|  |  |
| --- | --- |
| 学号 | 10214518314 |



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-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 | 熊友旺 |
| 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。

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**

[摘要 6](#_Toc22)

[1. 分析目标 7](#_Toc7179)

[2.分析流程 7](#_Toc274)

[3.分析实施 7](#_Toc9528)

[3.1数据获取 7](#_Toc1346)

[3.2数据预处理 8](#_Toc27936)

[3.3 数据标准化 9](#_Toc5448)

[4.数据分析 9](#_Toc22138)

[4.1 关注996的程序员主要来自哪些公司 9](#_Toc6929)

[4.2 关注996的程序员主要来自哪些高校 10](#_Toc21812)

[4.3 关注996的程序员主要来自于哪些城市 11](#_Toc28100)

[4.4 关注996的程序员的平均粉丝数、关注数、仓库数 12](#_Toc29466)

[4.5 程序员关注的问题 14](#_Toc18996)

[5.思考 15](#_Toc21256)

[6.总结 15](#_Toc983)

[参考文献 16](#_Toc28281)

[课程设计成绩评定表 17](#_Toc23190)

# 摘要

在互联网技术不断升级的今天，数据获取变的越来越容易，大数据行业越来越火，基于海量数据进行分析的数据分析人才逐渐成为各大企业追逐的宠儿。大数据推动起数据的深度分析与挖掘，使得研究结果更加客观，并且多维。

实习是每个大学生必须拥有的一段经历，它使我们在实践中了解社会，让我们学到了很多在课堂上根本就学不到的知识，受益匪浅，也打开了视野，增长了见识，为我们以后进一步走向社会打下坚实的基础。996工作制，是指一种"早上9点上班，晚上9点下班，每周工作6天"的用工制度，有时也被用来指代一系列资方要求劳方延长工时而不额外给薪的工作制度。本文针对所提供的相关课题的数据进行996工作制的状况的分析。

# 分析目标

①关注996的程序员主要来自哪些公司

②关注996的程序员主要来自哪些高校

③关注996的程序员主要来自于哪些城市

④关注996的程序员的平均粉丝数、关注数、仓库数

⑤这些人关注的问题都有哪些

# 2.分析流程

①数据获取:找出需要用作数据分析的数据并进行获取。

②数据预处理:对获取数据进行去重、缺失值处理等一系列数据清洗操作。

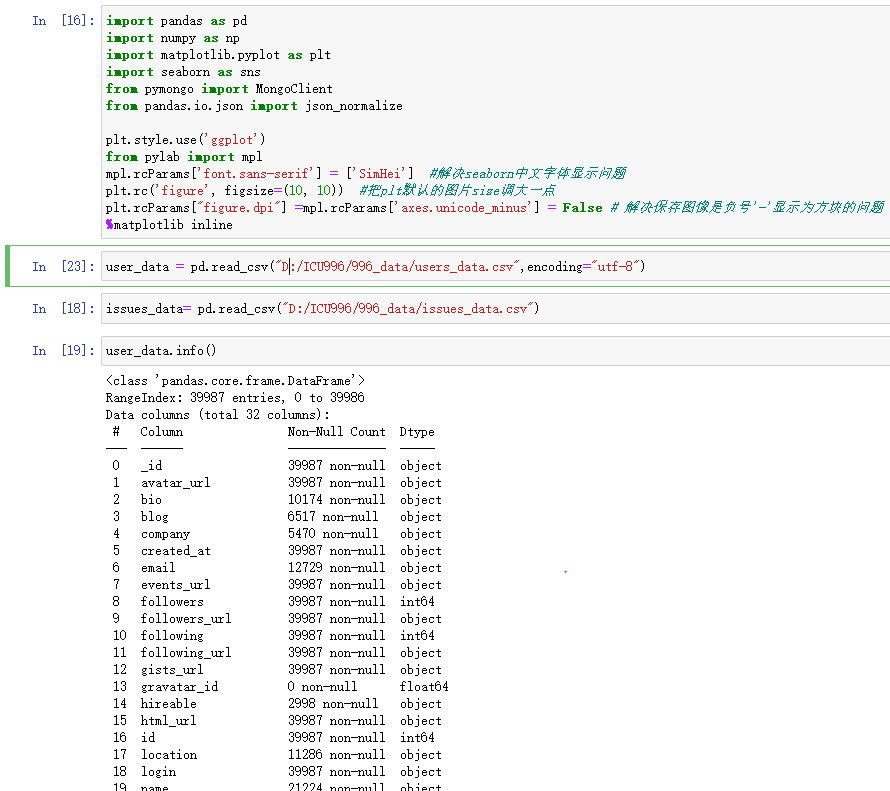
③数据可视化:文字不如数据，数据不如图片，可视化是对于数据最好的展现方式，让人一目了然。

