

# 15分钟，教你用Python爬网站数据，并用BI可视化分析！

作为一名在数据行业打拼了两年多的数据分析师，虽然目前收入还算ok，但每每想起房价，男儿三十还未立，内心就不免彷徨不已~

两年时间里曾经换过一份工作，一直都是从事大数据相关的行业。目前是一家企业的BI工程师，主要工作就是给业务部门出报表和业务分析报告。

回想自己过去的工作成绩也还算是不错的，多次通过自己分析告，解决了业务的疑难杂症，领导们各种离不开。

但安逸久了总会有点莫名的慌张，所以我所在的这个岗位未来会有多大发展空间，十年之后我能成为什么样的人呢？自己的收入空间还有多少？

一番惆怅之后，别再问路在何方了，于是抄起自己的“家伙”，花了一小会时间爬了智联招聘上BI岗位的数据信息，做了个分析。

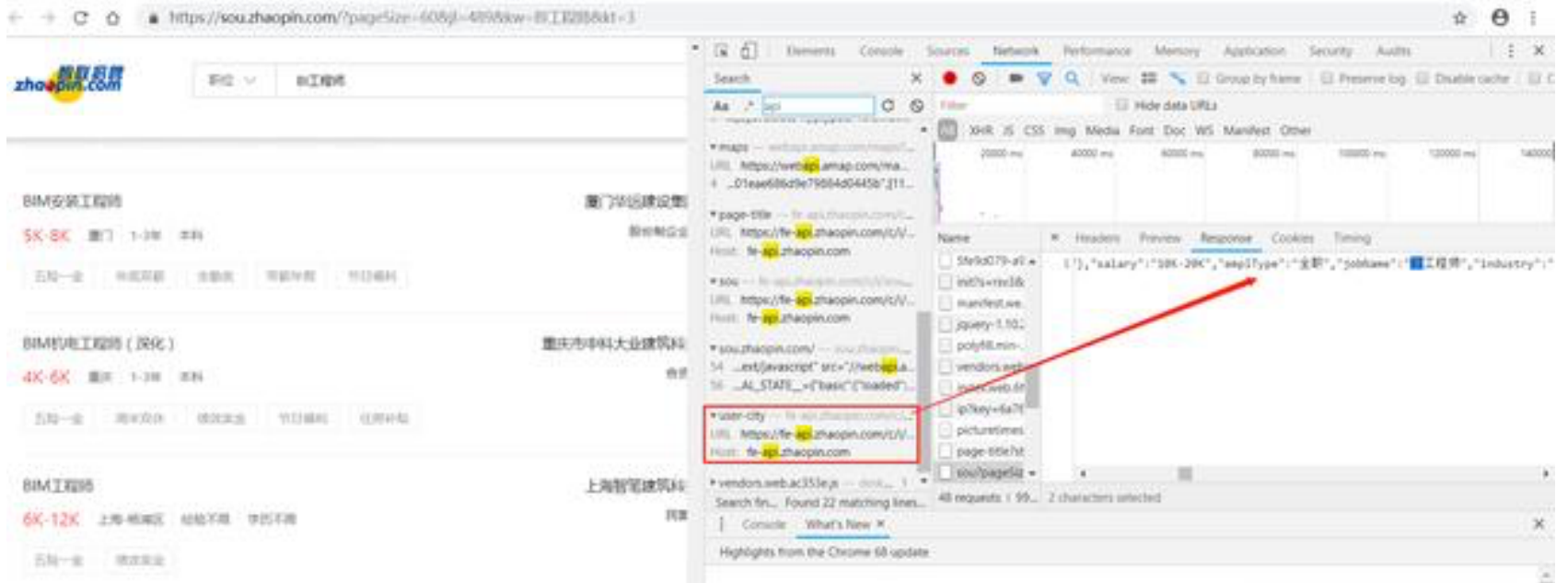
PS：所用工具为Python+BI

数据分析的过程如同烧一顿饭，先要数据采集（买菜），然后数据建模（配菜）、数据清洗（洗菜）、数据分析（做菜）、数据可视化（摆盘上菜）。

所以第一步，要采集/选择数据。

## 一、Python爬取智联招聘岗位信息（附源码）

选择智联招聘，通过Python来进行“BI工程师”的关键数据信息的爬取，这里大家也可以试着爬取自己岗位的关键词，如“数据分析师”、“java开发工程师”等。经过F12分析调试，数据是以JSON的形式存储的，可以通过智联招聘提供的接口调用返回。



