，获得的星数也仅在2到5万之间。有人称996ICU为“GitHub星数上升最快的项目”，在舆论声中，大家开始将这个项目定义为“程序员的一次反抗”。

## 3.2 用户群体

关注996的程序员群体。

## 3.3 使用技术说明

基于 Python的数据清洗分析项目。

* 展示形式

对相关数据进行一些直观展示。

* 数据集规模

数据集的范围为GitHub上面的一个名为996ICU的项目。

* 数据集的质量

涉及了程序员的看法、所在公司等。

* 业务架构

业务架构分为2层：数据分析处理，可视化。

## 3.4 项目需求概述

新冠疫情大屏展示项目功能图解如下：

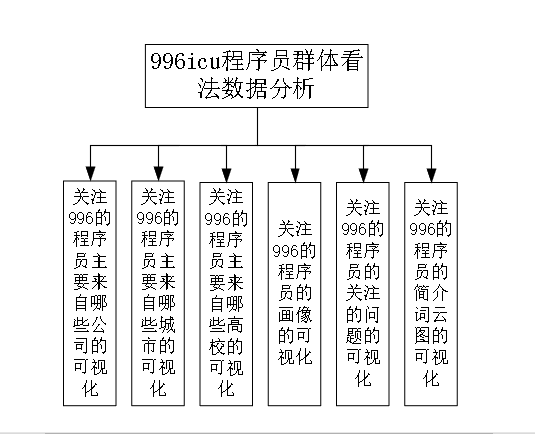


图3.4

## 3.5 总体设计

### 3.5.1 数据流程图



图3.5

# 4.具体项目及代码展示

## 4.1 关注996的程序员主要来自哪些公司的可视化

### 4.1.1 效果图

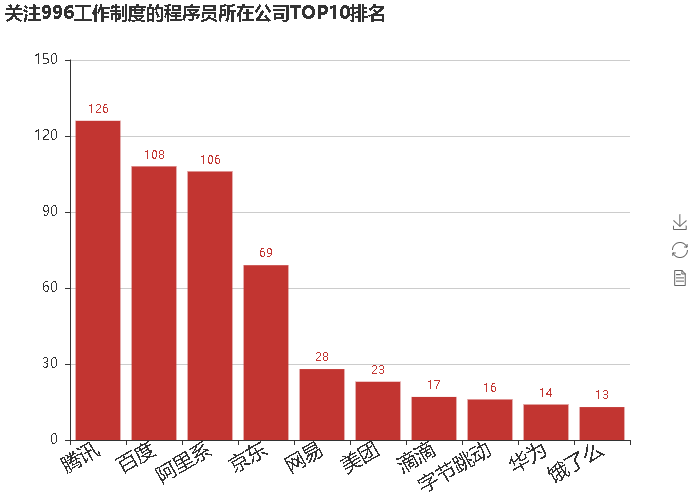


图4.1

### 4.1.2 代码展示

user\_data['company'].value\_counts()[:50]

user\_data['company'].value\_counts()[:100].index

def get\_company(data):

for com in ['tencent', '腾讯','Tencent']:

if com in data:

return '腾讯'

for com1 in ['baidu', '百度','Baidu','BAIDU']:

if com1 in data:

return '百度'

for com2 in ['Alibaba', '淘宝', 'Taobao', 'Alipay', '阿里巴巴', 'Aliyun', '阿里云','Alibaba Group']:

if com2 in data:

return '阿里系'

for com3 in ['JD', 'jd', '京东']:

if com3 in data:

return '京东'

for com4 in ['netease', 'NetEase', '网易']:

if com4 in data:

return '网易'

for com5 in ['Meituan', '美团']:

if com5 in data:

return '美团'

for com6 in ['bytedance', '字节', '头条']:

if com6 in data:

return '字节跳动'

for com7 in ['eleme', '饿了么']:

if com7 in data:

return '饿了么'

for com8 in ['Huawei', '华为']:

if com8 in data:

return '华为'

for com9 in['didi', 'DiDi', '滴滴', '嘀嘀']:

if com9 in data:

return '滴滴'

user\_data['company\_top'] = user\_data.loc[user\_data['company'].notna(), 'company'].apply(get\_company)

top10\_com = user\_data['company\_top'].value\_counts()

from pyecharts import Bar

bar = Bar("关注996工作制度的程序员所在公司TOP10排名", width = 700,height=500)

bar.add("", top10\_com.index, top10\_com.values, is\_stack=True,

xaxis\_label\_textsize=20, yaxis\_label\_textsize=14, is\_label\_show=True,

xaxis\_rotate=30)

bar

## 4.2 关注996的程序员的关注的问题的可视化

### 4.2.1 效果图



图4.2

### 4.2.2 代码展示

issues\_data.info()

import jieba

from collections import Counter

from pyecharts import WordCloud

jieba.add\_word('996')

jieba.add\_word('996制度')

jieba.add\_word('ICU')

swords = [x.strip() for x in open ('D:\ICU996\996\_data\stopwords.txt',encoding="utf-8")]

def plot\_word\_cloud(data, swords):

text = ''.join(data['title'])

words = list(jieba.cut(text))

ex\_sw\_words = []

for word in words:

if len(word)>1 and (word not in swords):

ex\_sw\_words.append(word)

c = Counter()

c = Counter(ex\_sw\_words)

wc\_data = pd.DataFrame({'word':list(c.keys()), 'counts':list(c.values())}).sort\_values(by='counts', ascending=False).head(100)

wordcloud = WordCloud(width=1300, height=620)

wordcloud.add("", wc\_data['word'], wc\_data['counts'], word\_size\_range=[20, 100])

return wordcloud

plot\_word\_cloud(issues\_data, swords=swords)

