|  |  |
| --- | --- |
| 学号 | 10212818106 |



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-吴某凡微博热点的分析 |
| 专 业 | \_\_\_\_\_\_\_\_\_软件工程\_\_\_\_\_\_\_\_\_\_\_\_ |
| 班 级 | \_\_\_\_\_\_\_\_\_软件1181 \_\_\_\_\_\_\_\_\_\_\_ |
| 姓 名 | \_\_\_\_\_\_\_\_\_\_周宇伦\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 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

**1.小组人员与分工1**

**2.编程环境1**

**3.程序设计1**

3.1开发包导入1

3.2数据导入1

3.3数据清洗2

3.4吴亦凡微博是否存在假流量2

3.4.1吴亦凡粉丝性别比例2

3.4.2吴亦凡真假流量转发量3

3.5大家对《大碗宽面》怎么看3

3.6有多少人拿吴亦凡和蔡徐坤做对比3

3.7有多少人开始路转粉4

3.7.1路转粉男女性别比例4

3.8评论的词云图5

**4.遇到的问题5**

**5.问题汇总6**

**6.小组总结6**

**7.个人总结7**

**参考文献8**

**课程设计评定表9**

1.小组人员与分工

项目小组人员由周宇伦(组长)，王志东，郭晶鑫组成，周宇伦主要负责对数据的清洗与处理、答辩等工作，王志东主要负责数据的收集和GitHub上传，郭晶鑫主要负责对处理后的数据进行汇总做成ppt。