那么我这边通过Python对智联招聘网站的数据进行解析，爬取了30页数据，并且将岗位名称、公司名称、薪水、所在城市、所属行业、学历要求、工作年限这些关键信息用CSV文件保存下来。

## 附上完整Python源码：

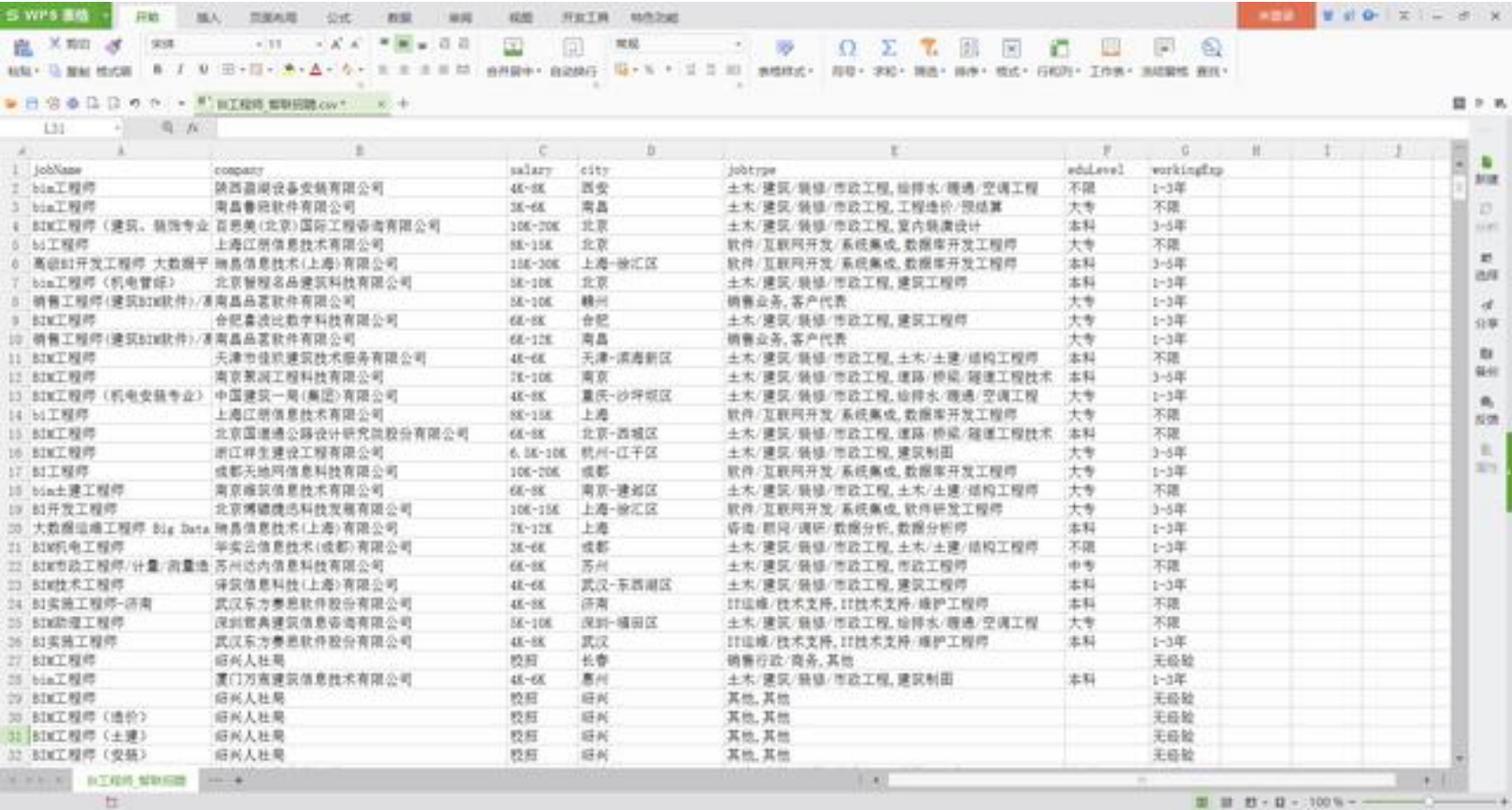
```
import requests
import json
import csv
from urllib.parse import urlencode
import time

def saveHtml(file_name,file_content): #保存content对象为html文件
    with open(file_name.replace('/', '_')+'.html','wb') as f:
        f.write(file_content)

def GetData(url,writer):#解析并将数据保存为csv文件
    response= requests.get(url)
    data=response.content
    saveHtml('zlpz',data) #保存html文件
    jsondata=json.loads(data)
    dataList=jsondata['data']['results']
    #print(jsondata)
    for dic in dataList:
        jobName=dic['jobName'] #岗位名称
        company=dic['company']['name'] #公司名称
        salary=dic['salary'] #薪水
        city=dic['city']['display'] #城市
        jobtype = dic['jobType']['display'] #所属行业
        eduLevel=dic['eduLevel']['name'] #学历要求
        workingExp=dic['workingExp']['name'] #工作经验
        print(jobName,company,salary,city,jobtype,eduLevel,workingExp)
```

```
writer.writerow([jobName,company,salary,city,jobtype,eduLevel,workingExp])
param={ 'start':0,
        'pageSize':60,
        'cityId':489,
        'workExperience':-1,
        'education':-1,
        'companyType': -1,
        'employmentType': -1,
        'jobWelfareTag': -1,
        'kw': 'BI工程师', #搜索关键词, 可以根据你需要爬取的岗位信息进行更换
        'kt': 3,
        'lastUrlQuery': {"p":1,"pageSize":"60","jl":"681","kw":"python","kt":"3"}
    }#参数配置
pages=range(1,31)#爬取1-30页数据
out_f = open('test.csv', 'w', newline='')
writer = csv.writer(out_f)
writer.writerow(['jobName','company','salary','city','jobtype','eduLevel','workingExp'])
for p in pages: #自动翻页
    param['start']=(p-1)*60
    param['lastUrlQuery']['p']=p
    url = 'https://fe-api.zhaopin.com/c/i/sou?' + urlencode(param)
    GetData(url,writer)
    time.sleep(3)#间隔休眠3秒, 防止IP被封
    print(p)
out_f.close()
```

