数据职业路在何方？

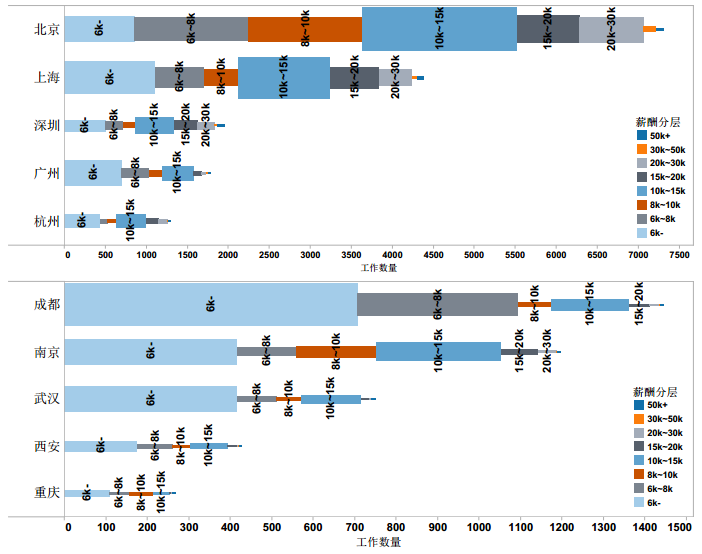
为了分析现有数据行业的就业情况，抓取前程无忧2016年9月20日~2016年10月10日的数据相关招聘职位145614条信息，并从中筛选出职位是数据分析、数据挖掘、文本挖掘、视觉学习、计算广告、数据开发、数据架构师等的招聘信息10576信息。

对比岗位数量在城市、薪酬分层、企业属性、职位要求等维度的分布情况，发现不同薪酬分层下的岗位数量，基本与城市发展趋势、行业汇聚、人才聚集呈现正比；一线城市仍然是岗位的集中地，但是二线城市中杭州、成都、南京也是可以选择的；企业属性上来看，岗位主要集中在民企，合资企业、和上市公司三大类；企业的规模上来比较，100~500之间的企业，成为主要的岗位提供者；行业分类上，岗位主要有大部分有电子商务、金融、互联网公司提供，同时，媒体行业、生物医药也具有一定需求。

薪酬档次与数据技术呈现紧密联系，基本可以分为三个层次：传统数据技术、hadoop分布式计算、spark全栈式计算。数据行业选择一项合适的技术，要一定程度上优于经验的积累。

一、城市分布

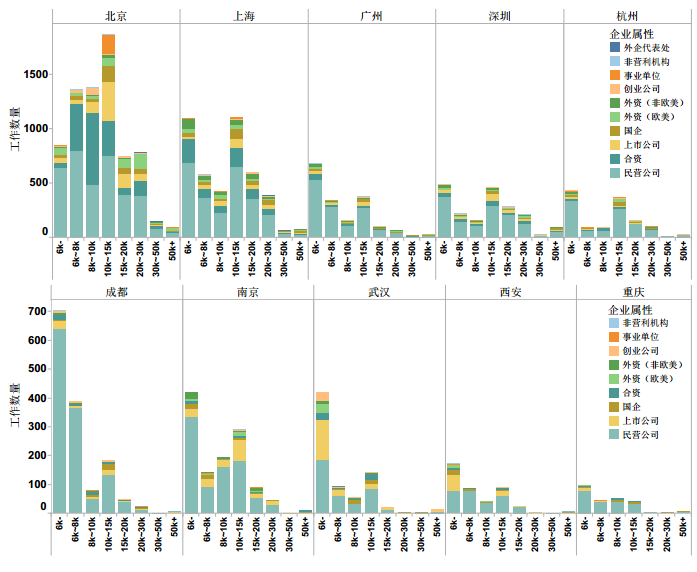
对比岗位数量在10个城市上分布（注：杭州归为一线城市），总体上岗位数量在各个城市的分布基本符合Zipf法则，一线城市中10k~15k占据主要部分，二线城市中主要以6k-为主。北京是政治中心，上海是金融中心，15k+的岗位数量与广深相比，占比相对较高。杭州拥有阿里、网易等互联网公司，电子商务居全国前列，与一线城市有一比的实力可期。二线城市中，以成都、南京位居前两位。