2.编程环境

硬件平台：联想拯救者

软件平台：windows10

编程环境：ipython3.7

开发平台：jupyter notebook

项目托管平台：GitHub

3.程序设计

3.1开发包导入

图3.1.1

 3.2数据导入

图3.2.1

图3.1.2

3.3数据清洗



图3.3.1

3.4吴亦凡微博是否存在假流量？

3.4.1吴亦凡粉丝性别比例

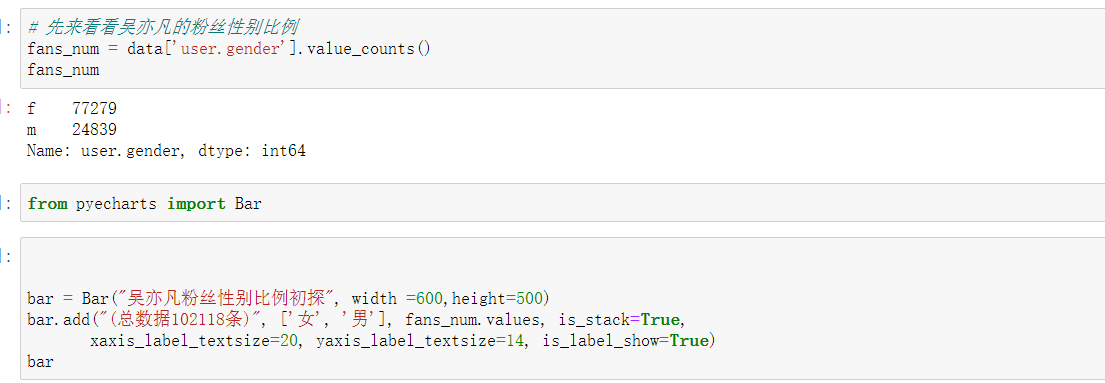


图3.4.1.1