经过一番编译调试，代码成功运行。



jobName	company	salary	city	jobtype	eduLevel	workingExp
bia工程师	陕西富源设备安装有限公司	4K-8K	西安	土木/建筑/装修/市政工程, 给排水/暖通/空调工程	不限	1-3年
bim工程师	南昌睿源软件有限公司	3K-6K	南昌	土木/建筑/装修/市政工程, 工程造价/预结算	大专	不限
BIM工程师 (建筑、装修专业)	百思美(北京)国际工程咨询有限公司	10K-20K	北京	土木/建筑/装修/市政工程, 室内装潢设计	本科	3-5年
bi工程师	上海江明信息技术有限公司	8K-15K	北京	软件/互联网开发/系统集成, 数据库开发工程师	大专	不限
高级BI开发工程师 大数据平	唯嘉信息技术(上海)有限公司	15K-30K	上海-徐汇区	软件/互联网开发/系统集成, 数据库开发工程师	本科	3-5年
bim工程师 (机电管综)	北京智程名品建筑科技有限公司	5K-10K	北京	土木/建筑/装修/市政工程, 建筑工程师	本科	1-3年
销售工程师(建筑BIM软件)/	南昌品茗软件有限公司	5K-10K	赣州	销售业务, 客户代表	大专	1-3年
BIM工程师	合肥嘉波比数字科技有限公司	6K-8K	合肥	土木/建筑/装修/市政工程, 建筑工程师	大专	1-3年
销售工程师(建筑BIM软件)/	南昌品茗软件有限公司	6K-12K	南昌	销售业务, 客户代表	大专	1-3年
BIM工程师	天津市佳联建筑技术服务有限公司	4K-6K	天津-滨海新区	土木/建筑/装修/市政工程, 土木/土建/结构工程师	本科	不限
BIM工程师	南京聚河工程科技有限公司	7K-10K	南京	土木/建筑/装修/市政工程, 道路/桥梁/隧道工程技术	本科	3-5年
BIM工程师 (机电安装专业)	中国建筑一局(集团)有限公司	4K-8K	重庆-沙坪坝区	土木/建筑/装修/市政工程, 给排水/暖通/空调工程	大专	1-3年
bi工程师	上海江明信息技术有限公司	8K-15K	上海	软件/互联网开发/系统集成, 数据库开发工程师	大专	不限
BIM工程师	北京国通公路设计研究院股份有限公司	6K-8K	北京-西城区	土木/建筑/装修/市政工程, 道路/桥梁/隧道工程技术	本科	不限
BIM工程师	浙江祥生建设工程有限公司	6.5K-10K	杭州-江干区	土木/建筑/装修/市政工程, 建筑制图	大专	3-5年
bi工程师	成都天地网络信息科技有限公司	10K-20K	成都	软件/互联网开发/系统集成, 数据库开发工程师	大专	1-3年
bi+土建工程师	南京维筑信息技术有限公司	6K-8K	南京-建邺区	土木/建筑/装修/市政工程, 土木/土建/结构工程师	大专	不限
bi开发工程师	北京博德德玛科技发展有限公司	10K-15K	上海-徐汇区	软件/互联网开发/系统集成, 软件研发工程师	大专	3-5年
大数据运维工程师 Big Data	唯嘉信息技术(上海)有限公司	7K-12K	上海	咨询/顾问/调研/数据分析, 数据分析师	本科	1-3年
BIM机电工程师	华安信息技术有限公司(成都)	3K-6K	成都	土木/建筑/装修/市政工程, 土木/土建/结构工程师	不限	1-3年
BIM市政工程师/计量/测量造	苏州达内信息科技有限公司	6K-8K	苏州	土木/建筑/装修/市政工程, 市政工程师	中专	不限
BIM技术工程师	保筑信息科技有限公司	4K-6K	武汉-东西湖区	土木/建筑/装修/市政工程, 建筑工程师	本科	1-3年
bi实施工程师-济南	武汉东方泰恩软件股份有限公司	4K-8K	济南	IT运维/技术支持, IT技术支持/维护工程师	本科	不限
BIM助理工程师	深圳智典建筑信息咨询有限公司	5K-10K	深圳-福田区	土木/建筑/装修/市政工程, 给排水/暖通/空调工程	大专	不限
bi实施工程师	武汉东方泰恩软件股份有限公司	4K-8K	武汉	IT运维/技术支持, IT技术支持/维护工程师	本科	1-3年
BIM工程师	绍兴人社局	校招	绍兴	销售行政/商务, 其他	无经验	
bim工程师	厦门万商建筑信息技术有限公司	4K-6K	惠州	土木/建筑/装修/市政工程, 建筑制图	本科	1-3年
BIM工程师	绍兴人社局	校招	绍兴	其他, 其他	无经验	
BIM工程师 (造价)	绍兴人社局	校招	绍兴	其他, 其他	无经验	
BIM工程师 (土建)	绍兴人社局	校招	绍兴	其他, 其他	无经验	
BIM工程师 (安装)	绍兴人社局	校招	绍兴	其他, 其他	无经验	

