|  |  |
| --- | --- |
| 学号 | 10212818111 |



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-地震的数据分析 |
| 专 业 | \_\_\_\_\_\_\_\_\_\_\_软件工程\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 班 级 | \_\_\_\_\_\_\_\_\_\_\_软件1181\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 姓 名 | \_\_\_\_\_\_ \_ \_唐国旗\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ |
| 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

# 

# 分析目的

对这次实习岗位数据分析的目的是，一方面查看各地区的公司对应届毕业生的招聘情况，还有个地区的工资待遇，公司福利等。分析这些数据一方面是发现毕业生就业问题，并且找到问题的根源，最终通过切实可行的方法解决存在的问题，另一方面通过对这些数据的分析，可以总结出目前毕业生面临的就业环境是怎么样的。通过对这些数据的总结就能知道毕业生现在的就业问题，

# 1开发环境与技术简介

## 1.1硬件平台

硬件平台使用的是64位计算机一台

## 1.2 GitHub网站

github是通过Git进行版本控制的软件源代码托管服务平台，于2008年4月10日正式上线。

github除了Git代码仓库托管及基本的Web管理界面以外，它还提供了一些方便社会化共同软件开发的功能，即一般人口中的社群功能，包括允许用户追踪其他用户、组织、软件库的动态软件代码的改动和bug提出评论等。

GitHub可以托管各种git库，并提供一个web界面，但与其它像或 Google Code这样的服务不同，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。

是的，我们正是这么计划的。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。

在GitHub，用户可以十分轻易地找到海量的开源代码。

## 1.3jupyter运行平台

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](https://baike.baidu.com/item/%20%E5%BA%94%E7%94%A8%E7%A8%8B%E5%BA%8F/5985445)，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](https://baike.baidu.com/item/markdown/3245829)。 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。

## 

## 1.4编程语言-python

Python是一种计算机程序设计语言，由吉多·范罗苏姆创造，第一版发布于1991年，可以视之为一种改良的LISP。Python的设计哲学强调代码的可读性和简洁的语法。相比于C++或Java，Python让开发者能够用更少的代码表达想法。

自从20世纪初Python语言诞生至今，它已被逐渐广泛应用于系统管理任务的处理和Web编程。

Python的创始人为Guido van Rossum。1989年圣诞节期间，在阿姆斯特丹，Guido为了打发圣诞节的无趣，决心开发一个新的脚本解释程序，作为ABC 语言的一种继承。之所以选中Python（大蟒蛇的意思）作为该编程语言的名字，是取自英国20世纪70年代首播的电视喜剧《蒙提.派森干的飞行马戏团》（Monty Python's Flying Circus）。

ABC是由Guido参加设计的一种教学语言。就Guido本人看来，ABC 这种语言非常优美和强大，是专门为非专业程序员设计的。但是ABC语言并没有成功，究其原因，Guido 认为是其非开放造成的。Guido 决心在Python 中避免这一错误。同时，他还想实现在ABC 中闪现过但未曾实现的东西。

就这样，Python在Guido手中诞生了。可以说，Python是从ABC发展起来，主要受到了Modula-3（另一种相当优美且强大的语言，为小型团体所设计的）的影响。并且结合了Unix shell和C的习惯。

Python[[2]](https://baike.sogou.com/v58828.htm?fromTitle=Python#quote2)已经成为最受欢迎的程序设计语言之一。自从2004年以后，python的使用率呈线性增长。2011年1月，它被TIOBE编程语言排行榜评为2010年度语言。

由于Python语言的简洁性、易读性以及可扩展性，在国外用Python做科学计算的研究机构日益增多，一些知名大学已经采用Python来教授程序设计课程。例如卡耐基梅隆大学的编程基础、麻省理工学院的计算机科学及编程导论就使用Python语言讲授。众多开源的科学计算软件包都提供了Python的调用接口，例如著名的计算机视觉库OpenCV、三维可视化库VTK、医学图像处理库ITK。而Python专用的科学计算扩展库就更多了，例如如下3个十分经典的科学计算扩展库：NumPy、SciPy和matplotlib，它们分别为Python提供了快速数组处理、数值运算以及绘图功能。因此Python语言及其众多的扩展库所构成的开发环境十分适合工程技术、科研人员处理实验数据、制作图表，甚至开发科学计算应用程序。

