广西大学行健文理学院

实践周作业

|  |  |
| --- | --- |
| 题目： | 计科岗位 |
|  | 就业形势数据分析 |

|  |  |
| --- | --- |
| 学 部： | 理工学部 |
| 专 业： | 计算机科学 |
| 班 级： | 182 |
| 学 号： | 1738940341 |
| 学生姓名： | 杨启玢 |
| 指导教师： | 罗泉 |

二〇二〇年十二月

摘要

2020年的疫情爆发，让大学生的就业面临了严峻的挑战。疫情期间，很多公司，商店被迫关门。很多学长学姐已经本已经找到了企业，却因为疫情被告知公司已经倒闭。随着很多公司的倒闭，社会提供的岗位也越来越少。所以，今年的实践周，老师特意安排我们用爬虫工具爬取招聘网站上的职位信息，并写下这篇论文。

但是目前的疫情只是得到了控制，但是疫情还没有完全过去，偶尔还是会反弹一下。但是总体已经趋于可控范围内。

这篇论文主要就java,linux,web前端,游戏服务端，还有算法这5个领域，从薪资,学历要求，工作经验要求,这三个方向去分析疫情稳定后，我们国家上海的计算机相关行业的就业形式。

我使用Python，beautifulsoup4从boss直聘上爬取了近1000条数据，数据量较少，由于boss直聘有反爬手段。每一个关键字爬取20页就已经是极限了，20页往后直接是不让访问的。正因为每个关键字只能爬20页，所以我一共爬了5个关键自。分别是java,linux.web前端,游戏服务端,算法。

关键词：就业形式，计算机

目录

[第一章 相关领域简介 3](#_Toc27152)

[1.1 java 3](#_Toc4304)

[第二章 相关数据简介 4](#_Toc14504)

[2.1数据分析的手段 4](#_Toc12999)

[2.1.1平均数 4](#_Toc23977)

[2.1.2 中位数 4](#_Toc1717)

[2.2.3 众数 4](#_Toc28196)

[2.1.3 图表呈现 5](#_Toc12557)

[第三章 数据分析 5](#_Toc23010)

[3.1 java岗位分析 5](#_Toc13095)

[3.1.1 薪资分布情况分析 5](#_Toc29378)

[3.1.2 数据离散情况分析 8](#_Toc15777)

[4.1 Linux岗位分析 8](#_Toc15391)

[4.1.1薪资分布情况分析 8](#_Toc28354)

[4.1.2离散情况分析 9](#_Toc10806)

[5.1 web前端分析 9](#_Toc9394)

[5.1.1薪资分布情况分析 9](#_Toc31203)

[5.1.2离散情况分析 11](#_Toc2768)

[6.1 游戏服务端岗位分析 11](#_Toc19253)

[6.1.1薪资分布情况分析 11](#_Toc628)

[6.1.2离散情况分析 12](#_Toc15778)

[7.1 游戏服务端岗位分析 12](#_Toc19147)

[7.1.1薪资分布情况分析 12](#_Toc741)

[致谢 14](#_Toc10436)

# 相关领域简介

## 1.1 java

java主要是指使用java语言进行开发的领域，目前java技术主要用在网站的服务器，以及安卓系统的开发上。由于其良好的规范，以及长期以来技术生态的发展，备受程序员的青睐。java语言拥有非常优秀的设计规范，完全为企业打造，所以社会对java程序员的需求量非常大。但是也由于太受欢迎，导致很多大学，和培训机构都在做相关的培训。java程序员的工资并不是特别的高，但也侧面的说明了java语言的受欢迎程度

1.2 Linux

Linux系统也许大家十分陌生，但是在服务器市场，它的地位是举足轻重的。Linux诞生于一个天才大学生的电脑上，他写完第一版的Linux内核后，将其源代码公布在了网络上。在其公布源代码后，立刻受到了很多从事计算机行业的人的关注。他们免费为Linux提供代码，后来在服务器市场上，因为其免费且可靠。所以在服务器市场备受青睐。

1.3 web前端

Web前端是一个大类，它包含的语言是和技术是比较多的。基础的开发语言包括javascript,html,css,jsp,asp。框架包括bootstrap,vue,elementPlus,等一系列的前端框架。web前端就是我们经常看到的网页。

