boss直聘招聘信息分析

班级：17信管

学号：2017254144

姓名：卢某某

目录

[一、摘要 2](#_Toc44873190)

[二、背景介绍 2](#_Toc44873191)

[三、文献综述 2](#_Toc44873192)

[四、研究目的与说明 2](#_Toc44873193)

[五、数据清洗、预处理 2](#_Toc44873194)

[六、探索性分析 3](#_Toc44873195)

[七、数据建模、分析模型介绍 3](#_Toc44873196)

[八、结果分析 3](#_Toc44873197)

[九、结论和建议 7](#_Toc44873198)

[十、参考资料 8](#_Toc44873199)

# 一、摘要

本文以boss直聘上的招聘信息作为来源，得到的数据集是关于10个互联网行业基本信息，通过python数据分析，得到各城市对于不同职位的学历、经验等需求，也通过薪资对比，找出什么行业薪资最高最有前景。

**关键词**：python。数据分析，数据集

# 二、背景介绍

现在的社会是一个高速发展的社会，科技发达，信息流通，人们之间的交流越来越密切，生活也越来越方便，大数据就是这个高科技时代的产物。阿里巴巴创办人马云来台演讲中就提到，未来的时代将不是IT时代，而是DT的时代，DT就是Data Technology数据科技，显示大数据对于阿里巴巴集团来说举足轻重。

随着大数据时代的到来，技术进步必定会带领工具的升级换代，企业和个人都需要一种新的方式简单去分析和获取网上的信息。那么python就是当下担任该重要角色的语言。

# 三、文献综述

（1）基于Python的软件技术人才招聘信息分析与实现——以前程无忧为例

综述：这篇文献比较简短，集中讲述了利用大数据分析招聘信息的分析和实现，使用了什么库，什么方法，操作流程和一些参考意义。

（2）基于前程无忧招聘信息的数据人才需求分析

综述： 数据来源，数据提取，数据清洗，数据分析，数据可视化，关于各种因素在招聘中的比重，比如，学历，经验，职责词云等，还在最后给出了技术学习需求图。

（3）其余三篇都是简短的文献，着重点都是在数据可视化下的数据分析，虽然行业和数据集不一样，但是描述方法都差不多，而其中刑事警察学院那篇还加上了数据爬虫的步骤。

# 四、研究目的与说明

目的：因为近日一直在上前端的课，临近毕业比较关注前端信息的招聘，还有统一一下前端人员应该掌握技能要求，以便在找实习时已经掌握和了解。所以恰巧进行数据分析。但是boss直聘网站有反爬机制，爬虫得需要登录还要利用cookie去进入，每次登录cookie都不一样，所以试了了几次都没有成功，就在网上找了数据集，作者本人对数据集也做了一些处理，例如把实习的工作去掉，因为实习有些按天计算，而且工资很低，一起分析的话不能准确保证接近真实情况。在此基础上，我做的数据清洗就是列名重命名还有一个重复行去除。我也学习了数据分析中的饼图画法，虽然最后用的比较多的是在慕课平台上学习到的pie（）函数。总体上说就是我想画什么图就去网上找图的教程。而知网上的五篇文献，最主要的作用是提供一个思路，因为多数是文字描述，并不会给代码出来。最后，因为前后画图时间不一致，学习情况也不一样，所以图会有不同情况。

# 五、数据清洗、预处理

更改列名：预设字典，原列名是字典的key，更改的列名是values，然后通过data.rename(columns=name\_dict,inplace=True)进行更改。

去除重复行：data.duplicated()返回每行的true和false，true代表有重复行，反之没有。data.duplicated().sum()，进行重复行计算。然后删除重复行并进行原数据覆盖data=data.drop\_duplicates()。

# 六、探索性分析

分组：groupby（）

折线图：

* matplotlib.pyplot库的Plt. plot (x, y, format\_string, \*\*kwargs)，也是静态图。
* seaborn的lineplot也可以画折线图，sns.lineplot(x,y,data,……)

