|  |
| --- |
|  |
| 网络数据采集与整理报告 |
| 专业：大数据Q1741  姓名：刘永芳  学号：17080166 |
|  |
|  |

统计学专业就业市场需求分析与培养建议

1. **背景简介**

近年来,数据分析专业人才需求高涨,各大高校纷纷开设大数据研究等相关专业,大数据时代对统计学的发展提出了更高的要求。本文基于python+selen

-ium的环境爬取前程无忧招聘网站上的信息来简易分析基于大数据时代下社会对于统计学专业人才的需求，进而提供对统计学人才的求职建议及可行的统计学人才培养建议。随着数字化的进程不断加快，人们越来越多地希望能够从大量的数据中总结出一些经验规律从而为后面的决策提供一些依据。统计学专业不是仅仅像其表面的文字表示，只是统计数字，而是包含了调查、收集、分析、预测等。应用的范围十分广泛。

湖北经济学院于2001年开设了统计学专业，2012年，统计学获湖北省省级重点(培育)学科，并在本方向下开设了统计学、应用统计学、经济统计学（金融统计方向）3个本科专业，并于自2014年起与湖北大学开展应用统计硕士研究生的联合培养工作。

**二、数据来源及抓取要点**

（一）数据来源

前程无忧网（https://search.51job.com）

1. 抓取要点

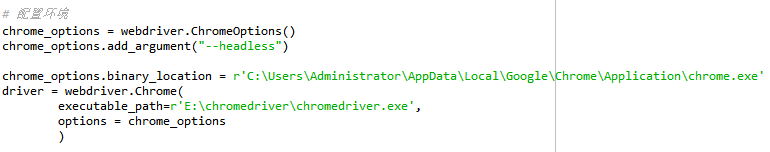


图1-1

1. 配置selenium与第三方浏览器chrome结合（图1-1）。Selenium自身不带浏览器，不支持浏览器的功能，所以得与第三方浏览器结合。



图1-2.1 图1-2.2

1. 进入爬取网页页面滚动屏幕获取职位名（如图1-2.1左列）通过职位名称进入详情页（如图1-2.2）。滚动屏幕是为了防止滑动才能出现职位名从而直接获取导致数据不全的情况发生。



图1-2.3

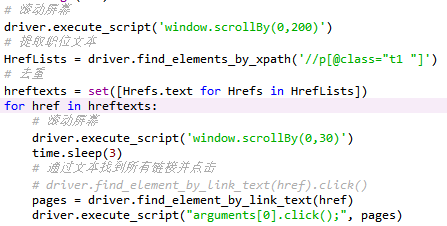


图1-2.4

但在用driver.find\_element\_by\_link\_text(href).click() 命令的时候出现了

图1-2.3的元素单击已拦截的问题，于是我将命令改写成图1-2.4就解决了上述问题。



图1-3.1 图1-3.2

1. 进入详情页后直接对数据进行爬取。但在点击后直接对数据进行爬取时出现了无法找到元素的错误，在检查确定我的解析地址无误后，考虑了driver当前所在窗口的问题，由于我的电脑在用chrome浏览器时点击元素会新开一个窗口，如果driver点击了新开的详情页窗口（如图1-2.2）后又回到了原窗口（如图1-2.1）就可能出现找不到元素的问题，于是我分别在爬取元素的命令前后添加了切换到当前打开窗口的命令（如图1-3.1）和回到原窗口（如图1-3.2）的命令，问题得到解决。



图1-4.1

1. 翻页到最后一页。我直接采用了获取到“下一页”文本进行点击的命令。（如图1-4.1）

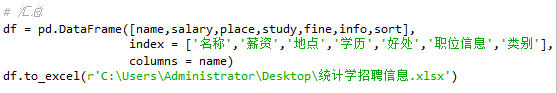


图1-5.1

1. 将数据汇总，将得到的数据保存。（如图1-5.1）

**三、结果分析**

通过对爬取的数据进行数据筛选和数据清洗操作，简单的对得到的数据进行描述性分析。进入大数据时代以来，数据分析成为了当今的热门词汇，于是我对数据分析员进行了较多的统计分析。