1.4 游戏后端

游戏后端是一个有门槛的领域，因为很少有相关机构去专门做游戏后端相关的开发。游戏后端使用的语言比较局限，一般是用c++，lua，erlang，java，go。其实这个在网上是不好调查的，我查了百度，一般都是推荐c++。但是据我了解，每个公司自有自己的一套体系，这跟项目领导的风格是息息相关的。例如我目前实习的公司就是用c++做的底层通信，lua做的上层逻辑。并没有全用c++，甚至c++只占很小的一部分。

1.5 算法

算法可以说是计算机领域的一块非常难啃的骨头，算法是计算机的核心。所以专门从事算法领域的门槛是最高的。算法非常的抽象难以理解，涉及的知识非常的多。因为这方面的人才比较缺乏，所以一般offer拿的也比较的高。

# 第二章 相关数据简介

## 2.1数据分析的手段

### 2.1.1平均数

平均数是指在一组数据中所有数据之和再除以数据的个数 平均数是统计中的一个重要概念.平均数一般是指算术平均数,也就是一组数据的和除以这组数据的个数所得的商.在统计中算术平均数常用于表示统计对象的一般水平,它是描述数据集中程度的一个统计量.既可以用它来反映一组数据的一般情况,也可以用它进行不同组数据的比较,以看出组与组之间的差别.用平均数表示一组数据的情况,有直观、简明的特点,所以在日常生活中经常用到,如平均速度、平均身高、平均产量、平均成绩等等.

使用平均数这个工具，可以让我们知道在上海，我们研究的计算机领域的平均工资。帮助我们从平均数的角度去了解整个数据的离散情况。我们可以通过研究平均数，来了解我们毕业后，从事相关行业所很可能拿到的offer。

### 2.1.2 中位数

中数是按顺序排列的一组数据中居于中间位置的数，即在这组数据中，有一半的数据比他大，有一半的数据比他小。同样，通过研究中位数，可以帮助我们了解我们毕业后，从事相关行业很有可能拿到的offer。

### 2.2.3 众数

众数的定义：一组数据中，出现次数最多的数就叫这组数据的众数。

众数的个数：一组数据中可以只有一个众数，也可以很多个，也可以没有众数。

众数的特点：用众数代表一组数据，可靠性较差。不过，众数不受极端数据的影响，并且求法简便。在一组数据中，如果个别数据有很大的变动，选择中位数表示这组数据的“集中趋势”就比较适合，选择众数不一定适合，选择平均数肯定不适合。

使用众数可以帮助我们了解大多数相关行业从事者的薪资水平，相较于平均数与中位数，它不受极端数据的影响，所以可能更具有代表性。

### 2.1.3 图表呈现

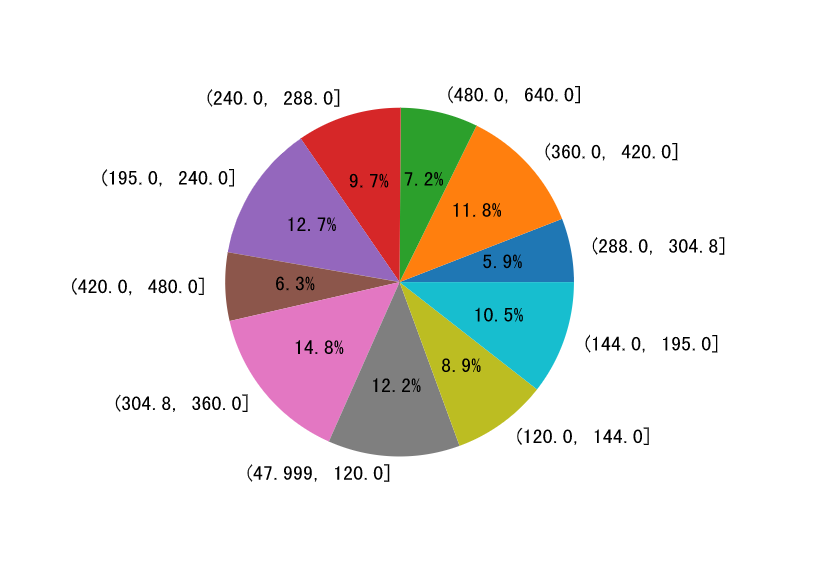
使用图表去分析数据，已经是一个历史悠久的数据分析手段了。特别是随着计算机软件的发展，已经有了excel这样的可以方便的根据数据画出图表的软件了。更高级的，还有用python对数据进行清洗和绘制图形。