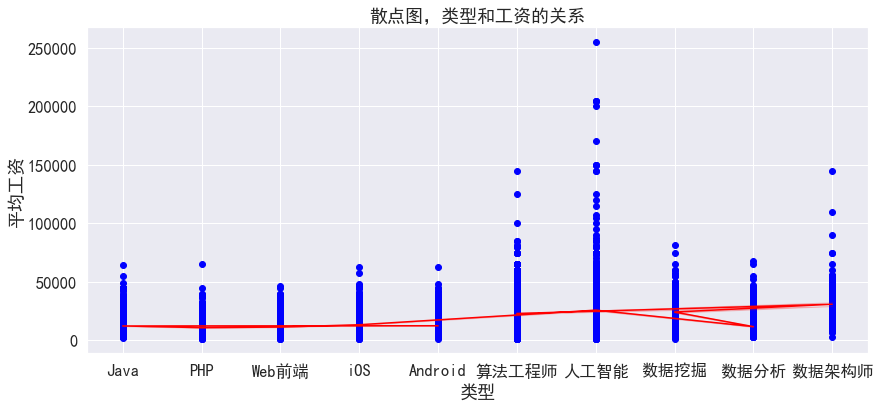
柱形图 ：

* seaborn的countplot(x,hue,data,……),以某一属性列进行数据分类加总的柱形图，属于静态图，鼠标点上去不会有变化。
* pyecharts的Bar是一个动态图，就是说鼠标点上去会动。还可以在侧边标签点击看数据。若想鼠标不滑过显示数字，只要把is\_label\_show=False即可。