1. 职位需求分析

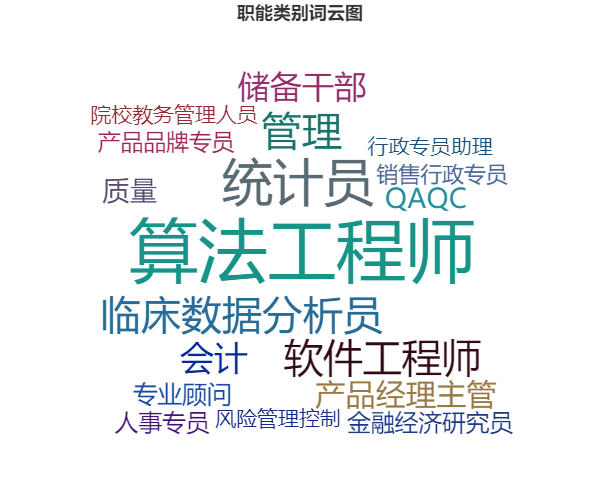


图2-1.1

用得到的数据通过Python绘制了词云图（如图2-1.1），可看出统计学招聘最多的是算法工程师，其次是统计员、软件工程师、管理和临床数据分析员，可推断出目前社会对统计学专业人才需求量相对较大的是算法工程师，大数据时代已然到来。

1. 职位需求薪资分析

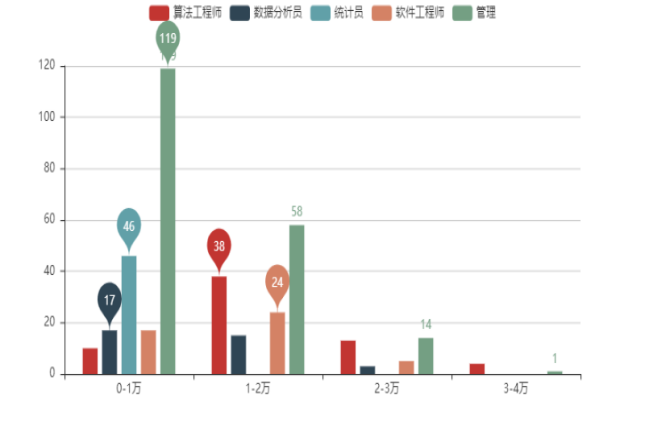


图2-2.1

通过职能类别词云图（图2-1.1）对前五类职能进行薪资统计对比分析来比较哪一类工作薪资相对较高，由图2-2.1可看出管理和统计员薪资在0-1万之间占比较大，算法工程师和软件工程师相对较高，在1-2万之间，数据分析师在0-1万和1-2万之间的比例相差不大，对比分析出目前统计学专业从事技术类工作的薪资相对较高。

