

拉勾网数据分析职位

► 分析报告

报告人：瞿 晨
日 期：2018/03

目录

1 分析背景与目的

2 数据来源

3 图表分析

4 结论

5 附录

分析背景与目的

随着人工智能、大数据、云计算等领域的兴起，企业所需要处理的数据量逐渐增大，传统的数据处理方法渐渐力不从心，配合数据库、计算机编程的需求慢慢提高，数据分析/数据挖掘职位的需求量与日俱增。

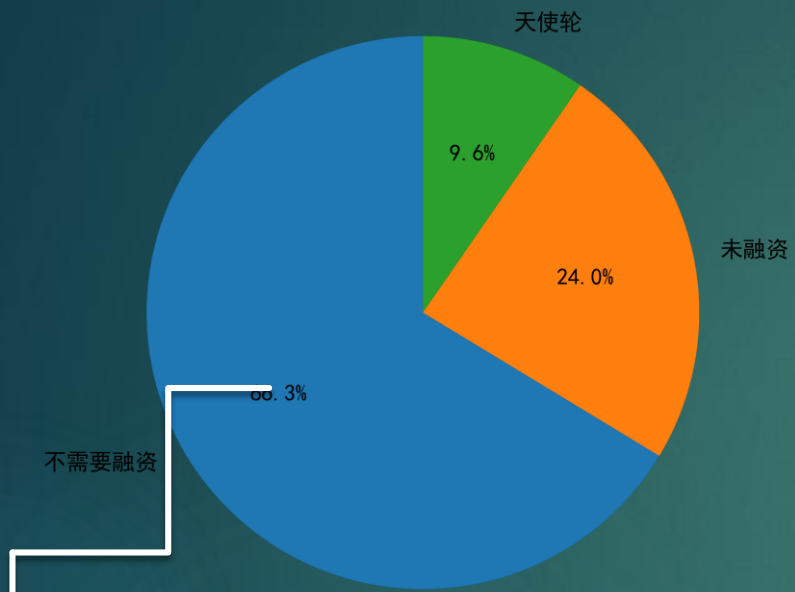
本篇报告旨在分析学历、城市、企业发展阶段、企业规模、薪资等特性对于数据分析职位的影响。

数据来源

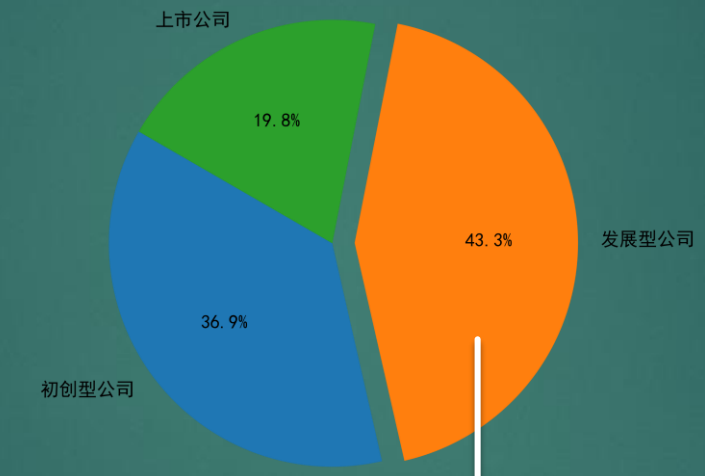
全部数据来源于拉勾网，使用Python Scrapy框架编写爬虫程序，获取拉勾网中北京、上海、深圳、广州、杭州、成都、武汉、南京、苏州、西安共十个站点的数据分析职位相关信息数据，共1883条。