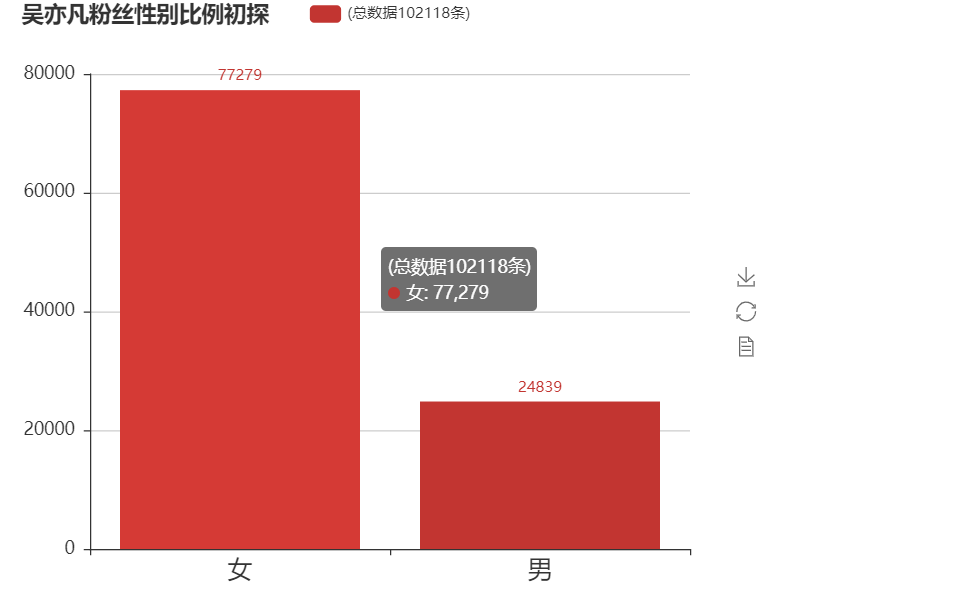


图3.4.1.2

3.4.2吴亦凡真假流量转发量

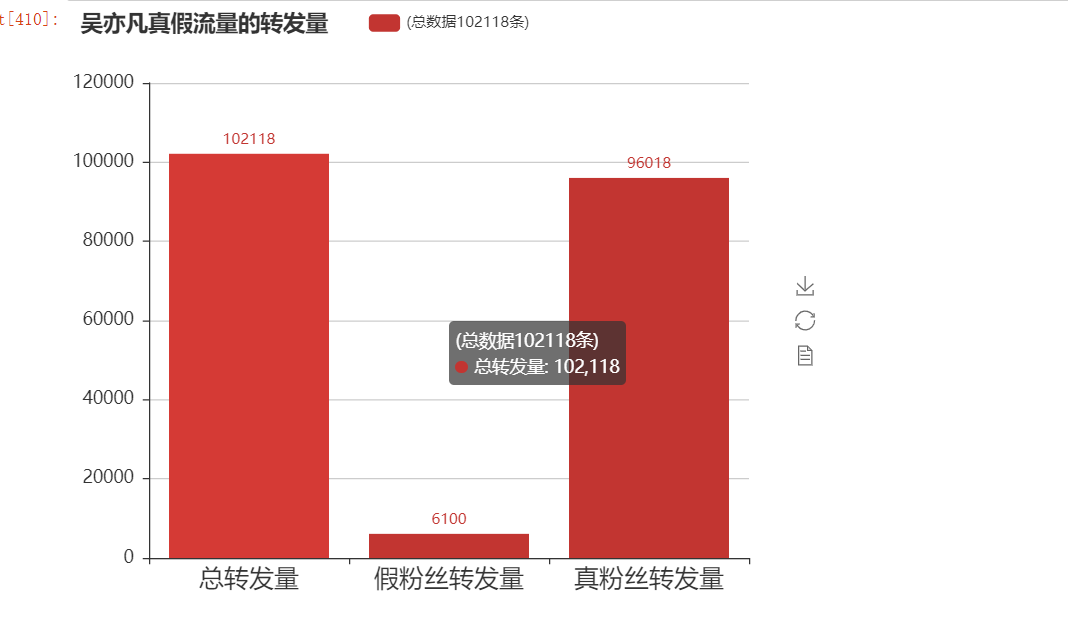


图3.4.2.1

3.5大家对于《大碗宽面》怎么看

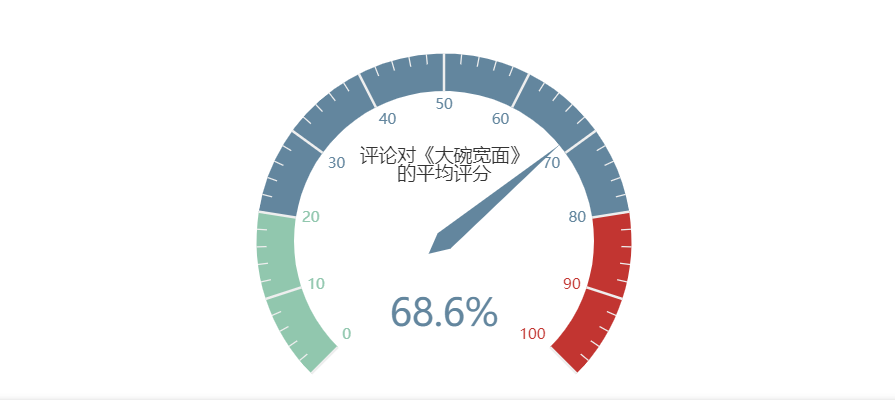


图3.5.1

3.6有多少人拿吴亦凡和蔡徐坤做对比

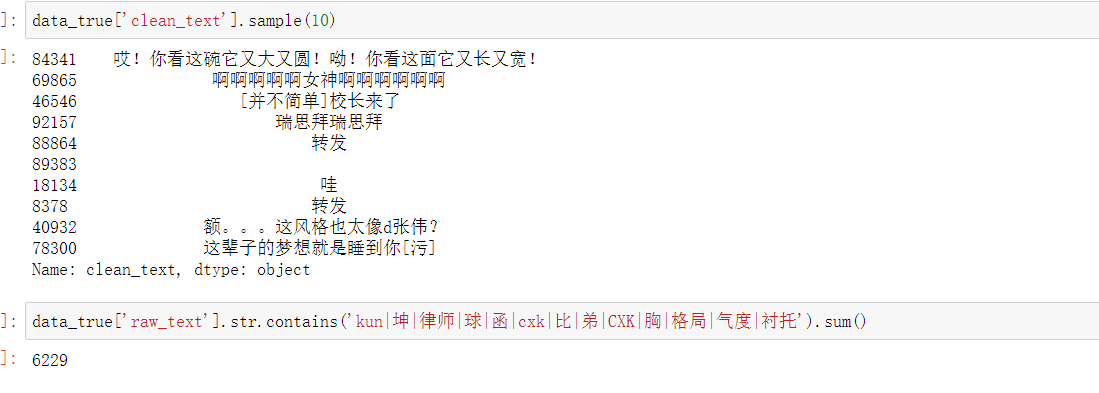


图3.6.1