④深入分析:对于数据深入分析，寻找不同因素之间与薪资之间的关系。

# 3.分析实施

## 3.1数据获取

由于数据老师已提供，这里仅放爬虫的代码，代码如下:



## 3.2数据预处理

对抓取完的数据，然后进行的是数据清洗。对user\_data.csv文件的数据清洗如图3.1所示，对issues\_data.csv文件的数据清洗如图3.2所示。

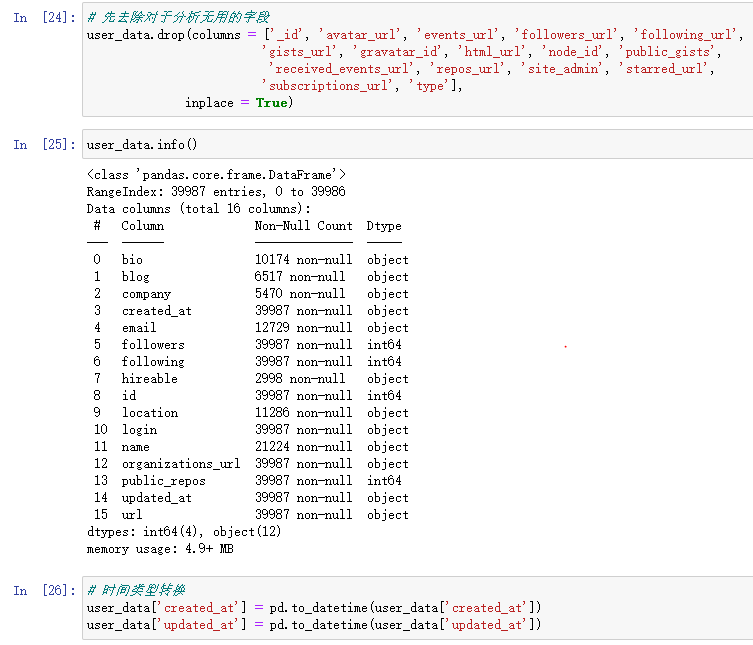


图3.1

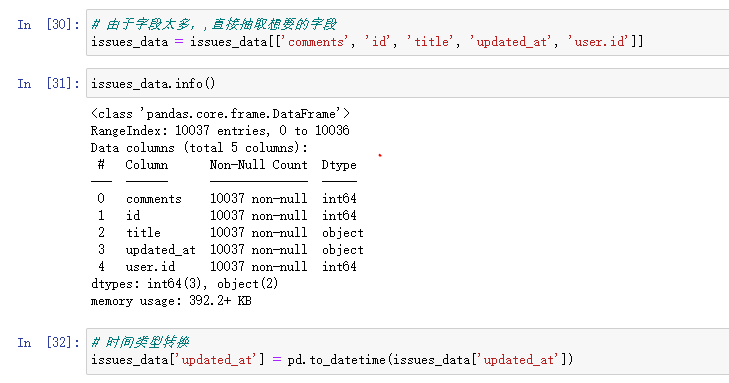


图3.2

## 3.3 数据标准化

标准化主要是对员工和所属公司、城市、学校等数据进行规范。所属公司、城市、学校抓取的结果是用“/”隔开的，所以先对其分列，在分别定义维度。

# 4.数据分析

## 4.1 关注996的程序员主要来自哪些公司

①通过Python代码进行统计人数排名最多的前50名公司，如图4.1

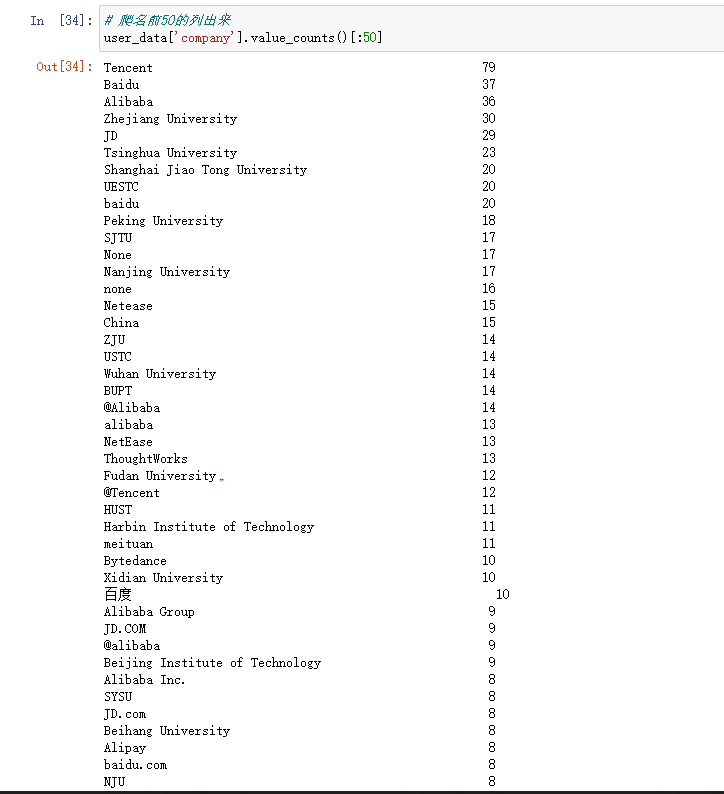


图4.1