饼图：

* pyecharts库的Pie函数，通过.groupby函数分类计总在转化为列表给Pie用，可以看出不同情况下各行业的占比，例如工作类型占比。

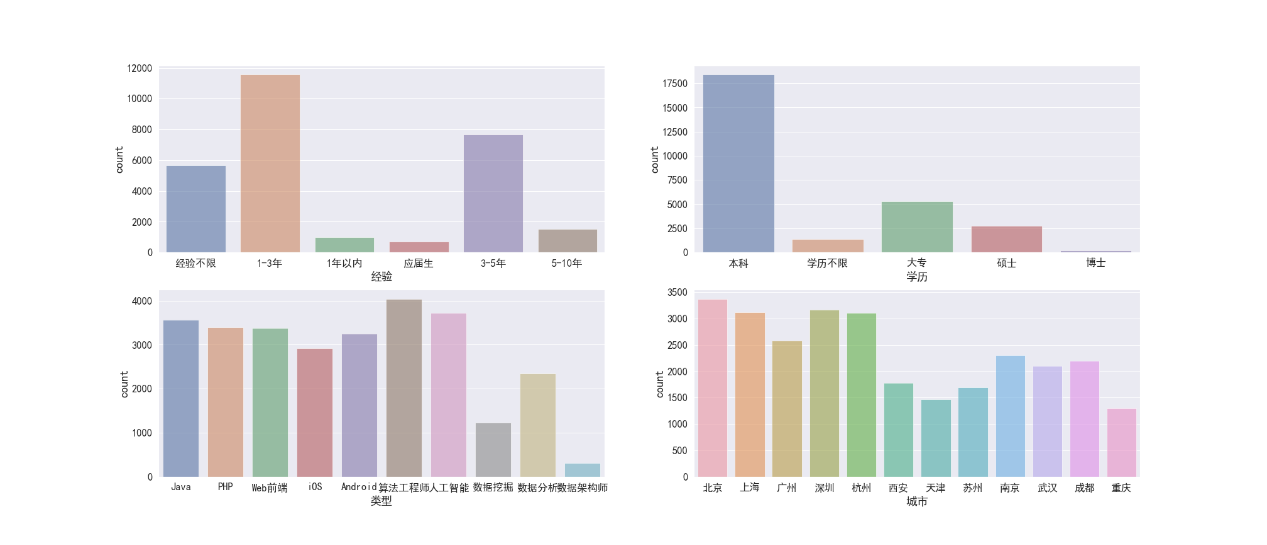
# 七、数据建模、分析模型介绍

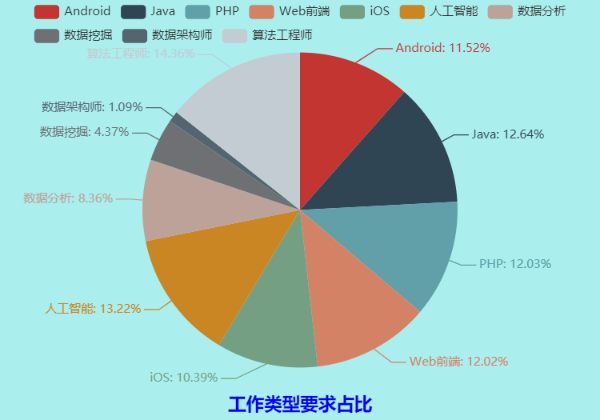


进行相关分析时，plt.scatter制作散点图加线性图，看不出什么关系，招聘信息是一个某一个时间点的信息，是一个静态数据量，暂时找不到关系和建模方法，水平不够。