3.7有多少人开始路转粉？

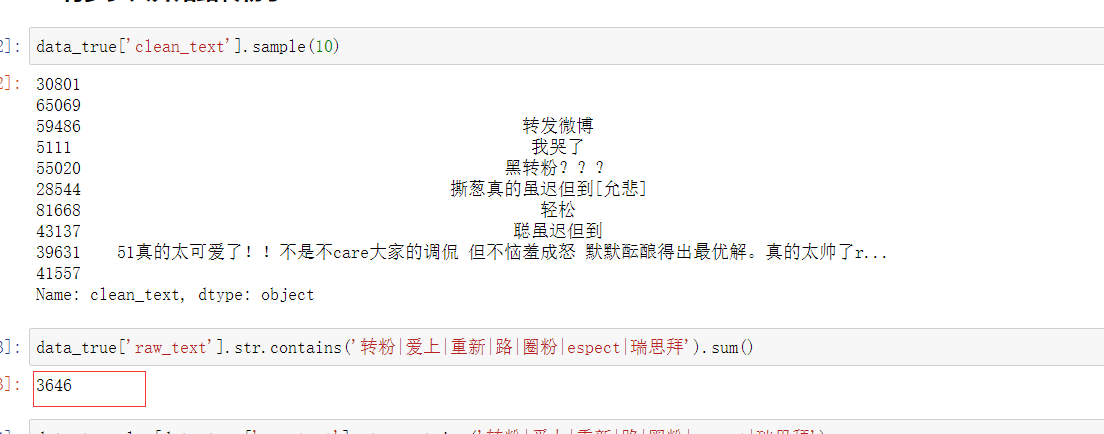


图3.7.1

3.7.1路转粉男女性别比例

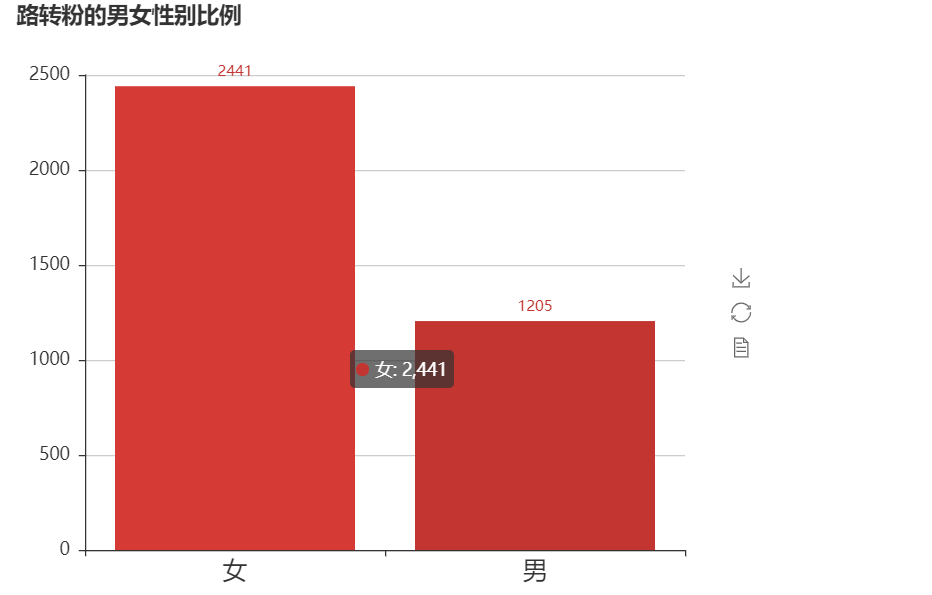


图3.7.1.1

3.8评论的词云图



图3.8.2

图3.8.1

4.遇到的问题

在调试代码时打开csv文件，数据中的空字符是以NaN的形式存在的，在筛选假粉丝的时候无法判断空字符串。经过调试发现只需在打开文件时将NaN转换为空字符串即可如下图所示。

