|  |  |
| --- | --- |
| 学号 | 10214518320 |



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-实习岗位状况的数据分析 |
| 专 业 | 软件工程 |
| 班 级 | 软件1182班 |
| 姓 名 | 于明轩 |
| 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.概要 5](#_Toc15057)

[1.1课题的背景 1](#_Toc7225)

[1.2课题研究的目的和意义 1](#_Toc29178)

[2.运行环境 1](#_Toc13947)

[3.分析目标 1](#_Toc15516)

[3.1命题预测 2](#_Toc129)

[3.2确定目标 2](#_Toc32359)

[4.内容分析 2](#_Toc673)

[5.信息读取 2](#_Toc11297)

[6.数据的清洗和处理 4](#_Toc15687)

[6.1数据清洗 4](#_Toc4181)

[6.2数据修改优化 5](#_Toc13659)

[7.数据分析 6](#_Toc5807)

[7.1探究目标 6](#_Toc16560)

[7.2学历与岗位数量 6](#_Toc23331)

[7.3工资与岗位的分析 9](#_Toc31509)

[7.4公司规模与岗位数量 12](#_Toc25417)

[7.5城市与岗位数量 15](#_Toc20672)

[8.总结 18](#_Toc23443)

9.[参考文献 19](#_Toc12579)

1.概要

## 1.1课题的背景

随着科学技术的飞速发展，对于当代大学生的要求和标准也越来越高。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。通过本次课程设计对实习岗位状况的数据分析，通过分析得出不同城市的岗位需求量，不同城市能够提供的薪资水平，大中小型企业对数据分析岗位的需求情况等关系。

## 1.2课题研究的目的和意义

通过实习岗位状况的分析，可以更好掌握到实习岗位的相关信息，同时对于大学毕业前夕就面临找工作和实习就业的选择，本次python课设可以对实习岗位进行大概的了解和分析，更好的为之后的实习就业做准备。数据的分析和处理可以明确的凸显实习岗位受影响的条件和因素，让我们尽量做出正确的选择。