②通过Python代码实现数据可视化，这里觉得选取50个有点多，我们这里取前10的数据进行可视化，代码如下,数据可视化之后如图4.2所示。

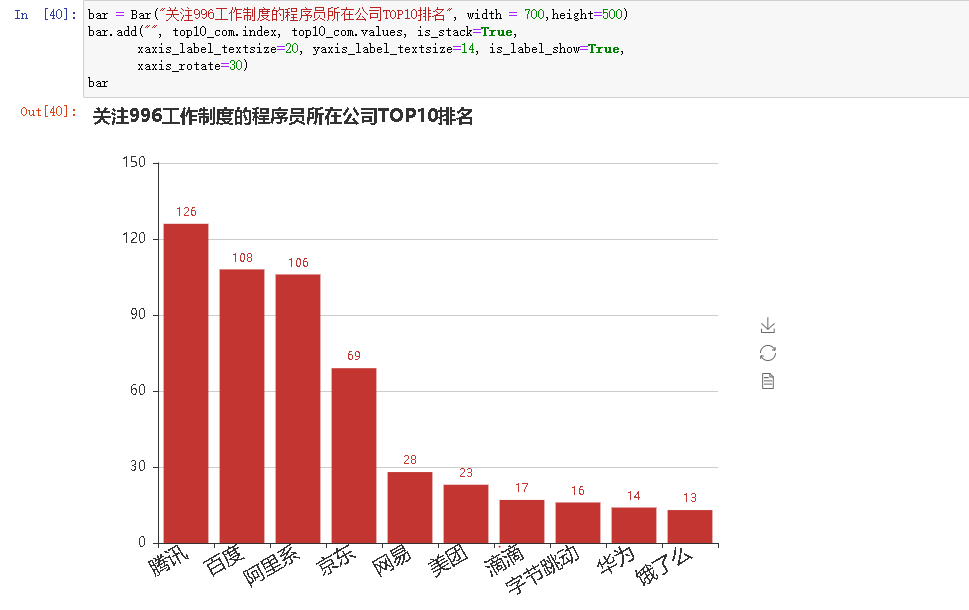


图4.2

结论: 从图中得知，腾讯、百度、阿里这三个公司的员工更加关注996工作制度。

## 4.2 关注996的程序员主要来自哪些高校

①通过Python代码进行统计关注996的程序员主要来自哪些高校，如图4.3所示。



图4.3

②通过Python代码实现数据可视化，我们这里取前10的数据进行可视化，代码如下,数据可视化之后如图4.4所示。

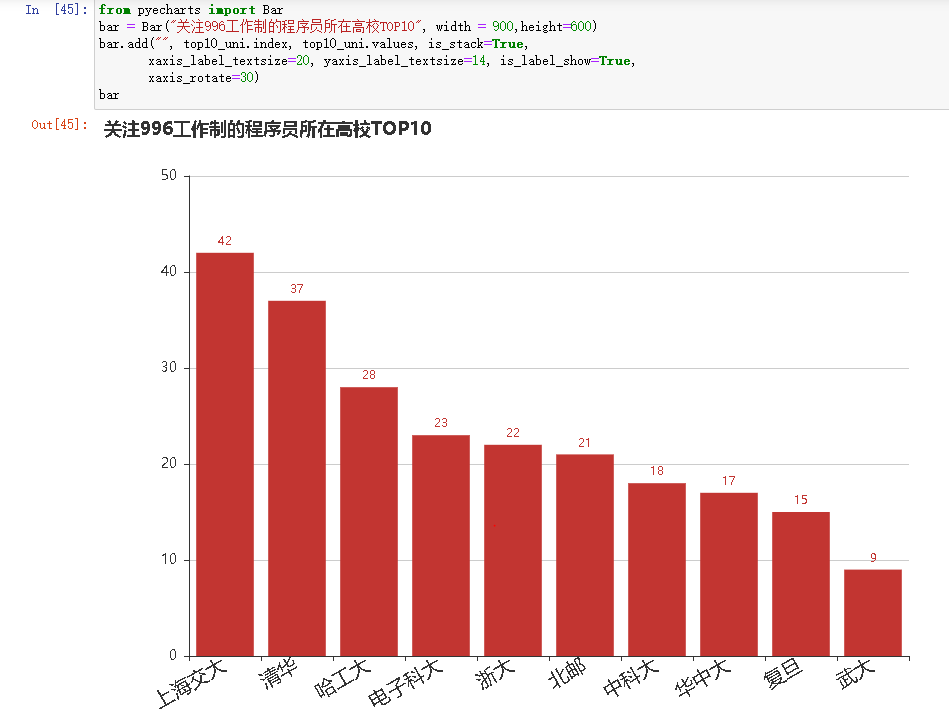


图4.4

结论: 上海交大、清华、哈工大占据了关注996工作制的程序员所在高校榜的前三。

## 4.3 关注996的程序员主要来自于哪些城市

①通过Python代码进行统计关注996的程序员主要来自哪些城市，如图4.5所示。

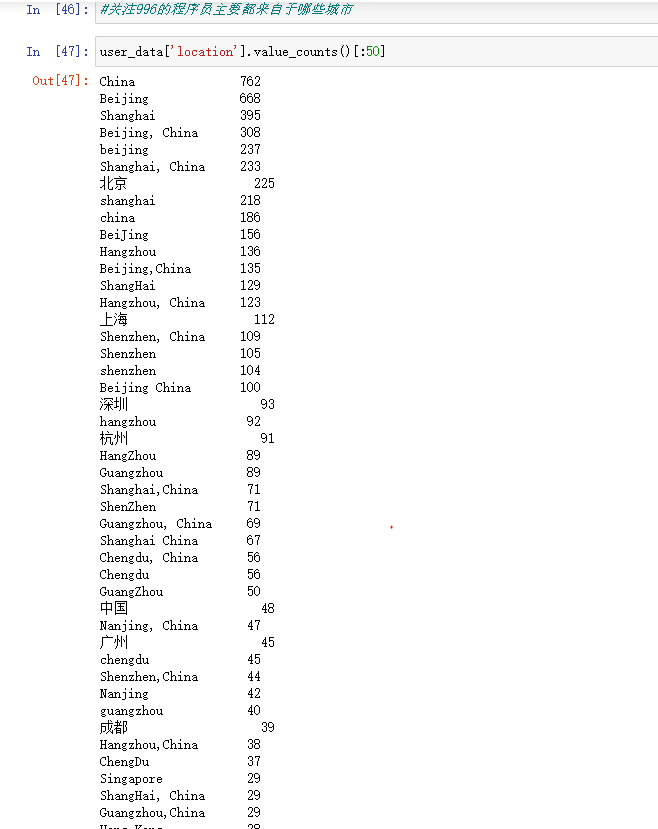


图4.5