# 第三章 数据分析

## 3.1 java岗位分析

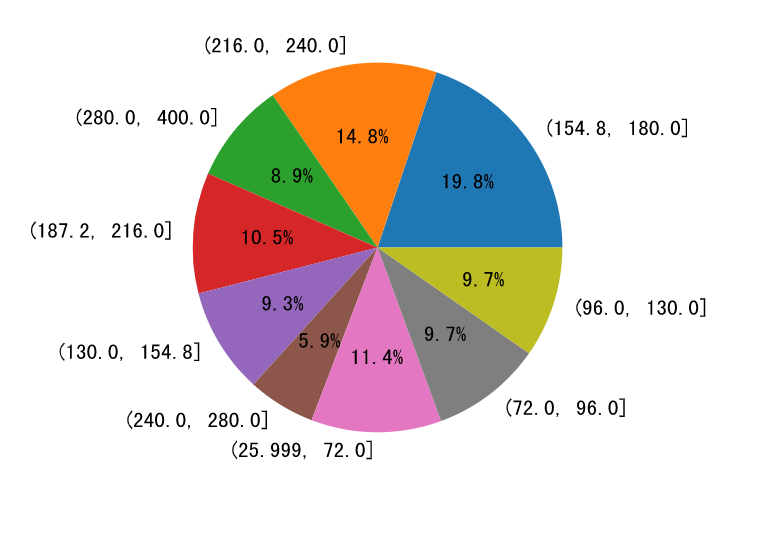
### 3.1.1 薪资分布情况分析

Boss直聘上给出的每一个职位的薪资是一个范围，而不是一个准确的数值。所以我取这个范围的最高值，和最低值。并给这两个数值分别命名为范围最高值和范围最低值。我对提供的每一个职位的薪资进行范围化(桶分)，我取了范围最高值和范围最低值中的最高值和最低值，并依据他们，将数据划分为十个部分。

首先，我们看一下范围最高值的薪资分布情况

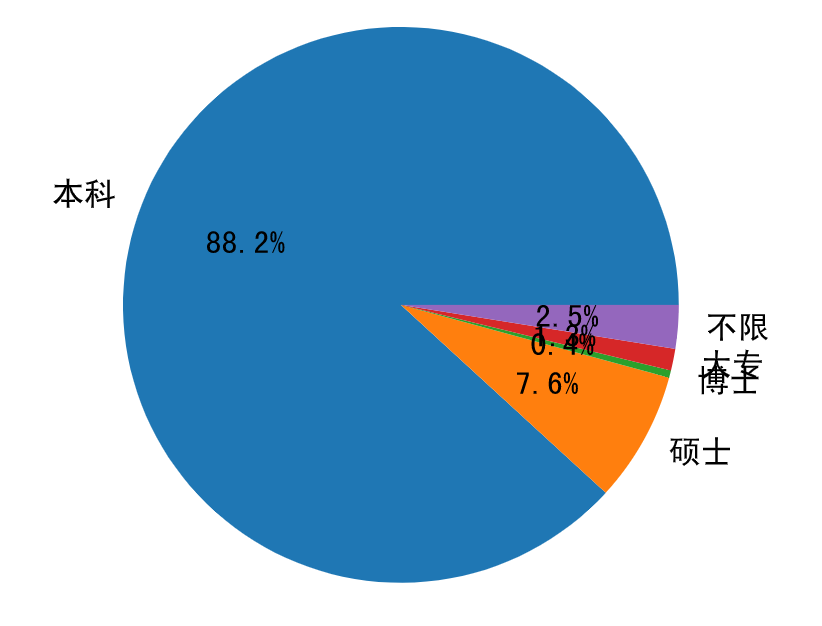
从图中我们可以看到，每年拿304k的人最多，其次是每年47k。最高年薪在480k，到640k之间，说明这个行业的上限非常的高。

再让我们看看范围最低薪资的分布情况



即使是最低范围薪资的情况不是十分的乐观，最低每年25k-72k的有11.4%，但是每年给到154k的占了近20%。工资在每年72k以上的占比竟然高达89.6。可以说，好好干很有前途。

我们再来看看学历要求



学历要求就比较高了，基本在大专之下就不用看了。大专已经成为了行业准入的及格线。

### 3.1.2 数据离散情况分析

最低范围年薪分布情况



最高范围年薪分布情况

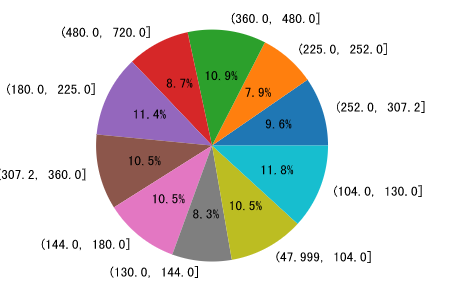


可以看到 无论是中位数，众数，还是平均值都在每年180k左右徘徊。只算最低范围情况就已经是这个结果了，而最高范围年薪的平均分布情况竟然再285k每年以上。在受极端值较小的中位数表现下，也有288k每年的高薪