# 2.运行环境

硬件环境：PC计算机，微机系列，内存在1G以上

软件环境：jupyter

# 

# 3.分析目标

对题目和目标进行分析和假设，筛选出相关条件和可能影响的因素。

## 3.1命题预测

①学历

身边人常说学历越高越好找工作，就越能找到好的工作岗位

②城市

城市越大越繁华，实习的岗位和机会就可能会更多

③薪资水平

薪资水平可能会反映实习岗位的重要程度和抢手程度

④公司规模

公司规模和公司底蕴文化可能会直接影响实习里岗位的前景和作用

## 3.2确定目标

①学历与岗位需求量

②城市与岗位需求量

③薪资水平分析

④公司规模分析

# 4.内容分析

①读取数据：首先读取全部数据并去掉重复

②筛选数据：获取需要的数据并进行去空处理

③分析数据：对提取的有用信息进行统计分析

④得出结论：通过分析得出相应结论并实现数据可视化

# 5.信息读取

读取csv文件中的信息，将信息设置为中文标签显示，通过文件中的类型读取job\_list.csv中有效字段并显示出来，如下图5.1所示





图5.1数据提取

# 6.数据的清洗和处理

## 6.1数据清洗

①对抓取完的数据，首先进行的是数据清洗。然后进行重复项删除，最后只留下唯一的数据。对job\_detailed.csv文件的数据清洗如图6.1所示



图6.1数据清洗

## 

## 6.2数据修改优化

②查询数据是否存在缺失项，同时为了对错误数据进行修改和优化，让数据更加直观准确，特将CSV文件导入数据库，通过数据库直接将数据进行修改和优化，最后再进行读取。

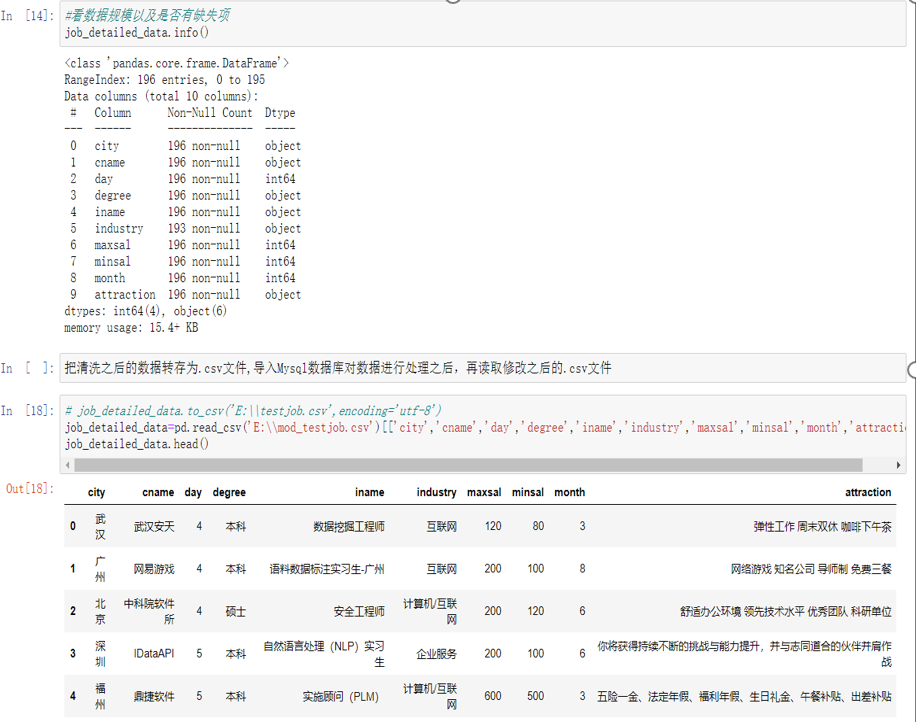


图6.2数据处理

# 7.数据分析

## 7.1探究目标

①学历与岗位数量

探究岗位与相关学历是否有要求或联系

②城市与岗位数量

验证城市与岗位数量的关系

③公司规模与岗位需求量

探究公司规模是否影响岗位需求量

④不同城市的岗位工资关系

数据分析不同岗位的工资区别

## 7.2学历与岗位数量

①统计学历符合岗位的数量

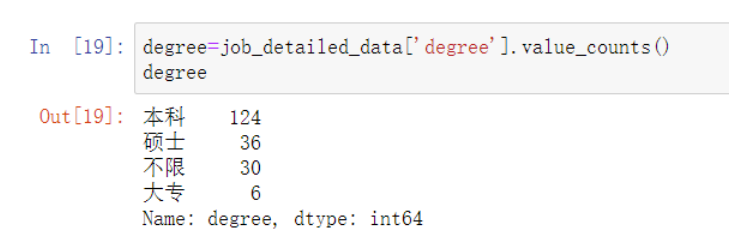


图7.2.1不同学历岗位数量图

②用python实现岗位数量和学历的可视化，柱状图如下图7.2.1所示

plt.figure(figsize=(10, 10), dpi=80)

x = degree[:15].index

y = degree[:15].values

p = plt.bar(x, y, label="num")

for x,y in zip(x,y):

plt.text(x,y, '%.0f'%y, ha='center',va='bottom')

plt.xlabel('学历')

plt.ylabel('岗位数量')

plt.title('学历与岗位数量')

plt.legend(loc="upper right")

plt.show()