拉勾网是一家专注互联网招聘的网站，超过10万家企业入驻。

图表分析——企业发展阶段分布

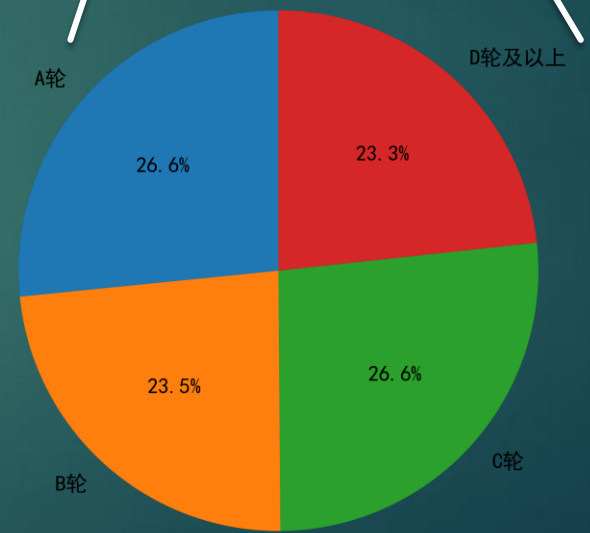


在初创型企业中，不需要融资的企业所占比重遥遥领先

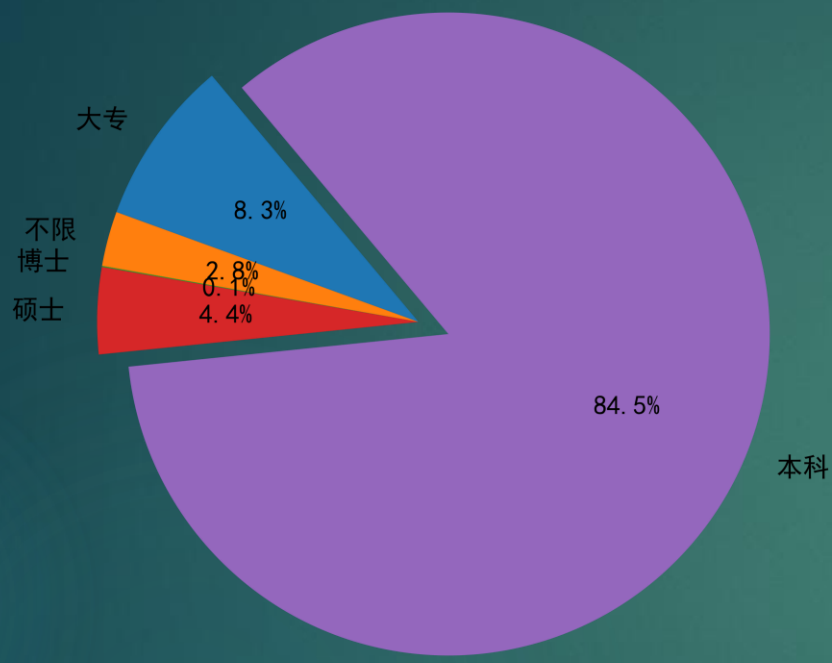


在全部企业中，发展型企业数量最多，上市公司数量最少

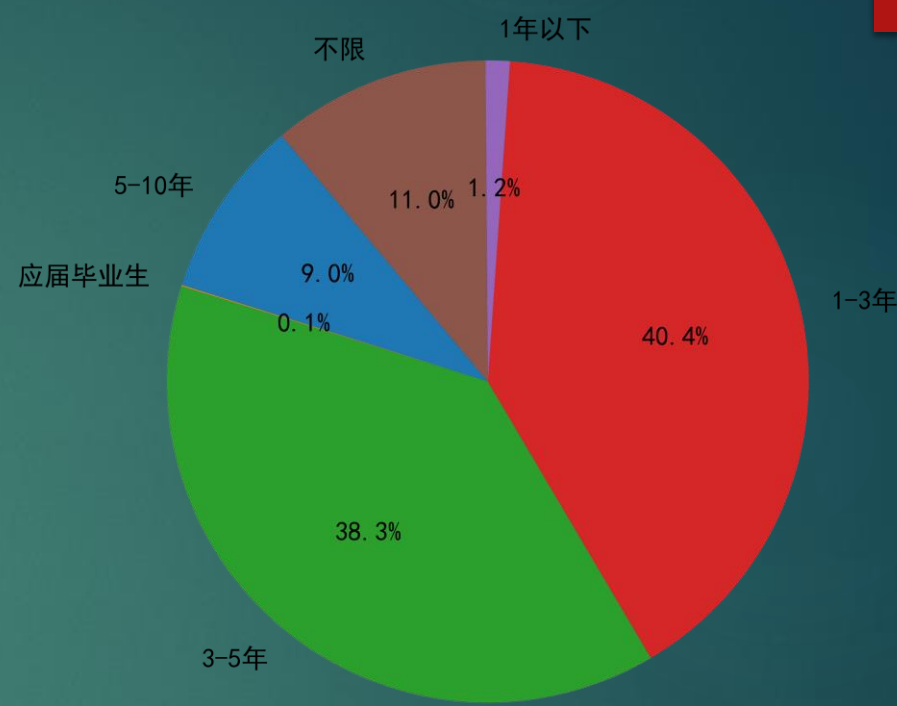
在发展型企业中，各轮融资阶段的数量相近，其中A、C轮要略多于B、D轮企业