# 八、结果分析

**整体情况需求情况**



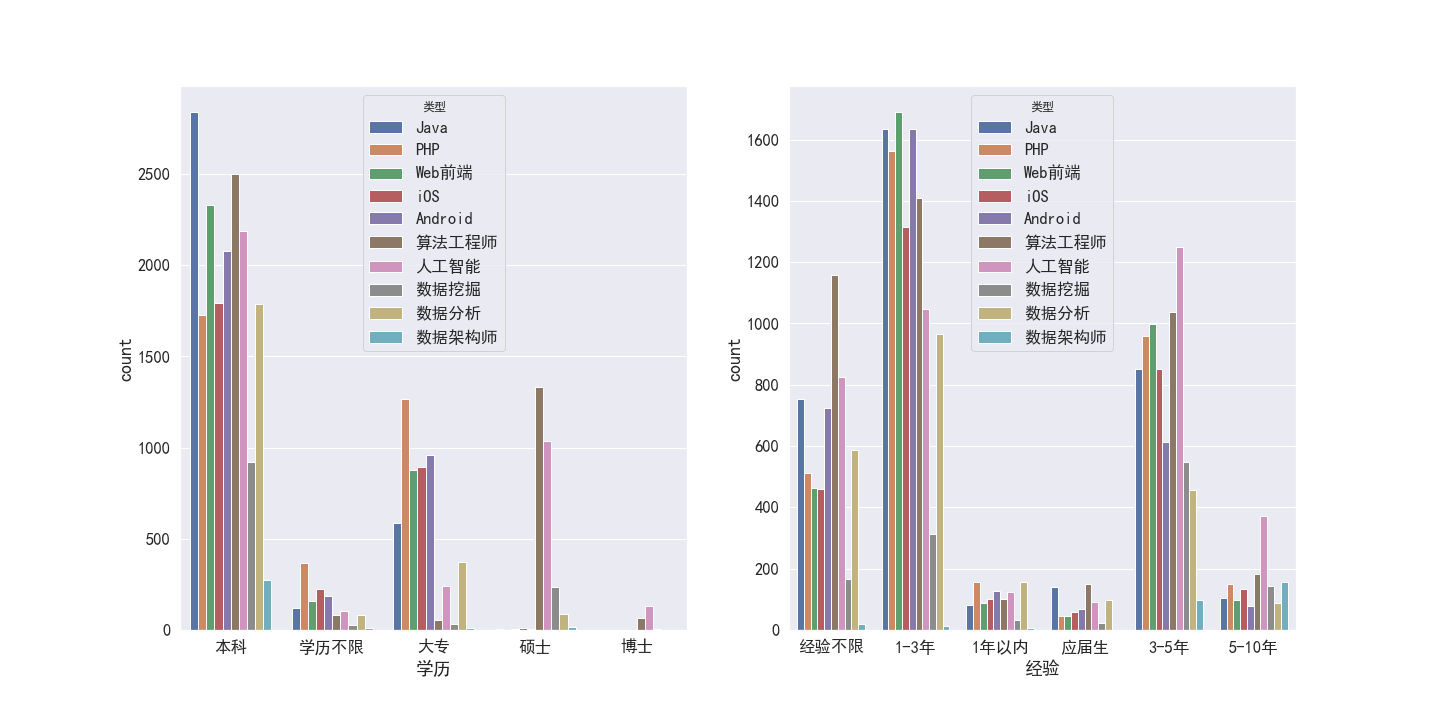


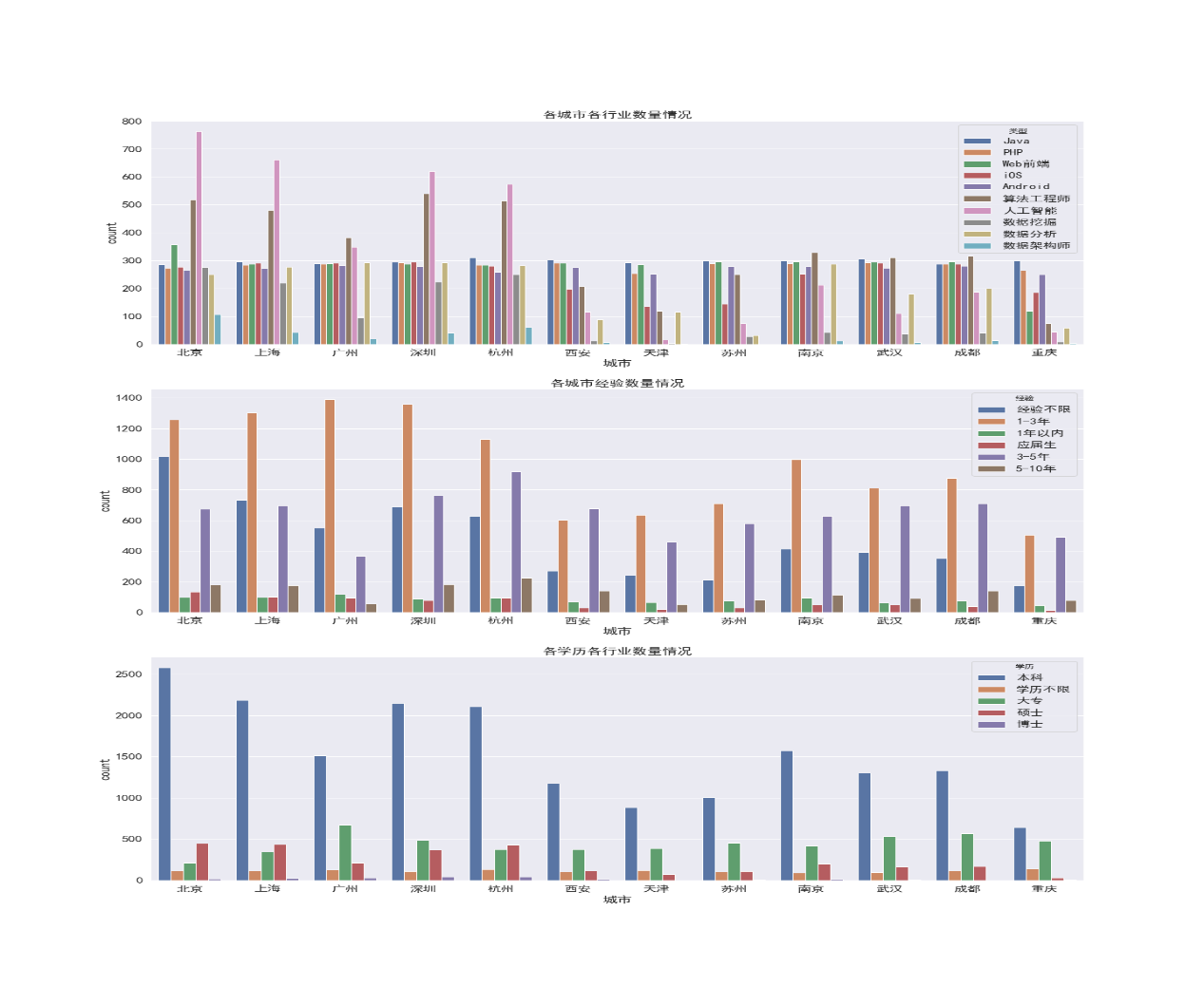
**结论**：从上面可以看出，普遍招聘人员要求工作经验基本是1-3年，后面是3-5年，与经验不限和应届生、一年以内而言，可以看出对于应聘者对于经验没有很严苛。