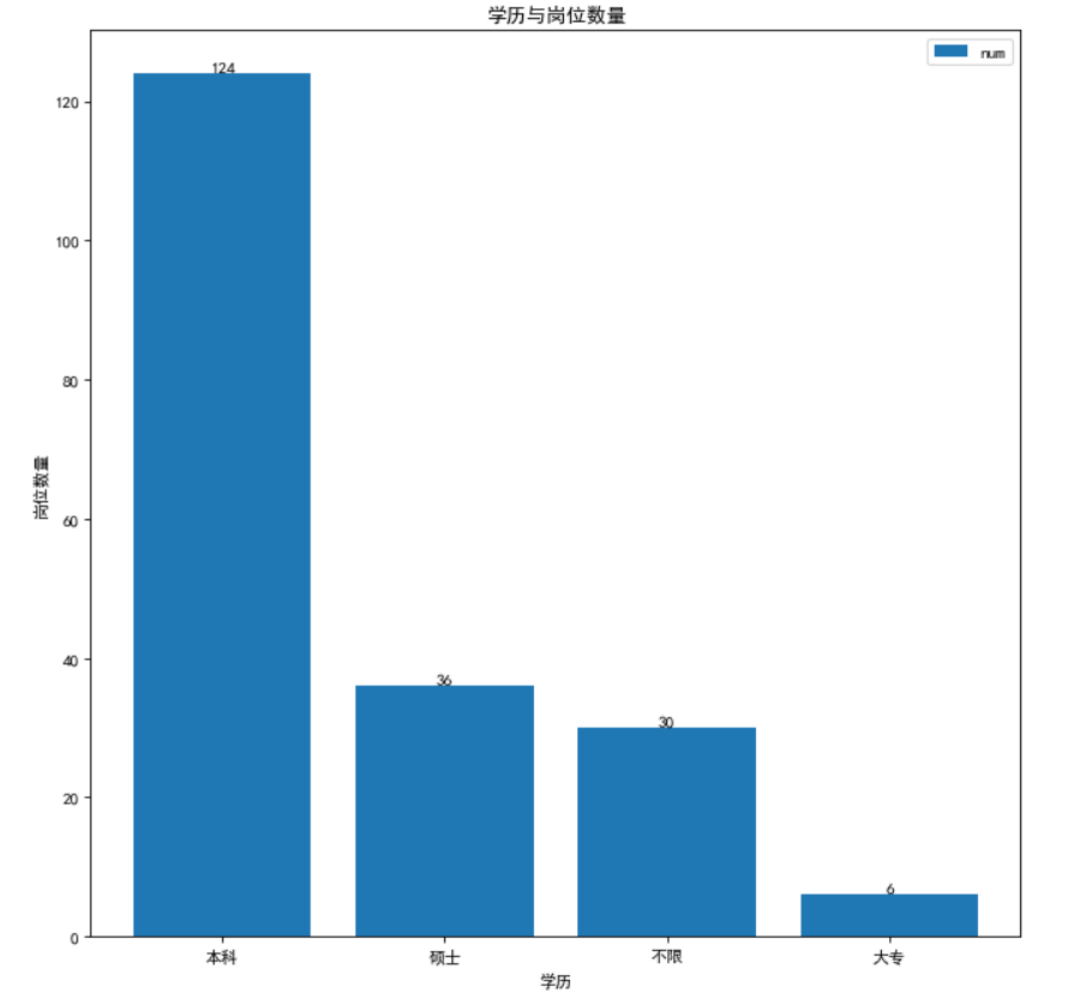


图7.2.2学历与岗位数量柱状图

③将柱状图用不同颜色表示，提高图形可视化，如下图7.2.3所示

plt.figure(figsize=(10, 10), dpi=80)

x = degree[:15].index

y = degree[:15].values

sns.barplot(x=x, y=y, data=job\_detailed\_data,capsize=.05)

plt.xlabel('学历')

plt.ylabel('岗位数量')

plt.title('学历与岗位数量')

plt.show()

