

爬完实习僧网站之后的一些简单分析

笔记本： 数据挖掘

创建时间： 2017/8/13 14:57

更新时间： 2017/8/13 16:05

作者： sunjune@qq.com

标签： python, 爬虫

因为想看一看，现在的互联网行业，招聘实习僧都是什么要求，以及这些企业的一些特点，我写了个爬虫，爬取了当前比较大且用户较多的实习信息发布网站--实习僧上所有互联网相关的实习信息。希望得到以下的一些信息：

- 企业的城市分布、薪资
- 实习的公司、薪资
- 各种类型的岗位数量
- 对学校学历专业的要求
- 编程语言的要求
- 对工具的要求
- 对个人品质的要求

爬虫的代码、原始数据，以及一些描述，放在了个人的Github上，欢迎使用和交流。地址：<https://github.com/sunjune/internshipAnalysis>

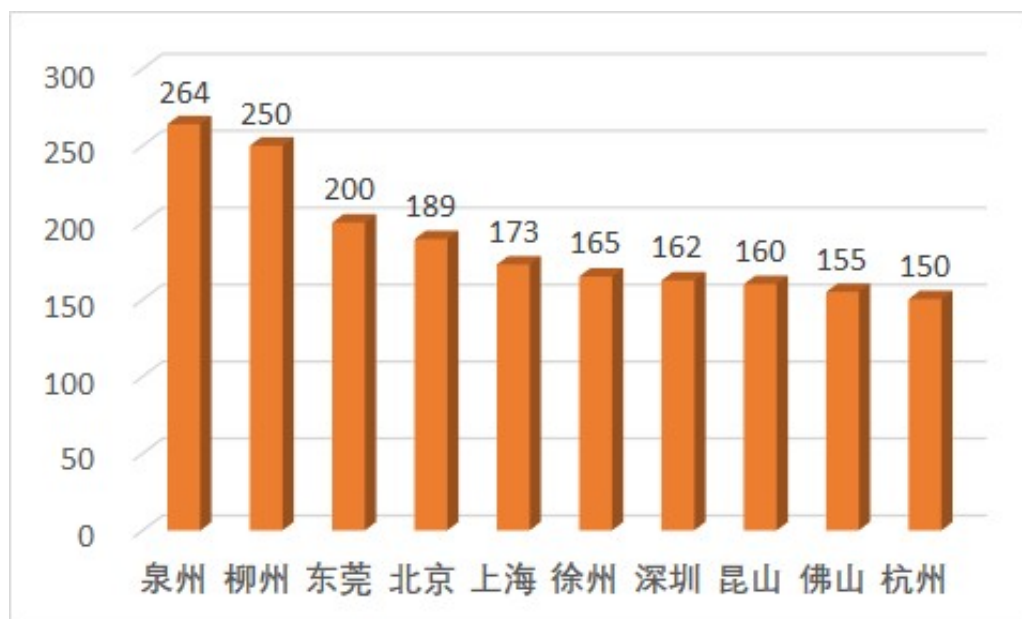
对数据进行分析，使用的主要是mysql和python，可视化用的excel。

* 不同城市的岗位数、平均日薪资



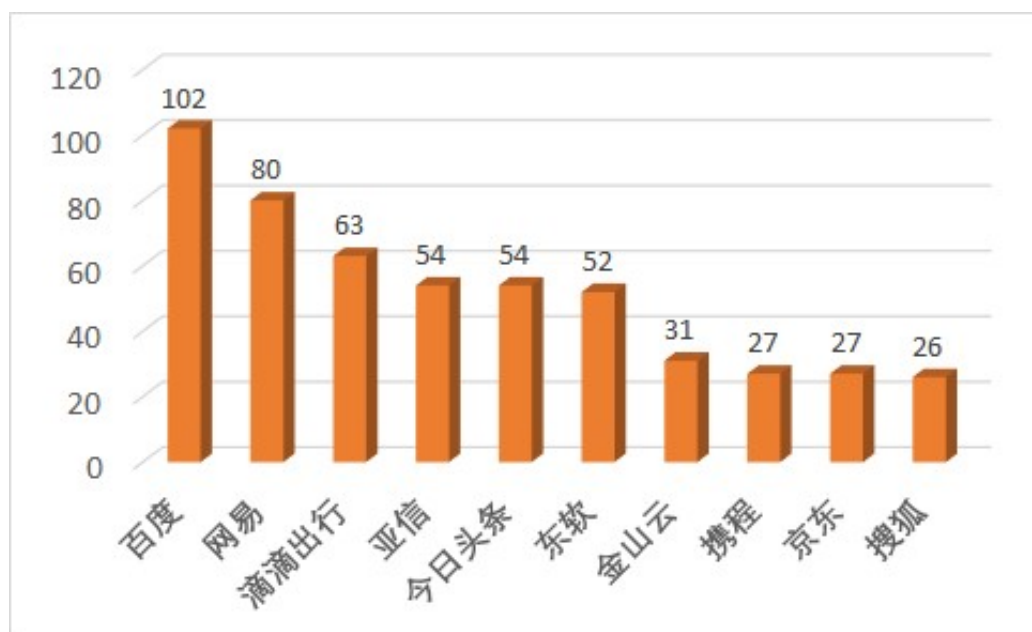
我们选取了实习岗位最多的十大城市，可以看到提供实习岗位数目最多的还是北上广深四大城

市，然而北京的实习数量，比上海广州深圳的总和还要多。作为二线城市代表的成都、杭州也有较多的实习岗位。



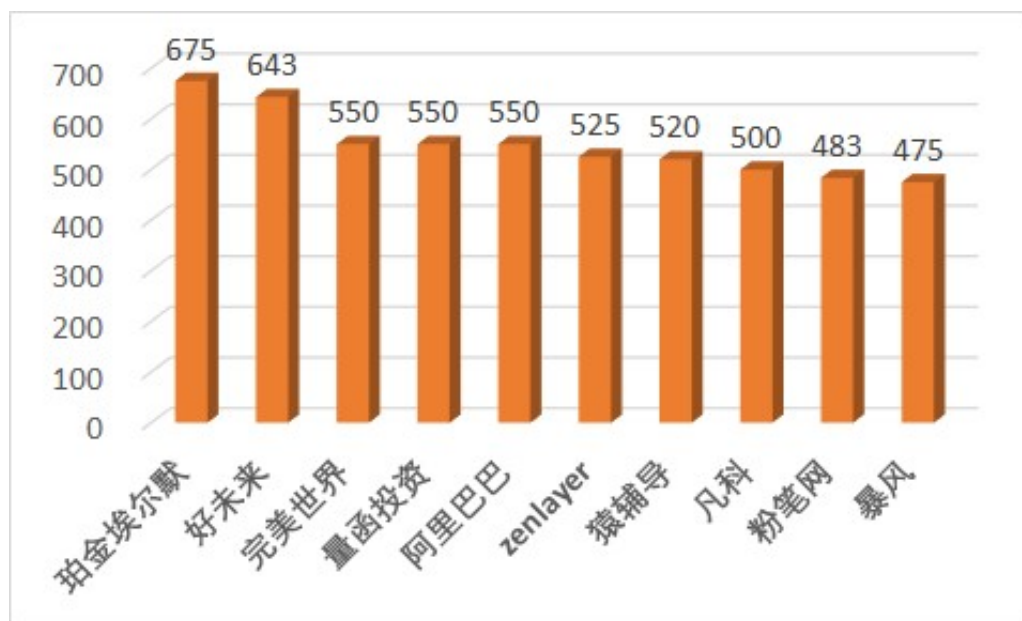
对于不同城市的平均日薪资水平，统计出来的结果，让人有些意外：泉州、柳州、东莞三个城市的平均实习薪资，排名前三位，接下来才是北京、上海、深圳这些互联网分布密集的城市。总体上来说，实习的薪资相差并不大，实际上80%的城市，平均日薪资都在100元以上，这从一个侧面说明，互联网的实习待遇还是不错的。