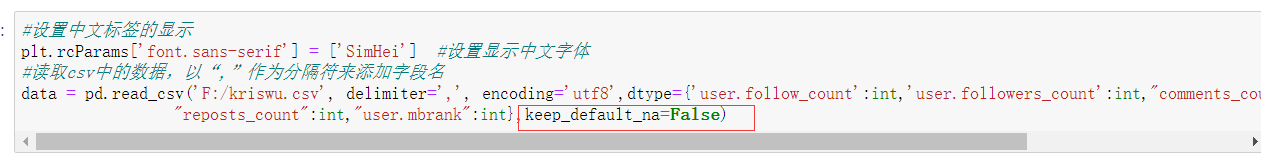


图4.1

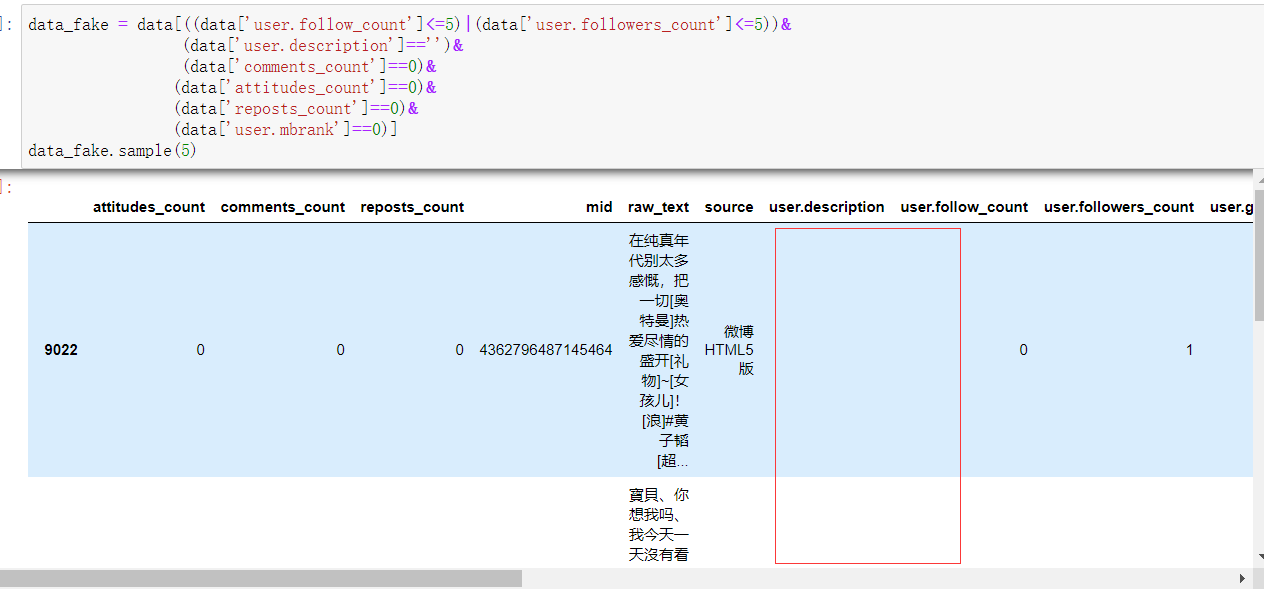


图4.2

5.问题汇总

1.吴亦凡该微博的转发是否存在假流量？

答：吴亦凡的微博确实存在假流量，一共转发了102118条，其中假粉丝转发6100条，占总数据的5.9%；真粉丝转发量96018条，占总数据的94%；真实粉丝转发量81872条，占总数据的80.17%。

2。大家对于《大碗宽面》怎么看？

答：《大碗宽面》的平均评分为68.6%，高于及格60%，大众应该是能接受的。

3。有多少人拿吴亦凡跟蔡徐坤做对比？

答：一共有6229个人拿吴亦凡和蔡徐坤做对比，占真粉丝转发量的6.4%

4。有多少人开始路转粉了？

答：一共有3646人路转粉，其中女生占2441人占总数据67%，男生1205人占总数据的33%。

5。评论的词云图

答：见图5-1

6.小组总结

这次Python课设让我们知道小组亦是团队、个人也许能发挥巨大作用，但团队能力更强。才开始我们在运行程序的时候遇到过许许多多在以前上课的时候没有遇到过得问题。通过小组讨论我们小组在网上查找资料一个脚步一个印的去解决，让我们学到了遇到问题如何解决并不是干着急！在课程设计的过程中，我们遇到了困难。也一齐分享了成功的喜悦。这次的课程设计对我们每个人来说都是一个挑战。课程设计中文档的撰写我们也有担心过，就是Python的数据处理这方面我们真的很担心，平时对这方面的知识接触的就不是很多，而且对于数据处理我们就更抓狂了。这时候小组的力量就体现出来了，各司其职，各尽其能。

7.个人总结

答：在这次课程设计中我学到了许多新的东西，同时加深了之前所学的知识，因为这种题型的课程设计是第一次做，稍微有些生疏，前期的困难还是比较多的，比如软件用的不太熟练，GitHub，jupyter等都是第一次使用，自己只能慢慢尝试摸索最终学会了使用，程序方面起初数据库没有调试好，之后选择了直接使用csv文件，代码方面经过自己不断的调试也完成了，不懂了就在百度上面找解决方法。总的来说这次课程设计收获良多，对自己有很大的帮助，对自身能力也提高了许多。这次课程设计也是圆满完成了，但是我仍然有很多方面有待提高在以后我也会更加认真的学习python这门语言。也感谢老师同学对我的帮助。

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】董付国.python可以这样学【M】.北京：清华大学出版社，2017.

【3】刘凌霞，吴海涛.21天学通python【M】.北京：电子工业出版社，2016.

**课程设计评定表**

|  |  |  |  |
| --- | --- | --- | --- |
| 课程设计题目 | 吴亦凡微博热点的分析 | | |
| 课程设计学生答辩或质疑记录：   1. 如何讲文件中的NAN转换为空字符串?   答：在打开csv文件的时候写上 keep\_default\_na=False代码即可。   1. 画柱状图的时候不显示如何解决？   答：由于ipython版本不兼容所以图像不显示，只需下载相对应的版本包即可。   1. 调用generic.py，为何发生python异常？   答：将generic.py里面的input()替换为raw\_input()，再进行调试！ | | | |
| **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 |  |
| 最终评定等级为：  指导老师签字：  2021 年 1 月 15日 | | | |