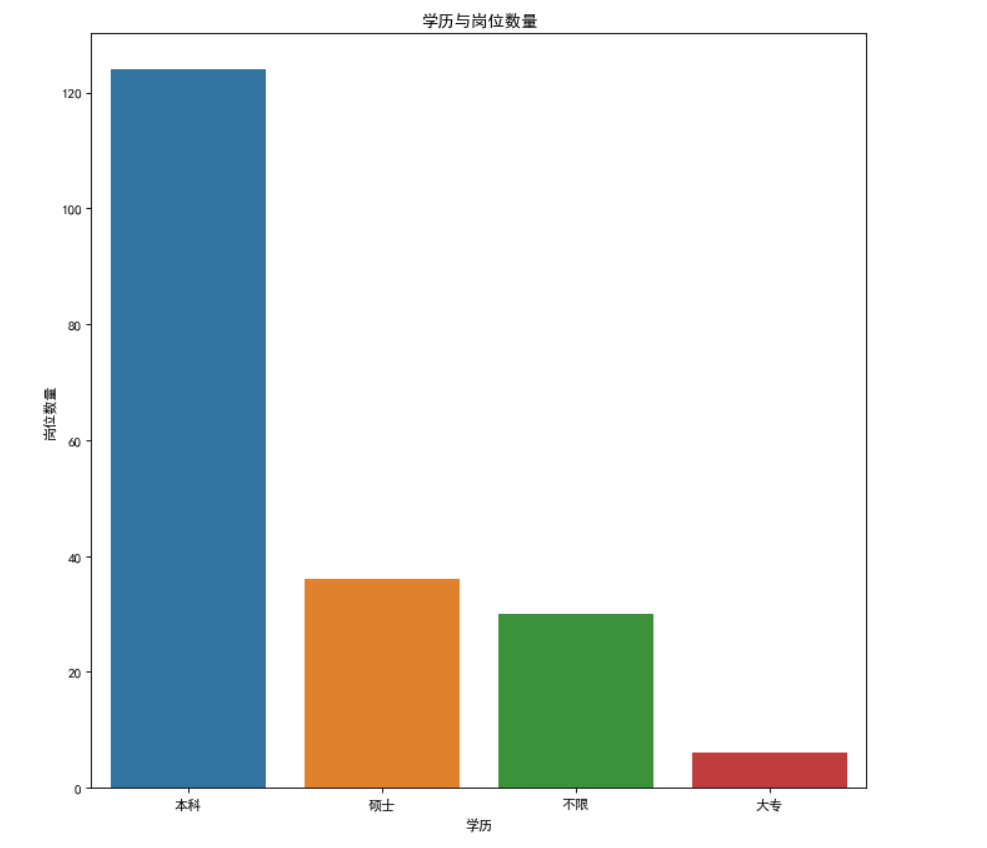


图7.2.3不同颜色学历与岗位数量柱状图

结论:岗位数量与学历有关，且岗位学历要求基本为本科及以上，学历越高越多合适岗位，本科以下专科岗位较少

## 7.3工资与岗位的分析

①不同城市岗位的每日最高最低工资



图7.3.1各地岗位单日最高、最低工资

②求平均日工资，更好的做对比



图7.3.2均日工资

③不同城市的月薪工资