（三）全国统计数据分析岗位需求分析

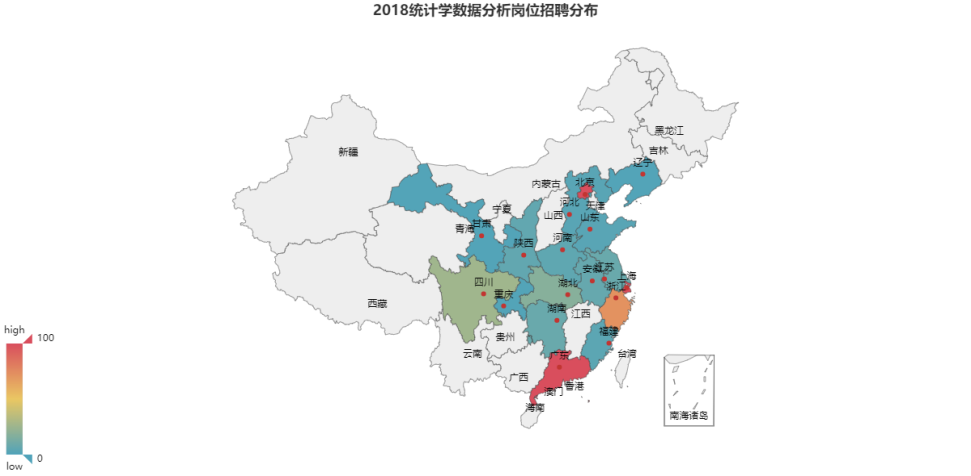
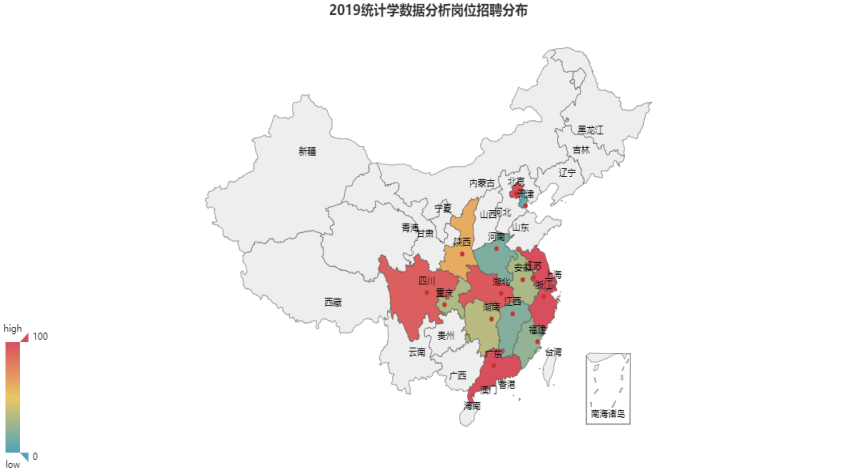


图2-3.1 图2-3.2

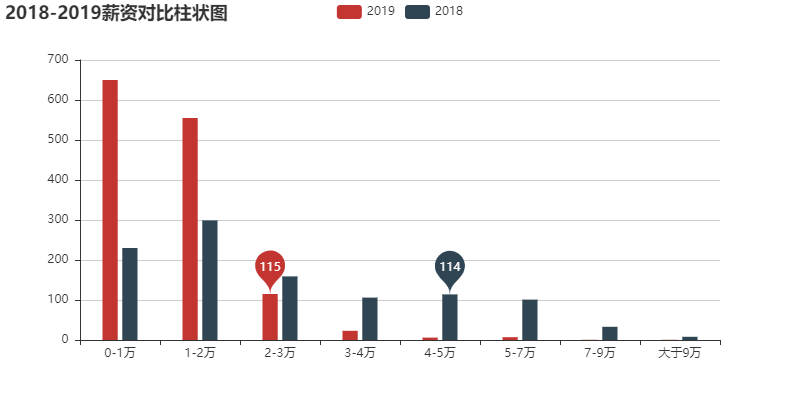


图2-3.3

对比分析2018年和2019年的全国需求分布图，2018年北京、上海、广州三个城市对数据分析人才的需求较大，到2019年四川、湖北、浙江、江苏等省份对人才的需求也在快速扩大，全国对数据分析人才的需求在不断增多。为了分析数据分析员薪资的趋势，于是对2018年和2019年的数据分析员的薪资进行分析（如图2-3.3），对比分析出数据分析员的薪资程下降趋势，2018年数据分析员薪资大部分在1-2万之间，今年则趋向于0-1万。