2018年3月，该语言作者在邮件列表上宣布Python 2.7将于2020年1月1日终止支持。用户如果想要在这个日期之后继续得到与Python 2.7有关的支持，则需要付费给商业供应商。

# 2程序设计

## 2.1数据分析库

### 2.1.1数据分析库导入见下图2.1-1

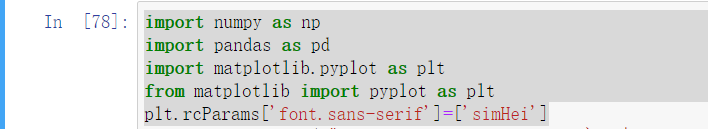


图2.1-1

### 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bai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 libraty。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数据表

### 2.2.1csv文件的读取2.2-1

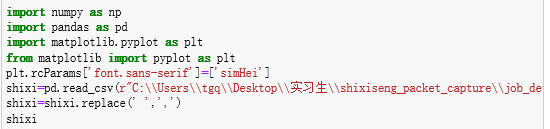


图2.2-1

## 2.3数据清洗

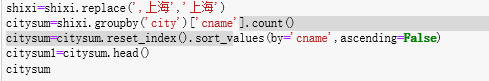
### 2.3.1数据清洗代码见下图2.1-1





## 2.4数据分析

### 2.4.1数据提取代码见下图2.4-1



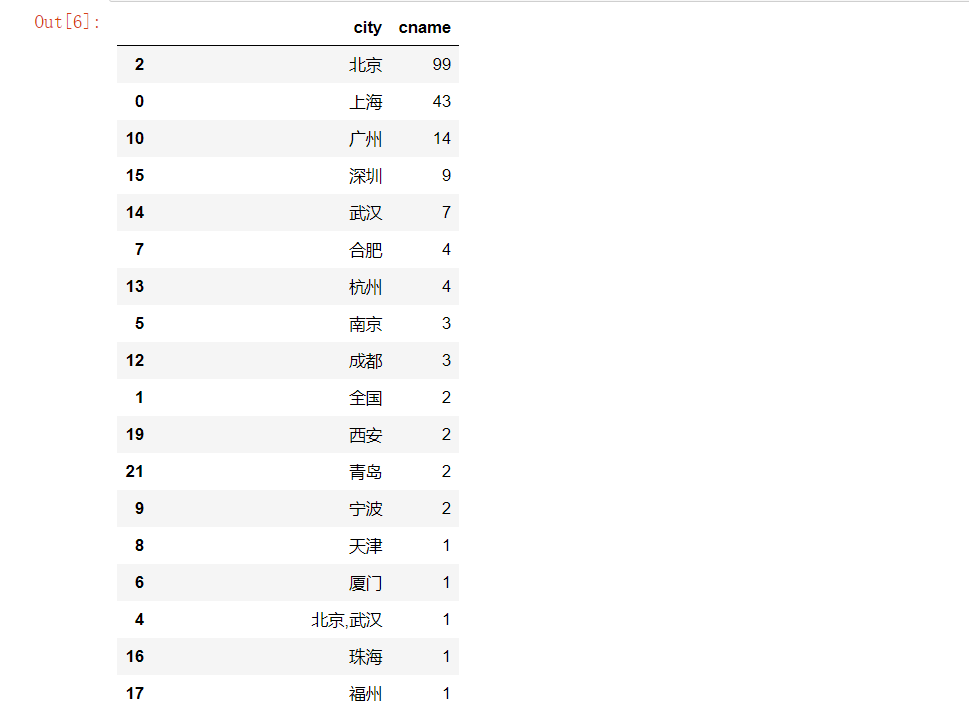


图2.4-1

### 2.4.2生成分析图见下图2.4-2

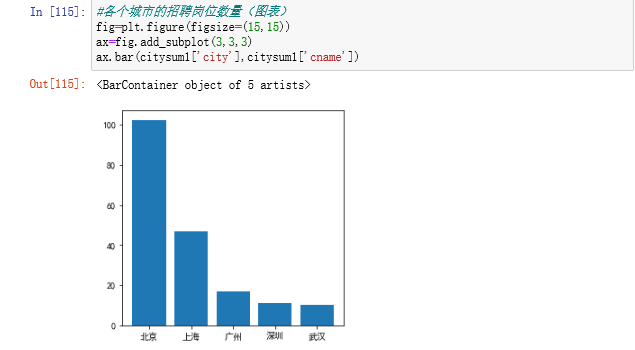


图2.4-2

## 2.5 统计全国招聘岗位工作天数

### 2.5.1统计图生成代码见下图2.5-1

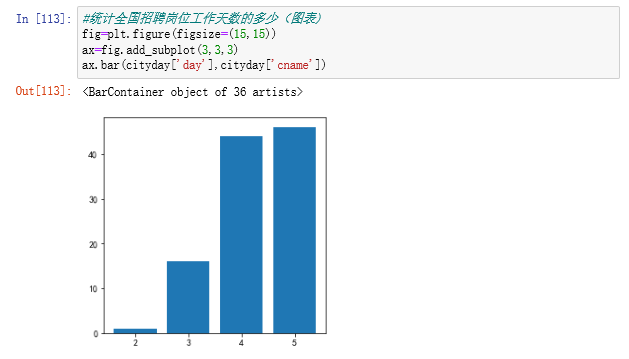


图2.5-1

## 2.6 全国岗位招聘要求统计图

### 2.6.1统计图生成代码见下图2.6-1

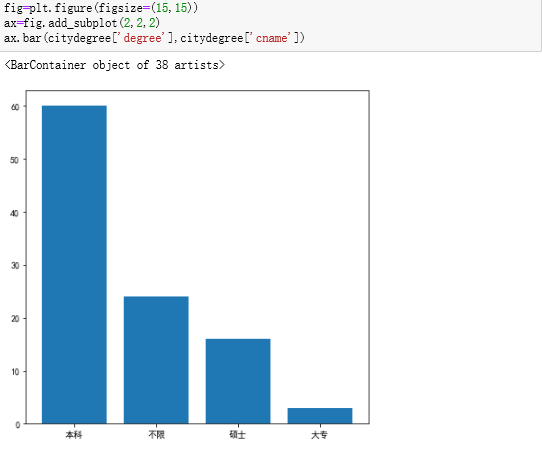


图2.6-1

## 2.7 各个城市岗位的平均最高和最低工资统计图

### 2.7.1统计图生成代码见下图2.7-1

IMG_256