* 不同公司提供的岗位数、平均日薪资



这里选取的是实习僧网站上，发布实习岗位数量前10的公司及其提供的岗位数，百度、网易、滴滴占据前三，头条、携程、京东、搜狐也出现在了前10榜单中，很奇怪没有见到BAT的另外两大巨头腾讯和阿里巴巴，可能这也跟不同公司发布信息的习惯有关系吧，可能他们就是不在这个

平台发布信息呢~



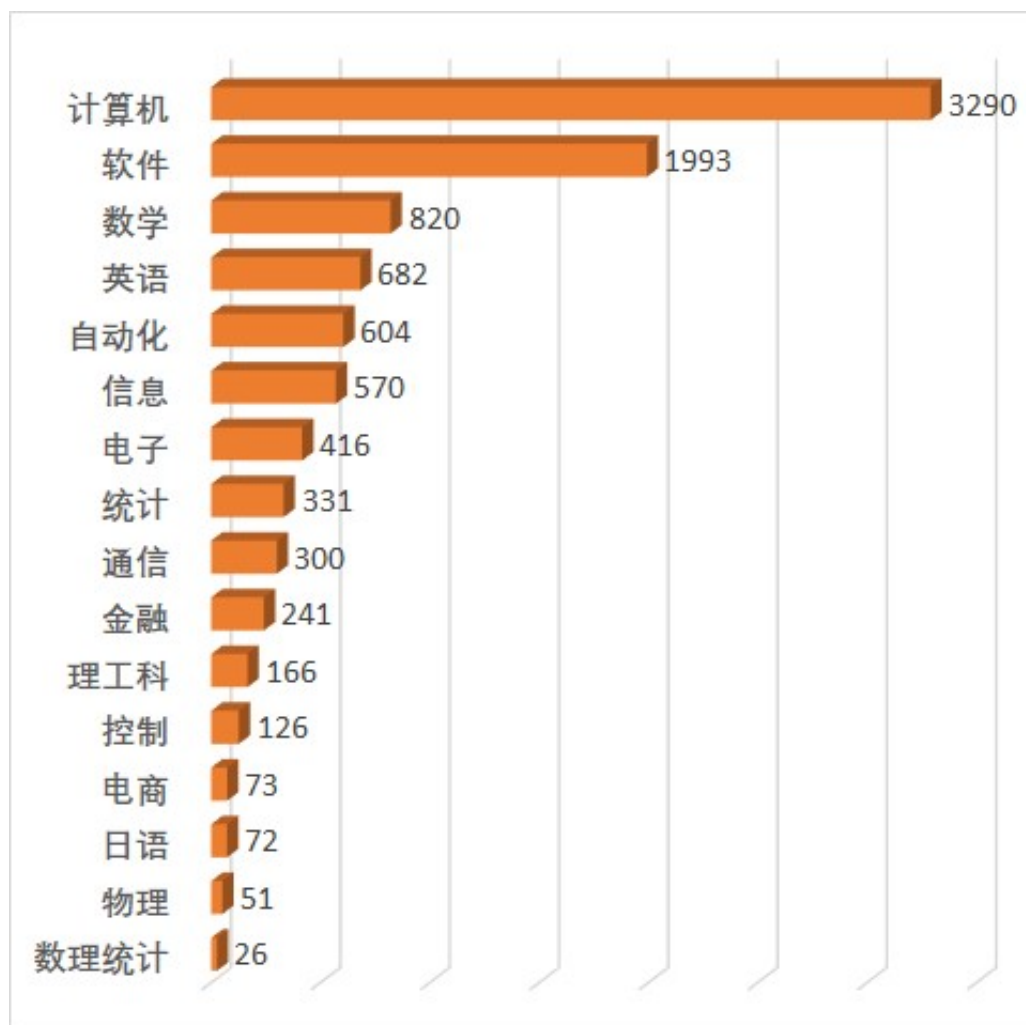
日薪资排名前几位的公司，基本都没有听说过，铂金埃尔默是做化学生物仪器的，好未来是做教育的（那个学而思配有就是他家的...），完美世界是做游戏的（比如诛仙这款游戏）。前10中有3家做教育，2家网络服务，1家做游戏，1家电商，1家金融，1家视频服务，1家做仪器设备。所以，什么行业赚钱，大家心里有数了.....当然中国家长的钱最好赚了....