通过分析数据分析员全国需求分布图和薪资对比图，数据分析员薪资下调可能是由于大量数据分析人才涌入市场以及数据分析人才水平参差不齐导致的。

（四）技能及其相关薪资分析

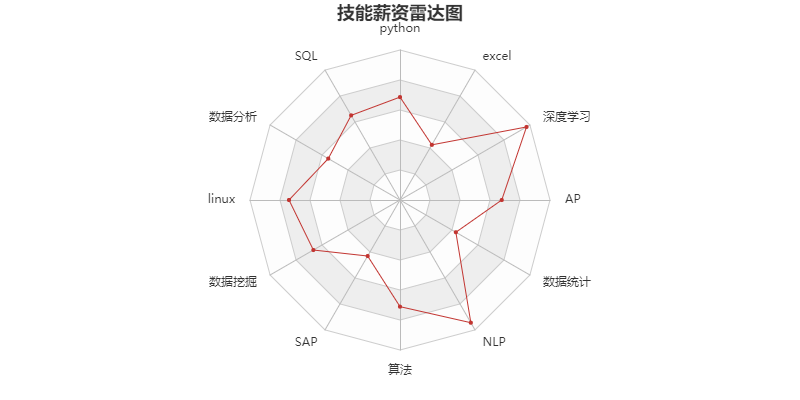
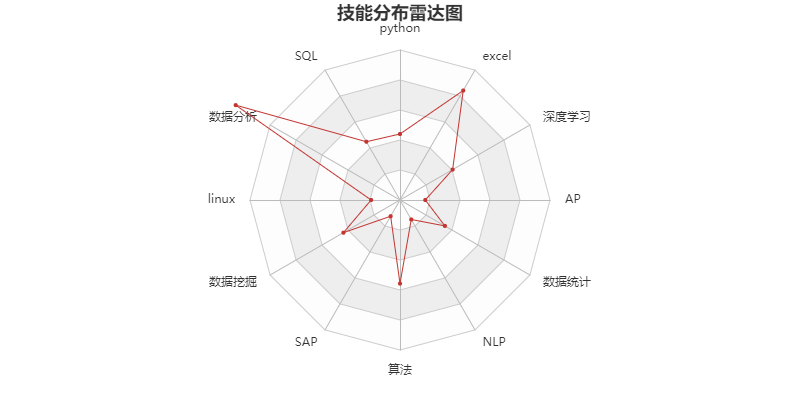


图2-4.1 图2-4.2

由技能分布了雷达图可看到在技能要求上职位对数据分析还是需求较多，其次是Excel、算法和深度学习，Excel是目前统计学乃至其他非专业学生必备技能就不多做分析，主要的是看数据分析和算法以及深度学习，这三中技能其实都与算法相关联，这也与我们所做词云图（如图2-1.1）相对应。

通过技能薪资雷达图可看出，掌握NLP（自然语言处理）和深度学习技能的薪资较高，可能是由于掌握这方面人才相对较少导致。数据分析的薪资相对与算法和深度学习来所薪资较少，但也在1万/月左右。

（五）招聘学历分析

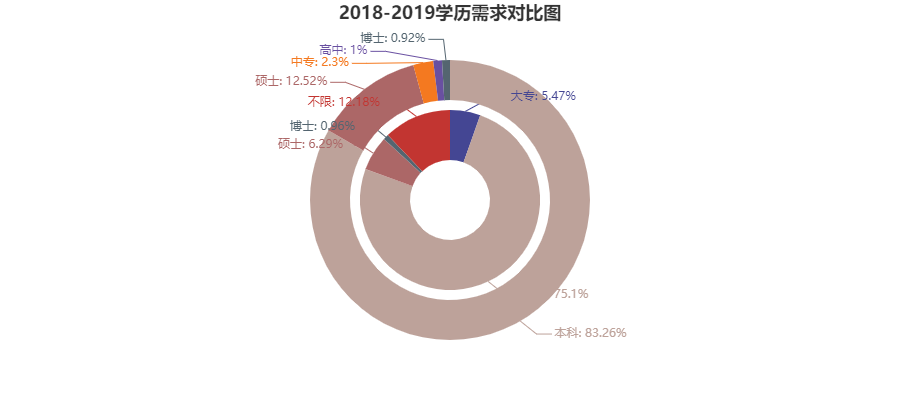


图2-5.1

分析2018-2019学历需求对比图（如图2-5.1）可看出，本科占比从75.1%增长到83.26%，硕士占比从6.29%增长到12.52%，社会对统计学人才的学历需求大部分为本科，其次是硕士。