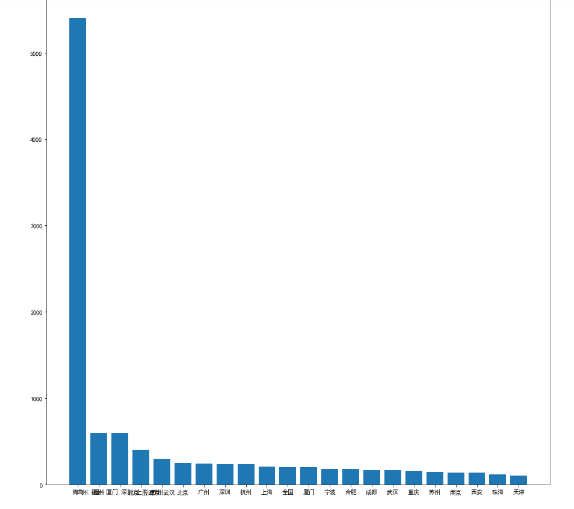


图2.7-1

IMG_256

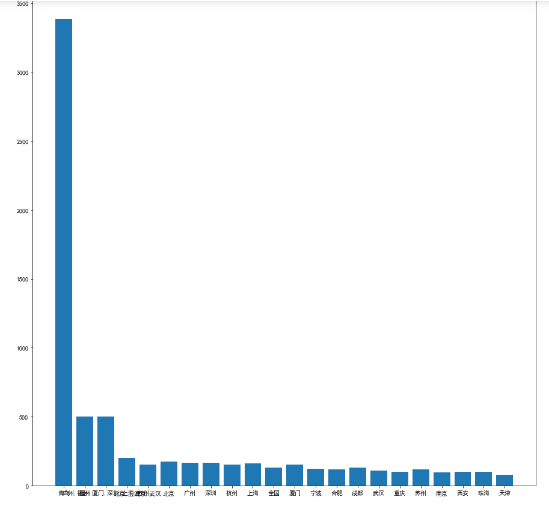


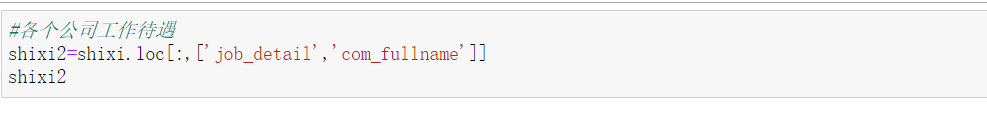
图2.7-2

## 2.8公司注册资本





## 2.9各个公司工作待遇





# 3数据分析

## 3.1引入问题

对于非北上广的普通院校应届生申请实习，相关单位会提供机会吗?相关网络言论见图3.1-1

这是网络上的回答，第一眼看的不是能力，而是做事要脚踏实地。

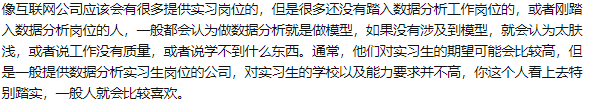


图3.1-1

## 3.2数图分析

我们先看各地方的岗位实习职位数量，见图3.2-1

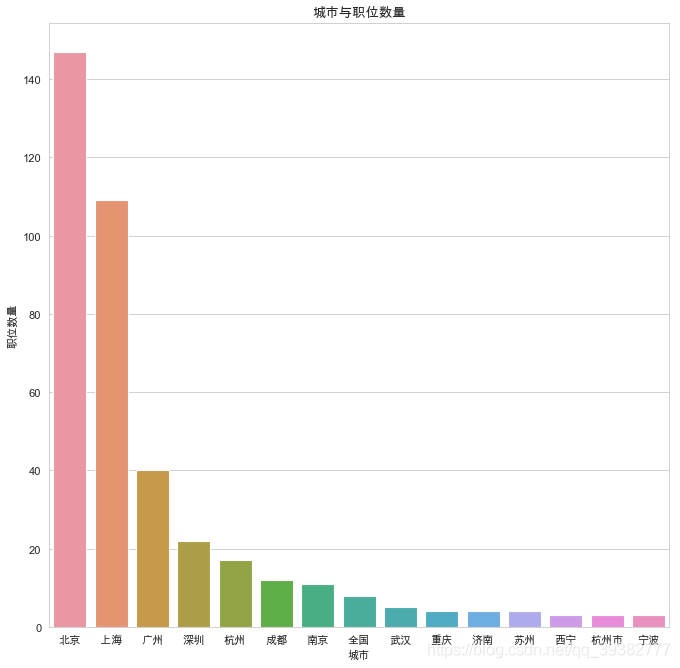
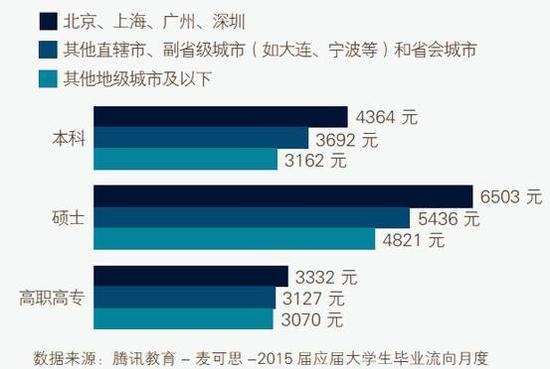