* 不同岗位实习数量

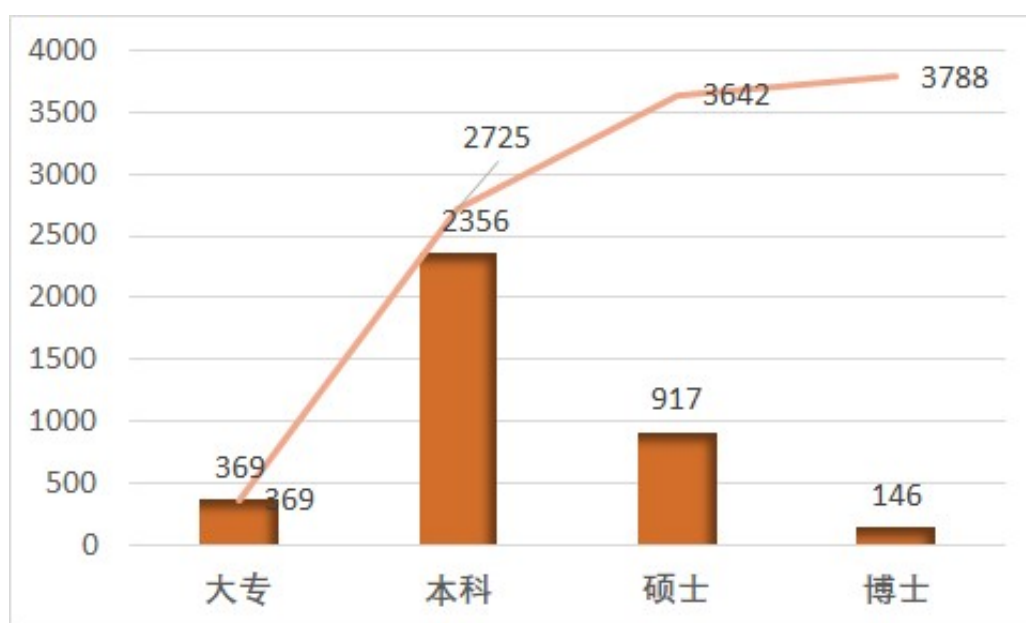


因为没有一个标准来对这些岗位进行分类，所以难免存在一些类别的交叉，但是从上面的结果来看，开发（程序员）必须是互联网的主流，然后是产品、测试、维护之类的岗位。具体细分开发岗位，可以看到前端的需求量是最大的，其次是算法、web、数据挖掘、数据库、机器学习、计算机视觉、游戏、网络、安卓、自然语言、ios、后台。总结起来就是，大部分人都去写网页、写算法了.....