（六）工作福利分析

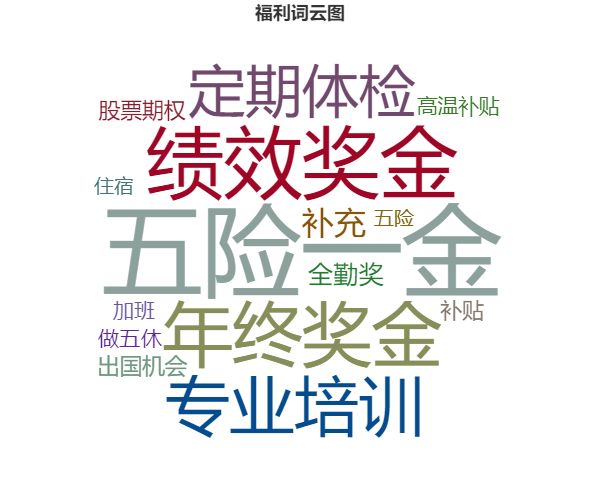


图2-6.1

通过福利词云图（如图2-6.1）可以看出公司对人才的福利吸引大多为五险一金、年终奖金、绩效奖金、定期体检和专业培训，少数公司还有股票期权、出国机会和高温补贴。

**四、意见与建议**

通过湖北经济学院信息管理与统计学院统计学专业人才培养方案可知，2016级和2017级的专业核心课程是：数学分析、大学英语、概率论、数理统计、统计学原理、抽样技术、市场调查、国民经济核算、时间序列分析、R语言程序设计、试验设计、计量经济学、多元统计分析、贝叶斯统计。而2018级人才培养方案专业核心课程去掉了数学分析、大学英语和试验设计这三门课程，增添了数学建模和机器学习两门课程。

结合湖北经济学院信息管理与统计学院统计学专业人才培养方案以及本次招聘网站数据分析对学生、老师和学校提出了一些意见与建议。对学生来说，学生应该注重算法的学习，了解算法的原理并且会运用算法实现数据分析，辅助学习工具Python和SQL数据库。根据现在需求市场来说，数据分析成为大多数工作的工作需求，由于数据分析员水平的参差不齐导致其薪资分布不均，想要从事相关工作的同学熟练的掌握各种算法及其原理并灵活运用可能薪资会达到一个较高水平。想要更深层次的学习的话考研也是一个不错的选择。对老师来说，应当注重学生的运用能力，理论与应用相结合才能使知识得到更加充分的运用。对学校来说，应当设置较为系统的课程体系。在课程设置的时候可以分为理论课和实践课，如果课时不够，可以分为一门理论课和一门实践课两门课来学习，中间相隔时间不超过一个学期。例如机器学习这门课程，当时我们学习时本该全是理论课，但老师利用了理论课的时间给我们上了实践课，虽然学会了Python算法的运用，但我们所学习的内容大大减少。

1. **代码**
2. [招聘数据爬取代码链接 （可点击打开）](https://github.com/AntikoFang/Python-Spyder/blob/master/%E7%BB%9F%E8%AE%A1%E5%AD%A6%E6%8B%9B%E8%81%98%E6%8A%A5%E5%91%8A/%E7%BB%9F%E8%AE%A1%E5%AD%A6%E6%8B%9B%E8%81%98.py)
3. [全部项目代码链接](https://github.com/AntikoFang/Python-Spyder/tree/master/%E7%BB%9F%E8%AE%A1%E5%AD%A6%E6%8B%9B%E8%81%98%E6%8A%A5%E5%91%8A)