②通过Python代码实现数据可视化，我们这里取前10的数据进行可视化，代码如下,数据可视化之后如图4.6所示。

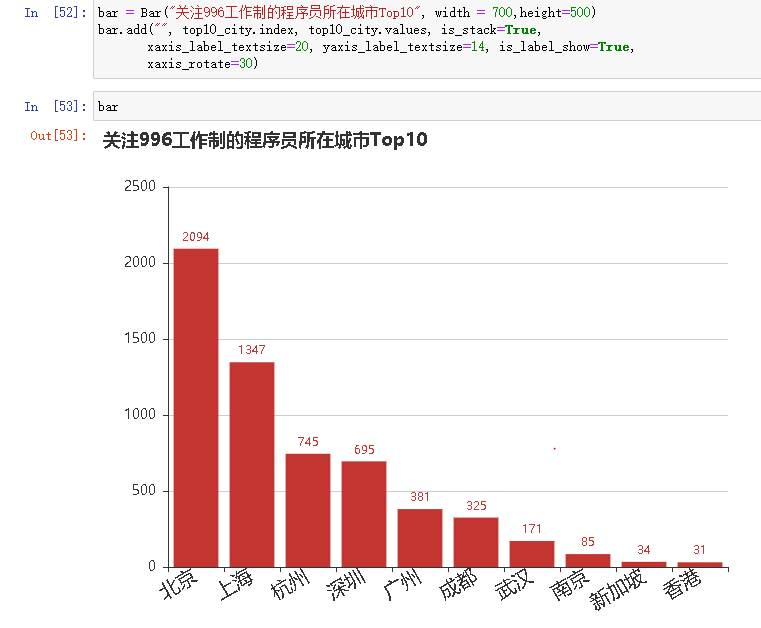


图4.6

结论: 北京、上海、杭州占据了关注996工作制的程序员所在城市榜的前三，其中北京已经超出上海近一半的人数。

## 4.4 关注996的程序员的平均粉丝数、关注数、仓库数

①通过Python代码进行统计关注996的程序员的平均粉丝数、关注数、仓库数，如图4.7所示。

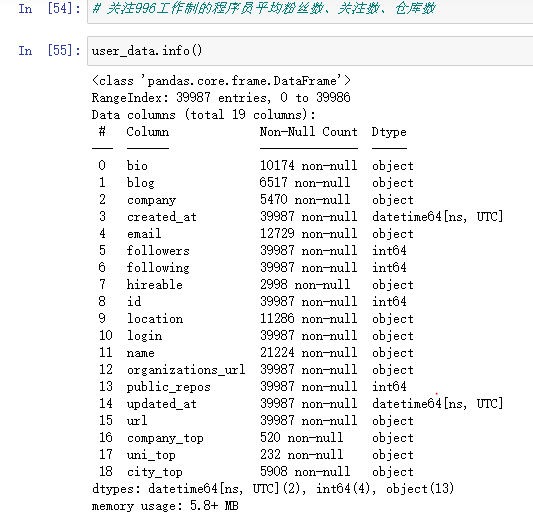


图4.7

②通过Python代码实现数据可视化，代码如下,数据可视化之后如图4.8所示。

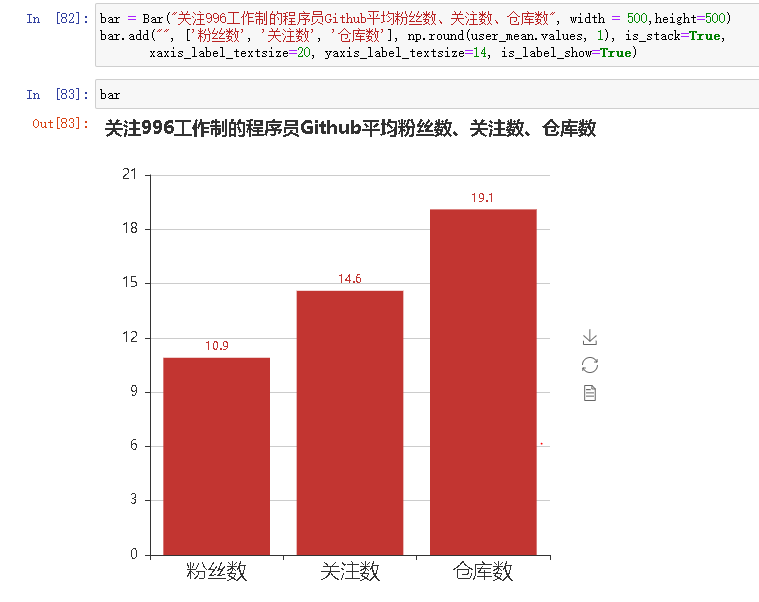


图4.8

## 4.5 程序员关注的问题

①通过Python代码进行统计程序员关注的问题有哪些，如图4.9所示。



图4.9

②通过Python代码实现对数据的云图可视化，代码如下,数据可视化之后如图4.10所示。

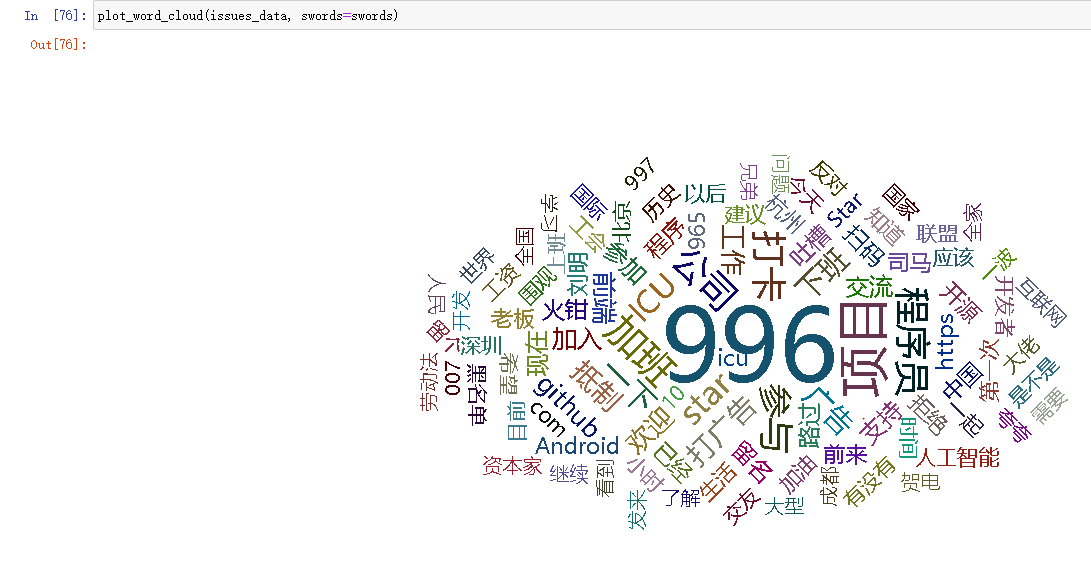