全部数据爬取完毕，一共1800条，保存在本地CSV文件中。

数据是爬到了，具体我想了解哪些信息呢：各城市的BI岗位需求情况以及薪资水平；薪水随工作经验的涨幅情况，以及有哪些具体的高薪岗。

由此可见，想要分析的角度很多，且看了源数据，还要做不少的数据处理。最简单快速出可视化的方法自然是用BI工具，来对数据做简单清洗加工，并呈现可视化。

BI能应付绝大多数场景的数据分析，尤其擅长多维数据切片，不需要建模；甚至数据清洗环节也能放在前端，通过过滤筛选、新建计算公式等来解决。最后呈现可视化，并可设计数据报告。

这里我用FineBI来做这样一份分析。

FineBI做分析大体是这样的流程：连接/导入数据——数据处理/清洗（过滤、筛选、新增公式列）——探索式分析——数据可视化——出报告。

## 二、数据清洗加工

### 1.薪水上下限分割：

将CSV文件数据导入FineBI中（新建数据链接，建立一个分析业务包，然后导入这张excel表）。因为薪水是以xxK-xxk（还有一些类似校招/薪资面议的数据）的形式进行存储的，我这边使用FineBI新增公式列（类似excel函数）将这些字符进行分割：

薪水下限（数值）：`left( indexofarray( split( salary,"-" ) ,1),find("K",INDEXOFARRAY( split(salary,"-" ) ,1))-1)`

薪水上限（含K字符）：`right( indexofarray( split(salary,"-" ) ,2),len(salary)- find("K",indexofarray(split(salary,"-"),2 ) ) )`