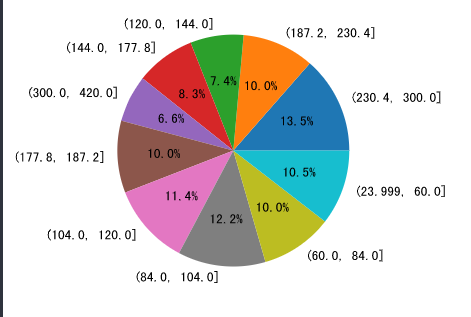
### 4.1 Linux岗位分析

### 4.1.1薪资分布情况分析

最高范围年薪分布情况



最低范围年薪分布情况



就Linux和java对比来看的话，他的下限和java差不多，上限也和java差不多，在这里没有什么好说的

### 4.1.2离散情况分析

最低年薪离散数据



最高年薪离散数据

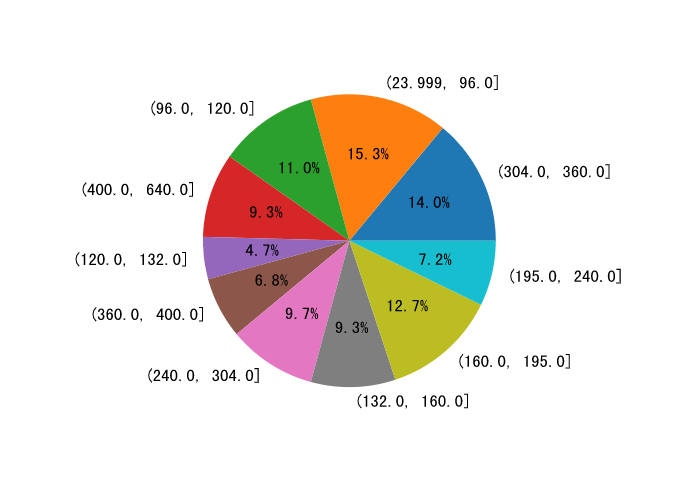


就离散情况来看，数值上没有java那么漂亮，但是也很不错了

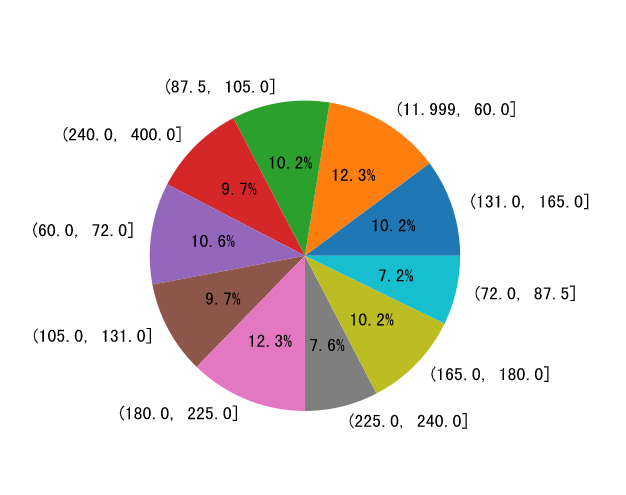
### 5.1 web前端分析

### 5.1.1薪资分布情况分析

最高范围年薪分布情况



最低范围年薪分布情况



和上面的差不多，但是最低年薪到了12k。前端的入门简单，但是要真的想有一个高薪，还是要有深入的研究的。

### 5.1.2离散情况分析

最低范围年薪离散情况



最高范围年薪离散情况

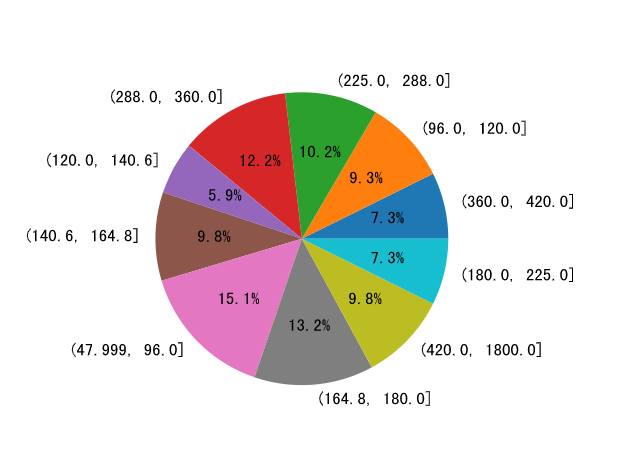


离散的表现没有前面两个强，最低值确实有点拉。这几个最低值数据极端值，没有什么代表性。而且就离散情况来看，各个数据离散程度太大。这里不表达意见。

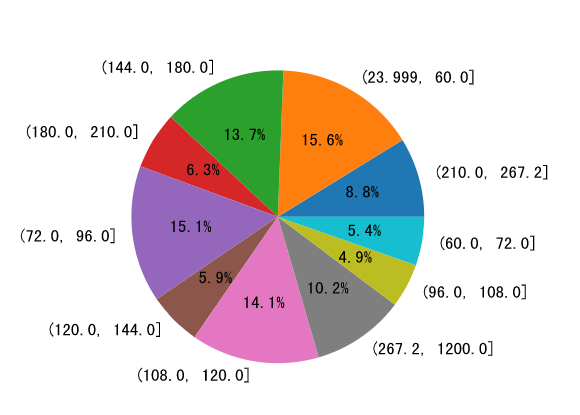
### 6.1 游戏服务端岗位分析

### 6.1.1薪资分布情况分析

最高范围年薪分布情况



最低范围年薪分布情况



这个最高值是真的亮眼啊1800k每年。可以说基本面强于上面任何一个，但是呢，这个行业在23k-60k每年的占15。

### 6.1.2离散情况分析

最低范围年薪离散情况



最高范围年薪离散情况



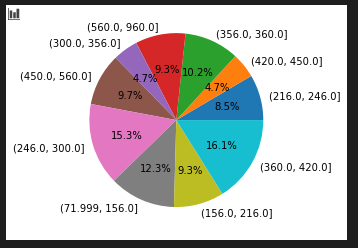