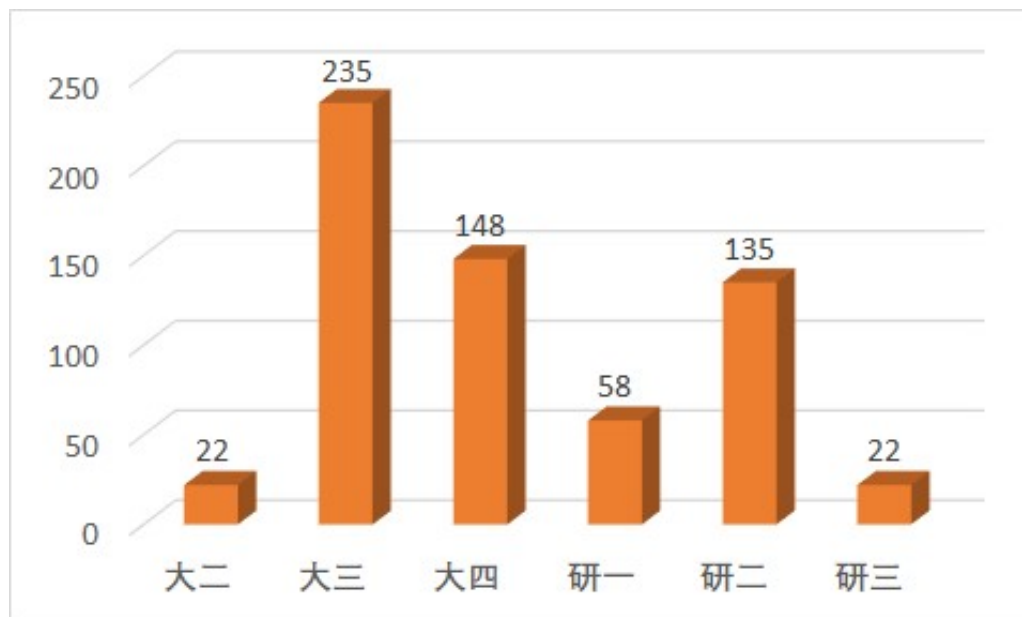
* 学校学历专业



综合起来，最受青睐的还是计算机、软件、数学这些专业的毕业生，当然互联网行业对专业的限制实际上并不严格，毕竟大家还是靠技术来吃饭的。

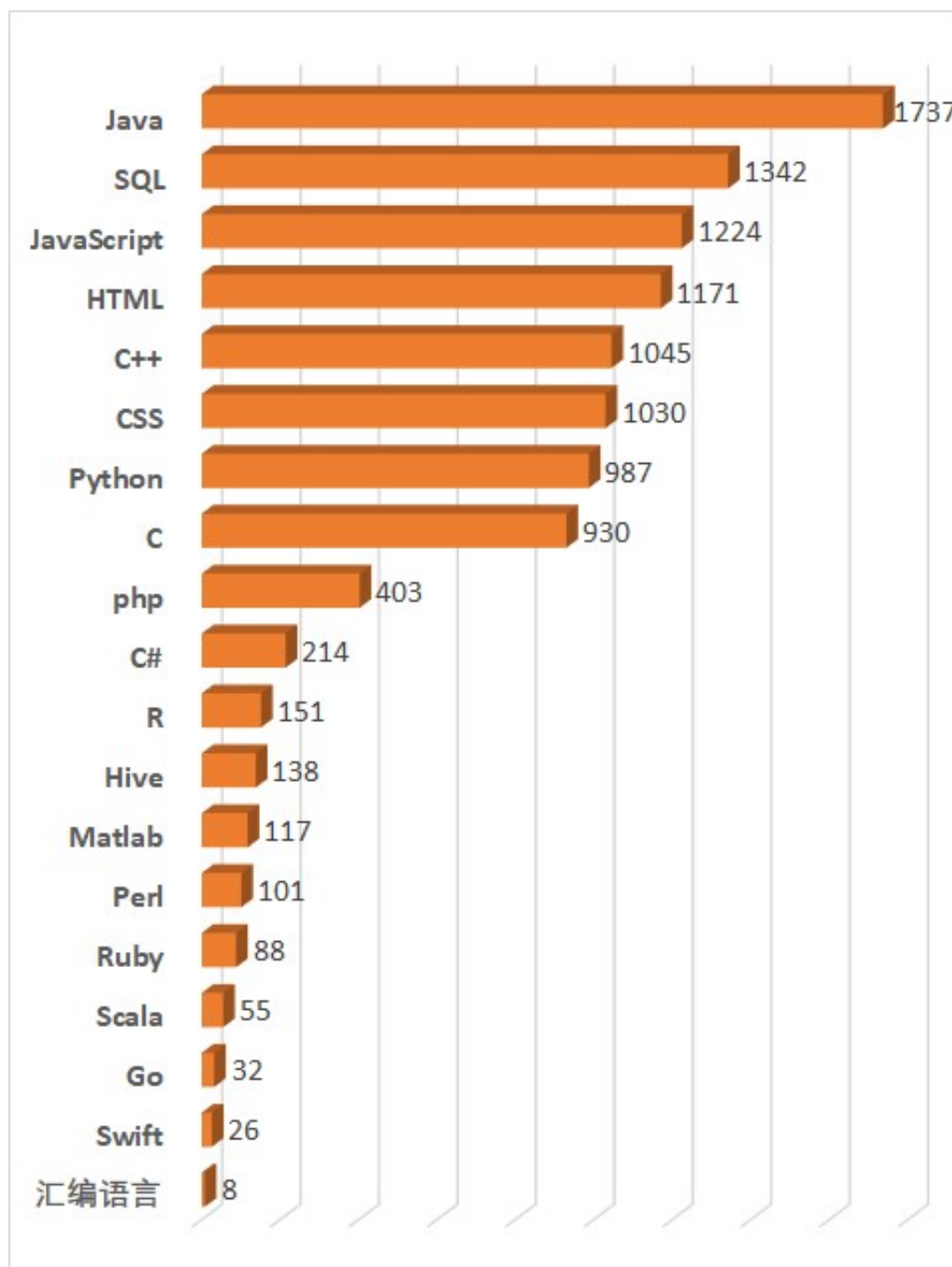