图表分析——学历与工作年限分布

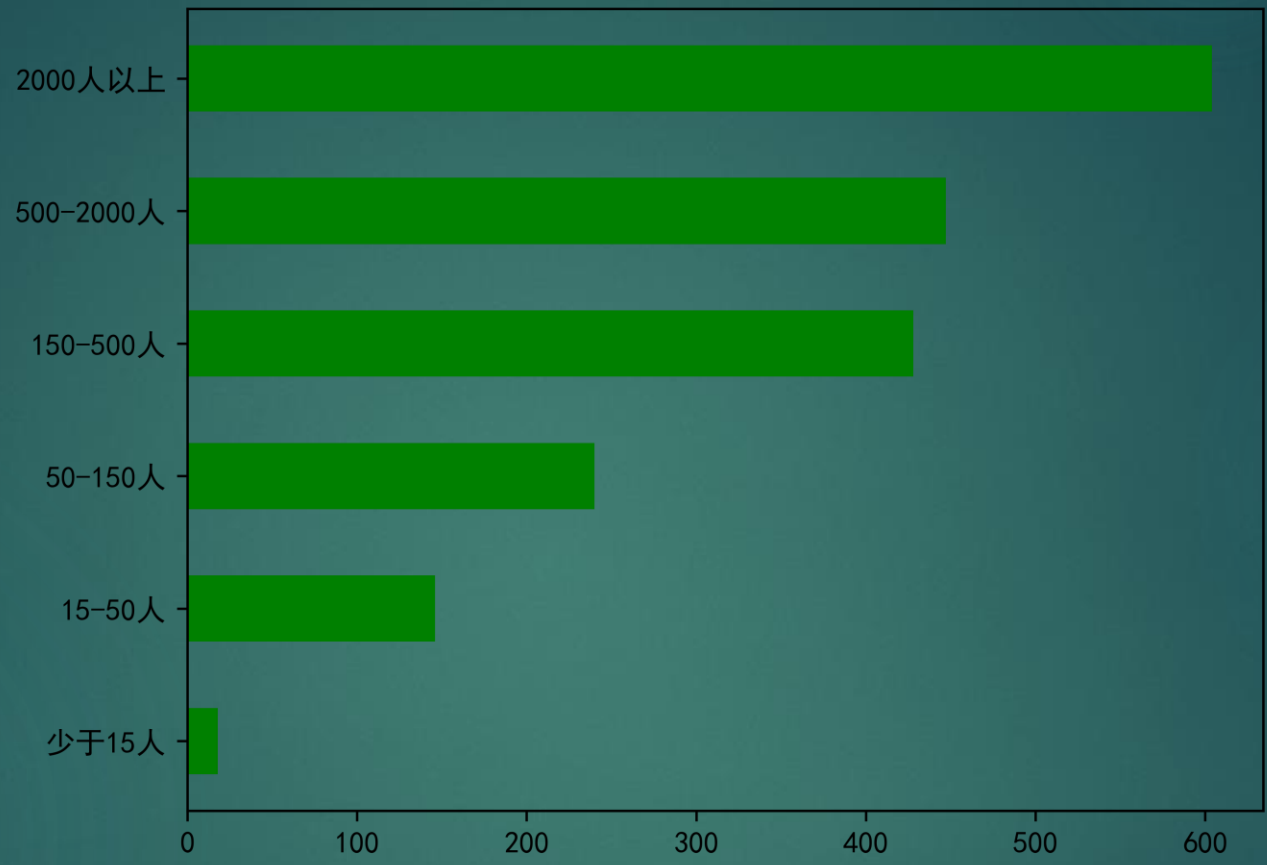


在数据分析职位中，学历要求本科的占有遥遥领先的地位84.5%，大多数企业要求本科，学历要求博士的企业低于0.1%，可以看出学历不会成为数据分析发展的天花板。



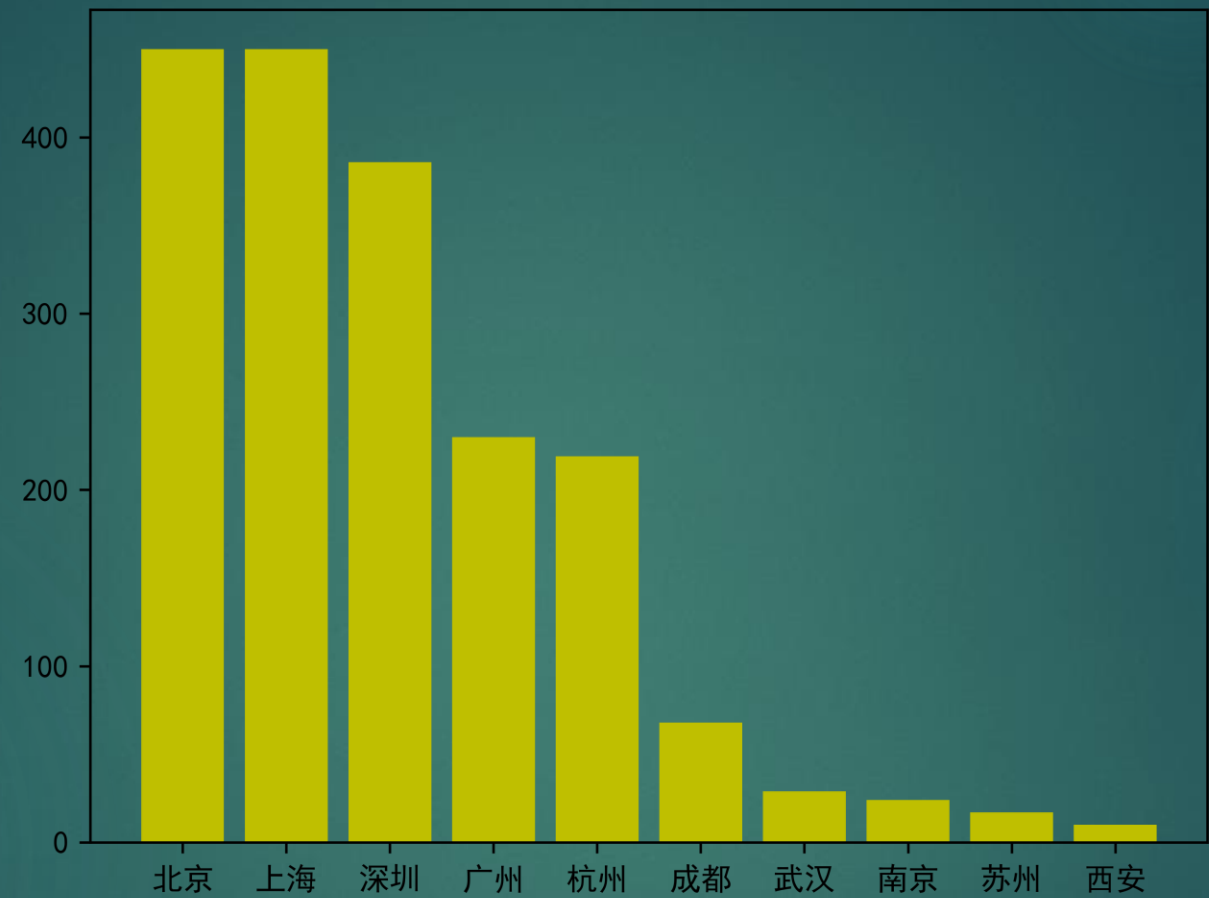
对工作年限的要求中，接近80%的企业需要1-3年和3-5年的员工，1年以下和应届生极少，而相对而言5-10年的需求量也同样并不是很高，仅为9%