二、公司情况分析

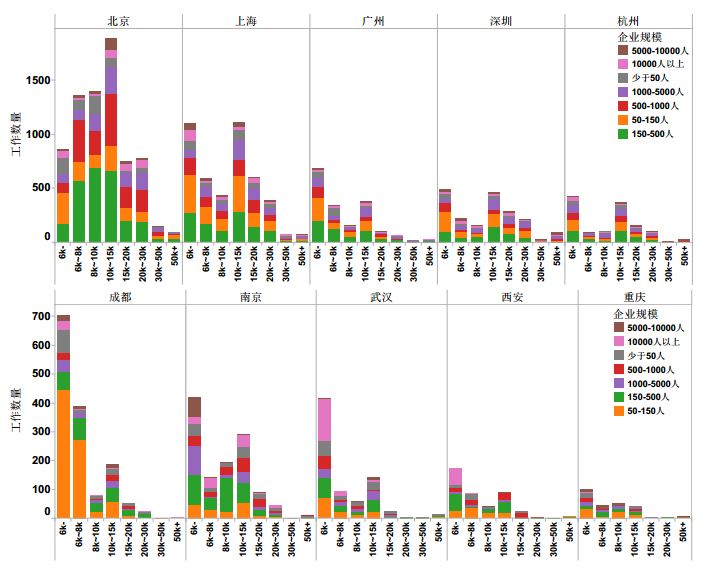
1. 企业属性

企业属性对比分析，一线城市和二线城市民营公司的占比都位列第一，所占比例均超过50%，二线城市比重更高。一线城市位居第二、三位的是合资企业和上市公司，二线城市则是上市公司国企。总体上来说，民营企业占据半边天。



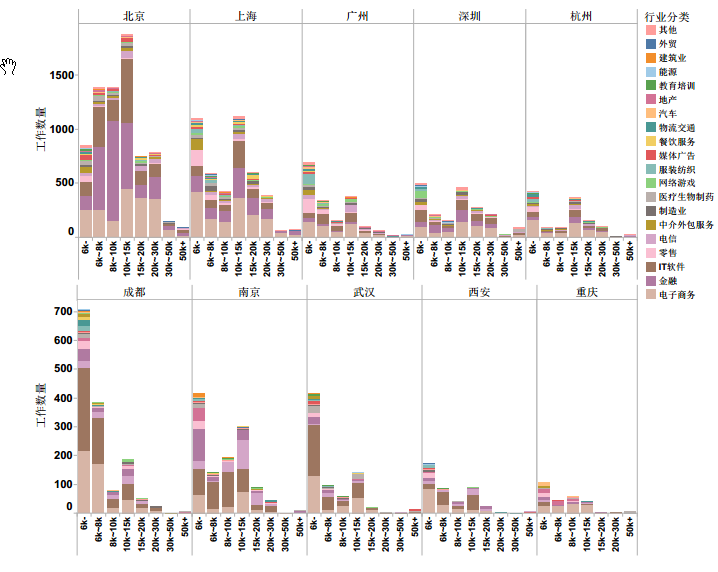
2. 企业规模

企业规模上分析，一线城市位居第一位的是150-500人，第二位是50-150人；二线城市，50-150人占比最大，150-500人次之，二线城市主要是初创企业为主。总体来说，无论一线还是二线城市，小于50人的企业，对数据职位的需求不大，说明数据分析只有在一定规模的企业，才能凸显重要性。



3. 行业分类

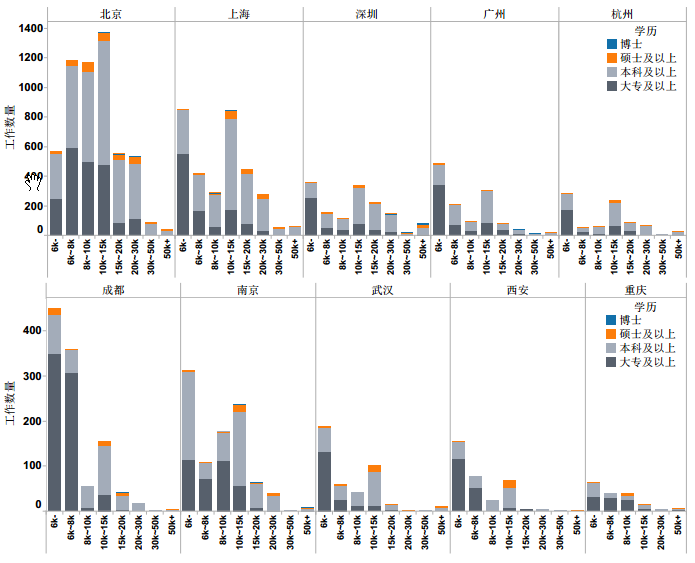
行业分类的角度来看，一、二线城市的区别不明显，占比位于前3位的分别是：电子商务、金融、IT软件，可以看出，数据职位主要集中在轻资产企业，如电子商务、IT软件，和基础数据相对完备的企业，如金融。