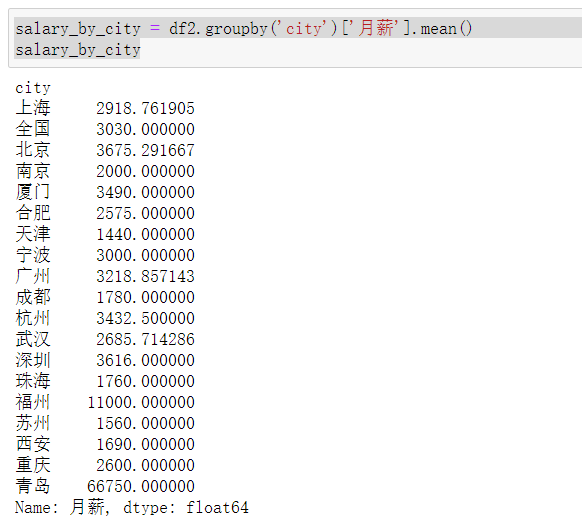


图7.3.3不同城市月薪

④不同城市岗位数量所占比



图7.3.4岗位数量所占比

⑤通过降序将城市月薪按从高到低排序

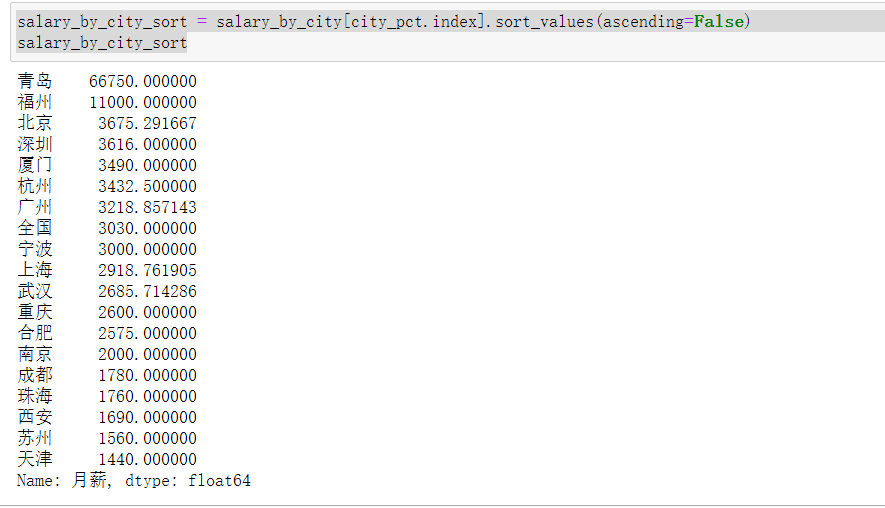


图7.3.5月薪排序

⑥城市与工资可视图

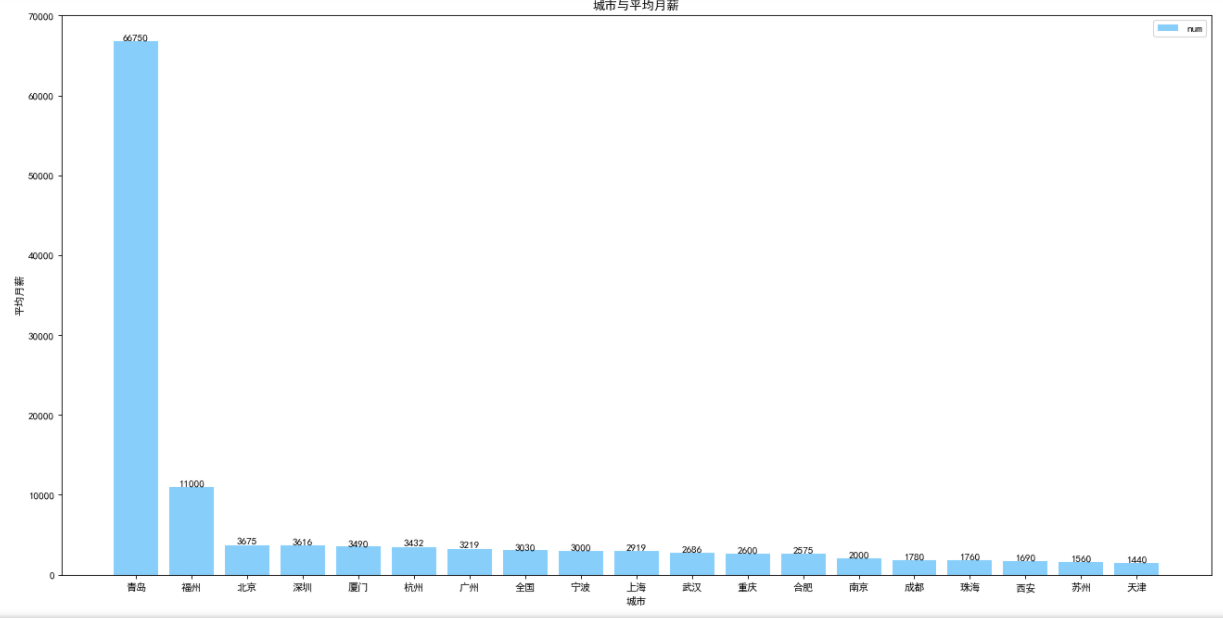