图表分析——需求量与公司规模



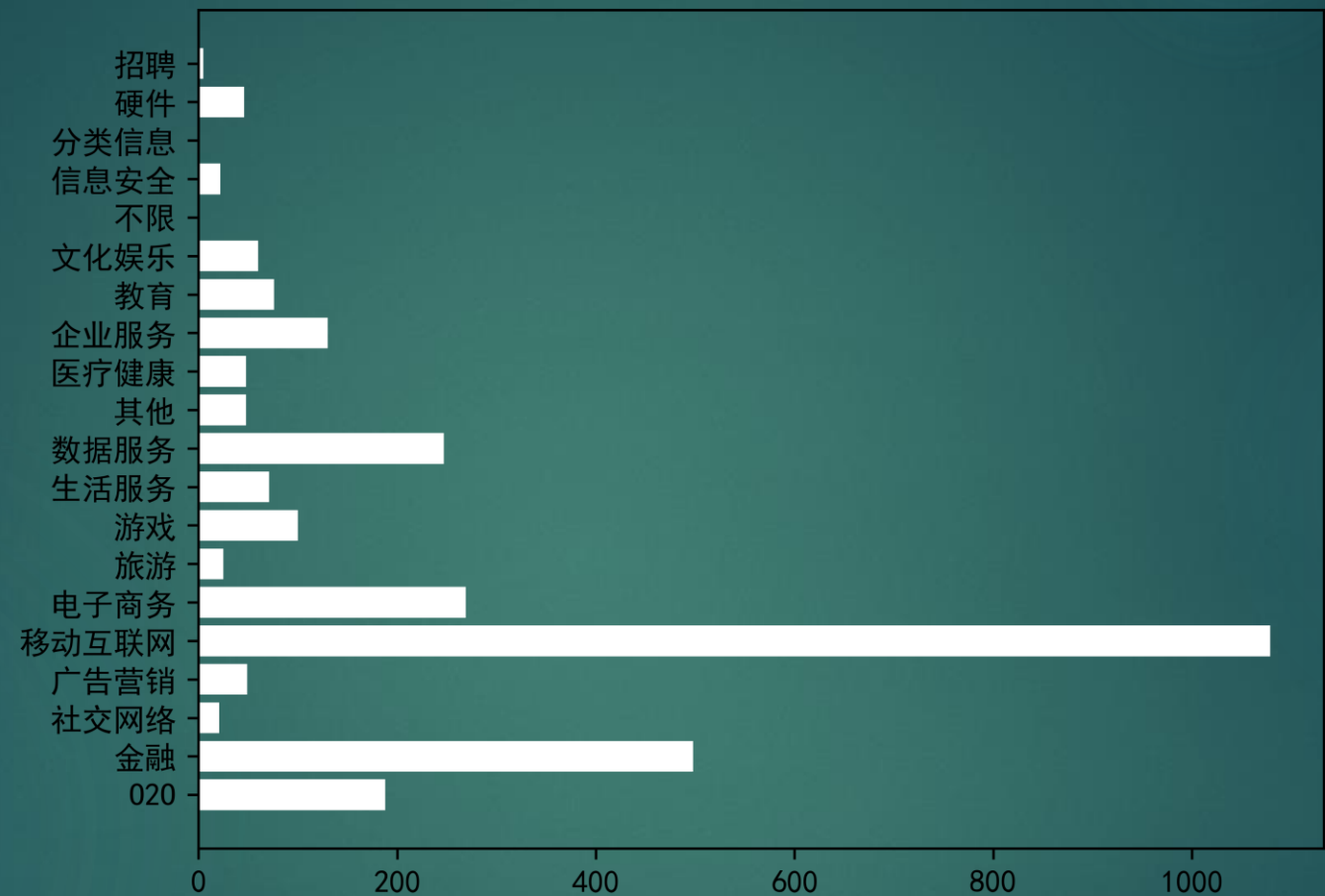
在需求数据分析职位的企业中，随着企业规模的增大，对于这一职位的需求程度也随之增加，图中可以看出需要数据分析的更多是大型企业，而小公司对这一职位的需求程度较低。