薪水上限（数值）：`left( 薪水上限(文本),find("K",薪水上限(文本))-1 )`

这样就得到每个岗位的数值格式的薪水区间了：



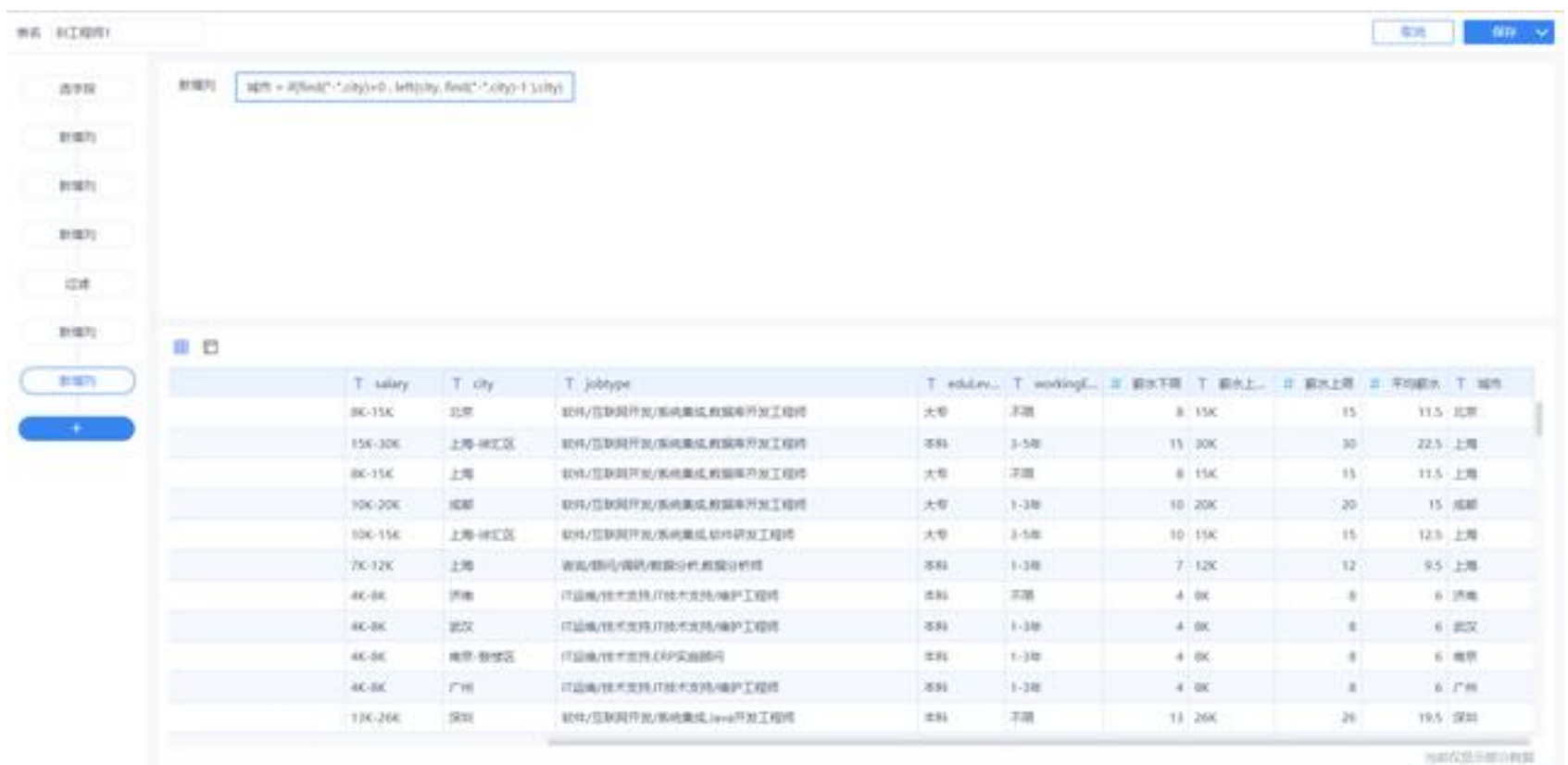




## 4.真实城市截取

由于城市字段存储有的数据为“城市-区域”格式，例如“上海-徐汇区”，为了方便分析每个城市的数据，最后新增列“城市”，截取“-”前面的真实城市数据。

城市：if(find("-",city)>0 , left(city, find("-",city)-1 ),city)



至此，18000多条数据差不多清洗完毕，食材已经全部准备好，下面可以正式开始数据可视化的美食下锅烹饪。

### 三、数据可视化

数据可视化可以说是很简单了，拖拽要分析的数据字段即可。

但是这里用finebi分析要理解一个思路。常规我们用excel做分析或者说做图表，是先选用钻则图表然后设定系列、数值。这里没有系列和数值的概念，只有横轴和竖轴。拖入什么字段，该字段就以该轴进行扩展，至于图表嘛，finebi会自动判别推荐。

我这边以各城市平均薪水/岗位数量分析为例给大家简单展示FineBI的可视化呈现过程。

1、横轴以“城市”字段扩展，展现两类数据。先是薪水值，拖拽到纵轴，默认对数值类的字段是汇总求和的。点击字段可直接对改字段修改计算、过滤等操作。