图7.3.6城市平均月薪柱状图

结论：不同城市的岗位薪资不同，且实习岗位和工资存在紧密关系。

## 7.4公司规模与岗位数量

①公司基本信息



图7.4.1公司信息

②统计不同规模大小公司数量

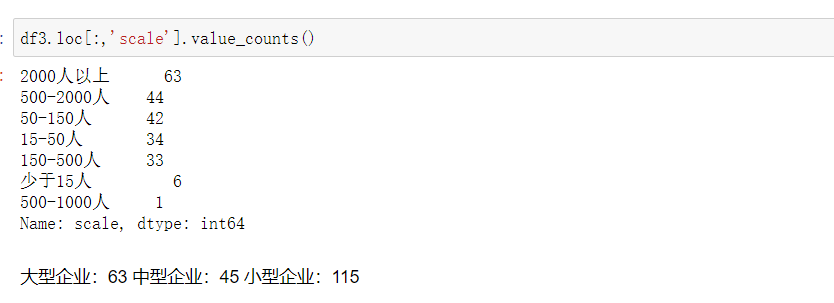


图7.4.2公司规模数量

③公司规模与岗位饼状图



图7.4.3饼状图

④将饼状图转化为环形可视图





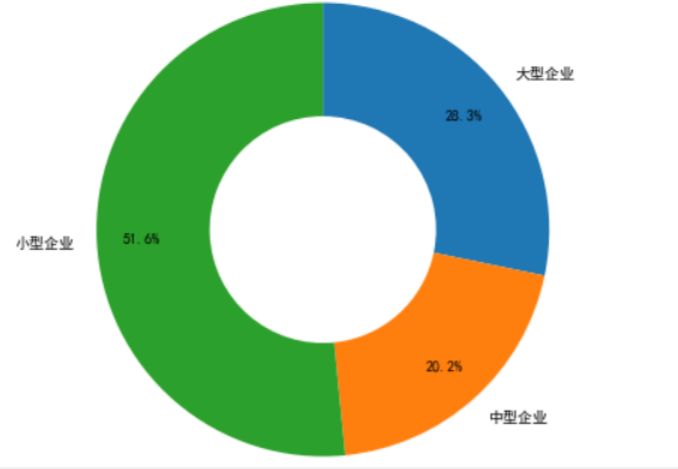


图7.4.4环形可视图

## 7.5城市与岗位数量

①统计城市与岗位数量



图7.5.1城市岗位图

②城市与岗位数量柱状图



图7.5.2运行代码

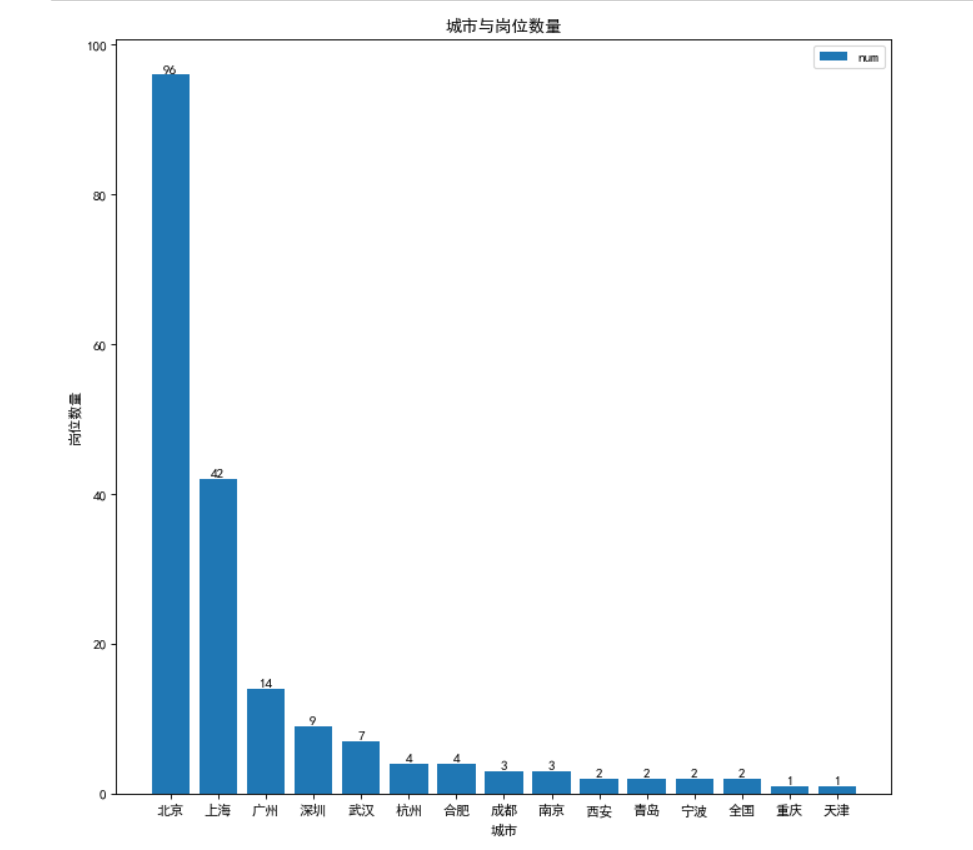
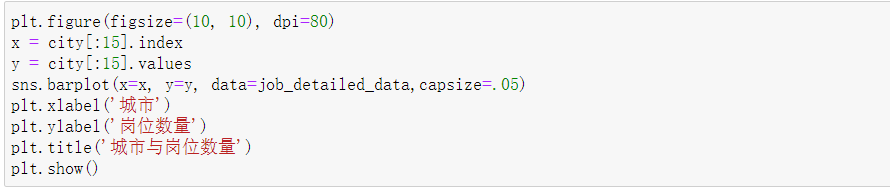