图表分析——需求量与城市



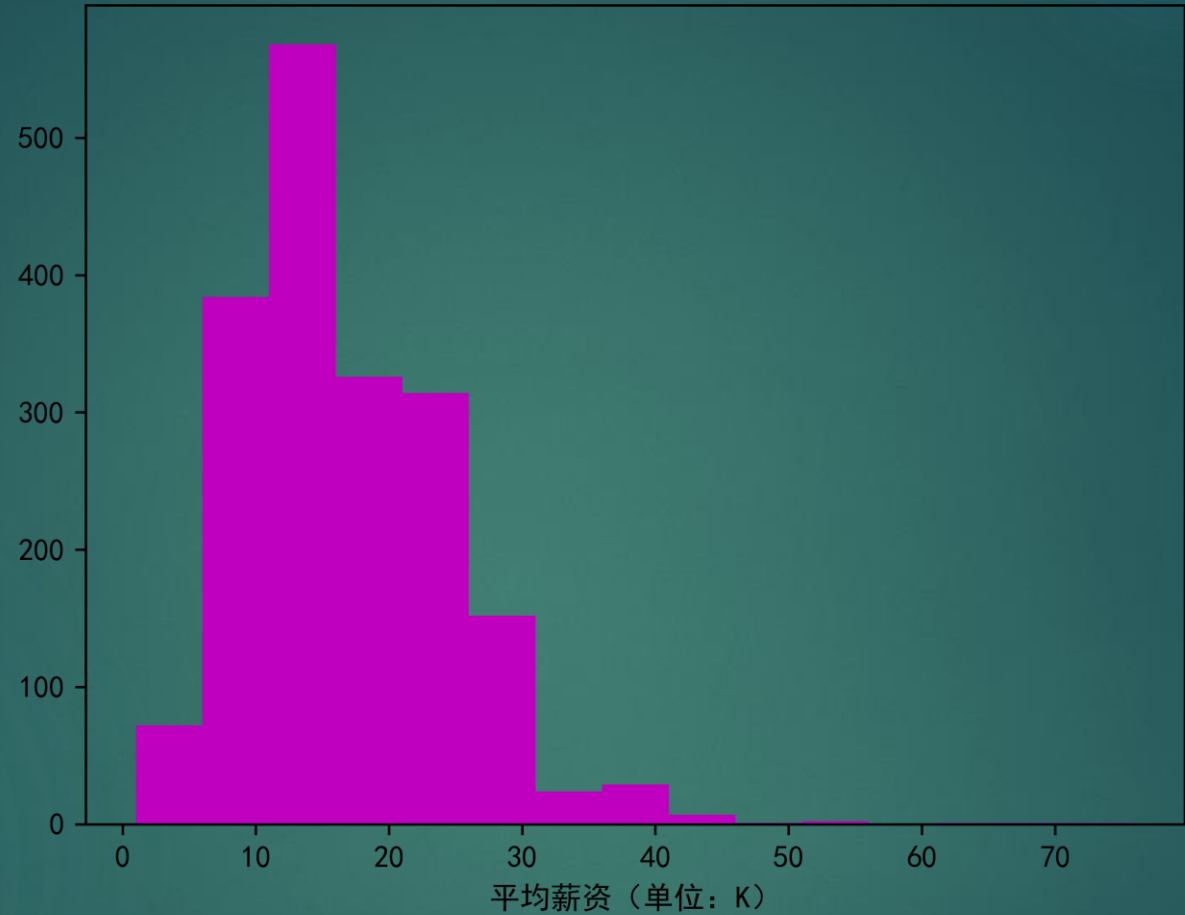
北京、上海、深圳对于数据分析岗位的需求量远远高于其他地区，尤其是北京、上海地区，其次是杭州和广州仅次于北上深地区，而除此以外的其他地区对于这一职位的需求则相对较低。