而在学历上讲，找工作最基本要本科学历和以上，学历不限占很少数，所以对于找工作而言，学历还是很重要的。

计算机行业总体情况，比较突出的是数据挖掘和数据架构师，因为需求数不高，其他行业总体差不多，其中算法工程师最受热门。就城市而言，沿海城市对于互联网行业需求高，而在沿海中，广州属于比较低水平。

**细分情况**



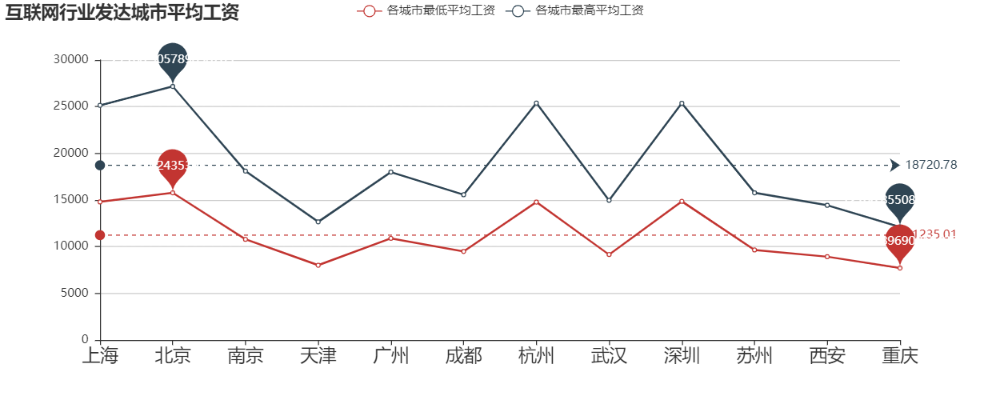


结论：这是一个在分类基础上再分类的数据图，本科阶段，java的需求量最高，次之是算法工程师。北京，上海，深圳，杭州的人工智能人才需求量突出，这也和现实情况相应衬，杭州-阿里巴巴公司总部地，上海-发展前沿，深圳-国家扶持，北京-首都。最低本科生学历要求突出，再次说明，学历是最基本的敲门砖，机会也多，读书是值得的。