图7.5.3城市与岗位数量柱状图

③对柱状图进行颜色填充



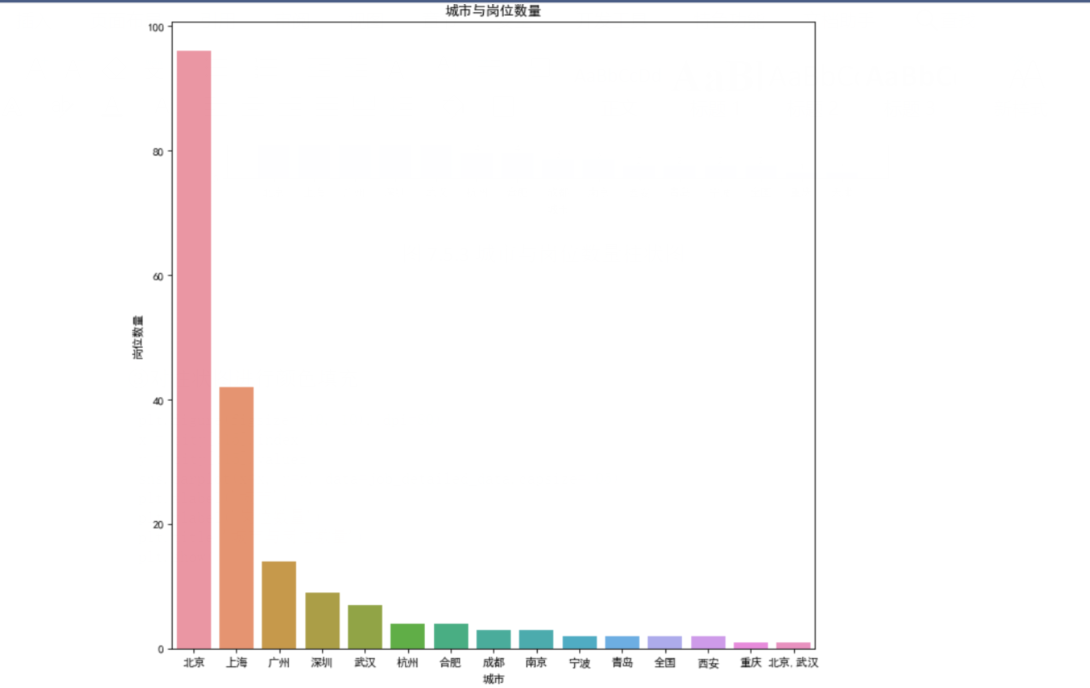


图7.5.4填充颜色柱状图

结论：城市的繁华程度影响岗位的需求量。一线城市（北上广深）特别是北上提供岗位数量比其他城市多，北京占到提供岗位的近一半。

# 

# 8.总结

这次python课程设计让我有了很深的感触和体会，虽然数据分析听起来并不是很难，但是在完成的过程中也有很多值得我们思考和学习的地方。在我看来课程设计需要的是课程所学知识的运用和延伸。十分感谢徐勇老师和同学们对我的指导，让我了解和学习了如何运用python语言对数据进行提取、清洗；如何使用语言做数据的分析和可视化图的绘制，让我们对于分析数据变得更加的快捷和直观。

徐勇老师非常细心的给我们讲了很多知识，可当自己真正动手去做时，才明白纸上得来终觉浅，得知此事要躬行。只有实践才能将理论更好的理解和贯通。通过此次课程设计，使我对于python的知识了解和掌握的更多，虽然过程中遇到了一些问题，但经过小组成员的认真思考、相互交流和鼓励，终于找出了原因所在，也暴露出了前期我在这方面的知识欠缺和经验不足。有时一些低级错误也是小组成员帮我检查和改正，同时教我如何理解和运用，最后通过实践出真知，亲自动手实践，使自己掌握的知识能马上运用，不再是纸上谈兵。  
 在分析数据过程中，我们不断发现错误，不断改正，不断领悟，不断获取。本身在践行“过而能改，善莫大焉”的知行观。这次python的分析数据终于顺利完成了，也能通过分析数据将课本和之前的知识进行温习巩固，在不太明白的地方进行查询资料和请教同学，终于将难关一个一个的攻克和解决。这次python课程设计也是我们这学期对此课程的一次综合性检测。通过这次做课程设计，我能发现自己学习上的漏洞，期间同学和老师也是十分细心的为我解答疑惑，才能使我顺利的完成本次课程设计。对于自身来说不能遇到问题就想到要退缩，一定要不厌其烦的发现问题所在，然后一一进行解决，只有这样，才能成功的做成想做的事，才能在今后的道路上劈荆斩棘，而不是知难而退，那样永远不可能收获成功。最后感谢徐勇老师的悉心教导和当我们遇到困难时耐心的为我们解决疑惑，没有徐勇老师的教导和帮助就没有我们的成长和进步！

# 9.参考文献

[1]董付国.Python可以这样学[M].北京:清华大学出版社，2017.

[2]刘宇宙.Python 3.5从零开始学[M].北京:清华大学出版社，2017.

[3]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)

[4]基于Python的数据可视化[J].罗博炜.信息记录材料.2019(12)

[5]董付国.python课程设计[M].北京:清华大学出版社，2015

[6]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.

**课程设计成绩评定表**

|  |  |  |  |
| --- | --- | --- | --- |
| 课程设计题目 | 实习岗位状况的数据分析的学习与实践 | | |
| 课程设计学生答辩或质疑记录：  1.怎样完成对数据的处理？  答：将读取的有效信息进行数据的清洗，将清洗后的数据存到.csv文件中，通过.csv文件将数据导入Mysql数据库，在数据库中通过语句对数据进行相应的处理。  2.python如何读取csv文件某列时保留空值？  答：可以用dict(data['Gender'].value\_counts(dropna=False))，就会保留空值。  3.数据清洗时的去重和异常如何处理？  答：直接删除是重复数据处理的主要方法，处理重复数据的方法drop\_duplicates。  异常处理：通过describe方法获取所有这些列的描述统计信息，排除那些金额数量为0的复数，经过筛选将那些异常值删除。 | | | |
| **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 | | | |