图表分析——需求量与行业领域



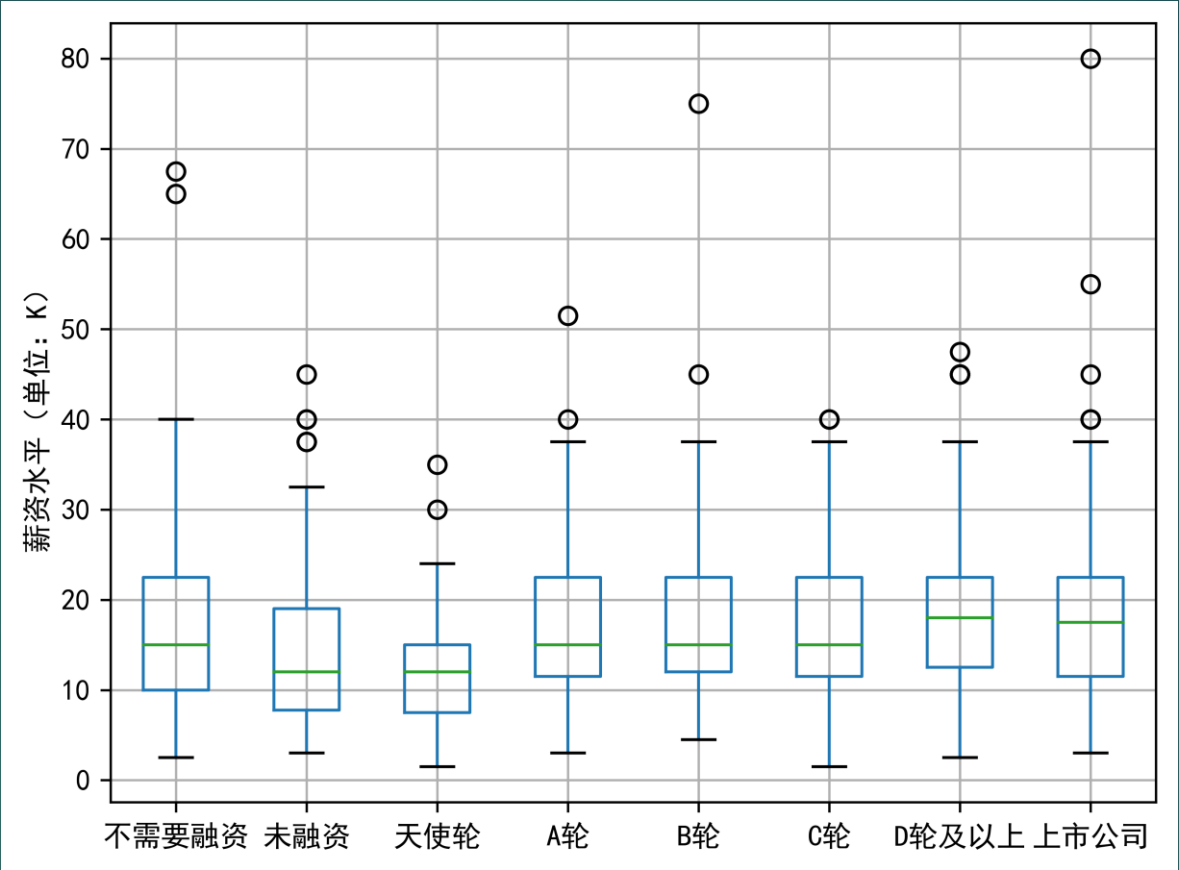
移动互联网企业对于数据分析职位的需求量远高于其他行业，而除此以外，金融、电子商务、数据服务这三大行业对于数据分析的需求也相对较大。

图表分析——薪资水平分布情况



数据分析职位的薪资水平大多数集中在5K-25K之间，高于或低于这个水平的数量只有总数的11.6%，数量最多的薪资水平是10K-15K之间，这是目前数据分析职位的主流薪资。