# 5.心得与体会

这一次我们进行了python的课程设计，学习和巩固这个学期所学习的知识，这次课程设计给我们最深刻的感受就是通过python这门课程让我们体会到了团队合作的重要性。首先作为一个团队，我们要处理好队员与队员之间的关系。团队不仅仅是把几个有共同目标的人凑到一起做事情，还会在团队存续期间乃至以后的生活中持续发挥影响。无论在工作中还是在生活中，都必须保证人之间的良好沟通。在这一点上，我感觉我们小组做的应该算是不错的，我们真的能够做到彼此真诚，能够为了完成一个共同的目标而一起努力。一滴水只有放进大海才会永不干涸。个人再完美，也不过是一滴水，唯有融入一个优秀的团队，才能获取源源不断的力量源泉。在我们小组合作期间，我们不仅仅是一起完成了工作，同时我们也更加的了解小组里的每一个人。团结就是力量，联合就有优势。融入团队并不是要磨平个人的棱角，团队就是一面凸透镜，它通过将分散的光线聚于一点，产生“1+1〉2”的强大效应。在我们小组之间建立了深厚的友谊关系。这是我感受较为深刻的地方。  
 注重项目团队合作，保证一个团体生机和活力，必须使得每一个成员能够相互支持和包容，成员间充分尊重对方意见，现在终于感受到了团队合作重要性，项目小组花了2个星期做这个，除了个人知识的融会贯通，更重要的是成员的沟通与协调。有人说，一个融洽的工作环境就是成功的一半。尤其是从事软件开发，团队就是发展的根源，更是一个好的系统的前提保证，一个人的思想是有限的，多远思维交织在一起，产生的力量是不可估量的。  
 本次课设看作是“理论与实践相结合的桥梁”.通过这次的课设和学习，我们知道了此次学习的目的，也清楚目前自己的不足，那就是缺乏相应的知识与经验，对所学的专业知识不能够很好地运用于实践操作.正所谓“百闻不如一见”，经过这次自身的切身实践，我才深切地理会到了“走出课堂，投身实践”的必要性.平时，我们只能在课堂上与老师一起纸上谈兵，思维的认识基本上只是局限于课本的范围之内，也许就是这个原因就导致我们对专业知识认识的片面性，使得我们只知所以然，而不知其之所以然！限制了我们网络知识水平的提高.但是课本上所学到的理论知识是为我们的实际操作指明了方向、提供相应的方法，真正的职业技巧是要我们从以后的实际工作中慢慢汲取的.而针对实际操作中遇到的一些特殊的问题，我们不能拘泥于课本，不可纯粹地“以本为本”.  
 有些知识点以前没有学过，但我们也没有去研究，实训时突然间觉得。我们真的有点无知，虽然现在去看依然可以解决问题，但要浪费许多时间，这一点是我必须在以后的学习中加以改进的地方，同时也要督促自己在学习的过程中不断的完善自我.另外一点，也是在实训中必不可少的部分，就是同学之间的互相帮助.所谓”当局者迷，旁观者清”，有些东西感觉自己做的是时候明明没什么错误，偏偏程序运行时就是有错误，让其他同学帮忙看了一下，发现其实是个很小的错误.所以说，相互帮助是很重要的一-点，这在以后的工作或生活中也是很关键俗话说：“要想为事业多添一把火，自己就得多添一捆材”。  
 此次课程设计，我们深深体会到了积累知识的重要性.在课设当中我们遇到了不少难题，但是经过我们大家的讨论和老师细心的一一指导，问题得到了解决.2个星期的课程设计结束了，收获颇丰，同时也更深刻的认识到要做一个合格的程序员并非我以前想像的那么容易，最重要的还是细致严谨.社会是不会要-一个一无是处的人的，所以我们要更多更快地从一个学生向工作者转变，总的来说我对这次实习还是比较满意的，它使我们学到了很多东西，为我们以后的学习做了引导，点明了方向。

# 8.参考文献

[1] 朱福喜.《python爬虫开发》（第二版）.科学出版社  
[2] 陈国君等.《python实战编程》（第二版）.清华大学出版社  
[3] Deitel.《python：从入门到实践》（第六版）.电子工业出版社    
[4] MaryCampione.《python语基础教程》（第四版）.机械工业出版社  
[5] Y.Daniel Liang.《Python语言程序设计进阶篇》（第六版）.机械工业出版社

**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如何对数据进行处理？  首先导入并筛选有效的数据，然后进行数据清洗，数据类型转换，之后构建模型，进行数据可视化的处理  ○2在做的过程中，应该注意什么？  在做的过程中，应尽可能多地观察数据  。  ○3在做的过程中遇到了哪些问题，如何解决的？  数据读取的时候，出现编码格式的问题 | | | |
| **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 | | | |