大部分实习，都要求本科学历，其次是硕士，所有，相对而言，进入互联网行业工作，本科学历基本足够了，当然硕士的起点会更高，对于博士学历，要求的并不多。



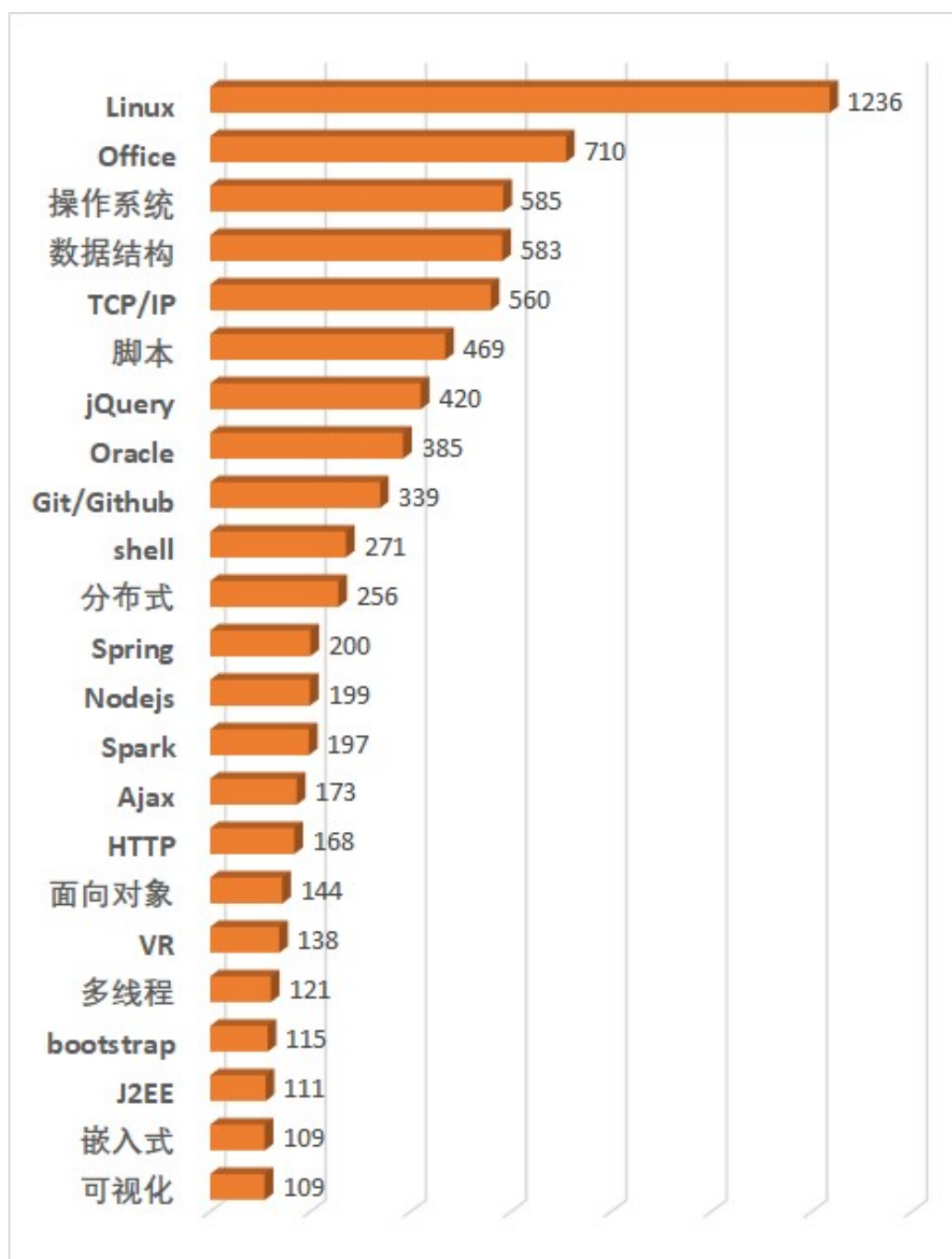
那对于学生年级的要求呢，本科生普遍要求大三、大四，研究生则是研一、研二，这反映了招聘的公司，实际上是想招聘校招的储备力量。