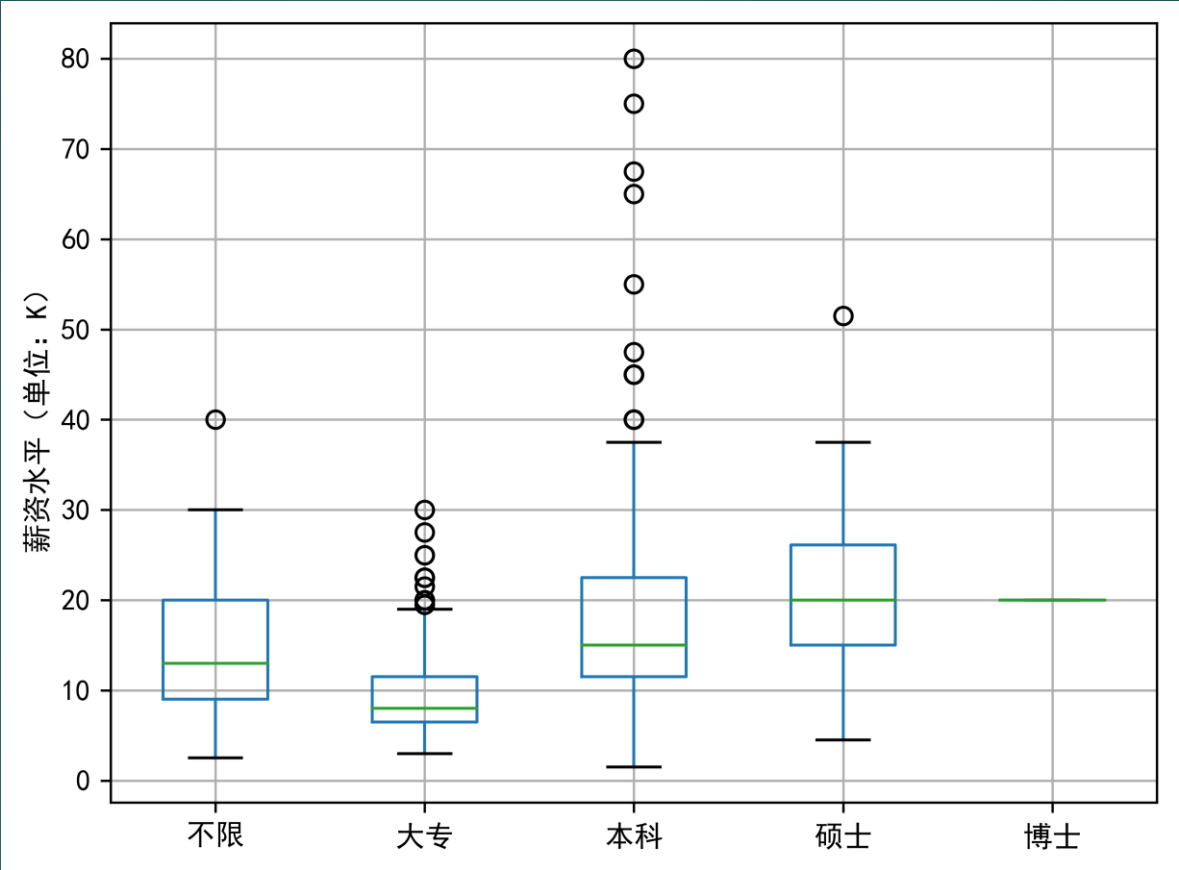
图表分析——薪资在不同企业发展阶段的分布



在所有的企业发展阶段当中，天使轮的薪资待遇明显低于其他阶段的企业，从区间来看A轮-上市阶段的企业薪资基本相同，考虑中位数则D轮-上市阶段的企业给出的薪资待遇最高。