图3.2-1

可以看到，北上广深等一线城市职位数量遥遥领先，而后面的城市职位数量就非常少，从图也可发现去大城市发展的人很多，竞争也很激烈。

以下是北上广深学历和工资平均对比图。



可见， 一线城市的职位平均工资要高一些，但是个城市生活水平质量也不同，每个人的选择也不同，但是大多数还是会选择一线城市。

各年级大学生关于实习的态度

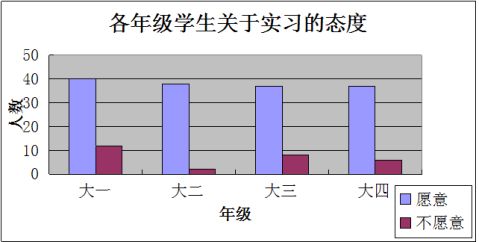


图3.2-3

绝大部分在校大学生希望参加实习。从调查数据中可以看出，对实习持赞成态度的学生占到了总数的84%，说明大部分本科生还是比较向往参加实习工作的。但是，各个年级的比例却有一些相对的差距。大一学生可能由于刚刚步入大学，对周边环境还不了解，赞成率仅为76.9%。大二学生对环境已逐渐熟悉，所以赞成率达到95.0%。而大三学生由于考研压力，反对率达到了18.8%。而大四学生的未来前景选择各异，赞成率为86.1%。综合上述，首先可以看出大部分学生倾向于参加实习工作，向往体验职场生活；其次可以看出，周边环境对学生态度有所影响；最后，对实习的态度产生影响的因素较多，因人而异。

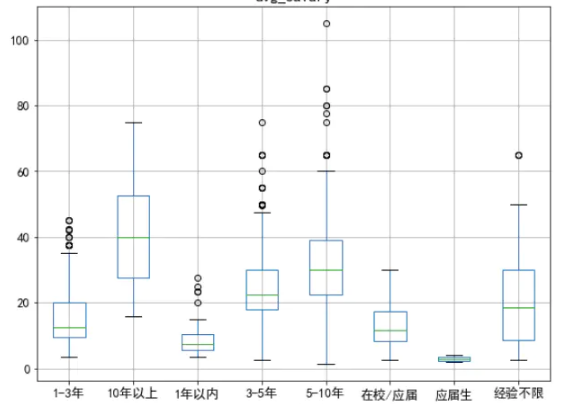


图3.2-4

工作经验上看，薪资的差距拉开的很大，应届生和有工作经验的

薪资对比，不在一个维度上，但这恰恰说明数据类岗位是可以通过个人能力的提高来获取更大的收益的，它是可成长的。你对技术、业务越熟悉，你的机会就越多，薪资就越高。

## 4 总结

　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。这次课程设计我做的最多的还是代码的编写，在代码编写的过程中，也遇到了许多问题，比如：原始数据的格式很乱处理，最后通过思考和网上查询资料解决了这些问题，我觉得思考问题的时候是最考验一个人的耐心与严谨态度的过程，仅有这样才能使分析的数据更准确、更美观。

　　二、经过课程设计，使我加深了对所学理论知识的理解与巩固，并能将课本上的纯理论应用到实践中，进一步加深了对知识的认识。同时，也有助于对其他知识的理解。此刻，我不但对书本上的知识更深入的理解，并且还学会了一些课本上没有的知识。

# 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.

|  |  |  |  |
| --- | --- | --- | --- |
| 课程设计题目 | 数据分析的学习与实践-实习岗位状况 | | |
| 程设计学生答辩或质疑记录：  Csv文件是怎么读取的？  答：利用read-csv（）函数对文件进行读取.  各个城市的岗位数量是怎么计算的？  答：利用groupby（）函数对每个城市按照名称进行计数  每一列的数据是怎么提取的？  答：利用loc（）函数对每一列进行提取 | | | |
| **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 | | | |

**课程设计成绩评定表**