三、职位要求

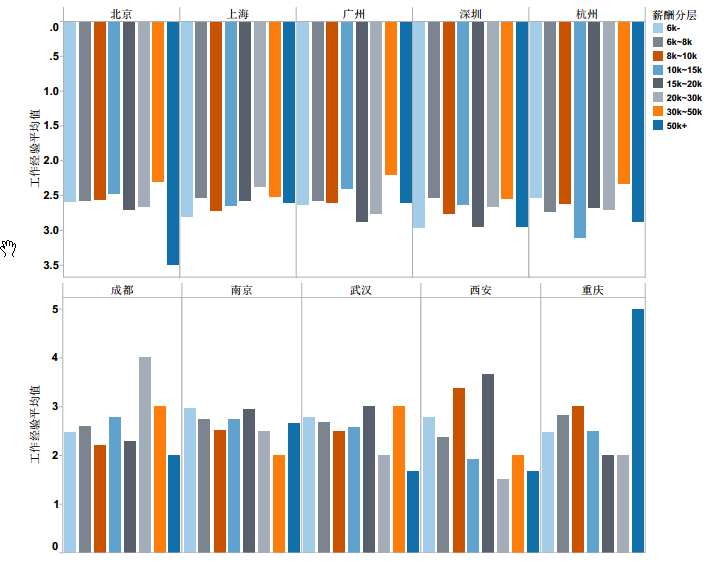
1. 学历要求

数据职位的学历要求来看，主要以大专、本科为主，大专及以上学历要求的占比随着薪酬的增加显著减少。

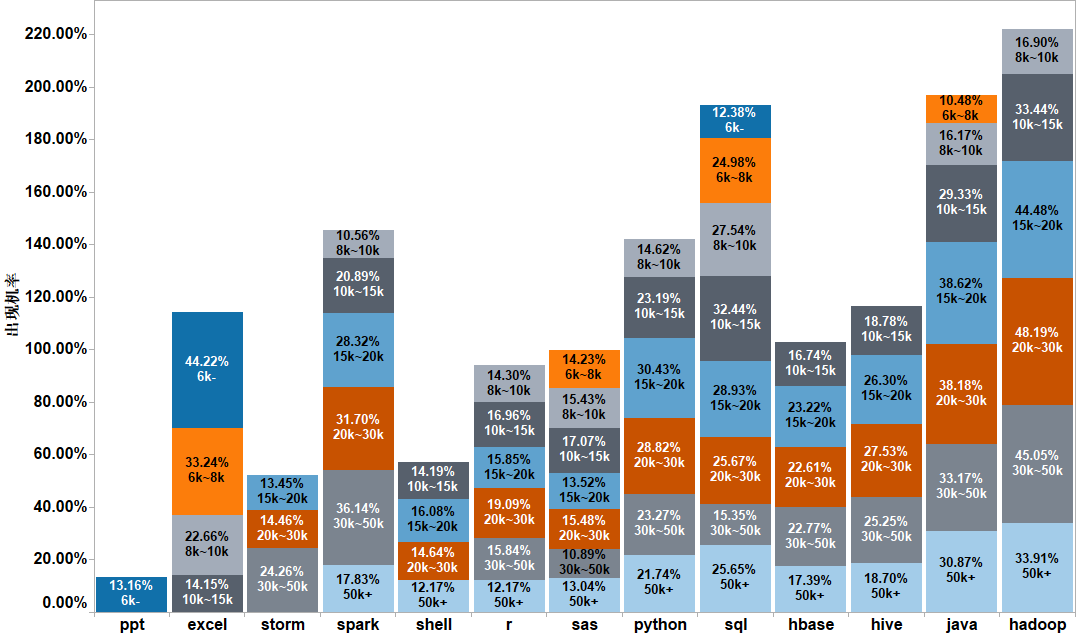


2. 工作经验

对比数据分析职位的经验要求，可以看出薪酬分层之间的差别不是很大，基本在3年左右，总体上，二线城市的工作经验要求略高于一线城市，侧面反映二线城市对于数据行业的经验更看重。



3. 工作技能



薪酬分层中出现的百分比机率（>10%）

横向对比分析，工作技能在各个薪酬分层中出现的比例可以看出，技术方面主要分为3个阶段。

第一个阶段：数据仓库+分析工具，主要通过建立数据仓库，以sql来访问数据库，然后使用分析工具python、sas、r来进行数据分析和数据挖掘。从上图可以看出，这阶段的技术在各个薪酬分层占比位于第二位，其中sql占比最大，而后排名是python、sas和r。

第二个阶段： hadoop（java）数据处理方案，核心设计是HDFS为存储方案，MapReduce为计算框架，其中hive为结构化查询工具、hbase为列式数据库，从上图可以明显看出，hadoop系列数据产品，总体在薪酬分层占比最高。

第三个阶段： spark（scala）的全栈式数据解决方案，为统一解决结构化查询、流计算、图计算和机器学习提供可能。由上图可以看出，spark、storm（流式数据处理）占据第三位置。