* 编程语言要求



从前面的岗位需求中，前端岗位需求旺盛，就不奇怪SQL、HTML、CSS、JavaScript这些语言的要求数排名靠前了。传统的Java、C++、Python、C、C#这五大金刚，除去前端和数据库语言之后，基本占据了主要地位。所以有时间的话，学一学前端还有数据库吧，毕竟容易找工作，Java、C++二选一，python和c简单，必会。

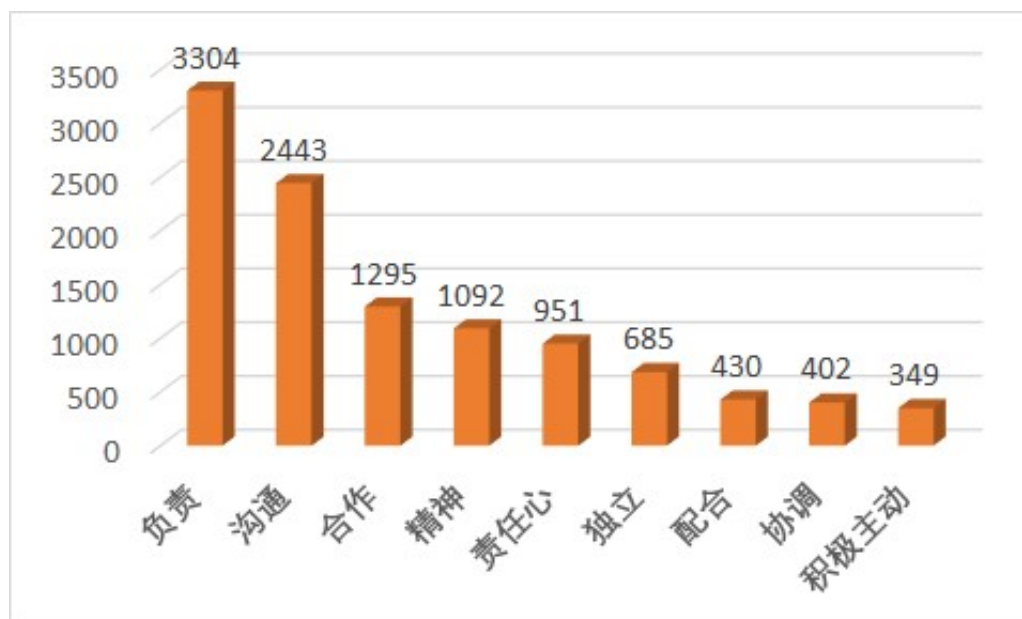
* 工具要求





对于工具这一类，具体包括编程工具、知识、框架之类的关键词，上面是要求最后的工具排行，基本所有的岗位都要求会使用Linux环境，office不知道是怎么乱入的，数据结构、操作系统、计算机网络这些基础知识需要了解，然后是脚本语言/shell，还有版本控制工具Git/Github的使用。

* 个人品质要求



这些没啥要说的，老生常谈了.....

总的来说，分析的结果还是比较符合我们的实际认知的，希望能对大家择业有所帮助。