此图来自官网，图中数据不是本次分析的数据，仅供参考

2、然后分析每个城市BI岗位的情况。将数据记录数这个指标拖入到纵轴展示。同样的方式，可以修改字段名。这里为了区分两者，将其修改为折线图，并且倒叙展示。







城市来看，国内排在前面的分别为深圳（14.72K）、上海（14.59K）、北京（14.51）、杭州（12.07K）、成都（11.13K）、广州（10.94K）。

3.从工作年限的平均薪水和岗位需求数量来看，工作5-10年的资深BI工程师的平均薪水可以达到20K以上（朝资深BI工程师方向奋斗！！！1年以下年限的计算出来的平均薪水虽然为19K，但是由于样本量只有3个，所以参考意义不大），其中大部分的工作需求年限为3-5年，平均薪水为14.24K。

4.从学历方面来看，最低学历要求主要以本科/大专为主，本科和大专学历要求的平均薪资分别为12.68K和11.97K（感觉差距并不大，过硬的技术实力可能才是企业最为看重的吧），博士和硕士学历要求很少。

5.看了一些高薪的招聘企业，最高的可以给到30K~40K的薪酬水平，其中主要是互联网、IT类公司为主。

醍醐灌顶，顿时有了奋斗的动力~知识就是财富，继续好好学习去吧，少年！！！！

**最后，附上本次爬取到的相关岗位明细的CSV数据（私心回复“csv”获取），如果大家也有分析自己岗位未来前景的想法，但是可能不熟悉数据可视化工具的，可以拿这份数据用FineBI练练手。**