图4.10

# 5.思考

此次课程设计主要是对之前学习的Python语言的简单运用和可视化图表的实践，真实的数据分析远不止这些，还需要扎实的教学、统计学基础，良好的数据敏感度，开拓但严谨的思维等，是一个终生学习的过程。

# 6.总结

进行本次Python课设的过程中，查阅了许多资料，在徐勇老师和同学的提点下，完成了996工作制的数据分析。在这个过程中，我学到了很多，如怎样去全面的看待并分析一个问题等，同时也在总体上巩固了自己的Python知识，发现自己之前从来没有像现在一样将零散的知识点结合在一起，完成了对996工作制的数据分析后，让我在整体上又认识了一次Python。

在这次的996工作制的数据分析中，有了平时学习的基础，完成本次Python课设的效率也因此大大提高，在设计过程中虽然遇到了一些问题，但经过多次的思考及反复的实践，检查出问题可能出现的原因，从而也暴露出了我在这方面的知识的欠缺和经验的不足，这使我意识到，自己学习还是只停留在表面，所以，我将努力在今后的时间里改正，并将所学的知识吃透、弄懂。

另外 ，通过这次课程设计我对996工作制有了更深的了解，首先，不可否认实行996可以增加工作时长，有利于企业利益最大化。但同时也应该考虑到员工的生活问题及情绪问题。其次，生活在这个时代我们是应该去奋斗，去拼搏。但奋斗≠996，为自己奋斗≠为公司奋斗。对于我们快毕业的大学生来说，本来是充满斗志，充满希望的，想着通过自己的奋斗干出一番事业。但强制996很容易使年轻人陷入成长性低，既繁琐又疲劳的日常工作中，没有时间自我学习提升，甚至陷入一种上班已经很努力了，下班只想放松的状态，这样只会蹉跎岁月。我不否认996工作制有其自身优点，不否定有人认可996工作制。但在我看来，996工作制弊大于利。

Python课设的不足望老师谅解，今后我会更加努力地巩固所学的知识，希望自己能够学以致用，不辜负老师的教导。

# 参考文献

[1]刘宇宙.Python 3.5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，郝宁波，吴海涛，21天学通Python[M].北京:电子工业出版社，2016

[4] 基于Python的数据可视化[J]. 罗博炜.  信息记录材料. 2019.12

[5] 基于Python的招聘网站信息爬取与数据分析[J]. 王芳.  信息技术与网络安全. 2019.08

**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-针对996工作，程序员群体的看法的分析 | | |
| 课程设计学生答辩或质疑记录：  ①数据处理的基本流程？  答：数据处理流程主要包括数据收集、数据预处理、数据存储、数据处理与分析、数据展示/数据可视化、数据应用等环节。  ②怎么实现数据可视化？  答：利用 Python 及 matplotlib 和 seaborn 库制作出多种多样的可视化图形。  ③课程设计的过程中遇到哪些问题，怎么解决？  答：刚开始在做数据爬取的时候总是会出现爬取不到数据的情况，后来检查发现是 文件路径的问题，容易把后缀名.csv弄掉，把后缀名加上就解决了这个问题。 | | | |
| **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、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 | | | |