Confidence(发展情况-->薪资)=66.8%

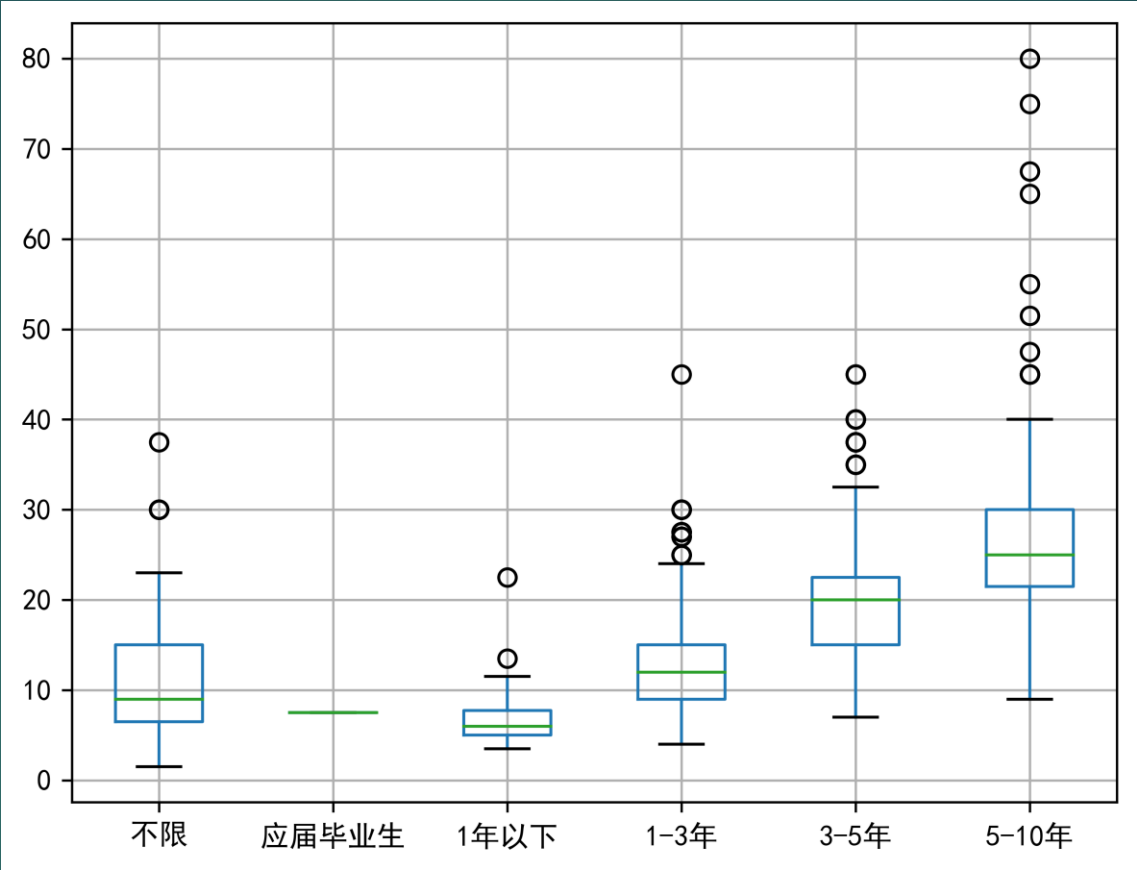
图表分析——薪资在不同学历的分布



数据分析职位的薪资随着学历的提高，平均薪资水平明显有着上升，但是从完整区间来看，拥有本科学历已经不会限制你在数据分析行业的发展。

Confidence(学历-->薪资)=66.7%

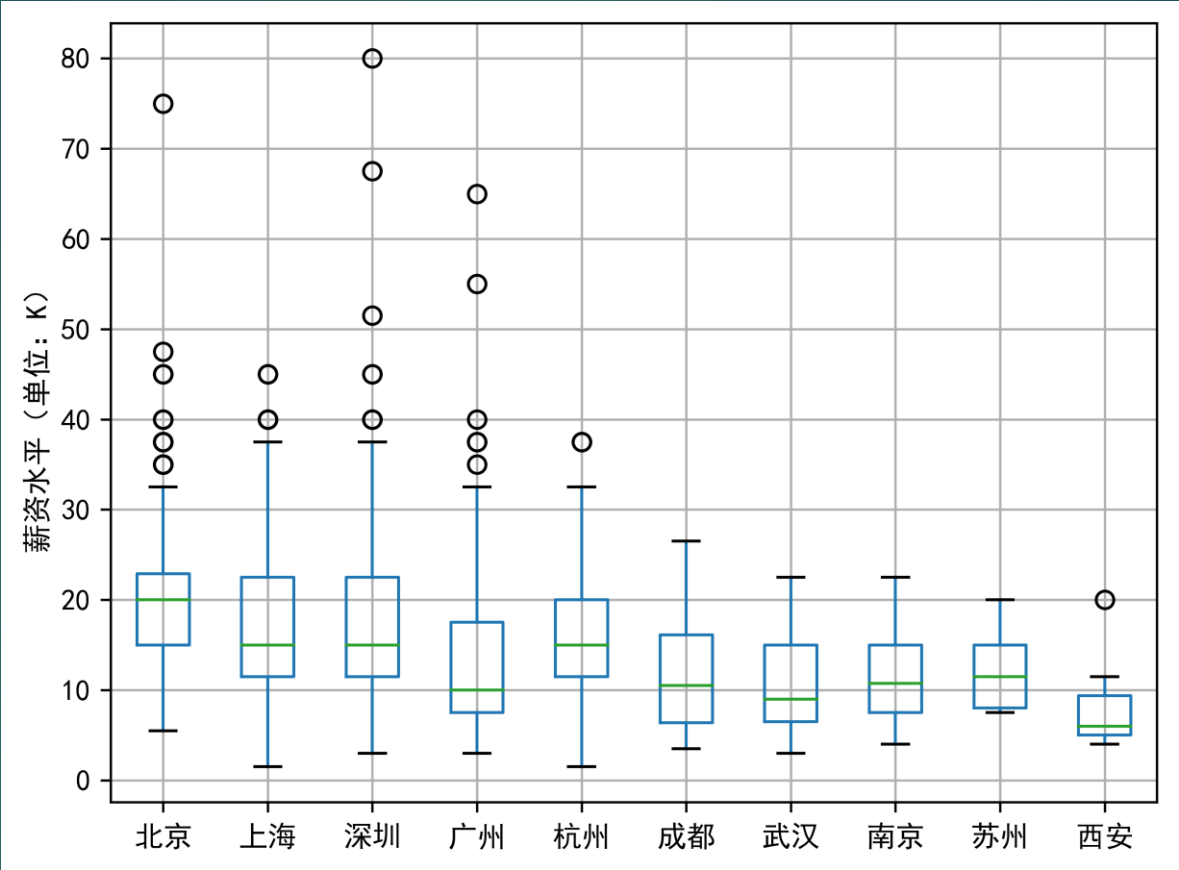
图表分析——薪资在不同工作年限的分布



随着工作年限的提高，薪资待遇呈现非常明显的线性增加，可以看出在数据分析这一行业，工作经验的积累程度对薪资来说是十分重要的影响因素。

Confidence(工作年限-->薪资)=67.1%

图表分析——薪资在不同城市的分布



北京、上海和杭州的平均薪酬要远高于其他城市，从整体区间来看广州同样处于领先地位，而在这其中北京的平均薪酬最高；

图表分析——词云



职位诱惑词云

词云表明企业用来吸引员工的除了薪资，更多的是五险一金、发展、平台、弹性工作等优势。



职位标签词云

企业提供的职位标签显示数据分析职位工作更多为遇到SPSS、金融、业务、运营、建模、数据管理等工作

结论

➤ 总结

1. 北上深杭的薪资待遇要高于其他城市，它们的需求量也明显高于其他城市；
2. 移动互联网、金融行业对于数据分析职位的需求程度要高于其他行业；
3. 平均薪资随着学历的增加而增加，但绝大多数岗位只要求本科学历；
4. 平均薪资随着工作年限的增加，同样也有明显呈线性地提高，大多数职位要求1-3年或3-5年的工作经验，相比于学历、城市、企业融资情况，工作年限对薪酬的影响更大；
5. D轮融资阶段和上市公司给出的薪酬中位数最高，天使轮最低；
6. 接近90%的企业给出的薪资水平在5K-25K，10-15K之间的数量最多；
7. 大部分企业吸引员工的内容是五险一金、发展空间和弹性工作制。

结论

➤ 不足

1. 本次数据的获取仅来自与拉勾网，数据量与数据类型均有所局限性，因此本报告不能完全代表全国所有企业的数据分析职位情况，仅作为数据参考；
2. 本篇报告是笔者学习Python数据分析之后，所进行的首个数据分析尝试，报告中参考了互联网中一些他人的分析思路，不足之处敬请谅解。

附录

- 数据处理及分析工具：

数据获取——Python Scrapy爬虫、MySQL数据库

数据清洗——Python

数据整理——Python pandas、Excel

数据建模——Apriori算法

数据可视化——Python matplotlib、wordcloud

- 原始数据



原始职位数据