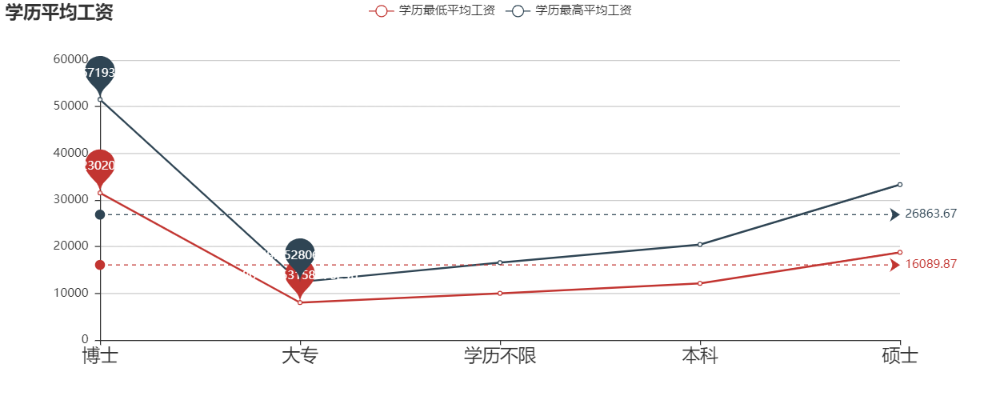
**薪资方面**



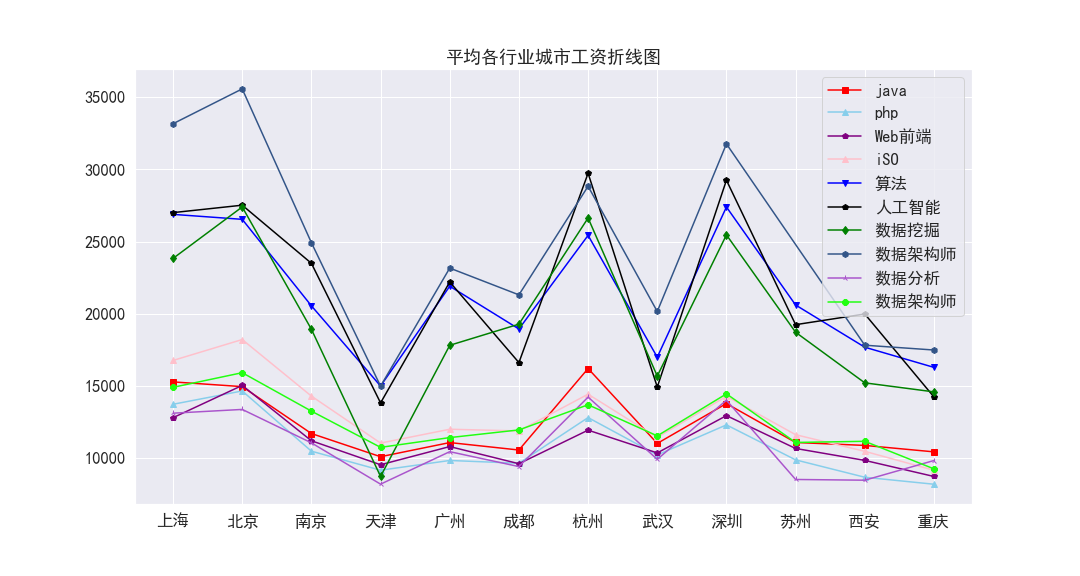
结论：人工智能，数据挖掘，数据架构师的薪资普遍高于其他行业，Android，java，php，web前端处于低水平，php的薪资处于最低水平，也验证了，php现在不怎么流行的现状。但是结合上面的数量而言，人工智能的需求数不高。



结论：上面数量分析中，北京，上海，深圳，杭州需求人工智能的人数高，也反映其城市科技含量高，而这些城市的互联网的平均工资也高，在这里不谈论物价水平，在这几个城市发展，会是不错的选择。



结论：除去学历不限，因为难以界定学历不限，所以总得来说学历越高工资越高，多读书对于自己是一个不错的投资。



结论：数据架构师在不同的城市的平均工资落差很大，在北京里，在所有的平均工资属于最高。在前面提到Android，java，php，web前端工资水平处于低水平，而在这个图中，这些行业在每个城市落差不大。哪些发展超一线城市，四大城市北京，上海，杭州，深圳工资普遍高于其他城市。