离散情况出乎我的预料，虽然最高年薪非常亮眼，但是最低范围年薪的表现并不是特别的亮眼，相反的是最高范围平均值反而是前面当中最高的一个。

### 7.1 游戏服务端岗位分析

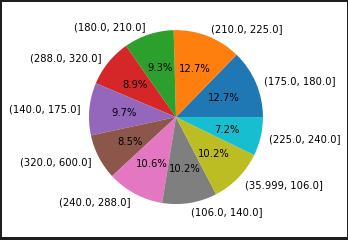
### 7.1.1薪资分布情况分析

终于到了我最感兴趣的算法领域了，老实说，我有点失望，但是呢也让我眼前一亮。

最高范围年薪分布情况



最低范围年薪分布情况



从图里面我们看见，算法的下限是我所研究的5个领域中最高的，但是，准入门槛也是最高的,这个我们后面说。算法的最高薪是这五个领域里面第二高的，我本期望他应该是最高的一个，毕竟算法工程师的要求是非常高的，但是最高的投入不一定就有最高的回报，薪资最高值的皇冠给了游戏服务端。

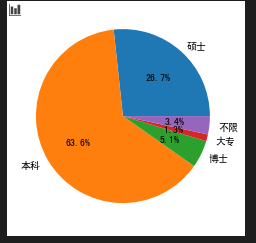
但是，算法有个不得不说的点，就是他的平均值是最高的，而且离散的程度也是最小的一个

让我们分别看看他的离散数据





可以看见，平均值分别为211k和347k。不但均值最高，而且离散程度也是最小的一个。拥有高的最低年薪，也拥有最高的平均值，最低的离散程度。毫无疑问，基本面最强。但是我们来看看学历要求。



如果说本科学历还能在上面四个领域打遍天下的话。在算法领域，可能本科学历就不会活得像前面那么轻松了。硕士要求高达26%。要知道，本科学历在前面4个领域可以说是可以横着走的。高学历的要求，也就是说准入门槛是最高的一个。

# 致谢

本论文是在罗全老师的悉心指导下完成的。罗全老师作为一名优秀的、经验丰富的教师，具有丰富的数据分析知识和教学经验，在整个论文实验和论文写作过程中，对我进行了耐心的指导和帮助，提出严格要求，引导我不断开阔思路，为我答疑解惑，鼓励我大胆创新，使我在这一段宝贵的时光中，既增长了知识、开阔了视野、锻炼了心态，又培养了良好的实验习惯和科研精神。在此，我向我的指导老师表示最诚挚的谢意！