也分析了web前端，发现和总体情况差不多，这里就不多以描述了。

# 九、结论和建议

boss招聘信息分析的并不全面，只是简单挑了几方面来说明情况，而且可能也没有选择最优图形的表示，比如比重方面。饼形图最合适，趋势方面折线图合适，在整个分析中，大部分用了折现加柱形的结合来说明情况，没有很好分类的说明，但总体分析已达到。

最后给自已一点建议，多读书多进步，整个互联网行业的技术更新比一般行业快，所以在当下学习到的东西未必在未来还能一样受用，就好比如php，现在就处于一个渐渐下降的过程，处于薪资水平最低，而在整体上看，人工智能正是当下发展快速的行业，虽需求量少，但是薪资水平高于其他行业，而在几年前或许还少有见人工智能，可见，互联网某一项技术不是永恒的，只有自己跟着进步，一点点更新自己的知识库，举个例子，java的jdk都更新到了第14版本了,总不能还用着jdk6的技术去开发吧！官方版本都在不断优化升级了。

# 十、参考资料

（tips：pyecharts和matplotlib库有版本的问题。）

数据集地址<https://blog.csdn.net/baidu_41871794/article/details/105785329?utm_medium=distribute.pc_relevant_t0.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase&depth_1-utm_source=distribute.pc_relevant_t0.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase>

学习groupby函数：<https://www.cnblogs.com/Yanjy-OnlyOne/p/11217802.html>

学习Bar函数：<https://pyecharts.readthedocs.io/zh/latest/zh-cn/charts_base/#bar>

1.【国家精品课】Python数据分析与展示（老师名字：嵩天）—北京理工大学

<https://www.icourse163.org/learn/BIT-1001870002?tid=1450349444#/learn/announce>

2.还选择性浏览：Python大数据分析—南京财经大学

<https://www.icourse163.org/learn/NJUE-1458311167?tid=1458808445#/learn/content?type=detail&id=1235401002&cid=1255088001&replay=true>

b站视频：<https://www.bilibili.com/video/BV1Ae411W7HA>

知网

注意点：我是借用了华南师范大学的同学的学生账号登入知网查阅资料。所以依这个链接进入需要登录才能显示。

文献地址：

1.【中国刑事警察学院—杨众】基于Python语言的招聘信息可视化分析

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637294710265220000&DBCODE=CJFQ&TABLEName=CJFDLAST2020&FileName=JSYW202002061&RESULT=1&SIGN=eaqDSBEM5fNivJTsrmnDxMYTUUU%3d>

2.【厦门医学院现代教育与信息技术中心—陈琛 方玉华】基于网络招聘信息的医药卫生类高职高专毕业生就业市场分析

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637294715467563750&DBCODE=CJFQ&TABLEName=CJFDLAST2020&FileName=YXSH202001035&RESULT=1&SIGN=wmbxxyyesvfR3bwhWqx56FyxZ60%3d>

3.【大连交通大学软件学院-殷丽凤 —张浩然】基于Python网上招聘信息的爬取和分析

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637294718532720000&DBCODE=CJFQ&TABLEName=CJFDLAST2019&FileName=GWDZ201920005&RESULT=1&SIGN=Cl5h2BNiV91JNK99W939jEK8jzY%3d>

4.【厦门大学嘉庚学院—朱思霖 郭丽清】基于前程无忧招聘信息的数据人才需求分析

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637294724537407500&DBCODE=CJFQ&TABLEName=CJFDLAST2019&FileName=WLWJ201908044&RESULT=1&SIGN=IhVdO6%2fcIOV0pzUhiunT6ynEH00%3d>

5.【长沙民政职业技术学院软件学院—王涛】基于Python的软件技术人才招聘信息分析与实现——以前程无忧为例

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637295554211178750&DBCODE=CJFQ&TABLEName=CJFDLAST2018&FileName=FJDN201811059&RESULT=1&SIGN=YgULuh94iSRpq%2fIg%2b0%2fSk3C7ZxI%3d>